

Article

Raising the Bar: Height Threshold and Grid Resolution Influence Repeatability of Crown Closure Estimation by Airborne Laser Surveys

Etienne Bellemare Racine^{1,*} , Antoine Leboeuf^{1,2}, Jean Bégin¹ 

¹ Département des sciences du bois et de la forêt, 2405 rue de la Terrasse, Université Laval, Québec, QC G1V 0A6, Canada;

² Ministère des Forêts, de la Faune et des Parcs, Direction des Inventaires Forestiers, 5700 4^e Avenue Ouest, Québec, QC G1H 6R1, Canada;

* Correspondence: Etienne.Bellemare-Racine.1@ulaval.ca

Abstract: Monitoring crown closure evolution using multi-temporal Light Detection and Ranging (LiDAR) surveys is a method that we expect to be increasingly adopted given the availability of LiDAR sensors and the accumulating survey archives. However, little attention was devoted to comparing crown closure estimates from independent surveys. Although survey parameters cannot be modified after the data collection, we speculate that the error associated to crown closure estimates comparison can be reduced by selecting optimal post-survey parameters. In this study, we compared crown closure estimates of three airborne LiDAR surveys from 2018 (40 pt/m²) used as a reference, and two lower-density surveys from 2016 (4.5 pt/m²) and 2018 (2 pt/m²). We studied the effect of the height threshold used to separate canopy points and the grid resolution, using skewness and variance of lagged difference of crown closure. Crown closure estimates using low height thresholds were more different across surveys, resulting in higher root mean squared error (RMSE), bias and more different variograms. Results show that optimal height threshold was 3 m and grid resolution was 25 m, although there was room for decision (RMSE of 7% and 5%, and bias of 4% and 0% for 2016 and 2018 low-density surveys).

Keywords: Boreal forest; Multi-temporal LiDAR remote sensing; Crown Closure Monitoring; Tree Density; Gap Fraction; Vertical canopy cover; Gap probability; Spatial Autocorrelation; Height threshold; Spatial grid resolution

1. Introduction

Airborne light detection and ranging (LiDAR) measures of stand crown closure are frequently among the most important variables to predict stand attributes and describe the forested habitats [1]. Monitoring crown closure across time through repeated LiDAR surveys is a method that should increasingly be adopted given the availability of LiDAR sensors and the accumulating archives. However, comparing LiDAR surveys across time and space is difficult when parameters change [2].

Comparing surveys is challenging because often multiple parameters vary simultaneously in addition to the evolution of the vegetation that is to be observed. It is therefore difficult to establish how much of the differences between two LiDAR surveys are due to vegetation evolution only.

In the absence of reference ground data, it can be difficult to confirm the validity of a remotely sensed observation. Practitioners such as forest managers are especially likely to face this dilemma when forest management objectives change, or questions are added to a follow-up survey. While it is difficult to measure the precision of a metric estimated by two LiDAR surveys without ground

observations, it is still possible to compare surveys conducted in a short time interval that did not allow for significant changes to understand the factors that can impact the results of the survey.

1.1. Crown Closure

We noted a number of tree density descriptors in the literature: crown closure, stand density, vegetation density, canopy density, (vertical, fractional) canopy cover, gap fraction, and LiDAR penetration. They essentially refer to the proportion of horizontal space occupied by tree stems or tree crowns depending if the survey is terrestrial or airborne [3–9]. In our study, we defined crown closure as the proportion of a surface covered with vegetation, measured as the ratio of first LiDAR returns above a threshold on the total number of first returns on an area. We further detail this definition in the methods section.

1.2. LiDAR measurement

Crown closure can be difficult to measure on the ground because unlike observations of individual tree mensuration, crown closure requires a large area of observation. Furthermore, ground plots locations are often registered within three to five meters, while LiDAR surveys can be matched within less than a meter using standard LiDAR survey protocol [10,11]. As a result, comparison between field estimates and LiDAR surveys can be biased or display inflated error.

Lidar was found to provide measures that matched ground observations of crown closure. Korhonen *et al.* [4] reported an RMSE of 3.7–7.0% and an overestimation of 4.6–3.7% of LiDAR estimates when compared to ground measurements. However, in the absence of repeated measures, it was not established which method, LiDAR or ground measurements, provided the most accurate estimates. In a study of the impact of LiDAR scan angles on crown closure estimates against ground measurements, Liu *et al.* [9] found an R^2 of 0.74, and 0.87 for measurements using a scan angle of 0–7°, and 7–23° respectively; they reported no difference in crown closure estimates for surveys with point densities higher than 5 pt/m².

LiDAR bias for crown closure estimates can be mostly attributed to four elements: (1) the absence of LiDAR pulse due to sensor design, sampling rate, scan pattern, scan angle, and aircraft altitude; (2) the topography which can induce bias in height and distort crowns; (3) the interaction of LiDAR pulses with different vegetation, and canopy architectures such as leaf and needles or canopy shape and angle; and (4) survey post-processing, such as ground point classification, height threshold, and area [2,9,12–14]. Our study concentrates on the impact of post-processing to increase the repeatability of LiDAR surveys.

1.3. Post-survey parameters

Many parameters of a LiDAR survey are decided before the survey, set during the acquisition, and cannot be changed subsequently (e.g. flight altitude). However some parameters that we'll call post-survey, can be modified after the realization of the survey. The choice of the grid resolution or the height threshold used to estimate crown closure can be decided based on the results of the survey. We used combinations of height thresholds and grid resolutions to quantify their impact on crown closure. We set two objectives for the optimization of the crown closure: the distribution of crown closure values should span the full range of possible values (0–100%) and describe the variation found in the study area.

1.3.1. Height threshold

The height threshold is used to separate canopy returns from non-canopy, and its value should be adapted to the environment studied [5]. The height threshold can be calibrated using ground data [15] or rely on the same height used during field observations [4]. Some studies used the ground points directly (no height threshold, or 0 m)[9,16,17]. While using a height threshold too high decreases the overall crown closure, conversely a height threshold too low could artificially increase the overall

crown closure. For that reason, finding the optimal height threshold might in turn provide a more balanced distribution of crown closures throughout the landscape and cover the full range of possible values: 0–100%.

1.3.2. Grid resolution

Crown closure estimates can change depending on the area used to compute it—often a raster grid composed of cells. A small spatial resolution (i.e. small grid cell size) artificially inflates the variability because they are subject to the idiosyncratic trajectory of the LiDAR reflections and tree location. Conversely, using a large area regresses values to the mean and hides the natural variation occurring in the study area. In both extreme grid resolutions, the information cannot be used to discriminate ecosystems. For that reason, there must be an optimal grid resolution to balance volatility, stability and bias [18].

Methods to estimate the optimal cell size for LiDAR surveys vary. For example, ESRI [8] state that the grid resolution should be at least 4 times the point spacing: a survey with 0.5 pt/m² yields a point spacing of 1.4 m/point and would require a minimal cell size of 5.7 m. The disadvantage of this method is that it evacuates the biological component of the crown closure and does not depend on species, crown size, nor vegetation height. White *et al.* [5] suggest that the cell size for forest inventory in general should be close to plot size and adapted to the average crown size, with typical sides of square cells of 20–25 m. Increased plot size provides a more uniform distribution of pulses, minimizes within pixel edge-effect, and improved the estimates by asymptotically increasing the overlapping area between the LiDAR data and the field observations [5,10,19]. On the other hand, comparing vector data such as forest stand polygons to large grid resolutions amplifies the edge-effect (i.e. pixelization). Smaller cell size also better reflects the variability of the study area, which is important for forest management [20]. For crown closure, the grid size has to be large enough to collect enough vegetation and ground points, even in very closed cover, to reflect the local crown closure. Hengl [18] suggested to use a variogram to select an appropriate cell size for point interpolation. Indeed, crown closure estimation can be seen as the interpolation of a highly sampled binary variable across space. The variogram can thus be used to observe the range of spatial autocorrelation and guide the decision of grid size, especially to provide an upper bound for the grid size. However, computing the variogram for an aggregate measure such as crown closure requires to decide on an aggregation measure prior to the calculation of the variogram, which can lead to a distorted understanding of the data.

1.4. Objectives

While LiDAR survey parameters are crucial for crown closure estimates, we have found that little attention was shed on the post-survey parameters that impact crown closure: the height threshold to separate canopy returns from non-canopy returns, and the area on which crown closure is calculated. In this study, we compare the crown closure estimates from three independent LiDAR surveys and propose a method to facilitate the comparison of crown closure across surveys by the optimization of the post-survey parameters. We look at the impact of the two parameters on RMSE and bias between surveys as well as the spatial autocorrelation structure of crown closure over the study area.

2. Methods

The study area partially overlaps the Grands-Jardins National Park, Québec, Canada (Figure 1). The forest is dominated by coniferous species, such as *Abies balsamea*, *Picea mariana*, and some deciduous such as *Betula papyrifera* and *Populus tremuloides*. The site is representative of a wide range of boreal forest conditions, from open black spruce to closed mixedwood forests. The growing seasons are short, between 80 and 100 days [21].

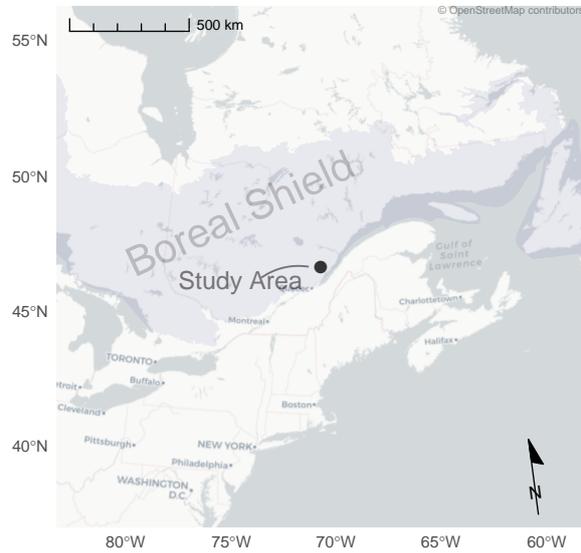


Figure 1. Study area is located in the Boreal Shield ecozone (light blue area) within the Grands-Jardins National Park, Québec, Canada.

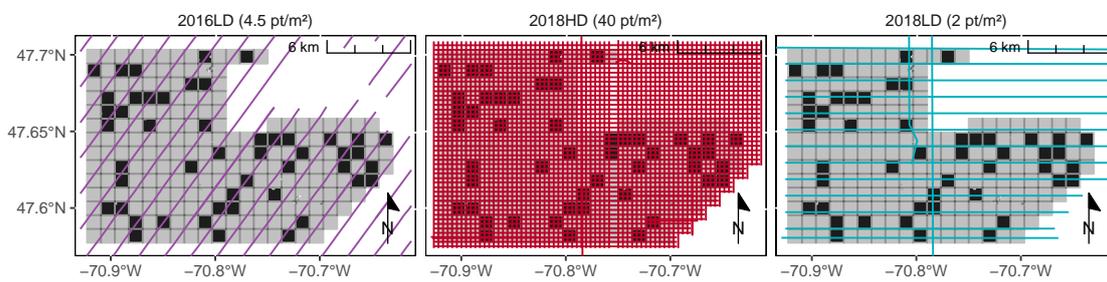


Figure 2. Flight lines (colored) of each survey over the study area (light grey), 1 km tiles holdout data for validation are in black.

2.1. LiDAR

Three LiDAR surveys were used for this study: one conducted in 2016 and two conducted in 2018. 2016 low-density (LD) survey (June 26 to August 15) nominal point density was 4.5 pt/m^2 , 2018LD survey (August 21 to October 5) was 2 pt/m^2 , and the high-density (HD) 2018 survey (August 24 to October 23) was 40 pt/m^2 (Table 1). The 2016LD survey had a north-east azimuth while the 2018LD survey has an east azimuth with a north crossing flight line for registration. The 2018HD survey has crossing flight lines pattern with both north and east azimuths (Figure 2). Apart from the point density, one of the most notable differences of the 2016 survey is that it was conducted earlier in the growing season than the 2018 surveys, and used a higher maximal scan angle, which can impact crown closure estimation [22]. Our study area is the intersection of the three independent LiDAR surveys. All points were classified with *lasground* [23] using default settings for the forest environment.

Table 1. LiDAR survey parameters

Year	Nominal density	Altitude	Speed	Sensor	Freq	Angle	Over	Divergence	Footprint	Median density
2018	40.0	1150	66.8	RIEGL VQ-780i	1000	20	50	0.25	0.1	38.9
2016	4.5	1932	92.6	RIEGL LMS Q-1560	800	30	30	0.25	0.3	4.6
2018	2.0	2300	66.9	RIEGL VQ-780i	350	20	30	0.25	0.4	2.1

Year: year flown

Nominal density: nominal first return density (pt/m^2)

Altitude: average sensor altitude above ground (m)

Speed: Flying speed (m/s)

Sensor: sensor model

Frequency: scan frequency (kHz)

Angle: maximal scan nadir angle ($^\circ$)

Over: Flight line overlap (%)

Divergence: Laser beam divergence (mrad)

Footprint: Average laser footprint diameter on the ground (\varnothing m)

Median density: median first return density (pt/m^2)

2.1.1. Sampling

We randomly sampled 5000 observation points within the study area to extract LiDAR metrics. We excluded two observations which presented insufficient LiDAR returns because they were either on the edge of at least one survey or included in a water body.

To minimize the impact of spatial autocorrelation on the estimation of the performance, we used random block sampling to create a spatially independent holdout [24]. We divided the study area in 1 km tiles to group observations: 1037 observations in the holdout tiles were selected to compute performance metrics (Figure 2).

2.1.2. Data Preparation

We computed the crown closure as the total number of first returns divided by the total number of first returns above the height threshold (Figure 3). Using only first returns was found to be more stable than using all returns or a canopy height model [4,10,25], although [9] used a weighting to use all returns. For each sample point, we used eleven height thresholds (0 to 10 m) and 36 resolutions (1–20 m

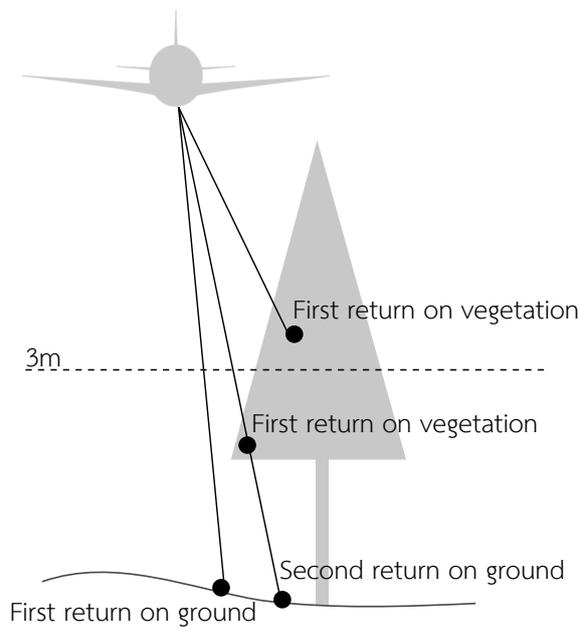


Figure 3. Illustration of crown closure. Each dot represents a LiDAR return and each stroke represents a pulse. The dashed line indicates a height threshold at 3 m. Crown closure is the ratio of the number of first returns above the threshold by the total number of first returns. In this example, crown closure above 3 m is: 1 first return above 3 meters / 3 first returns = 33%.

by 1 m increment, then 25–100 m by 5 m increment) on a squared pixel centered on the sampling point of observation.

2.1.3. Skewness

We needed to select a height threshold that is neither too high nor too low. In order to select the right height threshold, we compared the crown closure distribution to maximize the symmetry of the crown closure distribution across our observations. Given that our study area was representative of the boreal forest conditions with closed mixed stands and open coniferous stands, we speculate that a symmetric distribution of crown closure should lead to a more representative indicator because it avoids clustered crown closures around 100% (threshold too low) or 0% (threshold too high).

2.1.4. Variance of lagged difference in crown closure

The size of the observation unit was found to be an important driver of accuracy and variability. To find the optimal plot size, Frazer *et al.* [10] recommended to increase it gradually until measurements stabilized; given the large number of parameters combinations and observations, we propose a formal indicator based on their suggestion.

One of our objectives was to reflect the variability of the forest cover, while still removing the noise from the individual tree measurements. We chose to use the variance of lagged difference in crown closure as an indicator of crown closure stability. For each pair of consecutive pixels at observation i and resolutions r in a $\delta = 5\text{ m}$ increment, we computed the crown closure $Z_i(r)$ difference $d_i(r)$ between the increased pixel area. Then for each pixel resolution r , we computed the variance of the lagged differences

$$d_i(r) = Z_i(r) - Z_i(r + \delta) \quad (1)$$

The variance V of the lagged difference at each resolution r is

$$V(r) = \frac{1}{n} \sum (d_i(r) - \bar{d}(r))^2 \quad (2)$$

with $\bar{d}(r)$ the average $d_i(r)$ of all i . We expected the variance to be rapidly decreasing as the pixel size increased and then stabilized close to 0 once the pixel size was sufficiently large to encompass enough area. It would mean that increasing the pixel by δ would not change crown closure $Z_i(r)$. We deliberately used a notation close to the one typically used in variography (see e.g. Dale and Fortin [26], §6.2.3).

2.2. Performance metrics

We compared LiDAR surveys crown closure estimates on the holdout data with the exact same post-survey parameters values using root mean squared error (RMSE) and mean bias:

$$RMSE = \sqrt{\frac{\sum^n (\hat{y}_i - y_i)^2}{n}} \quad (3)$$

$$Bias = \frac{\sum^n (\hat{y}_i - y_i)}{n} \quad (4)$$

Where y_i is the observation i of n from the reference 2018 HD survey, \hat{y}_i is the observation from the low density survey.

We performed an analysis of variance on the difference $(\hat{y}_i - y_i)$ for each combination of resolution and height threshold. To protect from Type I error, we used a Bonferroni-corrected $\alpha \approx 0.0001$ to detect differences between surveys. The null hypothesis was that the difference between surveys were zero, meaning the estimated crown closure from low-density surveys was not different from the reference survey.

2.3. Spatial autocorrelation

We also looked at the spatial autocorrelation of crown closure by computing a variogram for each combination of height threshold and grid resolution on the holdout data. The empirical variogram expresses the average variance of pairs of points in a given range of distance. Typically, in the presence of spatial autocorrelation, the variance is lower for nearby points and stabilizes once it reaches a distance at which no correlation is observed. Since crown closure is driven in part by spatial processes, we can expect to observe spatial autocorrelation in the variogram. The variogram can be described using three parameters: the *nugget*, which is the variance at 0 m, a combination of the local variability and measurement error; the practical *range* is the distance at which we cannot detect any spatial autocorrelation; the *sill* is the semivariance reached by independent pairs of points, once the semivariance reaches a plateau [26,27]. An empirical variogram of spatially decorrelated values does not increase with distance; semivariance moves randomly and the trend is flat.

We approximated the empirical variogram as a spherical model using a damped least-squares approximation from the *gstat* package [28,29]. We used an initial seed range of 250 m, and adjusted a few seeds when the algorithm failed to converge to reasonable values.

We used the R software [30] and the *lidR* R package [31] for our analysis and the extraction of lidar metrics. Type 3 skewness was computed using the *e1071* R package [32].

3. Results

The height threshold modified the skewness of the crown closure distribution (Figure 4A). Lower height thresholds of 0–3 m were associated to a negative skew of the crown closure in the study area (more observations toward closed cover), while higher thresholds of 4–10 m were associated to a positive skew (more observations toward open cover). The 3–4 m height thresholds minimized the absolute skewness of the crown closure distribution for all surveys, even if we excluded the

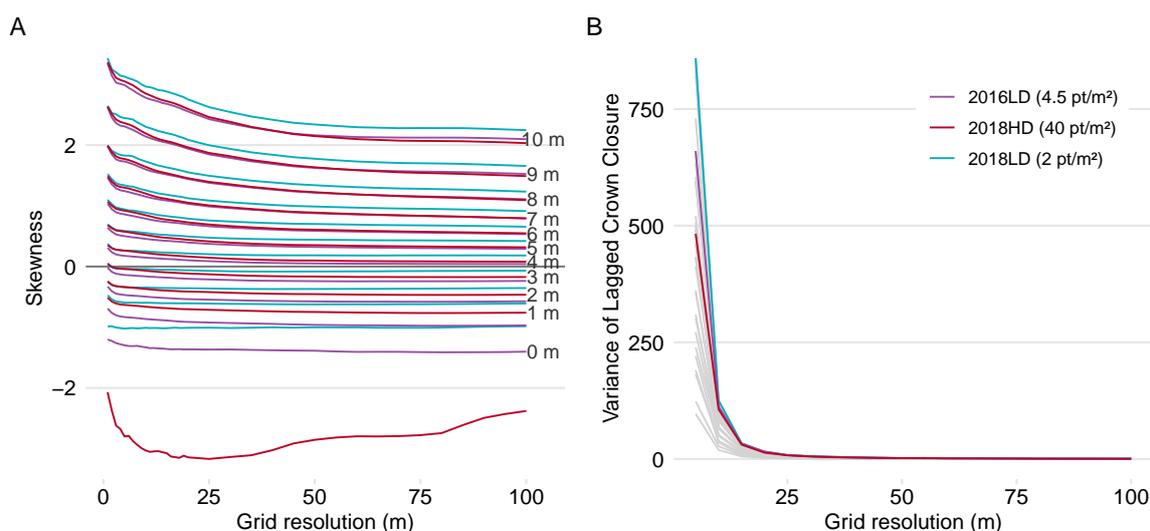


Figure 4. Effect of post-survey parameters. A: effect of resolution on crown closure skewness for 10 height thresholds. B: effect of grid resolution on variance of lagged crown closure difference with 3 m threshold (other thresholds are displayed in grey).

non-forested areas. Increased grid resolution reduced the absolute skewness of crown closure, especially for extreme thresholds (0 m and 7–10 m), except for the 2018HD and 2016LD surveys at 0 m height threshold. The effect of an increased grid spatial resolution (hereinafter referred to grid resolution) on skewness was larger for height thresholds that were more skewed, such as the 0 and 10 m thresholds. For thresholds that produced the most symmetrical distributions (3–4 m), the effect of grid resolution appeared linear for resolutions larger than around 6 m.

Most surveys skewness were similar at a given threshold and grid resolution, but the 0 m height threshold created notably large differences across surveys compared to other threshold values. The 2018LD survey skewness was systematically higher than the two other surveys for all height thresholds. The difference between surveys was more apparent for more extreme height thresholds (0 m, and 7–10 m), while all surveys seemed to reach an asymptotic difference at larger grid resolutions.

Increased grid resolution reduced the variance of lagged difference of crown closure across the study area (Figure 4B). The difference in variance declined rapidly by increasing the grid resolution until 25 m where variance stabilized near zero. The height thresholds that had the highest symmetry (lowest absolute skewness, i.e. 3–4 m) captured the highest variance and were more impacted by the grid resolution, while extreme height thresholds (0 and 10 m) presented smaller changes in variability.

3.1. Comparison with 2018 HD

Using the same post-survey parameters values for each comparison, we found that maximal bias between surveys was at 0 m threshold (Figure 5 top row). At 1 m height threshold, both survey bias decreased to 5.6 and -0.5% for 2016LD and 2018LD respectively. The 2016LD survey was significantly different from 2018HD survey ($p < \alpha$), except for resolution ≤ 2 m with height threshold ≥ 5 m (see supplemental material for detailed results). Using the 0 m height threshold reversed the bias for the 2016LD survey, which became negative. Overall, resolution had little impact on the bias especially for resolutions larger than 7 m. The bias slightly changed for small resolutions and stabilized at around 7 m resolution. The 2018LD survey was not statistically different from the 2018HD survey for height thresholds ≥ 1 m ($p < \alpha$).

Higher height thresholds were associated to a lower RMSE, with an important decline between the 0–1 m (Figure 5 bottom row). For grid resolutions smaller than 7 m, the RMSE remained above 10%

for thresholds <2 m. The increased grid resolution reduced the RMSE with a steep decline in 1–7 m resolutions.

3.2. Spatial autocorrelation

The empirical variogram of crown closure was influenced by the grid resolution and the height threshold selection (Figure 6). Small grid resolutions displayed weak structure of spatial correlation on crown closure, increased variance and larger differences between surveys, especially when combined with low height thresholds of 0–1 m. The 2018LD survey exhibited similar variogram than 2016LD at lower height thresholds and gradually shifted toward the 2018HD as the threshold was raised. The 2018HD survey semivariance (sill and nugget) was generally lower, especially at resolutions ≤ 2 m, or 0 m height threshold. Furthermore, the 0 m height threshold also amplified differences between surveys and decreased the observed sill and nugget of spatial autocorrelation.

Using damped least-squares fitting of a spherical variogram, we observed ranges of 100–350 m excluding the 0 and 10 m height thresholds. Surveys had mostly similar ranges for identical combinations of parameters (global $R^2=0.85$), however resolutions >50 m had much lower correlation between ranges. The mean absolute difference of range between surveys stabilized at around 12 m (not shown).

3.3. Parameter selection

The selection of the grid resolution and height threshold must generally be performed without a reference survey. For that reason, we evaluated the relevance of the skewness and lagged variance of crown closure difference to select a combination of parameters that produced comparable surveys.

Given our study area, and the objective we set for our crown closure estimators, we selected the 3 m height threshold (Figure 4A). It provides a balanced crown closure distribution and capture the most natural variance of the crown closure within the study area. The 4 m height threshold would also have been a good choice, but considering that the vegetation height rarely exceeds 20 m, we favored a lower threshold.

The 25 m grid resolution was the smallest unit that achieved complete stabilization of the crown closure (Figure 4B). Practically, it is often convenient to select a grid resolution that divides evenly into 100 [5], so resolutions of 20 and 25 m seemed good choices and could have been matched to ground plot area or preexisting grid resolutions.

With a 3 m height threshold and a 25 m grid resolution, we observed for the holdout data that the 2018HD and LD surveys were very similar with an R^2 of 0.98, RMSE of 5%, and no bias (Figure 7). The 2016LD survey was more different with RMSE to 7% (R^2 was 0.97), and bias of 4%, which indicated cover opening between 2016 and 2018.

4. Discussion

We found that the selection of a height threshold and grid resolution can improve the repeatability of the crown closure measurements in multi-temporal LiDAR surveys. While the grid resolution had a very large impact on performance indicators, it was mostly limited to small resolutions (<7 m) that are less common in area-based approaches. It is however more common to use low height thresholds (<3 m) which produced volatile crown closure estimates, especially when only ground returns were used (0 m height threshold).

4.1. Repeatability

The error between the 2018 surveys was 5% RMSE and 0% bias, while 2 years-apart surveys had a 7% RMSE and 4% bias by using a height threshold of 3 m and a grid resolution of 25 m. The decision of using a slightly different grid resolution or the 4 m height threshold had an impact on bias and RMSE of less than 1%, but using a much smaller grid (e.g. 1 m), or a much lower height threshold (e.g. 0 m) made the survey much less comparable because it inflated the bias beyond 10% and the RMSE close

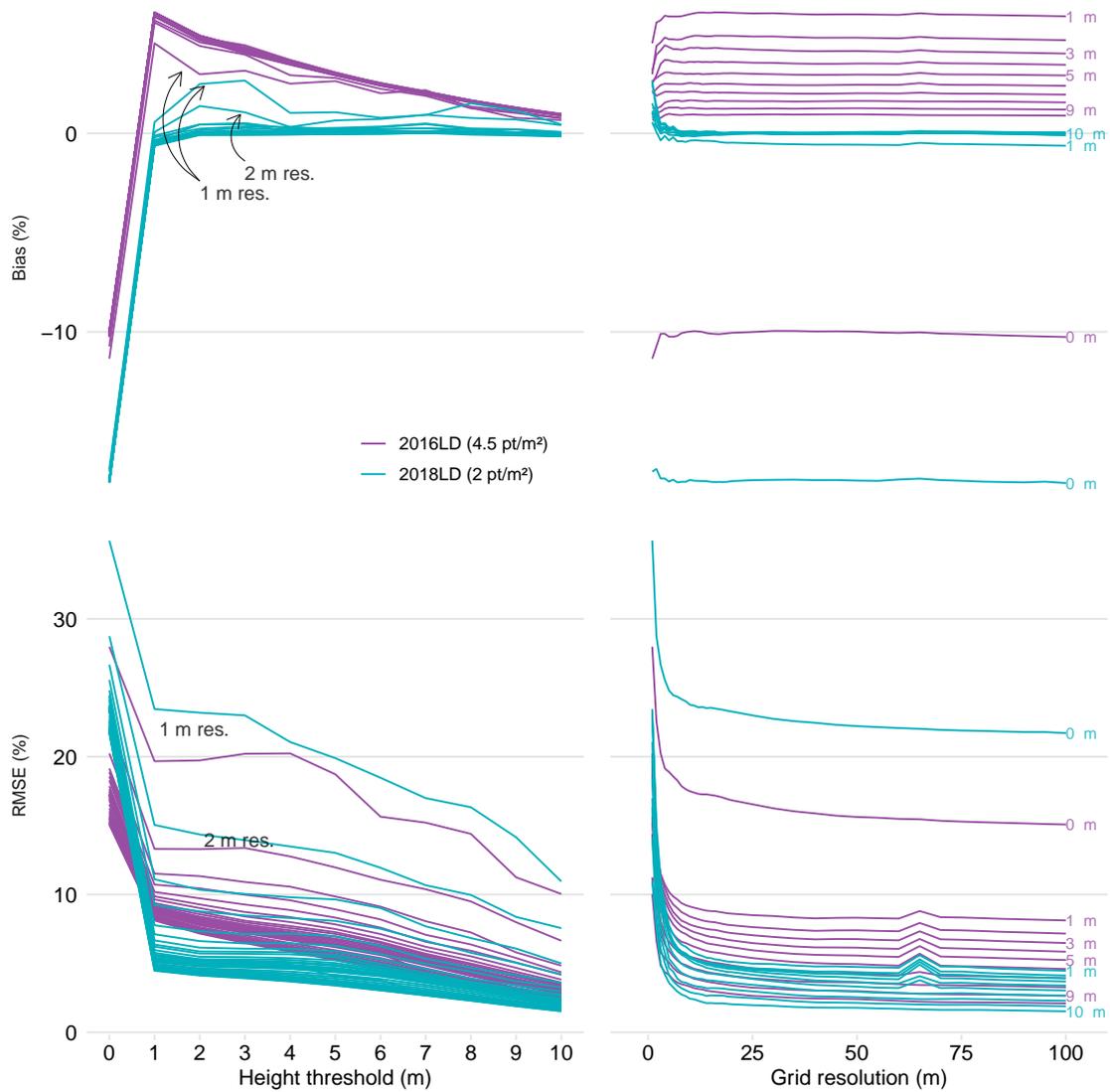


Figure 5. Effect of post-survey parameters on difference between surveys. Left: Effect of height threshold on bias (top) and RMSE (bottom) against the 2018 HD survey. Right: Effect of grid size on bias (top) and RMSE (bottom) against the 2018 HD survey.

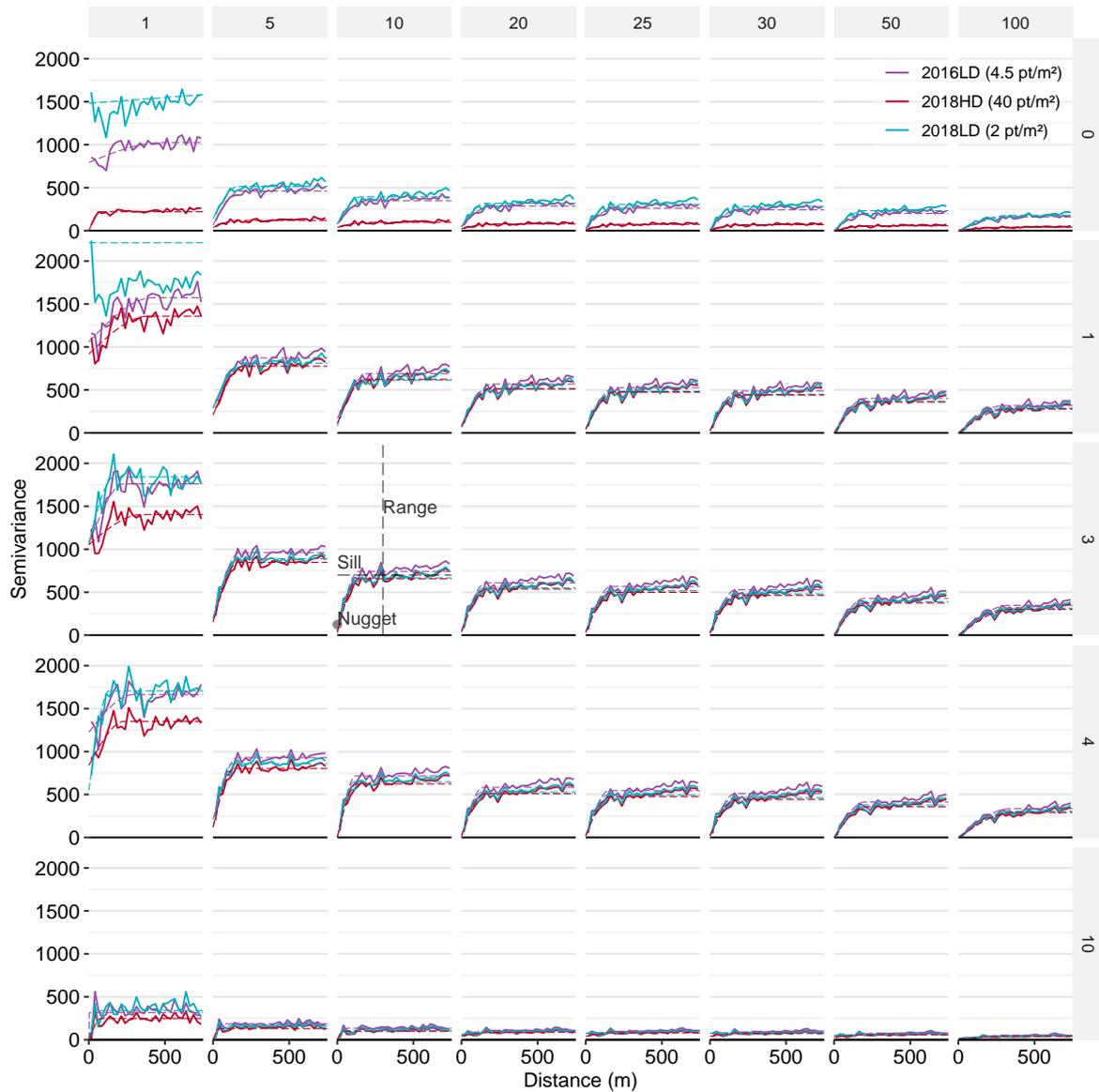


Figure 6. Experimental variogram of crown closure for selected grid resolutions (columns) and height thresholds (rows) using holdout data. The dashed lines show the corresponding fitted theoretical spherical variogram. Range, sill and nugget are illustrated for reference.

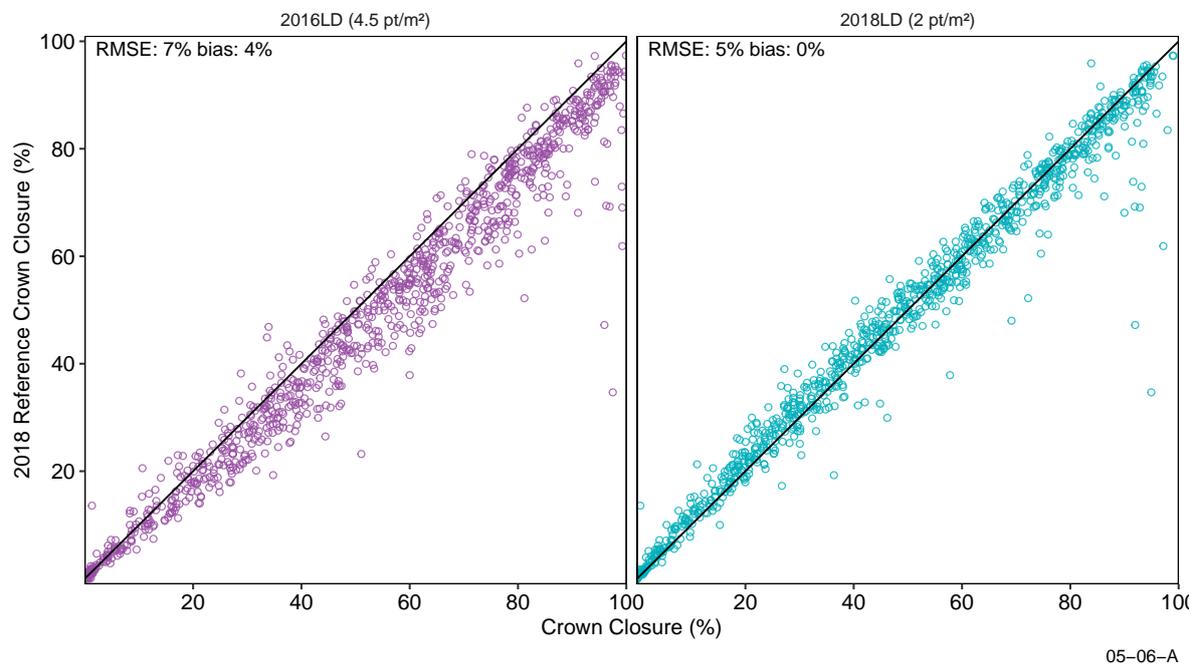


Figure 7. Comparison of crown closure from the holdout data against the 2018HD LiDAR survey using a 3 m height threshold and 25 m grid resolution.

to 30%. We conclude that multi-temporal LiDAR surveys can provide a repeatable measurement of crown closure, but the selection of the appropriate post-survey parameters is an important decision.

The general 4% decrease in crown closure between 2016LD and 2018HD suggested that the forest opened, which we doubted. Forest openings can be attributed to individual or group mortality, epidemics, windthrow, logging, or other non-replacing stand disturbances [33,34]. While these processes take place in the area, a general opening of the forest cover does not seem to correspond to the ecological processes in place and we did not observe any important logging activity. Furthermore, while some non-replacing stand disturbances can be small, most of them would have affected the complete area of the grid (1/16 ha). However early insect infestation, such as the spruce budworm known to be endemic in the area, could have reduced the foliage. The 2016LD survey was also conducted earlier in the growing season than the 2018 surveys. A different timing could lead to measuring different canopy and understory leaf development and modify the crown closure measurement. Higher maximal scan angles can also explain the overestimation of crown closure in 2016LD [4,9,22]. While transformations can be applied to the raw point cloud to correct for biased crown closure estimates [9,22], the relation with the post-survey parameters remains to be studied.

The differences in the variograms (Figure 6) illustrated that the choice of parameters also impacted how each survey reflected the spatial structure of crown closure. Low thresholds or small resolutions were more likely to produce structures of spatial correlation that were less repeatable across LiDAR surveys. Using a 0 m threshold created differences in the observed spatial organization of crown closure that remained even at 100 m resolution.

Korhonen *et al.* [4] reported an R^2 of 0.73 when they compared one-year apart crown closures using a height threshold of 1.3 m and resolutions between 20 and 40 m (to match ground plot sizes). In comparison, the R^2 for 2016LD and 2018LD at 1 m height threshold were 0.96 and 0.97 respectively. These differences could be explained by vegetation growth, disturbances, and lidar parameters.

4.2. Height threshold

The height threshold had an impact on repeatability of crown closure measurements across surveys. The absence of a threshold (0 m) produced volatile estimates of crown closure. As a result, low height thresholds, especially when computed on a small grid size, produced crown closure estimates with a higher bias, RMSE, and semi-variance when compared to the reference LiDAR survey.

The height threshold skewed the distribution of crown closures for the study area. Lower height threshold produced more closed cover observations, but more importantly, they almost never produced open areas. While in itself this could reflect the lack of open areas, it over represented closed canopy stands and prevented the use of the full range of possible crown closure values. When modeling forest attributes, for forest volume estimates for example, the selection of variables that can contribute to a prediction is central, and most methods from generalized linear models to machine learning models generally greatly benefit from variables that span a wider range.

We hypothesize that one of the reasons for the volatility of crown closure measurements with a low height threshold is that the ground classification algorithm might have an increasingly large effect as the threshold approached the ground, especially at 0 m, where the points reaching 0 m are strictly those identified by the algorithm. This could explain the very large differences between the three surveys at 0 m, although they were all classified using the same algorithm. In addition to the inherent sensor errors, the identification of ground points can add around 15–20 cm of error (RMSE) depending on the slope of the terrain and the algorithm used [14].

4.3. Grid resolution

Grid resolution did not modify bias between surveys except at resolutions smaller than 10 m that were more variable. RMSE was inflated for grid resolutions smaller than 7 m, and generally decreased with larger grid resolutions. Extending grid resolutions beyond 25 m did not change the crown closure estimates as observed with the variance of the lagged crown closure difference. There were no formal

indicators to constrain the grid resolution beyond the reduction of variation that we could rely on. All our indicators improved when the grid resolution improved, which could lead to the selection of larger resolutions and reduce the expression of variation on the landscape. Hengl [18] suggested the Shannon's information criteria to balance pixel variability and information, but further work would be required to adapt it to crown closure measurements.

4.4. Skewness and Variance of lagged difference

Using the variance of lagged difference and skewness can guide the selection of a height threshold and grid resolution. They present the advantage of being available in the absence of a reference survey, which is not the case of bias and RMSE. This is convenient when comparing archived LiDAR surveys that do not have common reference ground data.

4.5. Practical implications

The influence of the height threshold on the resulting crown closure should prompt a discussion and ideally an evaluation of the optimal selection of crown closure height and criteria for which to optimize. Further work is required to compare the effect of post-survey parameters on field observations and assess the practicality of raising the height threshold to 3 m, or increasing the grid resolution for example.

The ecology of the study area should also influence the selection of the post-survey parameters. For example, dimensions of the crowns and the maximal height of the vegetation should probably affect how surveys can be compared. Larger crowns could impose a larger grid resolution; areas with taller canopy and understory could require a higher height threshold. The ground point classification algorithm performance can depend on the environment where it used. For example, Moudrý *et al.* [14] found that low vegetation could be classified as ground points in steppes. Furthermore, the height threshold could have to be established in function of the season and phenology: a thicker understory in the summer and sparser in spring and fall, might require an adapted height threshold. We should also consider the possibility that post-survey parameters could have a larger impact in different study areas with different crown size and vegetation height, for example. The parameters selected in this study (3 m height threshold and 25 m grid resolution) could potentially apply to similar ecosystems within the boreal forest, and beyond.

Given the impact of small grid size on the error induced between surveys, we recommend a grid size of at least 10 m and a height threshold of at least 2 m to provide reliable crown closure estimates in ecosystems comparable to this study, although we found 3 m height threshold and 25 m grid resolution best suited our objectives. When in need of a smaller grid resolution, for compatibility with other data for example, the use of a moving window could be explored.

In this study, crown closure estimates benefited from a raised height threshold: it made the measurement more comparable to other surveys. Increased grid resolution also provided more repeatable estimates, but decreased the variability of the territory as illustrated by the variograms (Figure 6). When comparing the 2016LD survey to the 2018HD survey, the conclusions on crown closure evolution would have been completely different had we chosen a 0 m or 1 m height threshold. At 0 m height threshold the negative bias showed vegetation cover closing, while other thresholds rather had a positive bias that showed a vegetation opening (Figure 5). This study shows the importance of both grid resolution and height threshold parameters for the repeatability of the crown closure measurements.

5. Conclusion

Comparing crown closure from multi-temporal LiDAR surveys introduces additional sources of error: different survey parameters, and potentially height threshold, and grid resolution. Finding an optimal height threshold and grid resolution can minimize the impact of changing LiDAR survey parameters. We found that using a 3 m threshold and 25 m grid resolution provided the right balance

to describe the variability of the study area and reduced crown closure to 7–5% RMSE and bias to 4–0%. These parameters could be reused in comparable ecosystems within the boreal shield ecozone.

Acknowledgments: We thank Jean-François Bourdon and Marc-Olivier Lemonde for their support in conducting this study. This work was partially funded through a research contract from the Ministère des Forêts, de la Faune et des Parcs du Québec. Map data from Figure 1 copyrighted OpenStreetMap contributors and available from <https://www.openstreetmap.org>

Author Contributions: ER, AL and JB conceived and designed the experiments; ER performed the experiments, the analysis and wrote the paper with AL and JB providing feedback and critical review. All authors have read and approved this version.

Conflicts of Interest: The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, an in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

LiDAR	Light Detection and Ranging
RMSE	Root Mean Squared Error
LD	Low Density
HD	High Density

References

1. FAO. Knowledge Reference for National Forest Assessments. p. 152.
2. Roussel, J.R.; Caspersen, J.; Béland, M.; Thomas, S.; Achim, A. Removing Bias from LiDAR-Based Estimates of Canopy Height: Accounting for the Effects of Pulse Density and Footprint Size. *198*, 1–16. doi:10.1016/j.rse.2017.05.032.
3. Holmgren, J.; Nilsson, M.; Olsson, H. Simulating the Effects of Lidar Scanning Angle for Estimation of Mean Tree Height and Canopy Closure. *29*, 623–632. doi:10.5589/m03-030.
4. Korhonen, L.; Korpela, I.; Heiskanen, J.; Maltamo, M. Airborne Discrete-Return LIDAR Data in the Estimation of Vertical Canopy Cover, Angular Canopy Closure and Leaf Area Index. *115*, 1065–1080. doi:10.1016/j.rse.2010.12.011.
5. White, J.C.; Wulder, M.A.; Varhola, A.; Vastaranta, M.; Coops, N.C.; Cook, B.D.; Pitt, D.; Woods, M. A Best Practices Guide for Generating Forest Inventory Attributes from Airborne Laser Scanning Data Using an Area-Based Approach. *89*, 722–723. doi:10.5558/tfc2013-132.
6. Alain, D.; Boudreault, J.F.; Brochu, M.A.; Larochelle, G.; Lusignan, M.; Philibert, Y.; Vaillancourt, P.; Berger, J.P.; Morin, P.; Dumais, M.; Fournier-Simoneau, L.G.; Plante, D.; Arsenault, J.; Gouin, M.P.; Joncas, J.; Lapointe, M.; Martel, J. *Norme d'inventaire écoforestier - placettes-échantillons temporaires*; Ministère des Forêts, de la Faune et des Parcs, Secteur des forêts, Direction des inventaires forestiers.
7. Burkhart, H.E.; Avery, T.E.; Bullock, B.P. *Forest Measurements*.
8. ESRI. Estimating Forest Canopy Density and Height—Help | ArcGIS Desktop. Available Online: <https://desktop.arcgis.com/en/arcmap/latest/manage-data/las-dataset/lidar-solutions-estimating-forest-density-and-height.l>. Accessed 2020-12-11.
9. Liu, J.; Skidmore, A.K.; Jones, S.; Wang, T.; Heurich, M.; Zhu, X.; Shi, Y. Large Off-Nadir Scan Angle of Airborne LiDAR Can Severely Affect the Estimates of Forest Structure Metrics. *136*, 13–25. doi:10.1016/j.isprsjprs.2017.12.004.
10. Frazer, G.; Magnussen, S.; Wulder, M.; Niemann, K. Simulated Impact of Sample Plot Size and Co-Registration Error on the Accuracy and Uncertainty of LiDAR-Derived Estimates of Forest Stand Biomass. *115*, 636–649. doi:10.1016/j.rse.2010.10.008.
11. Muss, J.D.; Mladenoff, D.J.; Townsend, P.A. A Pseudo-Waveform Technique to Assess Forest Structure Using Discrete Lidar Data. *115*, 824–835. doi:10.1016/j.rse.2010.11.008.
12. Chasmer, L.; Hopkinson, C.; Treitz, P. Investigating Laser Pulse Penetration through a Conifer Canopy by Integrating Airborne and Terrestrial Lidar. *32*, 116–125. doi:10.5589/m06-011.

13. Wasser, L.; Day, R.; Chasmer, L.; Taylor, A. Influence of Vegetation Structure on Lidar-Derived Canopy Height and Fractional Cover in Forested Riparian Buffers During Leaf-Off and Leaf-On Conditions. *8*, [23382966]. doi:10.1371/journal.pone.0054776.
14. Moudrý, V.; Klápště, P.; Fogl, M.; Gdulová, K.; Barták, V.; Urban, R. Assessment of LiDAR Ground Filtering Algorithms for Determining Ground Surface of Non-Natural Terrain Overgrown with Forest and Steppe Vegetation. *150*, 107047. doi:10.1016/j.measurement.2019.107047.
15. Nyström, M.; Holmgren, J.; Olsson, H. Prediction of Tree Biomass in the Forest–Tundra Ecotone Using Airborne Laser Scanning. *123*, 271–279. doi:10.1016/j.rse.2012.03.008.
16. Woods, M.; Lim, K.; Treitz, P. Predicting Forest Stand Variables from LiDAR Data in the Great Lakes – St. Lawrence Forest of Ontario. *84*, 13.
17. Woods, M.; Pitt, D.; Penner, M.; Lim, K.; Nesbitt, D.; Etheridge, D.; Treitz, P. Operational Implementation of a LiDAR Inventory in Boreal Ontario. *87*, 512–528.
18. Hengl, T. Finding the Right Pixel Size. *32*, 1283–1298. doi:10.1016/j.cageo.2005.11.008.
19. Naesset, E. Practical Large-Scale Forest Stand Inventory Using a Small-Footprint Airborne Scanning Laser. *19*, 164–179. doi:10.1080/02827580310019257.
20. Pretzsch, H. *Forest Dynamics, Growth and Yield : From Measurement to Model*; Springer.
21. Pedlar, J.H.; McKenney, D.W.; Lawrence, K.; Papadopol, P.; Hutchinson, M.F.; Price, D. A Comparison of Two Approaches for Generating Spatial Models of Growing-Season Variables for Canada. *54*, 506–518. doi:10.1175/JAMC-D-14-0045.1.
22. Roussel, J.R.; Béland, M.; Caspersen, J.; Achim, A. A Mathematical Framework to Describe the Effect of Beam Incidence Angle on Metrics Derived from Airborne LiDAR: The Case of Forest Canopies Approaching Turbid Medium Behaviour. *209*, 824–834. doi:10.1016/j.rse.2017.12.006.
23. LAStools. Efficient LiDAR Processing Software.
24. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillera-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schröder, B.; Thuiller, W.; Warton, D.I.; Wintle, B.A.; Hartig, F.; Dormann, C.F. Cross-Validation Strategies for Data with Temporal, Spatial, Hierarchical, or Phylogenetic Structure. *40*, 913–929. doi:10.1111/ecog.02881.
25. Bater, C.W.; Wulder, M.A.; Coops, N.C.; Nelson, R.F.; Hilker, T.; Nasset, E. Stability of Sample-Based Scanning-LiDAR-Derived Vegetation Metrics for Forest Monitoring. *49*, 2385–2392. doi:10.1109/TGRS.2010.2099232.
26. Dale, M.R.T.; Fortin, M.J. *Spatial Analysis: A Guide for Ecologists*; Cambridge University Press.
27. Bivand, R.S.; Pebesma, E.J.; Gómez-Rubio, V. *Applied Spatial Data Analysis with R*; Springer New York.
28. Pebesma, E. Multivariable Geostatistics in S: The Gstat Package. *30*, 683–691. doi:10.1016/j.cageo.2004.03.012.
29. Pebesma, E.; Heuvelink, G. Spatio-Temporal Interpolation Using Gstat. *8*, 204–218.
30. R Core Team. *R: A Language and Environment for Statistical Computing Version 3.6.3*; R Foundation for Statistical Computing.
31. Roussel, J.R.; Auty, D.; Coops, N.C.; Tompalski, P.; Goodbody, T.R.; Meador, A.S.; Bourdon, J.F.; de Boissieu, F.; Achim, A. lidR: An R Package for Analysis of Airborne Laser Scanning (ALS) Data. *251*, 112061. doi:10.1016/j.rse.2020.112061.
32. Meyer, D.; Dimitriadou, E.; Hornik, K.; Weingessel, A.; Leisch, F. *E1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*.
33. Kneeshaw, D.D.; Bergeron, Y. Canopy Gap Characteristics and Tree Replacement in the Southeastern Boreal Forest. *79*, 783–794. doi:10.1890/0012-9658(1998)079[0783:CGCATR]2.0.CO;2.
34. Coops, N.C.; Shang, C.; Wulder, M.A.; White, J.C.; Hermosilla, T. Change in Forest Condition: Characterizing Non-Stand Replacing Disturbances Using Time Series Satellite Imagery. *474*, 118370. doi:10.1016/j.foreco.2020.118370.

Sample Availability: The original data used in this study can be obtained upon request to the Ministère des Forêts, de la Faune et des Parcs du Québec. The processed data used for the analysis as well as the code for analysis is available from the project repository <https://github.com/etiennebr/lidar-crown-closure-paper>