Lidar calibration and validation for geometric-optical modeling with Landsat imagery

Authors:

Gang Chen [a], Michael A. Wulder [a,*], Joanne C. White [a], Thomas Hilker [b], and Nicholas C. Coops [c]

Affiliations:

[a] *Canadian Forest Service (Pacific Forestry Centre), Natural Resources Canada, 506 West Burnside Road, Victoria, British Columbia, V8Z 1M5, Canada*
[b] *Biospheric Sciences Branch, NASA Goddard Space Flight Center, 8800 Greenbelt Road, Greenbelt, Maryland, 20771, USA*
[c] *Integrated Remote Sensing Studio, Department of Forest Resources Management, Faculty of Forestry, University of British Columbia, 2424 Main Mall, Vancouver, British Columbia, V6T 1Z4, Canada*

* Corresponding author: Mike Wulder. E-mail address: mwulder@nrcan.gc.ca

Pre-print of published version.
Reference:
Chen, G., M.A. Wulder, J.C. White, T.H. Hilker, and N.C. Coops. (2012). Lidar calibration and validation for geometric-optical modeling with Landsat. Remote Sensing of Environment. (Accepted May 30, 2012).

DOI.
http://dx.doi.org/10.1016/j.rse.2012.05.026

Disclaimer:

The PDF document is a copy of the final version of this manuscript that was subsequently accepted by the journal for publication. The paper has been through peer review, but it has not been subject to any additional copy-editing or journal specific formatting (so will look different from the final version of record, which may be accessed following the DOI above depending on your access situation).

Abstract

There is a paucity of detailed and timely forest inventory information available for Canada's large, remote northern boreal forests. The Canadian National Forest Inventory program has derived a limited set of attributes from a Landsat-based land cover product representing circa year 2000 conditions. Of the required inventory attributes, forest vertical structure (e.g., tree height) is critical for terrestrial biomass and carbon modeling and to date, is unavailable for these remote areas. In this study, we develop a large-area, fine-scale (25 m) mapping solution to estimate tree height (mean, dominant, and Lorey's height) across Canada's northern forests by integrating lidar data (representing 0.27% of the study area), and Landsat imagery (representing 100% of the study area), using a geometric-optical modeling technique. First, spectral mixture analysis (SMA) was used to extract image endmembers and generate fraction images. Second, lidar data were used to calibrate the inverted geometric-optical model by adjusting the model's three key fractional inputs: sunlit crown, sunlit background, and shade fraction, based upon the SMA derived images. The heterogeneity of the study area, spanning 2.16 million ha, made it challenging to directly and accurately decompose mixed Landsat image pixels into the canopy and background fractions used for the Li-Strahler geometric-optical model inversion. As a result we developed a novel method to use the lidar plot data to facilitate the calculation of these fractions in an accurate and automated manner. The average estimation errors for mean, dominant, and Lorey's height were 4.9 m, 4.1 m, and 4.7 m, respectively when compared to the lidar data, with the best result achieved using dominant tree height, where the average error was 3.5 m for over 80% of the forested area. Using this approach of optical remotely sensed data calibrated and validated with lidar height estimates we generate and evaluate wall-to-wall estimates of tree height that can subsequently be used as inputs for biomass and carbon modeling.

Keywords: Tree height; Landsat; Lidar plots; Li-Strahler geometric-optical model; Large-area

1. Introduction

Boreal forests are one of world's largest biomes, responsible for approximately 22% of terrestrial carbon stored in the global forests (Pan et al., 2011). Compared to tropical forests, the boreal forests contain almost twice the amount of carbon per unit area (Potapov et al., 2011), most of which is contained in soil organic matter. Forest management and reporting activities require accurate, timely, and consistent information which typically supports forest inventories. However, much of the boreal forest ecosystem occurs at high latitudes (45º to 65º) (Brandt, 2009) where human access is limited and industrial forest management is typically not practiced due to low forest productivity, small trees, and long distances to markets (Wulder et al., 2007). While there is some elasticity to the managed forest area constraints such as fuel costs, presence of road networks, and timber values, climate and productivity remain the limiting factors in these northern forests and the subarctic and cold continental climate is a major impediment to human activities in this biome (Potapov et al., 2008). Outside of the southern boreal where industrial activities are practiced, natural ecosystem processes tend to dominate the northern portions of the Canadian boreal forest (Andrew et al., 2012), with low populations and few roads (Wulder et al., 2007). As such, forest inventory in the northern boreal is logistically challenging, labor intensive, and expensive. For example, Andersen et al. (2009) reported an average cost of $6,000 USD to establish one ground plot on the Kenai Peninsula of Alaska. Similar costs have been found in Canada, based on plot installation activities for the National Forest Inventory (NFI). As a result, an alternate means of collecting data and characterizing forest conditions is highly desired.

The boreal forests of Canada extend from Newfoundland across Canada into Alaska, with its northern extent generally considered to be at the southern limit of the tundra, and its southern extent coincident with the northern limit of temperate forests (recognizing variations at both extremes according to local topography, climate, and edaphic conditions) (Brandt, 2009). Much of British Columbia is considered hemi-boreal, and other areas excluded from the boreal include eastern Maritime forests, central Canadian mixed woods, and the agricultural zones of the west. To date, large areas of Canada's boreal forests lack detailed and timely forest inventory information as they are not monitored by provincial or territorial resource management agencies. In these areas, Canada's federal NFI program has relied on a Landsat-based land cover product generated by the Earth Observation for Sustainable Developments of forests (EOSD) project, representing circa 2000 conditions, to provide relevant information for this northern portion of the Canadian boreal (Gillis et al., 2005; Wulder et al., 2008b). While current protocols have integrated very high spatial resolution satellite data as a surrogate for aerial photography in the north (Falkowski et al., 2009; Mora et al., 2010), additional information is required to provide spatially explicit forest structural information.

Forest vertical structural information is critical for forest monitoring and inventory. While optical imagery has been shown to be able to effectively capture forest area (Haapanen et al., 2004), cover types (Wulder et al., 2008b), and change in area (Stehman, 2009), knowledge of forest presence/absence is only part of the information required by monitoring and inventory programs. Forest vertical structure provides information on "how much", to complement the more readily produced "where is" information. Knowledge of tree height, for instance, provides information on volume, biomass, and—in conjunction with modeling—age. As such, forest vertical structure (especially tree height) is a critical NFI attribute. Tree height is known to correlate strongly with the biomass and carbon in forest ecosystems (Flanagan and Johnsen, 1995; Lefsky et al., 2002; Ni-Meister et al., 2010). Although the term "tree height" is used, apart from dedicated single tree applications (e.g., Gougeon and Leckie, 2006; Forzieri et al., 2009), it is typically an areal generalization of height that is generated, either stand height portrayed within an inventory polygon, or an average height attributed to a particular image pixel. In this research (as typical elsewhere) we refer to tree height as related to the areal generalization of height over a pixel. Further, from an inventory perspective, stand height is defined as the height of the leading species (i.e., the species with the greatest proportion of basal area in the stand) in the tallest horizontal stratum in the stand, with the goal of capturing the characteristics of the larger trees in a stand, indicative of the merchantable volume present. Pixel based approaches typically do not partition height by strata.

In recent years, lidar (light detection and ranging) technology flown on airborne platforms has become increasingly standardized and established for measuring forest vertical structure with a high degree of accuracy (Lim et al., 2003; Zhao et al., 2011), with notable relevance to forest management (Wulder et al., 2008a). Compared to field mensuration, lidar provides a relatively cost-efficient solution to estimate tree height; however, wall-to-wall airborne lidar surveys over large areas remain expensive (Chen and Hay, 2011a). In terms of accuracy, Næsset and Okland (2002) found multiple lidar measures of the same trees to have less variance than from multiple, manual, field measures. To reduce data acquisition costs whilst collecting useful estimates of forest vertical structure, recent efforts have focused on the development of integrated models using samples of lidar data to represent large areas (for a review, see Wulder et al., 2012), often using wall-to-wall optical remotely sensed data to create strata to support the spatial extension of estimates (e.g., Boudreau et al., 2008). Successful applications of lidar to sample and represent a population require sufficient sampling density and appropriate placement of samples, among other issues. However, the accuracy of lidar measures for predicting forest structure offers unique opportunities to also consider using the lidar attributes (calibrated against ground data) as calibration and validation data in empirical approaches to generate structural attributes from optical imagery. Further, a sample density that is too sparse for robust population level estimates may be sufficiently large and have value in conferring local structural conditions.

Following on the ideas above, lidar data can be viewed as providing plot-like information, enabling model calibration and validation. As such, models can be developed to spatially extend lidar-measured attributes over larger extents using image analysis. This type of approach has been applied by others with the research divided into two groups: those that apply parametric approaches (Chen et al., 2011; Hilker et al., 2008; Hudak et al., 2002; Hyde et al., 2006; Wulder and Seemann, 2003), and those that do not (Chen et al., 2012; McInerney et al., 2010; Stojanova et al., 2010). A typical parametric approach uses multiple regression which defines relationships between image spectral metrics and lidar-measured tree height. Although widely used and easy to interpret, this empirical approach often lacks the ability to characterize forest complexity, especially at fine spatial scales (Chen and Hay, 2011b), and most of these approaches were undertaken at the stand level, where internal stand forest structural variability has been reduced (Wulder and Seemann, 2003; Hilker et al., 2008). Alternatively, non-parametric machine learning techniques, such as support vector machines, have demonstrated superior performances over classic regression analysis for estimating tree height (Chen and Hay, 2011b; Zhao et al., 2011). However, most of these tools, as used in remote sensing studies of forests, have a black-box nature that can prohibit users from defining (or understanding) the relationship between model inputs and outputs.

While the integration of lidar and optical data through these parametric and non-parametric approaches to generating and extending forest vertical attributes over large areas has been well established, geometric-optical (GO) approaches provide an alternative which has proved useful for estimation of vegetation biophysical parameters (Chen et al., 2000; Chopping et al., 2006; Franklin and Turner, 1992; Liang, 2007; Peddle et al., 2003; Zeng et al., 2008) from canopy reflectance. The advantage of a geometric-optical (hereafter, GO model) approach over aforementioned approaches is that these kinds of models are typically based on physical principles of the geometric structures of discontinuous canopies which, in theory, allow these approaches to be used in a more generic fashion and across ecosystems. As the most widely used GO model, the Li-Strahler model estimates the fractions of four forest components (sunlit canopy, sunlit background, shaded canopy, and shaded background) as a function of tree size (e.g., height, crown dimensions) and tree density (Li and Strahler, 1985; 1992). The premise of the approach is that, if the fractions of these components (sunlit and shaded crown and background) can be accurately extracted from the remotely sensed imagery, it is possible to predict tree height through inversion of the model.

Implementation of the Li-Strahler model in a northern boreal environment presents unique considerations. First, northern boreal forest stands are distributed on less productive sites with relatively low tree densities, resulting in background components that typically exhibit high spectral heterogeneity, as they are typically characterized by bare soils and/or low vegetation. This heterogeneity causes difficulties in the accurate extraction of the four forest fractions used for GO models. Second, there are typically few field plots available in these areas, confounding model calibration and validation. Third, previous studies have found that the Li-Strahler model gives more reliable estimates of forest cover than height, and that the non-linearity of the model inversion process may introduce errors into height estimates (Woodcock et al., 1994; 1997); however, these studies were limited by the amount of data available for independent calibration. Lidar data provides a potential means to address these issues, by providing a large number of detailed forest structural measurements with which to calibrate the inverted Li-Strahler GO model by adjusting the input fractions, thereby enabling a more accurate estimation of tree height.

The primary objective of this study is therefore to develop a large-area, fine-scale (25 m) mapping approach for estimation of tree height in the Canada's northern forest environment. To do so, we propose to integrate sample of airborne lidar plots and Landsat data, within the Li-Strahler GO model. Here, lidar plots refer to a defined, spatially discrete areas similar in size to a ground plot from which tree characteristics (i.e., tree height, crown dimensions), as well as metrics generalizing the plot-wide vertical structural conditions, are extracted from lidar data.

2. Materials

2.1. Study area

While the focus of our study is the northern boreal we develop and demonstrate the approach at a study site in western Manitoba (centered at: 55°54′N, 99°30′W), Canada, covering an area of approximately 2.16 million ha (Fig. 1). As a typical region in the Boreal Shield, the largest of Canada's ecozones (Environment Canada, 2000), the area is characterized by forest, wetlands, and lakes, with wildfire the dominant agent of disturbance (Stocks et al., 2003). The forest is dominated by conifers, including black spruce (*Picea mariana*), white spruce (*Picea glauca*), balsam fir (*Abies balsamea*), jack pine (*Pinus banksiana*), as well as a small proportion of deciduous trees, such as paper birch (*Betula papyrifera*), trembling aspen (*Populus tremuloides*), and balsam poplar (*Populus balsamifera*). Mean annual temperatures range between -15°C in January and 17°C in July, and mean annual precipitation is around 400 mm (Environment Canada, 2000). Topographically, the site has an average elevation of 273 m above sea level, ranging from 45 m to 494 m with a mean slope of 4°.

[*INSERT FIG. 1*]

2.2. Lidar plots

Lidar data were collected from a national-scale forest lidar acquisition campaign performed during the summer of 2010, when a series of 34 individual transects with a total length of more than 24,000 km were flown across eight ecozones and 13 Universal Transverse Mercator (UTM) zones of Canada. A segment of one lidar transect (approximately 145 km long) was used to acquire forest structure variability for the study site (Fig. 1). Lidar data were acquired using a discrete return lidar system (Optech ALTM 3100). The survey specifications included a flying height of 1,200 m above ground level, a pulse repetition frequency of 70 kHz, scan angles of +/-15º, and a nominal pulse density of ~3 returns/m^2. Following the data acquisition, lidar metrics (e.g., percentile values) and CHM (canopy height model) were calculated for 625 m^2 (25-by-25 m) plots that fell within the approximately 400 m lidar swath using FUSION (McGaughey, 2000). Only non-ground returns greater than 2 m in height were used for metric calculation. There were 92,800 25-by-25 m lidar plots representing approximately 0.27% of the study area.

With plot-level lidar metrics (e.g., percentile values) and field data, Bater et al. (2011) developed multiple linear regression models to estimate three plot-level height attributes: mean height, dominant height, and Lorey's height, resulting in low RMSEs (root mean square errors) of approximately 1.5 m. Mean height is calculated as the arithmetic mean height of all trees in the plot, and dominant height is calculated as the arithmetic mean height of the four tallest trees in the plot. Lorey's height is calculated by multiplying, for each tree in the plot, the tree height by its basal area, then summing these values and dividing the total by the total basal area of the plot. Each of these height measures has importance to forest inventory and management, hence our interest in their estimation.

After excluding non-treed areas from the sample, 81,045 lidar plots remained for use as reference data. Summary statistics for these plots are provided in Table 1. We extracted 10% of the lidar plots for calibration, and another 10% from the remainder for validation. We used a stratified random selection strategy (Husch et al., 2002), where the 10% samples were taken from each 1.0 m height interval from the minimum to the maximum height. Compared to a simple random sampling, this method considered all tree height strata, ensuring a more reasonable representation of the population. Fig. 2

shows a comparison of three dominant tree height histograms derived from all lidar data, and the selected calibration and validation data. All three histograms are highly correlated with each other (R>0.95).

[*INSERT TABLE 1*]

[*INSERT FIG. 2*]

2.3. Landsat imagery

A Landsat-5 Thematic Mapper (TM) scene (Path 34, Row 21) of the study site was acquired on June 5, 2010 from the USGS archive in L1T format (orthocorrected). The cloud-free area of the image was used to determine the extent of the study area. Compared to low spatial resolution (i.e., 250/500/1000 m MODIS) and high spatial resolution optical satellites (e.g., 0.6/2.4 m QuickBird), Landsat (30 m) provides an appropriate resolution for balancing the need to collect fine-scale forest information with the requirements for reducing data acquisition and processing costs for large-area forest inventory and management. Compared to satellites with similar resolutions (e.g., 23 m IRS), Landsat offers free and open access to data (Wulder et al., 2011). In this study, Landsat bands 1-5 and 7 were converted to at-sensor radiances, which were then converted to top-of-atmosphere (TOA) reflectance (Chander et al., 2009). To facilitate the comparison between Landsat imagery and lidar plots, the Landsat scene was resampled from the original 30 m to 25 m (consistent with the lidar plot size) using the nearest neighbor method. The image was acquired with the solar zenith angle of 35.5º and a solar azimuth angle of 152.9º.

2.4. ASTER elevation data

The elevation data utilized in this study was collected by ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) as part of the GDEM (Global Digital Elevation Model) project (version 1). The accuracy of GDEM version 1 at the 95% confidence level is 21.31 m (RMSE 10.87 m) (ASTER GDEM Validation Team, 2009), although a lower error is expected in our study area due to the flat topography with a mean slope of 4.02 degrees. The DEM (30 m resolution) was resampled to the resolution of 25 m using the nearest neighbor method to be consistent with the resolution of Landsat image and the size of lidar plots. Aspect and slope images were also generated to represent the study area.

3. Methods

In this section, we first summarize the overall methodological flow, with dedicated subsections below providing required detail. Forested areas within the study site were delineated using a supervised classification of the Landsat imagery. Tree height information from lidar plots was used to aid in the collection of training data, by distinguishing various ground features, such as trees from non-treed low vegetation. We then calculated the forest fractional components (sunlit crown, sunlit background and shade fractions), which are critical inputs of the inverted Li-Strahler GO model. This was completed by using spectral mixture analysis (SMA) and lidar plots. To do so, the first step was to collect spectral endmembers, the reflectance spectra of the 'pure' pixels (Keshava and Mustard, 2002). This was followed by the calculation of fraction images, corresponding to the abundance of all endmembers. However, due to the high spectral variability resulting from the aforementioned heterogeneity of these forests, these SMA fractions may not accurately represent the inputs of the inverted Li-Strahler model. For example, two or more SMA endmembers may correspond to the same fractional component (e.g., a sunlit background may be represented by exposed soils, low vegetation, or the mixture thereof). Tree heights derived from lidar plots were used to calibrate the inverted Li-Strahler GO model by tuning the three fractional components, with results validated by independent lidar plots. The flowchart in Fig. 3 summarizes these steps; while the following sub-sections provide greater detail and explanation.

[*INSERT FIG. 3*]

3.1. Forest extraction

In this study, we applied a supervised maximum likelihood classification algorithm (Richards, 1999) to Landsat multispectral bands 3, 4, 5, and 7 to generate four major classes: forest, water, exposed land, and low vegetation (including shrubs, herbs and bryoids), as per Canada's EOSD circa 2000 land cover product (Wulder et al., 2008b). To reduce the impact of haze present when the image was acquired, bands 1 and 2 were excluded from our analysis (Chavez, 1988). Training data for each of the classes were selected manually. The forest and low vegetation classes were difficult to distinguish due to their spectral similarity and the lack of field measurements. To address this issue, we used a 1 m CHM derived from the lidar data to guide the selection of appropriate training samples by verifying the structure differences between trees and low vegetation (Fig. 4).

[*INSERT FIG. 4*]

3.2. Fraction calculation

Spectral mixture analysis (SMA) is one of the most popular techniques used to address the spectral heterogeneity present in remote sensing pixels, and it has been widely used in forestry applications (e.g., Peddle et al., 1999; Zeng et al., 2008). SMA estimates the proportions of pure components within each mixed pixel, which typically contains more than one ground cover type (Somers et al., 2011). In this study, SMA was applied to generate sub-pixel fraction images for the pixels in the forest class (as identified in the previous section).

One prerequisite to successful pixel unmixing using SMA is the selection of representative endmembers (Somers et al., 2011; Tompkins et al., 1997). In this study, endmembers were derived from the Landsat multispectral image (bands 3, 4, 5 and 7), rather than a spectral library, enabling ease of association with image features (Franke et al., 2009; Rashed et al., 2003). In this study, endmembers were selected using the Sequential Maximum Angle Convex Cone (SMACC) algorithm, which finds extreme vectors (i.e., endmembers) that cannot be represented by a positive linear combination of other vectors (Gruninger et al., 2004), similar to principal components analysis. A key benefit of the SMACC algorithm is that it is fast, has no requirement of *a priori* knowledge of the study area, and exhibits reliable performance, which is desirable for large-area applications. The number of endmembers was determined based on two criteria: (i) the SMACC relative error tolerance was 0.01, and (ii) the SMACC relative error began to converge markedly when additional endmembers were used. Fig. 5 shows a relative error plot, where five endmembers were finally selected, as the addition of more endmembers only marginally increased the SMACC performance, although greatly increasing the computational expense. The corresponding fraction (i.e., abundance) images were simultaneously generated based upon the assumption that the spectrum of a mixed pixel is a linear combination of the endmember spectra weighted by their area fractions, and the fractions of each pixel are constrained to sum to one (Gruninger et al., 2004):

$$H_{i,j} = \sum_{n}^{N} S_{i,n} F_{n,j}^{N} + R_{i,j}^{N} \tag{1}$$

Where, $H_{i,j}$ is the reflectance in the i-th channel of the j-th pixel, n is the endmember indices from 1 to the expansion length N, S is a matrix that contains the endmember spectra, F is a matrix that contains the fractional contribution (between 0 and 1) of each basis endmember spectrum to each pixel, and R is an error term.

[*INSERT FIG. 5*]

3.3. Geometric-optical model inversion

The Li-Strahler GO model simulates the complex relationship between sunlit and shaded canopy and background, and tree density and canopy geometric structure (h – the height from the ground to mid-crown, b – the half crown height, and r – the half crown width) (Li and Strahler, 1992). In the forward mode, the inputs are tree density and canopy geometric structure, producing outputs of the fractions of the four image components. By inverting the model, the data inputs and outputs are switched and we are able to predict canopy vertical structure (e.g., tree height: h+b) from the four fractions. Typically, shaded canopy and shaded background fractions are grouped for assessing forest canopy structure under an assumption that both components have the same reflectance (Peddle et al., 1999). It is therefore critical to obtain accurate fractions of the three components of sunlit canopy, sunlit background, and shade.

While unsupervised SMA methods such as SMACC are more practical to apply over large areas, one of the disadvantages is that the endmembers are not assigned to one of the three inputs of the inverted Li-Strahler model. Additionally, more SMA derived endmembers may represent the same GO model required image component. For example, the two types of backgrounds covered by bare soils and low vegetation have distinct reflectance, possibly leading to the definition of two endmembers. This issue can potentially be mitigated by using lidar-measured tree height to calibrate the inverted Li-Strahler model by tuning the model's input fractions. In this study, tree height was estimated in the following four main steps:

(1) We assigned the SMACC generated five endmembers to the three image components (i.e., sunlit crown, sunlit background and shade). Consequently, the corresponding five SMACC fraction images were combined to simulate the three inputs of the inverted Li-Strahler model. Since SMACC is an unsupervised method, it remains unclear which endmember should belong to which image component. To solve this issue, we evaluated all the possible assignment combinations—a total number of 150.

(2) For each assignment combination, we performed the Li-Strahler model inversion using the combined fractions, topographic data (aspect and slope images), and solar and viewing angles. Mathematically, the inversion of the Li-Strahler model is a non-linear minimization problem that can be solved through iterative adjustment of estimated a-priori inputs (Verstraete et al., 1996). The inversion problem can be defined as the minimum of a cost function C:

$$C = \sqrt{\sum_{i=1}^{n} [F_i(R) - F_i(T)]^2} \quad (2)$$

where, $F(R)$ are the fractions extracted from the spectral reflectance (i.e., combined SMA fractions), $F(T)$ represents the fractions calculated using the Li-Strahler model in forward mode with tree structure parameters (including tree height) as inputs, and n is the number of pixels. Different optimization algorithms are available; in this study, we selected a trust-region-reflective algorithm based on the interior-reflective Newton method to determine the "best" tree height (Coleman and Li, 1996; Coleman et al., 2002). Lidar measured tree structure information was also used to constrain the algorithm and avoid unrealistic results. For example, the lidar measured tree heights (Table 1) were used as *a priori* knowledge to force the model inversion to generate height estimates within the same range.

(3) The tree height estimates from all the possible fraction combinations were evaluated using lidar training plots, where the "best" combination was considered to result in the lowest average estimation error for height.

(4) The corresponding estimated tree heights were further validated using independently selected lidar validation plots.

The use of lidar plots provided an accurate way for calibrating the inverted Li-Strahler model through an automatic adjustment of the three fractional inputs. As indicated in section 2.2, the lidar calibration data included estimates of mean height, dominant height, and Lorey's height for each lidar plot. These three different measures of tree height were estimated individually from the GO model by running the process described above three times.

4. Results and Discussion

4.1. Endmember spectra and fraction images

In Fig. 6 we present the spectral reflectance of the five SMA extracted endmembers (E1, E2, E3, E4, and E5). By following the methods in Section 3.3, we found the best combination of these fraction images: sunlit crown—(E1), sunlit background—(E2+E3), and shade— (E4+E5), which were all used for estimating mean height, dominant height, and Lorey's height. Fig. 6 shows that the spectra of E1, E4 and E5 have a major difference in the NIR (near-infrared: 760-900 nm) band, where, as expected, the sunlit crown has a higher reflectivity contribution than the shade. E4 is different from E5 with a slightly higher reflectance in the NIR band. One possible reason could be that E5 was closer to the ground and it was therefore mixed with a higher percentage of low vegetation (e.g., shrub and grass) and/or soils, which typically have a higher water content than tree leaves, enabling a stronger absorption of NIR. Compared to the sunlit crown and shade endmembers, it is more apparent that E2 and E3 belong to the sunlit background, especially in the shortwave-infrared (band 5, SWIR-1: 1550-1750 nm and band 7, SWIR-2: 2080-2350 nm) bands, as they have higher spectral reflectance than those from trees and shade. The difference between E2 and E3 may be due to the distinctions in soil properties (e.g., moisture content).

[*INSERT FIG. 6*]

The fraction images (of the full study site) corresponding to the three components of sunlit crown, sunlit background, and shade are shown in Fig. 7. Sparse forest stands tended to have lower sunlit crown and shade fractions due to the lack of trees than dense forest stands, while the sunlit background fractions are higher. A typical example of the sparse forests in Fig. 7 is a large patch close to the center of the study area. For a viewing of the site, please refer to Fig. 1, where the patch is in light green tone representing a group of forest stands regenerating after wildfire. Fig. 7 illustrates that the findings from these fraction images are consistent with the assumption, as the regenerated forest stands have low sunlit crown and shade fractions and are dominated by a large portion of sunlit backgrounds.

[*INSERT FIG. 7*]

4.2. Tree height

Fig. 8 presents both the estimation errors (grey bars) and the area percentages (lines) of different tree height classes for (a) mean height, (b) dominant height, and (c) Lorey's height. The height classes (with 3-m interval) are: HT1 (2-5 m), HT2 (5-8 m), HT3 (8-11 m), HT4 (11-14 m), HT5 (14-17 m), HT6 (17-20 m), HT7 (20-23 m), HT8 (13-27 m), and HT9 (27-30 m). The 3-m interval was selected to aid in the interpretation of the estimation errors. The average estimation errors (i.e., RMSEs) are 4.9 m, 4.1 and 4.7 m for each of mean, dominant, and Lorey's height, respectively. The best model performance (i.e., lowest RMSE) was achieved for dominant tree height, which could be caused by two factors: (i) the signals received by Landsat were biased to the upper level of forest canopies; and (ii) the dominant trees have a greater likelihood of intercepting the laser pulses as noted by Popescu et al. (2002). Model performance for Lorey's height, which is mean tree height weighted by the basal areas of all trees, was better than for simply averaged mean height and worse than for dominant height.

[*INSERT FIG. 8*]

Errors were not uniform across all height classes. In particular, relatively small (i.e., HT1-2) and tall trees (HT7-9) have errors of over 6.0 m for all the three height measures (Fig. 8). A potential reason could be that small trees are typically located within sparse forest stands and the background in these stands is a major contributor to the spectral reflectance. Additionally, forest background is rarely only covered by bare soils, but rather is often a mixture of soils and low vegetation. The heterogeneous background tends to exhibit strong reflectance anisotropy, causing difficulties for the Li-Strahler GO model to estimate tree height with nadir data (Chopping et al., 2006; Gemmell, 2000). For tall trees, which are often mature, the sunlit crown tends to show a darker tone than other trees. This may have introduced errors in the fraction extraction using SMA. However, we noted that these five height classes accounted for only a small portion of the site, i.e., 2.9%, 15.4% and 8.1% for mean height, dominant height, and Lorey's height. The majority of the trees (over 80% by area) in the forested area ranged from 8 to 20 m (classes HT3-6), where the tree vertical structure was reasonably well estimated (Fig. 8). Especially for the three classes of HT4, HT5 and HT6, the height estimation errors were between 2.1 m and 3.9 m. In the case of estimating dominant tree height, the mean error was 3.5 m for 81.2% of the forested area, which shows comparable performance with other studies in the similar forest environment at the stand level. For example, Wulder and Seemann (2003) reported a height estimation error of 3.3 m using Landsat imagery to estimate lidar-measured canopy height in Saskatchewan, Canada. The average stand size used in their study was 14.2 ha. Similarly, Hilker et al. (2008) updated forest inventory attributes using high-spatial-resolution QuickBird imagery and a lidar transect in British Columbia, Canada. Their height prediction resulted in an error of 3.5 m using inventory polygons, which were typically larger than 2.0 ha in that area. More recently, Chen et al., (2012) applied similar data types of QuickBird imagery and lidar transects to estimate tree height in a Quebec study site. By incorporating machine learning techniques, they obtained an error of 3.4 m at the plot level of 0.04 ha. Compared to the estimates at the large forest stand level, height variability in our study was better retained using small plots (25-by-25 m), with the wall-to-wall estimates of tree height presented in Fig. 9. Wildfire boundaries have been overlaid on the height output to aid in illustrating the capacity of the wall-to-wall estimates to inform on stand vertical structure. Inset A in Fig. 9 shows an undisturbed forest and wetland area, where tree heights (right) typically increase with increasing distance from wetland features (as seen in the RGB Landsat image using a composite of bands 5, 4 and 3 on the left; shorter trees are clustered around wetland features). Inset B illustrates an area that was burned by wildfire in 1995. Outside the area of the fire, the natural variation in vertical structure as a function of site and topography is observed. Inside the fire perimeter, the spatial variability in fire impacts are evident, where areas of moderately tall trees that were not consumed by fire remain, along with areas of regenerating forests. Fig. 9 indicates that the height estimates from this study can be further linked with other landscape ecology research (e.g., wetlands and wildfire) as an important environmental variable.

[*INSERT FIG. 9*]

Given the results of previous research by Woodcock et al. (1994 and 1997) we had modest expectations for the capacity of a GO model driven approach to estimate tree heights, as these previous studies indicate that the GO model's estimates of forest cover are more reliable than its estimates of tree size. The iterative approach we have applied in this current research, using a large sample of plot-like data (from the lidar) to calibrate, validate, and re-calibrate as required, may have enabled the improved estimates of height achieved in this study. Furthermore, of the studies indicated above [e.g., Hilker et al. (2008) and Chen et al. (2012)], the prediction errors on height estimates are lower (i.e., 3.5 m and 3.4 m) than those reported herein, but these studies used high spatial resolution imagery. While difficult to compare across studies (e.g., the spatial precision of the estimates will differ), it is worth noting that the mapping effort per unit area is less for Landsat with the larger imaging footprint (185 × 185 km) in contrast to the high spatial resolution imagery, with typical extents of 10 × 10 km.

The intention is that for areas of Canada where spatially explicit forest inventory information is lacking, image derived estimates could fulfill this information need. For instance, the carbon budget model used by the Government of Canada to represent the forest sector—the Carbon Budget Model of the Canadian Forest Sector (CBM-CFS3)—operates using stand-based inventory data (Natural Resources Canada, 2012). The generation of pixel-based structural attributes corresponding to the required model inputs, generalized using image segmentation to replicate forest stands, may form a basis for using CBM-CFS3 over these remote locations. Currently, Canada meets national and international carbon reporting objectives by focusing on the managed forest using CBM-CFS3. Additional Earth-system science questions that encompass the entire forested area of Canada in a consistent and transparent manner follow the generation of remotely sensed inputs to aid in model parameterization.

5. Conclusions

Outside of Canada's managed forest area, such as the more northern forests, there can be a lack of detailed and timely forest inventory information. To reduce data costs while collecting large-area, fine-scale (25 m) grid-based tree height estimates over Canada's northern forests, we have developed a novel mapping solution integrating lidar plots, representing 0.27% of the study area, Landsat imagery, and the Li-Strahler geometric-optical model. As it is challenging to accurately identify the image endmembers required by the Li-Strahler model, lidar data were used to calibrate the three critical model inputs of sunlit crown, sunlit background, and shade fractions. We have evaluated the model performance for estimating three measures of plot-level tree height: mean, dominant, and Lorey's height. Based upon the results of this study, we found that the three forest components (i.e., sunlit crown, sunlit background, and shade) were spectrally heterogeneous. By applying the spectral mixture analysis to the Landsat imagery, we found five endmembers, rather than three, which best represented the conditions present in this boreal study site. The best result (lowest tree height estimation error) shows that these endmembers corresponded to one type of sunlit crown, two types of sunlit backgrounds, and two types of shade. Here, lidar plots were used as calibration data, which helped assign the endmembers to individual forest components in an accurate and automated manner. This also reduces the possibility of introducing errors caused by an inaccurate human interpretation of the endmembers. We also found that the average estimation errors were 4.9 m, 4.1, and 4.7 m for mean height, dominant height, and Lorey's height, respectively. The best result was achieved for characterizing the dominant tree height, where the average error was 3.5 m for 81.2% of the forested area. It should be noted that the non-linear GO model inversion process may have introduced errors, which can be reduced by using more accurate *a priori* knowledge to better constrain tree parameters. The coregistration between lidar data and Landsat image at the pixel level was another likely source of error due to their different data acquisition geometries. Future research will evaluate object-based approaches using polygons as basic units in tree height estimation to reduce this error. However, learning from and augmenting other studies in similar forest environments, our approach presents an advancement towards the characterization of wall-to-wall forest vertical structure over large areas using Landsat imagery and airborne (or perhaps satellite) lidar sample datasets in a time- and cost-efficient manner

Acknowledgements

This research has been funded by a Natural Sciences and Engineering Research Council (NSERC) Visiting Fellowship award to Dr. Gang Chen. Aspects of this research were undertaken as part of the "EcoMonitor: Northern Ecosystem Climate Change Monitoring" project jointly funded by the Canadian Space Agency (CSA) Government Related Initiatives Program (GRIP) and the Canadian Forest Service (CFS) of Natural Resources Canada.

References

Andersen, H. E., Barrett, T., Winterberger, K., Strunk, J., & Temesgen, H. (2009). Estimating forest biomass on the western lowlands of the Kenai Peninsula of Alaska using airborne lidar and field plot data in a model-assisted sampling design. In: *Proceedings of Extending forest inventory and monitoring, IUFRO Division 4*, Quebec City, Canada, May 19-22, 5 p.

Andrew, M.E., Wulder, M.A., & Coops, N.C. (2012). De facto protected areas of the Canadian boreal forest. *Biological Conservation*. DOI: http://dx.doi.org/10.1016/j.biocon.2011.11.029.

ASTER GDEM Validation Team (2009). *ASTER Global DEM Validation: Summary Report. June 2009*. Available online (accessed February 24, 2012): http://www.ersdac.or.jp/GDEM/E/image/ASTERGDEM_ValidationSummaryReport_Ver1.pdf.

Bater, C.W., Wulder, M.A., Coops, N.C., Hopkinson, C., Coggins, S.B., Arsenault, E., Beaudoin, A., Guindon, L., Hall, R.J., Villemaire, P., & Woods, M. (2011). Model development for the estimation of aboveground biomass using a lidar-based sample of Canada's boreal forest. *SilviLaser 2011*, Oct. 16-19, 2011, Hobart, Tasmania, Australia.

Boudreau, J., Nelson, R. F., Margolis, H. A., & Beaudoin, A. (2008). Regional aboveground forest biomass using airborne and spaceborne LiDAR in Quebec. *Remote Sensing of Environment*, 112, 3876–3890.

Brandt, J.P. (2009). The extent of the North American boreal zone. *Environmental Reviews*, 17, 101−161.

Chander, G., Markham, B.L., & Helder, D.L. (2009). Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sensing of Environment*, 113, 893−903.

Chavez, P.S. (1988). An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. *Remote Sensing of Environment*, 24, 459−479.

Chen, G., & Hay, G.J. (2011a). An airborne lidar sampling strategy to model forest canopy height from Quickbird imagery and GEOBIA. *Remote Sensing of Environment*, 115, 1532−1542.

Chen, G., & Hay, G.J. (2011b). A support vector regression approach to estimate forest biophysical parameters at the object level using airborne lidar transects and quick bird data. *Photogrammetric Engineering and Remote Sensing*, 77, 733−741.

Chen, G., Hay, G.J., Castilla, G., St-Onge, B., & Powers, R. (2011). A multiscale geographic object-based image analysis (GEOBIA) to estimate lidar-measured forest canopy height using Quickbird imagery. *International Journal of Geographic Information Science*, 25, 877−893.

Chen, G., Hay, G.J., & St-Onge, B. (2012). A GEOBIA framework to estimate forest parameters from lidar transects, Quickbird imagery and machine learning: A case study in Quebec, Canada. *International Journal of Applied Earth Observations and Geoinformation*, 15, 28−37.

Chen, J. M., Li, X., Nilson, T., & Strahler, A. (2000). Recent advances in geometrical optical modelling and its applications. *Remote Sensing Reviews*, 18, 227−262.

Chopping, M. J., Su, L., Laliberte, A., Rango, A., Peters, D.P.C., & Kollikkathara, N. (2006). Mapping shrub abundance in desert grasslands using geometric-optical modeling and multi-angle remote sensing with CHRIS/Proba. *Remote Sensing of Environment*, 104, 62−73.

Coleman, T.F., & Li, Y.Y. (1996). An interior trust region approach for nonlinear minimization subject to bounds. *Siam Journal on Optimization*, 6, 418−445.

Coleman, T.F., Liu, J.G., & Yuan, W. (2002). A new trust-region algorithm for equality constrained optimization. *Computational Optimization and Applications*, 21, 177−199.

Environment Canada (2000). *Ecological Assessment of the Boreal Shield Ecozone*. Available online (accessed February 24, 2012): http://www.ec.gc.ca/Publications/1F4C0C47-4E18-4988-8514-A842EED6F774%5CEcologicalAssessmentOfTheBorealShieldEcozone.pdf.

Falkowski, M. J., Wulder, M. A., White, J. C., & Gillis, M. D. (2009). Supporting large-area, sample-based forest inventories with very high spatial resolution satellite imagery. *Progress in Physical Geography*, 33, 403–423.

Flanagan, L.B., & Johnsen, K.H. (1995). Genetic variation in carbon isotope discrimination and its relationship to growth under field conditions in full-sib families of *Piceamariana*. *Canadian Journal of Forest Research*, 25, 39−47.

Forzieri, G., Guarnieri, L., Vivoni, E. R., Castelli, F., & Preti, F. (2009). Multiple attribute decision making for individual tree detection using high-resolution laser scanning. *Forest Ecology and Management*, 258, 2501–2510.

Franke, J., Roberts, D.A., Halligan, K., & Menz, G. (2009). Hierarchical Multiple Endmember Spectral Mixture Analysis (MESMA) of hyperspectral imagery for urban environments. *Remote Sensing of Environment*, 113, 1712−1723.

Franklin, J., & Turner, D.L. (1992). The application of a geometric optical canopy reflectance model to semiarid shrub vegetation. *IEEE Transactions on Geoscience and Remote Sensing*, 30, 293–301.

Gemmell, F. (2000). Testing the Utility of Multi-angle Spectral Data for Reducing the Effects of Background Spectral Variations in Forest Reflectance Model Inversion. *Remote Sensing of Environment*, 72, 46−63.

Gillis, M.D., Omule, A.Y., & Brierley, T. (2005). Monitoring Canada's forests: the National Forest Inventory. *The Forestry Chronicle*, 81, 214−221.

Gougeon, F. A., & Leckie D. G. (2006). The individual tree crown approach applied to IKONOS images of a coniferous plantation area. *Photogrammetric Engineering and Remote Sensing*, 72, 1287–1297.

Gruninger, J., Ratkowski, A.J., & Hoke, M.L. (2004). The Sequential Maximum Angle Convex Cone (SMACC) endmember model. In *SPIE Proceeding, Algorithms for Multispectral and Hyper-spectral and Ultraspectral Imagery*, 5425-1, 122−135.

Haapanen, R., Ek, A. R., Bauer, M. E., & Finley, A. O. (2004). Delineation of forest/nonforest land use classes using nearest neighbor methods. *Remote Sensing of Environment*, 89, 265–271.

Hilker, T., Wulder, M.A., & Coops, N.C. (2008). Update of forest inventory data with lidar and high spatial resolution satellite imagery. *Canadian Journal of Remote Sensing*, 34, 5−12.

Hudak, A.T., Lefsky, M.A., Cohen, W.B., & Berterreche, M. (2002). Integration of LIDAR and Landsat ETM+ data for estimating and mapping forest canopy height. *Remote Sensing of Environment*, 82, 397−416.

Hyde, P., Dubayah, P., Walker, W., Blair, J.B., Hofton, M., & Hunsaker, C. (2006). Mapping forest structure for wildlife habitat analysis using multi-sensor (LiDAR, SAR/InSAR, ETM+, Quickbird) synergy. *Remote Sensing of Environment*, 102, 26–36.

Keshava, N., & Mustard, J.F. (2002). Spectral unmixing. *IEEE Signal Processing Magazine*, 19, 44−57.

Lefsky, M. A., Cohen, W. B., Harding, D. J., Parkers, G. G., Acker, S. A., & Gower, S. T. (2002). Lidar remote sensing of above-ground biomass in three biomes. *Global Ecology & Biogeography*, 11, 393−399.

Li, X., & Strahler, A. H. (1985). Geometric-optical modelling of a conifer forest canopy. *IEEE Transactions on Geoscience and Remote Sensing*, GE-23, 705−721.

Li, X., & Strahler, A. H. (1992). Geometrical-optical bidirectional reflectance modelling of the discrete-crown vegetation canopy: Effect of crown shape and mutual shadowing. *IEEE Transactions on Geoscience and Remote Sensing*, GE-30, 276−292.

Liang, S. (2007). Recent developments in estimating land surface biogeophysical variables from optical remote sensing. *Progress in Physical Geography*, 31, 501−516.

Lim, K., Treitz, P., Baldwin, K., Morrison, I., & Green, J. (2003). Lidar remote sensing of biophysical properties of tolerant northern hardwood forests. *Canadian Journal of Remote Sensing*, 29, 658−678.

McGaughey, R. J. (2010). *FUSION/LDV: Software for LIDAR data analysis and Visualization.* Available online (accessed February 24, 2012): http://forsys.cfr.washington.edu/fusion/fusionlatest.html.

McInerney, D. O., Suarez-Minguez, J., Valbuena, R., & Nieuwenhuis, M. (2010). Forest canopy height retrieval using LiDAR data, medium-resolution satellite imagery and kNN estimation in Aberfoyle, Scotland. *Forestry*, 83, 195–206.

Mora, B., Wulder, M. A., & White, J. C. (2010). Segment-constrained regression tree estimation of forest stand height from very high spatial resolution panchromatic imagery over a boreal environment. *Remote Sensing of Environment*, 114, 2474–2484.

Næsset, E. & Okland, T. (2002). Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve. *Remote Sensing of Environment*, 79, 105–115.

Natural Resources Canada (2012). *Carbon budget model*. Available online (accessed February 24, 2012): http://cfs.nrcan.gc.ca/pages/94.

Ni-Meister, W., Lee, S., Strahler, A.H., Woodcock, C.E., Schaaf, C., Yao, T., Ranson, K.J., Sun, G., & Blair, J.B. (2010). Assessing general relationships between aboveground biomass and vegetation structure parameters for improved carbon estimate from lidar remote sensing. *Journal of Geophysical Research*, 115, G00E11.

Pan, Y., Birdsey, R. A., Fang, J., Houghton, R., Kauppi, P. E., Kurz, W. A., Phillips, O. L., Shvidenko, A., Lewis, S. L., Canadell, J. G., Ciais, P., Jackson, R. B., Pacala, S., McGuire, A. D., Piao, S., Rautiainen, A., Sitch, S., & Hayes, D. (2011). A Large and Persistent Carbon Sink in the World's Forests. *Science*, 333, 988–993.

Peddle, D.R., Franklin, S.E., Johnson, R.L., Lavigne, M.B., & Wulder, M.A. (2003). Structural change detection in a disturbed conifer forest using a geometric optical reflectance model in multiple-forward mode. *IEEE Transactions on Geoscience and Remote Sensing*, 41, 163–166.

Peddle, D.R., Hall, F.G., & LeDrew, E.F. (1999). Spectral mixture analysis and geometric optical reflectance modeling of boreal forest biophysical structure. *Remote Sensing of Environment*, 67, 288–297.

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

Potapov, P., Yaroshenko, A., Turubanova, S., Dubinin, M., Laestadius, L., Thies, C., Aksenov, D., Egorov, A., Yesipova, Y., Glushkov, I., Karpachevskiy, M., Kostikova, A., Manisha, A., Tsybikova, E., & Zhuravleva, I. (2008). Mapping the world's intact forest landscapes by remote sensing. *Ecology and Society*, 13, 51.

Potapov, P., Turubanova, S., & Hansen, M.C. (2011). Regional-scale boreal forest cover and change mapping using Landsat data composites for European Russia. *Remote Sensing of Environment*, 115, 548–516.

Rashed, T., Weeks, J. R., Roberts, D. A., Rogan, J., & Powell, R. L. (2003). Measuring the physical composition of urban morphology using multiple endmember spectral mixturemodels. *Photogrammetric Engineering and Remote Sensing*, 69, 1011−1020.

Richards, J.A. (1999). *Remote Sensing Digital Image Analysis*, Springer-Verlag, Berlin, p. 240.

Somers, B., Asner, G.P., Tits, L., & Coppin, P. (2011). Endmember variability in Spectral Mixture Analysis: A review. *Remote Sensing of Environment*, 115, 1603–1616.

Stehman, S. V. (2009). Model-Assisted Estimation as a Unifying Framework for Estimating the Area of Land Cover and Land-Cover Change from Remote Sensing. *Remote Sensing of Environment*, 113, 2455–2462.

Stocks, B.J., Mason, J.A., Todd, J.B., Bosch, E.M., Wotton, B.M., Amiro, B.D., Flannigan, M.D., Hirsch, K.G., Logan, K.A., Martell, D.L., Skinner, W.R. (2003). Large forest fires in Canada, 1959–1997. *Journal of Geophysical Research*, 107, 8149, doi:10.1029/2001JD000484.

Stojanova, D., Panov, P., Gjorgjioski, V., Kobler, A., & Dzeroski, S. (2010). Estimating vegetation height and canopy cover from remotely sensed data with machine learning. *Ecological Informatics*, 5, 256–266.

Tompkins, S., Mustard, J. F., Pieters, C. M., & Forsyth, D. W. (1997). Optimization of endmembers for spectral mixture analysis. *Remote Sensing of Environment*, 59, 472−489.

Verstraete, M.M., Pinty, B., & Myneni, R.B. (1996). Potential and limitations of information extraction on the terrestrial biosphere from satellite remote sensing. *Remote Sensing of Environment*, 58, 201−214.

Woodcock, C., Collins, J., Gopal, A., Jakabhazy, V., Li, X., Macomber, S., Ryherd, S., Wu, Y., Harward, V., Levitan, J., & Warbington, R. (1994). Mapping forest vegetation using Landsat TM imagery and a canopy reflectance model. *Remote Sensing of Environment*, 50, 240−254.

Woodcock, C., Collins, J., Jakabhazy, V., Li, X., Macomber, S., & Wu, Y. (1997). Inversion of the Li-Strahler canopy reflectance model for mapping forest structure. *IEEE Transactions on Geoscience and Remote Sensing*, 35, 405−414.

Wulder, M.A., & Seemann, D. (2003). Forest inventory height update through the integration of LIDAR data with segmented Landsat imagery. *Canadian Journal of Remote Sensing*, 29, 536–543.

Wulder, M. A., Campbell, C., White, J. C., Flannigan, M., & Campbell, I. D. (2007). National Circumstances in the International Circumboreal Community. *The Forestry Chronicle*, 83, 539–556.

Wulder, M. A., Bater C. W., Coops, N. C., Hilker, T., & White, J. C. (2008a). The role of LiDAR in sustainable forest management. *The Forestry Chronicle*, 84, 807–826.

Wulder, M.A., White, J.C., Cranny, M., Hall, R.J., Luther, J.E., Beaudoin, A., Goodenough, D.G., & Dechka, J.A. (2008b). Monitoring Canada's forests. Part 1 : Completion of the EOSD land cover project. *Canadian Journal of Remote Sensing*, 34, 549–562.

Wulder, M.A., White, J.C., Masek, J.G., Dwyer, J., & Roy, D.P. (2011). Continuity of Landsat observations: Short term considerations. *Remote Sensing of Environment*, 115, 747–751.

Wulder, M.A., White, J.C., Nelson, R.F., Næsset, E., Ørka, H.O., Coops, N.C., Hilker, T., Bater, C.W., & Gobakken, T. (2012). Lidar sampling for large-area forest characterization: A review. *Remote Sensing of Environment*. DOI: http://dx.doi.org/10.1016/j.rse.2012.02.001.

Zeng, Y., Schaepman, M.E., Wu, B., Clevers, J., & Bregt, A. (2008). Scaling-based forest structural change detection using an inverted geometric-optical model in the Three Gorges region of China. *Remote Sensing of Environment*, 112, 4261–4271.

Zhao, K., Popescu, S.C., Meng, X., Pang, Y., & Agca, M. (2011). Characterizing forest canopy structure with lidar composite metrics and machine learning. *Remote Sensing of Environment*, 115, 1978–1996.

Tables:

Table 1

Summary statistics for estimates of mean, dominant, and Lorey's height in 81,045 forested lidar plots.

Tree height type	Minimum (m)	Maximum (m)	Mean (m)	Median (m)	Standard deviation (m)
Mean height	5.3	18.5	10.2	10.0	2.1
Dominant height	3.4	27.2	12.5	12.2	4.1
Lorey's height	3.6	23.6	10.3	10.0	3.1

Figures:

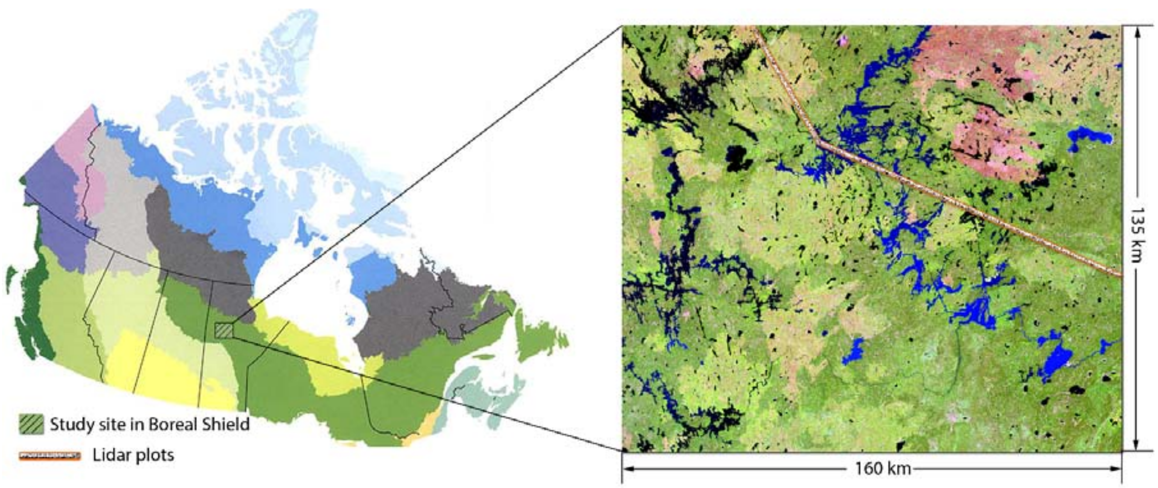

Figure 1. Study area (left) located in western Manitoba, a part of the Boreal Shield ecozone of Canada. The Landsat image (right) is from a color composite using shortwave-infrared, near-infrared, and red bands, and is partially covered by lidar plots (0.27% coverage).

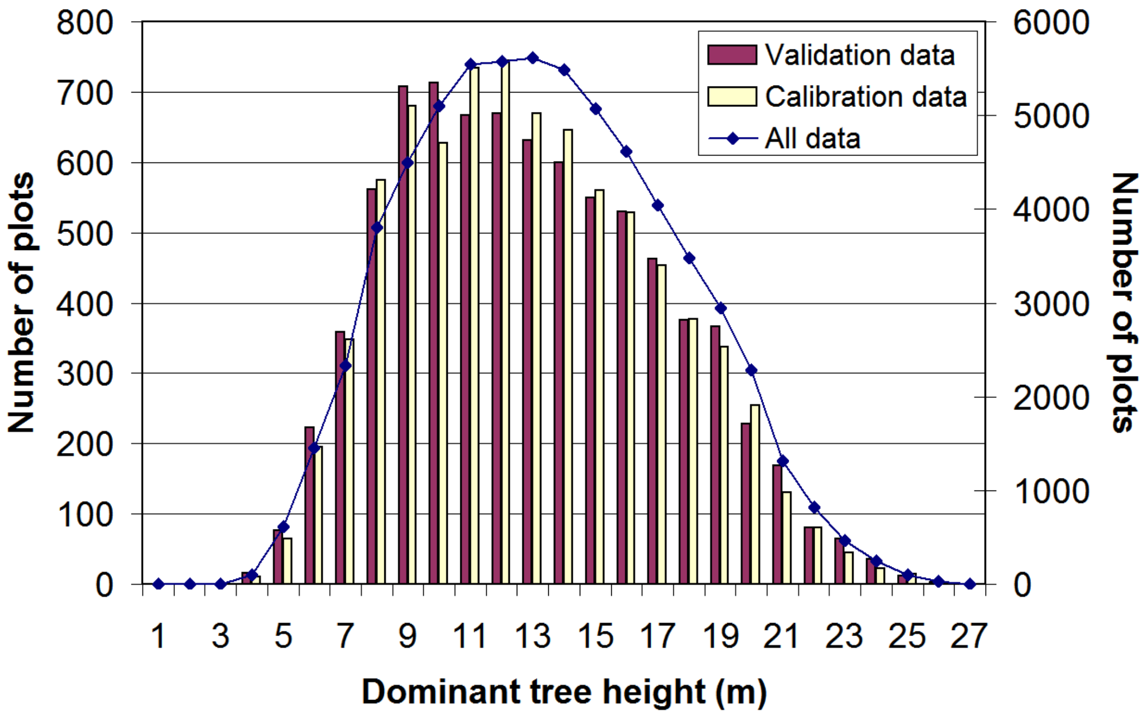

Figure 2. Comparison of three dominant tree height histograms derived from all lidar data, and randomly selected calibration and validation data.

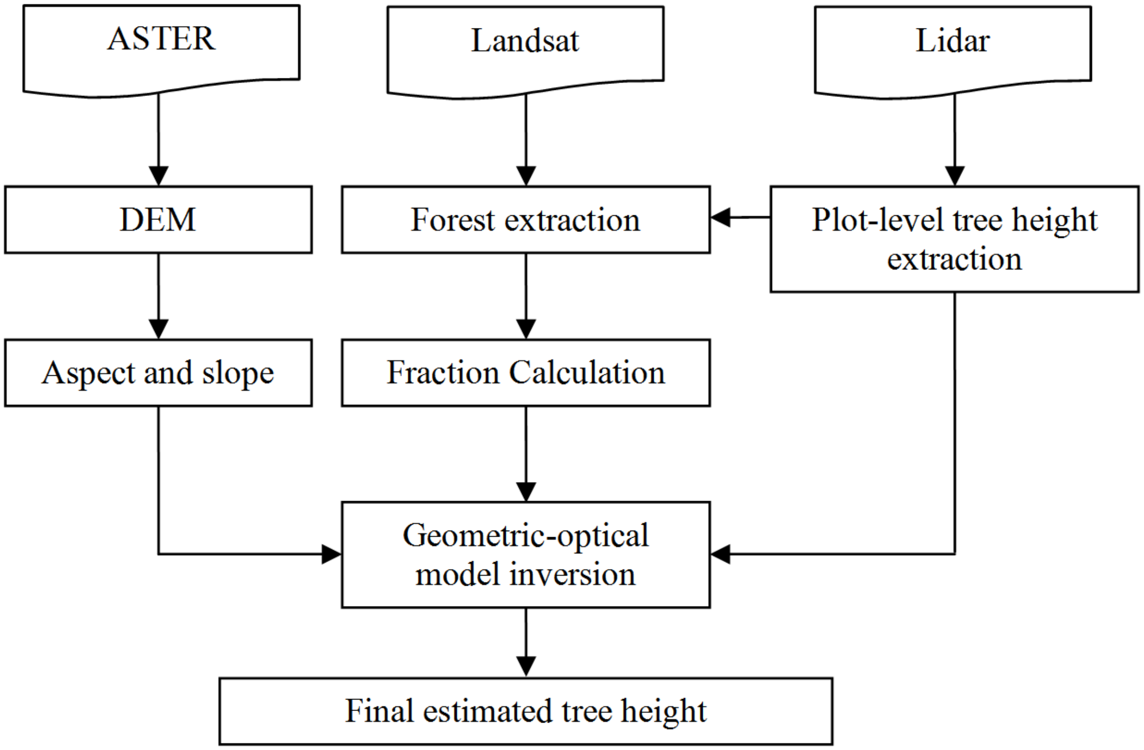

Figure 3. A flowchart of the overall approach for estimating tree height using remote sensing data and a geometric-optical model.

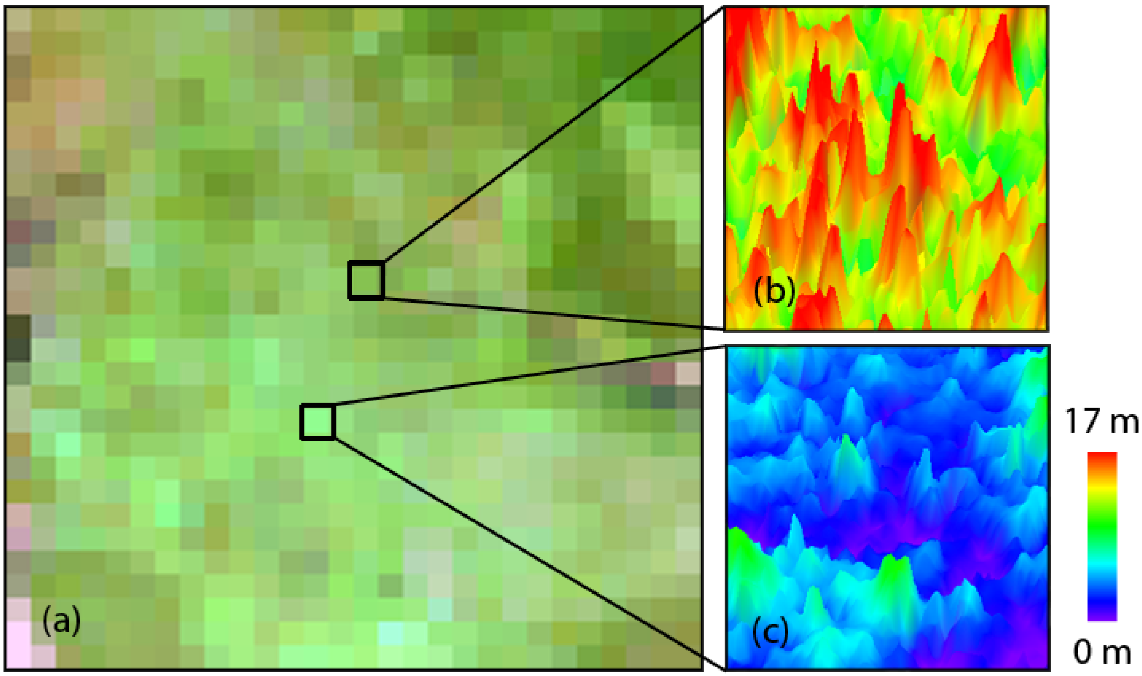

Figure 4. An example of using lidar CHM (canopy height model) in 3D view to distinguish (a) two Landsat pixels representing regions dominated by (b) trees and (c) low vegetation.

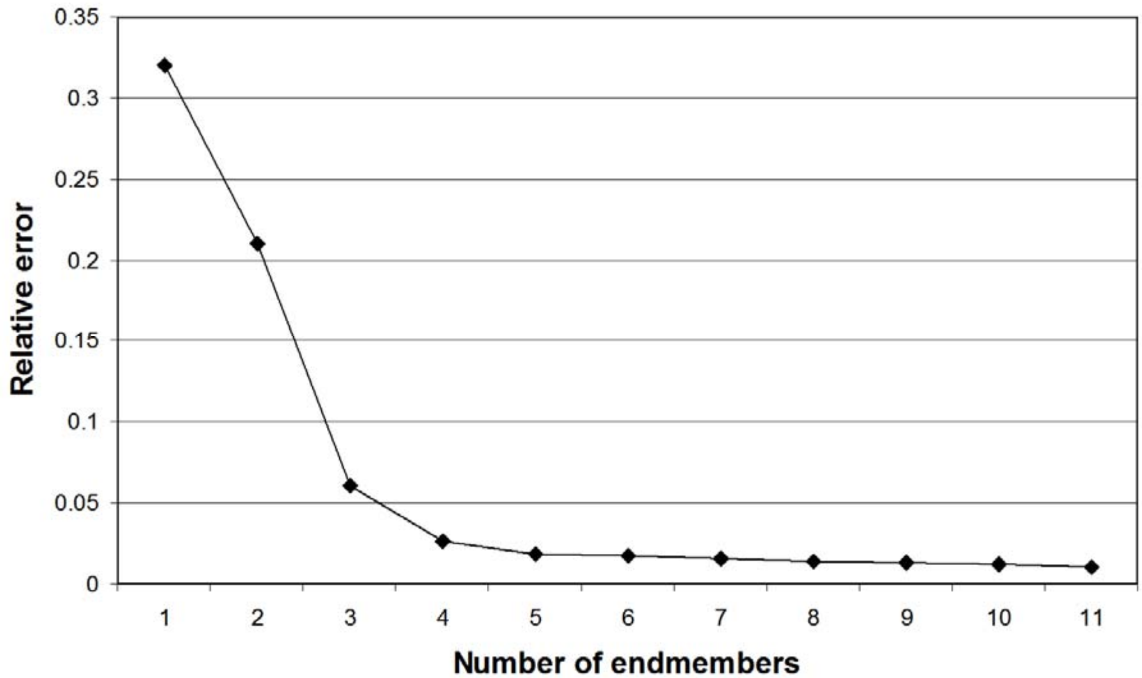

Figure 5. SMACC model relative errors calculated using different numbers of endmembers.

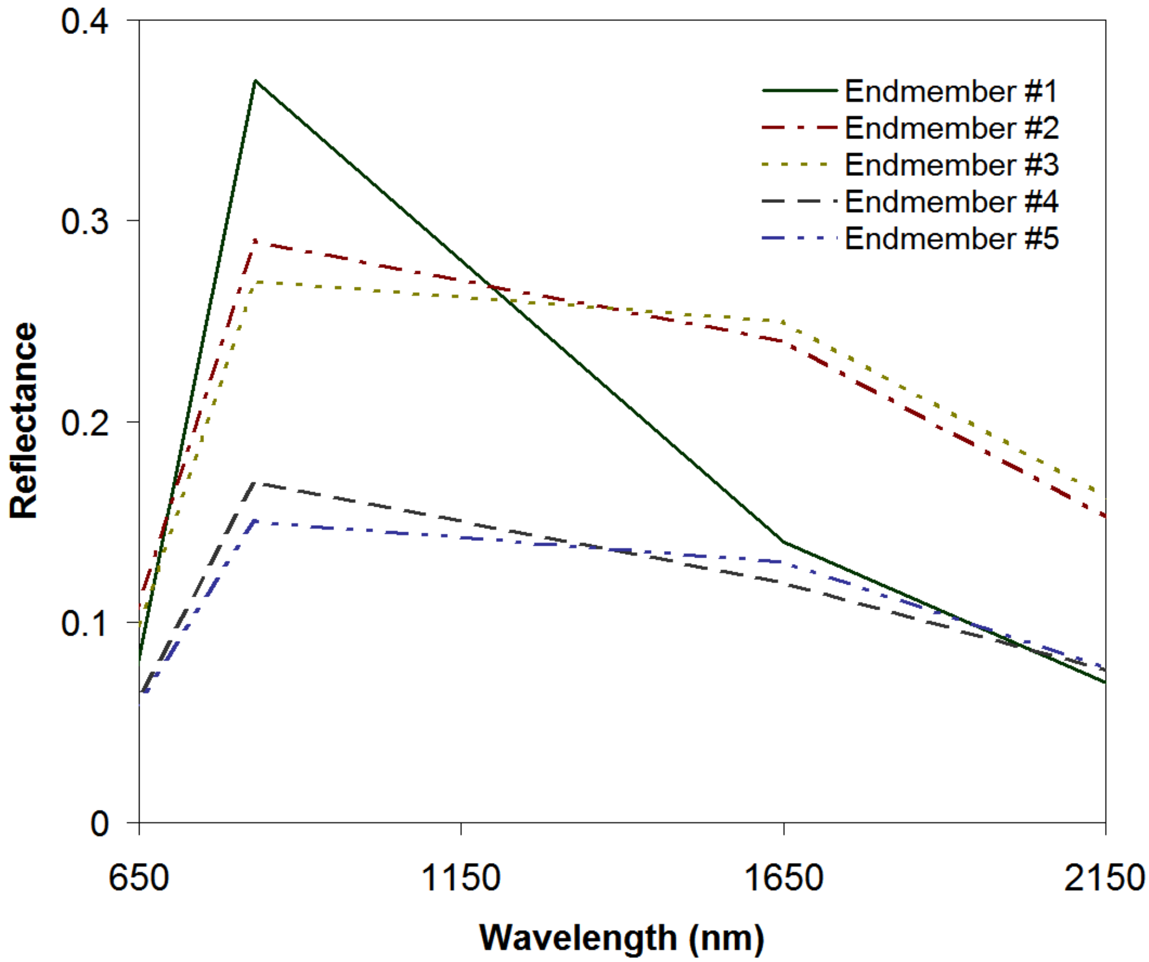

Figure 6. Endmember spectra derived from Landsat shortwave-infrared, near-infrared and red bands. Endmember #1 belongs to the sunlit crown, endmembers #2 and #3 belong to the sunlit background, and endmembers #4 and #5 belong to the shade fraction.

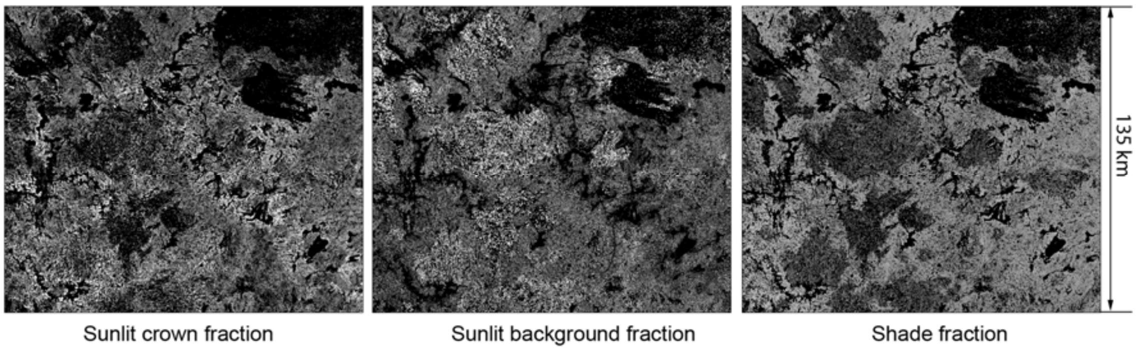

Figure 7. Fraction images of the three forest components of sunlit crown, sunlit background and shade, where grey tones represent values from low (black) to high (bright).

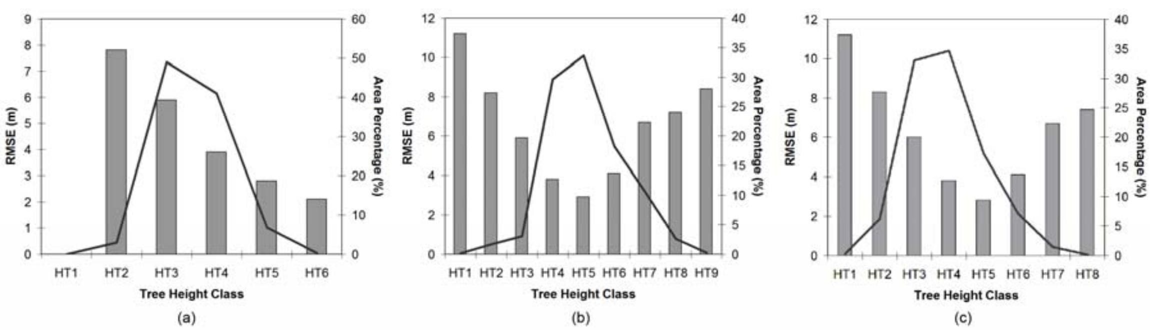

Figure 8. Estimation errors (grey bars) and area percentages (lines) of different tree height classes for (a) mean height, (b) dominant height, and (c) Lorey's height. The height classes (with 3 m interval) are: HT1 (2-5 m), HT2 (5-8 m), HT3 (8-11 m), HT4 (11-14 m), HT5 (14-17 m), HT6 (17-20 m), HT7 (20-23 m), HT8 (23-27 m), and HT9 (27-30 m).

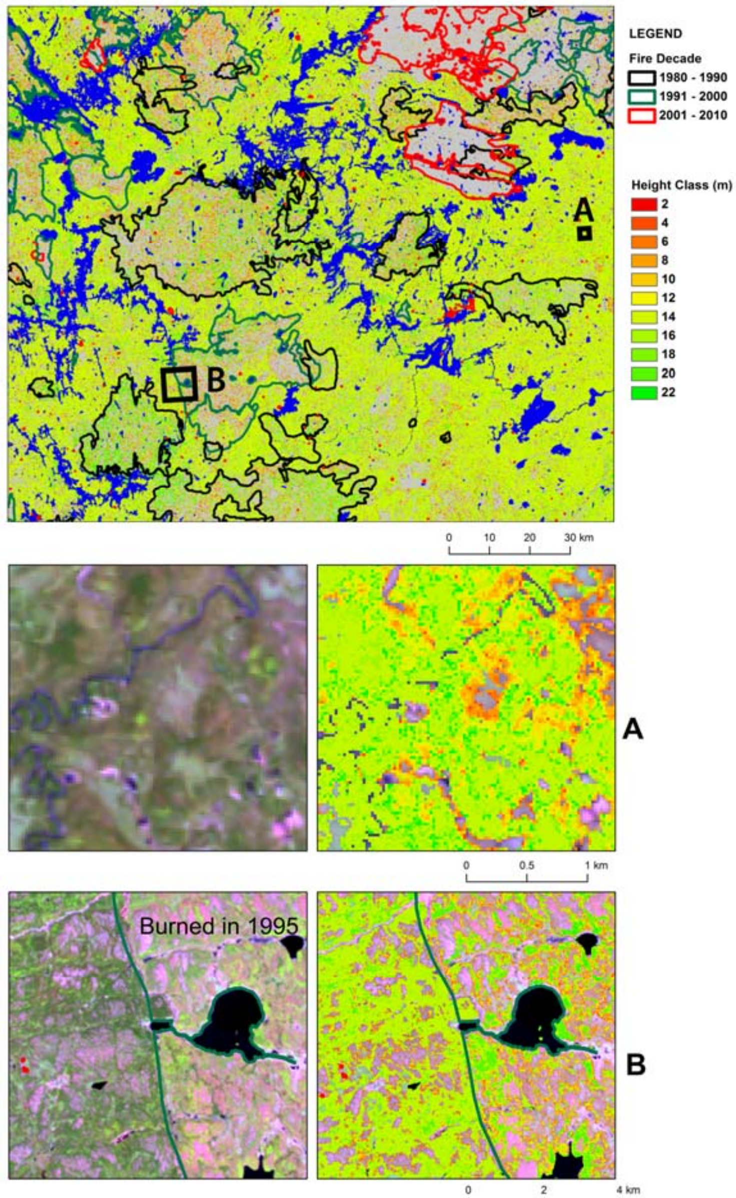

Figure 9. Wall-to-wall estimates of tree height generated from inversion of the Li-Strahler geometric-optical model, overlaid by wildfire boundaries. Insets A and B illustrate the level of detail afforded by the fine-resolution (25 m) forest height estimates. Inset A is a forest and wetland area where variability in height is a function of distance to wetland features. Inset B illustrates variability inside and outside a > 15-year old fire perimeter.