

Article

Modeling the Effect of the Spatial Pattern of Airborne Lidar Returns on the Prediction and the Uncertainty of Timber Merchantable Volume

Sarah Yoga ^{1,*}, Jean Bégin ¹ , Benoît St-Onge ² and Martin Riopel ¹

¹ Department of Wood and Forest Sciences, Université Laval, Quebec City, QC G1V 0A6, Canada; jean.begin@sbf.ulaval.ca (J.B.); martin.riopel.1@ulaval.ca (M.R.)

² Department of Geography, Université du Québec à Montréal, Montréal, QC H3C 3P8, Canada; st-onge.benoit@uqam.ca

* Correspondence: sarah.yoga-bengbate.1@ulaval.ca

Academic Editors: Lars T. Waser and Prasad S. Thenkabail

Received: 1 June 2017; Accepted: 1 August 2017; Published: 6 August 2017

Abstract: Lidar data are regularly used to characterize forest structures. In this study, we determine the effects of three lidar attributes (density, spacing, scanning angle) on the accuracy and the uncertainty of timber merchantable volume estimates of balsam fir stands (*Abies balsamea* (L.) Mill.) in eastern Canada. We used lidar point clouds to compute predictor variables of the merchantable volume in a nonlinear model. The best model included the mean height of first returns, the proportion of first returns below 2 m and the canopy surface roughness index. Our analysis shows a high correlation between lidar and field data of 119 plots (pseudo- $R^2 = 0.91$), however, residuals were heteroscedastic. More precise parameter estimates were obtained by adding to the model a variance function of variables describing the mean height of returns and the skewness of the area distribution of triangulated lidar returns. The residual standard deviation was better estimated ($3.7 \text{ m}^3 \text{ ha}^{-1}$ multiplied by the variance function versus $28.0 \text{ m}^3 \text{ ha}^{-1}$). We found no effect of density on the predictions (p -value = 0.74). This suggests that the height and the spatial pattern of returns, rather than the density, should be considered to better assess the uncertainty of merchantable volume estimates.

Keywords: lidar; timber merchantable volume; spatial distribution; return density; return spacing; scanning angle; residual variance

1. Introduction

Airborne Laser Scanning (ALS or lidar) has proven to be a useful 3d tool for the measurement of forest attributes over the past decades. It generates point clouds which are a three-dimensional representation of the volumetric interaction between pulse photons and illuminated objects. The spatial distribution of returns within the point cloud depends on the characteristics of the lidar systems and the target object [1,2]. The scan mechanisms (e.g., oscillating mirror, rotating mirror, nutating mirror) mounted on lidar scanners produce different scanning patterns of pulses on the ground [1]. For example, the bi-directional scan mechanism of oscillating mirrors produces a see-saw scanning pattern where pulses tend to be more homogeneously distributed along the center of the flight line than at the borders. Consequently, an alternation of local clumps of returns and local gaps will be observed within the point cloud [3]. Scanning angle is another setting which influences the spatial distribution pattern. For example, high-scanning angles increase the distance traveled by the incident pulse. Objects located further from the lidar sensor are likely to be occluded or undetected [4], thus producing an irregular spatial distribution of returns.

The effects of the surface characteristics of scanned objects on the scanning pattern are not controlled by the user. Gatzliolis et al. [1], for example, demonstrated that higher return densities are recorded in forest stands compared to green pastures under similar lidar settings. Korpela et al. [2] noted that leaf size and orientation and foliage density affect the intensity of lidar returns. Balsa-Barreiro et al. [5] noted that topography and land cover influence the density and spacing of returns.

Metrics (e.g., height metrics, canopy cover metrics), derived from point clouds are often used to model forest attributes such as volume [6,7]. Hence, highly correlated models between lidar data and field volume have been achieved in many studies [8–11]. Studies on lidar modeling focus on investigating the optimal explanatory variables that best explain the variability of the forest attributes. The relationship between the explanatory and dependent variable is obtained by means of a regression model. Least square procedures are commonly used to estimate the parameters of the regression model [12]. These procedures focus on minimizing the sum of squared residuals. Heteroscedasticity occurs in a model when the variance of the residuals is not constant [12,13]. Some possible causes of heteroscedasticity can be measurement errors of the data, autocorrelation or misspecification of the model. This can lead to a significant loss of precision where the estimated parameters of the model are inefficient, although unbiased. The standard errors and confidence intervals are also unreliable. In other words, the forest attribute can still be predicted, but the predictions are uncertain. In an operational context, uncertain predictions of a forest attribute such as the merchantable volume have a direct effect on the evaluation of its economic value [14]. Nonetheless, heteroscedasticity may be corrected using different methods [12,15,16]. Variable transformation for example modifies the measurement scale of variables in a model and heteroscedasticity may then be corrected [12,17]. Variance modeling identifies the sources of heteroscedasticity in a model and attempts to better estimate the model uncertainty through a variance function [15].

Few studies can be found on the effects of lidar on the uncertainty of predictive models (e.g., [4,18,19]). Yet, with the increasing use of lidar in forestry, reliable estimates of forest attributes derived from lidar are necessary. There is, therefore, a need to investigate whether factors like the spatial pattern of returns can also influence the predictions or the uncertainty of modeled forest attributes. This study, therefore, aims to analyze the effects of three lidar attributes (return density, return spacing, scanning angle) on the predictions and the uncertainty of merchantable volume estimates.

2. Materials and Methods

2.1. Study Area

The study was conducted at the Montmorency Forest which is a teaching and research forest facility of the Université Laval. The forest is located in central Quebec, Canada (47.3° N, 71.1° W), about 70 km north of Quebec City (Figure 1). The study site covered an area of 66 km². Elevation at the site ranged from approximately 460 to 1040 m above sea level. The average annual precipitation was 1589 mm, and the mean annual temperature was 0.3 °C [20]. The site lies on the Laurentian Plateau and is part of the boreal forest domain. Balsam fir (*Abies balsamea* [L.] Miller), black spruce (*Picea mariana* [Miller] BSP) and white spruce (*Pinus glauca* [Moench] Voss) are the dominant conifers found in the forest. White birch (*Betula papyrifera* Marshall) and trembling aspen (*Populus tremuloides*) are also common [21].

2.2. Field Plot Data

The Montmorency Forest has a network of circular permanent plots (radius = 11.28 m, area = 400 m²). The plots center positions have been georeferenced using a dual-frequency antenna GPS (Trimble Yuma, accuracy ~30 cm) and were not post-processed. Plots are re-measured on a five-year cycle. In this study, we used field data that were collected between 2007 and 2014. A sample of 119 plots was selected following these criteria: the dominant species was balsam fir, the last cut was before 1996 and the minimum dominant height was 7 m. The merchantable trees with a diameter at

breast height (dbh) above 9 cm were measured using a diameter tape. Tree height was measured on eight sample trees per plot using a Vertex hypsometer. A height-diameter relationship was adjusted for each plot (nonlinear mixed effects model) and then used to calculate the height of the other trees. The merchantable volume MV was computed using height and diameter based equations as described in [22]. The total MV was computed as the sum of MVs for all trees. Twenty-four field plots had been initially inventoried in 2011, the same year as the lidar survey. Attributes of the other plots were estimated by interpolating field measurements between two dates of inventory. Table 1 presents a summary of the field data for 2011.

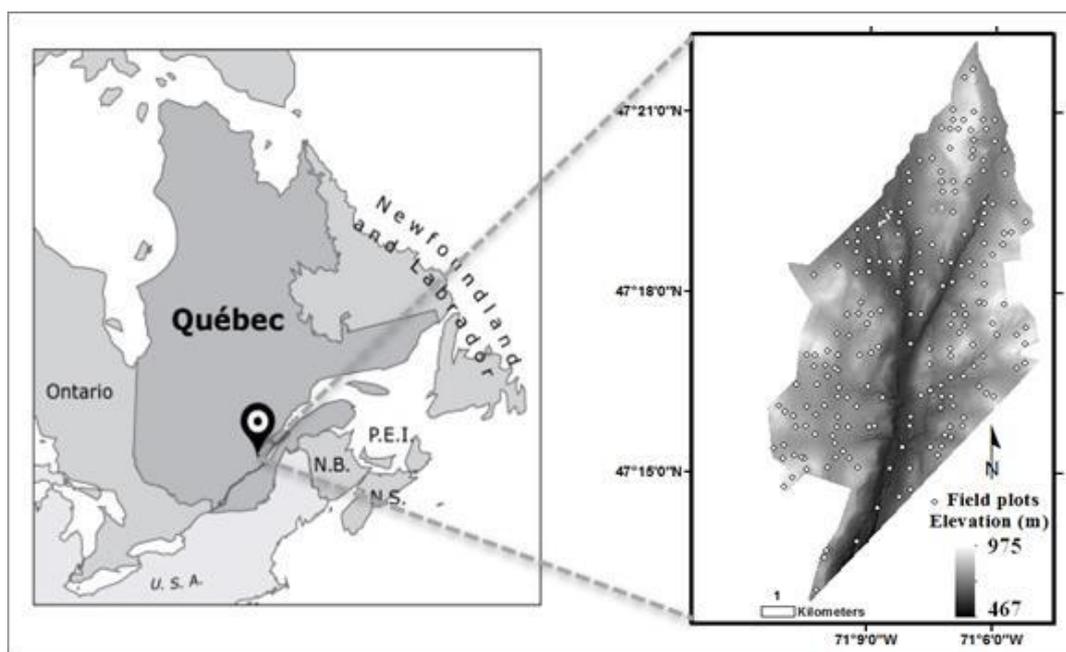


Figure 1. Locations of the Montmorency Research Forest study site and the field plots over a lidar DTM.

Table 1. Summary of field plot measurements for the Montmorency Forest study site ^a.

Field Attributes	Mean	Std.	Range
Diameter at breast height (cm)	21.9	5.2	13.3–38.0
Dominant height (m)	15.1	4.2	9.2–26.6
Density (trees ha ⁻¹)	1433.0	536.0	117.0–3078.0
Basal area (m ² ha ⁻¹)	25.8	10.9	3.7–49.5
Merchantable volume (m ³ ha ⁻¹)	144.1	90.2	21.0–411.9

^a Std. = standard deviation.

2.3. Lidar Data

Airborne laser data were acquired in August 2011 with an ALTM 3100 discrete return sensor (Teledyne Optech, Vaughan, ON, Canada). The average flying altitude was 1000 m above ground. The pulse repetition rate was of 100 kHz with a 50% strip overlap. The scanning angle ranged from 0° to 24° relative to the nadir. The sensor could record up to four measurements per pulse. Each field plot was covered by two to four lidar strips. The average above-ground return density was 6.4 m⁻². Information for each return was recorded in las files (version 1.0). The returns classified as ground (by the vendor) were interpolated to construct a digital terrain model using the LAStools software [6]. Non-ground returns were normalized by measuring the elevation difference with the underlying terrain model. Table 2 presents a summary of the lidar data of the study site.

Table 2. Summary of the lidar data for the Montmorency Forest study site (119 plots) ^a.

Lidar Attributes	Mean	Std.	Range
Mean_h (m)	5.9	2.2	1.7–11.5
Above-ground density (m ⁻²)	6.4	1.8	3.2–12.5
Above-ground spacing (dm)	4	1	3–6
Scanning angle (°)	9.1	4.7	0.0–24.0
Skew_area	1.4	0.5	0.4–3.9
Prop_2m (%)	15.3	13.9	0.2–63.5
Ri	2.7	0.7	1.6–5.2

^a Std. = standard deviation, Mean_h = mean height of first returns, Skew_area = skewness of the area distribution of triangulated returns, Prop_2m = proportion of first returns below 2 m, Ri = rumple index.

2.4. Data Analysis

The data analysis is organized in six sections: (Section 2.4.1) describes the nonlinear generalized least squares regression; (Section 2.4.2) provides an analysis of the spatial distribution of returns using an example field plot; (Section 2.4.3) calculates the lidar variables from the point clouds; (Section 2.4.4) describes the merchantable volume modelling; (Section 2.4.5) analyses the residual variance; and (Section 2.4.6) provides the model validation.

2.4.1. Generalized Least Squares in Nonlinear Regression

Heteroscedasticity occurs in a nonlinear regression (NLIN-GLS) model when the error variance is not constant over all the observations [15]:

$$y_i = f(x_i; \beta) + \varepsilon_i \text{ with } \text{var}(\varepsilon_i) = \sigma^2 g^2(z_i; \delta) \quad (1)$$

where y_i is the value of the response variable measured on the i th experimental unit, and is expressed as a known function $f(\cdot)$ of some explanatory variables (x_i) and parameters (β) plus a random error term (ε_i). In this model, the heteroscedasticity is considered by the fact that the variance σ^2 of this latter term is not constant, but depends on a known function $g(\cdot)$ of some covariates z_i and parameters δ .

Equation (1) was used to predict the merchantable volume of our study site. The variance functions $g(\cdot)$ that we have considered had one of the following expressions [15]:

$$g(z_i; \delta) = \begin{cases} |z_{1i}|^\delta \\ \exp(\delta z_{1i}) \\ |z_{1i}|^{\delta_1} |z_{2i}|^{\delta_2} \\ \exp(\delta_1 z_{1i}) \exp(\delta_2 z_{2i}) \\ |z_{1i}|^{\delta_1} \exp(\delta_2 z_{2i}) \end{cases} \quad (2)$$

The first variance function is a power function of the absolute value of the covariate z_i , the second one is an exponential function of this covariate, and the others are various combinations of the previous ones.

The SP group of covariates, which will be described later, have been included in this model either in the fixed part through the $f(\cdot)$ function or in the random part through the $g(\cdot)$ function.

2.4.2. Analysis of the Spatial Distribution of Returns

The return density, return spacing and scanning angle play an important role in the spatial distribution of returns. To provide a visual example of this, we have extracted a sample of the 2d distribution of all the returns of a given field plot that had been covered by three lidar strips (#8, #9 and #10) during the survey (Figure 2). The strips had an average return density of 2.15, 2.19 and 2.04 m⁻² respectively, an average return spacing of 6.8, 6.8 and 7.0 dm and an average scanning angle of 16°, 1.6°, and 9.4°. The returns were plotted in a 2d geographic grid and a line segment was drawn between

consecutive returns to show the scanning sequence. Strips #8 and #10 had a heterogeneous spatial distribution of returns. Local clumps of returns alternated with local gaps. Conversely, strip #9 had a more homogeneous spatial distribution of returns throughout the plot. When the three strips were pooled together at the plot level, the return spacing drastically decreased (4.0 dm). However, local gaps were still observable within the plot.

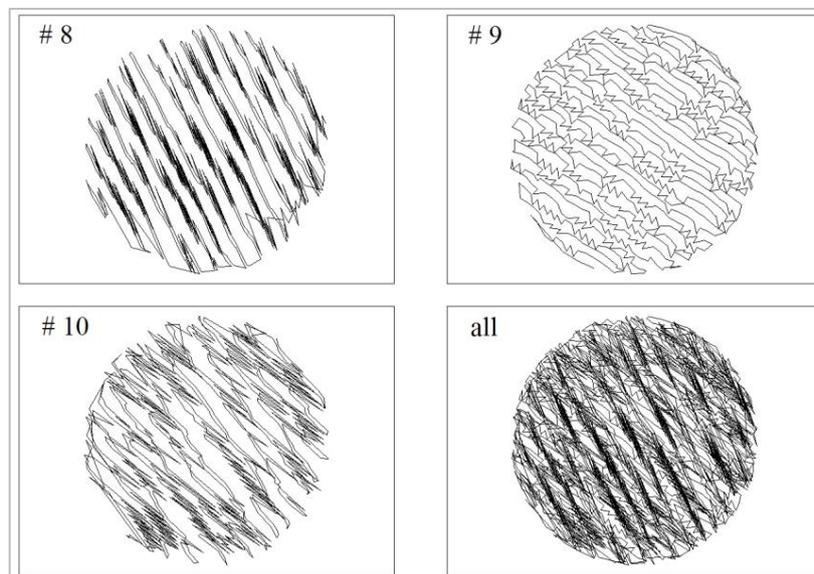


Figure 2. Spatial distribution pattern of returns for the field plot #28 surveyed by three lidar strips. Dark areas depict zones of clumps of returns and white areas depict gap zones where there are no returns.

We triangulated neighboring returns into a 3d network to analyze the spacing distribution. This was done using the Delaunay triangulation algorithm of ArcGIS 10.3 (® ESRI). The area of each triangle was calculated. Figure 3 shows the area distribution of triangulated returns for the previous example field plot. Strips #8 and #10 which had a heterogeneous pattern (see Figure 2) produced an asymmetric right-skewed distribution. The areas ranged from 0.2 to 211.44 dm² and 0.68 to 217.29 dm² respectively. Conversely, strip #9, which had a homogeneous pattern, produced a symmetric leptokurtic distribution. The areas ranged from 0.14 to 67.12 dm². The heterogeneous pattern observed, when pooling all strips together, induced an asymmetric right-skewed mesokurtic distribution. The number of small triangles increased. The areas ranged from 0.06 to 53.14 dm². This example shows that an irregular distribution of returns persisted locally even with three overlapping strips. Strips were always pooled together in the following analyses.

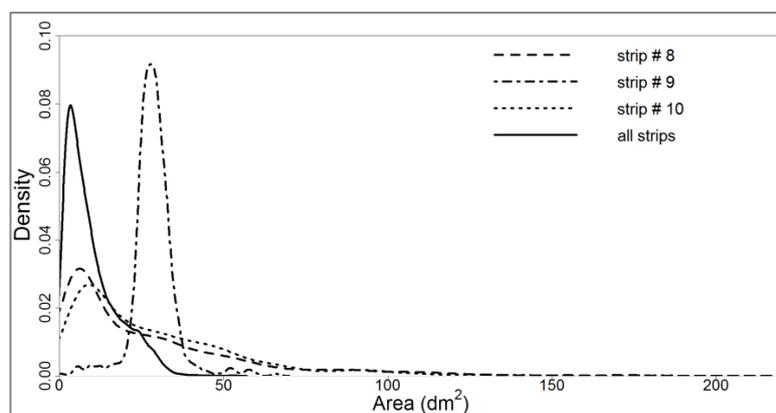


Figure 3. Area distribution of triangulated returns for the field plot #28 surveyed by three lidar strips.

2.4.3. Lidar Variable Generation

Lidar returns were clipped to the extent of each plot spatial boundary to extract the corresponding point cloud. Metrics were computed both at the strip and the plot levels using Fusion software [7]. They were classified into three standard groups of explanatory variables: canopy height variables (percentiles of height, mean, mode), canopy density variables (proportion of first returns above or below a 2 m, 5 m or 7 m threshold) and canopy structural heterogeneity variables (standard deviation, variance, coefficient of variation, canopy surface roughness index). The canopy surface roughness index, hereafter called rumple index, was computed as the ratio of the canopy outer surface area to the ground surface area as measured by the lidar-derived canopy surface and digital terrain models in a $1\text{ m} \times 1\text{ m}$ grid [23].

A fourth group of explanatory variables, SP, was computed to describe the spatial distribution pattern of returns. Variables included in this group were: density variable (computed as the number of above ground returns of a given plot divided by its area), area distribution variables (mean, standard deviation, minimum, maximum, skewness and kurtosis of the area distribution of triangulated returns) and scanning angle variables (mean, mode, standard deviation, minimum, and maximum of the scanning angle distribution).

2.4.4. Establishing the Merchantable Volume Model

The basic merchantable volume model was built at the plot level using one variable per standard group iteratively. Nonlinear relationships were observed between the field MV and the lidar canopy density/canopy heterogeneity variables. Hence, each subset of variables was combined into the NLIN-GLS described in Equation (1):

$$MV_i = \exp(\beta_0) * H_i^{\beta_1} * CD_i^{\beta_2} * CSH_i^{\beta_3} + \varepsilon_i \quad (3)$$

where MV is the merchantable volume, H the canopy height variable, CD the canopy density variable, CSH the canopy structural heterogeneity variable and ε the residual term.

The SP variables were then added to Equation (3):

$$MV_i = \exp(\beta_0) * H_i^{\beta_1} * CD_i^{\beta_2} * CSH_i^{\beta_3} * SP_i^{\beta_4} + \varepsilon_i \quad (4)$$

The best subsets of variables were selected based on the corrected Akaike Information Criterion AICc, the residual standard deviation RSD, the level of significance of the model p -value, and the coefficient of determination pseudo- R^2 . Variable multicollinearity in the model was tested by measuring the variance inflator factor VIF with the threshold value set to 10 [24]. The statistical analysis was done with R software [25]. We used the “nlme” package [26] for the model fitting.

2.4.5. Residual Variance

The absolute residuals were regressed against the predicted MVs to verify the heteroscedasticity of the residuals. A significant p -value of the MV parameter was an indication of heteroscedasticity. The MV uncertainty was then assessed by iteratively adding an SP variable to the error variance function (random part of the model):

$$MV_i = \exp(\beta_0) * H_i^{\beta_1} * CD_i^{\beta_2} * CSH_i^{\beta_3} + \varepsilon_i \text{ with } \text{var}(\varepsilon_i) = \sigma^2 g^2(SP_i; \delta) \quad (5)$$

The regression parameters were estimated using an NLIN-GLS regression. The AICc, RSD, p -value and the 95% confidence intervals (95% CI) were computed to compare models.

2.4.6. Model Validation

The leave-one-out cross-validation was used to assess the accuracy of the merchantable volume models. A new dataset was created by removing one field plot iteratively. The MV model was fitted

to the new dataset (training data) and used to predict the MV of the removed field plot (test data). The procedure was repeated until predicted values were obtained for all the field plots.

3. Results

We found a very good correlation between field and predicted merchantable volumes for balsam fir stands even before the introduction of the SP variables (pseudo- $R^2 = 0.91$, Table 3, and Figure 4). The best predictors were the mean height of first returns (Mean_h), the proportion of first returns below 2 m (Prop_2m) and the rumple index (Ri). The VIF factor was below 10. The RSD was $28.0 \text{ m}^3 \text{ ha}^{-1}$ for the regression model and $28.98 \text{ m}^3 \text{ ha}^{-1}$ after cross-validation. The best equation was:

$$MV = \exp(2.04) * \text{Mean_h}^{1.33} * \text{Prop_2m}^{-0.08} * \text{Ri}^{0.64} + \epsilon \tag{6}$$

Table 3. Parameters of the merchantable volume models before and after the addition of the variance function ^a.

	Variables	Before (no Variance Function)				After (with Variance Function)			
		Value	Std. Err.	p-Value	95% CI	Value	Std. Err.	p-Value	95% CI
Parameters of the fixed part of the model									
β_0	Intercept	2.04	0.17	<0.001	[1.77–2.42]	1.97	0.16	<0.001	[1.76–2.34]
β_1	Mean_h	1.33	0.10	<0.001	[1.14–1.52]	1.33	0.11	<0.001	[1.12–1.55]
β_2	Prop_2m	−0.08	0.03	0.020	[−0.13–−0.02]	−0.07	0.03	0.020	[−0.07–−0.02]
β_3	Ri	0.64	0.12	<0.001	[0.40–0.88]	0.71	0.13	<0.001	[0.64–0.87]
Parameters of the variance function									
δ_1	Mean_h	—	—	—	—	0.85		<0.001	[0.50–1.20]
δ_2	Skew_area	—	—	—	—	0.31		<0.001	[0.05–0.57]

^a Variance function = $\text{Mean_h}^{\delta_1} * \exp(\delta_2 * \text{Skew_area})$. Mean_h = mean height of first returns, Prop_2m = proportion of first returns below 2 m, Ri = rumple index, Skew_area = skewness of the area distribution of triangulated returns, Std. Err. = standard error.

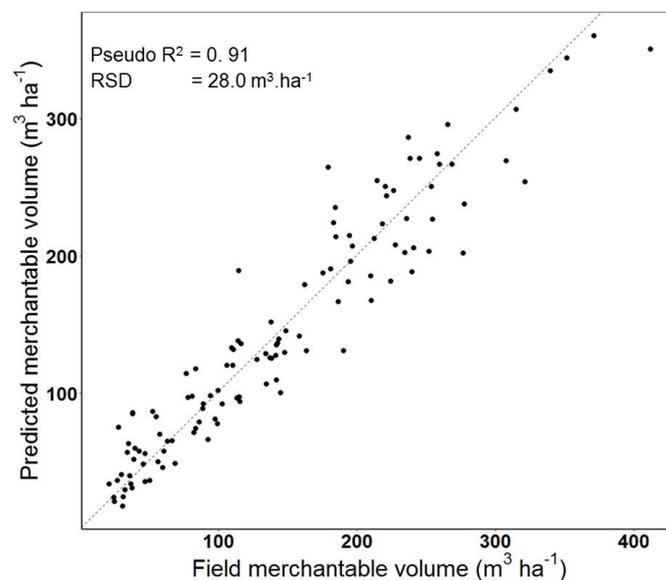


Figure 4. Scatter plot of predicted merchantable volume versus field merchantable volume (observed) of plots.

Figure 5 shows plots of the model residuals versus the predicted MV, Mean_h and Skew_area variables. Residuals had an outward-opening funnel form: they were more variable as the predicted

MVs increased. The MV parameter was significant ($p = 0.02$) when regressing the absolute residuals against the predicted MVs. This indicated the presence of heteroscedasticity and needed to be corrected.

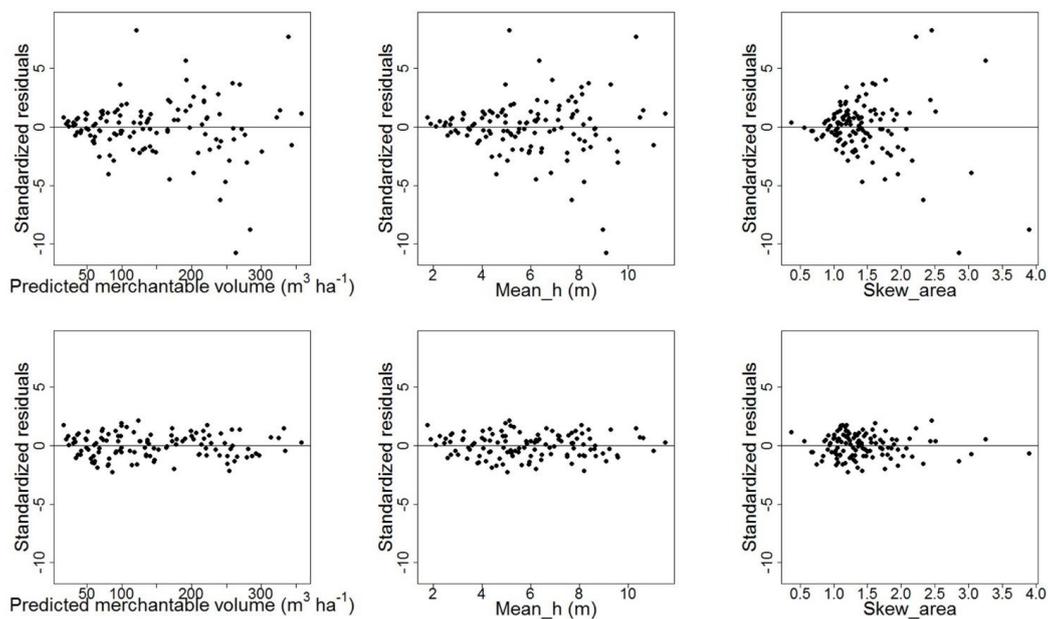


Figure 5. Scatter plots of standardized model residuals versus predicted merchantable volume (MV), observed mean heights of first returns (Mean_h) and skewness of the area distribution of triangulated returns (Skew_area) of plots. The upper scatter plots show the relationship before addition of the variance function to the model (see Equation (3)) and the lower scatter plots, after the addition (see Equation (5)).

The spatial distribution of returns barely influenced the predicted MV. We found no significant improvement when adding an SP predictor to the fixed part of the MV model. Table 4 shows a model comparison between the basic MV model and 3 MV models built with an additional SP variable:

$$MV = \exp (2.01) * \text{Mean_h}^{1.34} * \text{Prop_2m}^{-0.08} * \text{Ri}^{0.68} * \text{Skew_area}^{-0.06} + \epsilon \quad (7)$$

$$MV = \exp (2.25) * \text{Mean_h}^{1.32} * \text{Prop_2m}^{-0.08} * \text{Ri}^{0.66} * \text{Mean_angle}^{-0.09} + \epsilon \quad (8)$$

$$MV = \exp (2.00) * \text{Mean_h}^{1.34} * \text{Prop_2m}^{-0.08} * \text{Ri}^{0.63} * \text{Density}^{0.02} + \epsilon \quad (9)$$

The additional SP variables did not improve the basic model (AICc = 1137.79, 1137.48 and 1139.32 respectively for Equations (7) – (9) versus 1137.23 for Equation (6)). The density variable (Equation (9)) was the least significant ($p = 0.74$). RSDs were however similar for all models.

Table 4. Comparison of four predictive merchantable volume MV models before (basic model) and after addition of a spatial distribution variable ^a.

MV Model	AICc	p-Value	Pseudo R ²	RSD
Basic model	1137.23	<0.001	0.91	28.0
Basic model + Density	1139.32	0.74 *	0.91	28.0
Basic model + Mean_angle	1137.79	0.21 *	0.91	28.0
Basic model + Skew_area	1137.48	0.17 *	0.91	27.9

^a Basic model = $\exp (2.04) * \text{Mean_h}^{1.33} * \text{Prop_2m}^{-0.08} * \text{Ri}^{0.64} + \epsilon$ Mean_h = mean height of first returns, Prop_2m = proportion of first returns below 2 m, Ri = rumple index, Density = return density, Mean_angle = mean scanning angle of returns, Skew_area = skewness of the area distribution of triangulated returns, RSD = Residual standard deviation. * The shown value represents the level of significance of the additional variable.

Table 5 shows a comparison between MV models before (basic model) and after addition of a variance function (corrected models). Adding SP variables to the random part of the MV model improved the precision. The corrected MV models had lower AICc value (1110.32, 1123.58 and 1106.72 versus 1137.23 for the basic model). Mean_h and Skew_area had the most effect on the unexplained variance as residuals were more variable when these variables increased (Figure 5, Table 6). The MV uncertainty was better assessed by adding the following variance function to Equation (6):

$$\text{var}(\varepsilon_i) = \sigma^2 * \left[\text{Mean}_h^{0.85} * \exp(0.31 * \text{Skew_area}) \right]^2 \quad (10)$$

The RSD, which was 28.0 m³ ha⁻¹ initially, was estimated to 3.7 m³ ha⁻¹ multiplied by the variance function, when considering the heteroscedasticity of residuals.

Table 5. Comparison of the merchantable volume MV models before (basic model) and after addition of a variance covariate ^a.

MV Model	AICc	p-Value	RSD
Basic model	1137.23	<0.001	28.0
Basic model + Mean_h variance covariate	1110.32	<0.001 *	9.0
Basic model + Skew_area variance covariate	1123.58	<0.001 *	13.5
Basic model + Mean_h + Skew_area variance covariates	1106.72	<0.001 *	3.7

^a Basic model = $\exp(2.04) * \text{Mean}_h^{1.33} * \text{Prop_2m}^{-0.08} * \text{Ri}^{0.64} + \varepsilon$ Mean_h = mean height of first returns, Prop_2m = proportion of first returns below 2 m, Ri = rumple index, Skew_area = skewness of the area distribution of triangulated returns, RSD = Residual standard deviation. * The shown value represents the level of significance of the additional variance function.

Table 6. Effects of Mean_h and Skew_area variables on the variance of predicted merchantable volume ^a.

	Skew_Area		Mean_h (m)				
	—	2.0	4.0	6.0	8.0	10.0	12.0
RSD (m ³ ha ⁻¹)	0.5	7.8	14.0	19.8	25.3	30.6	35.7
	1.0	9.1	16.4	23.1	29.5	35.7	41.7
	1.5	10.6	19.1	27.0	34.5	41.7	48.7
	2.0	12.4	22.3	31.5	40.3	48.7	56.9
	2.5	14.5	26.1	36.8	47.0	56.9	66.4
	3.0	16.9	30.5	43.0	54.9	66.4	77.5
	3.5	19.7	35.6	50.2	64.1	77.5	90.5
	4.0	23.0	41.5	58.6	74.9	90.5	105.7

^a RSD = Residual standard deviation.

4. Discussion

This study aimed to develop an efficient lidar model to predict merchantable volume in balsam fir stands and to better assess the model uncertainty. As uncertainty increases, the model parameters can still be estimated but the confidence intervals are unreliable [12,27]. Uncertain predictions induce measurement errors in the volume of stands and consequently on the evaluation of their economic value [14]. The uncertainty of merchantable volume should, therefore, be correctly estimated.

As expected, highly significant model for predicting the MV could be developed using standard lidar metrics (pseudo-R² = 0.91, RSD = 28.0 m³ ha⁻¹). This is consistent with other studies indicating that lidar variables can accurately predict forest structure attributes in the eastern Canadian boreal forest [11,28,29]. The optimal explanatory variables of the model were the mean height of first returns, the proportion of first returns below 2 m and the rumple index. Mean height is a good predictor of volume as shown in many studies (e.g.: [8,11]). The variable was positively correlated to the MV in the model as shown in Table 3.

The proportion of first returns below 2 m variable could be correlated with the fraction of canopy gaps. Canopy gaps are small openings of stands where there are smaller or no trees [30]. They are often found in mature stands, and reduce the overall stand volume. Several plots of our study site were within mature forest stands. The variable was negatively correlated to the MV in our model as shown in Table 3.

The rumple index measures the canopy surface roughness or structural complexity. It has also been defined as a 3d measure of canopy heterogeneity [23]. It can be used to characterize multi-cohorts stands such as those present in our study site. Mature trees, which have bigger volume, are more likely to be present in multi-aged stands [31,32]. The variable was positively correlated to the MV in the model as shown in Table 3.

Figure 2 shows that the spatial distribution of returns can remain irregular despite a high return density or the presence of overlapping strips. The consequences of an irregular spacing pattern are directly evident when building, for example, canopy surface models or digital elevation models [3,33,34]. In our case study, the irregular pattern induced an alternation of local clumps (overscanned areas) and local gaps (areas without data) within the point cloud. This pattern did not affect the accuracy of the model as adding an SP variable to the fixed part did not improve the predictions. However, the residuals were affected by the pattern. This shows that an irregular spatial distribution of returns can increase the uncertainty of predictive models.

The variance function helped to identify the sources of heteroscedasticity of residuals. Residuals were more heteroscedastic with increases of the mean height of returns (Mean_h) and the skewness of the area distribution of triangulated returns (Skew_area); see Figure 5 and Table 6. This suggests a higher MV uncertainty in high stands (e.g., mature forest stands) and with irregular lidar scanning patterns. Natural high stands tend to have a more heterogeneous canopy height structure compared to low stands [35]. An irregular scanning pattern over these stands could over or under characterize the height distribution substantially. Yet these stands are preferentially cut during forest operations. Their MV, therefore, need to be precisely predicted. Conversely, low stands, tend to have a more homogeneous canopy height structure. The over or under characterization of the height distribution due to the irregular scanning pattern would be less substantial.

We further tested the robustness of the corrected MV model for a low-density scenario like in some operational context (pulse density $\leq 2 \text{ m}^{-2}$). We obtained similar results to the previous analyses. The RSD was $27.4 \text{ m}^3 \text{ ha}^{-1}$ initially and was estimated to $3.7 \text{ m}^3 \text{ ha}^{-1}$ multiplied by the variance function when considering the heteroscedasticity of residuals. Mean_h and Skew_area remained the best combination of covariates for the variance function. The spacing distribution of returns should also be considered when planning lidar surveys or, at least, when analyzing acquired lidar data.

Pooling lidar strips together at the plot level increased the return density (6.4 m^{-2} on average). However, our analysis shows that the return density did not influence the prediction of the merchantable volume estimates for the study site ($p = 0.74$). This result is in accordance with other studies (e.g., [10,11,36]) done even in low return density contexts.

We tested variable transformation to correct for heteroscedasticity. The response and explanatory variables of the MV model were strictly positive (see Tables 1 and 2), and three variables (MV, Prop_2m, and Ri) had a right-skewed distribution. We thus tested a square root, a cube root, and a logarithmic transformation. The heteroscedasticity was best reduced with the logarithmic transformation. The p -value for the predicted MV parameter was improved, however, it remained significant (0.01). The model had the following form:

$$\log(\text{MV}) = 1.99 + 1.18 * \log(\text{Mean_h}) - 0.08 * \log(\text{Prop_2m}) + 0.95 * \log(\text{Ri}) + \epsilon$$

Acquiring lidar data with a small scanning angle would also be an ideal solution to obtain more homogeneous data but the acquisition costs would also be greatly increased. Other authors have chosen to exclude from the point clouds either pulses transmitted at high scanning angles [18,37] or irregularly spaced sectors within strips [3]. However, these changes would substantially alter the

point cloud structure. A statistical approach, as the one we have proposed, has the advantage of alleviating these problems while entailing no additional costs. We recommend the use of an additional spatial distribution variable when modeling forest attributes models from lidar point clouds having an irregular spatial distribution of returns.

A practical application of our study is the establishment of reliable uncertainty maps of predicted merchantable volumes. Users can then confidently assess the reliability of the estimates and consequently better plan their harvesting operations.

5. Conclusions

While lidar return density and scanning angle have little impact on the timber merchantable volume parameter estimates, an irregular spacing among returns influences a model uncertainty. Residuals of a standard merchantable volume predictive model were found to be correlated with the mean height of returns and a variable accounting for the irregularity of spatial distribution of returns. We included these variables in a variance function to correct for the heteroscedasticity. The residual standard deviation was better estimated ($3.7 \text{ m}^3 \text{ ha}^{-1}$ multiplied by a variance function versus $28.0 \text{ m}^3 \text{ ha}^{-1}$ without the variance function). Therefore, the mean height of returns and the return spatial distributions should be considered for reliable estimates of forest attributes.

Acknowledgments: The authors would like to thank the Fonds de recherche du Québec—Nature et technologies (FRQNT) and the Ministère des Forêts, de la Faune et des Parcs (MFFP) of the province of Quebec, Canada for the financial support of this research work. We would like to thank the Montmorency Forest for the access to the data used in this study. We are also grateful to the Applied Geomatics Research Group (AGRG, Nova Scotia Community College, Canada), Christopher Hopkinson (University of Lethbridge, Canada), the Canadian Consortium for Lidar Environmental Applications Research (C-CLEAR, Centre of Geographic Sciences, Canada) for providing the lidar data at a reduced cost. Special thanks to Marc-Olivier Lemonde (MFFP, Canada) for his assistance in the quality control of the lidar data and to Gaétan Daigle (Université Laval, Canada) for his advice on the statistical analyses. We are also grateful to the anonymous reviewers for their valuable comments on the manuscript.

Author Contributions: Sarah Yoga, Jean Bégin, and Benoît St-Onge proposed the theoretical framework of the study. Jean Bégin contributed statistical expertise and Benoît St-Onge, remote sensing expertise. Martin Riopel carried out the processing of the field data. Sarah Yoga conducted the processing of the lidar data and the statistical analyses. Sarah Yoga wrote the manuscript. All authors contributed to the interpretation of results and the editing of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gatzliolis, D.; Andersen, H.-E. *A Guide to LIDAR Data Acquisition and Processing for the Forests of the Pacific Northwest*; U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station: Washington, DC, USA, 2008.
2. Korpela, I.; Ørka, H.O.; Maltamo, M.; Tokola, T.; Hyypä, J. Tree species classification using airborne LiDAR—Effects of stand and tree parameters, downsizing of training set, intensity normalization, and sensor type. *Silva Fenn.* **2010**, *44*, 319–339. [[CrossRef](#)]
3. Balsa-Barreiro, J.; Pere, J.P.; Lerma, J.L. Airborne Light Detection and Ranging (LiDAR) point density analysis. *Sci. Res. Essays* **2012**, *7*, 3010–3019. [[CrossRef](#)]
4. Ehlert, D.; Heisig, M. Sources of angle-dependent errors in terrestrial laser scanner-based crop stand measurement. *Comput. Electron. Agric.* **2013**, *93*, 10–16. [[CrossRef](#)]
5. Balsa-Barreiro, J.; Lerma, J.L. Empirical study of variation in lidar point density over different land covers. *Int. J. Remote Sens.* **2014**, *35*, 3372–3383. [[CrossRef](#)]
6. Isenburg, M. *LAStools—Efficient Lidar Processing Software*, version 150526, unlicensed. Available online: <http://www.rapidlasso.com/LAStools> (accessed on 1 May 2017).
7. McGaughey, R.J. *FUSION/LDV: Software for LIDAR Data Analysis and Visualization*; version 3.42.; USDA Forest Service: Washington, DC, USA, 2014; Volume 175.

8. Bouvier, M.; Durrieu, S.; Fournier, R.A.; Renaud, J.-P. Generalizing predictive models of forest inventory attributes using an area-based approach with airborne lidar data. *Remote Sens. Environ.* **2015**, *156*, 322–334. [[CrossRef](#)]
9. Maltamo, M.; Eerikäinen, K.; Packalén, P.; Hyypä, J. Estimation of stem volume using laser scanning-based canopy height metrics. *Forestry* **2006**, *79*, 217–229. [[CrossRef](#)]
10. Næsset, E. Accuracy of forest inventory using airborne laser scanning: Evaluating the first nordic full-scale operational project. *Scand. J. For. Res.* **2004**, *19*, 554–557. [[CrossRef](#)]
11. Treitz, P.; Lim, K.; Woods, M.; Pitt, D.; Nesbitt, D.; Etheridge, D. LiDAR sampling density for forest resource inventories in Ontario, Canada. *Remote Sens.* **2012**, *4*, 830–848. [[CrossRef](#)]
12. Beal, S.L.; Sheiner, L.B. Heteroscedastic nonlinear regression. *Technometrics* **1988**, *30*, 327–338. [[CrossRef](#)]
13. Wolter, K.M. *Introduction to Variance Estimation*; Springer Science & Business Media: Berlin, Germany, 2007; pp. 384–389.
14. Holopainen, M.; Mäkinen, A.; Rasinmäki, J.; Hyytiäinen, K.; Bayazidi, S.; Vastaranta, M.; Pietilä, I. Uncertainty in forest net present value estimations. *Forests* **2010**, *1*, 177–193. [[CrossRef](#)]
15. Pinheiro, J.; Bates, D. *Mixed-Effects Models in S and S-PLUS*; Statistics and Computing Series; Springer: New York, NY, USA, 2000.
16. Worrall, J.J.; Egeland, L.; Eager, T.; Mask, R.A.; Johnson, E.W.; Kemp, P.A.; Shepperd, W.D. Rapid mortality of *Populus tremuloides* in southwestern Colorado, USA. *For. Ecol. Manag.* **2008**, *255*, 686–696. [[CrossRef](#)]
17. Ruppert, D. Transformation and weighting. In *The Work of Raymond J. Carroll*; Springer: New York, NY, USA, 2014; pp. 155–194.
18. Næsset, E. Effects of different flying altitudes on biophysical stand properties estimated from canopy height and density measured with a small-footprint airborne scanning laser. *Remote Sens. Environ.* **2004**, *91*, 243–255. [[CrossRef](#)]
19. Morsdorf, F.; Frey, O.; Meier, E.; Itten, K.I.; Allgoewer, B. Assessment of the influence of flying altitude and scan angle on biophysical vegetation products derived from airborne laser scanning. *Int. J. Remote Sens.* **2008**, *29*, 1387–1406. [[CrossRef](#)]
20. Canada Environment and Natural Resources. 1971–2000 Climate Normals & Averages. Available online: http://climate.weather.gc.ca/climate_normals/index_e.html (accessed on 1 May 2017).
21. Bélanger, L. La forêt mosaïque comme stratégie de conservation de la biodiversité de la sapinière boréale de l'Est: l'Expérience de la forêt Montmorency—Mosaic cutting as a biodiversity conservation strategy in eastern boreal balsam fir forests: The case study of the Montmorency forest. *Nat. Can.* **2001**, *125*, 18–25.
22. Fortin, M.; DeBlois, J.; Bernier, S.; Blais, G. Mise au point d'un tarif de cubage général pour les forêts québécoises: Une approche pour mieux évaluer l'incertitude associée aux prévisions—Establishing a general cubic volume table for Québec forests: An approach to better assess prediction uncertainties. *For. Chron.* **2007**, *83*, 754–765.
23. Kane, R.; McGaughey, R.J.; Bakker, J.D.; Gersonde, R.F.; Lutz, J.A.; Franklin, J.F. Comparisons between field- and Lidar-based measures of stand structural complexity. *Can. J. For. Res.* **2010**, *40*, 761–773. [[CrossRef](#)]
24. Van Breugel, P. rvif: A GRASS GIS Addon for Stepwise Variance Inflation Factor Variable Selection. Available online: https://www.researchgate.net/publication/301560324_rvif_A_GRASS_GIS_addon_for_stepwise_variance_inflation_factor_variable_selection (accessed on 1 May 2017).
25. R Core Team. R: A Language and Environment for Statistical Computing. Available online: <https://rdrr.io/cran/nlme/> (accessed on 1 May 2017).
26. Pinheiro, J.; Bates, D.; DebRoy, S.; Sarkar, D.; R Core Team. nlme: Linear and Nonlinear Mixed Effects Models, version 3.1-131. Available online: <http://www.gbif.org/resource/81287> (accessed on 1 May 2017).
27. Robinson, A.P.; Hamann, J.D. *Forest Analytics with R: An Introduction*; Springer: New York, NY, USA, 2011; p. 339.
28. Luther, J.E.; Skinner, R.; Fournier, R.A.; van Lier, O.R.; Bowers, W.W.; Côté, J.-F.; Hopkinson, C.; Moulton, T. Predicting wood quantity and quality attributes of balsam fir and black spruce using airborne laser scanner data. *Forestry* **2013**, *87*, 313–326. [[CrossRef](#)]
29. Woods, M.; Pitt, D.; Penner, M.; Lim, K.; Nesbitt, D.; Etheridge, D.; Treitz, P. Operational implementation of a LiDAR inventory in Boreal Ontario. *For. Chron.* **2011**, *87*, 512–528. [[CrossRef](#)]

30. St-Onge, B.; Vepakomma, U.; Sénécal, J.-F.; Kneeshaw, D.; Doyon, F. Canopy gap detection and analysis with airborne laser scanning. In *Forestry Applications of Airborne Laser Scanning*; Springer: New York, NY, USA, 2014; pp. 419–437.
31. Laiho, O.; Lähde, E.; Pukkala, T. Uneven-vs even-aged management in Finnish boreal forests. *Forestry* **2011**, *84*, 547–556. [[CrossRef](#)]
32. Ohara, K.L. Dynamics and stocking-level relationships of multi-aged ponderosa pine stands. *For. Sci.* **1996**, *33*, a0001.
33. Puetz, A.M.; Olsen, R.C.; Anderson, B. Effects of lidar point density on bare earth extraction and DEM creation. In *SPIE Defense, Security, and Sensing*; International Society for Optics and Photonics: Bellingham, WA, USA, 2009.
34. Vaze, J.; Teng, J.; Spencer, G. Impact of DEM accuracy and resolution on topographic indices. *Environ. Model. Softw.* **2010**, *25*, 1086–1098. [[CrossRef](#)]
35. Spies, T.A.; Franklin, J.F. The structure of natural young, mature, and old-growth Douglas-fir forests in Oregon and Washington. In *Wildlife and Vegetation of Unmanaged Douglas-Fir Forests*; U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station: Washington, DC, USA, 1991; pp. 91–109.
36. Thomas, V.; Treitz, P.; McCaughey, J.; Morrison, I. Mapping stand-level forest biophysical variables for a mixedwood boreal forest using lidar: An examination of scanning density. *Can. J. For. Res.* **2006**, *36*, 34–47. [[CrossRef](#)]
37. Næsset, E. 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 Sens. Environ.* **2005**, *98*, 356–370. [[CrossRef](#)]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).