Processing discrete-return profiling lidar data to estimate canopy closure for large-area forest mapping and management

Adam J. McLane, Gregory J. McDermid, and Michael A. Wulder

Abstract. We performed a series of empirical experiments designed to refine the processing of discrete-return profiling light detection and ranging (lidar) data for the purpose of estimating canopy closure across a broad range of forest conditions in west-central Alberta, Canada. The following three methodological conclusions were obtained: (*i*) a new line-segment method based on the ratio of overstory segment distance to total distance outperformed alternative point-count techniques described previously in the literature; (*ii*) an absolute overstory–understory threshold of 1.4 m generated the best models overall and appeared to extend well across a range of forest types; and (*iii*) stratification by species composition (hardwood, softwood, and mixedwood) or moisture regime (upland and wetland) was of little influence in alternate models, suggesting good portability of these methods across a broad variety of forest conditions. A $k = n$ cross-validation approach produced an average root mean square error (RMSE) of 7.2% for the best model with no systematic bias. In addition to contributing to the identification of sound methodological practices, these results successfully reconciled the conceptual differences between *canopy closure*, measured through the use of ground-based optical tools, and *canopy cover*, captured remotely with lidar, revealing a direct linear relationship between the two attributes.

Résumé. Nous avons réalisé une série d'expériences empiriques conçue pour raffiner le traitement des données de profileur lidar (détection et télémétrie par la lumière) à retours discrets afin d'estimer la fermeture du couvert à travers une large gamme de conditions forestières dans le centre ouest de l'Alberta, au Canada. Les trois conclusions méthodologiques suivantes ont été dérivées : (*i*) une nouvelle méthode de segment de ligne basée sur le ratio de la distance du segment de l'étage supérieur par rapport à la distance totale a mieux performé que les techniques alternatives de comptage de points décrites précédemment dans la littérature; (*ii*) un seuil absolu de 1,4 m de l'étage supérieur–l'étage inférieur a permis de générer globalement les meilleurs modèles et a semblé bien s'appliquer à toute une gamme de types de forêt; et (*iii*) la stratification selon la composition des espèces (feuillus, conifères et forêt mixte) ou le régime d'humidité du sol (hautes terres et terres humides) avait peu d'influence dans les modèles alternatifs suggérant ainsi la portabilité de ces méthodes dans une large gamme de conditions forestières. Une approche de validation croisée $k = n$ a permis de produire une valeur moyenne de RMSE de 7,2 % pour le meilleur modèle et cela sans biais systématique. En plus de contribuer à l'identification de pratiques méthodologiques saines, ces résultats ont permis de réconcilier avec succès les différences conceptuelles entre la *fermeture du couvert*, mesurée par le biais d'outils optiques au sol, et le *couvert*, capturé par télédétection lidar, révélant ainsi une relation linéaire directe entre ces deux attributs.
[Traduit par la Rédaction]

Introduction

Although forest inventories are collected primarily for strategic planning and operational management purposes, they are also relied upon to support a broad variety of additional reporting and modeling activities, including the assessment of carbon stocks (e.g., Grierson et al., 1992; Tate et al., 1997; Clark, 2002; Karjalainen et al., 2002; Hu et al., 2007; Woodall and Liknes, 2008; Wulder et al., 2008), biodiversity status (e.g.,

Noss, 1990; Parthasarathy and Karthikeyan, 1997; Ayyappan and Parthasarathy, 1999; Mani and Parthasarathy, 2005; Travaglini et al., 2007; Winter et al., 2008), and habitat conditions (e.g., Franklin et al., 2001; 2002; Bond et al., 2004; Schulte et al., 2005; Zielinski et al., 2006). Common to each of these applications is a need for forest structural attribution, including canopy closure. Although the current trend in forest inventory practice is towards an enhanced focus on capturing an extensive range of vegetation attributes (Siry et al., 2005),

A.J. McLane[1] and G.J. McDermid. Foothills Facility for Remote Sensing and GIScience, Department of Geography, University of Calgary, 2500 University Drive NW, Calgary, AB T2N 1N4, Canada.

M.A. Wulder. Pacific Forestry Centre, Canadian Forest Service, Natural Resources Canada, 506 West Burnside Road, Victoria, BC V8Z 1M5, Canada.

[1]Corresponding author (e-mail: ajmclane@ucalgary.ca).

updates often occur over long time intervals, and spatial coverage can be coarse or incomplete (Lund, 2004). Light detection and ranging (lidar) provides a means of estimating and subsequently extending forest inventory attributes for the purpose of producing spatially exhaustive information on detailed aspects of forest structure (e.g., Hudak et al., 2002; Wulder and Seemann, 2003), which can contribute to the mitigation of forest inventory or other data coverage limitations.

Canopy closure is defined as the proportion of the sky hemisphere obscured by vegetation when viewed from a single point (Jennings et al., 1999) and is of particular interest to forest ecologists studying the variability of understory environments (Minckler et al., 1973; Vitousek and Denslow, 1986), light regimes (Canham et al., 1990; 1994), reforestation patterns (Gray and Spies, 1996), sapling growth rates (Wright et al., 1998), and regeneration (Nicotra et al., 1999). Canopy closure has traditionally been measured in the field through the use of ground-based optical tools such as the spherical densiometer (Lemmon, 1956; 1957) and hemispherical photography (Frazer et al., 2001).

Similar to *canopy closure* is the concept of *canopy cover*, defined as the proportion of the forest floor covered by the vertical projection of the tree crowns (Jennings et al., 1999). It has been used to predict stand volume (Philip, 1994) and estimate precipitation interception by forests (Molicova and Hubert, 1994). Canopy cover can be measured in the field by looking vertically upwards and recording whether or not the forest obscures the sky. Thus, given a number of observations, the proportion of sky obscured gives an estimate of forest canopy cover. These measurements can be taken without instrumentation (e.g., Vales and Bunnell, 1988), but instruments such as the gimbal balance (Walter and Soos, 1962) and the sighting tube (Johansson, 1985) are commonly used to avoid nonvertical bias and enhance repeatability. Although individual observations of canopy cover are straightforward and efficient, a very large sample size (in excess of 100 observations) is required to obtain accurate estimates over an area, regardless of its size (Jennings et al., 1999). The amount of time required to collect these estimates suggests a limited role for field measurements of canopy cover in forestry mapping.

The most important difference between canopy closure and canopy cover is that the former measurements integrate information over the sky hemisphere above one point on the ground, whereas the latter assesses the presence of canopy vertically above a sample of points (Jennings et al., 1999). Despite the conceptual differences between the two, both attributes are descriptors of canopy density and structure and should be related. A strong relationship would permit the use of canopy cover estimates derived from lidar for describing and explaining canopy closure measurements made on the ground. Development of this relationship is important to address because logistical constraints can often limit the acquisition of field-based canopy closure measurements, and the acquisition of canopy cover in the field is a resource-heavy endeavor that can be prohibitively expensive. In addition, an improved understanding and reconciliation of the two parameters would

be valuable to confirm the relationship between the two distinct but closely related attributes.

It is common practice to acquire scanned data from small-footprint, discrete-return lidar systems, but these systems are also capable of obtaining transects (profiles). Lidar profiles may also be collected from dedicated systems (e.g., Nelson et al., 2003). The key difference between profiles and scanned datasets is that profiles constitute a single corridor of returns at nadir along the flight path, whereas scanned datasets have a dispersion of points over a given scan angle off nadir and therefore result in greater areal coverage (Lim et al., 2003). Although scanning datasets certainly have their utility, the extended time of acquisition and the large physical size of datasets (and therefore increased computing power and lengthy preprocessing and postprocessing) place practical limitations on the size of study areas that can be reasonably managed. Under many situations, profiling datasets provide an attractive alternative, since they can be acquired when flying higher, which increases the speed of acquisition, resulting in greater coverage in a given amount of time. The configuration produces smaller data volumes per distance covered and therefore more affordable datasets.

Profiling lidar data have been used to effectively relate ground-measured biomass to lidar-generated biomass (e.g., Nelson et al., 2004; Boudreau et al., 2008) and in fusion exercises with segmented Landsat Enhanced Thematic Mapper Plus (ETM+) imagery for canopy attribute change characterization (Wulder et al., 2007). These studies demonstrate the ability to link plot-based measures with lidar profiles to enable the extrapolation of structure over large areas in a cost-effective and timely manner. However, a variety of issues must be addressed before attempting to implement these strategies consistently across large, diverse study sites. Specifically, we require a robust strategy for separating *overstory* from *understory* over diverse forest conditions and a consistent and transparent approach for selecting overstory–understory thresholds. Also, although previous studies have calculated canopy cover as the fraction of canopy returns over a given unit area (Nelson et al., 1984; Morsdorf et al., 2006; Chasmer et al., 2008) or the number of returns per height category above a threshold (Ritchie et al., 1992; 1995; 1996) with varying levels of success, unexplained variation in the relationship between lidar-derived attributes and field measurements suggests that more work on the topic remains. In addition, it would be useful to establish the utility of stratification across differing forest types as a function of moisture conditions and (or) canopy occlusion of the lidar signal.

The goal of the research reported in this paper was to identify, develop, and communicate a methodological strategy for processing discrete-return profiling lidar data for the purpose of estimating canopy closure across a wide range of forest conditions in western Alberta, Canada. The work involved addressing four specific research objectives related to the handling and processing of profiling lidar data: (1) determining the best methodological approach for extracting estimates of canopy cover from lidar transects;
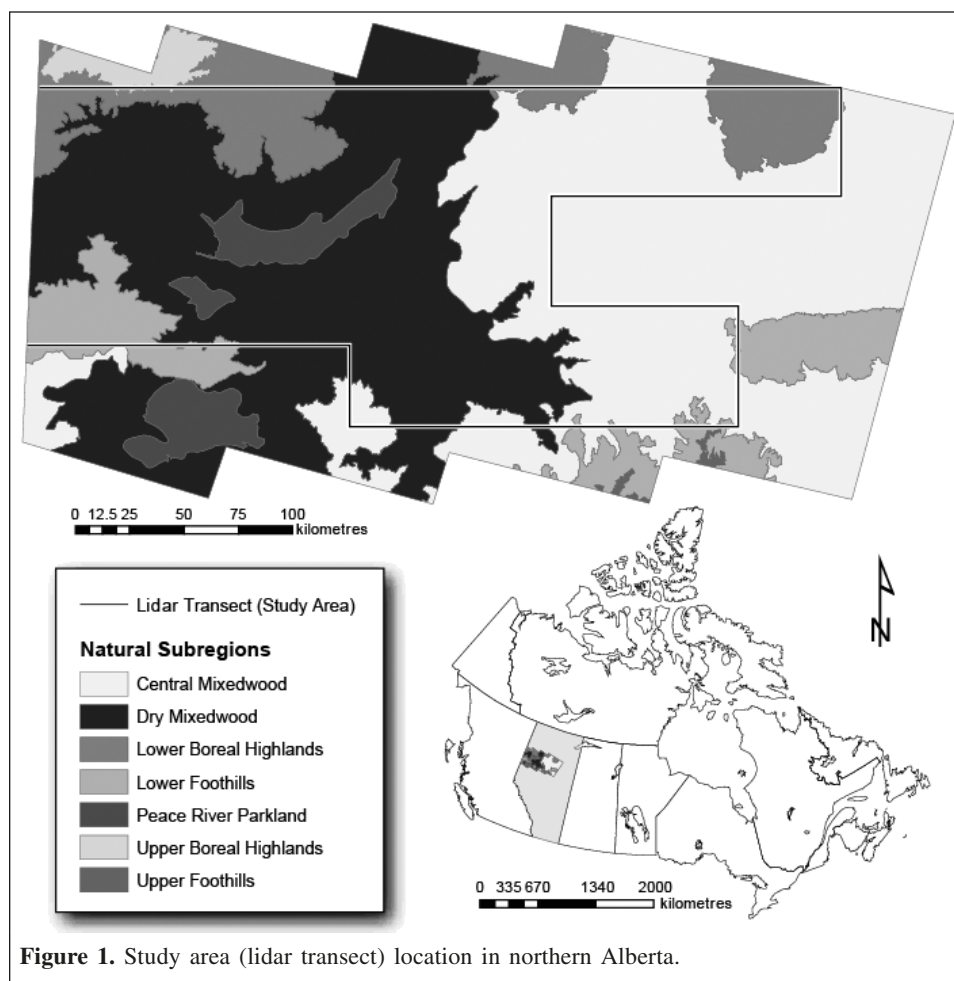
Figure 1. Study area (lidar transect) location in northern Alberta.

(2) establishing a transparent and reliable means for identifying the threshold strategy for separating overstory and understory points; (3) evaluating the value of stratification by species composition (hardwood, softwood, and mixedwood) and (or) moisture regime (upland and wetland) over diverse forest types; and (4) examining the strength of the relationship between lidar-derived measurements of canopy cover and field-based estimates of canopy closure.

Methods

Study area

The study area for this research is within the spatial extent of a 1200 km long lidar transect located in west-central Alberta, Canada, acquired on 19 and 20 August 2006 (**Figure 1**). The flight path was designed to sample the structural diversity of forests present in the area and traversed seven natural subregions defined by the Natural Regions Committee (2006): Central Mixedwood, Dry Mixedwood, Lower Boreal Highlands, Lower Foothills, Peace River Parkland, Upper Boreal Highlands, and Upper Foothills (**Table 1**). The forested areas of the transect contain hardwood species such as balsam poplar (*Populus balsamifera*), trembling aspen (*Populus tremuloides*),

and white birch (*Betula papyrifera*) and softwood species such as black spruce (*Picea mariana*), lodgepole pine (*Pinus contorta*), and white spruce (*Picea glauca*). Mixed stands are also common (Resource Information Management Branch, 2005), as are both upland and wetland areas.

Lidar data acquisition and preprocessing

Distance ranges from lidar (in reference to sea level) were acquired along the transect using a Riegl USA LMS-Q140i-80 sensor on board a fixed-wing aircraft flown by Laser Imaging Technologies of Calgary, Alberta, Canada. The LMS-Q140i-80 sensor is a discrete-return device that uses a rotating polygon mirror to scan the target surface in a parallel manner, up to a maximum scan angle range of 80°. The pulse is considered a class 1 laser (human-eye safe), has a beam divergence of approximately 3 mrad, and is emitted at 10 kHz at a near-infrared wavelength (900 nm). The average absolute measurement accuracy is typically ±0.025 m, the distance-dependent error is ≲20 ppm, and the relative measurement accuracy is ±0.003 m (Riegl USA, 2002). The mean flying altitude was approximately 150 m. From the instrument parameters specified and flight characteristics followed, a point density of approximately 2.5 returns/m^2 was achieved, and the

Table 1. Characteristics of the natural subregions present in the study area.

Natural subregion	Geology and landforms	Vegetation
Lower Foothills	Rolling topography consisting of moraine deposits over folded bedrock; extensive organic deposits in valleys and wet depressions	White spruce, black spruce, lodgepole pine, balsam fir, aspen poplar, paper birch, balsam poplar, buffaloberry, juniper, Labrador tea, fireweed, feathermoss, dwarf birch, and peat moss
Upper Foothills	Strongly rolling topography with frequent bedrock outcrops; ground moraine over bedrock with some colluvium on steep terrain	White spruce, black spruce, lodgepole pine, buffaloberry, bunchberry, Labrador tea, fireweed, feathermoss, dwarf birch, and peat moss
Central Mixedwood	Gently undulating plains with minor inclusion of hummocky uplands	Aspen poplar, white spruce, jack pine, black spruce, green alder, northern rice grass, Rocky Mountain fescue, dryland sedges, plains wormwood, Saskatoon berry, bearberry, blueberry, prickly rose, wild lily of the valley, and hairy wild rye
Dry Mixedwood	Low-relief topography consisting of ground moraine and sandy outwash plain	Aspen poplar, balsam polar, white spruce, balsam fir, jack pine, black spruce, tamarack, cranberry, red-osier dogwood, feathermoss, bearberry, lichen, Labrador tea, peat moss, and sedge
Lower Boreal Highlands	Gentle to strongly sloping lower elevations; some undulating and hummocky upland areas	Lodgepole pine, jack pine, hybrids, aspen poplar, white birch, green alder, bearberry, Labrador tea, blueberry, and bog cranberry
Upper Boreal Highlands	Steeply sloping dissected plateaus and undulating and hummocky upper plateau surfaces; medium-textured glacial tills present	Lodgepole pine, jack pine, hybrids, aspen poplar, balsam poplar, black spruce, white birch, green alder, beaked willow, Scouler's willow, bearberry, Labrador tea, blueberry, and bog cranberry
Peace River Parkland	Broad, gently rolling plains with scattered uplands and steep river valleys	White spruce, aspen poplar, balsam polar, sedges, western snowberry, wood rose, intermediate oat grass, old man's whiskers, low goldenrod, western porcupine grass, pasture sage, columbia needle grass, June grass, green needle grass, and pale comandra

mean footprint diameter was approximately 45 cm. Only first returns were used in this analysis, since it has been established that the combination of first and last returns does not provide any additional information regarding field-measured canopy cover than does first returns alone (Morsdorf et al., 2006). Initial processing of the lidar range files and global positioning system (GPS) was handled by Laser Imaging Technologies, with the final data delivery containing LAS files of x, y, and z coordinates and incidence angles. To obtain a narrow lidar profile along the study area transect, we thinned the data to select only those returns with an incidence angle of less than one quarter of a degree, a threshold that maintained the density inherent in the scanned lidar data. We assumed that there was minimal systematic bank-angle-dependent variation from the fixed-wing aircraft. This assumption was bolstered by examining the extent of the lidar profiles in the x and y domain only, a process that revealed very little deviation from centre.

Field data collection and processing

The lidar data acquired for this research were supported by spatially and temporally coincident ground estimates of canopy closure obtained in the field using hemispherical photography. We selected field sites using a stratified random sampling approach (Husch et al., 2003) to reduce sampling bias and limited our efforts to homogeneous stands at least 1 ha in size located between 60 and 300 m of known access features in an attempt to account for positional errors and reduce logistical constraints surrounding travel. Field site locations were measured using a Garmin GPSMAP 60 hand-held device, with positional errors recorded at each location not exceeding 3 m, according to error information provided by the device. It should be noted that according to the US Forest Service, the true GPS error for a hand-held system similar to this one was in the range of 6.4–37.2 m, depending on system configuration and forest canopy structure (Chamberlain, 2002), although this was not verified. To ensure the acquisition of a sample that represented the range of forest structural conditions present, we matched the pattern of tasseled cap wetness values (Kauth and Thomas, 1976) extracted from Landsat thematic mapper imagery of the sample sites to the distribution of wetness values observed across the entire study area, under the assumption that wetness had a positive relationship with structural complexity of the forest canopy and the optical depth of water in leaves (Hunt, 1991; Cohen and Spies, 1992; Cohen et al., 1995), and thereby provided an effective gradient across which to sample. Field sites were distributed proportionally across each natural subregion occurring within the transect, subject to the constraints described.

Our ground protocol involved measurements of vegetation composition (tree species), moisture regime (upland, wetland),

and forest structure (canopy closure and basal area) using standard forest mensuration techniques at 70 field sites. Upon arrival at a given site, five hemispherical photographs were taken at systematically distributed measurement stations located within a 30 m × 30 m plot, such that images were acquired 17 m from the plot centre in each of the intermediate cardinal directions. We used a Nikon CoolPix 8700 digital camera with an FC-E8 fish-eye lens, with most of the photography performed during diffuse-light conditions. The few images that were taken in direct light conditions were processed separately, with the sun masked out. The time of acquisition for each photograph was recorded and used to stratify images prior to processing, following the recommendations of Zhang et al. (2005). The five digital photographs in each plot were analyzed in WinSCANOPY (Regent Instruments Inc., Ottawa, Ont.), where gap fraction was calculated to estimate canopy closure. The software performs a classification of grey-level values using a threshold between plant and sky pixels selected using an image histogram. Gap fraction calculations were restricted to the area within 40° from zenith maximum. The five resulting canopy closure estimates were then averaged to determine a single measurement for each field site. Tree species composition was calculated as the proportion of species by basal area in a variable-area prism sweep conducted from the centre of each plot (Husch et al., 2003) and categorized as softwood (greater than 70% conifer species), hardwood (greater than 70% broadleaf species), or mixedwood (between 31% and 69% hardwood or softwood species).

Data integration and analysis

We extracted 70 lidar profile segments 30 m in length for processing, each one corresponding spatially to the location of the sample plots measured in the field. We inspected Canadian digital elevation data (CDED) at our plot locations and found that over 95% of plot locations had slopes less than 3°, indicating that ground elevation was stable over field site locations. Under this assumption, we defined the understory surface elevation for each profile segment as the minimum elevation recorded over that interval. We scaled the remaining elevation values on the basis of this minimum value, thereby generating a representative vertical profile of the local area running through each sample site (see Wulder et al., 2007 for additional details).

To determine the best methods for characterizing canopy closure from profiling lidar data (objective 1), we compared three different processing techniques: (*i*) a histogram method outlined by Ritchie et al. (1992), (*ii*) a point-count method described by Nelson et al. (1984), and (*iii*) a line-segment method described herein. The histogram method of Ritchie et al. determines canopy cover by counting the number of laser measurements in a height category and dividing by the total number of laser measurements for a given segment of the profile, thereby providing a frequency distribution of cover values at given height categories. The point-count method

described by Nelson et al. calculates canopy cover simply by dividing the number of canopy hits by the total number of hits.

The line-segment method introduced here interpolates the distance between lidar points and sums the length of the line segments above and below an overstory–understory threshold, an approach fundamentally different from previous efforts that focused only on counting the number of returns. The procedure is illustrated graphically in **Figure 2**. If the locations of p1 and p2 are known, then the distances *a* and *b* can be calculated through subtraction, after which distances *c* and *d* can be calculated using basic trigonometry. Once determined, the lengths representing overstory segments (distance *c*) and understory segments (distance *d*) are tallied along the entire transect. The ratio of overstory vegetation segments to total distance can be considered a lidar-derived horizontal projection of the vertical structure of the canopy and used to estimate canopy cover. It is important to note that the line-segment method does not assume an even distribution of returns, since returns are treated as locations with *x* and *z* coordinates (*z* being height above an understory surface), and uses trigonometry to determine the horizontal projection of the vertical structure of the profile. The placement of returns does not affect trigonometry calculations and therefore does not affect the estimation of canopy cover.

To determine the best local threshold for separating overstory from understory (objective 2), we tested each processing strategy using a range of both absolute and proportional height thresholds. The absolute thresholds ranged from 0.1 m to the maximum height of the canopy, and the proportional thresholds ranged from 1% to 100% of maximum canopy height. We summarized the relationships between the lidar-based canopy cover estimates and the corresponding field measurements using a series of simple linear regression models. We also appraised the value of stratifying the sample sites by species composition (i.e., softwood, hardwood, and mixedwood) or site moisture regime (i.e., wetland and upland) (objective 3) by repeating the regression analyses on both stratified and unstratified datasets. The result was a total of 17 100 regression models that provided the empirical results required for subsequent performance evaluation.

The best model for each trial was determined following Ferguson (1981), who outlined a strategy for testing the significance of the difference between two correlation coefficients. We calculated *t* values for comparison with critical *t* values to make a decision regarding the null hypothesis of no difference between the models. Once we established the top overall strategy for estimating canopy cover from profiling lidar in our study area (i.e., the best methodological procedure, top thresholding tactic, and best stratification approach), we performed a leave-one-out cross-validation procedure as has been done in other studies including machine learning (Elisseeff and Pontil, 2003), bioinformatics (Simon et al., 2003), and, recently, remote sensing (e.g., Van Der Heijden et al., 2007; Brovelli et al., 2008).

Leave-one-out cross-validation is a special case of the *k*-fold cross-validation method (Stone, 1974; Geisser, 1975), a process that involves portioning the original dataset into *k* subsets of
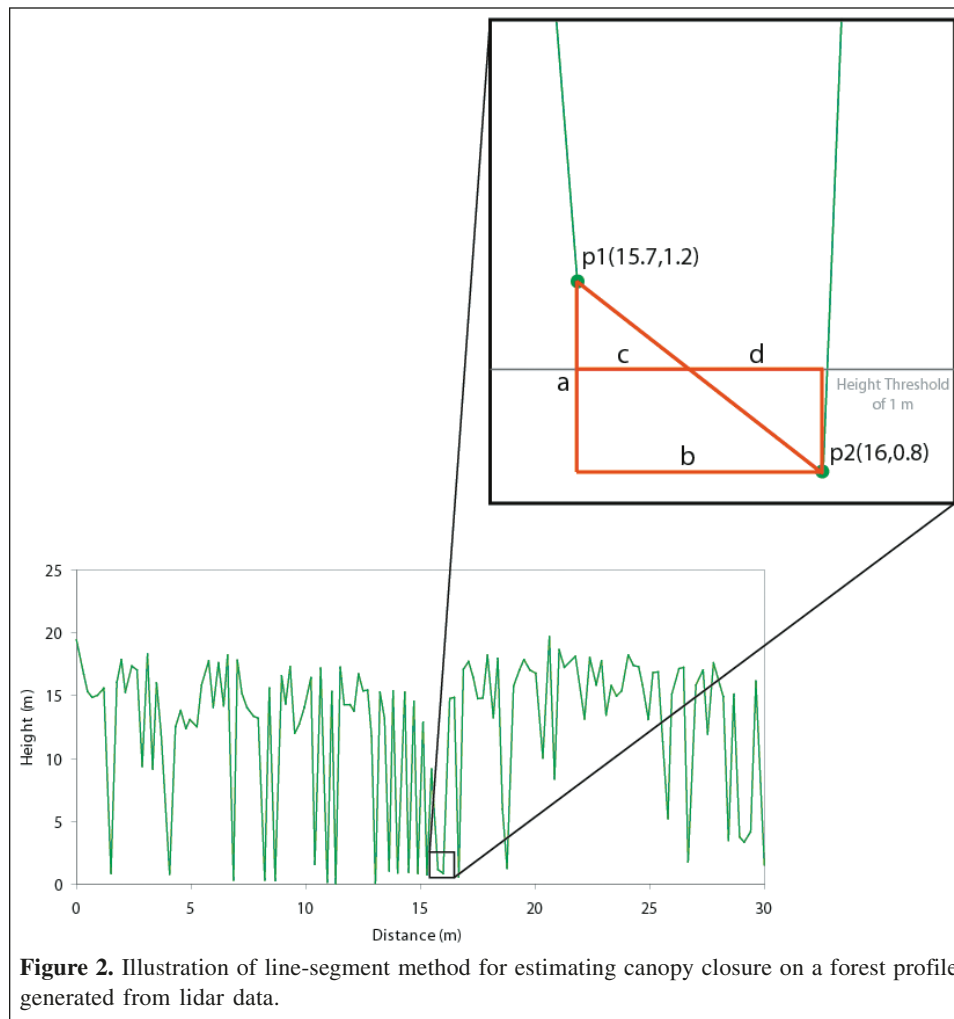
Figure 2. Illustration of line-segment method for estimating canopy closure on a forest profile generated from lidar data.

equal size, whereby the model is trained k times using each subset as validation and the remaining data as training. The leave-one-out cross-validation is a k-fold cross-validation computed where $k = n$, with n being the size of the original dataset. The purpose here was to provide an independent assessment of our capacity to estimate field-based measurements of canopy closure with lidar-derived measurements of canopy cover (objective 4). We calculated root mean square errors (RMSEs) for each k and reported the average. We estimated bias in the best model by calculating systematic error (SE) using an approach recommended by Wu (2004):

$$\text{SE} = \frac{\sum_{i=1}^{N}(\hat{c}_i - c_i)}{N} \tag{1}$$

where $\hat{c}_i$ is the estimated value of canopy cover, c_i is the reference value of canopy cover, and N is the total number of samples. SE calculates the effects of systematic errors: a positive systematic value would indicate overestimation and bias, and a negative value would indicate underestimation and bias.

Results

Histogram, point-count, and line-segment approaches to calculating canopy cover

Line plots of the coefficients of determination r^2 arising from the regression analyses performed between lidar-derived cover and hemispherical photography derived estimates of canopy closure using the line-segment, point-count, and histogram methods are found in **Figures 3**, **4**, and **5**, respectively. The plots illustrate the observed trends in the strength of each method and corresponding field measurements, given varying stratification strategies and overstory–understory thresholds. The key values from these plots are summarized in **Table 2**, illustrating that the line-segment method represented the best approach overall for estimating canopy closure, outperforming both the histogram and point-count methods described previously in the literature. The highest peak in the r^2 distribution of the line-segment method was 0.76 (RMSE = 6.9%), higher than the $r^2 = 0.73$ peaks observed for both the histogram and point-count techniques (RMSE = 7.2%) (**Table 2**). In addition to having the highest peak overall, the
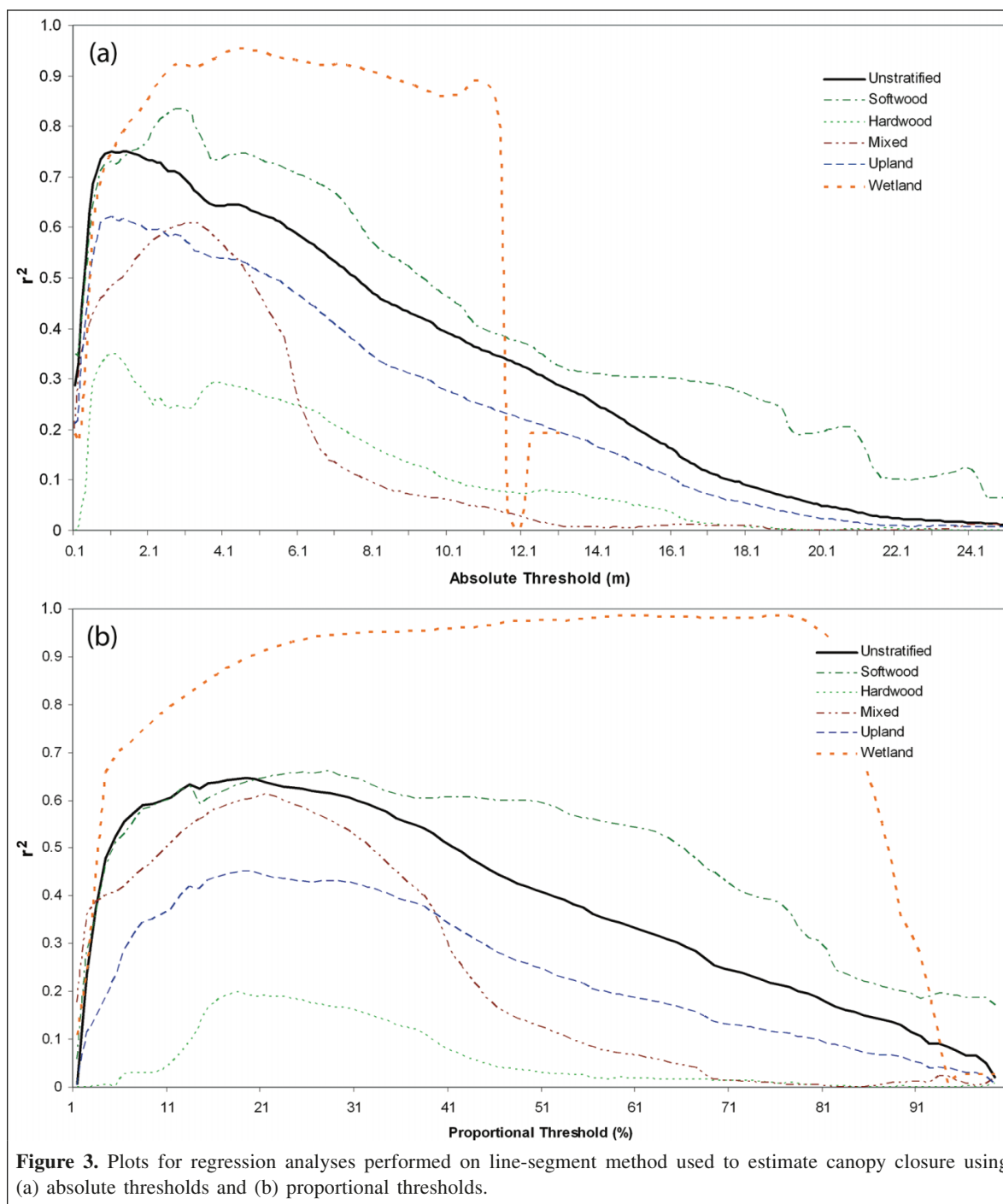
Figure 3. Plots for regression analyses performed on line-segment method used to estimate canopy closure using (a) absolute thresholds and (b) proportional thresholds.

line-segment method consistently generated the best r^2 values for any given threshold and stratification configuration (see **Figures 3–5**). The largely comparable coefficients of determination and RMSE values for all the canopy closure measurement approaches indicate the robustness of the relationship present between the field-measured canopy closure and the lidar-based remote measurements.

Separating overstory from understory

We found that an absolute threshold of 1.4 m was the best height for separating overstory from understory over the range of forest conditions present in our study area (**Figures 3–5**), generating a peak r^2 value of 0.76 (RMSE = 6.9%; **Table 2**). Although the 1.4 m threshold was not universally optimal (for example, a proportional threshold of 68% (of total canopy
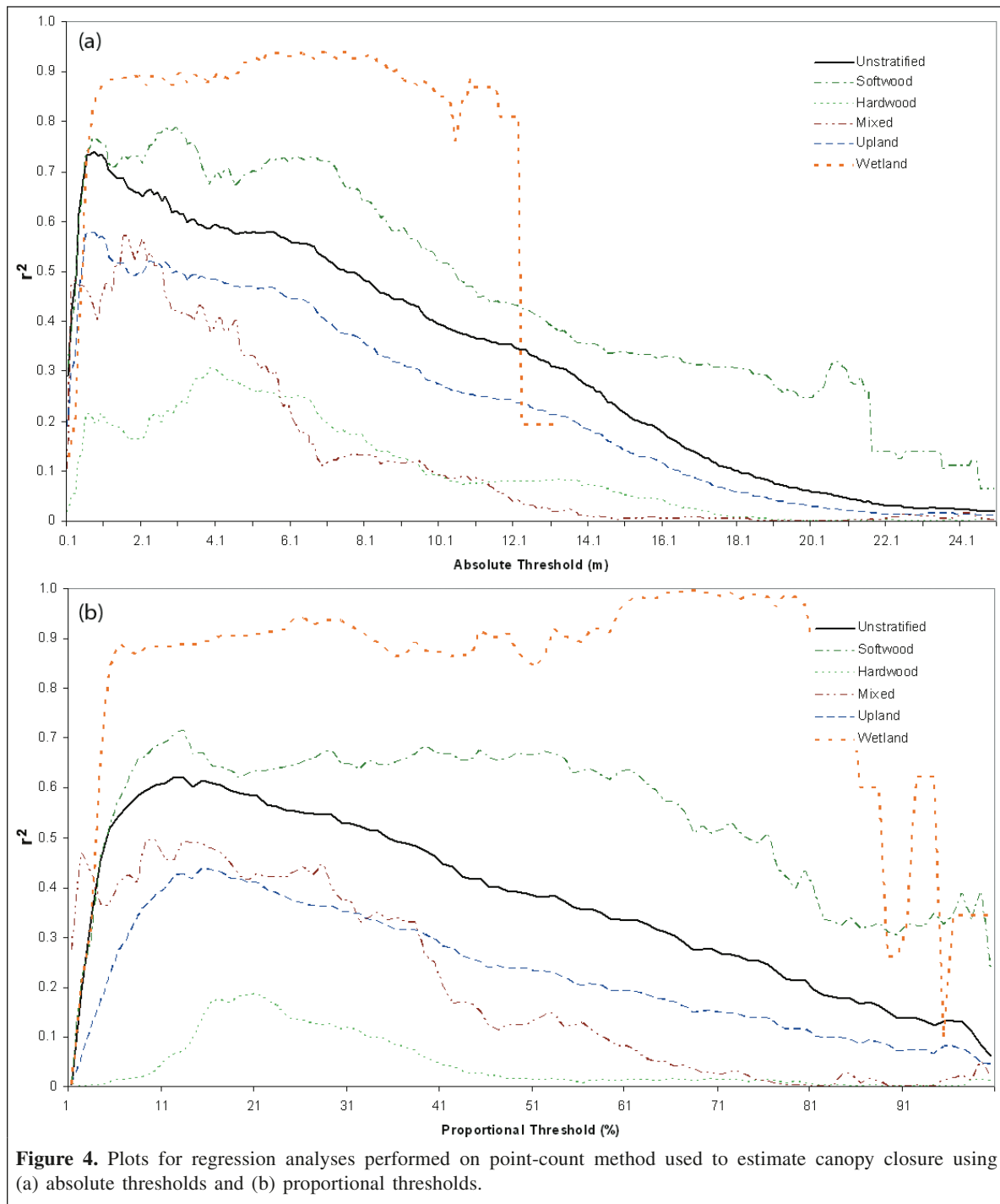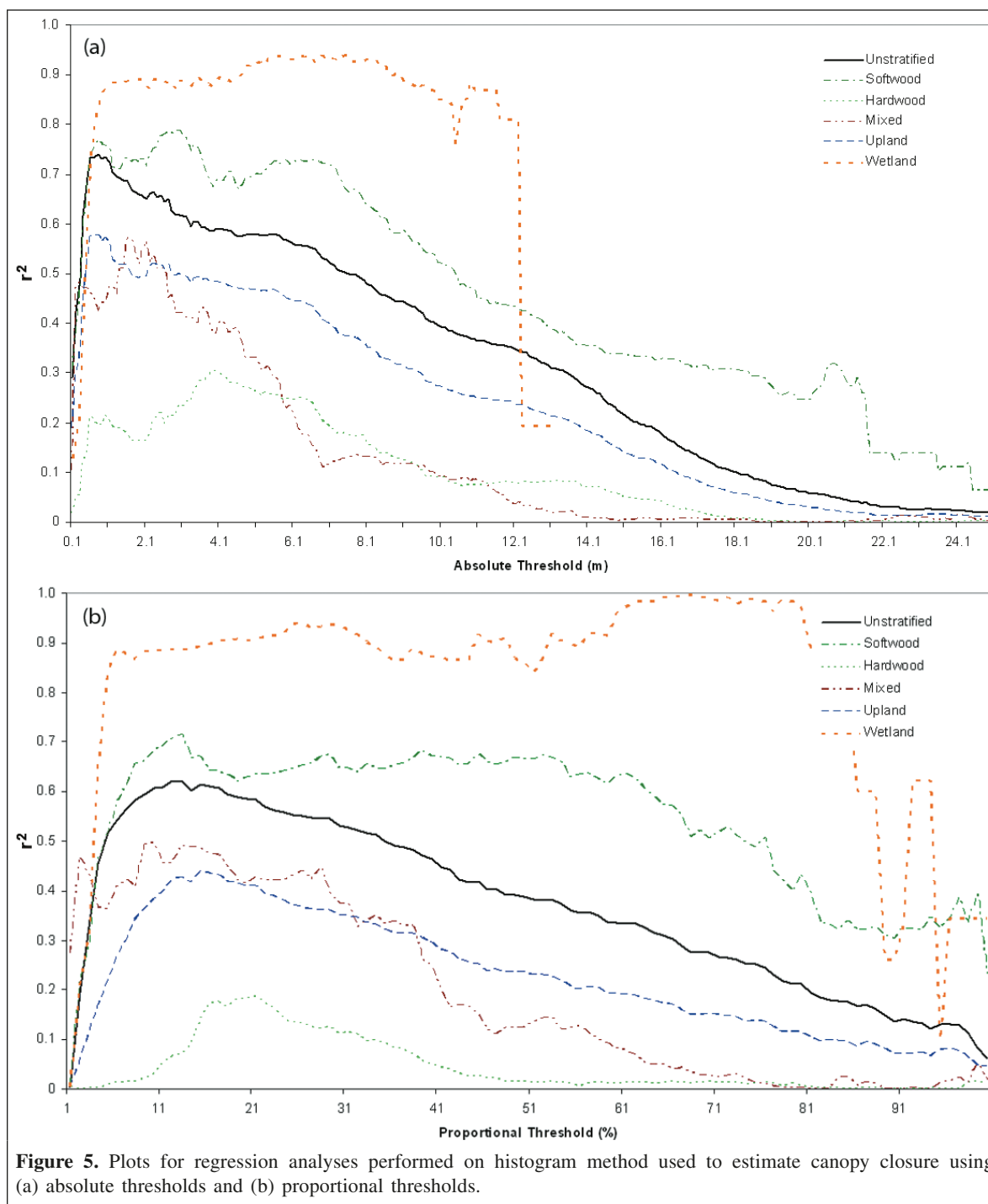
Figure 4. Plots for regression analyses performed on point-count method used to estimate canopy closure using (a) absolute thresholds and (b) proportional thresholds.

height) produced better models in wetland areas), the approach provided the best results overall. The top proportional threshold was observed at 15% of the total canopy height and generated an r^2 value of 0.61 (RMSE = 8.7%).

The value of stratification

Regression models from the unstratified datasets tended to perform better than those generated by models stratified by either moisture regime (upland, wetland) or species composition (softwood, hardwood, mixedwood) (**Figures 3–5**; **Table 2**), suggesting that there is limited value to be gained through stratifying by either of these two criteria. As noted previously, the highest peak in the r^2 distribution of the line-segment method was 0.75 for unstratified data (**Figure 3a**). Although a close inspection of **Figure 3a** reveals that some individual strata performed better, specifically softwood with a

Figure 5. Plots for regression analyses performed on histogram method used to estimate canopy closure using (a) absolute thresholds and (b) proportional thresholds.

peak r^2 of 0.83 (RMSE = 5.6%) and wetland with a peak r^2 of 0.92 (RMSE = 2.1%), the stratified trials were judged inferior overall because of the lower r^2 values from the corresponding hardwood ($r^2 = 0.24$, RMSE = 13.1%), mixedwood ($r^2 = 0.61$, RMSE = 8.8%), and upland ($r^2 = 0.57$, RMSE = 9.1%) strata.

Accuracy and bias

Results of determining bias for the best model overall (line-segment method using an absolute threshold of 1.4 m and

unstratified) revealed an SE of 0.003%. Therefore, no significant systematic bias estimation exists for the best model. The RMSE calculated between observed and predicted values of canopy closure reveals a value of 6.9%.

Cross-validation results

Canopy cover was calculated from the corresponding lidar segments running though the plots using the line-segment method and a 1.4 m threshold. We found the average RMSE

Table 2. Summary table showing the peak r^2 values observed in the line plots describing the regression analyses relating lidar-derived canopy cover to field-estimated canopy closure.

Method	Threshold type	Threshold value	Field data stratification	r^2	RMSE (%)
Line segment	Absolute	1.4 m	Unstratified	0.76	6.9
Point count	Absolute	0.8 m	Unstratified	0.73	7.2
Histogram	Absolute	0.8 m	Unstratified	0.73	7.2
Line segment	Proportional	19%	Unstratified	0.65	8.7
Point count	Proportional	15%	Unstratified	0.61	8.7
Point count	Proportional	68%	Wetland	0.99	1.8
Histogram	Proportional	68%	Unstratified	0.61	8.7
Line segment	Absolute	1.1 m	Upland	0.62	8.6
Line segment	Absolute	1.1 m	Hardwood	0.35	12.3
Line segment	Absolute	2.9 m	Softwood	0.84	5.9
Line segment	Proportional	21%	Mixedwood	0.61	8.8

from the $k = n$ trials to be 7.2%. These results indicate that a strong relationship exists between lidar- and field-derived estimates, suggesting that canopy closure can be reliably obtained across the study area using profiling lidar data processed with the line-segment method.

Discussion

Our empirical experiments demonstrated the better performance of the line-segment method over established point-count and histogram techniques described previously in the literature for extracting estimates of canopy closure from profiling lidar transects obtained across a diverse sample of forest types in west-central Alberta, Canada. The results suggest that interpolating between height samples of a target surface yields a more accurate estimate of vertical canopy projection than that obtained from discrete samples alone. This finding supports previous studies that have employed canopy height profiles and canopy height models to derive other relevant forest inventory information (e.g., Harding et al., 2001; Popescu et al., 2002; Van Aardt et al., 2008). However, unexplained variability remains and may be related to a variety of factors, including a theoretical difference between canopy cover and canopy closure, system noise – measurement errors, GPS error, and the nature of the first-return lidar data. In future research, investigation on the use of multiple returns in a similar type of analysis would be beneficial, since more information on the structure of the canopy may be available. It would also be useful to verify the true accuracy of hand-held GPS units compared with that of survey-grade units as outlined by Chamberlain (2002) and determine how errors influence the estimation of canopy closure from lidar.

We found the best strategy for separating overstory from understory in this analysis was through the use of an absolute threshold, rather than a proportional one. This was somewhat unexpected, given the range of forest canopy heights existing in the study area and the theoretical flexibility offered by a variable threshold. It is possible that our assumption of flat topography within a given 30 m transect introduced errors in

maximum tree height estimations that negatively impacted the ability to calculate stable proportional thresholds. However, the success of absolute thresholds mirrors that reported by other authors (Morsdorf et al., 2006; Chasmer et al., 2008), though their studies were not comparative in nature. Our observation that an absolute threshold of 1.4 m was the most suitable over most of the sites is likely due to the fact that this is the height at which the coincident hemispherical photographs were obtained. Our field protocol dictated the acquisition of hemispherical photographs at a height of 1.3 m (breast height), giving a difference of just 10 cm.

The superior performance of the proportional threshold at 66% of the maximum height in wetland areas is an interesting observation. Wetland areas in this region can be dominated by low, sparse black spruce stands; low, dense black spruce stands; and tall, open, and mature black spruce and larch stands. The variability between wetland forest regions makes these areas more challenging to characterize and therefore more suitable to a proportional threshold.

Stratifying sample plots on the basis of moisture regime (upland, wetland) or species composition (softwood, hardwood, mixedwood) revealed no benefit overall. This again was somewhat surprising, given our expectation that methods and (or) thresholds that performed well in softwood forests might not be readily transferable to hardwood or mixedwood environments, given the difficulty in penetrating hardwood canopies with lidar documented elsewhere (e.g., Nelson et al., 2004). However, although we believe that the hardwood penetration issue accounts for at least some of the unexplained variability in our best models, the hardwood forests in our study area did not appear to be dense enough to prevent at least one return from penetrating to the understory over the 30 m horizontal profile distance used to characterize each sample plot. The apparent ability of at least one return to penetrate to understory level (aided no doubt by the relatively high density of the data) rendered the methods transferable across all the forest types in the sample, a finding that could simplify the processing strategies adopted by future projects in these forests.

A strong, statistically significant, and unbiased linear relationship was found to exist between lidar-derived estimates of canopy cover and field-based measurements of canopy closure, despite the theoretical differenced between the two attributes. As a result, these findings indicate that profiling lidar systems can be successfully employed to estimate the widely used attribute of canopy closure in forests typified by those in the present study area, despite the fact that these instruments actually acquire measurements that are more suitable for measuring canopy cover. This finding supports previous studies that have related canopy cover from lidar to gap fractions measured from hemispherical photography in other environments (e.g., Riaño et al., 2004; Takahashi et al., 2006; Hanssen and Solberg, 2007), though the issue had not been tested before explicitly. This is an important result, since it affirms the ability of lidar to generate fine-scale estimates of this important structural descriptor that might not be otherwise possible using traditional field methods over large, diverse study areas. In addition, these findings could serve to reduce the demand for expensive ground-based observations of canopy closure in locations where high-density, small-footprint or profiling lidar data are available.

Conclusion

A line-segment method for processing discrete-return profiling light detection and ranging (lidar) data was found to provide the best and most consistent strategy for estimating canopy closure over diverse forest types in west-central Alberta, Canada. The strategy works best when applied using an absolute overstory–understory threshold of 1.4 m, and we recommend that future applications use breast height (approximately 1.4 m) as an absolute threshold for calculating overstory canopy closure. The strategy appeared to extend well across the full variety of forests sampled in this study, suggesting that little value is to be gained through prestratification on the basis of species composition (hardwood, softwood, and mixedwood) or moisture regime (upland and wetland). These results successfully reconcile the conceptual differences between canopy closure, which is measured in situ through optical instruments, and canopy cover, which is measured remotely by lidar. This finding is important, since it confirms the ability of lidar instruments to estimate an attribute strongly related to plant growth and survival characteristics, thereby extending the benefits of efficiency, repeatability, and accessibility offered by remote sensing technology. Our results suggest that the strategies reported in this research identify the utility of profiling lidar for the measurement of canopy closure, which can be successfully applied to the creation of fine-scale, spatially exhaustive estimates of canopy closure over large, diverse study areas, aiding the mitigation of forest inventory or data coverage limitations, and supporting the goals of forest monitoring and reporting programs.

References

Ayyappan, N., and Parthasarathy, N. 1999. Biodiversity inventory of trees in a large-scale permanent plot of tropical evergreen forest at Varagalaiar, Anamalais, Western Ghats, India. *Biodiversity and Conservation*, Vol. 8, No. 11, pp. 1533–1554.

Bond, M.L., Seamans, M.E., and Gutierrez, R.J. 2004. Modeling nesting habitat selection of California spotted owls (*Strix occidentalis occidentalis*) in the central Sierra Nevada using standard forest inventory metrics. *Forest Science*, Vol. 50, No. 6, pp. 773–780.

Boudreau, J., Nelson, R.F., Margolis, H.A., Beaudoin, A., Guindon, L., and Kimes, D.S. 2008. Regional aboveground forest biomass using airborne and spaceborne LiDAR in Quebec. *Remote Sensing of Environment*, Vol. 112, pp. 3876–3890.

Brovelli, M.A., Crespi, M., Fratarcangeli, F., Giannone, F., and Realini, E. 2008. Accuracy assessment of high resolution satellite imagery orientation by leave-one-out method. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 63, pp. 427–440.

Canham, C.D., Denslow, J.S., Platt, W.J., Runkle, J.R., Spies, T.A., and White, P.S. 1990. Light regimes beneath closed canopies and tree-fall gaps in temperate and tropical forests. *Canadian Journal of Forest Research*, Vol. 20, No. 5, pp. 620–631.

Canham, C.D., Finzi, A.C., Pacala, S.W., and Burbank, D.H. 1994. Causes and consequences of resource heterogeneity in forests — interspecific variation in light transmission by canopy trees. *Canadian Journal of Forest Research*, Vol. 24, No. 2, pp. 337–349.

Chamberlain, K. 2002. *Performance testing of the Garmin GPSMAP 76 global positioning system receiver.* USDA Forest Service, Vallejo, Calif. GPS Information and Receiver Performance Reports [online]. Available from www.fs.fed.us/database/gps/mtdc/map76/garmin_gpsmap_76_rev.pdf [cited 9 April 2009].

Chasmer, L., Hopkinson, C., Treitz, P., McCaughey, H., Barr, A., and Black, A. 2008. A lidar-based hierarchical approach for assessing MODIS fPAR. *Remote Sensing of Environment*, Vol. 112, pp. 4344–4357.

Clark, D.A. 2002. Are tropical forests an important carbon sink? Reanalysis of the long-term plot data. *Ecological Applications*, Vol. 12, No. 1, pp. 3–7.

Cohen, W.B., and Spies, T.A. 1992. Estimating structural attributes of Douglas-fir western hemlock forest stands from Landsat and SPOT imagery. *Remote Sensing of Environment*, Vol. 41, No. 1, pp. 1–17.

Cohen, W.B., Spies, T.A., and Fiorella, M. 1995. Estimating the age and structure of forests in a multi-ownership landscape of western Oregon, USA. *International Journal of Remote Sensing*, Vol. 16, No. 4, pp. 721–746.

Elisseeff, A., and Pontil, M. 2003. Leave-one-out error and stability of learning algorithms with applications. In *Advances in learning theory: methods, models and applications*. Edited by J.A.K. Suykens, G. Horvath, S. Basu, C. Micchelli, and J. Vandewalle. IOS Press, Amsterdam. NATO Science Series III: Computer and Systems Sciences, Vol. 190, pp. 111–130.

Ferguson, G.A. 1981. *Statistical analysis in psychology and education*. 5th ed. McGraw-Hill, New York. 185 pp.

Franklin, S.E., Stenhouse, G.B., Hansen, M.J., Popplewell, C.C., Dechka, J.A., and Peddle, D.R. 2001. An integrated decision tree approach (IDTA) to mapping landcover using satellite remote sensing in support of grizzly bear habitat analysis in the Alberta Yellowhead ecosystem. *Canadian Journal of Remote Sensing*, Vol. 27, No. 6, pp. 579–592.

Franklin, S.E., Peddle, D.R., Dechka, J.A., and Stenhouse, G.B. 2002. Evidential reasoning with Landsat TM, DEM and GIS data for landcover classification in support of grizzly bear habitat mapping. *International Journal of Remote Sensing*, Vol. 23, No. 21, pp. 4633–4652.

Frazer, G.W., Fournier, R.A., Trofymow, J.A., and Hall, R.J. 2001. A comparison of digital and film fisheye photography for analysis of forest canopy structure and gap light transmission. *Agricultural and Forest Meteorology*, Vol. 109, No. 4, pp. 249–263.

Geisser, S. 1975. The predictive sample reuse method with applications. *Journal of the American Statistical Association*, Vol. 70, pp. 320–328.

Gray, A.N., and Spies, T.A. 1996. Gap size, within-gap position and canopy structure effects on conifer seedling establishment. *Journal of Ecology*, Vol. 84, No. 5, pp. 635–645.

Grierson, P.F., Adams, M.A., and Attiwill, P.M. 1992. Estimates of carbon storage in the aboveground biomass of Victoria's forests. *Australian Journal of Botany*, Vol. 40, No. 4–5, pp. 631–640.

Hanssen, K.H., and Solberg, S. 2007. Assessment of defoliation during a pine sawfly outbreak: calibration of airborne laser scanning data with hemispherical photography. *Forest Ecology and Management*, Vol. 250, pp. 9–16.

Harding, D.J., Lefsky, M.A., Parker, G.G., and Blair, J.B. 2001. Laser altimeter canopy height profiles: methods and validation for closed-canopy, broadleaf forests. *Remote Sensing of Environment*, Vol. 76, pp. 283–297.

Hu, H.Q., Yuan-Chun, L., and Jiao, Y. 2007. Estimation of the carbon storage of forest vegetation and carbon emission from forest fires in Heilongjiang Province, China. *Journal of Forestry Research (Harbin)*, Vol. 18, No. 1, pp. 17–22.

Hudak, A.T., Lefsky, M.A., Cohen, W.B., and Berterretche, M. 2002. Integration of lidar and Landsat ETM+ data for estimating and mapping forest canopy height. *Remote Sensing of Environment*, Vol. 82, pp. 397–416.

Hunt, E.R. 1991. Airborne remote-sensing of canopy water thickness scaled from leaf spectrometer data. *International Journal of Remote Sensing*, Vol. 12, No. 3, pp. 643–649.

Husch, B., Miller, C.I., and Beers, T.W. 2003. *Forest mensuration*. 4th ed. McGraw-Hill, New York.

Jennings, S.B., Brown, N.D., and Sheil, D. 1999. Assessing forest canopies and understorey illumination: canopy closure, canopy cover and other measures. *Forestry*, Vol. 72, No. 1, pp. 59–73.

Johansson, T. 1985. Estimating canopy density by the vertical tube method. *Forest Ecology and Management*, Vol. 11, pp. 139–144.

Karjalainen, T., Pussinen, A., Liski, J., Nabuurs, G.J., Erhard, M., Eggers, T., Sonntag, M., and Mohren, G.M.J. 2002. An approach towards an estimate of the impact of forest management and climate change on the European forest sector carbon budget: Germany as a case study. *Forest Ecology and Management*, Vol. 162, No. 1, pp. 87–103.

Kauth, R.J., and Thomas, G.S. 1976. The tasseled cap — a graphic description of the spectral–temporal development of agricultural crops as seen by LANDSAT. In *Proceedings of the Symposium on Machine Processing of Remotely Sensed Data*, 29 June – 1 July 1976, Purdue University, West Lafayette, Ind. Edited by P.H. Swain, D.B. Morrison, and D.E. Parks. IEEE, New York. Vol. 4B, pp. 41–51.

Lemmon, P.E. 1956. A spherical densiometer for estimating forest overstory density. *Forestry Science*, Vol. 2, pp. 415–420.

Lemmon, P.E. 1957. A new instrument of measuring forest overstory density. *Journal of Forestry*, Vol. 55, pp. 667–669.

Lim, K., Treitz, P., Wulder, M.A., St-Onge, B., and Flood, M. 2003. LiDAR remote sensing of forest structure. *Progress in Physical Geography*, Vol. 27, No. 1, pp. 88–106.

Lund, H.G. 2004. *Considerations for developing U.S. standard definitions of forest and rangeland*. Forest Information Services, Gainesville, Va. 108 pp.

Mani, S., and Parthasarathy, N. 2005. Biodiversity assessment of trees in five inland tropical dry evergreen forests of peninsular India. *Systematics and Biodiversity*, Vol. 3, No. 1, pp. 1–12.

Minckler, L.S., Woerheide, J.D., and Schlesinger, R.C. 1973. *Light, soil moisture, and tree reproduction in hardwood forest openings*. US Department of Agriculture, St. Paul, Minn. USDA Forest Service, North Central Forest Experiment Station, Research Paper NC-89.

Molicova, H., and Hubert, P. 1994. Canopy influence on rainfall fields: microscale structure in tropical forests. *Journal of Applied Meteorology*, Vol. 33, No. 12, pp. 1464–1467.

Morsdorf, F., Kotz, B., Meier, E., Itten, K., and Allgower, B. 2006. Estimation of LAI and fractional cover from small footprint airborne laser scanning data based on gap fraction. *Remote Sensing of Environment*, Vol. 104, pp. 50–61.

Natural Regions Committee. 2006. *Natural regions and subregions of Alberta*. Compiled by D.J. Downing and W.W. Pettapiece. Natural Regions Committee, Alberta Department of Environment, Edmonton, Alta. Publication T/852.

Nelson, R., Krabill, W., and Maclean, G. 1984. Determining forest canopy characteristics using airborne laser data. *Remote Sensing of Environment*, Vol. 15, No. 3, pp. 201–212.

Nelson, R., Parker, G., and Hom, M. 2003. A portable airborne laser system for forest inventory. *Photogrammetric Engineering & Remote Sensing*, Vol. 69, No. 3, pp. 267–273.

Nelson, R., Short, A., and Valenti, M. 2004. Measuring biomass and carbon in Delaware using an airborne profiling LIDAR. *Scandinavian Journal of Forest Research*, Vol. 20, pp. 283–284.

Nicotra, A.B., Chazdon, R.L., and Iriarte, S.V.B. 1999. Spatial heterogeneity of light and woody seedling regeneration in tropical wet forests. *Ecology*, Vol. 80, pp. 1908–1926.

Noss, R.F. 1990. Indicators for monitoring biodiversity — a hierarchical approach. *Conservation Biology*, Vol. 4, No. 4, pp. 355–364.

Parthasarathy, N., and Karthikeyan, R. 1997. Plant biodiversity inventory and conservation of two tropical dry evergreen forests on the Coromandel coast, south India. *Biodiversity and Conservation*, Vol. 6, No. 8, pp. 1063–1083.

Philip, M.S. 1994. *Measuring trees and forests*. 2nd ed. CAB International, Wallingford, UK.

Popescu, S.C., Wynne, R.H., and Nelson, R.F. 2002. Estimating plot-level tree heights with lidar: local filtering with a canopy-height based variable window size. *Computers and Electronics in Agriculture*, Vol. 37, pp. 71–95.

Resource Information Management Branch. 2005. *Alberta vegetation inventory standards, version 2.1.1*. Resource Information Management Branch, Alberta Sustainable Resource Development, Edmonton, Alta. 11 pp.

Riaño, D., Valladares, F., Condés, S., and Chuvieco, E. 2004. Estimation of leaf area index and covered ground from airborne laser scanner (LiDAR) in two contrasting forests. *Agricultural and Forest Meteorology*, Vol. 124, pp. 269–275.

Riegl USA. 2002. *Technical documentation and user instructions: mirror scanner LMS-Q140i-80*. Riegl USA, Orlando, Fla.

Ritchie, J.C., Everitt, J.H., Escobar, D.E., Jackson, T.J., and Davis, M.R. 1992. Airborne laser measurements of rangeland canopy cover and distribution. *Journal of Range Management*, Vol. 45, No. 2, pp. 189–193.

Ritchie, J.C., Humes, K.S., and Weltz, M.A. 1995. Laser altimeter measurements at Walnut-Gulch Watershed, Arizona. *Journal of Soil and Water Conservation*, Vol. 50, No. 5, pp. 440–442.

Ritchie, J.C., Menenti, M., and Weltz, M.A. 1996. Measurements of land surface features using an airborne laser altimeter: the HAPEX-Sahel experiment. *International Journal of Remote Sensing*, Vol. 17, No. 18, pp. 3705–3724.

Schulte, L.A., Pidgeon, A.M., and Mladenoff, D.J. 2005. One hundred fifty years of change in forest bird breeding habitat: estimates of species distributions. *Conservation Biology*, Vol. 19, No. 6, pp. 1944–1956.

Simon, R., Dobbin, K., and McShane, L.M. 2003. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *JNCI Journal of the National Cancer Institute*, Vol. 95, No. 1, pp. 14–18.

Siry, J.P., Cubage, F.W., and Ahmed, M.R. 2005. Sustainable forest management: global trends and opportunities. *Forest Policy and Economics*, Vol. 7, pp. 551–561.

Stone, M. 1974. Cross-validation and multinomial prediction. *Biometrika*, Vol. 61, No. 3, pp. 509–515.

Takahashi, T., Yamamoto, K., Miyachi, Y., Senda, Y., and Tsuzuku, M. 2006. The penetration rate of laser pulses transmitted from a small footprint airborne LiDAR: a case study in closed canopy, middle-aged pure sugi (*Cryptomeria japonica* D. Don) and hinoki cypress (*Chamaecyparis obtuse* Sieb. et Zucc.) stands in Japan. *Journal of Forest Research*, Vol. 11, pp. 117–123.

Tate, K.R., Giltrap, D.J., Claydon, J.J., Newsome, P.F., Atkinson, I.A.E., Taylor, M.D., and Lee, R. 1997. Organic carbon stocks in New Zealand's terrestrial ecosystems. *Journal of the Royal Society of New Zealand*, Vol. 27, No. 3, pp. 315–335.

Travaglini, D., Barbati, A., Chirici, G., Lombardi, F., Marchetti, M., and Corona, P. 2007. Forest inventory for supporting plant biodiversity assessment — forest BIOTA data on deadwood monitoring in Europe. *Plant Biosystems*, Vol. 141, No. 2, pp. 222–230.

Vales, D.J., and Bunnell, F.L. 1988. Comparison of methods for estimating forest overstory cover. I. Observer effects. *Canadian Journal of Forest Research*, Vol. 18, No. 5, pp. 606–609.

Van Aardt, J.A.N., Wynne, R.H., and Scrivani, J.A. 2008. LiDAR-based mapping of forest volume and biomass by taxonomic group using structurally homogeneous segments. *Photogrammetric Engineering & Remote Sensing*, Vol. 74, No. 8, pp. 1033–1044.

Van Der Heijden, G.W.A.M., Clevers, J.G.P.W., and Schut, A.G.T. 2007. Combining close-range and remote sensing for local assessment of biophysical characteristics of arable land. *International Journal of Remote Sensing*, Vol. 28, No. 24, pp. 5485–5502.

Vitousek, P.M., and Denslow, J.S. 1986. Nitrogen and phosphorus availability in treefall gaps of a lowland tropical rain-forest. *Journal of Ecology*, Vol. 74, No. 4, pp. 1167–1178.

Walter, J., and Soos, J. 1962. *The Gimbal sight for the projection of crown radius*. Faculty of Forestry, University of British Columbia, Vancouver, B.C. Research Note 39.

Winter, S., Chirici, G., McRoberts, R.E., Hauk, E., and Tomppo, E. 2008. Possibilities for harmonizing national forest inventory data for use in forest biodiversity assessments. *Forestry*, Vol. 81, No. 1, pp. 33–44.

Woodall, C.W., and Liknes, G.C. 2008. Relationships between forest fine and coarse woody debris carbon stocks across latitudinal gradients in the United States as an indicator of climate change effects. *Ecological Indicators*, Vol. 8, No. 5, pp. 686–690.

Wright, E.F., Coates, K.D., Canham, C.D., and Bartemucci, P. 1998. Species variability in growth response to light across climatic regions in northwestern British Columbia. *Canadian Journal of Forest Research*, Vol. 28, No. 6, pp. 871–886.

Wu, C. 2004. Normalized spectral mixture analysis for monitoring urban composition using ETM+ imagery. *Remote Sensing of Environment*, Vol. 93, pp. 480–492.

Wulder, M.A., and Seemann, D. 2003. Forest inventory height update through the integration of lidar data with segmented Landsat imagery. *Canadian Journal of Remote Sensing*, Vol. 29, No. 5, pp. 536–543.

Wulder, M.A., Han, T., White, J.C., Sweda, T., and Tsuzuki, H. 2007. Integrating profiling LIDAR with Landsat data for regional boreal forest canopy attribute estimation and change characterization. *Remote Sensing of Environment*, Vol. 110, pp. 123–137.

Wulder, M.A., White, J.C., Fournier, R.A., Luther, J.E., and Magnussen, S. 2008. Spatially explicit large area biomass estimation: three approaches using forest inventory and remotely sensed imagery in a GIS. *Sensors*, Vol. 8, No. 1, pp. 529–560.

Zhang, Y., Chen, J., and Miller, J. 2005. Determining digital hemispherical photograph exposure for leaf area index estimation. *Agricultural and Forest Meteorology*, Vol. 133, pp. 166–181.

Zielinski, W.J., Truex, R.L., Dunk, J.R., and Gaman, T. 2006. Using forest inventory data to assess fisher resting habitat suitability in California. *Ecological Applications*, Vol. 16, No. 3, pp. 1010–1025.