Segment-constrained regression tree estimation of forest stand height from very high spatial resolution panchromatic imagery over a boreal environment

Brice Mora, Michael A. Wulder*, Joanne C. White

Canadian Forest Service, Pacific Forestry Centre, Natural Resources Canada, Victoria, BC, Canada V8V 1M5

1. Introduction

Tree height is a fundamental attribute for describing forest stands, as well as a critical parameter for indicating site quality (Véga & St-Onge, 2009; Wulder et al., 2009) and for estimating stand-level volume and biomass (Boudewyn et al., 2007; Falkowski et al., 2009). Forest inventory stand heights are typically interpreted from aerial photography (Avery & Burkart, 2002; Hall, 2003), supplemented with field calibration data. Although the accuracy requirements for stand height estimates vary from one inventory to another, allowable error rates typically range between 10 and 15% (Rhody, 1965; Kayitakire et al., 2006). The accurate estimation of stand height from air photos is difficult in areas of dense forest where the ground is not visible (St-Onge et al., 2008). Alternative approaches for estimating stand height may be required in remote areas where forest management practices are not spatially exhaustive or where it is logistically difficult or not common practice to acquire aerial photography.

Canada’s forests comprise 10% of global forest cover and occupy approximately 40% of Canada’s landmass (Wulder et al., 2008a). Canada implements a multiphase, plot-based National Forest Inventory (NFI) for assessment and monitoring of forests. Approximately 1% of Canada’s landmass is sampled in the NFI’s first phase using a systematic network of 19,000 photo plots located on a 20 by 20 km sampling intensity grid, with each plot being 2 by 2 km in size. Within the photo plot, photo interpreters manually delineate forest stand boundaries from 1:20,000 scale aerial photography and then, using a combination of manual air photo interpretation and allometric models, determine a suite of required forest inventory attributes. In the second phase of the NFI,

* Corresponding author. Tel.: +1 250 363 6090; fax: +1 250 363 0775.
E-mail address: mike.wulder@nrcan.gc.ca (M.A. Wulder).

Crown Copyright © 2010 Published by Elsevier Inc. All rights reserved.

doi:10.1016/j.rse.2010.05.022
Several different remotely sensed data sources have been explored for forest height estimation. For example, C-band data from the Shuttle Radar Topography Mission (SRTM) and ancillary data have been used to estimate stand height in uniform stands of red and Austrian pine (Brown et al., 2010). However, radar data have limited capacity to facilitate stand delineation and estimation of other forest inventory attributes of interest (Kenyi et al., 2010), restricting the data's utility in a forest inventory context. Furthermore, tree density, tree structure, and ground slope can influence forest parameter estimation from radar data (Garestier et al., 2009). The estimation of stand height with airborne lidar data has been the subject of extensive research (Lim et al., 2003) and lidar is now used in some operational forestry contexts. To date, the only spaceborne lidar system is the Geoscience Laser Altimeter System (GLAS). Some studies have investigated the use of GLAS for estimating canopy height, with varying levels of success (Lefsky et al., 2007; Duncanson et al., 2010). While airborne lidar data may represent the state-of-the-art for estimation of forest stand height (e.g., Nässet, 1997; Nässet & Økland, 2002), lidar can be an expensive monitoring option for extensive forest areas unless a sampling approach is adopted (Wulder & Seemann, 2003). Furthermore, airborne lidar faces certain acquisition constraints (Wulder et al., 2008d), and lidar estimates of stand height may be impacted by complex terrain, steep slopes, and high canopy cover (Gatzios et al., 2010). Andersen et al. (2006) also comment and remind that ground-based measures have error typically with a range of 1 to 10%, illuminating the difficulty in obtaining definitive accuracy for lidar—or in our case other measures—of tree height.

In Canada's northern region, financial and logistical constraints often preclude the acquisition of aerial photography. In response to this information gap, the NFI has used a Landsat-based land cover product, Earth Observation for Sustainable Development of Forests or EODS (Wulder et al., 2008a), to provide a limited number of the required forest inventory attributes in areas where aerial photography has not been acquired (i.e., cover type, density, volume, and biomass) (Gillis et al., 2005). In addition, a framework has been developed that employs Very High Spatial Resolution (VHSR; < 1 m) remotely sensed imagery to support Canada's NFI programme, particularly in the north. It is envisioned that the use of VHSR will help to fill the data gap in the north, while at the same time, improving the consistency of attribute estimation between northern and southern photo plots. VHSR optical images provide spatial and spectral information that is similar to aerial photography and may be used for manual or semi-automated interpretation of forest inventory attributes (Wulder et al., 2008b). VHSR images are being acquired by an increasing number of commercial sensors and have a high geometric fidelity. As identified in Falkowski et al. (2009), “incorporating VHSR satellite imagery into existing large-area, sample-based forest inventory frameworks may provide a means to increase overall inventory efficiency and precision.”

1.1. Background

Several studies have demonstrated data and methods that have potential utility for estimating stand height using VHSR imagery. Hirata (2008) estimated stand density and stand volume in Japanese cedar (Cryptomeria japonica) and Japanese cypress (Chamaecyparis obtusa) stands using segmented QuickBird panchromatic imagery. Crown areas were calculated from the QuickBird imagery and these crown areas were used in an allometric relationship to estimate diameter-at-breast height (DBH). These DBH estimates were used in conjunction with height-diameter curves (determined by site quality) to estimate individual tree heights. From this information, volumes were calculated and compared to field-based estimates. Although the relationship between the derived and field-measured DBH values is reported (R = 0.78, p < 0.001), the reliability of the height estimates (e.g., RMSE) were not reported in this study.

Texture measures generated from 1-m IKONOS image grey levels were used to estimate stand top heights in even-aged Norway spruce (Picea abies (L.) Karst.) stands (Kayitakire et al., 2006). The authors found a strong relationship between estimated heights and field-measured heights (R² = 0.76; RMSE = 2.060; p < 0.001). In a similar study, Chubey et al. (2006) used regression trees and a suite of 87 segment-level metrics from IKONOS-2 multispectral imagery to estimate stand height (among other variables) into one of four broad height classes. The accuracy of the height estimation, assessed using independent validation data, was found to be 49%.

1.2. Objectives

As indicated by these aforementioned studies, VHSR imagery can be used as a data source for forest inventory and assessment. The overall goal of this communication is to present a method for the automated estimation of stand height from VHSR QuickBird panchromatic imagery. It is envisioned that such a process would enable the consistent estimation of an important forest inventory attribute that is a critical input for a number of other modeled inventory attributes, such as volume and biomass. As such, the methods presented are intended to augment Canada's NFI VHSR framework (Falkowski et al., 2009). Using a regression tree approach, stand heights were estimated over four sample locations in the Yukon Territory, Canada. The accuracy of the estimated heights was assessed by a comparison to heights that were manually interpreted from the QuickBird imagery. The effects of varying calibration data set sizes and input parameters were also assessed. Although the process described is in support of Canada's NFI VHSR framework, the challenges identified are informative for stand-based forest inventories in general.

2. Methods

2.1. Study area

Four study sites located in the southern Yukon Territory were selected for possibility of access and related availability of QuickBird imagery (Fig. 1). The size of the study sites ranged from 625 to 2400 ha (Table 1). All of the study sites were located in the Boreal Cordillera Ecozone (Ecological Stratification Working Group, 1995), which is characterized by a climate ranging from cold and sub-humid to semi-arid. Mean annual temperatures range between 1 °C and 5.5 °C and mean annual precipitation ranges from less than 300 mm in valleys shadowed by coastal mountain ranges, to more than 1500 mm at higher elevations. The topography of the Boreal Cordillera Ecozone includes mountains and extensive plateaus, separated by wide valleys and lowlands. Glaciation, erosion, solifluction, and eolian and volcanic ash deposition have altered the original topography. Glacial drift, colluvium, and outcrops are the most common surface materials. Permafrost is widespread in the more northern areas of the ecozone. Depending on local conditions, tree species include white spruce (Picea glauca), black spruce (Picea mariana), alpine fir (Abies lasiocarpa), lodgepole pine (Pinus contorta), trembling aspen (Populus tremuloides), balsam poplar (Populus balsamifera), and white birch (Betula papyrifera). Forest disturbances in the Yukon Territory are primarily the result of wildfire, insects, and, to a lesser extent, forest harvesting.

2.2. Data

2.2.1. QuickBird imagery

An 8 km by 8 km panchromatic (0.45–0.90 μm) QuickBird-2 image with a 0.61 m spatial resolution was acquired for each study site (see
acquisition parameters listed in Table 1). As a result of the QuickBird sensor’s variable cross-track and in-track viewing capability, the sensor has a temporal resolution that will vary according to the latitude of the end-user’s area of interest and their tolerance for an off-nadir viewing angle. For instance, at 50° N latitude a revisit of 4 days (with up to 25° off-nadir) to 7 days (with up to 15° off-nadir), may be expected (DigitalGlobe, 2005). VHSR imagery such as QuickBird enables the detection of individual tree characteristics, which in turn can provide improved estimates of many forest inventory attributes (Wulder, 1998). Use of this data is not without challenges, owing to a lack of established methods for processing (Falkowski et al., 2009) and the complex interactions between sun-sensor-surface geometry.

![Study area located in the Yukon Territory, Canada. The QuickBird image locations are also noted.](image-url)
and forest structural characteristics, particularly at more northerly latitudes (Wulder et al., 2008c).

2.3. Image pre-processing

The QuickBird images were delivered as 11-bit data, but were converted to 16-bit unsigned, resulting in a theoretical range of grey level values from 0 to 65,536. The images were converted to top-of-atmosphere radiance as per Krause (2003) and were then ortho-rectified using a 15 m panchromatic Landsat-7 ETM+ orthoimage (Wulder et al., 2002a). All four sites were either on flat or gentle slopes (i.e. less than 5%). The average RMSE for the orthorectification process was 5 m.

2.4. Image segmentation

Segmentation was used to delineate units with homogeneous forest conditions, analogous to forest stands delineated by manual interpretation (Wulder et al., 2008b). To avoid over-segmentation as a result of the high spatial resolution of the QuickBird image (Wang et al., 2004), the segmentation process was applied to a median filtered version of the original orthorectified QuickBird panchromatic band. The median filter applied to the images had a window size that was either 7-by-7 or 15-by-15 pixels. A median filter was chosen as it will likely produce more homogeneous image segments and may reduce the amount of convolution in the final segmented stand boundaries (Falkowski et al., 2009). The segmentation was performed using Definiens Cognition Network Technology® (Baatz & Schäpe, 2000; Definiens Imaging, 2004). Through experimentation and development of an NFI protocol for segmentation (Henley et al., 2009; Falkowski et al., 2009), a set of initial segmentation parameters were determined: scale = 1200, color = 0.3, and compactness = 0.9 (Henley et al. 2009). The size of the filter window and the parameters were adjusted as required for each image in order to account for differences in land cover composition, as well as differences in vegetation structure and distribution. The final segments were reviewed manually to ensure quality (Wulder et al., 2008b).

2.5. Manual image interpretation

Photo interpreters certified by the British Columbia provincial government (Ministry of Forests and Range, 2009) manually interpreted and attributed (from the QuickBird image) each of the forested segments according to National Forest Inventory Photo Plot standards (Natural Resources Canada, 2004). Key attributes that were interpreted include species composition, age, crown closure, and height (Gillis et al., 2005). The manually interpreted heights were used for subsequent model calibration and validation.

2.6. Image classification

A land cover classification using four broad cover classes (forest, herb, exposed land, and water) was performed to determine the main land cover component within each segment, with the objective of excluding any non-forest segments from further analysis. The supervised fuzzy classifier in Definiens Cognition Network Technology® software was used to assign the cover type. This step was accomplished by manually establishing image grey level thresholds for each cover class. A subsequent iteration of the classifier was used to assign the forested class to either a coniferous or broadleaf subclass. Class definitions are those specified for the National Forest Inventory and Earth Observation for Sustainable Development (EOSD) of Forests (Wulder & Nelson, 2003).

2.7. Tree crown delineation

The spatial resolution of panchromatic QuickBird imagery (pixels sized 0.61 m) has been demonstrated to produce reliable tree crown delineation (Gougeon et al., 2003; Ozdemir, 2008), thereby enabling the inclusion of crown-based metrics into our model. The Individual Tree Crown (ITC) algorithm by Gougeon (1995) was chosen to delineate tree crowns within each of the forested segments and is available as an extension of the image processing software PCI Geomatica. The ITC method requires an upper and lower grey level threshold in order to determine whether a pixel represents a portion of a tree crown or represents the surrounding shadow or understory. As recommended in Gougeon (1995), we applied a 3-by-3-pixel averaging filter to the panchromatic image prior to using the ITC algorithm.

2.8. Calculation of stand-level metrics

Stand-level measures of a panchromatic image’s grey levels are known to be conditioned by canopy structural attributes such as crown closure, tree height, and stand type (Parker et al., 1995; Asner et al., 2003). Furthermore, local topography and image acquisition parameters also influence image grey levels (Itten et al., 1992; Leckie et al., 1992; Wulder et al., 2008c). In this study, stand-level summary measures of image grey levels were used as inputs for regression trees in the estimation of stand height. The following statistics were calculated for the image grey levels in each forested segment: majority, minority, median, mean, standard deviation, range, and the number of unique values of all pixels in the segment (variety). Segment grey-level values were also examined to remove any outliers from subsequent analysis (i.e., segments containing grey-level values that were 3 or more standard deviations from the segment mean grey-level).

From the crowns delineated with ITC, segment-level estimates of crown closure, mean crown size, and the 25th, 50th, and 75th percentiles of crown size distribution were generated. Forested segments with a crown closure less than 10% were considered non-forest (Wulder & Nelson, 2003; Boudewyn et al., 2007) and were excluded from further analysis.

2.9. Regression tree development

A regression tree approach was selected for this analysis because it is non-parametric, it can accommodate both discrete (i.e., stand type) and continuous variables, and it can account for non-linear relationships.
between model inputs and output targets. Moreover regression trees tend to be robust to errors in both the independent and dependent variables (Breiman et al., 1998). Regression trees function by recursively partitioning a dataset into increasingly homogenous subsets. Regression trees were implemented using the R software (R Development Core Team, 2005) and the package tree (Ripley, 2009).

2.9.5. Model validation
Each of the stand segments used as inputs to the regression tree were characterized by the thirteen aforementioned stand-level statistics. The data set, consisting of 189 forested segments, was randomly split into separate calibration (70% of segments) and validation (30% of segments) sets. In order to ensure a non-biased evaluation of the results, we used a Multi Response Permutation Procedure (Mielle & Berry, 2001) to evaluate the degree to which the calibration and validation data were representative of the entire data set. This non-parametric method tests the hypothesis of no difference between two or more data sets for a range of parameters (i.e., the stand-level metrics used as inputs to the regression tree).

2.9.2. Selection of input metrics
Since some of the stand-level metrics involve significant computational overhead to produce (i.e., crown-based metrics), a Hill–Smith test (Hill & Smith, 1976) was applied to the entire data set to identify those input metrics with the strongest positive or negative correlations to the manually interpreted stand heights. The objective of this test was to identify an optimal set of inputs to the regression tree, thereby reducing the number of input metrics required and resulting in a more efficient protocol for modelling stand height from the QuickBird-based metrics. The Hill–Smith test is multivariate and measures correlations between both categorical (e.g., stand type) and numerical variables (e.g., mean crown size). The regression trees were run twice: once using all of the input metrics, and once using only the optimal inputs. The resulting height estimates were assessed using the same set of validation data, and then compared.

2.9.3. Regression tree parameters
For each regression tree, a K-fold cross-validation (K = 10) was processed, followed by a tree pruning stage applied according to best practices regarding tree size (McLachlan et al., 2004). We averaged the estimated mean stand heights from each iteration to produce the final mean stand height values.

2.9.4. Optimizing the size of the calibration data set
Manual interpretation of imagery for forest inventory is time consuming, costly, and inconsistent for some attributes (Wayman et al., 2001; McRoberts et al., 2002). Consequently, as part of this study, we wanted to assess the impact of reducing the number of manually interpreted stands required for calibration of the regression tree model. We measured the impact of the calibration sample size on stand height estimation accuracy by varying the size of the calibration dataset, which ranged from 10% to 100% of the forested segments, in 10% increments. For each iteration, the segments used for calibration were selected at random and the regression tree was applied; this process was repeated 100 times for each calibration sample size.

2.9.5. Model validation
Manually interpreted heights were used for validation. The coefficient of determination ($R^2$), p-value, and root mean square error (RMSE) were calculated to characterize the quality of the model and the results. Similarly, scatterplots of modeled versus manually interpreted heights, and scatterplots of residuals were generated and examined. A Welch test (Welch, 1947) was used to compare the photo interpreted and the modeled stand heights (for all calibration data set sizes). This statistical test was selected because the distributions for mean stand heights were normal and variances for photo interpreted and modeled heights were significantly different.

Many applications that use stand height to model other attributes (e.g., timber supply and carbon budget modelling) will group this metric into 5 m height classes prior to analysis (Trofymow et al., 2008, Wulder et al., 2009). Therefore to further assess the reliability of our height estimates for this purpose, we categorized our mean stand heights for the validation sample into the NFI 5-metre height classes (using the results from the regression tree model generated from the 30% sample of calibration data). We similarly categorized the manually interpreted heights into the same classification hierarchy and compared these to our estimated height classes.

3. Results and discussion

3.1. Image segmentation and classification

The segmentation parameters we used resulted in segments that complied with the NFI’s requirement for a minimum polygon size of 2 ha for forested polygons (Natural Resources Canada, 2004) (Fig. 2a). A total of 426 segments were generated, of which 68% were forested with a mean segment size of 4.9 ha (Table 2). Approximately 76% of the forested segments were coniferous and 24% were broadleaf. Overall, the coniferous segments had a larger mean segment size at 5.14 ha, compared to 4.18 ha for the broadleaf segments.

3.2. Manual image interpretation and tree crown delineation

All of the forested segments at all four sites ($N = 291$) were manually interpreted according to NFI standards (Natural Resources Canada, 2004). Individual tree crowns in the forested segments, as

Fig. 2. a) Stand delineation (white lines) with QuickBird imagery in background, b) Individual tree crown delineation using the ITC suite, with tree crown objects indicated in black. Within each inset is a focus graphic to show the QuickBird panchromatic imagery and resultant tree crowns.
delineated by the ITC program, had a mean size of 8.2 m² (Table 3, Fig. 2b). Given that the proportion of stands dominated by coniferous species (77%) is greater than the proportion of stands dominated by broadleaf species (23%), the mean crown size for coniferous dominated stands is closer to the mean crown size for the full data set. The distributions of crown sizes for both broadleaf and coniferous stands were normal (Jarque-Bera test, with p-values of 0.36 and 0.17 for broadleaf and coniferous respectively), and a Fisher test indicated that the crown size variances for each stand type were not equal (p-value = 1.15 e-10). Consequently, we conducted a Welch test to compare the mean crown size for each stand type, and found that there was a significant difference (p-value = 2.71 e-7) between the mean crown size in broadleaf and coniferous stands.

There is a known relationship between crown diameter and tree height (Peper et al., 2001; Ayars, 2004; Morsdorf et al., 2004), and between crown diameter and tree bole diameter (at breast height (DBH)) (Zhang, 1997; Bechtold 2004, Hemery et al. 2005). Hirata (2008) used an understanding of these relationships to estimate both DBH and height from individual tree crowns delineated from QuickBird panchromatic imagery. Likewise, by including tree crown parameters as inputs to our regression tree, we sought to similarly exploit these known relationships.

3.3. Calculation of stand-level metrics

Stand-level metrics were calculated and used to screen for outliers. Table 4 summarizes the metric statistics over the entire set of delineated stands. Of the 291 forested segments identified, 37 segments were excluded from further analysis because they contained grey level outliers (i.e., grey levels that were 3 or more standard deviations from the segment mean grey level). As with any regression-based approach, the detection and removal of outliers is an important prerequisite for establishing a robust model. An additional 65 segments with less than 10% crown closure were also excluded from further analysis as these segments would be considered non-forest according to NFI specifications (Natural Resources Canada, 2004). In total, 189 forested stands remained for subsequent analysis and height estimation. The segmentation process produced forested stands with crown closure values that had a relatively limited range (i.e., between 40 and 55%), mirroring the photo interpreted crown closure conditions present over the study area.

3.4. Regression tree estimation of stand height

3.4.1. Model calibration

The 189 forested segments were partitioned into separate calibration (N = 132) and validation (N = 57) data sets and the MRPP test indicated a chance-corrected within-group agreement of -1.8 10^{-3} and a p-value of 0.65, demonstrating that there were no significant differences between the calibration and validation data sets (considering all of the stand-level metrics).

3.4.2. Selection of input metrics

The results of the Hill–Smith test indicated that the 25th, 50th, and 75th percentiles of the segment-level crown size distribution, as well as the mean segment crown size, were the variables with the strongest correlation to the manually interpreted mean stand heights. Fig. 3 illustrates the correlations between variables in the space created by the two first principal components (72% of the variance is explained in this space).

Two different scenarios were explored for the automated estimation of stand height with the regression tree. The first scenario used all thirteen stand-level metrics as inputs, while the second scenario used only those four inputs that were most strongly correlated with the manually interpreted stand heights. The independent set of validation data was used to assess the performance of the regression tree models. The RMSE and R² values (with the associated p-value) for both scenarios are summarized in Table 5. Higher R² values and lower RMSE values were obtained with the full set of input parameters, regardless of the size of the calibration data set. This result suggests that although the identified optimal subset of input metrics may be more strongly correlated with the manually interpreted stand heights, this subset is insufficient for generating robust estimates of stand height.

The main stand-level metrics used in the regression tree, by frequency of occurrence and in decreasing order were: majority, standard deviation, minority, range, and mean grey level, as well as mean crown size. These results confirm the need to use both segment-based grey-level metrics as well as crown-based information to accurately model mean stand height. Multiple selection of a given metric in the same tree occurred in 15% of cases, with all trees having between two and four metrics, regardless of the regression tree structure. Similar to the range of photo interpreted crown closure

<table>
<thead>
<tr>
<th>Stand type (coniferous or broadleaf)</th>
<th>Minority</th>
<th>Majority</th>
<th>Median</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Variety</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>152</td>
<td>210</td>
<td>239</td>
<td>245</td>
<td>69</td>
<td>480</td>
</tr>
<tr>
<td>Range</td>
<td>148</td>
<td>56</td>
<td>36</td>
<td>34</td>
<td>16</td>
<td>116</td>
</tr>
<tr>
<td>Range</td>
<td>460</td>
<td>8.2</td>
<td>3.7</td>
<td>6.2</td>
<td>10.5</td>
<td>6.6</td>
</tr>
<tr>
<td>Crown closure (%)</td>
<td>460</td>
<td>8.2</td>
<td>3.7</td>
<td>6.2</td>
<td>10.5</td>
<td>6.6</td>
</tr>
<tr>
<td>Mean crown size (m²)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25th percentile of crown size</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>distribution (m²)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50th percentile of crown size</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>distribution (m²)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75th percentile of crown size</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
values, the segmentation process produced forest stands that had a
limited range of crown closure values (i.e., between 40 and 55%); as a
result, there was little correlation between crown closure and
interpreted stand height (Fig. 3). Crown closure was not selected in
any of the regression tree models.

Stand type (coniferous versus broadleaf) was only selected in 0.5% of
regression trees, indicating that stand type does not play a signifi-
cant role in tree height estimation. We performed a Student’s t-test on the
mean residual distributions from both stand types for the 30%
calibration sample size (all 13 input metrics) to further explore the
effect of stand type on the estimation of mean stand height. The t-test
was chosen as residual distributions were found to be normal ($\chi^2$ of 0.1
and 0.3) according to the Jarque-Bera test (Jarque and Bera, 1987) and
variances were found to be equal ($p$-value of 0.70 for the Fisher test). No
significant difference between residuals was found ($p$-value of 0.15).
This result confirms that in this environment, the stand type had no
significant impact on tree height estimation. As a result, we would not
recommend that the calibration data be stratified by stand type, but
rather that the calibration data is selected to represent the distribution
of mean stand heights in the study area.

3.4.3. Optimizing the size of the calibration data set

At the outset of our study, we had set aside 132 stands for model
calibration. We then assessed the impact that the size of the
calibration data set had on the estimates of mean stand height that
were generated from the regression tree. Initially, we used 10% of the
calibration data to train the model, increasing the sample size in 10%
increments. Note that the small size of the 10% sample resulted in the
systematic generation of single nodes in the regression tree and as a
result, the $R^2$ value could not be calculated as the standard deviation
of the estimated height distribution was null. The model $R^2$ values
decreased as the size of the calibration data set increased (Table 5). In
the case where all metrics were used as input, regression trees using
20, 30, or 40% of the calibration data were significant ($p<0.01$); using
only the optimal metrics as input, all of the regression models were
significant ($p<0.01$) (Table 5). Height estimates generated from the

![Fig. 3. Correlation circle displaying the results of the Hill–Smith test applied to all the stand-level metrics (Table 2) versus the manually interpreted heights.](image)

Table 5

<table>
<thead>
<tr>
<th>Calibration sample size (%)</th>
<th>All metrics</th>
<th>Optimal metrics only</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2$</td>
<td>RMSE</td>
</tr>
<tr>
<td>10</td>
<td>NA</td>
<td>2.84</td>
</tr>
<tr>
<td>20</td>
<td>0.53**</td>
<td>2.71</td>
</tr>
<tr>
<td>30</td>
<td>0.49**</td>
<td>2.68</td>
</tr>
<tr>
<td>40</td>
<td>0.49*</td>
<td>2.62</td>
</tr>
<tr>
<td>50</td>
<td>0.46</td>
<td>2.65</td>
</tr>
<tr>
<td>60</td>
<td>0.44*</td>
<td>2.66</td>
</tr>
<tr>
<td>70</td>
<td></td>
<td>2.63</td>
</tr>
<tr>
<td>80</td>
<td></td>
<td>2.59</td>
</tr>
<tr>
<td>90</td>
<td></td>
<td>3.10</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>2.92</td>
</tr>
<tr>
<td></td>
<td>NA</td>
<td>2.96</td>
</tr>
<tr>
<td></td>
<td>0.41***</td>
<td>2.94</td>
</tr>
<tr>
<td></td>
<td>0.38**</td>
<td>2.92</td>
</tr>
<tr>
<td></td>
<td>0.34**</td>
<td>2.91</td>
</tr>
<tr>
<td></td>
<td>0.31**</td>
<td>2.95</td>
</tr>
<tr>
<td></td>
<td>0.30**</td>
<td>2.93</td>
</tr>
<tr>
<td></td>
<td>0.29**</td>
<td>2.90</td>
</tr>
<tr>
<td></td>
<td>0.31**</td>
<td>3.13</td>
</tr>
</tbody>
</table>

*p<0.05; **p<0.01.
30% sample of calibration data (all input metrics) had an $R^2$ of 0.5 and an RMSE of 2.71 m, compared to an $R^2$ of 0.26 and an RMSE of 3.10 m that resulted from using 100% of the calibration data. It is generally understood that the accuracy of decision tree models tends to increase with increasing calibration sample size until a certain threshold is reached, at which point accuracy will begin to decrease; however, this threshold is ambiguous and dependent on the particular application and the source data used (Pal & Mather, 2003). The RMSEs of our models do not follow such a clear trend; however, since the approach we present for estimating mean stand height from VHSR requires some minimal amount of calibration data, there is a trade-off to be made between the amount of calibration data acquired through manual interpretation and the robustness of the model estimates. Furthermore, it is important to note that the manually interpreted heights that were used as calibration data are not without error. Error rates for the manual interpretation of forest inventory stand heights are typically reported to be between 10 and 15% (e.g., Rhody, 1965; Kayitakire et al. 2006). Such errors in the calibration data will have an impact on model estimates.

### Table 6

<table>
<thead>
<tr>
<th>Calibration sample size (%)</th>
<th>All metrics</th>
<th>Optimal metrics only</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>NA</td>
<td>0.52</td>
</tr>
<tr>
<td>20</td>
<td>0.75</td>
<td>0.49</td>
</tr>
<tr>
<td>30</td>
<td>0.86</td>
<td>0.50</td>
</tr>
<tr>
<td>40</td>
<td>0.79</td>
<td>0.40</td>
</tr>
<tr>
<td>50</td>
<td>0.69</td>
<td>0.52</td>
</tr>
<tr>
<td>60</td>
<td>0.83</td>
<td>0.57</td>
</tr>
<tr>
<td>70</td>
<td>0.70</td>
<td>0.52</td>
</tr>
<tr>
<td>80</td>
<td>0.79</td>
<td>0.55</td>
</tr>
<tr>
<td>90</td>
<td>0.79</td>
<td>0.53</td>
</tr>
<tr>
<td>100</td>
<td>0.80</td>
<td>0.56</td>
</tr>
</tbody>
</table>

3.4.4. Model validation

The comparison of the photo interpreted and the estimated heights with the Welch test indicated that the manually interpreted and modeled height distributions were not significantly different ($p > 0.05$) (Table 6). Fig. 4 illustrates the manually interpreted mean stand heights versus the estimated heights for the 30% calibration sample size case. Figs. 5 and 6 are plots of the manually interpreted heights against the absolute residuals, and the residual values for the regression tree generated from the 30% calibration sample size, respectively. The lowest residuals were for stands with mean heights ranging from 20 to 23 m (Fig. 5). As indicated in Figs. 5 and 6, the residuals were not randomly distributed: the regression tree overestimated height for stands with a manually interpreted small mean stand height and underestimated height for stands with a manually interpreted large mean stand height. Although the largest residuals were associated with the smaller trees, this error would have less
Without bias (Eid et al., 2005). While the use of in situ measurements would certainly strengthen the methodology we have presented, a model that is reliant on ground measurements would not be practical given Canada’s vast northern forest area.

4. Conclusion

In this study, we modeled mean stand height using metrics generated from panchromatic QuickBird images as inputs to a regression tree. We compared these estimates to mean stand heights that were manually interpreted from the same QuickBird image. The study showed that low RMSEs can be obtained with smaller amounts of calibration data. This suggests that there may be efficiencies that are possible in the implementation of such an approach to augment existing NFI protocols. The results of this study also demonstrated that both stand-level metrics and information on tree crown sizes are necessary to obtain robust estimates of mean stand height. Furthermore, crown closure was not found to be informative to the regression tree models and stand type (e.g., coniferous, broadleaf) had no influence on the performance of the regression models. Statistically, the best model had an $R^2$ of 0.53 and an RMSE of 2.84 m, and 84.6% of our mean stand height estimates were within the level of acceptable error specified in the British Columbia forest inventory standards. Estimated and manually interpreted heights were reclassified into five height classes and compared; classes corresponded for 54% of stands assessed, and all stands had an estimated height class within ±1 class of their actual class. This study demonstrates the potential of VHSR panchromatic imagery for acquiring estimates of mean stand heights in remote or inaccessible forest areas. Given the difficulty in using optical remotely sensed data to relate a vertically distributed attribute—stand height—the results are positive and encouraging. Future work may focus first on using ground-based data for the calibration and validation of the method, and secondly on integrating additional data sources and/or metrics into the regression tree models to further improve the estimation of mean stand height.

Acknowledgements

This research was 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. Dr. François Gougeon, of the CFS, is thanked for access and assistance with the ITC software used in this study.
References


