Scaling plot to stand-level lidar to province in a hierarchical approach to map forest biomass in Nova Scotia

Chris Hopkinson, David Colville, Danik Bourdeau, Suzanne Monette & Robert Maher

Applied Geomatics Research Group,
Centre of Geographic Sciences,
Lawrencetown, Nova Scotia,
B0S 1M0, Canada
Chris.hopkinson@nscc.ca

Abstract

This paper presents a study that used lidar transect, plot and wide area polygon sample data collected across Nova Scotia, Canada from 2005 to 2010 to calibrate and extrapolate above ground forest biomass from permanent sample plots (PSPs) to forest stand polygons to the entire Province. The whole tree dry biomass estimate for the total forest resource inventory (FRI) database in Nova Scotia is ~ 373 x 10^6 tonnes ±39%. Where lidar coverage exists, biomass is modelled at the 25 m grid cell resolution, which is a great improvement over the previous ecoregion level estimates, allowing for more effective operational stand management. Given the large spatio-temporal domain of the data sources, one of the major challenges faced in this study was temporal latency between coincident field, lidar and GIS data inputs, which was a significant contributor to the overall level of uncertainty in the result.

1. Introduction

The amount and range of biomass stored within a forested stand is an indicator of its status and ecosystem functioning (Brown, 2002). In Atlantic Canada, revenues from sawlog and pulp wood forestry products, critically important to the rural economy, have been in a steady decline in recent years (APEC, 2008). At the same time, public energy utilities have been rising to the dual challenge of meeting growing energy demands while attempting to reduce greenhouse gas emissions. The Province of Nova Scotia, for example, has committed to 25% renewable energy supply by 2015 and 40% by 2020, and biomass is seen as a potential viable source of long-term carbon-neutral alternative energy to supplement more traditional sources (NSDE, 2010). For all of these reasons, the ability to map forest biomass in Nova Scotia at a scale appropriate for land management has economic, ecological, environmental value.

2. Data sources

The Province of Nova Scotia is a little over 50,000 km^2 and of this area >80% is forested. The forests of Nova Scotia are catalogued and monitored by the NS Department of Natural Resources (DNR). There are two publicly available and spatially explicit datasets that describe these resources and have been used as the basis for modelling in this project: a PSP database and a FRI GIS database. The PSP database details the attributes of all trees with a stem diameter at breast height (DBH) > 9.1 cm within 3250 plots of 11.3 m radius covering the whole Province. The PSPs are randomly established throughout the forests of Nova Scotia and cover a 400 m^2 circular area. About half of the 3250 plots were established from 1965 to 1970, while the rest were established between 1998 and 2002. About 650 plots are revisited every year to ensure a five year rotation for each plot. Within the plot, living and dead trees are numbered and several attributes are recorded for each tree including height, DBH, species, signs of disease, cause of death etc.
The FRI database describing the total forest coverage within Nova Scotia contains approximately 1.1 million stand polygons that are delineated from aerial photographs and intended to describe regions of stand similarity within contiguous parcels of land. The FRI database is primarily updated from air photos collected on a ten year rotation. Using paired photographs, interpreters can see in three dimensions and then delineate homogeneous stands of trees to interpret crown closure, stand height, species and land capability. Satellite imagery is used in between aerial photo years to update the FRI for noticeable change, such as clear-cutting. Both PSPs and FRI stands are being continuously updated on a revolving basis as opposed to updating the whole Province at one time.

Figure 1: Top - AGRG Lidar survey polygons and sampling transects from 2000 to 2010 within Nova Scotia. Bottom – spatially coincident permanent sample plots that are within two years of lidar survey.

While the DNR GIS databases contain extensive stand-level inventory data covering the Province, the conversion of these data into meaningful estimates of available and sustainable
biomass energy requires calibration (Townsend, 2008). At the local scale, PSP data collected in the field allow reasonably accurate calculation of biomass over small areas (Lambert et al. 2005), which have been used to provide coarse estimates of biomass for Nova Scotia down to the ecoregion scale (Townsend, 2008). However, to derive more spatially explicit estimates of biomass at the typical management unit or stand-scale is challenged by the heterogeneity displayed by Acadian mixed wood forests.

To scale between PSP and FRI data layers to develop a spatial model of biomass representing both spatial domains requires a data source that can sample the canopy structure within a plot and allow for effective aggregation at the stand scale. Airborne lidar data have been shown time and again to be ideally suited to the task of plot- and stand-level canopy structure and biomass modelling (Lim et al. 2003). Lidar data have been collected across Nova Scotia by the Applied Geomatics Research Group since 2000 (Figure 1). Several polygons and sample transects covering ~ 10,000 km² or ~20% of the area of Nova Scotia have been mapped using an Airborne Laser terrain mapper (ALTM) 3100C (Optech Inc. Toronto, Canada). Only ~ 50% of the data were suitable for this study, as a threshold of ~ 1 point/m² was applied to ensure a high density of data for subsequent model generation. Of the 3250 provincial PSPs, 281 were spatially coincident with lidar cover, and of these 99 were culled after applying a 2 year temporal buffer. The data sources and associated modelling approaches are described in Table 1.

Table 1: Data sources and attributes used, domains of spatial representation and notes on how the data were used in the Provincial biomass modeling approach

<table>
<thead>
<tr>
<th>Raw data</th>
<th>Spatial model scale</th>
<th>Data attributes</th>
<th>Modeling approach</th>
<th>Modeling purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSP attribute db [258 / 3250] 8%</td>
<td>Plot [400 m²] (1.3 km²)</td>
<td>Height, DBH, species, stem count</td>
<td>Based on Lambert et al. (2005). Species divided into hardwood / softwood</td>
<td>Generate biomass ‘ground truth’ at plot scale</td>
</tr>
<tr>
<td>Lidar point cloud [1000km² / 50,000km²] &gt;2%</td>
<td>Lidar survey coverage [-1 m point sampling to 500 km² polygon] (~ 10,000 km²)</td>
<td>Height percentiles &amp; vertical distribution ratios</td>
<td>Linear, quadratic and JGLS regression models with &lt; 2 variables to predict bole / whole tree biomass</td>
<td>Calibrate FRI stand data by extrapolating PSP-based lidar model</td>
</tr>
<tr>
<td>GIS stand polygon [2639 / 1.1 million] 2.4%</td>
<td>FRI stand [-0.01 km² to 10 km²] (~ 42,000 km²)</td>
<td>Mean canopy height &amp; closure</td>
<td>Linear, quadratic and JGLS regression models of bole / whole tree biomass</td>
<td>Simulate stand level biomass and aggregate up to Province</td>
</tr>
</tbody>
</table>

3. Summary of methods

After initial quality control of the coincident lidar and PSP data, there were 182 PSPs between the years 2005 and 2010 available to train and test a lidar-based model of biomass. PSP data were used to derive ground truth estimates of bole and whole tree dry biomass through the application of a robust individual tree biomass model that was constructed from plot-level sample data collected across Canada (Lambert et al. 2005). The lidar biomass data were then used to train an FRI-based model using attributes from 1873 stand polygons which could be applied to the entire Province. Lidar data metrics collected over 13 different survey missions using the same ALTM 3100C sensor were extracted for each of the PSPs. Summary statistics extracted using FUSION (McGaughey, 2010) describing the vertical within-plot lidar frequency distributions and point cloud ratios were used to describe canopy height and cover attributes.
These lidar ‘metrics’ were then correlated with the associated PSP biomass estimates to construct predictive models of biomass. The lidar-based maps of biomass were segmented into corresponding FRI polygons, which were then scaled up to the Province using a new model based on the FRI stand attributes of canopy height and closure.

4. Results & discussion

4.1 Plot-level biomass

Statistical descriptions of each PSP lidar point cloud dataset generated in the FUSION software (McGaughey, 2010) were tested for inter-correlation and suitability for use in multivariate biomass modelling (Table 2). As expected, all height-based frequency distribution metrics demonstrated high inter-correlation, as did most ratio-based metrics. While other derivatives of the frequency distribution are possible, it was decided to keep the PSP biomass modelling approach simple to allow for maximum transferability across diverse lidar datasets. Consequently, models tests were limited to two variables; one height-based and one ratio-based metric, as these demonstrated the least inter-correlation. Furthermore, height metrics are an index of canopy height (e.g. Naesset, 1997) while ratio metrics are an index of canopy cover (e.g. Hopkinson and Chasmer, 2008). These two attributes are logical indices of the two physical dimensions (height and width) that are fundamental to volume, and therefore, biomass calculations.

Table 2: Correlation matrix of selected PSP lidar point cloud frequency distribution attributes extracted from FUSION. Shaded cells denote correlations between height and ratio metrics. Bold values illustrate weakest inter-correlation and therefore suitability for multivariate modeling.

<table>
<thead>
<tr>
<th></th>
<th>ElevMean</th>
<th>ElevStdDev</th>
<th>ElevP70</th>
<th>ElevP75</th>
<th>ElevP90</th>
<th>ElevP95</th>
<th>ElevP99</th>
<th>Perc1stReturns&gt;Mean</th>
<th>PercAllReturns&gt;Mean</th>
<th>AllReturns&gt;Mean/Total1stReturns*100</th>
<th>Perc1stReturns&gt;1.50</th>
<th>PercAllReturns&gt;1.50</th>
<th>AllReturns&gt;1.50/Total1stReturns*100</th>
</tr>
</thead>
<tbody>
<tr>
<td>ElevMean</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ElevStdDev</td>
<td>0.81</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ElevP70</td>
<td>0.97</td>
<td>0.89</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ElevP75</td>
<td>0.97</td>
<td>0.91</td>
<td>0.97</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ElevP90</td>
<td>0.93</td>
<td>0.95</td>
<td>0.95</td>
<td>0.97</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ElevP95</td>
<td>0.91</td>
<td>0.95</td>
<td>0.93</td>
<td>0.95</td>
<td>0.99</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ElevP99</td>
<td>0.86</td>
<td>0.93</td>
<td>0.87</td>
<td>0.90</td>
<td>0.96</td>
<td>0.98</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perc1stReturns&gt;Mean</td>
<td>0.76</td>
<td>0.58</td>
<td>0.78</td>
<td>0.75</td>
<td>0.62</td>
<td>0.58</td>
<td>0.51</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PercAllReturns&gt;Mean</td>
<td>0.72</td>
<td>0.45</td>
<td>0.71</td>
<td>0.67</td>
<td>0.53</td>
<td>0.49</td>
<td>0.41</td>
<td>0.93</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AllReturns&gt;Mean/Total1stReturns*100</td>
<td>0.80</td>
<td>0.72</td>
<td>0.85</td>
<td>0.83</td>
<td>0.74</td>
<td>0.70</td>
<td>0.64</td>
<td>0.97</td>
<td>0.86</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perc1stReturns&gt;1.50</td>
<td>0.71</td>
<td>0.46</td>
<td>0.69</td>
<td>0.67</td>
<td>0.58</td>
<td>0.56</td>
<td>0.49</td>
<td>0.83</td>
<td>0.84</td>
<td>0.79</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PercAllReturns&gt;1.50</td>
<td>0.70</td>
<td>0.34</td>
<td>0.63</td>
<td>0.61</td>
<td>0.53</td>
<td>0.51</td>
<td>0.45</td>
<td>0.64</td>
<td>0.73</td>
<td>0.59</td>
<td>0.91</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>AllReturns&gt;1.50/Total1stReturns*100</td>
<td>0.86</td>
<td>0.66</td>
<td>0.85</td>
<td>0.84</td>
<td>0.78</td>
<td>0.76</td>
<td>0.70</td>
<td>0.82</td>
<td>0.75</td>
<td>0.86</td>
<td>0.91</td>
<td>0.85</td>
<td>1</td>
</tr>
</tbody>
</table>

Using PSP data and the biomass model of Lambert *et al.* (2005), several lidar biomass models were trained and tested for whole tree and bole. For the lidar model, species information was ignored but the ratio of softwood to hardwood stems was considered in PSP model training. Root mean square error (RMSE) for the best polynomial regression whole tree and bole lidar model remained approximately the same in both test datasets at ~ 26%. The explanation of variance in the test data was greater for bole biomass estimates at 75%, than whole tree biomass at 63%. All models were significant at the 99% level of confidence. These results indicated that...
the lidar models were robust enough to provide stand-level summaries.

4.2 Stand-level biomass

A challenge that became immediately apparent during the process of relating lidar biomass estimates to FRI stand-level attributes, was the temporal latency between the two datasets (FRI data dating back to 1990s in extreme cases, while lidar data ranging from 2005 to 2010). This latency was most evident when comparing stand-level FRI mean tree height with the mean maximum of the lidar data aggregated into 25 m grid cells (Figure 2). Using canopy height as an indicator, quality control procedures were put in place to systematically remove the most obvious outliers (due to growth and clearcuts) using objective height and date criteria. However, even after this quality control process, the latency between lidar and FRI still has the potential to propagate uncertainty into the model. The nature of this error is such that any stand growth, decay, thinning or clear cut occurring following the last FRI stand update and preceding the associated lidar acquisition will lead to divergence between the lidar and FRI attributes. As long as the forests are in a state of dynamic equilibrium (i.e. the Provincial forest resource as a whole is neither expanding nor contracting), then these stand-level biases will not necessarily lead to a systematic bias in the overall population statistics. In practical terms, this means that we expect the model to display a high level of variance at the stand scale but when aggregating biomass estimates to larger and larger spatial domains, there should be a level of compensation between over- and under-estimates.

Given the accepted level of uncertainty in the models at this stage of the analysis, no attempt was made to develop highly sophisticated or complex regression models between stand-level lidar biomass and FRI stand attributes. Crown closure and mean canopy height were chosen as the FRI attributes to be used in the stand biomass models as they most closely resembled the lidar metrics used in the previous modelling step. Similar to the PSP results, the RMSE in stand biomass approximated 27% both for whole tree and bole. However, the explanation of variance dropped to 41% and 43%, respectively, most likely a large function of the temporal latency issues described above.

4.3 Nova Scotia’s biomass

Using the hierarchical model development approach summarised above, we derived six estimates of total provincial biomass; three for bole (stem wood) and three for whole tree. The
three modelling approaches did not differ in terms of the data used at each stage of model development, rather the differences are simply in terms of the algorithm construction; ranging from simple single variable linear regression to dual variable quadratic and a further dual variable model that mimicked the structure of the model proposed by Lambert et al. (2005). Given the Lambert et al. (2005) model was used to derive the ‘ground truth’ plot-level dry biomass estimates from which the rest of the lidar and stand-level predictions are based, the model results are to be considered more reliable if expressed as dry biomass. The range of values for total bole biomass within the Province ranged from 253 – 260 x 10^6 dry tonnes, while whole tree biomass ranged from 365 – 373 x 10^6 dry tonnes (Figure 3). Bole biomass is the number to refer to if only stems are to be used as fuel wood and tree tops are to be left in situ for nutrient recycling. For the sake of comparison, the Nova Scotia Department of Natural Resources (DNR) estimate of total living merchantable tree stem biomass in the Province is 309 x 10^6 dry tonnes (Townsend, 2008). The Canadian Forest Service (CFS) has also developed a largely satellite image-based estimate of total above ground biomass at the 1km pixel resolution (Hall et al. 2010), and when this is aggregated to the Province scale a value of 362 x 10^6 dry tonnes results.

![Figure 3: Map of predicted forest stand-level biomass across Nova Scotia](image)

Both the DNR and the CFS estimates of total above ground biomass for the Province of Nova Scotia are lower (by 17% and 3%, respectively) than that generated using the lidar scaling approach described here. The DNR biomass value refers to living whole tree dry biomass but it should be noted that this only considers stems with DBH > 9.1 cm. The whole tree biomass estimate generated in this study includes standing dead stems and given lidar cannot differentiate stems based on DBH it is likely the estimate is further inflated relative to the DNR estimate by inclusion of small and immature tree stems. The primary difference between this and the DNR and CFS approaches is that the approach described here allows calculation of
biomass at the stand scale and is useful for operational planning and decision making. Based on a conversion ratio of 140% for dry to green biomass for typical Acadian mixed wood species (e.g. Shelton and Shapiro, 1976), the estimates above provide values of around 357 x 10^6 green tonnes for total bole biomass and 514 x 10^6 green tonnes for whole tree biomass.

Given the approach described uses three modelling steps, each building on the previous, uncertainty will propagate throughout. The RMSE values observed in the model results at each stage, demonstrated errors in the 15% to 30% range. Propagating the RMSE at the PSP, lidar and FRI modelling steps in quadrature compounds to an overall stand- and Province-level error of ~ 39%. This assumes that all errors are random and there is no significant bias.

5. Conclusions

For the time period from 2005 and 2010, 182 PSPs were used to train and test a lidar forest biomass model. This model was then used to train a FRI model from 1869 coincident stands, which was aggregated across all 1.1 million stands in Nova Scotia to arrive at a total above ground forest biomass estimate for the province. This biomass estimate can be expressed several ways but the whole tree dry biomass estimate is ~ 373 x 10^6 tonnes ±39%. Where lidar data are available in the Province (about 20% of the land surface area) a spatially explicit estimate of biomass can be generated at the 25 m grid cell resolution. In other areas, biomass can be estimated at the stand-level. The spatial resolution of these estimates constitutes an improvement over previous biomass estimates that were available at the ecoregion (DNR; Townsend, 2008) or 1 km pixel (CFS; Hall et al, 2010) resolutions. These results, therefore, can be used to aid in either stand- or within stand-level forest management practices and in informing forest biomass energy policy in Nova Scotia.

Modeling biomass over such a large area is not without challenges. Greatest of these is obtaining useable model calibration and validation data. In this study, DNR PSP and FRI data were all that were available at the scale required. Both data sources were limited in terms of temporal compatibility with the lidar data that were used to scale between the two. Up to two years of latency in the PSP data is less than ideal given forests grow, die and are managed. However, this was less problematic than the > 10 years of latency for some of the FRI stands. The time discrepancy will introduce larger errors for younger stands and for those that have been clear cut. While objective criteria were used to mitigate such occurrences it is impossible to remove all such instances without manual selection and verification of each stand. Such an approach is not practical at this scale so a substantial amount of model uncertainty remains. However, it is assumed that temporal discrepancies will cause both over- and under-estimation of stand-level biomass, such that there will be a level of compensation.

Acknowledgements

Many thanks to Allyson Fox and Laura Chasmer for their dedication to flying long airborne survey missions and assisting with data processing. Thanks are extended to Dr. Ron Hall of the Canadian Forest Service for providing a 1 km resolution map of total biomass for Nova Scotia. Nova Scotia Power Inc. is acknowledged for funding this project and the Canada Foundation of Innovation for funding AGRG’s lidar laboratory. Nova Scotia Department of Natural Resources provided the raw PSP and FRI GIS data layers, while Cory Isenor, Collin Gillis and Heather Morrison assisted with field data collection. Dr Bob McGaughey of the US Department of Agriculture is gratefully acknowledged for his support with FUSION.
References


