Remote Sensing of Environment 112 (2008) 3876–3890

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Regional aboveground forest biomass using airborne and spaceborne LiDAR in Québec

Jonathan Boudreau [a], Ross F. Nelson [b,*], Hank A. Margolis [a], André Beaudoin [c], Luc Guindon [c], Daniel S. Kimes [b]

[a] Centre d'Étude de la Forêt, Faculté de Foresterie et de Géomatique, Université Laval, Québec, Québec, Canada G1K 7P4
[b] Biospheric Sciences Branch, Code 614.4, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA
[c] Laurentian Forestry Centre, Natural Resources Canada, Canadian Forest Service, Québec, Québec, Canada G1V 4C7

ARTICLE INFO

Article history:
Received 15 January 2008
Received in revised form 3 June 2008
Accepted 15 June 2008

Keywords:
Biomass estimation
Forest carbon stocks
Geoscience Laser Altimeter System (GLAS)
ICESat
LiDAR
Portable Airborne Laser Systems (PALS)

ABSTRACT

Aboveground dry biomass was estimated for the 1.3 M km^2 forested area south of the treeline in the eastern Canadian province of Québec by combining data from an airborne and spaceborne LiDAR, a Landsat ETM+ land cover map, a Shuttle Radar Topographic Mission (SRTM) digital elevation model, ground inventory plots, and vegetation zone maps. Plot-level biomass was calculated using allometric relationships between tree attributes and biomass. A small footprint portable laser profiler then flew over these inventory plots to develop a generic airborne LiDAR-based biomass equation (R^2=0.65, n=207). The same airborne LiDAR system flew along four portions of orbits of the ICESat Geoscience Laser Altimeter System (GLAS). A square-root transformed equation was developed to predict airborne profiling LiDAR estimates of aboveground dry biomass from GLAS waveform parameters combined with an SRTM slope index (R^2=0.59, n=1325). Using the 104,044 quality-filtered GLAS pulses obtained during autumn 2003 from 97 orbits over the study area, we then predicted aboveground dry biomass for the main vegetation areas of Québec as well as for the entire Province south of the treeline. Including cover type covariances both within and between GLAS orbits increased standard errors of the estimates by two to five times at the vegetation zone level and as much as threefold at the provincial level. Aboveground biomass for the whole study area averaged 39.0±2.2 (standard error) Mg ha^{-1} and totalled 4.9±0.3 Pg. Biomass distributions were 12.6% northern hardwoods, 12.6% northern mixedwood, 38.4% commercial boreal, 13% non-commercial boreal, 14.2% taiga, and 9.2% treed tundra. Non-commercial forests represented 36% of the estimated aboveground biomass, thus highlighting the importance of remote northern forests to C sequestration. This study has shown that space-based forest inventories of northern forests could be an efficient way of estimating the amount, distribution, and uncertainty of aboveground biomass and carbon stocks at large spatial scales.

1. Introduction

Concerns about global climate change have highlighted the importance of finding efficient ways of quantifying terrestrial carbon stocks at regional, continental, and global scales. Inventories conducted at two points in time provide a potentially straightforward means of calculating the gain or loss of terrestrial aboveground carbon for a given period (Nelson et al., 2003a). Such an approach will be valuable for monitoring changes in carbon stocks over multi-year periods due to changes in climate, natural disturbance, and/or land use activities. Such carbon inventories could also be useful for helping to evaluate claims that additional carbon has been sequestered in vegetation over large land areas as a result of mitigation efforts designed to reduce overall greenhouse gas emissions. Such inventories could also be useful in assessing the impacts of climate change, natural disturbances, deforestation, and/or forest management on the amount and distribution of regional carbon stocks and provide useful information for ecosystem management and biodiversity conservation. Furthermore, a reliable method for calculating spatially explicit changes in carbon stocks would place a powerful constraint on model–data fusion techniques that combine top–down inverse modeling with bottom–up ecosystem process modeling to optimize the calculation of the spatial distribution of carbon sources and sinks for monthly time steps at continental scales (Raupach et al., 2005).

Ground inventories link tree metrics to their biomass using allometric relationships and then extrapolate these estimates to forest stands and to entire forest ecosystems (Botkin & Simpson, 1990; Botkin et al., 1993; Fournier et al., 2003; Penner et al., 1997). However, the lack of a ground-based forest inventory in remote areas and the inconsistency of inventory methods between management units over large regions are major constraints to obtaining reliable large-scale biomass estimates using ground-based techniques. Moreover, ground inventory campaigns require a huge effort over large areas and are not

* Corresponding author.
E-mail address: Ross.F.Nelson@nasa.gov (R.F. Nelson).

0034-4257/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2008.06.003

well suited for detecting changes because a single measurement campaign can extend over several years.

A potentially interesting way of obtaining a nearly instantaneous overview of terrestrial carbon stocks is to use Earth observation satellites that acquire information over large areas at regular intervals (e.g., Castro et al., 2003). However, Hyde et al. (2006) report that passive multispectral and hyperspectral sensors are of limited use especially in dense forests since they have difficulty penetrating beyond upper canopy layers and interferometric synthetic aperture radar (InSAR) does not seem to reach the desired accuracy unless a forest is structurally homogenous and has relatively low biomass. Light Detection And Ranging (LiDAR), on the other hand, is perhaps the most promising remote sensing technology for estimating biomass since it directly measures vertical forest structure by measuring the distance between the sensor and a target surface (Lefsky et al., 2002; Means et al., 1999).

LiDAR makes use of the close link between vegetation height and other biophysical characteristics of the vegetation (Dubayah & Drake, 2000; Reutebuch et al., 2005). Even though tree diameter at breast height (DBH) is the primary predictor used to estimate aboveground biomass of ground plots, Nelson et al. (2003a, 2004) demonstrated that tree height obtained from airborne lidar can also be a good predictor of biomass for large area averages. Such small footprint airborne LiDAR systems are available on a commercial basis and are now used at the operational level in forest resource inventories (Næsset, 2007; Næsset et al., 2004; Thomas et al., 2006). For example, Næsset and Goabakken (2008) were able to explain 88% and 85% of the variability in aboveground and belowground biomass, respectively, for 1395 sample plots in the coniferous boreal zone of Norway. Airborne systems usually record only the first and last return of the emitted energy that is reflected by either the vegetation or the underlying topography and have footprints of sub-meter size. Other studies have demonstrated that waveform LiDAR systems with footprint sizes of ten to 25 m can also provide accurate data to map canopy structure and biomass (Blair et al., 1999; Means et al., 1999; Drake et al., 2002).

The application of airborne lidars to forest biomass assessment still has its challenges. Estimating the biomass of hardwoods is more difficult than it is for conifers (Popescu et al., 2003; Næsset, 2004; Lefsky et al., 1999a,b). This is probably related to growth form, since tree heights are more strongly related to the volume and biomass of trees that exhibit apical dominance. Deliquescent tree forms, e.g., hardwoods, put greater amounts of woody biomass into lateral branches, thereby making the height–volume or height–biomass relationship noisier. Clark et al. (2004) found that slope, stand density, stand structure, and difficulties with determining ground elevation complicated the relationship between lidar and ground estimated tree heights for a tropical landscape in Central America. They showed a stronger correlation of lidar metrics with average stand height than for individual trees and the correlations were particularly low for trees in irregular, old growth forests and were higher for uniform plantations. Detecting changes in biomass between two points in time is particularly challenging since the estimate can be influenced by differences in instruments, geolocation, slope, and ground detection between the two sample acquisitions.

The Geoscience Laser Altimeter System (GLAS), onboard the Ice, Cloud and land Elevation Satellite (ICESat), is a full waveform digitizing LiDAR system with a nominal footprint size of ~65 m that acquires information on topography and the vertical structure of the vegetation (Harding & Carabajal, 2005; Brenner et al., 2003). While designed primarily for mapping the polar ice sheets (Zwally et al., 2003), GLAS data have been shown to have a fairly close correlation with canopy height and aboveground biomass measured on ground plots in tropical and temperate forests (Lefsky et al., 2005; Helmer & Lefsky, 2006). As the ICESat satellite acquires a consistent dataset that samples all of the Earth's vegetated surfaces on a 91-day repeat cycle, it is possible that these data can be used to obtain both regional and global estimates of aboveground biomass. However, multiple scattering of laser energy in canopies is a complex process that is dependent on wavelength, the amount of transmission through foliage, and the three-dimensional distribution of canopy components (Brenner et al., 2003). Furthermore, GLAS was designed to sample large, non-contiguous footprints within which slope and rough topography can confound the energy return over vegetated areas, thus making it more difficult to relate the signal to the structural characteristics of the vegetation (Harding & Carabajal, 2005). Variations in laser power over time are an additional confounding factor (Harding & Carabajal, 2005). Therefore, it is important to evaluate the accuracy, precision, and sources of uncertainty involved in using GLAS for large-scale biomass estimation in different regions of the world.

This article reports on such an evaluation using the 1.27 million km^2 of northern temperate and boreal forest in the eastern Canadian province of Québec. Our basic approach was (a) to relate treed aboveground dry biomass (Mg ha^{-1}) obtained from a series of 207 ground plots located in different parts of the study area to height measurements obtained from an airborne lidar system; (b) to fly the airborne lidar along several GLAS transects to relate the airborne biomass estimates to the 1325 quality-filtered GLAS waveforms with a model that used a square-root transformation of the dependent variable; and (c) to then use the 104,044 quality-filtered GLAS pulses available from 97 orbits over Québec during autumn 2003 and a Landsat-derived cover type map to estimate the treed aboveground biomass and its associated uncertainty for each cover type in the study area. Finally, we compare our estimates with other aboveground biomass estimates available in the literature and discuss how our results can be used to develop approaches for reliably estimating biomass at biome and global scales.

2. Data

2.1. ICESat GLAS laser altimetry data

Since January 2002, the ICESat satellite has been orbiting the Earth at an altitude of 600 km with a 94° inclination. During most of its operating life, it has been programmed for a 91-day orbital repeat cycle. The GLAS sensor onboard ICESat acquires altimetry data using a 1064-nm laser operating at 40 Hz, resulting in a nominal footprint of ~65 m diameter on the Earth's surface with each pulse separated by 172 m postings (Schutz et al., 2005). The laser pulse at ground level is actually in the form of an ellipse and its size varies over time as a function of the power output from the laser (Harding & Carabajal, 2005). The waveform recording sensors obtain data on the time variations in the intensity of the energy returned from a laser pulse, thus providing information on the vertical distribution of the canopy structure (Lefsky et al., 2002). The GLA14 product (http://nsidc.org/data/icesat/data.html) extracts information from these complex waveforms recorded over vegetated land by fitting up to six Gaussian distributions to the waveform to describe different waveform peaks that correspond to different features of the vertical structure of the vegetation and the underlying topography (Harding & Carabajal, 2005). Fig. 1 shows a typical waveform return profile over a boreal forest site having modest topographic relief.

Several variables related to vegetation structure can be extracted from the GLAS waveforms. The total waveform length (wflen), corresponding to the distance between signal beginning and signal end, provides information on vegetation height although it is combined with the effect of topographic slope (Lefsky et al., 2005). The position of the peaks from the six Gaussian distributions that are fit to the waveform can help to distinguish canopy from ground topography. Harding and Carabajal (2005) proposed using the distance from the signal beginning to the last peak (h_{14}) to estimate the maximum canopy height, the last peak representing the ground

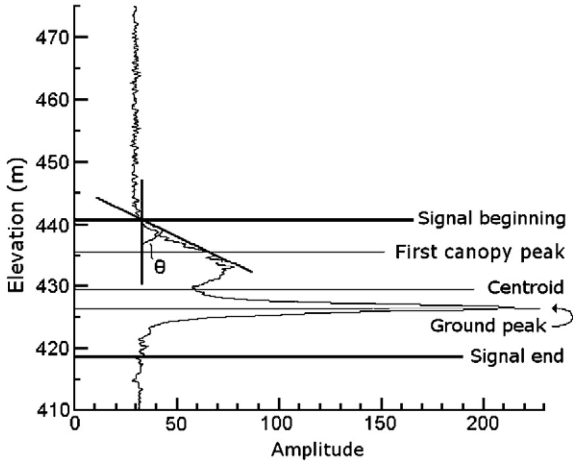

Fig. 1. A typical waveform profile for a GLAS shot over dense conifer in Québec's boreal forest showing some of the different measurements derived from the waveform. First canopy peak refers to the height of the first Gaussian peak calculated from the GLAS waveform and should correspond to the uppermost portion of the canopy. The leading edge, (θ), is defined as the slope between signal beginning and this first Gaussian canopy peak. The leading edge is used in the PALS-GLAS regression described in Section 3.2.

topography. In some cases, however, the last peak is not a good representation of the ground location, e.g., when the last peak has low amplitude and another peak with relatively high amplitude is close to it. In such cases, the higher amplitude peak should probably be used to represent ground height (alt_h_{14}). The centroid of the waveform corresponds to the height where half of the return energy is above and half below (centroid) (Ranson et al., 2004a,b). The front slope angle (θ) is defined as the angle between the vertical vector representing zero height and the vector running between signal beginning and the first peak representing the canopy (Fig. 1).

In this study, a GLAS dataset (version 26 of the Laser 2a operational period) was collected over a larger acquisition window (September 27 to November 18, 2003) within which we selected a subset (October 9 to 21, 2003) to use with the airborne data. The larger dataset contained 104,044 GLAS records from 97 orbits. The subset of data set consisted of four sections of the satellite's ground track orbit over Québec for a total of 1325 footprints across the study area from north to south.

2.2. Airborne laser profiler data

The Portable Airborne Laser System (PALS) is a profiling LiDAR designed to fly aboard a small helicopter or fixed-wing aircraft (Nelson et al., 2003b). The system digitizes the first and last return of the 905 nm laser recording at 200 Hz, for nominal footprints of 0.45 m spaced by 0.25 m postings when the sensor is flying at 200 m altitude (Nelson et al., 2003b). PALS has proven its capacity to estimate canopy heights and predict biomass in various forested cover types and densities (Nelson et al., 2003b, 2004). Nelson et al. (2003a,b) found a RMSE of 0.55 m when applying PALS to measure the height of buildings where we know precisely where the laser pulse stopped its downward trajectory. Since forests are irregular porous surfaces, the RSME estimate of building height is probably the most accurate way to estimate the precision of PALS, although this RMSE obviously also includes error in our ground measurement of building height. Fig. 2 shows the laser profile from PALS for the land area captured in the waveform shown in Fig. 1. The PALS profiles provide the data required over a given flight segment to derive the mean height for all laser shots (h_a) and for canopy shots only (h_c), the quadratic mean height of all shots (h_{qa}) and for canopy shots only (h_{qc}), the mean height of the three tallest trees (mh3), their variance, the cover density (g), and the decile heights ($h10$, $h20$,..., $h100$). PALS is also equipped with a differential Global Positioning System (dGPS) receiver and a charge-coupled device (CCD) video camera with GPS video titling. The GPS receiver was a Garmin GPS V and the dGPS corrections were provided by the Canada-wide dGPS Service (CDGPS) for the data acquired north of 50°N latitude, and by the Wide Area Augmentation System (WAAS), for differential positioning south of 50°N.

In August 2005, an aircraft equipped with the PALS system flew over portions of four GLAS orbits within the study area, acquiring over 15 million data records. In total, over 4600-km of GLAS transects were flown by PALS. PALS data were also collected over 207 ground inventory plots (see next section for plot descriptions). After completion of the flights, all the PALS data were processed, an operation that involved separating canopy hits from ground hits. This semi-automated operation was conducted using a procedure (Nelson et al., 2003a,b) developed with the Interactive Data Language (IDL) programming language. The ground hits in each PALS transect were sampled using a moving window that retains only the lowest shots and a spline was fit to this sampling of ground hits to reconstruct the ground surface. An operator then manually adjusted the spline to represent the ground as realistically as possible. The difference between the height of the upper canopy and the spline curve allows us to estimate the canopy height at any point along a transect. A threshold of 2-m was set to reject records of low height as this would likely be the result of minor variations in micro-topography (Naesset, 1997). Fig. 3a and b present an overview of the LiDAR data available in the study area and Fig. 4 illustrates how the different layers of information were overlaid.

2.3. Forest inventory plot data

A database of ground inventory plots was assembled at the Laurentian Forestry Centre (LFC) of the Canadian Forest Service (CFS). It regrouped data from two sources, temporary sample plots (TSPs)

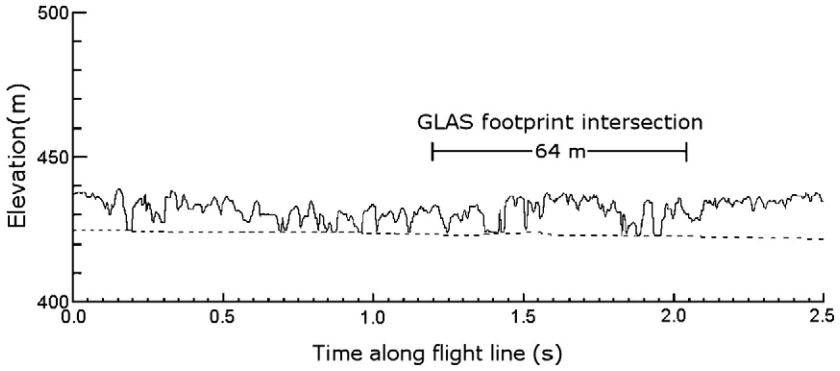

Fig. 2. A PALS profile over the GLAS footprint shown in Fig. 1. The dashed line represents the ground spline fit to the last pulse returns received by the profiling system.

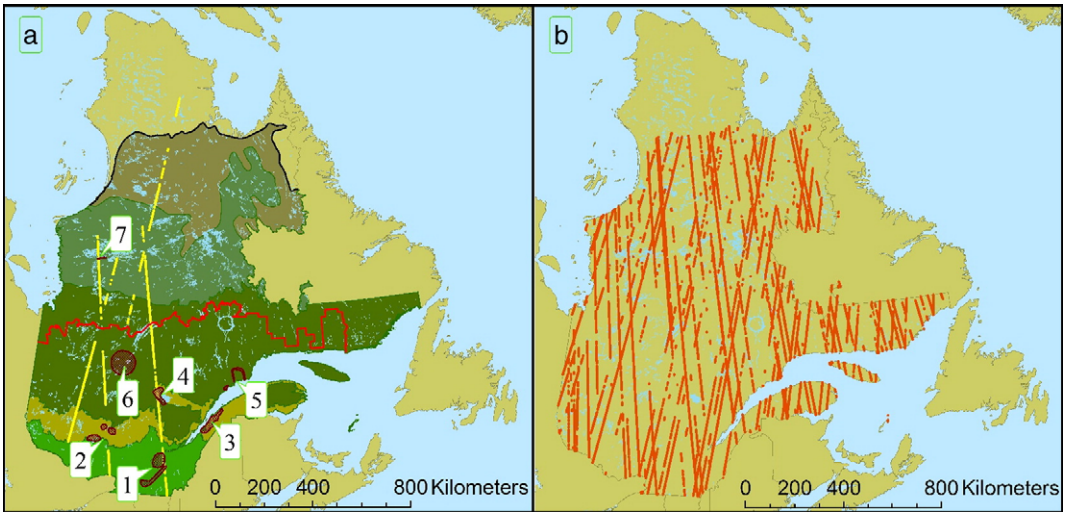

Fig. 3. Overview of the study area. (a) Regions where ground plots were sampled by PALS (brown spots), GLAS transects sampled by PALS (yellow lines), limit of the commercial forest (red line), and northern treeline (black line). The background shows, in different colors, the five vegetation zones: (from south to north) northern hardwood, northern mixedwood, boreal, taiga, and treed tundra. Ground plot regions are given a number where 1 = Trois-Rivières, 2 = Mont-Laurier, 3 = Rivière-du-Loup, 4 = Lac-Saint-Jean, 5 = Baie-Comeau, 6 = Chibougamau, 7 = Radisson. (b) All spaceborne GLAS transects from the 2a acquisition (orange lines). All transects were used in the current study.

from the Ministère des Ressources Naturelles et de la Faune et des Parcs du Québec (MRNFPQ) and TSPs collected within the Canadian Forest Service's EOSD (Earth Observation for Sustainable Development of forests) and ECOLEAP (Extended Collaboration to Link Ecophysiology and Forest Productivity) projects. Provincial inventory plots are selected according to a rigorous statistical stratification procedure for all the merchantable forests in Québec. Within this larger sample, seven areas were selected in different parts of the province for aircraft measurements of ground plots. The idea was to sample as many ecozones as feasible given several significant logistical constraints related to flight planning. This included the need to have a significant number of recently measured ground plots within a limited area that were located on fairly flat terrain (<10% slope) with no dangerous topographic features (cliffs, mountains) in the way of the aircraft. Since the PALS data are acquired at low flying altitudes (~200 m), areas with steep slopes perpendicular to the flight line had to be avoided and proximity to local airports had an influence on the final choice of plots.

Within a given sampling area, the plots were selected to obtain as wide a range of biomass as possible from the available plots. The objective for this portion of the study was to develop the ground-to-air regression equations; it was *not* to conduct a comprehensive and representative aircraft and plot sampling of Québec's forests. This latter objective was accomplished with the GLAS-PALS sampling and overall GLAS analysis described in Sections 3.2 and 3.3. Table 1 shows the number of plots flown by PALS and their characteristics for the seven different regions of the study area.

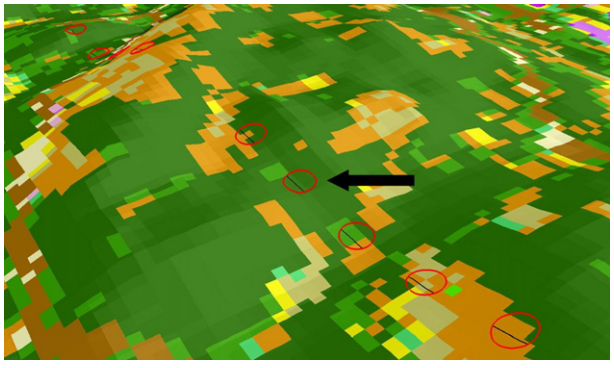

Fig. 4. Representation of the GLAS footprints on the ground (red circles) with intersection of the PALS flight lines (black lines). The DEM is a resampling of the SRTM DEM from 90-m to 30-m represented with a vertical exaggeration of 4 and a hillshade effect. The EOSD Landsat ETM+ land cover map is shown in the background. The black arrow points at the GLAS shot represented in Fig. 1.

In the areas of Québec where the forest is commercially exploited (south of the red line in Fig. 3a), the ground plot forest inventory was conducted primarily by the MRNFPQ. 17,832 temporary sample plots (TSPs) located with a GPS are available for the years 1998 to 2003. Each of these circular plots have a radius of 11.28-m, for a total area of 400-m^2. Species and diameter at breast height (DBH) of trees with a DBH larger than 9 cm were recorded. In the center of the plot, a sub-plot of radius 3.57 m (40 m^2) was established. In this sub-plot, trees with a DBH smaller than 9 cm were counted and DBH and species were recorded. TSPs are located along transect lines, each transect line usually contained five TSPs, although there could be up to ten TSPs under some circumstances. The center of the plots was located with a GPS with an error that can reach as much as 15-m. Additional information on these plots is available in MRNFPQ (2003a).

Several TSPs were jointly established by the federally sponsored EOSD and ECOLEAP projects. These plots were mostly designed to calibrate and validate methods to estimate aboveground biomass using optical data within the EOSD biomass mapping component (http://www.eosd.cfs.nrcan.gc.ca/biomass/index_e.html) and were therefore highly suitable for the purpose of the present study. The protocol used for these plots is based on the MRNFPQ ground inventory methodology except that the center sub-plot was 3.99-m radius (50 m^2) and the DBH limit for small trees was 5 cm as many plots were established in non-commercial coniferous stands with many small trees. Plot centers were located using a handheld GPS with differential CDGPS corrections, with an accuracy of 10-m. The EOSD plots selected for this study were located either (a) in the vicinity of the Fluxnet-Canada mature site (Eastern Old Black Spruce) flux tower near Chibougamau, Québec (49.7°N, 74.3°W) (Bergeron et al., 2007) and measured during 2003 and 2004 or (b) near Radisson, Québec (53.8°N, 77.6°W) and measured from 2002 to 2004.

2.4. Landsat-7 ETM+ land cover map

The EOSD project produced a land cover map of the forested lands of Canada primarily from Landsat 7 ETM+ data (Wulder et al., 2003). The final product, completed in 2006, is a 25-m resolution raster map

Table 1
Description of ground inventory plots for each ground plot study area

Region		# of plots	Average biomass (Mg ha^{-1})	Average DBH (cm)	Modal height class (m)	Dominant tree species
Name	no.					
Trois-Rivières	1	4	127.8	4.2	7–12	Sugar maple, beech, birch, aspen, basswood, ash
Mont-Laurier	2	26	63.3	4.1	7–12	Sugar maple, yellow birch, beech, eastern hemlock
Rivière-du-Loup	3	11	84.2	4.8	12–17	Balsam fir, yellow birch, white spruce, sugar maple
Lac St-Jean	4	36	61.4	5.3	4–7	Balsam fir, yellow birch, white spruce, sugar maple
Baie-Comeau	5	37	81.9	6.4	7–12	Balsam fir, white birch, white spruce, black spruce, jack pine
Chibougamau	6	38	81.2	6.1	7–12	Black spruce, balsam fir, white birch, aspen poplar
Radisson	7	26	34.5	4.5	7–12	Black spruce

available through a worldwide web data portal in 1:250,000 map sheets in GeoTiff format (http://eosd.cfs.nrcan.gc.ca). In provinces outside of Québec, the maps were produced from transformation and translations of already undergoing provincial projects. While some of the Canada's coverage was produced using K-means unsupervised classifications (Wulder et al., 2003), the 95 Landsat scenes in Québec were classified in a slightly different manner using the Enhancement-Classification Methodology (ECM), a method suitable for large scale mapping based on image enhancement, filtering and visual interpretation for the progressive class labelling of unsupervised k-means clusters down to the desired number of useful thematic classes (Beaubien et al., 1999). The images were all acquired in the peak growing season of 1998 to 2003. This required the purchase of around 35% of the 95 images for the specific use of the EOSD classification for Québec. To fill in some gaps due to clouds and get full coverage of the area, a few scenes from Landsat-5 TM were also integrated into the classifications. Out of the 22 land cover classes, there are two classes of shrubs, three classes of wetlands (including one with shrubs and one with trees), and nine classes of forest. These nine forest classes include coniferous, broadleaf, and mixedwood forests each with three density levels.

The accuracy of EOSD land cover classification in Québec is not yet well characterized. Consequently, inaccurate classifications are a potential source of error in the current study that we are not able to quantify. However, a field test in the sub-arctic area around Radisson (53.8°N, 77.6°W) showed 78% correct classification. Validation studies are on-going in the Abitibi region.

2.5. Vegetation zone map

This vegetation product was produced by the MRNFPQ and delineates the major vegetation zones and bioclimatic domains of Québec (MRNFPQ, 2003b). The delineations are determined by climatic factors as influenced by both latitude and elevation for which Québec is divided into six vegetation zones. The northern temperate forest includes two of these zones, (1) the deciduous forest in the south, characterized by sugar maple (*Acer saccharum* Marsh.), yellow birch (*Betula alleghaniensis* Britton), bitternut hickory (*Carya cordiformis* (Wangenh.) K. Koch), and basswood (*Tilia americana* L.), and (2) the northern mixedwood forest represented primarily by mixtures of yellow birch and balsam fir (*Abies balsamea* (L.) Mill.). The boreal area is represented by three vegetation zones. (1) The continuous boreal forest in the southern portion of this area consists primarily of mixed stands of white birch (*Betula papyrifera* Marsh.) and balsam fir as well as stands of either black spruce (*Picea mariana* (Mill.) B. S. P.) or jack pine (*Pinus banksiana* Lamb.) that are sometimes mixed with trembling aspen (*Populus tremuloides* Michx.). (2) The taiga, is dominated by lichen-black spruce stands and jack pine stands to a lesser extent. (3) The northern portion of this area is the forested tundra of which the northern limit is the northern treeline, denoted by the black line circumscribing the treed tundra vegetation zone in Fig. 3a (see also: http://www.mrnf.gouv.qc.ca/forets/connaissances/connaissances-inventaire-zones-carte.jsp). The sixth vegetation zone is the arctic, which was not included in this study based on prior knowledge of the absence of trees in this area.

The red line in Fig. 3 represents the northern limit of commercial operations in the boreal forest. Therefore, the continuous boreal forest was then divided into two portions, south and north of this red line. Thus, for reporting purposes, we used a total of six vegetation zones: (1) northern hardwood, (2) northern mixedwood, (3) boreal commercial, (4) boreal non-commercial, (5) taiga, and (6) treed tundra.

2.6. SRTM digital elevation model

To investigate the ability of GLAS to detect vegetation height, information on ground topography was required. The Shuttle Radar Topography Mission (SRTM) took place during 11 days in February, 2000 and acquired full coverage elevation data of the land area between 60°N and 56°S latitude (Sun et al., 2003). The resulting 90-m resolution digital elevation model (DEM) covers our entire study area. Information on slopes can be derived from this DEM using a 3×3 moving window. To determine the effect of topography on the GLAS waveforms, Lefsky et al. (2005) used a terrain index, defined as the difference between the maximum and minimum SRTM elevation value (range) in a 3×3 window centered on the pixel of interest. We followed this methodology to calculate both slope (in degrees) and the terrain index (range), in meters.

3. Methods

3.1. Developing airborne lidar biomass estimates

Biomass was calculated for each ground sample plot using a subset of the 33 species-specific equations developed by Lambert et al. (2005) for which diameter at breast height (dbh) is the independent variable. There were 38 species present in the ground plots that were selected for PALS overflights of which 22 had species-specific equations. These 22 species accounted for more than 98.5% of the biomass. For the other 16 tree species, we used substitute equations that were selected from the closest taxonomic group for which a species-specific equation was available. All of the species which required substitute equations were fairly rare and were not species that typically attain large size (e.g. *Acer pensylvanicum*).

Those sections of the PALS transects that crossed the 11.3-m radius ground inventory plots were extracted and used to calculate the LiDAR variables. From the total 413 plot crossings, a filtered dataset was constructed consisting of 207 plot crossings. The selected crossings were all within 10 m of the plot center and had homogenous land cover within 30 m of the center as determined by the video recorder mounted on board of the aircraft. Since the ground plots were inventoried between 2000 and 2002 but the PALS flight were conducted in 2005, particular attention was paid to detecting changes due to fire, logging, windthrow, insects and diseases that might have occurred between these dates. Such plots were removed based on an obvious disagreement between the ground plot data and the laser measurements and/or video records.

This ground plot filtering process was also necessary because of location errors in both the ground data and the laser data. Due to the time delay in recording the aircraft position in the laser data stream, the latitudes and longitudes recorded in the PALS data were delayed by about 0.65 s. At a nominal flight speed of 50 m/s, the cumulative delays amount to an along-track location error of ~30 to 35 m. This near-systematic error was mitigated by shifting the records along the track according to the current flight speed and orientation. The aircraft attitude represents another source of error in the PALS-ground location. PALS does not record information on the roll, pitch, and yaw of the aircraft, although the pilot was instructed to cross each plot with wings level to mitigate this unknown source of error. Finally, there is also an unknown error associated with the location of the ground plots that we believe could be as high as 15 m in some cases.

The plots were represented in a Geographic Information System (GIS) as 11.3-m radius circles centered on the plot coordinates (Fig. 4). For each plot crossed by PALS, the PALS height variables were extracted (see Data section). Vegetation zone and EOSD land cover were also extracted for each plot. This resulted in a dataset that combined ground plot biomass, PALS metrics describing forest structure, and qualitative plot information such as the land cover type and the stand density class. This dataset was used to develop two sets of parametrically-derived simple linear regressions to predict aboveground biomass using SAS Version 9.0. The first regression was a generic biomass equation developed using all 207 crossings across all vegetation zones and cover types. The second was a set of biomass equations, stratified by vegetation zone and, when possible, cover type (i.e., deciduous, mixedwood, conifer) within vegetation zone. Because of limited sample sizes, the Northern Hardwood and Mixedwood vegetation zones were combined into a joint Northern Hardwood-Mixedwood zone (NHM) for the stratified equations.

3.2. Relating PALS biomass estimates to GLAS waveform parameters

The footprints from the GLAS laser shots along the four selected orbits of ICESat were incorporated into the GIS as circles centered on the footprint center (Fig. 4). A diameter of 64-m was used since this is the nominal footprint size for ICESat acquisitions L1–L2c, L3a and L3b (Abshire et al., 2005). Each acquisition represents a period ranging from 33 to 38 days when GLAS was collecting data. The PALS transects along these orbits were overlaid on to the GLAS footprints and the various height statistics for the PALS transects within each footprint were derived. From these PALS measurements, PALS estimates of biomass were calculated for each GLAS footprint so we could develop a new regression equation between the PALS-derived biomass and various GLAS metrics (described below). Vegetation zones were also extracted for each GLAS footprint along with the EOSD Landsat ETM+ land cover and SRTM DEM information.

The first step in filtering those GLAS data that were used to develop the predictive regressions was to remove records associated with water and those records where no land cover information was available due to clouds obscuring the ETM+ observations. Second, pulse returns falling in the maximum range bin of 255 were removed. Such pulses are associated with cloud contamination. Third, a maximum vegetation height of 40 m was set for both the PALS ($\bar{h}_{qa}$) and GLAS (h_{14}) data. Although 40 m was the maximum tree height recorded in the entire TSP database for the study area, trees higher than 30 m are very rare in Québec. Fourth, only shots falling within a 100% homogenous land cover were retained. Finally, 20 pulse returns had a value of wflen greater than 50 m and these pulses were removed since they tended to greatly skew the relationships. Note that the variable, wflen, is the total waveform height including the full spread of the ground return, whereas h_{14} is the distance from signal beginning to the top of the ground peak. They probably differ most when there is an effect of local topography that does not show up in the SRTM map.

Regressions were developed between GLAS h_{14} and PALS $\bar{h}_{qa}$ (see Data section) to determine the impact of the selected filters on GLAS data quality. Using SAS, regressions were formulated to link PALS estimates of aboveground dry biomass to GLAS waveform parameters. Dry biomass refers to the weight of the plant matter exclusive of any water in the tissues. It is the standard biomass measurement used in ecology since water contents can vary dramatically over short periods of time after harvest and before weighing and it is closely related to the carbon stocks.

Lefsky et al. (2005) used a parametric multivariate regression model based on the total waveform extent (wflen, from GLAS), terrain index (range, from SRTM), and the leading edge extent where this latter variable is described as the vertical distance between signal start and the topmost peak of the waveform. In our data set, variable selection procedures identified the slope of the leading edge (fslope) as being more informative than the leading edge extent. The fslope variable (θ), defined by the slope between signal beginning and the first Gaussian canopy peak, provides information on canopy density and the vertical variability of the upper canopy (Fig. 1). Several GLAS vegetation height variables were compared (wflen, h_{14}, *alt_*h_{14} and centroid). The *range* and θ variables were included in the regression, both separately and combined. The best possible subset algorithm in SAS selected a model that used SRTM range, θ, and wflen to estimate canopy height. The other variables, such as several combinations of

Table 2
The stratified and the generic regression equations for PALS biomass based on regressions between biomass derived from ground inventory plots and PALS data acquired over these plots

		PALS metric	Slope	Intercept	R^2	RMSE Mg ha^{-1}	n	Ground Inventory Biomass		
								CV (%)	Min. Mg ha^{-1}	Max. Mg ha^{-1}
Vegetaion areas/cover type	Generic	$\bar{h}_{qa}$	10.15	−2.53	0.65	25.5	207	62.3	0.2	213.9
	Northern hardwood-Mixedwood/deciduous	$h70$	9.61	−19.91	0.73	20.6	37	79.6	0.2	137.6
	Northern hardwood-Mixedwood/mixedwood	$(\bar{h}_{qc})$ (g)	0.10	2.48	0.51	38.7	24	65.8	11.4	213.9
	Northern hardwood-mixedwood/conifer	$\bar{h}_{qc}$	10.45	−15.99	0.79	18.0	23	51.1	17.3	167.9
	Boreal/deciduous	$(\bar{h}_{a})$ (g)	0.08	25.37	0.51	26.7	10	37.6	36.3	149.8
	Boreal/mixedwood	$(\bar{h}_{qa})$ (g)	0.09	14.24	0.59	22.8	26	42.1	21.2	148.4
	Boreal/conifer	$\bar{h}_{qa}$	10.88	0.09	0.68	23.2	61	52.2	21.6	165.8
	Taiga	$\bar{h}_{qa}$	14.90	−12.54	0.72	14.2	26	76.6	6.6	106.2

$\bar{h}_{qa}$ = Quadratic mean height of all pulses, in meters.
$h70$ = 7th decile height, in meters, i.e., height below which 70% of the pulses fall.
$\bar{h}_{qc}$ = Quadratic mean height for canopy hits >3 m, in meters.
(g) = Crown closure, in percent, i.e., [(number of canopy hits >3 m)/total number of pulses]*100.
(h_a) = Mean height of all pulses, in meters.
CV = Coefficient of variation, i.e., standard deviation/mean biomass.
The PALS metrics were calculated for PALS pulses falling within the 22.6-m diameter circles representing each ground plot.

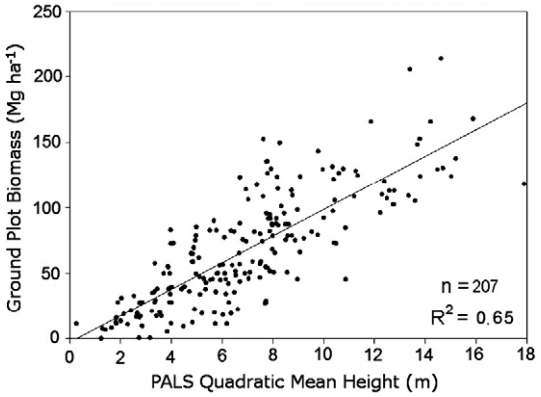

Fig. 5. Quadratic mean height (h_{qa}) derived from the airborne lidar (PALS) vs. ground measured biomass for 207 field plots located in seven different regions of Québec.

the six Gaussian peaks fitted to the waveform, were rejected because of collinearity issues with wflen.

Residual analysis of the model indicated strong heteroskedasticity. A square-root transformation of the dependent variable (PALS biomass) was employed to mitigate this issue (Gregoire et al., in press). A log transformation was also tested, but plots of residuals indicated that the square-root transformation was much better at resolving heteroskedasticity and better linearized the PALS biomass-GLAS height relationship. Attempts were made to produce stratified PALS-GLAS equations, but a paucity of data in some strata and poor fits in other strata ($R^2 \sim 0.1$) precluded the use of these stratified equations. Consequently, we decided to work with one single non-stratified PALS-GLAS equation.

This generic equation was then used to develop biomass estimates for the entire study area. A $\sqrt{\text{biomass}}$ estimate was calculated for each GLAS pulse and the following back-transformation was applied, as per Eq. 14 in Gregoire et al. (in press):

$$\hat{b}_G = \left(\sqrt{\hat{b}_G}\right)^2 + \hat{s}^2_{reg} \quad (1)$$

where $\hat{b}_G$ = GLAS pulse estimate of aboveground dry biomass, in Mg ha^{-1} and $\hat{s}_{reg}$ = root mean square error associated with the regression of the generic equation. Gregoire et al. (in press) noted that this back-transformation is conservatively biased but that in most cases, the bias is near zero. Furthermore, the results of the unbiased back-transformation that they propose differed very little from those obtained using Eq. (1), e.g., mean values differed by less than 0.1% for the data used in our study (results not shown).

We attempted to consider the impact of prediction error from the regression on biomass estimates by randomly adding noise to the dataset, thus creating new GLAS data sets that included this noise. This is described in further detail at the end of the following section.

3.3. Extrapolating GLAS biomass estimates to the study area

GLAS estimates of aboveground dry biomass were computed for all 104,044 filtered GLAS shots available over the study area from the Laser 2a operational period (e.g., <40 m height, homogenous land cover, saturation index <1, etc.). For each of the 97 GLAS orbits, biomass estimates were computed for each Landsat EOSD cover type within each vegetation zone. At this phase, the sections of the GLAS orbits within each cover type and each vegetation zone became the basic sampling unit.

The biomass estimates for each Landsat cover type within each of the six vegetation zones were obtained from a weighted sum of the GLAS flight lines falling within each cover type for a given vegetation zone. The weights were based on the number of GLAS pulses within each orbit for each cover type within each vegetation zone. The equations that we used are based on equations presented in Nelson et al. (2004), modified to reflect the fact that we are working with individual GLAS pulses within an orbit instead of lengths of profiling laser segments within flight lines. Note that for pulses within a cover type within an orbit, spatial autocorrelation is not an issue because we take the average for the orbit and this average value for the orbit is the sampling unit in subsequent analyses. Standard errors would have been much lower if we had used individual shots as the sampling unit, but then autocorrelation would have been a problem. For estimates made between cover types within an orbit, there is a positive covariance (autocorrelation), so we calculated the covariances within and among orbits as described below.

The equations used for calculating biomass estimates and their associated variances are described in Appendix A. The procedure involved calculating (a) biomass ha^{-1} for a given cover type, by orbit, within a given vegetation zone (Eq. (A-1)); (b) biomass estimates for cover types within a vegetation zone (Eq. (A-2)); (c) the variance of the cover type estimates (Eq. (A-3)); (d) biomass estimates for vegetation zones within the Province (Eq. (A-4)); (e) variances (Eq. (A-5)) and covariances (Eqs. (A-6) and (A-7)) for the vegetation zone estimates; (f) Provincial-level biomass estimates (Eq. (A-8)) and (g) Provincial-level variances calculated either without inclusion of covariances (Eq. (A-9)) or with inclusion of covariances (Eqs. (A-10) and (A-11)).

For the study area estimates, we also took advantage of the information provided by the Landsat land cover classification to identify GLAS shots that should be excluded because they are associated with a cover type that does not normally have significant biomass, e.g., ice, bare ground.

Table 3
Regressions between GLAS height and PALS height after the application of different filters

Filters	n	Slope	Intercept	R^2	RMSE	GLAS height		
						CV (%)	Min (m)	Max (m)
–	8841	0.24	1.22	0.14	6.9	78.5	0.6	81.1
Land cover class is water or no data	7352	0.24	1.71	0.12	7.5	69.9	0.6	81.1
Saturation index>1	6435	0.24	1.78	0.14	7.0	70.7	0.6	81.1
PALS $\bar{h}_a$, $\bar{h}_c$, $\bar{h}_{qa}$, $\bar{h}_{qc}$ and mh3>40 m	6406	0.21	1.92	0.28	4.0	70.8	0.6	81.1
GLAS h14>40 m	6139	0.35	0.08	0.45	3.4	58.8	0.6	40.0
Pure land cover class within footprint	1345	0.41	−0.60	0.52	3.5	58.2	0.8	39.9
GLAS wflen>50 m	1325	0.41	−0.67	0.54	3.4	58.0	0.8	39.9

$\bar{h}_a$ = PALS mean height of all pulses, in meters.
$\bar{h}_c$ = PALS mean height for canopy hits >3 m, in meters.
$\bar{h}_{qa}$ = PALS quadratic mean height of all pulses, in meters.
$\bar{h}_{qc}$ = PALS quadratic mean height for canopy hits >3 m, in meters.
mh3 = PALS mean height of the three highest hits, in meters. h14 = Distance from signal beginning to ground peak from GLAS waveform, in meters.
wflen = Distance from signal beginning to signal ending from GLAS waveform, in meters.
The GLAS height used for this regression was the distance from signal beginning to ground peak (h14). PALS height is the quadratic mean height (h_{qa}) of all pulses within a 64-m diameter circular GLAS footprint.

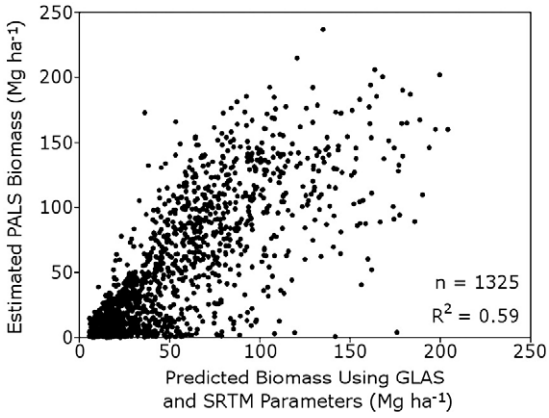

Fig. 6. Biomass estimated using PALS-ground equations for GLAS pulses sampled by PALS (Table 2) versus biomass estimates for the same pulses developed from the regression using the GLAS and SRTM parameters (Eq. (2)).

In this biomass statistical generalization process, we used a parametrically-derived, implicitly linear equation to relate PALS estimates of biomass to GLAS waveform parameters. The prediction error, or scatter about the regression line, may introduce an important source of variation compared to the standard errors of the estimates of biomass that are normally reported. This prediction error represents, for each GLAS waveform used in the scaling-up process, an uncertainty in predicted biomass. To account for the influence of this additional uncertainty, we perturbed each GLAS estimate of biomass by adding prediction error (Nelson et al., 2008). The IDL function RANDOMN was used to generate normally-distributed random numbers with a mean of zero and variance of one. This random value was multiplied by the RMSE of the predictive regression equation and by a prediction error term reported by Myers (1990, pg. 112) and utilized by Nelson et al. (2008). This random error, when added to the predicted values of biomass, resulted in a new dataset with the biomass variability at the individual shot level characterized by this prediction error. However, introducing prediction error into a square-root transformed model can result in a positive bias in the estimates (see Discussion, Section 5.3).

4. Results

4.1. Ground plots to PALS

We developed two sets of equations to link the PALS data records to the ground plot biomass (Table 2); (1) a generic equation using $\bar{h}_{qa}$ that has an R^2 of 0.65 (Fig. 5) and (2) seven equations stratified for different combinations of vegetation zones and cover types. This latter set of equations used several PALS height variables ($h70$, $\bar{h}_{qc}$, $\bar{h}_a$, and $\bar{h}_{qa}$) sometimes combined with crown closure (g) and yielded R^2 values between 0.51 for deciduous boreal stands to 0.73 for deciduous stands in the NHM vegetation zone (Table 2).

4.2. PALS to GLAS

Different filters were applied to the initial GLAS Acquisition 2a dataset along the orbits selected for aircraft sampling so as to improve data quality. These filters improved the R^2 of the regressions between GLAS h_{14} and PALS $\bar{h}_{qa}$ from as low as 0.14 up to 0.54 (Table 3). These filters reduced the number of GLAS-PALS data records available, dropping them from 8841 records to 1325.

The filtered GLAS dataset ($n=1325$) was used to build one generic equation to link PALS estimates of aboveground dry biomass measured for the GLAS pulses using a generic equation that related PALS biomass to different GLAS and SRTM variables (Fig. 6, Eq. (2)):

$$\sqrt{\hat{b}_{PG}} = 0.27\text{wflen} - 0.83\theta - 0.06\,range + 2.67,\ n = 1325,\ R^2 = 0.59 \quad (2)$$

where $\hat{b}_{PG}$ = is the dependent variable used for calculating the regression parameters, i.e., the PALS estimate of aboveground dry biomass in Mg ha^{-1} for GLAS pulses sampled by PALS and calculated using generic PALS-ground equations; wflen = total GLAS waveform extent, i.e. the distance from signal beginning to signal end, in meters; θ = GLAS waveform front rising slope angle, in radians; and range = the SRTM DEM terrain index, in meters.

4.3. Extrapolating biomass estimates to the study area

The statistical generalization process provided estimates of aboveground dry biomass by vegetation zone and for the entire Province south of the treeline. This was done via a series of preliminary calculations for each vegetation zone using a number of different assumptions (described below) for calculating means and variances and reporting only for forested Landsat-derived EOSD land cover types (Table 4). The first column in Table 4 is the result of an extrapolation based on biomass calculated directly from the filtered provincial-wide GLAS records using Eq. (2) and without accounting for prediction error or for covariance as described in Eqs. (A-6), (A-7) and (A-9). Biomass varied from 21.4 to 63.3 Mg ha^{-1} in the different vegetation zones and the standard errors ranged from 0.5 and 1.3 Mg ha^{-1} (Table 4). When these estimates for the individual vegetation zones were scaled-up to the provincial level, the mean biomass was 39.0 Mg ha^{-1} and the standard error decreased to 0.3 Mg ha^{-1}.

Table 4
Preliminary and final GLAS estimates of aboveground dry biomass and their standard errors for the major vegetation areas and for the entire Province of Québec south of the treeline

		Mean biomass (Mg ha^{-1})			Area (Mha)	Total biomass[4] (Mg × 10^6)	Percentage of the total biomass (%)
		1	2	3			
Vegetation zones	Northern hardwood	56.6±1.3	60.7±1.3	56.6±1.4	11.0	621.3±15.7	12.6
	Northern mixedwood	63.3±0.7	69.7±0.8	63.3±2.3	9.8	620.9±22.1	12.6
	Boreal (commercial)	50.6±0.6	56.1±0.6	50.6±2.3	37.5	1897.3±84.3	38.4
	Boreal (non-commercial)	36.2±0.6	41.4±0.6	36.2±0.7	17.7	640.1±13.2	13.0
	Taiga	23.8±0.5	28.2±0.5	23.8±0.7	29.5	702.0±19.8	14.2
	Treed Tundra	21.4±0.5	25.6±0.5	21.4±1.1	21.3	456.8±22.7	9.2
	Province (South of treeline)	39.0±0.3	43.9±0.3	39.0±2.2	126.8	4938.2±274.8	100.0

1. Without prediction error from the regression and without covariance.
2. With prediction error from the regression, without covariance.
3. Without prediction error from the regression and with covariance.
4. Developed from results in column 3.
These results are reported from Eq. (2) and reporting only for forested Landsat-derived land cover types. The effects of adding regression noise (prediction error) on the estimates is shown in column 2, and the effects including covariance within and between orbits is shown in column 3. The calculations were based on a PALS-GLAS regression model that applied a square-root transformation to the dependent variable as shown in Eq. (2).

Table 5
GLAS estimates of aboveground dry biomass (Mg ha^{-1}), standard errors, and the number of GLAS orbits intercepting each stratum (n), by vegetation zone, for all EOSD forested land cover classes

		Vegetation zones					
		Northern hardwood	Northern mixedwood	Boreal (commercial)	Boreal (non-commercial)	Taiga	Treed tundra
EOSD forested cover types	High shrubs (Height>2 m)	63.8±3.9 (27)	49.6±2.7 (39)	39.9±2.2 (47)	42.7±3.1 (31)	47.7±2.6 (49)	40.7±2.3 (44)
	Low shrubs (Height < 2 m)	37.5±6.2 (16)	48.7±3.6 (33)	41.1±1.8 (56)	33.0±1.3 (58)	32.8±1.7 (56)	32.7±1.4 (48)
	Wetland treed	45.5±10.2 (8)	50.4±8.6 (8)	24.2±1.2 (21)	21.1±0.6 (11)	23.1±1.5 (21)	22.3±3.5 (6)
	Wetland shrubs	32.6±4.4 (11)	34.3±4.4 (12)	23.2±1.6 (33)	21.5±1.7 (25)	23.2±1.0 (35)	41.1±7.3 (14)
	Dense conifer (density>60%)	65.3±2.5 (24)	74.7±2.0 (41)	72.6±1.4 (64)	66.2±1.8 (56)	47.4±2.3 (46)	42.2±4.6 (14)
	Open conifer (density 26–60%)	57.4±4.8 (21)	65.7±2.5 (40)	58.3±1.6 (65)	50.2±1.4 (67)	38.0±1.4 (56)	33.6±1.4 (42)
	Sparse conifer (density 10–25%)	24.8±5.9 (3)	57.3±6.5(14)	37.7±1.4 (56)	37.5±1.5 (67)	34.9±1.5 (55)	32.9±1.2 (48)
	Dense deciduous (density>60%)	95.2±4.0 (27)	88.6±2.5 (42)	64.0±3.7 (41)	72.7±3.7 (21)	59.4±9.2 (9)	48.8±26.0 (2)
	Open deciduous (density 26–60%)	78.9±4.0 (27)	72.2±3.3 (39)	55.4±3.1 (39)	56.8±9.7 (10)	30.1±7.0 (3)	–
	Sparse deciduous (density 10–25%)	–	–	–	–	–	–
	Dense mixed (density>60%)	85.5±3.3 (29)	82.4±1.5 (47)	73.3±1.7 (55)	71.6±3.2 (32)	45.6±11.1 (4)	26.3±4.9 (3)
	Open mixed (density 26–60%)	67.5±3.3 (27)	71.6±2.0 (45)	60.3±1.9 (57)	55.3±2.7 (44)	55.6±5.5 (16)	26.3±1.8 (2)
	Sparse mixed (density 10–25%)	–	–	23.2±3.5 (3)	–	27.0±2.0 (4)	–

The calculations were based on a PALS-GLAS regression model that applied a square-root transformation to the dependent variable as shown in Eq. (2).

Prediction error was then added to each GLAS pulse estimate of biomass. The second column of the Table 4 shows the effect of this noise with biomass estimates now varying from 25.6 to 69.7 Mg ha^{-1} with standard errors between 0.5 and 1.3 Mg ha^{-1}. For the entire Province south of treeline (1.3 M km^2), we now estimate 43.9 Mg ha^{-1} of aboveground dry biomass with a standard error remaining essentially stable at 0.3 Mg ha^{-1}. It is important to note that this result is anomalous. The addition of prediction error should change mean estimates very little (~0) and, in general, increase the variances of those mean estimates. Just the opposite occurs in column 2. The reason behind this result is discussed in Section 5.3.

The third column in Table 4 shows the biomass estimates when the covariance between and within the flight lines within cover types at the vegetation zone level and between the cover types within vegetation zones at the provincial level is included (Eqs. (A-6), (A-7), and (A-9)). The estimates do not include the prediction error and are the same as those in column 1 (Table 4). However, the standard errors increased, reaching values ranging from 0.7 to 2.3 Mg ha^{-1} for the vegetation zones and 2.2 Mg ha^{-1} for the Province.

When the estimates from column 3 in Table 4 are multiplied by the appropriate surface areas, we obtain estimates of the total aboveground dry biomass for each vegetation zone and for the entire Province south of the treeline, i.e., 4.9 (±0.3 SE) Pg aboveground dry biomass (Table 4).

In the extrapolation process, intermediate estimates of aboveground dry biomass were computed by vegetation zone for each land cover type (Table 5). The values presented in this table were obtained using Eq. (A-8) where a weighted average of biomass was calculated

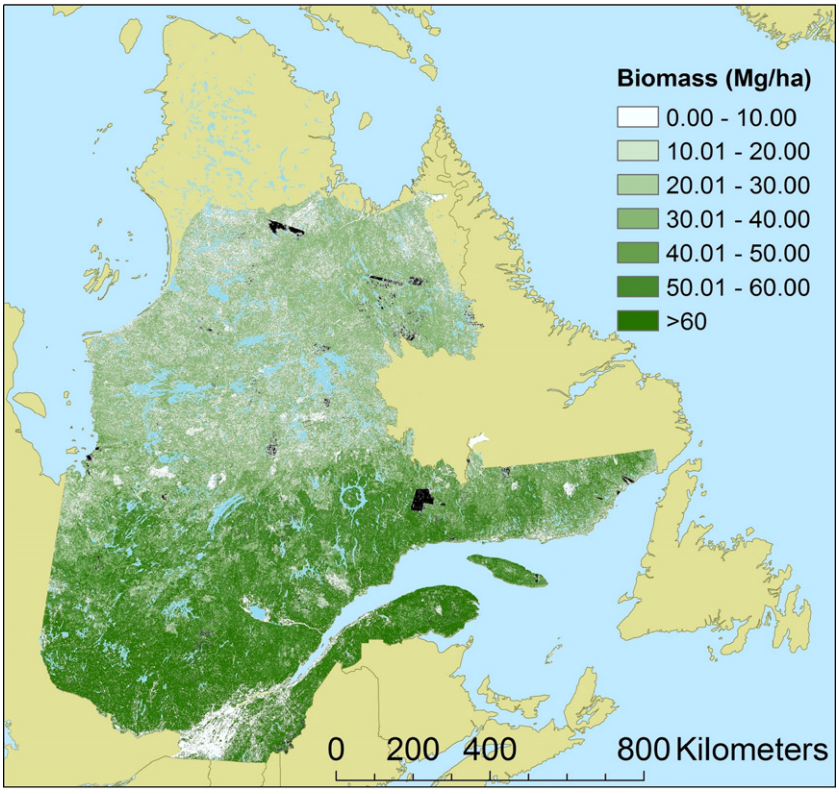

Fig. 7. Map of aboveground dry biomass for the Province of Québec. Black areas indicate areas with no land cover data. Blue areas indicate open water.

Table 6
GLAS estimates of mean aboveground dry biomass and their standard errors for commercial and non-commercial forested areas and for the entire Province south of treeline as well as comparison with other sources in the literature

		Mean biomass (Mg ha^{-1})
Source	This Study	
	Commercial forests	53.9±2.6
	Non-commercial forests	26.3±1.1
	Province (south of treeline)	39.0±2.2
	Boreal forest	34.9±2.2
	Northern temperate forest	59.8±1.7
	Other data sources	
	Québec commercial forests (MRNFPQ Inventory)	83.7±0.3
	North American boreal forest[a]	41.8±10.1
	Laurentian highlands[b]	68.0±33.3
	St. Lawrence Lowlands[b]	58.0±72.5
	Eastern North America Temperate Deciduous Forest[b]	80.5±13.8
	Canada's commercial forests[c]	90.9
	Québec's commercial forests[c]	70

[a] Botkin and Simpson (1990), errors determined from 95% C.I.
[b] Botkin et al. (1993), errors determined from 95% C.I.
[c] Penner et al. (1997).

using the cover type within each vegetation zone within each orbit, with covariance included. The maximum biomass had a mean of 95.2 Mg ha^{-1} for the dense deciduous stands of the northern hardwood vegetation zone, while the minimum biomass was 21.1 Mg ha^{-1} in the treed wetlands in the non-commercial portion of the boreal forest. The standard errors were sometimes quite high, reaching 26.0 Mg ha^{-1} for the open deciduous stands of the treed tundra. In general, large standard errors are associated with relatively rare cover types, i.e., land cover classes not frequently found in the landscape. For example, only two orbits crossed this open deciduous cover type in this treed tundra zone.

The biomass estimates for the study area were applied to reclassify the Landsat EOSD land cover map by vegetation zone so as to create a map of estimated aboveground dry biomass for the entire Province of Québec south of treeline (Fig. 7). This method allowed us to calculate the average biomass for different sub-categories such as the commercial or non-commercial forests (north or south of the red line in Fig. 3a) or for boreal forest only. These results are compared to other biomass results found in the literature for Québec and other parts of the boreal forest (Table 6).

A biomass estimate was also developed from the 17,382 MRNFPQ ground plots. The estimates were first calculated for each cover type within the three vegetation zones of the portion of the study area in which forest is commercially exploited. Weighted sums of the estimates for each cover type within each vegetation zone were computed to obtain vegetation zone estimates. Weights were calculated from the area covered by each cover type in each vegetation zone. In a similar way, a provincial estimate was computed by calculating weighted sums of the vegetation zone estimates using weights developed from the area of each vegetation zone. The 17,832 plots were treated assuming that the plots represent a random sample, i.e., assuming that the plots in different strata are independent, and not including covariance. Our GLAS-based estimates for the commercial forest are lower than that from the MRNFPQ provincial inventory and from Penner et al. (1997), but this is understandable since our methodology does not measure the same population as these other two studies (see Discussion). Our estimates for the Province's boreal forests are also lower, but closer to the estimates of Botkin and Simpson (1990) for the North American boreal forest (Table 6).

5. Discussion

We developed a method to evaluate aboveground dry biomass of the boreal and northern hardwood forests of Québec by combining airborne lidar sampling, ground plots, and the GLAS large footprint lidar. The results demonstrate the power of our approach for large-scale biomass estimation but also allow us to assess some of the different sources of error and uncertainty.

5.1. Linking ground plots, PALS and GLAS

The use of PALS airborne sampling was crucial for linking the ground plot data to the GLAS satellite data because of the lack of ground plot data sets within GLAS footprints. Even though PALS acquires data only along a narrow flight line, its capacity to extract information on regional forest biomass has been well demonstrated by the fairly strong correlations (Table 2) between the different ground plot and PALS variables (see also Nelson et al., 2003a). During our aircraft campaign, PALS was flown two times over 65 GLAS footprints allowing analysis of the variability in the PALS data over the same GLAS footprint. This allowed us to assess the degree to which a profiling airborne laser altimeter illuminating contiguous 0.45 m footprints along a line crossing a much larger GLAS footprint actually represents the GLAS waveform. A regression was run between PALS $\bar{h}_{qa}$ of the first and the second flight over these footprints and this analysis produced a R^2 of 0.74. This result suggests that PALS measurements are repeatable on an area the size of a GLAS footprint. However, the fact that 26% of the variation remains unexplained suggests the extent to which reliance on a profiler (as opposed to a scanner) adds noise to the regression relationships.

This result, along with the regression between PALS and ground inventory plots, could be improved by better quantification, or mitigation, of the positioning error of the PALS shots. A quantification of this error would be a significant improvement in future studies using this type of measurement approach. A better approach would be to use scanning LiDAR to collect data over both the ground plots and along the GLAS transects, but such data are more expensive to obtain, and were beyond the means of this investigation. Also, a sampling design for TSPs that placed them in more homogenous areas could improve the biomass estimates. Most of the TSPs from MRNFPQ are currently located in heterogeneous stands with the intent to represent the variability of the forest with a minimum of plots. On the other hand, the EOSD plots are located within homogenous areas to decrease the impact of GPS location errors. These latter types of plots are better suited for geospatial studies. Unfortunately, the EOSD plots were all located only in the boreal forest areas around Radisson and Chibougamau, so the provincial TSPs were essential for the objectives of the current study.

The stratified PALS equations demonstrated the capacity of this sensor to estimate biomass over a wide range of cover types of different heights and densities, thus demonstrating the versatility of the LiDAR technology for biological applications and confirming previous results achieved with this sensor (Nelson et al., 2004). On the other hand, some of the stratified equations have extreme intercept values resulting in either high or negative biomass values when the related PALS metric values were close to zero.

These R^2 values obtained for the stratified regressions between the ground plot biomass and the PALS metrics (Table 2) are typical of what is obtained for lidar profilers for forests. A major reason for the remaining uncertainty is that lidar profilers only obtain estimates of tree heights, they are not effective for estimating stem diameter which can be an important independent variable for predicting tree volume and biomass. In the current study, we used the generic (non-stratified) plot-to-PALS equation (R^2=0.65) for applying PALS to GLAS. We cannot assess the implications of the prediction error of the regressions between ground plots and PALS on the large-scale estimates, although our results suggest that the impact of the prediction error for the PALS and GLAS regression on regional biomass estimates was small (e.g., standard errors in column 1 versus column 2 of Table 4).

The equation developed to link PALS estimates of biomass to GLAS variables provided a unique opportunity to assess aboveground dry

biomass in Québec. Several filters had to be applied to the GLAS records to get to this final generic equation (Table 3, Eq. (2), Fig. 6). Among these filters, an upper limit to vegetation height ($h_{14} < 40$ m) had to be applied, revealing that even after filtering for saturation index, significant noise remained in the data. This filter allowed us to significantly reduce the noise in the data, while the number of GLAS records remove was minor (Table 3). The use of 100% pure EOSD land cover classes within the GLAS footprints also improved the results significantly (Table 3). Indeed, biomass in heterogeneous stands seems to be harder to capture using variables related to vegetation height (Clark et al., 2004). Land cover heterogeneity also likely resulted in some classification errors in the land cover map.

The generic equation uses the variable, range, representing the SRTM terrain index. This variable quantifies the elevation difference between the highest and lowest nine 90 m SRTM pixels surrounding the GLAS pulse. Integration of the DEM with GLAS data, as proposed by Lefsky et al. (2005), requires a lot of data processing and introduces errors from various sources. In fact, the DEM from SRTM is affected by the vegetation height itself (Sun et al., 2003). Improved methods to better describe ground topography should be considered in future work, particularly variables derived from the waveform itself. Lefsky et al. (2007) described an alternate method to infer and mitigate the effects of topography on the GLAS waveform response. They reported results using the extent (in meters) of the trailing edge of the waveform. Such an approach has the potential to move spaceborne LiDAR analyses away from the need for ancillary data sets such as regional DEMs or SRTM data. However, having multiple waveform-derived variables in the same equation may lead to collinearity problems.

5.2. Extrapolation to larger spatial scales

The extrapolation process described in this article is a statistically robust approach for using a generic GLAS equation to predict biomass and its variability over large areas. Despite the R^2 value of 0.59 for the filtered generic PALS-GLAS equation, prediction error from this regression is not necessarily a significant source of uncertainty in the regional biomass estimates. If the prediction error in the regression estimates respects the basic assumptions of linear regression by being homogenously distributed and having a mean of zero, the cumulative regression prediction error generated from a large number of samples should tend to cancel itself out.

The impact of adding regression noise to determine the impact of prediction error from the regression on biomass estimates induced differences in the overall biomass estimates for each vegetation zone and for the entire Province of around 10 to 20% (Table 4), even though we believe that changes to regional mean estimates should be near zero. The inclusion of the prediction error is not well suited to a square-root model since the addition of a negative prediction error to a low biomass estimate occasionally results in a negative sqrt (biomass) value. Some proportion of these negative values (~4 to 6% of the 104,044 GLAS shots) result in biomass estimates that are negative and back-transformation of these negative values results in a positive bias. Consequently, we did not include the prediction error from the regressions in the final method for calculating the large area biomasses. Nevertheless, the use of the square-root model remains a valid choice for this study since it was the only model that mitigated heteroskedasticity. Even though prediction error from the regressions was not included in the large-scale biomass estimates, we did include the covariance among cover types within and between GLAS orbits and the covariance among vegetation zones (Table 4) in the estimates. We believe that this solution represents the most realistic estimator of sampling variation.

The estimates of biomass by vegetation zones for all EOSD forested land covers allowed us to construct a biomass map for the entire study area (Fig. 7). The biomass map created is derived directly from the original land cover map in that biomass values were derived for each land cover class in each vegetation zone. This map is not designed to provide accurate estimates of biomass at a local scale. However, it is useful for describing and quantifying the distribution of aboveground biomass over the entire area of the Province and within each vegetation zone. Previously, such estimates were only available for the commercial forest area, only 46% of the forested area of the Province and, according to our results, containing only 64% of the Province's total aboveground biomass (Table 4).

5.3. Detecting changes in biomass over time

Assessing our ability to detect changes in biomass between two points in time using lidar is an important area of research that lies largely beyond the scope of the current study. Issues that need to be addressed include differences in instruments used, phenological differences, geolocation problems, slope issues, ground detection under increasingly dense canopies, etc.). We do not believe that differences with ground-based tree height measurements are a major limitation to the application of lidar for large-scale biomass estimation since our capacity to assess tree height from the ground has significant limitations. Ultimately, whether estimates of changes in biomass obtained by repeated lidar measurements are “good enough” for a given purpose is a value judgment made in a given specific context. The precision required for detection of change will depend on what a person is trying to estimate (see Tables 5 and 6 for different estimates made at different scales). From Table 5, we see that 10 Mg differences in aboveground biomass could be detectable for a given cover type with around 1 m difference in average height, something that could be within the accuracy level of most current airborne lidar systems. For larger scale estimates, even smaller height differences could be used to infer significant biomass changes because of the large number of samples.

5.4. Comparing biomass estimates to other data sources

We derived a biomass estimate uniquely from the MRNFPQ ground plot data gathered for our study. This resulted in an estimate of 81.9 Mg ha^{-1} with a standard error of 0.5 Mg ha^{-1} for commercial forests of Québec (Table 6). Penner et al. (1997) also developed an estimate of biomass based on available forest inventory. Her results were in the same range as what is found using the MRNFPQ plots database, although no estimates of errors were provided (Table 6). In both these cases, the sampling strategy was not well suited to obtaining an estimate of aboveground biomass over a large area. These types of ground plot inventories are designed to determine the commercial potential of forests in a local area at a given point in time and consequently the ground inventory plots are usually skewed to areas with large amounts of forest biomass. To get a statistically reliable estimate of biomass over a large area, the sampling must be designed to cover all the strata in a systematic way (Botkin and Simpson, 1990; Botkin et al., 1993). Interestingly, the Botkin and Simpson (1990) estimates of biomass for the North American boreal forest are much closer to the results of our study for the boreal forest of Québec (Table 6). In the Northern Temperate forest zone in our study, our results were again closer (Table 6) to those reported for the Laurentian Highlands and St. Lawrence Lowlands in Botkin et al. (1993) (Table 6).

5.5. Timing of data acquisitions

While the airborne laser dataset was acquired in August 2005, the corresponding satellite LiDAR dataset was acquired in October 2003. This difference of two years between the two acquisitions could introduce errors in the sampling for areas where disturbance caused the biomass to change dramatically between the two sampling dates.

However, disturbed areas over these two years represent less than 1% of the entire territory (Natural Resources of Canada, 2004, 2005, 2006) and so we do not consider this to be an important source of error. The principal drawback of the time difference between the two datasets likely resides in the seasonal differences. PALS data was acquired during the growing season, allowing a good identification of the top of the canopy. In open and sparse forests, the ground was almost always very easy to identify. In dense broadleaf forests, on the other hand, it was sometimes more time consuming to identify the ground, but it was feasible.

In contrast, GLAS data was acquired during the period of leaf fall. The height of the top of canopy might have been underestimated for those stands characterized by broadleaf trees during or after leaf fall (Hollaus et al., 2006). There is a need to quantify the impact of the presence or absence of leaves on GLAS data as this might be a source of inconsistency in the regression between airborne and spaceborne LiDAR data. Another reason why PALS data were acquired during the growing season was to avoid snow. In October, in northern latitudes or at higher elevations, snow can already be present on the ground. As GLAS data was acquired over a broad range of elevations and latitudes, there is a chance that snow was present in some areas and this would have biased our estimates downward.

In the absence of airborne lidar data, it is still possible to make use of GLAS data for forest biomass estimation by taking ground plot measurements where GLAS pulses have struck the ground. This approach, combined with MODIS land cover data, has been successfully used to estimate timber volume in Siberia (Ranson et al., 2007).

5.6. Accessory data sets

Data from multiple sources were used in the analyses described in this paper, i.e., LiDAR data, EOSD Landsat ETM+ land cover, and SRTM DEM. The land cover map is an unsupervised classification product that allowed us to investigate the stratification and reporting of results over the entire study area. This classification contains errors (preliminary accuracy assessment using digital aerial photos over various regions in Québec indicates a maximum overall classification accuracy around 75–80%) and the filter for preserving GLAS shots within homogenous patch of land cover was intended to partially offset these classification errors. As discussed previously, LiDAR data can be affected by topographic slope and this can introduce a bias in the estimates of vegetation structure. It is difficult to further improve these accessory data sets since they are already highly refined products. On the other hand, the fusion of PALS or GLAS data with raw multispectral or SAR data represents an interesting research opportunity for further study (Slatton et al., 2001).

5.7. Future sensor development

The GLAS instrument onboard ICESat offers an unprecedented opportunity to work with a uniform, near-global LiDAR dataset. Even if the principal objective of the ICESat mission is to retrieve information on ice elevation, we have shown that the secondary objective of monitoring land and vegetation can also be achieved to a fair extent. There is a clear need for a spaceborne LiDAR mission that is specifically designed for biomass applications. An increase in sampling density along an orbital track and a reduction of the distance between orbits would contribute to improve the sampling density for vegetation assessment. As demonstrated in this study, more orbits within a stratum reduce the variability of biomass estimates significantly. The reduction of the size of the footprint would also improve sampling accuracy. It has already been demonstrated that 25-m wide airborne LiDAR footprints have the ability to describe vegetation structure accurately (Blair et al., 1999; Drake et al., 2002) and a smaller footprint will mitigate the effects of topography. Finally, data quality would likely be improved if it were obtained during the growing season.

6. Conclusions

This study has demonstrated that GLAS data can be used to achieve large-scale monitoring of aboveground dry biomass at large spatial scales. The 4.9 (±0.27) Pg of aboveground biomass in Québec south of the treeline represents a C content equal to about 32% of the global annual fossil fuel CO_2 emissions in 2004 and is 6.6 and 54.4 times greater than the 2005 emissions for Canada and Québec, respectively. Furthermore, commercial forests represented only 64% of the total aboveground biomass, thus highlighting the importance of remote northern forests to C sequestration. Spaceborne LiDAR data has a strong potential to provide global information on vegetation state. We have demonstrated that by using an airborne profiling laser system to link ground data to the GLAS data, space-based forest inventories of northern forests can be efficiently used to provide aboveground biomass information at large spatial scales for applications to forest management, carbon cycle science, and assessing the impacts of global change.

Acknowledgments

This study was supported by the NASA Carbon Cycle Science Program, the Natural Science and Engineering Research Council of Canada (NSERC), the Canadian Foundation for Climate and Atmospheric Sciences (CFCAS), and the Canadian Forest Service (CFS). We are also grateful for the support offered by the UMBC-GEST Program and Caelum Research Corporation. The provincial ground sample plots were acquired through the MRNFPQ forest inventory program who kindly shared this valuable dataset for the study through the ECOLEAP project. The authors would also like to thank Stephen Côté and Philippe Villemaire from the Laurentian Forestry Centre (CFS) as well as Guoqing Sun from the Biospheric Sciences Branch of NASA Goddard Space Flight Centre for their important contributions to the data processing. This project could not have been realized without the statistical support from Huor Ung of the Laurentian Forestry Centre and from Tim Gregoire of the Yale School of Forestry and Environmental Studies. We also thank the very capable C-182/PALS pilot, Ryan Collins, for flying us safely from 45°N to 62.5°N and back, with many stops in between.

Appendix A

Calculation of biomass and the associated variances and covariances was conducted as follows.

(a) Biomass per hectare (Mg ha^{-1}) for the kth orbit, jth cover type, ith vegetation zone:

$$\hat{b}_{ijk} = \frac{\sum_{p=1}^{n_{ijk}} \hat{b}_{ijkp}}{n_{ijk}} \quad (A-1)$$

where $\hat{b}_{ijk}$=estimate of biomass ha^{-1} on the kth orbit, jth cover type, and ith vegetation zone, n_{ijk}=number of GLAS pulses along the kth orbit which intercepts the jth cover type in the ith vegetation zone, and $\hat{b}_{ijkp}$=biomass (Mg ha^{-1}) of the pth GLAS pulse on the kth orbit in the jth cover type in vegetation zone i, where $\hat{b}_{ijkp}$ is a regression estimate based on the GLAS waveform and SRTM variables.

(b) Cover type estimates within a vegetation zone:

$$\hat{b}_{ij} = \sum_{k=1}^{n_{ij}} w_{ijk}\, \hat{b}_{ijk} \quad (A-2)$$

where $\hat{b}_{ij}$=biomass per hectare (Mg ha^{-1}) in the jth cover type, ith vegetation zone, n_{ij}=number of orbits intercepting the jth cover type

in the ith vegetation zone. Note that n_{ij} may be less than the number of orbits actually acquired over a vegetation zone i; in fact this may be a regular occurrence for rare cover types. w_{ijk} = weight of the kth orbit that intercepts the jth cover type in vegetation zone i,

$$w_{ijk} = \frac{n_{ijk}}{\sum_{k=1}^{n_{ij}} n_{ijk}} \qquad \sum_{k=1}^{n_{ij}} w_{ijk} = 1.0$$

(c) Variance of the cover type estimates:

$$\hat{\mathrm{var}}\left(\hat{b}_{ij}\right) = \frac{\sum_{k=1}^{n_{ij}} w_{ijk}\left(\hat{b}_{ijk} - \hat{b}_{ij}\right)^2}{n_{ij} - 1} \qquad (A-3)$$

This equation is based on an equation presented in DeVries (1986), as reported in Nelson et al. (2004). The $\sqrt{\hat{\mathrm{var}}\left(\hat{b}_{ij}\right)}$ reports the standard error (not the standard deviation) of $\hat{b}_{ij}$.

Estimates of total biomass are calculated by multiplying per hectare estimates of the cover type by a_{ij}, the area of cover type j in vegetation zone i, based on the Landsat land cover classification.

(d) Vegetation zone estimates within the Province:

$$\hat{b}_i = \sum_{j=1}^{n_i} w_{ij}\,\hat{b}_{ij} \qquad (A-4)$$

where $\hat{b}_i$ = biomass per hectare in vegetation zone i, n_i = number of cover types in vegetation zone i,

$$w_{ij} = \frac{a_{ij}}{a_i}, \quad \sum_{j=1}^{n_i} w_{ij} = 1.0,$$

a_{ij} = area (ha) of cover type j in vegetation zone i, and a_i = area (ha) of vegetation zone i.

(e) Variance of the vegetation zone estimates:

$$\hat{\mathrm{var}}\left(\hat{b}_i\right) = \sum_{j=1}^{n_i} w_{ij}^2\left(\hat{\mathrm{var}}\left(\hat{b}_{ij}\right)\right) \qquad (A-5)$$

Eq. (A-5) would be valid if GLAS observations between cover types within vegetation zones were independent within and between orbits. Such an argument for independence is weak given the close proximity of GLAS pulses along and, occasionally, between orbits. To account for this expected lack of independence, we add a weighted covariance term to Eq. (A-5).

$$\hat{\mathrm{var}}\left(\hat{b}_i\right) = \sum_{j=1}^{n_i} w_{ij}^2\,\hat{\mathrm{var}}\left(\hat{b}_{ij}\right) + 2\sum_{j=1}^{n_i-1}\sum_{m=j+1}^{n_i} w_{ij}\,w_{im}\,\hat{\mathrm{cov}}\left(\hat{b}_{ij},\hat{b}_{im}\right) \qquad (A-6)$$

where $\hat{b}_{ij}$ and $\hat{b}_{im}$ = biomass ha^{-1} in cover types j and m, respectively, w_{ij} and w_{im} = area weights associated with cover types j and m, as defined in Eq. (A-4),

$$\hat{\mathrm{cov}}\left(\hat{b}_{ij},\hat{b}_{im}\right) = \frac{\sum_{k=1}^{n_{i(j,m)}}\sum_{l=1}^{n_{i(j,m)}}\left(\hat{b}_{ijk} - \hat{b}_{ij}\right)\left(\hat{b}_{iml} - \hat{b}_{im}\right)}{n_{i(j,m)}^2 - 1} \qquad (A-7)$$

and $n_{i(j,\,m)}$ = the number of orbits in vegetation zone i that intercept both cover types j and m.

Variances computed with and without covariances (i.e., Eqs. (A-5) and (A-6)) are compared to see to what degree covariances, in fact, need to be taken into account (Table 4).

The last variable, $n_{i(j,\,m)}$, brings up two important points. First, variances and associated covariances are calculated based only on observations from orbits that have intercepted the cover type(s) of interest. An orbit that does not intercept a particular cover type is not treated as an observation of zero biomass, rather it is treated as "no information" and excluded from the calculation in question. Second, given the first point and the fact that, *a priori*, we do not know the exact location of a GLAS orbital track or the location of a particular pulse within that orbit, the GLAS sample is considered a post-stratified sample. This means that many of the variables found in these equations, e.g., the n's and w's, are random variables rather than fixed quantities that are known or set prior to the collection of the sample. Therefore, we condition our sampling on the particular post-stratified sample that we selected. This allows us to treat n's and w's as fixed quantities even though, in fact, they are random variables.

As a result of this condition, however, all inferences and conclusions drawn from study results are limited to the specific set of flight lines, orbits, and to the particular post-stratification processing used in this study. Despite this limitation on inference, we present these results in the expectation that studies conducted employing similar techniques will note similar trends regarding the large impact of including covariances on overall variances and the difficulty of estimating precise biomass estimates for rare cover types.

(f) Provincial biomass estimates:

$$\hat{b} = \sum_{i=1}^{n} w_i\,\hat{b}_i \qquad (A-8)$$

where $w_i = \frac{a_i}{a}$ $\sum_{i=1}^{n} w_i = 1.0,$ n = number of vegetation zones in the Province, and a = area (ha) of the Province (south of treeline).

(g) Variance of the Provincial biomass estimate:

$$\hat{\mathrm{var}}\left(\hat{b}\right) = \sum_{i=1}^{n} w_i^2\left(\hat{\mathrm{var}}\left(\hat{b}_i\right)\right) \qquad (A-9)$$

As with Eq. (A-5), Eq. (A-9) can be calculated assuming that GLAS-based biomass estimates between vegetation zones are independent, or covariances can be incorporated to account for non-independent sample observations. Provincial-level variances which include the covariance terms are calculated as follows, and both are reported in Table 4.

$$\hat{\mathrm{var}}\left(\hat{b}\right) = \sum_{i=1}^{n} w_i^2\left(\hat{\mathrm{var}}\left(\hat{b}_i\right)\right) + 2\sum_{i=1}^{n-1}\sum_{q=i+1}^{n} w_i\,w_q\,\hat{\mathrm{cov}}\left(\hat{b}_i,\hat{b}_q\right) \qquad (A-10)$$

where $\hat{b}_i$ and $\hat{b}_q$ = biomass ha^{-1} in vegetation zones i and q, respectively, w_i and w_q = area weights associated with vegetation zones i and q, as defined for Eq. (A-8),

$$\hat{\mathrm{cov}}\left(\hat{b}_i,\hat{b}_q\right) = \frac{\sum_{k=1}^{n_{i,q)}}\sum_{l=1}^{n_{i,q)}}\left(\hat{b}_{ik} - \hat{b}_i^*\right)\left(\hat{b}_{ql} - \hat{b}_q^*\right)}{n_{(i,q)}^2 - 1} \qquad (A-11)$$

$\hat{b}_{ik}$ = average biomass ha^{-1} on the kth orbit in the ith vegetation zone, across all n_i cover types,

$$\hat{b}_{ik} = \frac{\sum_{j=1}^{n_i}\sum_{p=1}^{n_{ijk}}\hat{b}_{ijkp}}{\sum_{j=1}^{n_i} n_{ijk}},$$

$\hat{b}_{ql}$ = average biomass ha^{-1} on the lth orbit in the qth vegetation zone, across all n_q cover types,

$$\hat{b}_{ql} = \frac{\sum_{j=1}^{n_q}\sum_{p=1}^{n_{qjl}} \hat{b}_{qjlp}}{\sum_{j=1}^{n_q} n_{qjl}},$$

$\hat{b}_i^*$ and $\hat{b}_q^*$ are estimates of biomass ha^{-1} for vegetation zones i and q calculated using the number of GLAS pulses as weighting factors rather than the estimated area of each land cover,

$$\hat{b}_i^* = \sum_{j=1}^{n_i} w_{ij}^* \hat{b}_{ij} \qquad \text{and} \qquad \hat{b}_q^* = \sum_{j=1}^{n_q} w_{qj}^* \hat{b}_{qj}$$

$$w_{ij}^* = \frac{\sum_{k=1}^{n_{ij}} n_{ijk}}{\sum_{j=1}^{n_i}\sum_{k=1}^{n_{ij}} n_{ijk}}, \quad w_{qj}^* = \frac{\sum_{l=1}^{n_{qj}} n_{qjl}}{\sum_{j=1}^{n_q}\sum_{l=1}^{n_{qj}} n_{qjl}}, \quad \sum_{j=1}^{n_i} w_{ij}^* = \sum_{j=1}^{n_q} w_{qj}^* = 1.0,$$

for six vegetation zones, 23 cover types, and 97 orbits, and $n_{(i,\,q)}$ = the number of orbits in the Province that intercept both vegetation zones i and q.

The substitution of $\hat{b}_i^*$ and $\hat{b}_q^*$ for $\hat{b}_i$ and $\hat{b}_q$, respectively, in Eq. (A-11) is done to ensure comparable weighting. In the covariance calculation, weighting must be uniform, i.e., weights must be based either on number of pulses per orbit or on Landsat ETM-based areas. Since our concern is within- and between-orbit biomass estimates, a pulse-based weighting system is necessary and therefore the vegetation zone estimates must also be calculated using the pulse-based weights. Mixing weights within the covariance calculation (i.e., differencing $\hat{b}_{ik}$ with $\hat{b}_i$ and $\hat{b}_{ql}$ with $\hat{b}_q$) led to inflated covariance estimates. If GLAS pulses were proportionally allocated to each cover type and vegetation zone based on their areas, then $\hat{b}_i$ and $\hat{b}_q$ could be substituted for $\hat{b}_i^*$ and $\hat{b}_q^*$ in Eq. (A-11), but this was not the case in this study. For whatever reason, e.g., clouds over selected cover types and/or cover type or vegetation zone characteristics that might lead to GLAS waveform saturation, GLAS pulses were not proportionally allocated. As a result, modified vegetation zone means ($\hat{b}_i^*$ and $\hat{b}_q^*$) were employed to calculate vegetation zone covariances.

Covariances can be positive or negative. In this study, situations occasionally arose where negative covariances overwhelmed the variance sums. This resulted in situations where the variance–covariance sums in Eqs. (A-6) or (A-10) were negative. If a particular variance turned negative with the addition of the covariance term, then the covariance term was ignored and the variance without the covariance term (Eqs. (A-5) and (A-9)) was reported.

References

Abshire, J. B., Sun, X., Riris, H., Sirota, J. M., McGarry, J. F., Palm, S., et al. (2005). Geoscience Laser Altimeter System (GLAS) on the ICESat Mission: On-orbit measurement performance. *Geophysical Research Letters*, *32*, L21S02.

Beaubien, J., Cihlar, J., Simarg, G., & Latifovic, R. (1999). Land cover from multiple thematic mapper scenes using a new enhancement-classification methodology. *Journal of Geophysical Research*, *104*(D22), 27909–27920.

Bergeron, O., Margolis, H. A., Black, T. A., Coursolle, C., Dunn, A. L., Barr, A. G., et al. (2007). Comparison of CO_2 fluxes over three boreal black spruce forests in Canada. *Global Change Biology*, *13*, 89–107.

Blair, J. B., Rabine, D. L., & Hofton, M. A. (1999). The Laser Vegetation Imaging Sensor: A medium-altitude, digitisation-only, airborne laser altimeter for mapping vegetation and topography. *ISPRS Journal of Photogrammetry & Remote Sensing*, *54*, 115–122.

Botkin, D. B., & Simpson, L. G. (1990). Biomass of the North America Boreal Forest: A step toward accurate global measures. *Biogeochemistry*, *9*, 161–174.

Botkin, D. B., Simpson, L. G., & Nisbet, R. A. (1993). Biomass and carbon storage of the North American deciduous forest. *Biogeochemistry*, *20*, 1–17.

Brenner, A. C., Zwally, H. J., Bentley, C. R., Csathó, B. M., Harding, D. J., Hofton, M. A., et al. (2003). Derivation of range and range distributions from laser pulse waveform analysis for surface elevations, roughness, slope, and vegetation heights. *GLAS ATBD Version 4.1* 92 pp.

Castro, K. L., Sanchez-Azofeifa, G. A., & Rivard, B. (2003). Monitoring secondary tropical forests using space-borne data: Implications for Central America. *International Journal of Remote Sensing*, *24*, 1853–1894.

Clark, M. L., Clark, D. B., & Roberts, D. A. (2004). Small-footprint lidar estimation of sub-canopy elevation and tree height in a tropical forest landscape. *Remote Sensing of Environment*, *91*, 68–89.

DeVries, P. G. (1986). *Sampling Theory for Forest Inventory.* Berlin: Springer-Verlag 420 pp.

Drake, J. B., Dubayah, R. O., Knox, R. G., Clark, D. B., & Blair, J. B. (2002). Sensitivity of large-footprint lidar to canopy structure and biomass in a neotropical rainforest. *Remote Sensing of Environment*, *81*, 378–392.

Dubayah, R. O., & Drake, J. B. (2000). Lidar remote sensing for forestry. *Journal of Forestry*, *98*, 44–46.

Fournier, R. A., Luther, J. E., Guindon, L., Lambert, M. -C., Piercy, D., Hall, R. J., et al. (2003). Mapping aboveground tree biomass at the stand level from inventory information: Test cases in Newfoundland and Québec. *Canadian Journal of Forest Research*, *33*, 1846–1863.

Gregoire, T.G., Lin, Q.F., Boudreau, J., & Nelson, R. (in press). Regression estimation following the square-root transformation of the response. *Forest Science*.

Harding, D. J., & Carabajal, C. C. (2005). ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure. *Geophysical Research Letters*, *32*, L21S10.

Helmer, E.H. & Lefsky, M.A. (2006), Forest canopy heights in Amazon River basin forests as estimated with the Geoscience Laser Altimeter System (GLAS). In: Aguirre-Bravo, C., Pellicane, P.J., Burns, D.P., Draggan, S., Eds. Monitoring Science and Technology Symposium: Unifying Knowledge for Sustainability in the Western Hemisphere Proceedings RMRS-P-42CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. p. 802–808.

Hollaus, M., Wagner, W., Eberhöfer, C., & Karel, W. (2006). Accuracy of large-scale canopy heights derived from LiDAR data under operational constraints in a complex alpine environment. *ISPRS Journal of Photogrammetry & Remote Sensing*, *60*, 323–338.

Hyde, P., Dubayah, R., 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*, 63–73.

Lambert, M. -C., Ung, C. -H., & Raulier, F. (2005). Canadian national tree aboveground biomass equations. *Canadian Journal of Forest Research*, *35*, 1996–2018.

Lefsky, M. A., Cohen, W. B., Acker, S. A., Parker, G. G., Spies, T. A., & Harding, D. (1999). Lidar remote sensing of the canopy structure and biophysical properties of Douglas-fir western hemlock forests. *Remote Sensing of Environment*, *70*, 339–361.

Lefsky, M. A., Harding, D., Cohen, W. B., Parker, G., & Shugart, H. H. (1999). Surface lidar remote sensing of basal area and biomass in deciduous forests of eastern Maryland, USA. *Remote Sensing of Environment*, *67*, 83–98.

Lefsky, M. A., Cohen, W. B., Parker, G. G., & Harding, D. J. (2002). Lidar remote sensing for ecosystem studies. *BioScience, 52, 1*, 19–30.

Lefsky, M. A., Harding, D. J., Keller, M., Cohen, W. B., Carabajal, C. C., Del Bom Espirito-Santo, F., et al. (2005). Estimates of forest canopy height and aboveground biomass using ICESat. *Geophysical Research Letters*, *32*, L22S02.

Lefsky, M.A., Keller, M., Pang, Y., deCamargo, P., & Hunter, M.O. (2007). Revised method for forest canopy height estimation from the Geosciences Laser Altimeter System waveforms. *Journal of Applied Remote Sensing, 1*, 013537.

Means, J. E., Acker, S. A., Harding, D. J., Blair, J. B., Lefsky, M. A., Cohen, W. B., et al. (1999). Use of large-footprint scanning airborne lidar to estimate forest stand characteristics in the Western Cascades of Oregon. *Remote Sensing of Environment*, *67*, 298–308.

Ministère des Ressources naturelles, de la Faune et des Parcs (MRNFPQ) (2003a), Normes d'inventaire forestier – Placettes-échantillons permanentes. Gouvernement du Québec, Ministère des Ressources naturelles, de la Faune et des Parcs, Direction des Inventaires Forestiers.

Ministère des Ressources naturelles, de la Faune et des Parcs (MRNFPQ) (2003b), Zones de végétation et domaines bioclimatiques du Québec. Gouvernement du Québec, Ministère des Ressources naturelles, de la Faune et des Parcs, Direction des Inventaires Forestiers, Code de diffusion, 2003–3043.

Myers, R. H. (1990). *Classical and modern regression with applications*, 2nd Ed. Boston, MA: PWS-Kent Publishing Co. 488 pg.

Natural Resources Canada (2004). *The state of Canada's forests 2003–2004.* : Government of Canada.

Natural Resources Canada (2005). *The state of Canada's forests 2004–2005.* : Government of Canada.

Natural Resources Canada (2006). *The state of Canada's forests 2005–2006.* : Government of Canada.

Naesset, E. (1997). Determination or mean tree height of forest stands using airborne laser scanner data. *ISPRS Journal of Photogrammetry & Remote Sensing*, *52*, 49–56.

Næsset, E. (2004). Practical large-scale forest stand inventory using a small-footprint airborne scanning laser. *Scandinavian Journal of Forest Research*, *19*, 164–179.

Næsset, E. (2007). Airborne laser scanning as a method in operational forest inventory: Status of accuracy assessments accomplished in Scandinavia. *Scandinavian Journal of Forest Research*, *22*(5), 433–442.

Naesset, E., & Gobakken, T. (2008). Estimation of above- and below-ground biomass across regions of the boreal forest zone using airborne laser. *Remote Sensing of Environment*, *112*, 3079–3090.

Næsset, E., Gobakken, T., Holmgren, J., Hyyppä, H., Hyyppä, J., Maltamo, M., et al. (2004). Laser scanning of forest resources: The Nordic Experience. *Scandinavian Journal of Forest Research*, *19*, 482–499.

Nelson, R., Valenti, M. A., Short, A., & Keller, C. (2003). A multiple resource inventory of Delaware using airborne laser data. *BioScience*, *53*(10), 981–992.

Nelson, R., Parker, G., & Hom, M. (2003). A portable airborne laser system for forest inventory. *Photogrammetric Engineering & Remote Sensing*, *69*, 267–273.

Nelson, R., Short, A., & Valenti, M. (2004). Measuring biomass and carbon in Delaware using an airborne profiling LIDAR. Scandinavian Journal of Forest Research, 19, 500–511. Erratum. 2005. *Scandinavian Journal of Forest Research*, *20*, 283–284.

Nelson, R., Næsset, E., Gobakken, T., Ståhl, G., & Gregoire, T. (2008). Regional forest inventory using an airborne profiling LiDAR. *Japanese Journal of Forest Planning*, *13*, 287–294.

Penner, M., Power, K., Muhairwe, C., Tellier, R., & Wang, Y. (1997). Canada's forest biomass resources: Deriving estimates from Canada's forest inventory. *Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre. Information Report BC-X-370* 33pp.

Popescu, S. C., Wynne, R. H., & Nelson, R. F. (2003). Measuring individual tree crown diameter with LiDAR and assessing its influence on estimating forest volume and biomass. *Canadian Journal of Remote Sensing*, *29*, 564–577.

Ranson, K. J., Sun, G., Kovacs, K., & Kharuk, V. I. (2004). Landcover attributes from ICESat GLAS data in central Siberia. *Geoscience and Remote Sensing Symposium, 2004. IGARSS '04. Proceedings. 2004 IEEE International*, *Vol. 2*. (pp. 753–756).

Ranson, K. J., Sun, G., Kovacs, K., & Kharuk, V. I. (2004). Use of ICESat GLAS data for forest disturbance studies in central Siberia. *Proceedings, International Geoscience and Remote Sensing Symposium (IGARSS), 2004. IEEE International*, *Vol. 3*. (pp. 1936–1939).

Ranson, K. J., Kimes, D., Sun, G., Nelson, R., Kharuk, V., & Montesano, P. (2007). Using MODIS and GLAS data to develop timber volume estimates in central Siberia. *Proceedings, International Geoscience and Remote Sensing Symposium (IGARSS), 2007. Barcelona, Spain* (pp. 2306–2309). doi:10.1109/IGARSS.2007.4423302

Raupach, M. R., Rayner, P. J., Barrette, D. J., Defries, R. S., Heimann, M., Ojima, D. S., et al. (2005). Model–data synthesis in terrestrial carbon observation: Methods, data requirements and data uncertainty specifications. *Global Change Biology*, *11*, 378–397. doi:10.1111/j.1365-2486.2005.00917.x

Reutebuch, S. E., Andersen, H. -E., & McGaughey, R. J. (2005). Light Detection and Ranging (LIDAR): An emerging tool for multiple resource inventory. *Journal of Forestry*, *103*, 286–292.

Schutz, B. E., Zwally, H. J., Shuman, C. A., Hancock, D., & DiMarzio, J. P. (2005). Overview of the ICESat mission. *Geophysical Research Letters*, *32*, L21S01.

Slatton, K. C., Crawford, M. M., & Evans, B. L. (2001). Fusing interferometric radar and laser altimeter data to estimate surface topography and vegetation heights. *IEEE Transactions on Geoscience and Remote Sensing*, *39*, 2470–2482.

Sun, G., Ranson, K. J., Kharuk, V. I., & Kovacs, K. (2003). Validation of surface height from shuttle radar topography mission using shuttle laser altimeter. *Remote Sensing of Environment*, *88*, 401–411.

Thomas, V., Treitz, P., McCaughey, J. H., & Morrison, I. (2006). Mapping stand-level forest biophysical variables for a mixedwood boreal forest using lidar: An examination of scanning density. *Canadian Journal of Forest Research*, *36*, 34–47.

Wulder, M. A., Dechka, J. A., Gillis, M. A., Luther, J. E., Hall, R. J., Beaudoin, A., & Franklin, S. E. (2003). Operational mapping of the land cover of the forested area of Canada with Landsat data: EOSD land cover program. *The Forestry Chronicle*, *79*, *6*, 1075–1083.

Zwally, H. J., Schutz, R., Bentley, C., Bufton, J., Herring, T., Minster, J., Spinhirne, J., & Thomas, R. (2003). GLAS/ICESat L2 global land surface altimetry data V026, 15 October to 18 November 2003. *Boulder, CO. National Snow and Ice Data Center. Digital media*.