Removing bias from LiDAR-based estimates of canopy height: accounting for the effects of pulse density and footprint size

Jean-Romain Roussel\textsuperscript{a,1,*}, John Caspersen\textsuperscript{b}, Martin Béland\textsuperscript{c}, Sean Thomas\textsuperscript{b}, Alexis Achim\textsuperscript{a}

\textsuperscript{a}Centre de recherche sur les matériaux renouvelables, Département des sciences du bois et de la forêt, Pavillon Gene-H.-Kruger, 2425 rue de la Terrasse, Université Laval, Québec, QC G1V 0A6, Canada
\textsuperscript{b}Faculty of Forestry, University of Toronto, Toronto, Canada
\textsuperscript{c}Department of Geomatic Sciences, Pavillon Louis-Jacques-Casault, 1055, avenue du Séminaire, Université Laval, Québec, G1V 0A6, Canada

Abstract

Airborne laser scanning (LiDAR) is used in forest inventories to quantify stand structure with three dimensional point clouds. However, the structure of point clouds depends not only on stand structure, but also on the LiDAR instrument, its settings, and the pattern of flight. The resulting variation between and within datasets (particularly variation in pulse density and footprint size) can induce spurious variation in LiDAR metrics such as maximum height ($h_{\text{max}}$) and mean height of the canopy surface model ($C_{\text{mean}}$). In this study, we first compare two LiDAR datasets acquired with different parameters, and observe that $h_{\text{max}}$ and $C_{\text{mean}}$ are 56 cm and 1.0 m higher, respectively, when calculated using the high-density dataset with a small footprint. Then, we present a model that explains the observed bias using probability theory, and allows us to recompute the metrics as if the density of pulses were infinite and the size of the two footprints were equivalent. The model is our first step in developing methods for correcting various LiDAR metrics that are used for area-based prediction of stand structure. Such methods may be particularly useful for monitoring forest growth over time, given that acquisition parameters often change between inventories.

Keywords: LiDAR, canopy height, LiDAR metrics, pulse density, footprint size, forest inventory, stand structure

1. Introduction

Airborne laser scanning (LiDAR) is a remote sensing technology for characterizing the surface of the earth using a cloud of georeferenced points. A single point records the height at which the emitted light was reflected back to the sensor with enough energy to generate a "spike of intensity". During the last two decades, the adoption of this technology has increased rapidly, along with the number of applications, particularly in the fields of topography and forest inventory. In the forestry sector, LiDAR has the potential to reduce the need for intensive ground-based measurement of stand structure, making it a valuable tool for "wall-to-wall" forest inventory and mapping (Thomas et al., 2006).

1.1. Prediction methods and their limits

The most common approach for describing forest structure is referred to as the “area-based approach” (ABA), because the point cloud is aggregated and summarized into LiDAR metrics that reflect the structure of the forest at the stand level (usually square pixels of 400 m$^2$) (Woods et al., 2011; White et al., 2013). This method is dependent on plot-based inventory data, which is used for the calibration of statistical models relating LiDAR metrics to variables of interest, such as stand height, stand wood volume, and stand aboveground biomass (e.g. Holmgren, 2004; Ioki et al., 2009; Lim et al., 2014; Bouvier et al., 2015).

The alternative “individual tree based approach” of delineating and measuring individual tree crowns is rapidly gaining in importance (e.g. Pyysalo and Hyvönen, 2002; Morsdorf et al., 2004; Reitberger et al., 2009; Kwak et al., 2010; Yao et al., 2012; Vega et al., 2014). However, despite the decreasing costs of data acquisition and the constant increase of computing power, the ABA remains the most practical approach for large-scale inventories because it needs lower point density and is therefore cheaper. For example, due to the large landbase of the Canadian province of Quebec, the Ministry of Forests, Wildlife and Parks (MFWP) has recently made the decision to run a province-wide survey at a low to medium pulse density ($\sim 2$ to 4 pulses/m$^2$). This will not be sufficient for delineating individual tree crowns in closed-crown forests, so we expect that the ABA will remain relevant for some years to come.

However, one drawback of the ABA is that the statistical models used cannot be generalized in every configuration. For example, when relating two metrics $X$ and $Y$ to a quantity of interest $Q$ by the equation $Q = aX^\beta Y^\gamma$, the model is not only specific to the forest type being sampled (Van Leeuwen and Nieuwenhuis, 2010; Coomes et al., 2017), because $a$, $\beta$ and $\gamma$ have been estimated using a local inventory, but is also
likely to be specific to the LiDAR campaign, because $X$ and $Y$ could be specific to the instrument, its settings, and the pattern of flight.

Beyond the bias potentially included in existing models, the fact that ABA-based descriptions of forest structure cannot be generalized is important because in practice this might limit the usage of LiDAR for wide-scale or multi-temporal inventory surveys in forestry. Datasets acquired from different flights, and often different providers, may not be perfectly compatible. In the operational context of the province-wide survey described above, statistical incompatibility of datasets acquired with different device parameters has been observed in contiguous areas leading to a spatial discontinuities of predictions at the exact boundary of the datasets using a metric derived from the canopy surface model that was expected to emulate a measure of stand height made in classical optical imagery (Ferland-Raymond B. & Lemonde M.-O. – MFWPQ, personal communication).

One way to avoid this issue when implementing the ABA on a large scale is to collect inventory data for each LiDAR survey, and to fit the statistical models separately. However, this is not ideal in the case of two contiguous datasets that share the same forest type. Also, such a solution implies a new ground inventory and a new calibration is necessary for each dataset, which is both time-consuming and costly. An ideal automated approach would involve the development of models that remain stable for any LiDAR settings and could therefore be applied to various datasets sampled at different times and by different providers.

One potential solution to this problem is to develop models using metrics that remain stable when acquisition parameters change. Such considerations are rarely presented in the literature, though Næsset (2004) reported that the height of first returns did not vary significantly with flight altitude or footprint diameter (footprint size ranged between 16 and 26 cm), while last returns were more sensitive to variation in footprint diameter. The most common practice is to process a large number of candidate metrics and aim for the highest possible goodness-of-fit by automatically selecting the best combination of usually 3 or 4 of them (for model parsimony) to predict a variable of interest. This approach generally includes little consideration for metric stability. Moreover, the intrinsic nature of LiDAR point clouds implies that there are endless possibilities to develop new variants of each metric, a fact that limits the possibility to make general assessments of their robustness.

A second solution is to examine the effect that acquisition parameters have on the structure of the point cloud, and hence on metrics and model predictions. This option has received more attention in the literature, particularly the influence of pulse density on model predictions (e.g. Lovell et al., 2005; Anderson et al., 2006; Thomas et al., 2006; Gobakken and Næsset, 2008; Lim et al., 2008; Pirotti and Tarolli, 2010; Jakubowski et al., 2013). Most of these studies reached the conclusion that pulse density has little or no effect on predictions because many statistical metrics remain stable when pulse density is artificially reduced (by definition of what a statistic is). Some studies concluded that pulse density affects the accuracy of the predictions without necessarily introducing bias (Magnusson et al., 2007; Magnussen et al., 2010; Ruiz et al., 2014). However, metrics such as maximum height and its derivations are not stable because they are not statistics. Models that rely on unstable metrics can yield biased predictions at low pulse densities (e.g. Nilsson, 1996; Næsset, 1997; Evans et al., 2001; Sadeghi et al., 2015) especially for multi-temporal or multi-provider datasets.

Prior studies generally use an empirical (data-driven) approach to test if acquisition parameters have a measurable effect on particular metrics. However, hypothesis-driven efforts dedicated to correcting the bias that such effects may cause have mainly been restricted to the normalization of signal intensity (e.g. Höfle and Pfeifer, 2007; Kukko et al., 2008). This approach can also be used to recompute LiDAR metrics as if they were obtained from an idealize “standard device”. Such a standardization method should yield the same metrics that would be obtained with an infinite pulse density, a null footprint size and a constant scan angle at nadir as it has been achieved for signal intensity.

1.2. The specific case of maximum height ($h_{\text{max}}$) and derived metrics

In this paper we focus on the metric $h_{\text{max}}$ expressed in two different ways. We derive a mathematical model for understanding how bias in $h_{\text{max}}$ varies as a function of pulse density, forest structure, and the scale at which it is computed (the window size). We also examine effect of the footprint size, and a derived metric called $C_{\text{mean}}$, which allows us to further examine the issue of scale dependency.

We examine two sources of variation in pulse density: variation between datasets and variation within datasets. Variation between datasets is mainly attributable to fixed differences in device and flight parameters. Finer scale variation within a single dataset is due to overlaps between flight-lines (twice as many pulses per square meter on average), and variation in aircraft speed and attitude (mainly pitch adjustments), which are rarely discussed in the literature. Aircraft pitch adjustments are unavoidable because of the need to maintain the specified altitude. Direction and speed corrections are also common and may result in local variations in pulse density. The local pulse density variations that result from pitch corrections create a clear geometric pattern perpendicular to the flight direction (Figure 1). Gatziolis and Andersen (2008) presented a similar pattern and highlighted the fact that its effects on predictions remain unknown.
2. Methods

2.1. Study area

The study area is located within the Haliburton Forest and Wildlife Reserve (fig. 2). The forest is a 32,000 ha privately owned property located in the Great Lakes - St. Lawrence Forest Region of central Ontario, Canada (45°13’ N, 78°35’ W). Elevation ranges from approximately 400 to 500 m above sea level. The forest is a mixture of hardwoods and conifers typical of northern hardwood forests, and sugar maple (Acer saccharum Marsh) is the dominant species, comprising 60% of the basal area. Most of the forest has been managed under selection silviculture for the past 50 years, and was selectively harvested before then. Thus, most of the stands are uneven-aged, with average canopy heights ranging from 20 to 25 m.

2.2. LiDAR data

Two separate LiDAR datasets were acquired in August 2009 with an Optech ALTM 3100 system. The first dataset covers the whole 320 km² of Haliburton forest (brown in figure 2), and was acquired with a standard pulse density (table 1). The second dataset is a small area of 68 ha (36 ha of forest, 32 of lake) within Haliburton forest (purple in figure 2) that was sampled with a high pulse density (table 1) by flying at a low altitude with a higher scan frequency.

The mature forest in this smaller area was sampled with higher density because it encompasses the Haliburton “megaplot” (13.5 ha), which is part of the CTFS-ForestGEO network of long-term forest dynamics research plots (Anderson-Teixeira et al., 2015). The area overflown twice and with large overlaps, which means that on average, each part of this large plot was overflown four times. Pulse density reached 26 pulses/m² on average, ranging from 15 to 80 pulses/m² (maps of pulse density are given in the supplementary materials fig. S1).

Table 1 lists the flight parameters for the two datasets. The acronym “HD” refers to the high density dataset, whereas “LMD” refers to the low to medium density dataset. This information was provided by the data provider as part of the documentation provided with the datasets. The LMD dataset encompasses all of Haliburton, but a subset of this data will be compared to the HD dataset from the megaplot, so in this context we will also refer to this subset as the LMD dataset.

The normalization of the datasets (i.e. the subtraction of the digital terrain model) was done by the provider and we had no access to the raw data. The method was based on triangular irregular network construction from returns classified as “ground”, although we could not obtain further details about the algorithm used to determine point classes.

<table>
<thead>
<tr>
<th></th>
<th>LMD</th>
<th>HD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altitude</td>
<td>1500 m</td>
<td>500 m</td>
</tr>
<tr>
<td>Overlap</td>
<td>30 %</td>
<td>50 %</td>
</tr>
<tr>
<td>Speed</td>
<td>120 kts</td>
<td>120 kts</td>
</tr>
<tr>
<td>Scan Frequency</td>
<td>36 Hz</td>
<td>70 Hz</td>
</tr>
<tr>
<td>System PRF</td>
<td>70 kHz</td>
<td>70 kHz</td>
</tr>
<tr>
<td>Scan half angle</td>
<td>16 °</td>
<td>10 °</td>
</tr>
<tr>
<td>Cross track resolution</td>
<td>0.89 m</td>
<td>0.40 m</td>
</tr>
<tr>
<td>Down track resolution</td>
<td>0.86 m</td>
<td>0.35 m</td>
</tr>
<tr>
<td>Point density</td>
<td>≈ 2 m²</td>
<td>≈ 28 m²</td>
</tr>
<tr>
<td>Pulse density</td>
<td>≈ 1.6 m²</td>
<td>≈ 26 m²</td>
</tr>
<tr>
<td>Footprint size</td>
<td>0.14 m²</td>
<td>0.015 m²</td>
</tr>
<tr>
<td>Area</td>
<td>30,000 ha</td>
<td>68 ha</td>
</tr>
</tbody>
</table>

Figure 1: Heat map of the variation in pulse density across a 4 km² area. Dark blue: low density; light blue and green: intermediate density; yellow and red: high density. Variation is due to overlap between adjacent flight lines (running from left to right) and aircraft pitch corrections, which cause the perpendicular stripes (running from top to bottom).

Figure 2: Map of study areas. Brown area: low to medium density (LMD) dataset encompassing Haliburton Forest. Purple area: high density (HD) dataset encompassing the megaplot (this area was sampled twice, once at high density and once at low to medium density).
2.3. Data processing

2.3.1. Data pre-processing

Lakes and wetlands were removed from the datasets to retain only forested areas. To do so, we used geographic data from the latest provincial cartography of Ontario, which matched the location of lakes and wetlands from our LiDAR datasets very closely.

2.3.2. Rasterization

As shown in Figure 3, we analyzed that data at three nested scales: plot pixel (400 m$^2$), which is commonly used to compute LiDAR metrics for area-based approaches; canopy pixel (4 m$^2$), which were used to compute the canopy surface model as a raster of canopy pixels; spot pixel (0.14 m$^2$), which approximate the size of an LMD footprint, allowing us to test the effect of the footprint size on LiDAR metrics.

![Figure 3: Nesting of plot, canopy, and footprint pixel. The plot pixel were used to compute LiDAR metrics, canopy pixel were used to compute the canopy surface model, the spot pixel were used to test the effect of the footprint size on LiDAR metrics.](image)

2.3.4. Canopy pixel: computing the canopy surface model

A canopy surface model was computed for both the HD and LMD datasets using the canopy pixels. We used the “local maximum” algorithm to identify the highest point in each 4 m$^2$ canopy pixel. This is the simplest algorithm that can be used to compute a canopy surface model, and has the advantage of being amenable to analysis. This algorithm is nothing more than the computation of $h_{\text{max}}$ for a smaller window size. Thus, calculating $C_{\text{mean}}$ enables us to address the question of scale dependency and the question of metrics indirectly linked to $h_{\text{max}}$. Moreover, the method is identical to that used by the provider to extract the canopy surface model. It therefore corresponds to a product that is used in practice.

A 2 × 2 m resolution was selected based on the pulse density of the LMD dataset: it was the highest possible resolution beyond which holes would start to occur in the canopy surface model. It was also the resolution used by the provider, but it remains only a choice made among other possibilities.

2.3.5. Footprint pixel: assessing the effect of footprint size

The footprint pixel enabled us to test whether beam divergence causes additional bias when estimating canopy height. Before describing how we assessed the effect of footprint size using the footprint pixels (see section 2.7.3), we must further explain the conceptual framework of our analysis, beginning with a simple observation.

2.4. A preliminary observation: comparison of the HD and LMD datasets

To assess the magnitude of bias, we used both the HD and LMD datasets to calculate the height metrics, $h_{\text{max}}$ and $C_{\text{mean}}$. The maximum and mean heights were 57 cm and 1.0 m greater, respectively, when calculated using the HD and LMD datasets to calculate the height metrics, respectively.

![Figure 4: Comparison of the height metrics calculated from the HD and LMD datasets, including 586 plot pixels (400 m$^2$) from the megaplot.](image)

The goal of this study is to identify the sources of this bias, and determine how they can be understood, modeled, and predicted. Using a model based on probability theory,
we describe these observations mathematically, as a function of the number of points used to sample a given area. We first describe sampling bias and our model of it from a theoretical perspective. Then, to validate the model, we develop a method for correcting for the effect of pulse density, and show that applying the correction to both the HD and LMD datasets effectively removes the bias, yielding the same height metrics for both datasets. This validation exercise demonstrates that our correction method allows us to recompute the height metrics as if the datasets had been sampled with an infinite pulse density.

2.5. A conceptual framework for understanding sampling bias

The complete mathematical development of the model is described in section 3. Here, we first present a conceptual framework for understanding various sources of sampling bias, using simple diagrams to illustrate the effects of pulse density, sampling area, and crown shape. Then, we describe our model of sampling bias, and explain how it was validated using the HD and LMD datasets.

Figure 5a illustrates how the bias between the observed maximum height ($\hat{h}_{max}$) and the true maximum ($h_{max}$), i.e. the actual highest point of the plot, increases as pulse density decreases. When 21 pulses reach the canopy, the observed maximum height ($\hat{h}_{max,1}$) underestimates the true maximum by the amount $\Delta h_1 = h_{max} - \hat{h}_{max,1}$. In contrast, when only 11 pulses reach the canopy (i.e. after removing every second pulse), the observed maximum height ($\hat{h}_{max,2}$) is even lower, and underestimates the true maximum by the amount $h_{max} - \hat{h}_{max,2} = \Delta h_1$.

Figure 5a also illustrates how the bias increases as the area sampled (x axis) decreases (and pulse density remains the same). A plot pixel (400 m$^2$) includes multiple large trees, so the probability of sampling near the apex of a large tree (near $h_{max}$) is relatively high, and the observed maximum height ($\hat{h}_{max,1}$) only underestimates the true maximum by the amount $\Delta h_1$. In contrast, when only one large tree is sampled (e.g. between 50 and 100 on the x axis), the probability of sampling near the apex is lower. As a result, the observed maximum height in a 50 m$^2$ plot underestimates the true height by an even greater amount $h_{max} - \hat{h}_{max,3} = \Delta h_1$. This bias is even more extreme when using canopy pixel (4 m$^2$) to compute the canopy surface model.

To account for both of the sources of bias illustrated in Figure 5, our model quantifies density-dependent bias at two distinct scales: canopy pixel and plot pixel. The basic form of the model is:

$$ h_{max} = \hat{h}_{max} + \epsilon(\rho) $$

where $h_{max}$ is the true maximum value of a pixel (either a canopy pixel or plot pixel), $\hat{h}_{max}$ is the observed maximum value of the pixel, and $\epsilon(\rho)$ is the modelled bias computed from the local pulse density ($\rho$), as described in section 3. This basic form also applies to calculating the bias of $C_{mean}$ since this metric is derived from a collection of $h_{max}$ computed in a narrow windows.

Sampling density and sampling area are not the only sources of bias. Comparing figure 5a to figure 5b demonstrates that the bias in conifer stands is expected to be larger than the bias in hardwood stands, all else being equal. This is because conifers have more conical crowns, forming a “rougher” canopy with larger variation in the observed maximum height. As described below, our model also accounts for the effect of canopy shape.

2.6. Quantifying canopy shape

We used the HD dataset to quantify canopy shape as accurately as possible. As shown in figure 6a, we used the original data to calculate the number of pulses that returned from each of many different height intervals, and thereby generated “canopy histograms” that provide both a visual and quantitative assessment of vertical variation in the height of return. The number of returns in each bin reflects the probability that a pulse returns from a given height, so the shape of the histogram reflects the vertical distribution of return heights.

Since our goal is to quantify the bias between the true maximum height and the observed value, our point of reference is not the ground but the true maximum height in a given pixel area. Thus, we standardized the histograms by subtracting the local maximum from the height of each return, such that local maximum equals zero, and all the other returns are negative. This is illustrated both for the plot pixel and the canopy pixel in figure 6b and figure 6c, respectively.

Because they are examples, the histograms in figure 6 were obtained using a subset of the HD dataset, but for the purpose of our analyses we used the entire HD dataset to generate one histogram for each scale (i.e. one for canopy pixel and one for plot pixel). These two histograms were used to quantify the average shape of the canopy in the megaplot, and ultimately the magnitude of bias when estimating the true canopy height, as explained in section 3.3.4.

2.7. Validation of the model

2.7.1. Comparing two corrected datasets: HD vs. LMD

We used the model (equation 1) to correct the height metrics calculated from the HD and LMD datasets (Figure 4). The goal was to validate the model by showing that adding the density-dependent error term to the estimated height yielded the same result for both the HD and LMD megaplots. For the model to be valid, the correction must remove the fixed difference in height between the two datasets (Figure 4), resulting in one-to-one relationship between the two datasets. The correction must also increase the goodness-of-fit of the relationship by taking into account secondary sources of density variation within the datasets, such as those attributable to speed and attitude variations.
Figure 5: Dependence of bias on pulse density, pixel area and canopy shape. $h_{\text{max}}$ is the true maximum height, $\hat{h}_{\text{max},1}$ is the observed maximum height in the following three scenarios: $\hat{h}_{\text{max},1}$ a pixel (400 m$^2$) sampled by 21 pulses; $\hat{h}_{\text{max},2}$ a pixel sampled by half as many pulses; $\hat{h}_{\text{max},3}$ a smaller pixel (between 50 and 100 on the x axis) sampled at the same density as scenario 1. As explained in the text, comparing panels (a) and (b) serves to illustrate how bias depends on canopy shape.
Figure 6: Canopy histograms generated using a subset of the HD dataset from the megaplot (a strip of 100×4 m). (a) Original data, (b) standardized at the plot pixel scale, (c) standardized at the canopy pixel scale.
2.7.2. Comparing corrected flightlines from the same dataset

Secondary sources of density variation were isolated by separating the flightlines of the entire LMD dataset, calculating the metrics for each flightline individually, and comparing the repeat estimates of canopy height obtained from plot pixels that were surveyed in two flightlines (and therefore have two independent estimates). For this analysis, our goal was to further validate the model by showing that adding the density-dependent error term to the estimated height reduced the expected bias between pairs of measurements.

For pixels sampled with the same pulse density in adjacent LMD flightlines, the average bias between repeat estimates should be approximately 0. However, there is appreciable variation in pulse density about the mean, which yielded a range of differences in pulse density among the 150,000 plot pixels included in the analysis. Prior to correction, we expected a positive correlation between the difference in height and the difference in pulse density. Adding the density-dependent error term should remove any such correlation, indicating that any residual difference between repeat estimates is unrelated to pulse density.

By applying the correction to the entire LMD dataset, we are assuming that the HD dataset is representative of Haliburton as a whole. In particular, we are assuming that the structure of the canopy in and around the megaplot is similar to that of Haliburton as a whole. While the HD dataset does encompass a fairly large area (36 ha), the average canopy shape may differ somewhat, given that the megaplot itself is largely comprised of old-growth forest (20 ha) that has never been harvested.

2.7.3. Footprint size as a potential source of residual bias

We cannot assume that variation in pulse density is the only source of bias, because Hirata (2004) showed that footprint size affects canopy height estimates. Even if the density-dependent correction described above were perfect, there may be appreciable residual bias between the HD and LMD datasets because the footprint of the LMD dataset is approximately ten times larger. Thus, we developed a method for testing whether beam divergence causes additional bias, using the spot pixels that approximate the size of one LMD footprint, yet contain multiple footprints from the HD dataset. The goal of the analysis is to compare the height of the local maximum to an estimated “equivalent height” of one LMD footprint, as explained in section (section 3).

2.8. Tools used

Data pre-processing and processing was done in the R programming environment (R Core Team, 2015). A purpose-built package named lidR was specifically developed for processing LiDAR data (Roussel and Auty, 2016). The source code for implementing our model is provided in the appendix.

3. A probabilistic model of bias

3.1. Notation

The following notation is used to describe the model:

- \( h_{\text{max}} \): true maximum height for a given area
- \( \hat{h}_{\text{max}} \): observed maximum height for a given area
- \( \hat{h} \): expected (or most probable) maximum height for a given area
- \( \mathcal{P}(E) \): probability of event \( E \)
- \( p \) or \( P \): letters used for a probability
- \( X \): a random variable

SI base units are used for numeric application of the model.

3.2. Quantifying bias using idealized canopy shapes

Section 2.5 described how discrete sampling leads to the underestimation of \( h_{\text{max}} \). This section demonstrates how to quantify the underestimation of \( h_{\text{max}} \) using a probabilistic model. Rather than simply presenting the mathematical derivation of the model, we use diagrams of idealized canopy shapes to illustrate how the bias can be quantified probabilistically.

The probabilistic nature of the underlying sampling process can also be understood by analogy with rolling loaded dice. In particular, when a canopy divided into \( k \)-height bins (Fig. 6) is sampled with \( n \) pulses, the expected maximum height is equivalent to the expected value when rolling a \( k \)-sided dice \( n \) times. The fact that the \( k \)-sides are not equally likely to land face up (in a loaded dice) is analogous to the fact that \( k \)-height bins are not equally likely to return a pulse (Fig. 6). The probabilistic nature of this sampling process should become clearer after reviewing the four cases below.

3.3. A perfectly flat canopy

If the canopy were a perfectly flat surface (Fig. 7a), the observed maximum height can only take one value. This canopy shape is represented by a histogram with one bin (shown on the righthand side of fig. 7a), indicating that a pulse can return from only one height (\( h_0 \)), with a probability (\( p_0 \)) of 1. In this simple case, the observed maximum height \( (\hat{h}_{\text{max}} = h_0) \) will always be the true maximum height \( (h_{\text{max}} = h_0) \), regardless number of pulses (\( n \)). Thus, the expected value of \( \hat{h}_{\text{max}} \), denoted by \( \bar{h} \) and expressed as a function of \( n \), is:

\[
\bar{h}(n) = p_0^n \times h_0 = h_0
\]  

The expected value (or most probable value) is the mean value that would be found if we sampled the surface an infinite number of times. Indeed, a computer simulation of the sampling process confirms this simple mathematical result (compare expected and simulated in figure 7b), which is hardly surprising in this trivial case, but serves to illustrate that our model captures the underlying sampling process,
both in this case and the non-trivial cases discussed further below (for all four cases, we ran 1,200 simulations, including 200 replicates at each of 60 sampling densities).

\[ X \sim \text{B}(n, p_0), \] so the probability that \( h_1 \) is the observed maximum height (\( \hat{h}_{\text{max}} \)) is the probability that at least one of the \( n \) pulses returns at height \( h_1 \):

\[ P(X < n) = 1 - P(X = n) \]
\[ = 1 - \frac{n}{n} p_0^n (1 - p_0)^{n-n} \]
\[ = 1 - p_0^n \]
\[ = 1 - (1 - p_1)^n \quad (4) \]

The expected value of \( \hat{h}_{\text{max}} \), expressed as function of the number of sampling points \( n \) is:

\[ \bar{h}(n) = P(X < h_1) h_1 + P(X = h_0) h_0 \]
\[ = (1 - p_0^n) h_1 + p_0^n h_0 \]
\[ = (1 - (1 - p_1)^n) h_1 + (1 - p_1)^n h_0 \quad (5) \]

Again, the expected value (or most probable value) is the mean value that would be observed if the canopy were repeatedly sampled with \( n \) pulses. Sometimes one or more of the pulses would return from the true maximum height \( h_1 \), but sometimes not, so on average there is bias.

As before, this is confirmed by a computer simulation: comparing the expected and simulated values in figure 8b shows that in both cases the observed maximum height first increases with pulse density, then approaches the true maximum (\( h_1 \)) asymptotically. Thus, approximately 40 pulses are required to observe the true maximum height with high probability. At lower pulse densities, one or more of the pulses may return from the true maximum height, but on average there is bias, since many pulses will return at \( h_0 \), such that \( \bar{h} < h_{\text{max}} \).

Our goal is to quantify the bias shown as a function of pulse density and canopy shape including more realistic canopy shapes. To do so, we must first introduce a more generic form of equation 5 that allows to write a generic form of the equation for canopies with more than one singularity. In particular, we need to re-express equation 5 using the following notation:

\[ p_k' = p_1 = \frac{p_0}{p_0 + p_1} \] (because \( p_0 + p_1 = 1 \)), then, let \( P_k^n \) be

\[ P_k^n = 1 - \left( 1 - \frac{p_k}{\sum_{i=0}^k p_i} \right)^n \quad (6) \]

with \( k \in \mathbb{N} \) and \( n \) still the number of points. We can see that:

\[ P_1^n = 1 - \left( 1 - \frac{p_1}{p_0 + p_1} \right)^n = 1 - (1 - p_1)^n \quad (7) \]
Thus, $P^n_i$ can be substituted for $p_1$, which is equal to $p'_1$ in equation 3 (single pulse) to obtain equation 5 ($n$ pulses):

$$\hat{h}(n) = P^n_i h_1 + (1 - P^n_i) h_0$$  \hspace{1cm} (8)

This generic form can also be expanded to quantify the expected maximum value when sampling canopies with more than two heights (see below).

### 3.3.2. A flat canopy with two singularities

If we add two singularities to an otherwise flat surface (fig. 9), the observed maximum height can take three values. In this case, the histogram includes a third bin, representing the probability $(p_2)$ that a pulse returns from $h_2$, the true maximum height in this case. If we sample this surface at random with a single pulse, the probability of missing the true maximum height is $1 - p_2$. If $h_2$ is missed, we have now two other possibilities i.e. finding $h_1$ or $h_0$. For a single sampling point missing $h_2$, the probability to find $h_1$ and $h_0$ becomes $p'_1 = \frac{p_1}{p_1 + p_2}$ and $p'_0 = 1 - \frac{p_1}{p_1 + p_2}$, respectively. Because $p'_2 = p_2 = \frac{p_2}{p_0 + p_1 + p_2}$, $\hat{h}(1)$ can be written:

$$\hat{h}(1) = p'_2 h_2 + (1 - p'_2) (p'_1 h_1 + (1 - p'_1) h_0)$$  \hspace{1cm} (9)

Again, we note that sampling with more than one pulse is a Binomial process. Thus, the expected value of $\hat{h}_{max}$ can be calculated by substituting each of the probabilities $(p'_i)$ with $P^n_i$, as we demonstrated for a canopy with one singularity (eq. 8). With two singularities, however, such a demonstration would be rather lengthy, so we only provide the final equation for the expected value of $\hat{h}_{max}$:

$$\hat{h}(n) = P^n_i h_1 + (1 - P^n_i) [P^n_k h_k + (1 - P^n_k) h_0]$$  \hspace{1cm} (10)

The expected and simulated values in figure 9a again show that the observed maximum height increases asymptotically with pulse density, but only 20 pulses are required to observed the true maximum with high probability. The curve is steeper in this case because singularity number 2 is wider, and because singularity number 1 provides another value closer to the real maximum than $h_0$, as explained in section 2.5.

### 3.3.3. A continuous canopy

As shown in figure 10a), a continuous canopy can be discretized using a histogram with $k$ bins, one for each height $(h_i)$ and probability $(p_i)$, where $i \in [0,k]$. When sampled with $n$ pulses, the expected value of $h_{max}$ is:

$$\hat{h}(n) = P^n_k h_k + (1 - P^n_k) [P^n_{k-1} h_{k-1} + (1 - P^n_{k-1}) (P^n_{k-2} h_{k-2} + \ldots)]$$  \hspace{1cm} (11)

This equation can be simplified using its recursive form. Let $\mathcal{H}$ be the set of couples height/probability: $\mathcal{H} = \{(h_i, p_i)\mid i \in [0,k]\}$. We can define:

$$H^n_i(\mathcal{H}) = \left\{ \begin{array}{ll} h_0 & \text{if } i = 0 \\ P^n_i h_i + (1 - P^n_i) H^n_{i-1} & \text{else} \end{array} \right.$$

$$H^n(\mathcal{H}) = H^n_{k}(\mathcal{H})$$  \hspace{1cm} (12)
Therefore:
\[
\hat{h}(n) = H^n_k(\{\delta\})
\] (13)
The agreement between the expected and simulated values in figure 10b shows that this recursive function can be used to calculate \(\hat{h}_{\text{max}}\) for realistic canopy shapes like that shown in figure 10a, just as we did for the idealized canopies in figures 7a, 8a and 9a.

3.3.4. Quantifying bias using standardized histograms
Comparing the expected and simulated values has demonstrated that we can calculate \(\hat{h}\) for any canopy shape, and that it varies as a function of both canopy shape and pulse density. Since \(h_{\text{max}}\) is the point of reference (section 2.6), the bias \(e\) (equation 1) is always negative and must be calculated using the standardized histograms. Let \(H_r\) be an histogram standardized with a resolution \(r:\)
\[
e_r(n) = H^n_k(\delta r)
\] (14)

3.4. The effect of footprint size
Including the recursive function in equation 1 isolates a second error term that quantifies the bias associated with footprint size \(\delta\):
\[
h_{\text{max}} = \hat{h}_{\text{max}} + H_k^n(\delta r) + \delta
\] (15)
For the HD dataset, we can assume that the footprint is small enough to have a negligible effect, so we fixed \(\delta\) at 0. However, we did estimate \(\delta\) for the LMD dataset (as explained further below), since the footprint is ten times larger in the LMD dataset.

Figure 11 illustrates why large footprints are expected to underestimate maximum height. The individual red columns represent pulses with small footprints, and the larger orange column in the background represents a pulse that is ten times larger. As shown to the right, the waveform of small footprints is Gaussian with a small standard deviation, so the returned height is rather accurate. In contrast, the large footprint underestimates the height of the local maximum returned by the smaller footprints (i.e. the height at which the orange waveform peaks is lower than the highest red peak). This phenomenon is described in detail by Hancock et al. (2015) and Disney et al. (2010). Note that the large footprint is represented by Gaussian distribution, though in practice it may not be Gaussian (see Hancock et al. (2015)).

We used the spot pixels to estimate the bias for the LMD dataset, since they approximate the size of one LMD footprint (0.14 m²), yet contain 3-15 pulses from the HD dataset. Our method consisted of subtracting the height of the local maximum obtained from the HD dataset from an estimated “equivalent height” of the LMD footprint. In other words, we estimated the average distance between the peak of the orange waveform and the peak of the highest red waveform (figure 11).

Assuming that canopy reflectivity did not change between the LMD and HD surveys, we estimated the height of the orange peak as the point at which the smaller footprints returned 50% of their total intensity (figure 12). This method probably does not provide the correct value in every case, but by computing it over a large number of spot pixels, it can be expected to provide a good estimation of the average “equivalent height”.

The Optech ALTM 3100 system used in this study emits pulses with a beam width (a function of pulse duration) of 1.02 m (Hancock et al., 2015). This width is defined as the distance between the points at which the power drops below 61% of the maximum. This implies that such pulses are unable to distinguish two distinct objects that are located less than 50 cm apart in the direction of the beam axis. Thus, a beam that is 1 meter wide would be unable to distinguish between the first 5 returns in red (figure 12), but it would be able to distinguish them from the other 3 beams in black, which were therefore be excluded when estimating the equivalent height of the orange beam. The value of \(\delta\) was assessed for each spot pixel by subtracting this equivalent height from the height of the highest sampling point among the HD pulses it contained.
Figure 11: The effect of footprint size on the observed maximum height ($\tilde{h}_{\text{max}}$). The red columns are small footprints, while the orange colour should be seen as a single column in the background belonging to a footprint that is ten times larger. The larger footprint has a broader waveform because the intensity is integrated over a larger surface area. As a result, the large footprint underestimates the maximum height recorded by the smaller footprints - i.e. the height at which the orange waveform peaks is lower than the highest red peak.

Figure 12: Illustration of the “equivalent height” of a large footprint (LMD dataset), as estimated from many of smaller footprints (HD dataset). The vertical lines show the intensity of pulses that returned from two sets of objects, one set that is higher in the canopy (red with round end), and one set that is lower in the canopy (black with square end). The integrals at the top show the corresponding waveforms for the small (plain red) and large (stripped orange) footprints. The peak of the orange waveform is the “equivalent height” of a large footprint, and is estimated as the point at which the smaller footprints have returned 50% of their total intensity (cumulative intensity is shown in dotted green). Note that a large footprint (1 m wide) is only able to distinguish two objects, as indicated by the two orange waveforms.

4. Results

4.1. Expected value of the bias as a function of the scale of observation

As expected, the average canopy shape differed substantially (figure 13) when calculated using the canopy (4 m$^2$) and plot (400 m$^2$) scales (histograms are called $H_4$ and $H_{400}$). These two histograms must be used in conjunction with equation 15 to calculate the difference in bias for these two scales of observations.

Figure 13: Canopy shape in canopy pixels (4 m$^2$, $H_4$) and plot pixels (400 m$^2$, $H_{400}$).

Results show that there was less bias when estimating the $h_{\text{max}}$ of plot pixels (figure 14). Approximately 10 pulses/m$^2$ (4 000 pulses) are required to estimate the $h_{\text{max}}$ of plot pixels with reasonable accuracy (mean bias < 10 cm). At 30
pulses/m² (12000 pulses), the bias is negligible.

A higher density of pulses is required for canopy pixels, even though they exhibit less variation in height. Approximately 20 pulses/m² (80 pulses) are required to estimate the \( h_{\text{max}} \) of plot pixels with a mean bias < 10 cm (figure 14).

4.2. Footprint size

The HD dataset included 1160000 spot pixels (0.14 m²), of which 193000 included a sufficient number of pulses (4 or more) to compute the equivalent height (eq. 15). On average, the equivalent height of an LMD footprint was 16 cm lower than the highest HD footprint (i.e. \( \delta = 16 \) cm).

4.3. Comparing two corrected datasets: HD vs. LMD, effect of device configuration

4.3.1. \( h_{\text{max}} \)

The original bias between the HD and LMD datasets was -57 cm, on average (figure 4a). We applied a correction using the black line in figure 14:

\[
h_{\text{max}} = \tilde{h}_{\text{max}} + H_{\text{LMD}}^p(h_{\text{HD}}) + \delta
\]

After correcting each plot pixel individually in the LMD dataset, \( h_{\text{max}} \) increased by 56 cm, on average, with a range of 41 cm to 75 cm (figure 15). In contrast, \( h_{\text{max}} \) only increased by 2 cm for the HD dataset, with a range of 1 cm to 8 cm. The bias between the two corrected datasets was reduced to 7 cm, on average.

The original goodness-of-fit (R²) between the HD and LMD datasets was 0.976. After correcting both datasets, the R² was increased to 0.977, and the RMSE of the regression was reduced to 1.17 cm from an initial value of 1.34 cm (figure 15).

4.3.2. \( C_{\text{mean}} \)

As explained in section 2.3.3, \( C_{\text{mean}} \) is computed using the local maxima from 100 canopy pixels (4 m²), each of which should be corrected using the red line in figure 14. However, correcting each canopy pixel individually is computationally demanding, and would be even more so if the canopy surface model were computed at a higher resolution (in other applications). For this reason, we chose to apply an average correction based on pulse density of the plot pixels. This way, the correction was the same for every canopy pixel and it was computed only once at the plot pixel scale:

\[
C_{\text{mean}} = \bar{C}_{\text{mean}} + H_{\text{LMD}}^p(h_{\text{HD}}) + \delta
\]  

(17)

The original bias between the HD and LMD datasets was -1 m, on average (figure 4b). After correcting each plot pixel individually in the LMD dataset, \( C_{\text{mean}} \) increased by 82 cm on average, with a range of 30 cm to 1.70 m. In contrast, \( C_{\text{mean}} \) only increased by 4 cm for the HD dataset, with a range of 2 cm to 14 cm. The bias between the two corrected datasets was reduced to 7 cm, on average.

The original goodness-of-fit (R²) between the HD and LMD datasets was 0.978. After correcting both datasets, the R² was increased to 0.983, and the RMSE of the regression was reduced to 31 cm from an initial value of 36 cm (figure 16).

To test the validity of using an average correction for each plot pixel, we also corrected each canopy pixel in the megaplot individually, then repeated the analyses described above. The results did not differ substantially from those obtained using an average correction at the plot scale (not shown), so we concluded that an average correction could be applied to the entire LMD dataset.
4.4. Comparing corrected flightlines from the same dataset: effect of aircraft attitude

On average, there was no difference between the repeat estimates of mean canopy height ($C_{mean}$) obtained from LMD flightlines that sampled the same plot pixels twice (Fig. 17a). Prior to correction, however, the difference in mean canopy height ($C_{mean}$) was positively correlated with the difference in pulse density, with the bias reaching more than 50 cm at either extreme. After correction, this correlation was largely removed (Fig 17b), indicating that any residual difference between repeat estimates is unrelated to aircraft speed and attitude.

![Graph showing comparison of corrected mean heights (Cmean) from HD and LMD datasets.](image)

Figure 16: Comparison of the corrected mean heights ($C_{mean}$) from the HD and LMD datasets, including the 586 plot pixels (400 m²) from the megaplot.

We obtained the same result for $h_{max}$ (not shown), but the correlation was weaker, because $h_{max}$ is less sensitive to the pulse density due to an effect of scale.

4.5. Accuracy of quantifying canopy shape with the LMD dataset

All the results above were generated using canopy histograms obtained from the HD dataset. To test whether a LMD dataset could be used instead (as in an application for which only one dataset is available), we repeated all the analyses after using LMD data from the megaplot to generate the histograms. The results were approximately the same, but the residual biases were slightly higher at 13 cm and 20 cm for $h_{max}$ and $C_{mean}$, respectively (results not shown).

4.6. Effect of scan angle

While our method removed most of the bias associated with variation in pulse density, there was considerable residual bias attributable to scan angle. Following the same procedure as that applied in section 4.4 for pulse density, we compared plots from separate flightlines based on their mean angle of incidence. Figure 18 shows a clear effect of scan angle, which is not accounted for in our correction method. This effect was not significant for $h_{max}$ (results not shown).

![Graph showing the effect of scan angle on canopy height difference.](image)

Figure 18: The effect of scan angle as revealed by the difference between repeat estimates of mean canopy height ($C_{mean}$). Rasters that were sampled in two flightlines were binned by the difference in scan angle, and box plots were used to visualize the correlation between the two differences, both before (a) and after (b) correction (150,000 observations).

5. Discussion

5.1. On the usage of $h_{max}$ and $C_{mean}$

The correction of $h_{max}$ and $C_{mean}$ proposed in this study was derived from the initial question we raised about metric normalization. Our capacity to understand and describe the underlying sampling process was the most important factor determining our choice of metrics for this study. Nevertheless, correcting biases for these metrics is also important in practice. Even if $h_{max}$ can be avoided in predictive models in favour of other less density dependent metrics such as lower percentiles, its use remains common whether it is in a direct or indirect form. The latter can occur, for example when metrics are defined based on a layerization of the point cloud. For example, Woods et al. (2008) defined a metric $d_n$ which can mathematically be expressed as:
5.2. Removing bias from estimates of canopy height

We have shown that our model can be used to remove the bias in maximum height ($h_{\text{max}}$) and the bias in the mean height of the canopy surface model ($C_{\text{mean}}$), both of which are substantial when the LiDAR data is collected at a low to medium pulse density. We have also shown that there is considerable variation in pulse density within a single dataset (figure 1), and that the resulting biases can be removed by our model as well.

The asymptotic relationship that we observed between bias and pulse density is similar to that observed by Hirata (2004), who found that the number of trees located using local maxima reaches an asymptote at 10 pulses/m$^2$ and higher, but decreases sharply below 3 or 4 pulses/m$^2$. Similar asymptotic relationships have also been described by Jakubowski et al. (2013) and Hansen et al. (2015). Our model also describes the underlying mechanism leading to a plot size dependency, as found by Hansen et al. (2015). However, these authors described the relationships empirically, whereas we modelled it based on probability theory. Furthermore, our model not only provides an a mechanistic explanation of the underlying sampling process, but also the means to correct the resulting bias. Our model focuses on two specific metrics, but each of the additional questions raised in these cited references remain driven by the probability theory, and are thus more likely to be understandable in a model rather than from descriptions of local observations.

Our conclusion is that the new metrics obtained from our analytical model are accurate and correspond to what would be computed if the data were sampled with an infinite pulse density. This assertion is reasonable if an error of 10 or 15 cm is considered acceptable. Our results do not suggest that the residual error can be further reduced using our method, but the gain in accuracy compared to using the raw data remains substantial. Our results imply that caution should be used when building predictive models from such uncorrected metrics. It is difficult to generalize their effects on predictions because they depend on multiple factors such as the model used, the plot size, the dataset used, the device settings, the forest type and the model calibration method. However, in a homogeneous and dense hardwood forest the effect is expected to be rather low. Conversely, in a sparse coniferous forest it could be more important.

5.3. Effect of footprint size

The effect of footprint size on height bias has received little attention in the literature. One of the few studies Hirata (2004) was conducted in mountainous terrain, and found that large footprints overestimate the maximum height recorded by smaller footprints, the opposite of what we found. However, this result may be specific to mountainous terrain, because it was explained based on geometric considerations related to topography and slopes. Furthermore, the footprint sizes were much larger, reaching 1.1 m$^2$, nearly ten times larger than the footprint of the LMD dataset. In experimental conditions closer to ours, Andersen et al. (2006) found a similar effect of underestimating tree height of few

\[ d_n = \int_0^n f(z) dz \]

with $n$ being an integer between 1 and 9 and $f(z)$ the probability distribution of points on the $z$ axis. In this article the term “maximum height” is never used and equation 18 is not provided, but a careful interpretation of the metric description leads us to state that each $d_n$ is biased because of the indirect use of $h_{\text{max}}$. This can be referred to as a second order usage of $h_{\text{max}}$.

The case of $C_{\text{mean}}$ is another example of indirect usage of $h_{\text{max}}$. We found only three other examples of this metric being used in the literature (Ruiz et al., 2014; Asner and Mascaro, 2014; Coomes et al., 2017). However, it remains an interesting metric with potential applicability in the development of predictive models of forest structure. Most metrics used in ABA models are unidimensional metrics derived from the $z$ coordinate only. But with the LiDAR point cloud being, at least, a tri-dimensional dataset it can be argued that it is reductive to use only one of them. Features extracted from the canopy surface model represent an easy and accessible way to extract information from the three spatial coordinates. For example, it can be used to extract information on the texture of the forest canopy. The interest of such metrics derived from canopy surface model is recognized and they have been used in the literature (e.g. Kane et al., 2010; Ruiz et al., 2014; Asner and Mascaro, 2014). However, our study shows that they must also be used and interpreted with caution. The choice of the algorithm used to compute the canopy surface model is not without consequences. The local maximum algorithm is a simple, easily implementable algorithm which is used in the recently developed itcSegment R package (Dalponte, 2016). A careful study of the source code for this package shows the canopy is computed using such an algorithm with a linear interpolation and smoothing as post process. Ruiz et al. (2014) computed a canopy surface model in the same way except for the use of an inverse distance weighting interpolation.

The availability of such tools implies that users have the possibility to derive various types of metrics from a canopy surface model. As highlighted in the introduction, such metrics are currently being provided in an operational context. In some cases, this may create bias issues which may be corrected using the probabilistic approach proposed in this study. Beyond this, our analysis of the behaviour of $h_{\text{max}}$ at different scales provides a case study that may help raise general awareness about the fact that various metrics can be more or less sensitive to device parametrization, forest structure, footprint size, plot size, etc. It also demonstrates that such variations may not always be trivial.
5.4.1. Predicting stand structure and monitoring growth

Our correction method produced good results both for $h_{\text{max}}$ and $C_{\text{mean}}$, indicating that it is reasonable to attribute the remaining bias to footprint size. However, the footprint correction should only be seen as a plausible explanation for the residual bias. Whether or not it is the real cause remains debatable. Indeed there could be additional error caused by the non-random distribution of pulses. Our model assumes that pulses are randomly and uniformly distributed in space, but in reality they follow a clear scanning pattern (a seesaw wave). This model assumption, required to make the mathematical development, may have an influence that we believe to be negligible compared to the gain in accuracy that we obtained. Another source of residual bias could come from the unknown pre-processing done by the provider. For example, the point classification step may have differed between the HD and LMD datasets.

5.4. Implications for the state of the art

5.4.1. Predicting stand structure and monitoring growth

We have shown that $h_{\text{max}}$ and $C_{\text{mean}}$ are systematically underestimated unless a sufficiently high pulse density is used to approach the asymptotic values. The density-dependence of LiDAR metrics may limit the applications of the aerial LiDAR technology, especially when two datasets sampled with different parameters need to be joined (different contracts for a large area) or compared (two datasets are sampled at a five-year interval to monitor forest growth).

Using the same pulse density in each inventory is not a solution to this problem, because pulse density changes substantially within a single dataset, as seen in figure 1. For example, the model predicts that $C_{\text{mean}}$ is 50 cm higher in overlaps where the pulse density is twice as high, whereas $h_{\text{max}}$ is 10 cm higher. Homogenizing the pulse density within a dataset could be a good way to avoid this problem, but removing points will introduce more uncertainty. Indeed, a metric can be seen as the single realization of a random variable, which is thus associated with a given uncertainty. A higher point density implies a lower uncertainty. Removing data willfully would not make much sense as it equates to adding noise in otherwise more accurate data. Therefore, a correction of metrics based on a hypothesis-driven approach appears preferable.

In practice, it is unlikely that a separate high density dataset would be available to generate the canopy histograms. We positop that this step could be achieved using local areas of high sampling density as a reference. The interface between overlaps and zones where aircraft pitch correction has further increased the sampling density could be used, for example. Since we found that a very high pulse density was not necessarily required (section 4.5), this solution should provide satisfying results. For larger areas than that used in our study, we suggest using a moving window to build a correction profile field, which would then be defined in any location. Hence, we consider that several solutions can be implemented to apply our model to any low density dataset acquired over any forest type. However, since our main focus was to present a formal description of the underlying sampling process, providing more specific guidance for practical applications is beyond the scope of this study.

5.4.2. Other high percentiles of height

The 99th percentile is often used instead of maximum height (e.g. Garcia et al., 2010; Singh et al., 2015; Goodwin et al., 2007), because it is both representative of maximum height while being robust to the noise caused by any possible outliers García et al. (2011). However, similarly to maximum height, the 99th percentile is also subject to underestimation, but by a smaller value due to the slightly higher probability of sampling it. The same logic of decreasing underestimation applies to the 98th percentile and each subsequent percentile. The bias is expected to decrease until the quantiles eventually become statistically stable. These hypotheses have been empirically tested in other results (fig 2 in supplementary materials) and we found that percentiles become very stable around the 90th percentile in our dataset, but the biases become positive after the 90th. While the methodology and the equations developed in this study could likely be transferred to the higher percentiles, the full mathematical development would undoubtedly be much more complex than for the limit case of the maximum height. The theoretical quantification of this effect for other percentiles is beyond the scope of this study.

5.5. Alternatives to the local maximum algorithm

We used the local maximum algorithm because it is the simplest method to compute the canopy surface model, and because it has the dual advantage of being amenable to analysis using probability theory, while also allowing an easy assessment of scale dependency. However, it is not necessarily the most widely used algorithm in practice.

Some algorithms modify the results obtained from the local maximum method (e.g. Popescu, 2007). They consist of computing the local maximum at a high resolution and filling the holes with an interpolation algorithm. Interpolation may render mathematical analysis more difficult to solve, but the preliminary considerations made in section 2.5 remain applicable. Obviously, a careful study of the effect of LiDAR parameters on interpolated canopy surface model would still be required.

Triangular irregular networks are also commonly used (e.g. Maltamo et al., 2004; Zhao et al., 2009; Asner and Mascaro, 2014). Metrics derived from this kind of representation may also be unstable, so we can also expect some artefacts and side effects. Thus, a careful study of the effect of LiDAR parameters on the canopy surface models produced by this type of algorithm would also be required.
5.6. A more complex issue than usually portrayed

In the majority of cases, the effect of pulse density or scan angle is studied in the literature in an overly simplistic way that does not correctly represent reality. We have shown that the problem is much more complex than what is suggested by a simple artificial reduction of pulse density. In reality, variations of pulse density are accompanied by variations of aircraft altitude, and therefore of footprint size. The footprint size changes the behaviour of the rays, allowing them to penetrate more or less easily into the canopy. Aircraft altitude changes are also be accompanied by variations in other parameters like aircraft speed, scan frequency, and emitted pulse frequency. Such changes modify the sampling pattern over the forest and the shape of the full waveform returns.

Furthermore, the results of empirical experiments are only valid locally, and they do not elucidate the underlying mechanisms. There is therefore a need for mathematical models that provide a mechanistic explanation of the underlying sampling process. Our model was only designed to recompute two metrics while accounting for two effects, and yet the model is still rather complex, despite the fact that we did not consider how the beams penetrate the canopy (we analysed only \( h_{\max} \) at different scales). Modelling the effect of pulse density and other parameters using all returns, for example, would be much more challenging because of penetrating beams.

Admittedly, we chose only two of many possible metrics because they were relatively easy to model. But there remains need to model other metrics, and not only as a function of pulse density, but also of the scan angle, the footprint size, the pulse duration or the scanning pattern.

6. Conclusion

The metrics used in an area based approach do not represent absolute values, meaning that they depend not only on forest structure but also the LiDAR device, its settings, and the pattern of flight. As a result, some metrics are systematically underestimated, and we have shown that the magnitude of bias depends on pulse density, canopy shape, observation scale and probably footprint size. Furthermore, we developed a model that explains the observed bias and allows us to recompute the metrics as if the density of pulses were infinite, while also controlling for the effect of footprint size and observation scale.

This is a first step towards developing what we refer to as a standardization method, that consists of recomputing metrics as if they were obtained using a “standard device” and “standard parameters”. It follows a similar approach to that currently used to correct for variations in signal intensity within and between datasets. The ultimate goal is to describe the behaviour of all metrics as a function of the most important device parameters, such as pulse density, scan angle, footprint size, pulse duration or emitted energy.

It is important to bear in mind, however, that data providers consider some information to be proprietary, such as the algorithms used to discretize the full waveform signal and classify the points, as well as some details about the sensors. These details inevitably introduce variability that is impossible to model if the information is not made available to the end-users.

Acknowledgement

This research was funded by AWARE (the Assessment of Wood Attributes from Remote Sensing NSERC CRDPI462973-14; grantee Prof Nicholas Coops, UBC), in collaboration with Canadian Wood Fiber Centre (CWFC), FP-Innovations and Tembec.

References


