Biomass mapping of Canadian northern boreal forests using a k-NN approach and sample plots from high resolution QuickBird images

A. Beaudoin[1], L. Guindon[1], A. Leboeuf[2], J.E. Luther[3],
C.H. Ung[1], S. Côté[1] & M.-C. Lambert[1]

[1] Canadian Forest Service - Laurentian Forestry Centre
e-mail: abeaudoin@cfl.forestry.ca
[2] Département de Géomatique Appliquée, Université de Sherbrooke, Canada
3 Canadian Forest Service - Atlantic Forestry Centre

ABSTRACT

The limited availability of ground sample plots (GSP) in huge and remote regions requires mapping methods with greater dependence on modeling and remote sensing. We developed and tested a method to map the biomass of black spruce (Picea mariana) stands of Canadian subarctic forest using a k-NN approach applied to Landsat Thematic Mapper imagery with sample plots generated from high resolution QuickBird images. Biomass was mapped locally using three QuickBird panchromatic images and a global regression model of above ground biomass as a function of shadow fraction (SF) of the images (reported in [1]). Herein, the QuickBird-derived biomass maps provided surrogates to traditional GSP for mapping biomass at the regional scale. The k-NN method was applied to the full extents of three Landsat images, representing three test sites. RMSE and bias were calculated using (i) GSP to estimate the combined error of scaling from the plot to the regional level and (ii) SSP to estimate the scaling error from application of the local biomass maps to the regional level. The RMSE of the regional scale biomass maps ranged from 10.1 to 19.6 t/ha with an overall RMSE of 17.2 t/ha based on the GSP. Bias estimates were only slightly positive with an overall bias of 2.8 t/ha for the three test sites. Application of the k-NN method using SSP produced good estimates of biomass over the three test sites with very low biases and relative errors in the order of 20-30% depending on the test-site. Further developments will consider extension of the method across large areas of Canadian subarctic forests. [1]

Key words: remote sensing, subarctic forests, biomass, Landsat, k-NN, shadow fraction.

1 INTRODUCTION

Forest biomass maps are needed in non-inventoried Canadian subarctic regions to address crucial northern issues related to carbon balance in a changing climate, and increasing exploitation pressure at the northern margin of the merchantable forests. Unfortunately, the number of ground sample plots (GSP) in northern Canada is very small due to the remoteness and the lack of infrastructure in these vast tracks of lands representing about 60% of Canada landmass and totalizing 4M km^2. These areas are mainly covered by open stands of black spruce (Picea mariana) with few other species, established over gently rolling or flat terrain. This context promotes the use of satellite remote sensing and modelling to estimate forest biomass, defined here as the total dry weight of above-ground trees per unit area (t/ha).

In this paper we report on the testing and implementation of the K-Nearest Neighbours (k-NN) method at a regional scale for mapping the biomass of northern black spruce (Picea mariana) stands. k-NN method is a non-parametric method that has been used extensively in the Scandinavian boreal forest context ([2];[3]) and may be applicable for some boreal regions in Canada. However, to provide good results, k-NN as many other methods require ground sample plots (GSP) representing the full range of forest conditions found within the mapping extent, which are lacking in our case. Therefore, we tested an alternative way of applying the k-NN method where Satellite Sample Plots (SSP) derived from QuickBird (QB) biomass maps [1] were used as surrogates to GSP. This study was conducted in the larger context of the Earth Observation for Sustainable Development of Forests (EOSD) project whose goal is to map the forest biomass of Canada using Landsat imagery ([4]; [5], submitted) and the Ecoleap project ([6]; [7]).

Paper No. 74, Proceedings of t the 26th Canadian Symposium on Remote Sensing, Wolfville, Nova Scotia, June 14-16, 2005

2 TEST SITES AND DATA

The study was carried on over 3 test sites located near three towns (Chibougamau: CH; Radisson: RA; Wabush: WA) in two subarctic ecozones in Quebec and Labrador, the eastern taiga shield and the boreal shield (Fig. 1). The test sites cover a wide biomass range and represent the dominant stand conditions of Northeastern boreal forests. Each test site is represented by a full Landsat 7 (L7) scene acquired between 1999 and 2003 under clear-sky conditions and during the growing season between mid-July and end of August, except for one L7 image acquired at the end of September. Each L7 scene was transformed into top-of-atmosphere reflectance and orthorectified with a resulting positional RMSE in the order of 15-20m.

SSP for each site were extracted from local scale biomass maps derived from corresponding QB high resolution images. The biomass maps were produced using a Northern Black spruce stand regression relationship established between biomass measured at each GSP plot and the shadow fraction extracted from each QB image [1]. The GSP dataset, used to develop and test the regression relationships between plot biomass and shadow fraction of the QB images, consisted of 108 circular GSP ($400m^2$) measured between 2002 and 2004. The plots were located within the 3 test sites and were dominated by black spruce (basal area > 50 %), the dominant forest type of the region. Biomass of each plot was estimated as the sum of individual tree above-ground biomass, which was estimated from DBH using allometric equation as in [8].

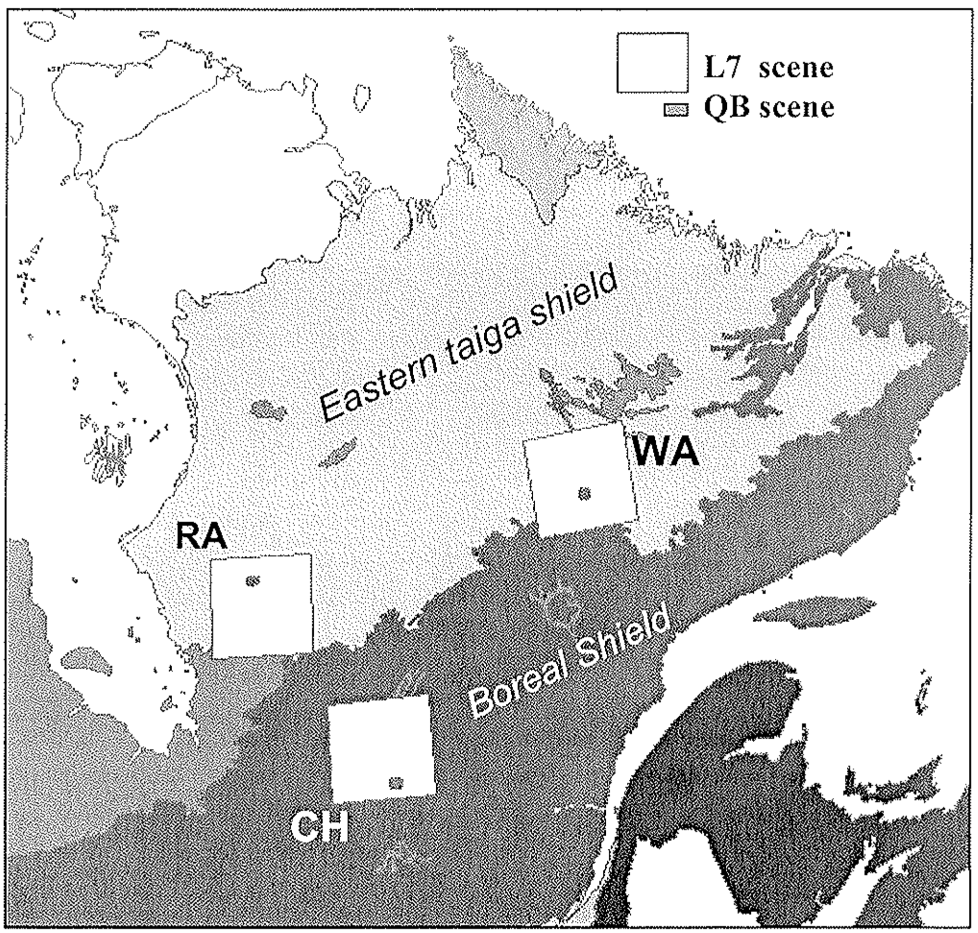

Figure 1. Location of the three test-sites in two northern ecozones, imaged by nested sets of QB and L7 images

3 METHOD

Biomass estimation with the k-NN method is given by a weighted average of biomass values from GSP which are spectrally the nearest from the considered pixel, as expressed by:

$$BIOknn_i = \sum_{k=1}^{K} W_k BIOgsp_k \Bigg/ \sum_{k=1}^{K} W_k \; for\, k \neq i \qquad where\; W_k = \frac{1}{d_k^j} \tag{1}$$

where $BIO_{k\text{-}NNi}$ is the k-NN biomass estimate for the i[th] Landsat pixel, k is the k[th] out of K nearest spectral neighbours, $BIOgsp_k$ is the GSP biomass, W_k is a weighting coefficient, d_k^j is the spectral Euclidian distance to which is applied a j power value, typically 0 to 2. In this application of the k-NN approach, we replaced the GSP in the above equation with SSP derived from high resolution QB images. SSP were derived by randomly sampling each local scale QB biomass map to obtain N SSP used as a training set (SSP_{cal}). We assumed that our local SSP_{cal} set selected in the forest stands found in the QB coverage was representative of the forest conditions encountered in the full Landsat scene. This assumption was tested by determining the representativeness of forest stands found in the QB coverage (50-100 km^2) relative to those of the L7 coverage (34,000 km^2) whereby land cover maps were derived from the L7 image using ECM classification methods [9]. The spatial occurrence (%) of five coniferous classes with variable crown closure (CC) classes (very sparse: 0-10%; sparse: 10-25%; open: 25-60%; dense: 60+%; treed bogs) for each QB coverage location was within 15% of the class occurrences within the related L7 coverage.

In order to optimize the parameterization of the k-NN equation, we tested combinations of N, K and j values within the following ranges: (i) K = 1 to 20, (ii) j = 0, 1 or 2, (iii) N =100, 300 and 600. Each set of values was tested with or without a coniferous forest mask to remove non-coniferous pixels. Further, each combination of input parameters was applied to all Landsat 30m pixels within the extent of the QB coverage, providing a biomass grid layer where each 30m cell had its biomass estimate given by the k-NN method, $BIO_{k\text{-}NN}$.

Quantitative error assessment was performed by comparing k-NN biomass with both plot-level GSP and SSP biomass estimates. Errors in the biomass estimates (RMSE and bias) were calculated using: (i) a 30% GSP_{val} set to estimate the combined error of scaling from the plot to the regional level, and (ii) the full SSP dataset to estimate the scaling error from application of the local biomass maps to the regional level. The combination of input parameters that minimized RMSE and bias were subsequently used to produce a final biomass map. $BIO_{k\text{-}NN}$ statistics (mean and standard deviation) were reported for five coniferous classes of different CC ranges.

4 RESULTS AND DISCUSSION

Concerning the optimal set of k-NN input parameters, it was found that (i) the number of SSP_{cal} was optimal around 600, (ii) there was no significant differences in RMSE and bias statistics for K>10 and (iii) the power of the weighting coefficient j had no significant impact. Moreover, the use of a coniferous stratification mask provided slightly poorer results than without using it. In the following, we report results using the following optimal parameter set: N=600, K=15, j=1 and no coniferous mask.

As a first qualitative assessment of the k-NN biomass estimates $BIO_{k\text{-}NN}$, mean and standard-deviation of mapped $BIO_{k\text{-}NN}$ values within each coniferous cover class, as seen in Fig. 2, exhibited the expected biomass quasi-linear increase with CC as reported in [10]. The 60-100 % CC class in Chibougamau showed a higher estimated biomass range compared to other sites because of denser stands more present in this southern site. Over-estimation was found in (i) some bogs or along lake shores but represent a small fraction of the test sites and (ii) on steep slopes opposite to the sun.

Fig. 3 shows $BIO_{k\text{-}NN}$ estimates compared to reference BIO_{gsp} values for each test site. Good agreement was found between $BIO_{k\text{-}NN}$ and BIO_{gsp} (R2=80%) with a slight under-estimation and increased variance at the higher end of the biomass range. RMSE and bias of the k-NN estimates determined using GSP and SSP are reported in Table 2. RMSE values determined using the GSP dataset were lowest for the WA site (10.1 t/ha) and highest for the CH site (18.9 t/ha) with an overall RMSE of 17.2 t/ha, which provided a relative error of about 20-30% depending on the site. Bias was less than 5 t/ha for all sites. Biomass estimation errors of the SSP (reported in [1]) were only slightly lower overall (15.3 t/ha), suggesting that the k-NN method did not bring significant additional errors in the regional estimates compared to the local ones. Overall bias estimates are slightly higher for the local compared to the regional estimates (4.2 and 2.8 t/ha, respectively).

RMSE and bias of the k-NN estimates based on the large set of SSP (all cells of the QB biomass map except those used to run the k-NN) report errors resulting from the local (QB) to regional scaling-up (L7). RMSE values are comparable to those of the local scale maps validated with the GSP (13.7 versus 15.3 from [1]). Bias tends toward 0 as the number of SSP samples increases and is consistent with observations reported in [2].

Table 2: Errors for $BIO_{k\text{-}NN}$ using GSP_{val} and SSP_{val} datasets

		Using GSP_{val}		*Using SSP_{val}*	
SITE	**Nb GSP_{val}**	**RMSE (t/ha)**	**Bias (t/ha)**	**RMSE (t/ha)**	**Bias (t/ha)**
CH	*9*	18.9	0.7	13.3	0.6
RA	*14*	19.6	4.7	9.1	0.1
WA	*9*	10.1	2.0	19.6	0.2
ALL	***32***	**17.2**	**2.8**	**13.75**	**0.31**

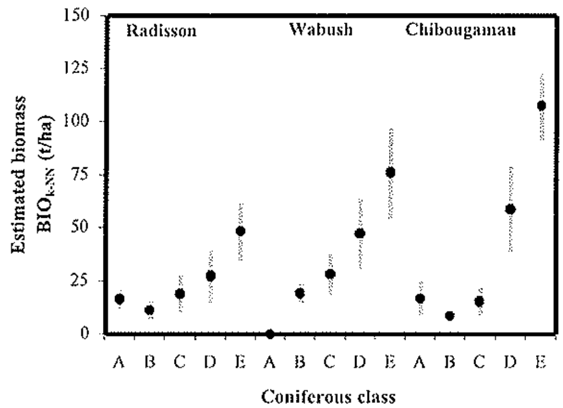

Figure 2 : $BIO_{k\text{-}NN}$ (mean ± 1 std) for each test site within each coniferous class of various CC ranges (A= treed bog,B=0 -10%,C=10 -25%,D=25 -60% and E=60-100% of crown closure).

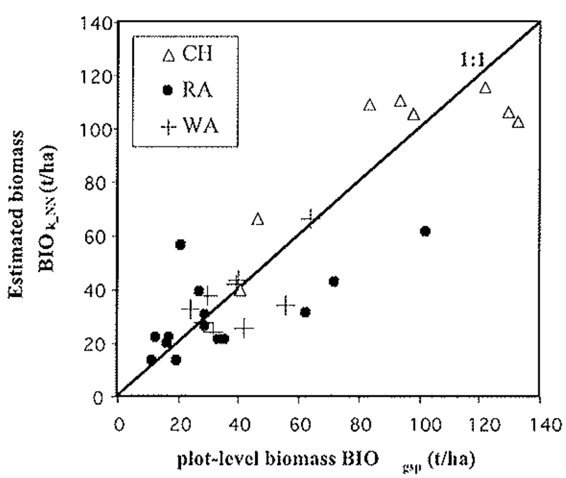

Figure 3: Estimated biomass $BIO_{k\text{-}NN}$ compared to GSP biomass BIO_{gsp}

Overall, application of the k-NN method using SSP produced good estimates of biomass over the three test sites with very low biases and relative errors of the order of 20-30% depending on test-site. An important consideration for the general application of the method is to ensure good representativeness of QB samples within the extent of the L7 scene. Further, co-registration errors between the QB and L7 images should be minimized as much as possible as the outlier $BIO_{k\text{-}NN}$ estimates were generally those for GSP located in sparse coniferous forests with highly variable understory (lichen, shrubs). This could be accomplished, in part, by excluding heterogeneous pixels from the SSP training set. In addition, terrain radiometric normalization may account for slope effects for areas of high relief. Finally, the acquisition lag between the QB and L7 acquisition must be taken into consideration as it may cause problems due to land cover changes (fires, logging, insect).

5 CONCLUSION

This paper presents an adaptation of the k-NN method for regional scale mapping of biomass in Canadian subarctic forests. Application of QB-based SSPs to implement the k-NN method on L7 images resulted in biomass estimates with very low biases and relative errors of the order of 20-30% depending on test-site. Overall RMSE and bias estimates were 17.2 and 2.8 t/ha respectively. Similar results were observed over three test sites representing a wide range of forest conditions. The method thus shows promise for expansion across northern Canada. Implementation would require a network of representative QB image samples scattered within a normalized mosaic of Landsat images for all the subarctic forest. Further work is required to test the method for other northern species (e.g. balsam fir, jack pine…)

ACKNOWLEDGEMENT

Support for this study was provided by the Earth Observation for Sustainable Development Project of the Canadian Space Agency's Government Related Initiatives Program and CFS Ecoleap Project. Sandra Labrecque, Richard Fournier and Ron Hall provided helpful insights through scientific discussions on biomass mapping in the Canadian context. We thank Philippe Villemaire for GIS and image processing assistance and S. Dagnault, J.-P. Bérubé, P. Villemaire, J. Boudreau, G. Simard, N. Laflamme, J. Donnelly and G. Strickland for field assistance in collecting ground truth data supporting this work.

REFERENCES

[1] LEBOEUF, A, BEAUDOIN, A., FOURNIER, R.A., GUINDON, L., LUTHER, J. E. AND LAMBERT, M.-C., 2005: Mapping tree biomass of northern boreal forest using Shadow fraction from Quickbird imagery. In proceedings of 26th Canadian Symposium on Remote Sensing. June 14-16, 2005, Wolfville, NS, 5p.

[2] FAZAKAS, Z., NILSSON, M. AND OLSSON, H., 1999: Regional forest biomass and wood volume estimation using satellite data and ancillary data. Agricultural and Forest Meteorology. 98-99, pp. 417-425.

[3] REESE, H., NILSSON, M., SANDSTRÖM, P., AND OLSSON, H., 2002: Applications using estimates of forest parameters derived from satellite and forest inventory data. Comp. Electr. Agric. 37, pp. 37–55.

[4] LUTHER, J. E. FOURNIER, R. A. HALL, R. J. UNG, C-H. GUINDON, L. PIERCEY, D. E., LAMBERT, M.-C AND BEAUDOIN, A., 2002: A strategy for mapping Canada's forest biomass with Landsat TM imagery. Proceedingsof IEEE International Geoscience and Remote Sensing Symposium and the 24th Canadian Symposium on Remote Sensing (CD-ROM). June 24-28, 2002, Toronto, ON. IEEE, Piscataway, New Jersey, 4 p.

[5] LABRECQUE, S., FOURNIER, R.A., LUTHER, J. E. AND PIERCEY, D. E., A comparison of four methods to map forest biomass from Landsat-TM and inventory data in western Newfoundland. Forest Ecology and Management (Submitted).

[6] BERNIER P.Y., FOURNIER R.A., UNG C.H., ROBITAILLE G., LAROCQUE G.R., LAVIGNE M.B., BOUTIN R., RAULIER F., PARÉ D., BEAUBIEN J. AND DELISLE C., 2002: Linking ecophysiology and forest productivity: An overview of the ECOLEAP project. The Forestry Chronicle. 75, 3, pp. 417-421.

[7] UNG, C.-H., BEAUDOIN, A. BERNIER, P.Y., BEAUBIEN, J., BÉRUBÉ, J.-P., DAGNAULT, S., GILLIS, M., GUINDON, L., LAMBERT, M.-C. AND RAULIER. F., 2002: Estimating Sub-Arctic Forest Carbon Stock: Concepts, and Methods. Natural Resources Canada, Canadian Forest Service. Internal Report, 65 p.

[8] OUELLET, D., 1983: Biomass equations for black spruce in Quebec. Environ. Can., Can. For. Serv., Laurentian Forestry Research Centre, Info. Report. 11p.

[9] BEAUBIEN, J., CIHLAR, J., SIMARD, G. AND LATIFOVIC, R., 1999: Land cover from multiple thematic mapper scenes using a new enhancement-classification methodology. Journal of Geophysical Research., 104, pp. 27909-27920.

[10] BEAUDOIN, A., GUINDON, L., LAMBERT, M.-C., UNG, C.H., SIMARD, G. LUTHER, J.E. AND FOURNIER, R.A., 2003: A method for scaling up biomass of Canadian subarctic forests from tree to landscape levels using ground plots, Quickbird and Landsat data. In Proceedings of 25th Canadian Symposium on Remote Sensing and 11ième Symposium de l'Association Québécoise de la Télédétection. October 14-16 2003, Montreal, QC, 10 p.