Model development for the estimation of aboveground biomass using a lidar-based sample of Canada's boreal forest

Christopher W. Bater[1,*], Michael A. Wulder[2,*], Nicholas C. Coops[1,*], Chris Hopkinson[3],
Samuel B. Coggins[4], Eric Arsenault[5], André Beaudoin[6], Luc Guindon[6], R.J. Hall[5],
Philippe Villemaire[6], & Murray Woods[7].

[1]Integrated Remote Sensing Studio, Department of Forest Resources Management,
University of British Columbia.
[2]Pacific Forestry Centre, Canadian Forest Service, Natural Resources Canada.
[3]Applied Geomatics Research Group, Centre of Geographic Sciences.
[4]Nisga'a Lisims Government.
[5]Northern Forestry Centre, Canadian Forest Service, Natural Resources Canada.
[6]Laurentian Forestry Centre, Canadian Forest Service, Natural Resources Canada.
[7]Southern Science and Information Section, Ontario Ministry of Natural Resources.
[*]For additional information please contact:
cbater@interchange.ubc.ca; nicholas.coops@ubc.ca; or mwulder@nrcan.gc.ca

1. Introduction

The northern forested areas of Canada are largely unmanaged and not subject to inventories with the same level of detail or regularity as southern forested regions. In an effort to augment monitoring and inventory activities, airborne light detection and ranging (lidar) has been employed to obtain plot-level information over a sample of Canada's northern forests. During the summer of 2010, a series of 34 transects were flown over a total length of more than 24,000 km, spanning the width of the Canadian landmass from Nova Scotia to the Yukon, and crossing eight ecozones and 13 UTM zones.

Following data acquisition, a suite of plot-level lidar vegetation metrics were calculated. To develop estimates of forest attributes such as biomass, however, field data were required from the range of conditions found across the region. Co-located field plots and lidar data are required for algorithm development. To that end, datasets were acquired from Quebec, Ontario and the Northwest Territories. The Quebec field plots were coincident with the lidar transects, existing lidar and field data were used for Ontario and Northwest Territories, with careful analysis required due to different lidar acquisition parameters. In this paper we describe the development of regression models for large area estimates of various tree aboveground biomass components using field and lidar datasets of uncommon provenance, with significant differences both in terms of the environments in which they were collected, and the characteristics of the field and lidar surveys. The equations developed are deemed suitable for application and extrapolation across the national series of lidar transects.

2. Methods

2.1 Field data

Field data were obtained from the Northwest Territories, Ontario and Quebec (Figure 1). The datasets are described in detail below. The equations are intended to be relevant over the boreal area of Canada's forest, not the entire forest area depicted.

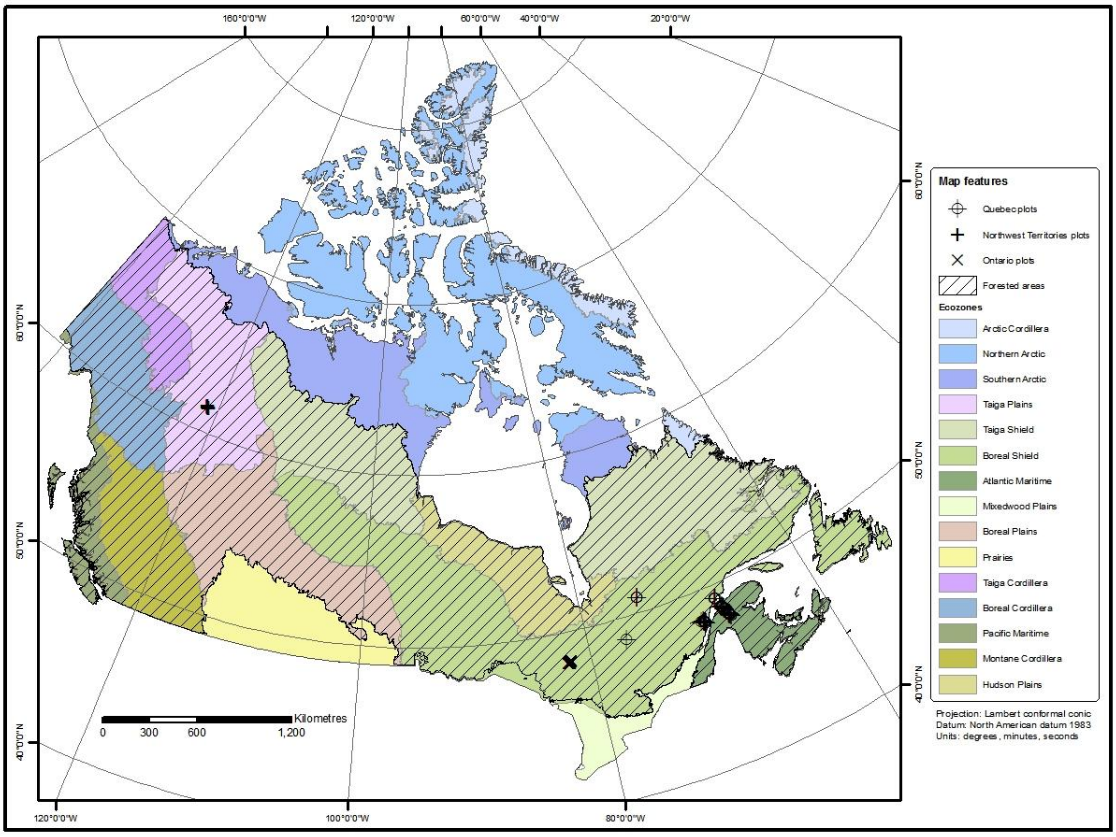

Figure 1. Canadian ecozones and locations of field plots by province.

2.1.1 Northwest Territories

Field data were provided by the Northern Forestry Centre, Natural Resources Canada, and consisted of individual tree measurements, including species, height, and diameter at breast height (DBH). Data were collected from square 20 x 20 m ($400 m^2$) plots.

2.1.2 Ontario

Field data were provided by the Ontario Ministry of Natural Resources and collaborators. As with those from the Northwest Territories, the data consisted of individual tree measurements, including species, height, and DBH collected from 11.28 m radius ($400 m^2$) circular plots located in the Romeo Malette Forest near the city of Timmins. Plots can be coarsely stratified into mixedwood, intolerant hardwoods, Jack pine, and black spruce.

2.1.3 Quebec

Field data were provided by: Direction des inventaires forestiers Forêt Québec, Ministère des Ressources Naturelles et de la Faune. Tree lists were provided for each plot by species and stems per hectare grouped in 2 cm DBH classes. Tree heights were measured for four to six representative trees per plot. Tree numbers for each DBH class were then calculated based on a $400 m^2$ plot, and hardwood/softwood-specific equations were developed to estimate height for all stems in the plots based on DBH.

2.1.4 Field-based biomass estimates

Biomass components were calculated using national DBH- and height-based “all species” equations originally provided by Lambert et al. (2005) and subsequently updated by Ung et al. (2008). The equations were developed using archival data collected through the Energy for the

Forest Research (ENFOR) program, and provide estimates for wood, bark, stem, branches, foliage, crown, and total biomass. National “all species” equations were selected because no spatially explicit species information is readily available for the majority of the lidar transects. The DBH- and height-based equations take the following form:

$$y_{wood} = \beta_{wood1} D^{\beta wood2} H^{\beta wood3} + e_{wood}$$

$$y_{bark} = \beta_{bark1} D^{\beta bark2} H^{\beta bark3} + e_{bark}$$

$$y_{stem} = \hat{y}_{wood} + \hat{y}_{bark} + e_{stem}$$

$$y_{foliage} = \beta_{foliage1} D^{\beta foliage2} H^{\beta foliage3} + e_{foliage}$$

$$y_{branches} = \beta_{branches1} D^{\beta branches2} H^{\beta branches3} + e_{branches}$$

$$y_{crown} = \hat{y}_{foliage} + \hat{y}_{branches} + e_{crown}$$

$$y_{total} = \hat{y}_{wood} + \hat{y}_{bark} + \hat{y}_{foliage} + \hat{y}_{branches} + e_{total}$$

Where y_i is the dry biomass component i of a living tree (kg); i is wood, bark, stem, foliage, branches, crown, or total; $\hat{y}_i$ is the prediction of y_i; D is the DBH (cm); H is tree height (m); $\beta_{jk}$are model parameters with coefficient estimates b_{jk}; j is wood, bark, foliage or branches; k = 1, 2 or 3; and e_i are error terms. Stem, crown, and total biomass are then obtained by summing their respective components (Ung et al., 2008). Model parameter estimates are provided in Table 1; the field-based biomass estimates are summarized in Table 2.

Table 1. Model parameter estimates for all-species biomass equations published by Ung et al. (2008)

Parameter	**Estimate**	**Standard error**
β_{wood1}	0.0283	0.0004
β_{wood2}	1.8298	0.0075
β_{wood3}	0.9546	0.0101
β_{bark1}	0.012	0.0003
β_{bark2}	1.6378	0.017
β_{bark3}	0.7746	0.0233
$\beta_{branches1}$	0.0338	0.0008
$\beta_{branches2}$	2.6624	0.0182
$\beta_{branches3}$	–0.5743	0.0233
$\beta_{foliage1}$	0.1699	0.0036
$\beta_{foliage2}$	2.3289	0.0184
$\beta_{foliage3}$	–1.1316	0.0235

Table 2. Summary of biomass components estimated using field data and national “all species” equations provided by Lambert et al. (2005) and Ung et al. (2008).

Province	**Wood biomass (kg)**			
	Mean	**Min**	**Max**	**Std.Dev.**
Northwest Territories	4,667	1,125	10,665	2,218
Quebec	2,646	24	7,312	1,650
Ontario	4,364	495	10,499	1,942
All Groups	4,073	24	10,665	2,070

Province	**Bark biomass (kg)**			
	Mean	**Min**	**Max**	**Std.Dev.**
Northwest Territories	643	184	1,378	269
Quebec	377	5	1,066	223
Ontario	621	85	1,298	245
All Groups	576	5	1,378	264

Province	**Stem biomass (kg)**			
	Mean	**Min**	**Max**	**Std.Dev.**
Northwest Territories	5,310	1,308	12,043	2,486
Quebec	3,023	28	8,378	1,870
Ontario	4,985	580	11,797	2,185
All Groups	4,649	28	12,043	2,333

Province	**Branch biomass (kg)**			
	Mean	**Min**	**Max**	**Std.Dev.**
Northwest Territories	840	250	1,977	363
Quebec	612	9	1,698	371
Ontario	864	96	1,864	344
All Groups	807	9	1,977	366

Province	**Foliage biomass (kg)**			
	Mean	**Min**	**Max**	**Std.Dev.**
Northwest Territories	188	81	352	59
Quebec	149	5	377	80
Ontario	232	35	664	92
All Groups	206	5	664	90

Province	**Crown biomass (kg)**			
	Mean	**Min**	**Max**	**Std.Dev.**
Northwest Territories	1,028	331	2,329	418
Quebec	761	13	1,938	445
Ontario	1,095	132	2,260	423
All Groups	1,014	13	2,329	444

Province	**Total tree biomass (kg)**			
	Mean	**Min**	**Max**	**Std.Dev.**
Northwest Territories	6,338	1,640	14,373	2,887
Quebec	3,784	42	10,261	2,312
Ontario	6,080	712	14,057	2,557
All Groups	5,663	42	14,373	2,739

2.2 Lidar data

2.2.1 Northwest Territories

Lidar data were provided by the Northern Forestry Centre, Natural Resources Canada. Data were acquired by the Applied Geomatics Research Group (AGRG) in August of 2007 using an Optech ALTM 3100. During the survey, the power supply for the laser diode was in a state of rapid but steady degradation, and thus the sampling efficiency of this site was well below optimal. This directly impacted the probabilities associated with ground and canopy returns (due to reduced pulse power) and data density was up to five times below what it should have been. Returns were classed as first, intermediate, last, or single. Because of the power supply issues, many points fell into the single return category and very few were classified into ground and non-ground returns. The AGRG undertook additional processing to create a ground return point dataset that was in turn used to create a 1 m raster of ground elevation. The lidar "all-returns" point data was then overlaid onto this grid and the ground elevations subtracted to generate a canopy point data set.

2.2.2 Ontario

Lidar data were provided by the Ontario Ministry of Natural Resources and collaborators. Data were in standard las format and classified as ground and non-ground returns.

2.2.3 Quebec

Lidar data were extracted from this project’s dataset where the transects and Quebec field plots intersected. The desired survey specifications included a flying height of 1,200 magl, a 70 kHz pulse repetition frequency, scan angles of +/-15^{o}, and a nominal pulse density of ~ 2.8 returns/m^{2}, with the understanding that flying conditions might necessitate lower or higher flying heights. Scan angle was generally kept fixed at 15^{o}.

2.2.4 Lidar data processing

Lidar data were processed using FUSION software (McGaughey, 2010). A suite of standard plot-level lidar metrics were calculated, including mean first return height; standard deviation, coefficient of variation, and the 95th percentile of first return heights; percentage of first returns above 2m; and percentage of first returns above the first return mean height . Plot sizes were 400 m^{2} and either circular or square depending on their province of origin. Table 3 provides a plot-level summary of return numbers by province.

Table 3. Summary of lidar return numbers within field plots.

Province	**Number of plots**	**First returns above 2m**		**All returns**	
		Mean	**Std.Dev**	**Mean**	**Std.Dev.**
Northwest Territories	40	88	55	189	97
Quebec	41	391	145	745	113
Ontario	120	1086	255	1881	405
All groups	201	746	474	1315	785

2.2 Statistical analysis

Given the large number of possible predictor variables (lidar metrics) which were often strongly intercorrelated, a subset of candidate predictors were selected based on their relatively low intercorrelations and their biological relevance (e.g. Li et al., 2008). Prior to analysis, both predictor variables and forest attributes were transformed to their natural logarithms. Best subsets multiple linear regression analyses were then preformed for each forest attribute. Akaike's Information Criterion (Akaike, 1973) was employed to select the most parsimonious models (Posada and Buckley, 2004). Following the development of regression models, outliers were assessed based on their standardized residuals and leverage values. If a large gap existed between the majority of leverage values and one or a few with very high values, the outliers were removed from the final regression model.

To assess the quality of the final models, the lidar-derived biomass estimates were back-transformed from their natural logarithms to arithmetic units using a bias correction factor (Sprugel, 1983). Bias, mean absolute error (MAE) and root-mean-square error (RMSE) were then calculated from the residuals between predicted and observed values.

3. Results

Results of the multiple linear regression analyses are shown in Tables 4 and 5, and Figure 2. The final regression models explained between 36% and 78% of the variance in the various biomass components. Of the biomass components, foliage biomass was the most poorly predicted, with an adjusted R^2 of 0.36. However, 78% and 76% of the variance in wood and total aboveground biomass were explained, respectively (Table 4).

Table 4. Summary statistics for the multiple linear regression models for the field-based biomass components (dependent variables) and lidar canopy height and cover metrics (predictors).

Dependent variable	**n plots**	**Multiple R**	**Multiple R^2**	**Adjusted R^2**	**df model**	**df residual**	***F***	***p***
Foliage biomass	197	0.61	0.37	0.36	2	194	56.08	0.00
Branch biomass	198	0.81	0.65	0.65	2	195	182.28	0.00
Crown biomass	198	0.77	0.60	0.59	2	195	145.63	0.00
Bark biomass	198	0.86	0.73	0.73	2	195	267.69	0.00
Wood biomass	198	0.88	0.78	0.78	3	194	231.02	0.00
Stem biomass	198	0.88	0.78	0.77	3	194	224.43	0.00
Total tree biomass	198	0.87	0.76	0.76	3	194	206.79	0.00

Bias was negligible for all models (Table 5). Total aboveground biomass had a bias of 56.8 kg, or less than 1% of the mean. Foliage biomass, the smallest biomass component, but also the most poorly predicted in terms of its precision, had a MAE of 52 kg, or 25% of the mean value observed in the plots. Conversely, the largest component, wood biomass, had an MAE of 749 kg, or 18% of the mean value observed in the plots, while total tree biomass had an MAE of 1029 kg, or 18% of the mean observed value (Table 5). Foliage biomass had an RMSE of 67 kg or 33%. Wood biomass had an RMSE of 990 kg, or 24% of the mean value observed in the plots, while total aboveground biomass had an RMSE of 1353 kg or 24% (Table 5).

Of the candidate lidar-derived predictor variables, the 95th percentile of first return heights and the percentage of first returns above the first return mean height were employed to estimate the crown biomass components, while mean first return height, the coefficient of variation of the first return heights, and the percentage of first returns above 2 m were employed to estimate the stem components and total tree biomass.

Table 5. Fit statistics for the multiple linear regression models for the field-based biomass components (dependent variables) and lidar canopy height and cover metrics (predictors). Field plots sizes were 200 m^2.

Dependent variable	**Units**	**BCF**[1]	**Bias**	**MAE**[2]	**RMSE**[3]
Foliage biomass	kg	1.059	0.87	51.96	67.22
Branch biomass	kg	1.043	4.83	163.38	218.60
Crown biomass	kg	1.045	5.61	212.79	282.12
Bark biomass	kg	1.04	7.00	105.84	136.98
Wood biomass	kg	1.039	51.82	749.33	990.40
Stem biomass	kg	1.039	58.56	854.28	1125.35
Total tree biomass	kg	1.038	56.80	1029.34	1353.21

[1] Bias correction factor (Sprugel, 1983)
[2] Mean absolute error
[3] Root mean square error

4. Discussion

The objective of this paper was to describe the development of large area estimates of tree biomass components by combining three different field and lidar datasets. The results indicate that, using biologically relevant lidar-derived predictor variables based on height, vertical structural complexity, and cover, over 70% of the variance in stem and total tree biomass can be explained; this despite significant differences in lidar acquisitions, particularly the dissimilarities in return densities (Table 3). In comparison, biomass estimates employing purposefully collected lidar and field data over comparably small areas typically explain between 70% and 90% of the variance in tree biomass (e.g. Popescu, 2007; Li et al., 2008; Næsset and Gobakken, 2008).

It should be noted that the usage of “all-species” national allometric equations to estimate biomass from field-measured height and DBH is not optimal, as prediction errors grow with increased model generalization (Hall and Case, 2008). Nonetheless, without additional information on species within the lidar transects, the equations do provide a means to estimate forest attributes of critical importance.

Acknowledgements

The authors thank D. Pitt (Canadian Wood Fibre Centre) , P. Treitz (Queen’s University), and personnel of the Northern Forestry Centre of Natural Resources Canada, the Ontario Ministry of Natural Resources, and the Direction des inventaires forestiers Forêt Québec, Ministère des Ressources Naturelles et de la Faune for providing insights as well as lidar and field data to support this initiative.

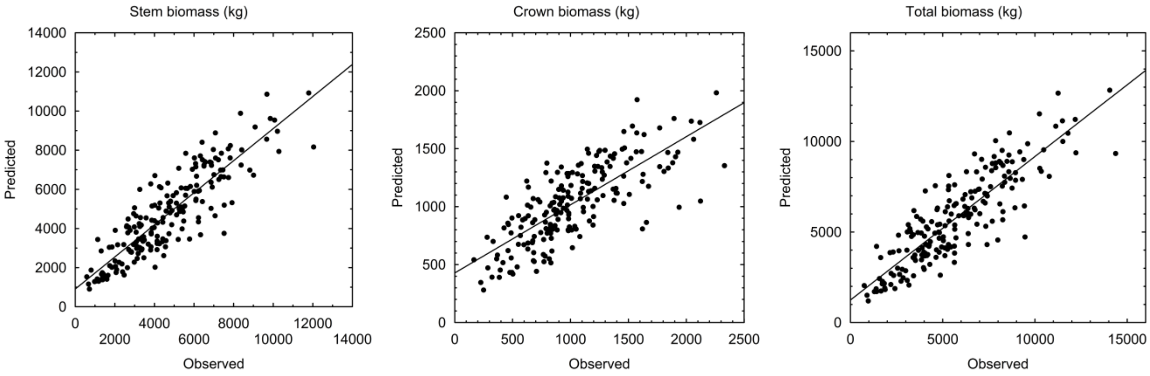

Figure 1. Observed and predicted values of stem, crown and total biomass components estimated using plot-level lidar metrics. Ground plots were 400 m^2.

References

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle, in: Second International Symposium on Information Theory. Academiai Kiado, Budapest, Hungary, pp. 267-281.

Hall, R.J., Case, B.S., 2008. Assessing prediction errors of generalized tree biomass and volume equations for the boreal forest region of west-central Canada. Canadian Journal of Forest Research 38, 878–889.

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

Li, Y., Andersen, H.E., McGaughey, R., 2008. A comparison of statistical methods for estimating forest biomass from light detection and ranging data. Western Journal of Applied Forestry 23, 223-231.

McGaughey, R.J., 2010. FUSION/LDV: Software for LIDAR data analysis and Visualization (No. Version 2.90). Pacific Northwest Research Station, Forest Service, United States Department of Agriculture.

Næsset, 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.

Popescu, S., 2007. Estimating biomass of individual pine trees using airborne lidar. Biomass and Bioenergy 31, 646–655.

Posada, D., Buckley, T.R., 2004. Model Selection and Model Averaging in Phylogenetics: Advantages of Akaike Information Criterion and Bayesian Approaches Over Likelihood Ratio Tests. Systematic Biology 53, 793 -808.

Sprugel, D.G., 1983. Correcting for bias in log-transformed allometric equations. Ecology 64, 209-210.

Ung, C.H., Bernier, P., Guo, X.J., 2008. Canadian national biomass equations: new parameter estimates that include British Columbia data. Canadian Journal of Forest Research 38, 1123-1132.