Estimating spatial variation in Alberta forest biomass from a combination of forest inventory and remote sensing data

J. Zhang1, S. Huang2, E. H. Hogg3, V. Lieffers1, Y. Qin4, and F. He1

1Department of Renewable Resources, University of Alberta, Edmonton, Alberta, T6G 2H1, Canada
2Forest Management Branch, Alberta Department of Environment and Sustainable Resource Development, 8th Floor, 9920-108 Street, Edmonton, Alberta, T5K 2M4, Canada
3Canadian Forest Service, Natural Resources Canada, Northern Forestry Centre, 5320-122 Street, Edmonton, Alberta, T6H 3S5, Canada
4Geographic Information Science Center of Excellence, South Dakota State University, Brookings, South Dakota, 57007 USA

Correspondence to: J. Zhang (zjian@ualberta.ca)

Received: 29 October 2013 – Published in Biogeosciences Discuss.: 4 December 2013
Revised: 17 March 2014 – Accepted: 13 April 2014 – Published: 27 May 2014

Abstract. Uncertainties in the estimation of tree biomass carbon storage across large areas pose challenges for the study of forest carbon cycling at regional and global scales. In this study, we attempted to estimate the present aboveground biomass (AGB) in Alberta, Canada, by taking advantage of a spatially explicit data set derived from a combination of forest inventory data from 1968 plots and spaceborne light detection and ranging (lidar) canopy height data. Ten climatic variables, together with elevation, were used for model development and assessment. Four approaches, including spatial interpolation, non-spatial and spatial regression models, and decision-tree-based modeling with random forests algorithm (a machine-learning technique), were compared to find the “best” estimates. We found that the random forests approach provided the best accuracy for biomass estimates. Non-spatial and spatial regression models gave estimates similar to random forests, while spatial interpolation greatly overestimated the biomass storage. Using random forests, the total AGB stock in Alberta forests was estimated to be \( 2.26 \times 10^9 \) Mg (megagram), with an average AGB density of \( 56.30 \pm 35.94 \) Mg ha\(^{-1}\). At the species level, three major tree species, lodgepole pine, trembling aspen and white spruce, stocked about \( 1.39 \times 10^9 \) Mg biomass, accounting for nearly 62\% of total estimated AGB. Spatial distribution of biomass varied with natural regions, land cover types, and species. Furthermore, the relative importance of predictor variables on determining biomass distribution varied with species. This study showed that the combination of ground-based inventory data, spaceborne lidar data, land cover classification, and climatic and environmental variables was an efficient way to estimate the quantity, distribution and variation of forest biomass carbon stocks across large regions.

1 Introduction

Forest ecosystems, accounting for over 80\% of terrestrial vegetation biomass, play a major role in balancing the regional and global carbon (C) budget and analyzing the fate of carbon dioxide produced by the burning of fossil fuels and forest harvesting (Dixon et al., 1994; Brown et al., 1997; Houghton et al., 2009). The accurate estimation of broad-scale biomass C stocks has been a focus of regional and global C cycle studies and has attracted the interest of researchers, forest managers and policymakers over the past half century. A proper assessment of actual and potential roles of forest ecosystems in the global C cycle requires accurate information about carbon storage and change over space and time (Botkin and Simpson, 1990). However, such accurate information has been lacking at regional and global scales.

A number of approaches have been developed to estimate the spatial distribution of biomass C stocks, ranging from allometric regression equations or biomass expansion...
factors (e.g., Brown, 1997; Cairns et al., 1997; Schroeder et al., 1997), local and regional scale forest inventories (Monserud et al., 2006; Blackard et al., 2008), simulation modeling (Tans et al., 1990; Ciais et al., 1995), to methods using only remote sensing or combined with inventory data (Hall et al., 2011; Myneni et al., 2001; Wulder et al., 2008; Yemshanov et al., 2012). However, the estimates obtained by these different approaches are often inconsistent. For example, Houghton et al. (2001) compared several biomass estimates for the Brazilian Amazon forests and found very low agreement across the estimates, with the values ranging from 39 to 93 gigatons (Gt) of carbon. Blackard et al. (2008) compared several estimates of C pools in living forest biomass of the continental U.S. forests and found that satellite-image-based estimation was two times higher than estimates based on inventory data.

Forest ground-based inventory laid out in a statistically sound design is considered to be the optimum approach to accurately and precisely measuring forest biomass C stocks (Schroeder et al., 1997; Ketterings et al., 2001; Brown, 2002). However, sampling a sufficient number of trees to represent the size and species distribution in a forest is extremely time-consuming and costly. The task becomes much harder for accurate estimation of biomass C stocks over large areas. For carbon estimation at the regional scale, most researchers tend to measure biomass on a few small, generally non-randomly selected plots, and use various prediction approaches (e.g., spatial interpolation techniques, and regression models), to estimate regional biomass C stocks based on observed values of these small sampling plots. However, inventories based on ground samplings are not free of problems. The first problem is related to the scarcity of ground-based inventory plots (Botkin and Simpson, 1990; Wulder et al., 2008; Pan et al., 2011). The lack of sufficient and high-quality sample plots has been identified as a major barrier to the development of robust biomass estimates and to the subsequent validation of these estimates (Wulder et al., 2008).

For example, in a recent report about global carbon storage, Pan et al. (2011) stated that estimates of C stocks are only limited to the 230 million hectares (Mha) of managed forest in Canada, leaving about 118 Mha of northern forests unaccounted for because of data paucity. The second problem is related to the fact that forest inventories tend to be conducted in forests that are considered to have commercial value, in other words, closed forests, with little regard to the open, drier forests, woodlands, or human-disturbed forests (Botkin and Simpson, 1990; Brown, 1997). This biased sampling design usually tends to overestimate biomass C stocks over large areas.

Light detection and ranging (lidar) is perhaps the most promising remote sensing technology for estimating biomass, because it directly measures vertical forest structure, such as canopy height and crown dimensions (Simard et al., 2011). Generally, lidar remote sensing has three platforms, including spaceborne, airborne, and ground-based platforms. While airborne or ground-based lidar methods have been intensely used for biomass-related measurements at the stand level or individual tree level, these methods are only feasible at local or small-regional scales, rarely at larger scales (Popescu et al., 2011). The main reason for this restriction is because the costs of airborne or ground-based lidar on data acquisition and analysis are still high to large extents (Popescu et al., 2011; Saatchi et al., 2011). For biomass and carbon estimation at the regional scale, spaceborne lidar with relatively low costs has advantages.

The boreal forest, containing large amounts of carbon in its biomass and soils, has been recognized as an important global contributor to the net balance of carbon exchange between the atmosphere and the biosphere (Kurz and Apps, 1999; Fyles et al., 2002; Pan et al., 2011). According to the Intergovernmental Panel on Climate Change (IPCC, 2007), climate warming in northern latitudes is occurring almost twice as rapidly as the global average. Climate warming in the boreal may be leading to increased frequency of wildfires (Harden et al., 2000), insect outbreaks (e.g., mountain pine beetle, Kurz et al., 2008) and regional drought events (Allen et al., 2010), thus influencing carbon stocks and dynamics (Kurz et al., 2008; Monserud et al., 2006; Pan et al., 2011). Since forest biomass is a key biophysical parameter in evaluating and modeling terrestrial carbon stocks and dynamics (Houghton et al., 2009), an accurate estimation of regional biomass is important for understanding boreal forests and their responses to climate warming. However, most of the previous studies for biomass estimation in the boreal were limited to the regions with high productivity and little disturbance (Botkin and Simpson, 1990). There is a lack of information about biomass in regions under other successional stages and different disturbance extents. In addition, for remote areas in northern boreal regions, few ground inventory data are available.

In this study, we estimated aboveground biomass stocks in the forest regions of Alberta, Canada, using recent forest inventory data from different forest monitoring networks and remote sensing data. Our inventory data had a large sample size, covered a broad range, and included different disturbance types, stand age groups, and successional stages. Our objectives were to (1) produce a spatially explicit data set of Alberta forest aboveground biomass stocks; (2) quantify the relative contributions of various predictor layers including climate variables, elevation and canopy height to the biomass stocks; and (3) assess the variability in estimation of biomass stocks using different techniques.

2 Methods

2.1 Study area

The forests of the Canadian province of Alberta (49–60° N, 110–120° W) cover an area of about 45 million hectare
(Mha), accounting for about 68% of the total area of the province. They encompass four natural regions: Boreal forest, Foothills, Rocky Mountains and Canadian Shield (Alberta Natural Regions Committee, 2006). These regions have short summers and long and cold winters. Mean annual temperature ranges from –2.6°C in the Canadian Shield to 1.7°C in the Foothills. Mean warmest month temperature ranges from 11.0°C in the Rocky Mountains to 16.6°C in the Canadian Shield, and mean coldest month temperature ranges from –25.1°C in the Canadian Shield to –11.7°C in the Rocky Mountains. Precipitation follows a summer-high continental pattern. Mean annual precipitation ranges from 380 mm in the Canadian Shield to about 800 mm in the Rocky Mountains. Elevations range from about 150 m near the Alberta–Northwest Territories border to over 3600 m in the Rocky Mountains. There is also large variation in climatic variables within the subregions of each natural region.

Variation in climate and topography in this area has produced a wide range of vegetation types across the province. In the Boreal region, deciduous aspen (Populus tremuloides), balsam poplar (Populus balsamifera), coniferous white spruce (Picea glauca), black spruce (Picea mariana) and jack pine (Pinus banksiana) forests are the dominant species. In the Foothills, mixed forests of aspen, lodgepole pine (Pinus contorta), white spruce and balsam poplar with variable understories are dominant on average sites at lower elevations, while at higher elevations lodgepole pine forests with less diverse understories are typical. In the Rocky Mountains, closed coniferous forests are dominant at lower elevations, and open coniferous stands and herbaceous alpine meadows are the major vegetation types at higher elevations. In the Canadian Shield, open jack pine, aspen and birch stands occur where the soil is sufficiently deep for retaining moisture and nutrients to sustain these species.

2.2 Data sources

We combined three different sources of ground-based inventory data for our current study, including 342 permanent sample plots (PSPs) from Alberta Environment and Sustainable Resource Development (ESRD), 635 PSPs from Weyerhaeuser Canada, 501 PSPs from West Fraser Mill Ltd., and 490 plots from Alberta Biodiversity Monitoring Institute (ABMI). In total, 1968 plots measured in the period 2000–2012 were selected to estimate current biomass carbon stock in the Alberta forest region (Fig. 1). For the selected plots with more than one census, only the latest inventory data was selected for the current analysis.

2.2.1 Permanent sample plots (PSPs)

The Alberta PSP network has maintained more than 2000 PSPs established and re-censused by the government and forest companies starting from the 1950s. Most PSPs were selected in forest regions with high productivity, and these plots were excluded from normal harvesting and other human disturbances. Plot sizes ranged from 400 m² (0.04 ha) to 8092 m² (0.81 ha) (mean: 0.12 ha). Within each PSP of ESRD, all living trees and standing dead trees (snags) with a tree height ≥ 1.3 m were tagged and recorded. Within each PSP of Weyerhaeuser Canada, all living trees and snags with DBH (diameter at breast height) ≥ 5 cm were measured. Within each PSP of West Fraser, all living trees and snags with DBH ≥ 7 cm were measured. These 1478 PSPs contained 206 213 living trees and 17 688 snags over the study period.

2.2.2 ABMI sampling plots

ABMI conducts a regional-scale, long-term monitoring program to track biodiversity status and trends in Alberta (http://www.ABMI.ca). ABMI collects information on thousands of terrestrial species and habitat structures at over one thousand sites spaced systematically on a 20 km grid evenly across the entire province. Terrestrial survey sites are established on each grid, with a random distance and directional offset of up to 5.5 km from this grid. Different from the PSP network, ABMI sampling plots were more randomly distributed and were thus more representative of the full range of forest
stand ages and disturbance regimes at the landscape level. The area of each ABMI plot is 1 hectare (100 × 100 m). On each site, all trees and snags with ≥ 25 cm DBH in four selected 25 × 25 m plots, all trees and snags with ≥ 7 cm DBH in four 10 × 10 m subplots, and all trees and snags in four 5 × 5 m further subplots were measured regardless of size. In total, 490 sampling plots with measurements for 36 059 living trees and 7046 snags were used in this study.

2.2.3 Canopy height data from spaceborne lidar

Spaceborne lidar top canopy height data for Alberta forest regions (Supplement A) were obtained from a global wall-to-wall canopy height map at 1 km spatial resolution (Simard et al., 2011). This map was produced by using the data acquired by the Geoscience Laser Altimeter System (GLAS), onboard the Ice, Cloud, and land Elevation Satellite (ICESat), in combination with seven global ancillary variables, which correspond to climate and vegetation characteristics. These variables included annual mean precipitation, precipitation seasonality, annual mean temperature, temperature seasonality, elevation, tree cover, and protection status.

2.2.4 Climatic variables

Climate data for Alberta forests were derived from the program CLIMATE WNA 4.70 (Wang et al., 2012). This program uses baseline climate data derived from monthly precipitation and temperature grids (Daly et al., 2008) based on interpolated climate data from weather stations for the period 1961–1990. The program includes a lapse-rate-based down-sampling to 1-km resolution and estimation of biologically relevant climatic variables. Based on input values for longitude and latitude of each inventory plot or each grid, we localized 10 climatic variables using the average values across the last 10 years (2000–2009) to describe local climatic conditions. The 10 climatic variables were as follows:

1. MAT: mean annual temperature (°C)
2. MWMT: mean warmest month temperature (°C)
3. MCMT: mean coldest month temperature (°C)
4. MAP: mean annual precipitation (mm)
5. MSP: mean summer (May to September) precipitation (mm)
6. AHM: annual heat: moisture index \((\text{MAT} + 10)/(\text{MAP}1000))\)
7. SHM: summer heat: moisture index \(((\text{MWMT})/(\text{MSP}1000)))\)
8. DD0: degree days below 0 °C, chilling degree days
9. DD5: degree days above 5 °C, growing degree days
10. DI: dryness index \((\text{DD5}/\text{MAP})\).

2.2.5 Alberta land cover map

The wall-to-wall land cover map of Alberta (ABMIw2wlcv2000v2.1) at 30 m spatial resolution was used for identifying forest lands in the study area (Supplement B, ABMI 2012). This map is a seamless GIS vector layer with nearly a million polygons describing the spatial distribution of land cover across Alberta, circa 2000, at the 1:125 000 scale. It consists of a mosaic of 977 556 non-overlapping polygons of various sizes, from 0.5 ha to thousands of ha. Each polygon represents a contiguous area relatively homogeneous in terms of land cover. The map is derived by applying a semantic and spatial generalization algorithm to combine two pre-existing land-cover products: the Canadian Forest Service’s Earth Observation for Sustainable Development (EOSD) map of the forested region, and Agriculture Agri-Food Canada’s map of the agricultural zone. This map consists of 11 land cover classes, including waters, snow/ice, rock/rubble, exposed land, developed, shrubland, grassland, agriculture, coniferous forest, broadleaf forest, and mixed forest. The overall accuracy of the map was estimated to be 75 % with 11 land cover classes (ABMI Remote Sensing Group, 2012).

2.2.6 Alberta natural region and subregion classification

To compare how tree biomass carbon stock varies in different forest regions, we used Alberta natural regions (NRs) and natural subregions (NSRs) classification system (Alberta Natural Regions Committee, 2006) as the basis for our comparisons. In Alberta, this system has informed provincial natural resource management activities since the 1970s. The current version of this system consists of 6 NRs and 21 NSRs. NRs, the largest mapped ecological units in this system, are defined geographically on the basis of landscape patterns, notably vegetation, soils and physiographic features. NSRs, subdivisions of a NR, are generally characterized by vegetation, climate, elevation, and latitudinal or physiographic differences within a given NR.

2.3 Data analysis

2.3.1 Estimation of aboveground biomass (AGB)

AGB was estimated for each living individual tree in all ground inventory plots using DBH- and height-based biomass allometric equations and tree-species-specific parameters provided by Lambert et al. (2005) and Ung et al. (2008). These equations were derived from thousands of trees sampled across Canada and allow the calculation of tree biomass carbon stock for different forest regions with high accuracy.
biomass (foliage, branches, stem bark, and stem wood) based on DBH measurements (for details see Lambert et al., 2005 and Ung et al., 2008). The form of the allometric equation is as follows:

\[ Y = \beta_1 D^{\beta_2} H^{\beta_3}, \]  

where \( Y \) is the biomass component of interest, diameter \( (D) \) is measured on each tree, height \( (H) \) is measured on a subsample tree in each plot, and \( \beta_1, \beta_2 \) and \( \beta_3 \) are parameters. For trees with missing height measure, the heights are estimated from local-species-specific height-diameter equations developed by Huang et al. (2009).

Since the three sources of PSP data had different minimum DBH cutoffs, we used the PSP data from ESRD to calculate average percentages of AGB at different DBH cutoffs. The percentages were used to calculate total AGB for Weyerhaeuser PSPs (0.4 % for trees with DBH < 5 cm) and West Fraser PSPs (0.9 % for trees with DBH < 7 cm). Total AGB of each PSP was summed up from all trees in each plot. Total aboveground biomass of each ABMI site was summed up from three parts: the biomass per hectare of trees \( \geq 25 \) cm DBH in the 25 \( \times \) 25 m plots, the biomass per hectare of trees \( 7-25 \) cm DBH in 10 \( \times \) 10 m subplots, and the biomass per hectare of trees < 7 cm DBH in 5 \( \times \) 5 m subplots.

### 2.3.2 Estimation of total biomass stock

Since total biomass stock has been a major concern of scientists, police makers and the public, it is important to report the estimation of total biomass stock. However, the detailed data for belowground biomass and debris biomass are sparse or not measured in our study region. Here, we used several published equations on the relationships between AGB and belowground biomass and debris biomass to estimate belowground and debris biomass.

We estimated belowground tree root biomass using previously developed regression equations developed for boreal forests by Li et al. (2003):

\[ BGB_s = 0.222 \cdot AGB_s \]  
\[ BGB_h = 1.576 \cdot AGB_h^{0.615}. \]

where \( BGB \) is the belowground biomass (coarse and fine roots), and \( AGB \) is the aboveground biomass; subscripts \( s \) and \( h \) are softwood and hardwood species groups, respectively.

To estimate debris biomass, we calculated the ratios of debris biomass (fine and coarse woody debris) to aboveground biomass for 90 study sites across Canada’s forest regions (Shaw et al., 2005). The average ratio of debris biomass to aboveground biomass was 5 %, which was used to estimate the debris biomass in the plots.

Estimates of belowground biomass, debris biomass and standing dead tree biomass were added to AGB to produce total biomass (including debris) estimates. The biomass carbon pool was calculated by multiplying a carbon biomass conversion factor of 0.5 to the total biomass (Schlesinger 1997). Because of the strong correlation between AGB and total biomass, we restricted our reporting to AGB in our main text. Total biomass estimates are reported in the supporting document (Supplement C and D).

### 2.3.3 AGB-environment correlations

We used simple Pearson correlations to explore covariation among AGB and 11 environmental variables. Because the presence of spatial autocorrelation in model residuals violates the assumption of data independence (Bini et al., 2009), Pearson correlations among AGB and biotic and abiotic variables were calculated after accounting for spatial autocorrelation using the R package MODTEST 1.4 (José Manuel Blanco Moreno, Universitat de Barcelona, Spain, personal communication, 2012).

### 2.3.4 Scaling up to the whole region

To get an accurate estimate of AGB distribution, four approaches were selected for our analysis, including spatial interpolation of direct field measurements, non-spatial regression model, spatial regression model, and decision-tree-based modeling with random forests algorithm (RF).

**Spatial interpolation methods:** These methods have been used for mapping forest variables (e.g. site index, standing volume, AGB, productivity, etc.) based on forest inventory data where these variables seemingly have spatial autocorrelation (e.g., Dungan, 1998; Freeman and Moisen, 2007; Viana et al., 2012). In this study, we compared several different approaches to find the “best” method for spatial interpolation of tree biomass. These approaches included ordinary kriging, standardized ordinary cokriging (with elevation as the covariate), inverse distance weighting, thin-plate smoothing splines, and partial thin-plate smoothing splines. A cross-validation analysis was used to evaluate effective parameters for these interpolation methods. The results with the highest \( R^2 \) in cross-validation analyses were finally selected. Kriging, cokriging and inverse distance weighting were calculated using the geostatistics software GS+ (http://www.gammadesign.com), and thin-plate smoothing splines were calculated using the R package “fields” (Fields Development Team 2006). After producing the biomass map for Alberta, we used the Alberta Natural Region GIS map to crop grassland and parkland regions, and the Alberta land cover map to crop the areas with the following land cover classes: waters, snow/ice, rock/rubble, exposed land, shrubland, grassland, and agriculture.
Non-spatial and spatial regression models: Two steps were used to estimate biomass stocks using canopy height data from spaceborne lidar. First, we used the data from the 1968 forest inventory plots to establish the relationships between total tree biomass and ground-measured top canopy height, climatic variables, and elevation. Both non-spatial multiple regression models (ordinary least squares, OLS) and spatial linear models (here “spatial simultaneous autoregressive error models (SARs)”, Kissling and Carl, 2008) were used. The SARs models allow the inclusion of the residual spatial autocorrelation of the data. Among these predictors, some of them were highly correlated. To reduce the risk of multi-collinearity, we used VIF (Variance Inflation Factors) for variable selection. The variables with VIF > 10, which represent high collinearity, were removed. The ‘best’ model is selected based on lower AIC (Akaike information criterion) and higher $R^2$. Second, we applied this selected model to estimate tree biomass density (Mg ha$^{-1}$) using lidar canopy height and other environmental variables in each 1 × 1 km grid in Alberta forest regions. All analyses were done using R language (R Core Team, 2013), and SARs were calculated using the R package “spdep” (version 0.5–33).

Decision-tree-based modeling with random forests algorithm (RF): This method is an ensemble machine learning technique, where many decision trees are constructed based on random sub-sampling of the given data set (Breiman, 2001). As one of the tree-based models, RF performs recursive partitioning of data sets, and makes no assumptions regarding the distribution of the input data. RF can capture non-linear relationships between the response variable (tree biomass in our study) and predictor variables (canopy height, climate, and other environmental variables in our study), and can deal with correlated variables while producing a low generalization error (Breiman, 2001). In addition, RF can be used to rank the importance of variables in a regression or classification problem in a natural way. In our study, this method was used to detect the relative importance of climate, topography and other environmental variables, and predict the distributions of forest biomass. All analyses were implemented in the R package “randomForest” (Liaw and Wiener, 2002).

2.3.5 Model accuracy assessment

Three well-known error statistics were calculated to measure the difference between the observed and predicted forest biomass, including mean absolute error (MAE), root mean-square error (RMSE), and the normalized root-mean-square error (NRMSE). They are defined as

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |\text{PRE}_i - \text{OBS}_i|$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\text{PRE}_i - \text{OBS}_i)^2}$$

$$\text{NRMSE} = 100 \times \frac{\text{RMSE}}{\text{OBS}_{\max} - \text{OBS}_{\min}}.$$  

where $\text{PRE}_i$ and $\text{OBS}_i$ denote the $i$th predicted and observed values, respectively. NRMSE is the RMSE divided by the range of observed values of a variable being predicted. The value is often expressed as a percentage, where lower values indicate less residual variance.

We randomly divided the 1968 ground inventory plots into training data (60 %) and testing data (40 %). These four approaches of AGB estimation were fitted with training data and evaluated with testing data. MAE, RMSE and NRMSE were calculated to assess model accuracy. This procedure was repeated 100 times, and the average values of these three model accuracy indicators were reported.

3 Results

3.1 Biomass variations among forest inventory plots

Direct field measurements yielded an estimate of 128.24 ± 76.64 Mg ha$^{-1}$ for the density of AGB for Alberta forests, with a range from nearly zero to 450.64 Mg ha$^{-1}$ in these inventory plots. For the PSP inventory plots only, the average biomass density estimate was 148.08 Mg ha$^{-1}$, which is more than double the density of 67.09 Mg ha$^{-1}$ for the ABMI inventory plots ($P<0.0001$, two-sample t test).

For forest inventory plots at the species level, the average AGB density estimates for lodgepole pine, trembling aspen, and white spruce were 75.79, 73.21, and 38.84 Mg ha$^{-1}$, respectively.

Based on our inventory data, we detected a large variation of AGB along forest stand ages (Fig. 2a, b). We classified these plots into four forest age groups (young, immature, mature, and old-growth forests). Old-growth forests (age > 120 years) and mature forests (80–120 years) had the highest average tree AGB, 148.76 and 148.26 Mg ha$^{-1}$, respectively. The average AGB density in immature forests (50–80 years) was 92.22 Mg ha$^{-1}$, and the average in young forests (< 50 years) was 48.28 Mg ha$^{-1}$.
3.2 AGB-environment correlations

The results of Pearson correlations after accounting for spatial autocorrelation showed that total AGB of each ground plot was strongly correlated with observed canopy height \((R^2 = 0.702, \ P < 0.001, \ \text{Table 1, Fig. 2c})\). Elevation also showed significant correlations with total AGB. Among the 10 climatic variables, most variables were highly correlated with others. MCMT (mean coldest month temperature) and DD0 (degree days below 0°C) had relatively stronger correlations with total AGB.

3.3 AGB estimates from four different approaches

We compared the results of four approaches for AGB estimation (Table 2, Fig. 3). The RF approach provided the best accuracy for AGB estimation \((R^2 = 0.62, \ \text{MAE} = 35.97 \text{ Mg ha}^{-1}, \ \text{RMSE} = 47.03 \text{ Mg ha}^{-1}, \ \text{NRMSE} = 62.40 \%)\) (Table 2). Non-spatial and spatial regression models performed nearly as well as the RF approach, while spatial interpolation had the poorest estimate \((R^2 = 0.30, \ \text{MAE} = 50.22 \text{ Mg ha}^{-1}, \ \text{RMSE} = 63.90 \text{ Mg ha}^{-1}, \ \text{NRMSE} = 84.20 \%)\). Total tree AGB estimation from spatial interpolation was \(4.68 \times 10^9 \text{ Mg} \), which was much larger than the...
Figure 3. The estimates of total AGB density (Mg ha\(^{-1}\)) using spatial interpolation, spatial multiple regression model, and decision-tree-based modeling with random forests algorithm (Projection: UTM zone = 11; spatial resolution: 1 km).

Table 2. Validation statistics for four different approaches for total tree AGB estimation.

<table>
<thead>
<tr>
<th>Methods for biomass estimation</th>
<th>(R^2)</th>
<th>MAE (Mg ha(^{-1}))</th>
<th>RMSE (Mg ha(^{-1}))</th>
<th>NRMSE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial interpolation</td>
<td>0.30</td>
<td>50.22</td>
<td>63.90</td>
<td>84.20</td>
</tr>
<tr>
<td>Non-spatial regression model</td>
<td>0.59</td>
<td>37.30</td>
<td>49.70</td>
<td>63.60</td>
</tr>
<tr>
<td>Spatial regression model</td>
<td>0.60</td>
<td>37.30</td>
<td>49.70</td>
<td>63.70</td>
</tr>
<tr>
<td>Decision-tree modeling with random forests algorithm</td>
<td>0.62</td>
<td>35.97</td>
<td>47.03</td>
<td>62.40</td>
</tr>
</tbody>
</table>

Notes: MAE: mean absolute error; RMSE: root mean square error; NRMSE: the normalized root mean square error.

Figure 4. Histogram of forest AGB density based on the estimate of decision-tree-based modeling.

estimates from the spatial regression model (2.13 \(\times 10^9\) Mg) and RF (2.26 \(\times 10^9\) Mg) (Fig. 3).

Using the RF approach, the estimated total AGB for all forest regions across Alberta was 2.26 \(\times 10^9\) Mg (Table 3, Fig. 3). The average AGB density in each 1 × 1 km grid was 56.30 ± 35.94 Mg ha\(^{-1}\). Around 23% of total forest areas had AGB densities between 40–60 Mg ha\(^{-1}\), and around 14% of total forest areas had AGB densities larger than 100 Mg ha\(^{-1}\) (Fig. 4).

Total tree AGB in the boreal region (RF approach) was about 1.30 \(\times 10^9\) Mg, accounting for 57.67% of total tree AGB in Alberta forests among the four main natural regions of Alberta (Table 3). The estimated AGB was about 0.57 \(\times 10^9\) Mg in the Foothills, 0.37 \(\times 10^9\) Mg in the Rocky Mountain, and 0.02 \(\times 10^9\) Mg in the Canadian Shield. Among the 14 natural subregions (Table 3), Central Mixedwood had the highest total tree AGB (0.66 \(\times 10^9\) Mg), followed by Lower Foothills, Subalpine and Lower Boreal Highlands.

The average AGB density of inventory plots across all regions was 128.24 Mg ha\(^{-1}\) (Table 3). The Foothills and Rocky Mountain natural regions had higher AGB densities of 143.35 and 141.75 Mg ha\(^{-1}\), respectively, than the others. Average AGB densities showed even greater variations

among subregions, from 13.18 Mg ha\(^{-1}\) in Boreal Subarctic to 147.60 Mg ha\(^{-1}\) in Lower Foothills.

Among three major land cover types in Alberta forests (Supplement B), coniferous forests stored 1.14 × 10\(^9\) Mg AGB, accounting for 50% of total tree AGB in Alberta forests, while broadleaf forests and mixed forests stored 0.62 × 10\(^9\) and 0.17 × 10\(^9\) Mg AGB, respectively.

### 3.4 AGB estimates of major tree species

Three major tree species, lodgepole pine, trembling aspen and white spruce, stocked about 1.39 × 10\(^9\) Mg AGB in total, accounting for 62% of total AGB in Alberta forests (Fig. 5, Table 4). Total AGB of lodgepole pine was 0.55 × 10\(^9\) Mg, and 85% of which is distributed in the Foothills and Rocky Mountain regions. For trembling aspen, total AGB was 0.50 × 10\(^9\) Mg, of which 78% is distributed in the Boreal region. For white spruce, total AGB was 0.35 × 10\(^9\) Mg, of which 58% is distributed in the Boreal region.

### 3.5 Variable importance on AGB distribution

Using the RF, we also assessed the importance of various predictor variables on AGB distribution (Fig. 6). Canopy height, which was directly related to AGB, had a major influence on AGB distribution at both stand and species levels. Elevation was also significantly correlated with AGB distribution of lodgepole pine, but not for other two species. Each of the 10 climatic variables had relatively weak effects on AGB distribution at the stand level, although MSP had a relatively stronger influence than other climatic variables. The three major tree species showed differing relationships with climatic variables. For lodgepole pine, MWMT, MCMT, DD0 and DD5 had stronger impacts on AGB than the other climatic variables. For trembling aspen, DI and DD0 were a little more important than the others. For white spruce, DD0 and MCMT had slightly stronger impacts on AGB than others.

### 4 Discussion

We reported on a large-scale, spatially explicit data set for presenting biomass storage in Alberta’s forest regions, derived from a combination of forest inventory data from 1968 plots, spaceborne lidar data, land cover classification, climate and other environmental variables. Using decision-tree-based approach with random forests algorithm, total AGB stock in the study region was estimated to be 2.26 × 10\(^9\) Mg, which is very close to Bonnor’s (1985) estimate (2.31 × 10\(^9\) Mg) based on volume inventory data, but is smaller than Penner et al.’s (1997) estimate (3.14 × 10\(^9\) Mg) (Table 5). The average AGB density was 56.30 Mg ha\(^{-1}\), which is close to Bonnor’s (1985) estimate (57 Mg ha\(^{-1}\)). This study showed that
the combination of multisource data could be a cost-effective way to estimate the amounts, distributions and variations of biomass carbon stocks across large regions with reasonable accuracy.

4.1 Comparison with previous biomass estimations

We summarized previous studies on boreal forest AGB estimation at different spatial extents (Table 5). At the global scale, estimates of total AGB for boreal forests ranged from $81.85 \times 10^9$ Mg (Cao and Woodward, 1998) to $129.41 \times 10^9$ Mg (Dixon et al., 1994). For Canadian forests, total biomass estimates varied from $15.53 \times 10^9$ Mg (Myhneni et al., 2001) to $41.43 \times 10^9$ Mg (Penner et al., 1997). In Alberta forest regions, our estimate ($2.26 \times 10^9$ Mg) using a decision-tree approach was very similar to the estimate of Bonnor (1985), but smaller than the estimate of Penner et al. (1997) (Table 5). Compared with other studies, our estimate of mean AGB density in Alberta was similar to the estimates reported by several studies at global and regional scales.
Figure 6. Relative importance of predictor variables for AGB estimation by decision-tree-based modeling with random forest algorithm. Variable importance is measured in mean decrease in accuracy, which is the decrease in accuracy of a classification after the variable has been randomly permuted. A higher mean decrease in accuracy means the variable contributes more to the accuracy of the classification.

Table 5. AGB and total biomass estimations in previous studies.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Study area</th>
<th>Area (Mha)</th>
<th>Methodology or data source</th>
<th>Total tree AGB ($\times 10^9$ Mg)</th>
<th>Mean AGB density (Mg ha$^{-1}$)</th>
<th>Total tree biomass ($\times 10^9$ Mg)</th>
<th>Mean biomass density (Mg ha$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dixon et al. (1994)</td>
<td>Boreal forests (Global)</td>
<td>1372</td>
<td>Inventory data (1987–1990)</td>
<td>129.4$^b$</td>
<td>94.12$^{b}$</td>
<td>176</td>
<td>128</td>
</tr>
<tr>
<td>Cao and Woodward (1998)</td>
<td>Boreal forests (Global)</td>
<td>1210</td>
<td>Predicted from a global carbon model (1990s)</td>
<td>81.85$^{b}$</td>
<td>67.63$^{b}$</td>
<td>111.32</td>
<td>92</td>
</tr>
<tr>
<td>Jarvis et al. (2001)</td>
<td>Boreal forests (Global)</td>
<td>1381</td>
<td>Inventory data (1990s)</td>
<td>84.55</td>
<td>61</td>
<td>114.99$^a$</td>
<td>83$^a$</td>
</tr>
<tr>
<td>Myneni et al. (2001)</td>
<td>Northern forests (Global)</td>
<td>1419.9</td>
<td>Remote sensing (NDVI; 1995–1999)</td>
<td>89.29$^a$</td>
<td>63.10$^b$</td>
<td>121.44</td>
<td>85.82</td>
</tr>
<tr>
<td>Pan et al. (2011) and Stinson et al. (2011)</td>
<td>Boreal forests (Global)</td>
<td>1135</td>
<td>Inventory data and statistical or process models (2007)</td>
<td>102.94$^b$</td>
<td>90.70$^{b}$</td>
<td>140$^c$</td>
<td>123.35$^c$</td>
</tr>
<tr>
<td>Bonnor (1985)</td>
<td>Canadian forests</td>
<td>440.7</td>
<td>Volume Inventory data (1981)</td>
<td>26.09</td>
<td>59</td>
<td>34.42$^a$</td>
<td>85.68$^a$</td>
</tr>
<tr>
<td>Dixon et al. (1994)</td>
<td>Canadian forests</td>
<td>436</td>
<td>Inventory data (1987–1990)</td>
<td>17.65$^b$</td>
<td>41.18$^{b}$</td>
<td>24</td>
<td>56</td>
</tr>
<tr>
<td>Penner et al. (1997)</td>
<td>Canadian forests</td>
<td>440.7</td>
<td>Volume Inventory data (1991)</td>
<td>41.43</td>
<td>94</td>
<td>56.34$^b$</td>
<td>127.84$^a$</td>
</tr>
<tr>
<td>Kurz and Apps (1999)</td>
<td>Canadian forests</td>
<td>404.2</td>
<td>Inventory data (1990s)</td>
<td>21.33$^{b}$</td>
<td>52.79$^{b}$</td>
<td>29.02$^c$</td>
<td>71.8$^c$</td>
</tr>
<tr>
<td>Pan et al. (2011) and Stinson et al. (2011)</td>
<td>Canadian forests</td>
<td>229.4</td>
<td>Inventory data and statistical or process models (2007)</td>
<td>27.94$^b$</td>
<td>122.06$^{c}$</td>
<td>38$^{c}$</td>
<td>165.65$^{c}$</td>
</tr>
<tr>
<td>Myneni et al. (2001)</td>
<td>Canadian forests</td>
<td>239.5</td>
<td>Remote sensing (NDVI; 1995–1999)</td>
<td>15.53$^{b}$</td>
<td>64.84$^{b}$</td>
<td>21.12</td>
<td>88.18</td>
</tr>
<tr>
<td>Liski and Kauppi (2000)</td>
<td>Canadian forests</td>
<td>244.6</td>
<td>Inventory data (mid-1990s)</td>
<td>20.24</td>
<td>82.8</td>
<td>27.53$^a$</td>
<td>112.61$^b$</td>
</tr>
<tr>
<td>Beaudoin et al. (2014)</td>
<td>Canadian forests</td>
<td>403</td>
<td>Remote sensing and photo plots (2000s)</td>
<td>25.77</td>
<td>63.94</td>
<td>35.05$^{a}$</td>
<td>86.96$^{a}$</td>
</tr>
<tr>
<td>Penner et al. (1997)</td>
<td>Alberta forests</td>
<td>40.3</td>
<td>Volume Inventory data (1991)</td>
<td>3.14</td>
<td>78</td>
<td>4.28$^{a}$</td>
<td>106.08$^{a}$</td>
</tr>
<tr>
<td>Bonnor (1985)</td>
<td>Alberta forests</td>
<td>40.3</td>
<td>Volume Inventory data (1981)</td>
<td>2.31</td>
<td>57</td>
<td>3.14</td>
<td>77.52</td>
</tr>
<tr>
<td>This study</td>
<td>Alberta forests</td>
<td>40.3</td>
<td>Inventory data (2000–2012) and lidar canopy height data (2006)</td>
<td>2.26</td>
<td>56.30</td>
<td>3.19</td>
<td>79.56</td>
</tr>
</tbody>
</table>

For the studies with aboveground biomass data only, belowground biomass is assumed to be 0.36 of the aboveground biomass (Jarvis et al., 2001).

For the studies with total biomass data only, aboveground biomass is assumed to be $0.74 (=1/(1+0.36))$ of the total biomass (Jarvis et al., 2001).

For the studies with carbon storage only, biomass is assumed to be two times that of carbon storage (Schlesinger, 1997).

scales, but was smaller than the estimates of some studies, such as Dixon et al. (1994), Pan et al. (2011) and Penner et al. (1997) (Table 5). Clearly, there is a huge disagreement among different estimates, but it is hard to compare them because of differences among data sources, estimation methodologies, and time periods of data collection. Another source of variation among studies is that there are major differences among the reported areas (Table 5) included under the categories of boreal (northern) forests and Canadian forests (some of which exclude more northerly, unmanaged forests) (Brandt, 2009).

Compared with these previous studies, our current study has at least two improvements and advantages: (1) multisource data: we combined the data from ground-based...
inventory, lidar, land cover, climate and other environmental variables, and provided a cost-effective scheme of mapping biomass stock for provincial- and national-scale assessments. Many previous studies used only a single data source, and did not consider the role of climate and other variables in their analyses. (2) The inclusion of spatially representative forest inventory plots: the lack of sufficient and unbiased sample plots has been identified as a major barrier to accurate estimation of biomass stocks in large areas (Botkin and Simpson, 1990; Brown, 1997; Wulder et al., 2008). In the present study, the two different sources of plot data showed significant differences in stand age structure and biomass distribution (Fig. 2). The PSP data were derived from undisturbed, relatively productive stands and thus gave much greater average values of biomass density than the ABMI plots, which include both disturbed and undisturbed sites. Further, the regular distribution of ABMI plots places some of them in peatlands, which generally were avoided in the PSP inventory. Thus, the use of PSP data alone would lead to the overestimation of biomass. In terms of the scope and sample sizes, the data used in this study are more comprehensive and extensive than previous data sets.

4.2 Comparison of different methods for biomass estimations

Selection of appropriate models plays a central role in estimating biomass and carbon stocks (Fang et al., 1998; Saatchi et al., 2011). Four different approaches, including spatial interpolation, non-spatial and spatial regression models, and decision-tree-based modeling with random forests algorithm (RF), were used to yield estimates of total AGB in our study area. We found that spatial interpolation greatly overestimated total AGB, while regression models and RF provided similar estimate with high accuracy. The overestimation by spatial interpolation might be related to the characteristics of the approach itself and the data we used.

First, the spatial interpolation approach assumes that spatial distribution of the variable we try to predict is a spatially continuous surface, and the near points generally receive higher weights than far away points. This approach is appropriate for the interpolation of some climate and topography variables, but for biomass and carbon, major errors may arise from discontinuities in the spatial distribution of biomass induced by disturbances and land uses such as agriculture (Supplement B).

Second, the spatial interpolation approach we used only considered one additional variable, which seriously constrains the ability to accurately predict. Although some techniques have been developed to consider multiple variables into spatial interpolation, they are still not available in most widely used geostatistics software. Furthermore, for most of the PSP plots placed on upland sites, these are intermixed with a fine-scale mosaic of forested peatlands with much lower biomass.

As a nonparametric approach, RF has shown some outstanding advantages in our study. This is also supported by previous studies for soil mapping (e.g., Grimm et al., 2008), biomass mapping in forests (Baccini et al., 2004; Neumann et al., 2011; Asner et al., 2013) and seafloor (Wei et al., 2010), and bird distribution modeling (Kreakie et al., 2012). The advantages of random forests include the ability to model high dimensional non-linear relationships, handling of categorical and continuous predictors, resistance to overfitting, relative robustness with respect to noise features, unbiased measure of error rate, and measures of variable importance (Breiman, 2001; Grimm et al., 2008). Therefore, by combining different predictor variables, this approach has a great potential for improving the estimation of forest biomass at regional and global scales.

4.3 Canopy height as an important determinant of biomass distribution

It is well known that canopy height is a critical indicator of forest site quality and growth potential (Kimmins, 2004; Fang et al., 1998). Also, canopy height is highly related to stand age and forest disturbance, both of which directly affect forest biomass and productivity. Using a large sample of forest inventory data, we detected a significant relationship between biomass and canopy height (Table 1, Fig. 2). The assessment of variable importance using the RF approach also showed that canopy height was the most important variable for determining biomass distribution in our study area (Fig. 6). However, canopy height has rarely been used in previous estimations of regional-scale biomass and carbon storage, because this information was not available over large areas in the past. The development of remote sensing techniques, especially lidar, has provided high- or medium-resolution canopy height products at both regional and global scales (Lefsky et al., 2010; Simard et al., 2011), and provides an opportunity to obtain more accurate estimates of biomass and carbon storage over large areas. For example, based on 1 km resolution spaceborne lidar canopy height data (Lefsky et al., 2010) and ground inventory data, Saatchi et al. (2011) mapped the total biomass carbon stocks in tropical regions across three continents with a forest area of 2.5 billion ha. Therefore, the integration of plot-based measurements of biomass with remotely sensed observations of canopy height can provide a cost-effective method for large-scale mapping. In addition, the lidar canopy height data are closely related to logging and fire history, allowing recently logged and burned sites to be more accurately accounted for in biomass carbon estimation.

The current study and that of Saatchi et al. (2011) in tropical forests have demonstrated the benefits of using spaceborne lidar canopy height data for biomass mapping. However, the coarse spatial resolution of spaceborne lidar data may pose problems for fine-scale biomass mapping. Recently, Bolton et al. (2013) investigated the agreement
between spaceborne lidar canopy height data (1 km resolution) and airborne lidar data (25 m resolution) in Canada’s boreal forests, and found that airborne-lidar-derived canopy heights were generally in good agreement with spaceborne lidar canopy height data we used in the current study. In the Boreal Plains ecozone, in which our study area is located, the RMSE (root mean square error) between spaceborne and airborne heights was 4.39 m (Bolton et al., 2013). Nevertheless, further improvements in accuracy of biomass estimation and mapping may be expected from the use of higher-resolution lidar data coupled with further advances in data processing techniques.

4.4 Biomass–climate relationships

Understanding biomass–climate relationships is important for biomass and carbon mapping under past and present conditions as well as for making future projections under a changing climate. Although climatic variables have been used in biomass estimations, we know relatively little about how climate influences variation in biomass stocks (Stegen et al., 2011). In this study, we found that climate explained relatively little of the observed, stand-level variation in Alberta forest biomass (Table 1, Fig. 6), which is consistent with Stegen et al.’s (2011) findings on biomass–climate relationships in temperate and tropical forests. Disturbance regime is likely a better predictor of biomass but these are often difficult to map at regional scales. Because canopy height is strongly influenced by the time since the last stand-replacing disturbance (e.g., fire), high-resolution lidar data can play an important role in estimating biomass and productivity at regional and national scales.

Species-level analysis on biomass–climate relationships showed that tree species respond differently to how climate affects biomass distribution (Fig. 6). For lodgepole pine, the mean warmest month temperature (MCMT), mean coldest month temperature (MWMT) and chilling degree days (DD0) played a more important role than other climatic variables. This strong correlation with degree days is also supported by previous studies on lodgepole pine site index study in Alberta forests (Monsrud et al., 2006). For trembling aspen, drought-related variable (dryness index, DI) were slightly more important than other climatic variables, which confirms previous studies about drought-related impacts on aspen stand dynamics (e.g., Hogg et al., 2008; Michaelian et al., 2011).

4.5 Total carbon stocks in Alberta forests

The present study reports on tree biomass in the forests of Alberta. However, the estimation and mapping of total carbon (C) storage also requires high quality data on soil C. Boreal forest ecosystems contain vast C stocks in soil, most of which is found in peatlands and permafrost soils (DeLuca and Boisvenue, 2012). Soil C in boreal ecosystems has been reported to account for about five times the total C in the standing biomass or about 85% of the total biome C (Malhi et al., 1999). The large-scale estimation of soil C stocks poses many challenges (Liu et al., 2013), and was thus not specifically included in the current study. However, based on the recent data set of North American soil organic carbon content at 0.25 degree resolution (Liu et al., 2013), total soil carbon stocks in Alberta’s forests are approximately $11.8 \times 10^9$ Mg, with a high proportion in peatlands (Vitt et al., 2000). Thus, our estimate of total tree biomass carbon ($1.59 \times 10^9$ Mg, 50% of total tree biomass, Supplement C) only accounted for 12% of estimated total carbon stocks ($13.39 \times 10^9$ Mg), while soil carbon accounted for 88%. Clearly, more efforts are needed to better understand spatial and temporal variation of biomass and soil carbon stocks in the boreal forest.

The Supplement related to this article is available online at doi:10.5194/bg-11-2793-2014-supplement.

Acknowledgements. We thank Alberta Ministry of Environment and Sustainable Resource Development (ESRD), Alberta Biodiversity Monitoring Institute (ABMI), Weyerhaeuser Canada, and West Fraser for graciously granting us access to their data sets. We also thank Prof. Paul Stoy, two anonymous reviewers, Xianli Wang, Frédéric Raulier, and Scott Nielsen for their constructive comments, which helped improve the manuscript. This work was supported by Alberta Innovates – BioSolutions, ESRD, ABMI, the Natural Sciences and Engineering Research Council of Canada (NSERC), and the National Natural Science Foundation of China (No. 41101342).

Edited by: P. Stoy

References


Penner, M., Power, K., Muhairwe, C., Tellier, R., and Wang, Y.: Canada’s forest biomass resources: deriving estimates from Canada’s forest inventory, Pacific Forestry Centre Victoria, Canada, 1997.


Stegen, J. C., Swenson, N. G., Enquist, B. J., White, E. P., Phillips, O. L., Jorgensen, P. M., Weiser, M. D., Mendoza, A. M.,


