Biomass estimation over a large area based on standwise forest inventory data and ASTER and MODIS satellite data: A possibility to verify carbon inventories

P. Muukkonen a,⁎, J. Heiskanen b

a Finnish Forest Research Institute, P.O. Box 18, FI-01301 Vantaa, Finland
b Department of Geography, University of Helsinki, P.O. Box 64, FI-00014 University of Helsinki, Finland

Received 2 June 2006; received in revised form 18 October 2006; accepted 21 October 2006

Abstract

According to the IPCC GPG (Intergovernmental Panel on Climate Change, Good Practice Guidance), remote sensing methods are especially suitable for independent verification of the national LULUCF (Land Use, Land-Use Change, and Forestry) carbon pool estimates, particularly the aboveground biomass. In the present study, we demonstrate the potential of standwise (forest stand is a homogenous forest unit with average size of 1–3 ha) forest inventory data, and ASTER and MODIS satellite data for estimating stand volume (m³ ha⁻¹) and aboveground biomass (t ha⁻¹) over a large area of boreal forests in southern Finland. The regression models, developed using standwise forest inventory data and standwise averages of moderate spatial resolution ASTER data (15 m×15 m), were utilized to estimate stand volume for coarse resolution MODIS pixels (250 m×250 m). The MODIS datasets for three 8-day periods produced slightly different predictions, but the averaged MODIS data produced the most accurate estimates. The inaccuracy in radiometric calibration between the datasets, the effect of gridding and compositing artifacts and phenological variability are the most probable reasons for this variability. Averaging of the several MODIS datasets seems to be one possibility to reduce bias. The estimates obtained were significantly close to the district-level mean values provided by the Finnish National Forest Inventory; the relative RMSE was 9.9%. The use of finer spatial resolution data is an essential step to integrate ground measurements with coarse spatial resolution data. Furthermore, the use of standwise forest inventory data reduces co-registration errors and helps in solving the scaling problem between the datasets. The approach employed here can be used for estimating the stand volume and biomass, and as required independent verification data.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Boreal forests; Carbon cycle; Forest vegetation

1. Introduction

Forests play an important role in global carbon cycling, since they are large pools of carbon as well as potential carbon sinks and sources to the atmosphere. Accurate estimation of forest biomass is required for greenhouse gas inventories and terrestrial carbon accounting. The needs for reporting carbon stocks and stock changes for the Kyoto Protocol have placed additional demands for accurate surveying methods that are verifiable, specific in time and space, and that cover large areas at acceptable cost (IPCC, 2003; Krankina et al., 2004; Patenaude et al., 2005; UNFCCC, 1997).

Remote sensing has opened an effective way to estimate forest biomass and carbon (Rosenqvist et al., 2003). According to the IPCC GPG (Intergovernmental Panel on Climate Change, Good Practice Guidance) (IPCC, 2003), remote sensing methods are especially suitable for verifying the national LULUCF (Land Use, Land-Use Change, and Forestry) carbon pool estimates — particularly the aboveground biomass. The purpose of verifying national greenhouse gas inventories is to establish their reliability and to monitor the accuracy of the numbers reported by independent means.

At continental and global scale biomass mapping, the coarse spatial resolution optical sensors, such as the NOAA AVHRR
(Dong et al., 2003; Häme et al., 1997) and Moderate Resolution Imaging Spectroradiometer (MODIS) (Baccini et al., 2004), have been useful due to the good trade-off between spatial resolution, image coverage and frequency in data acquisition (Lu, 2006). However, for quantifying biomass at local to regional scales, data provided by finer spatial resolution instruments, such as Landsat TM (Fazakas et al., 1999; Häme et al., 1997; Krankina et al., 2004; Tomppo et al., 2002; Turner et al., 2004) and Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) (Muukkonen & Heiskanen, 2005) are required.

Biomass estimation over large areas using coarse spatial resolution data has been limited because of the mixed pixels and the huge difference between the support of ground reference data and pixel size of the satellite data (Lu, 2006). Mixed pixels mean that due to the relatively small mean forest stand (forest stand is a homogenous unit) size (1–3 ha) (Hyyppä & Hyyppä, 2001; Poso, 1983), the coarse resolution pixels usually receive response from several stands, which makes the direct biomass estimation problematic. Typically, finer spatial resolution satellite data has been used as an intermediate step when relating ground reference data with coarser spatial resolution data, usually by regression techniques (Häme et al., 1997; Iverson et al., 1994; Tomppo et al., 2002). Häme et al. (1997) concluded that regression models derived using ground reference data and Landsat TM satellite data can be utilized

Table 1
Descriptive statistics for forests of the Finnish Forestry Centres (Finnish Statistical Yearbook of Forestry, 2001; Korhonen et al., 2000a,b,c, 2001; Tomppo et al., 1998a, b, 1999a,b,c,d, 2000, 2001)

<table>
<thead>
<tr>
<th>Forestry Centre</th>
<th>0</th>
<th>1a</th>
<th>1b</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total area (1000 ha)</td>
<td>153</td>
<td>669</td>
<td>696</td>
<td>1736</td>
<td>1430</td>
<td>1078</td>
<td>1227</td>
<td>1444</td>
<td>1945</td>
<td>1658</td>
<td>1651</td>
<td>1778</td>
</tr>
<tr>
<td>Forestry landa (1000 ha)</td>
<td>89</td>
<td>401</td>
<td>495</td>
<td>1064</td>
<td>954</td>
<td>796</td>
<td>911</td>
<td>1246</td>
<td>1375</td>
<td>1397</td>
<td>1333</td>
<td>1537</td>
</tr>
<tr>
<td>Proportion of upland soil forests from all forestry land (%)</td>
<td>99.4</td>
<td>98.9</td>
<td>80.6</td>
<td>77.7</td>
<td>83.8</td>
<td>81.1</td>
<td>83.6</td>
<td>74.2</td>
<td>59.5</td>
<td>73.5</td>
<td>84.2</td>
<td>64.9</td>
</tr>
<tr>
<td>Proportion of bare rock forests from all upland soil forests (%)</td>
<td>55.5</td>
<td>26.0</td>
<td>6.0</td>
<td>10.2</td>
<td>4.3</td>
<td>5.2</td>
<td>5.2</td>
<td>4.4</td>
<td>3.6</td>
<td>2.7</td>
<td>2.3</td>
<td>0.9</td>
</tr>
<tr>
<td>Mean stand volume (m³ ha⁻¹)</td>
<td>131.5</td>
<td>152.1</td>
<td>108.8</td>
<td>147.3</td>
<td>156.3</td>
<td>146.0</td>
<td>141.4</td>
<td>141.5</td>
<td>103.0</td>
<td>124.6</td>
<td>121.0</td>
<td>111.7</td>
</tr>
</tbody>
</table>

a Consists of forest land and scrub land.
with NOAA AVHRR data. Furthermore, Tomppo et al. (2002) showed that models for estimating stand volume and aboveground biomass from Landsat TM data may be used as an intermediate step between ground measurements and coarse resolution IRS-1C Wide Field Sensors (WiFS) data.

In this study, the regression models by Muukkonen and Heiskanen (2005) developed using standwise forest inventory data and moderate resolution ASTER data were utilized to estimate biomass with coarse resolution MODIS data for a large area. The study demonstrates one possible approach to integrate multiscale standwise forest inventory data, and ASTER and MODIS data for estimating biomass in boreal forests. The aim is to provide required independent verification data for carbon inventories.

2. Material and methods

2.1. Study area

The study area covers 11 out of 14 (labeled as 0–10) regional Finnish Forestry Centres (Fig. 1 & Table 1). It is mainly characterized by coniferous forests. The Forestry Centres 11–13 were excluded since the forests and climate of northern Finland are clearly different from those in southern Finland. Forests in this southern boreal zone consist of several tree species, typically Norway spruce (Picea abies), Scots pine (Pinus sylvestris) and broad-leaved birches (Betula pendula and B. pubescens). The most common understory species are the dwarf shrubs bilberry (Vaccinium myrtillus) and lingonberry (V. vitisidaea). The climate of the study area is not extremely cold compared with that of similar latitudes elsewhere on Earth. Temperatures normally decrease towards the north, the growing season becoming shorter and the effective temperature sum smaller, whereas from west to east the climatic trend is from oceanic to continental (Heikkinen, 2005).

2.2. Remote sensing data and its processing

While the size of the forest stands and therefore the standwise inventory data is too small to integrate it directly with coarse resolution MODIS data, the standwise averages of higher resolution ASTER data were used for developing regression models, which were utilized with MODIS data. This reduces the effect of mixed pixels since the average reflectance of forest stands is more pure than that of 250 m resolution MODIS pixels. This also avoids the averaging of the ground reference data. In the present study, we employed the non-linear regression models of Muukkonen and Heiskanen (2005)

\[
Y_{\text{trees}} = \exp(26.80) \cdot (1 + \text{RED})^{-2877.39} \cdot \text{NIR}^{7.09} \times \exp(2739.64 \cdot \text{RED}) \cdot \exp(-42.73 \cdot \text{NIR}), (r^2 0.56)
\]  

\[
Y_{\text{all vegetation}} = \exp(26.29) \cdot (1 + \text{RED})^{-2907.02} \cdot \text{NIR}^{6.90} \times \exp(2770.31 \cdot \text{RED}) \cdot \exp(-41.73 \cdot \text{NIR}), (r^2 0.56)
\]

using red and near infrared (NIR) reflectance (ASTER bands 2 and 3, respectively) as predictors, applied pixel-by-pixel to the MODIS reflectance data over southern Finland in order to estimate aboveground biomass of trees and aboveground biomass of all forest vegetation (t ha\(^{-1}\)) (see Fig. 2 for an overview of the estimation process). The models are based on standwise forest inventory data and standwise averages of ASTER data (Fig. 3). Although the regression models are based on a relatively small study area (two ASTER images covering totally 60 km × 120 km with field data of 3700 ha) (Fig. 1), it is adequately representative of managed and unmanaged forests in southern Finland.

In this study, we utilized red and near infrared (NIR) spectral bands of ASTER (bands 2 and 3) and MODIS (bands 1 and 2) (Fig. 4). The spatial resolution of ASTER bands is 15 m and the spatial resolution of MODIS bands is 250 m. We used ASTER product AST_07 (Abrams, 2000) and MODLAND product MOD09Q1 providing surface reflectance data (Justice et al., 1998, 2002). MODIS data was downloaded for three 8-day periods for growing season 2001 (July 4th–11th, August 13th–20th, August 21st–28th) while the ASTER data was acquired June 26th). MODIS Level 1 data is geolocated to the sub-pixel accuracy, the geolocation accuracy being approximately 50 m at nadir (Wolfe et al., 2002). The pixelwise average of these three periods was also calculated and used for estimations.

Because of the differences in ASTER and MODIS spectral bandwidths (Fig. 4), particularly between NIR bands, the bands
were calibrated using linear regression analysis (Häme et al., 1997). The following linear models were used:

\[ y_{\text{ASTER}}^{(2)} = -0.001 + 1.004 \cdot x_{\text{MODIS}}^{(1)}, \quad (r^2 0.44) \]  

\[ y_{\text{ASTER}}^{(3)} = 0.018 + 0.898 \cdot x_{\text{MODIS}}^{(2)}, \quad (r^2 0.63). \]

The terms \( a \) and \( b \) of these linear models were calculated from:

\[ a = \frac{\sigma(y)}{\sigma(x)} \]  

\[ b = \bar{y} - a \cdot \bar{x} \]

where \( \bar{y} \) and \( \bar{x} \) are the means of the variables \( y \) and \( x \), and \( \sigma(y) \) and \( \sigma(x) \) the standard deviations, respectively (Curran & Hay, 1986). The parameterizations of the linear models are based on the overlay of ASTER and MODIS data for all pixels in the study area.

To compare our results according to independent nation-wide field measurements based on Finnish National Forest Inventory (NFI) data we also predicted the stand volume (m\(^3\) ha\(^{-1}\)) by using the non-linear regression model of Muukkonen and Heiskanen (2005):

\[ y^{(\text{volume})} = \exp(24.79) \cdot (1 + \text{RED})^{-675.01} \cdot \text{NIR}^{6.33} \times \exp(588.65 \cdot \text{RED}) \cdot \exp(-39.43 \cdot \text{NIR}), \quad (r^2 0.55). \]

We used nation-wide stand volume measurements since direct biomass measurements are not available. Predicted stand volumes were compared to the NFI volume estimates for Finnish Forestry Centres (Korhonen et al., 2000a,b,c, 2001; Tomppo et al., 1998a, 1999a,b,c,d, 2000, 2001).

2.3. Forest mask dataset

A forest mask consisting of forest land on upland soils was derived from the Finnish CORINE Land Cover 2000 product at a 25 m grid size (CLC2000-Finland, 2005). The geometric accuracy of the IMAGE2000 data used for producing CLC2000 map is high, the average RMSE being 10 m. Forests and transitional woodlands/shrubs (crown basal area of 10–30%) on mineral and rocky soil were included in the forest mask. These classes correspond to the NFI categories of pure forest land and other wooded land (with an annual growth rate of 1 m\(^3\) a\(^{-1}\)) on upland soils. The MODIS pixels consisting of upland soil forest were separated from the pixels with similar reflectance characteristics by calculating the fractional forest cover for pixels and using a threshold value of 80%. In other words, in the mask more than 80% of the pixel area consists of upland soil forest land. The relatively high threshold value was chosen in order to reduce the effect of mixed pixels.

3. Results and discussion

3.1. Validation of volume estimates

The regression models developed using standwise forest inventory data and ASTER satellite data (Muukkonen & Heiskanen, 2005) were successfully utilized with MODIS
data for estimating stand volume (m$^3$ ha$^{-1}$) (Fig. 5a and b). The most accurate estimates were produced by averaged MODIS data. Fig. 5b demonstrates the differences in the estimation accuracy using three 8-day MODIS composites. The differences in the accuracy are relatively large and the estimation errors are considerably lower when using the average of three composites instead of the single datasets. The first MODIS dataset (July 4th–11th) consistently produced underestimates, the second (August 13th–20th) produced overestimates and the results of the third dataset (August 21st–28th) depended on the Forestry Centre. The averaging removes these systematic and non-systematic differences effectively. The inaccuracy in radiometric correction of the atmospheric effects between the datasets is the most probable reason for the differences. The averaging might also reduce the effect of phenological variations and gridding and compositing artifacts of the composite data. The results suggest that data product has to be chosen very carefully. Averaging of the several composites seems to be one possibility to reduce bias.

The difference between NFI and best MODIS estimates varied between $-16.0$ and $10.6$ m$^3$ ha$^{-1}$ and their relative counterparts between $-12.7\%$ and $8.0\%$. For whole study area (southern Finland) the estimation error (NFI estimate–MODIS estimate) was $-4.0$ m$^3$ ha$^{-1}$ ($-3.6\%$). This difference is quite small when considering that the estimation was done over a large area in southern Finland while the models of Muukkonen and Heiskanen (2005) are based on a relatively small amount of ground reference data. In the study of Tomppo et al. (2002), the mean difference was $-3.0$ m$^3$ ha$^{-1}$ ($-3.5\%$), which is in the same magnitude as the results of this study. They tested the simultaneous use of moderate resolution Landsat TM (25 m×25 m) and coarse resolution IRS-1C WiFS satellite data (188.3 m×188.3 m) for the Forestry Centre labelled by “1b” in Fig. 1.

**Table 2**

Comparison of MODIS stand volume (m$^3$ ha$^{-1}$) predictions and NFI stand volume measurements

<table>
<thead>
<tr>
<th>MODIS</th>
<th>RMSE$^a$</th>
<th>RMSE$^b$(%)</th>
<th>Bias$^c$</th>
<th>Bias$^d$(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10.1</td>
<td>7.6</td>
<td>-3.9</td>
<td>-2.9</td>
</tr>
<tr>
<td></td>
<td>1.38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.190</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$ RMSE = $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y})^2}$, where $\hat{y}$ is the predicted value, $y$ is the NFI estimate, and $n$ is the number of Forestry Centres.

$^b$ RMSE$^r$ = $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2} \times 100$, where $\bar{y}$ is the mean of the observed values.

$^c$ Bias = $\frac{1}{n}\sum_{i=1}^{n} (y_i-\hat{y})$

$^d$ Bias$^r$ = $\frac{1}{n}\sum_{i=1}^{n} \frac{(y_i-\hat{y})}{\bar{y}} \times 100$

$^e$ Significance of bias with degrees of freedom $n-1$ is based on

$t = \frac{\text{Bias}}{\text{Bias}_r} = \sqrt{\frac{n}{\sum_{i=1}^{n}(y_i-\hat{y})^2}}$

(Ranta et al., 1999).

**Fig. 5.** Comparison of estimated stand volume and NFI stand volume. The estimates were produced by averaged MODIS data (sub-figure a). The Finnish Forestry Centre network is presented in Fig. 1. The sub-figure b shows that estimates produced by different MODIS datasets provided clearly different results.

**Fig. 6.** Difference between stand volume estimates and NFI measurements. The signs “+” and “−” represent the over- and underestimation, respectively.
Table 2 shows the RMSE and bias for stand volume estimates. The RMSE for this large area should be considered as fairly good. Although, the coefficients of determination \( r^2 \) were low in the regression analysis of ground reference data and ASTER data (Muukkonen & Heiskanen, 2005), the best MODIS stand volume estimates for the Forestry Centre level were quite satisfactory.

The \( t \)-test indicates that MODIS estimates are not statistically significantly different from the NFI estimates (Table 2 and Fig. 5a). We studied the correlation coefficients between the stand volume estimates and descriptive characteristics of the Forestry Centres (see Table 1 for descriptive characteristics). Furthermore, we also studied such characteristic as dominant tree species and forest fertility levels. However, there was no statistically sound evidence that the prediction error is dependent on these characteristics (for prediction error see Fig. 6).

### 3.2. Aboveground biomass estimates

Estimates of the average aboveground biomass of trees and the average aboveground biomass of all vegetation including trees and understorey vegetation \( (t \text{ ha}^{-1}) \) are shown in Table 3. Furthermore, Fig. 7 shows the biomass map for the whole study area. The average aboveground biomass of all forest vegetation growing in upland mineral soils was estimated to be approximately 85 \( t \text{ ha}^{-1} \) for the area of Forestry Centres 0–10 in southern Finland. The corresponding value only for aboveground tree biomass was 83 \( t \text{ ha}^{-1} \). As reported in the Table 3, our biomass estimates for the Forestry Centre 1b are slightly higher than previous estimate of Tomppo et al. (2002). However, our biomass estimates for the whole of southern Finland are rather close to the estimates of Liski et al. (2006) which are based on NFI volume measurements and biomass conversion.

The accuracy assessment of biomass estimates is often limited by the lack of appropriate data (Lu, 2006). In this study, the stand volume \( (\text{m}^3 \text{ ha}^{-1}) \) estimates of the Finnish NFI provided a good source of data for validation. In addition, Muukkonen and Heiskanen (2005) have concluded that the biomass and stand volume predictions have equal reliability.

### 3.3. Applicability of the method

The results indicate that models developed for estimating stand volume and biomass based on standwise forest inventory data and moderate resolution ASTER data (Muukkonen & Heiskanen, 2005) can be utilized also to the coarse resolution MODIS data. The demonstrated approach can be used as a cost-effective tool to produce preliminary biomass estimates for large areas where more accurate national or large scale forest inventories do not exist. However, although the estimates were reasonable when averaged for large areas, the pixel level estimates could have low accuracy. Furthermore, the method requires a reliable forest mask, which is not always available.

The Finnish NFI based large scale carbon stock estimates are considered to be quite reliable, but IPCC Good Practise...
Guidance (IPCC, 2003) has put emphasis to the development of independent verification methods. The approach illustrated in this study can be used as verification data since it is based on independent ground reference data. The independency is one requirement of the useful verification data. The verification method should also cover the same range of forest types and management regimes as estimates to be verified. The presented method fills this requirement since the forest mask is comparable to the definition of forest in the Finnish NFI. Furthermore, the study area used for developing regression models is representative for managed and unmanaged forests of southern Finland.

The biomass estimation using coarse resolution remote sensing data has been limited because of the huge difference between the support of ground reference data and pixel size of the remote sensing data (Lu, 2006). For example, the errors in the image registration and location of the sample plots produce high estimation errors at the pixel level since the field plots are typically small (Mäkelä & Pekkarinen, 2004). We managed this problem by using standwise forest inventory data instead of plotwise measurements normally used in remote sensing application of forestry. The area of a forest stand in southern Finland is usually between 1 and 3 ha (Poso et al., 1987). The area of forest stands is still too small to integrate standwise forest inventory data directly with coarse resolution MODIS data. Therefore, standwise averages of higher resolution ASTER data were used for developing regression models, which were successfully utilized with MODIS data. The standwise averages of higher resolution data correspond to the homogeneous pixels of coarse resolution data and provide the connection between the ground reference data and coarse resolution satellite data.

Acknowledgements

The authors would like to thank the suppliers of the ASTER and MODIS data. Thanks also go to the following five Finnish foundations for financial support of the work of Dr. Muukkonen: Marjatta & Eino Kollin säätiö, Metsämiesten säätiö, Otto A. Malmin hajotusrahasto, Mikko Kaloisen säätiö and Helsingin yliopiston matematiikan ja luonnontieteiden rahasto. We are also grateful for Dr. Raisa Mäkipää (Finnish Forest Research Institute) and Prof. Petri Pellikka (Department of Geography, University of Helsinki) for their valuable comments.

References


