Estimating Wood Volume and Basal Area in Forest Compartments by Combining Satellite Image Data with Field Data

Johan Holmgren
Estimating Wood Volume and Basal Area in Forest Compartments by Combining Satellite Image Data with Field Data

Johan Holmgren

Arbetsrapport 47 1998

Master thesis in remote sensing
Supervisor: Mats Nilsson, Swedish University of Agricultural Scienses
Estimating Wood Volume and Basal Area in Forest Compartments by Combining Satellite Image Data with Field Data

ABSTRACT

Landsat TM satellite image data and field data from the National Forest Inventory (NFI) were combined using the kNN method. Wood volume and basal area estimations were done for an area in western Sweden. For each pixel in the inventory area, the distance in feature space was calculated to each reference (NFI) plot. The estimation for each pixel in the inventory area could be determined by choosing attribute values for the k closest neighbours in feature space and weighting them with the inverse squared distances to the estimation pixel. Two distance functions were used; Mahalanobis distance and prediction difference. The pixels were aggregated into 296 validation compartments and an average for wood volume and basal area was derived. Validation data for the compartments was acquired from a field inventory conducted by STORA Forest and Timber. The main tree species were Scots pine (Pinus sylvestris), Norway spruce (Picea abies) and birch (Betula spp.). Volume and basal area could be estimated with a standard error of 21% by combining satellite image and NFI data for compartments in the study area that had a volume between 100 and 300 m³/ha. The standard error was high and wood volume was strongly overestimated for volumes less than 100 m³/ha. Wood volume was underestimated by 33% for volumes greater than 300 m³/ha. Neither correcting of reflectance for topography nor adding temperature sum improved wood volume estimation accuracy in the validation compartments. However, wood volume and basal area estimation accuracy improved if site index, age and mean tree height were included. The standard error was then 12% for compartments with a volume between 100 and 300 m³/ha. Using Prediction difference reduced the RMSE, compared to using Mahalanobis distance, if mean tree height was included as predicting variable.

Key words: Satellite image, forest inventory, compartment information, forest variables, data integration.
PREFACE

This study is a MSc thesis in forestry at the Department of Forest Resources Management and Geomatics, SLU, Umeå. STORA Forest and Timber was host for this project, financing satellite data, Digital Elevation Model (DEM) and gave a reward for the work. Jan Gustafsson, my contact person at STORA Forest and Timber, suggested this project and discussed the needs of information in forestry. The Swedish Land Survey kindly gave a special offer for the DEM. Dr. Ola Lindgren gave information about the survey used for validation. The following people all work at the Department of Forest Resources and Geomatics, SLU, Umeå. Dr. Mats Nilsson was my supervisor. Steve Joyce wrote programs, helped with image processing, and together with Asst. Prof. Sören Holm, came up with the linear algebra for the correlation-weighted distance function. Olle Hagner helped me with image processing. Prof. Håkan Olsson and Hampus Holmström gave comments on the report in general. Heather Reese corrected English spelling and grammar.

Umeå, September 1998

Johan Holmgren
TABLE OF CONTENTS

INTRODUCTION...................................................................................................................... 2

MATERIAL AND METHODS ................................................................................................. 4

Study area ................................................................................................................................ 4

Satellite image data ............................................................................................................. 4

Sampling spectral values ..................................................................................................... 4

Reference data ..................................................................................................................... 5

Ancillary data ....................................................................................................................... 5

Validation data .................................................................................................................... 6

Topography correction ........................................................................................................ 6

Temperature sum .................................................................................................................. 7

Estimation algorithms ......................................................................................................... 8

Validation procedure .......................................................................................................... 9

RESULT .................................................................................................................................. 12

DISCUSSION ....................................................................................................................... 15

REFERENCES ..................................................................................................................... 17
INTRODUCTION

Accurate information about the state of the forest is needed to make good decisions in forestry management planning. Some examples of different decisions are: strategic decisions about timber policies, decisions made to allocate forest operations in time and space and operational decisions such as work plans for forest operations. The forest compartment is the smallest unit for which decision-supporting information is collected and stored (Lindgren, 1984). Besides field inventories, several techniques have been developed for retrieving data about forest variables on a compartment level.

Aerial photographs have been used for forestry mapping since the middle of this century. Visual interpretation of aerial photographs to delineate forest into uniform compartments has been commonly used. Information about the type of forest has been supplemented by information about tree size, usually height, together with compartment density (Duggin et al., 1989). Measurements of tree height with stereo plotters and aerial photographs have been used operationally in Swedish forestry simultaneously with compartment delineation (Åge, 1985).

The first civil satellites monitoring the earth's resources, which were launched in the beginning of the 1970s, delivered data with 80 m pixels. The Landsat TM sensors, first launched in 1982, had higher spatial and spectral resolution than earlier sensors. The TM sensor has seven spectral bands in the visible, short wave infrared and thermal spectrum with a picture element (pixel) corresponding to 30 by 30 m on the ground for six of the bands (Kramer, 1996). Low spatial resolution provides the possibility to create relatively simple algorithms for estimating stand variables because each pixel will represent the average reflection from an area with a number of trees. Because of the large size of each image, training data can be used from a large area, for example, from national forest inventories. Satellite sensors are planned with a spatial resolution of 5-10 m (Konecny, 1996). The positioning of reference data needs to be more accurate if high-resolution sensors are used. Satellite navigation systems provide the possibility to determine exact locations of field plots. High-resolution sensors receive signals from each pixel that represents the reflectance from a part of a single tree, making the procedure to estimate variables, describing stand structure, more complex. Furthermore, if the high-resolution sensor is airborne, the geometry is more complex due to the viewing and illumination conditions.

In Finland and Sweden, field data from the National Forest Inventory (NFI) have been used as reference data for estimation of wood volume using satellite spectral data. The obtained estimation accuracy has been found to be limited regarding the pixel level, but accuracy increases if the pixel-wise estimates are aggregated into compartments (Hagner, 1990; Muinonen and Tokola, 1990) or landscape (Iverson et al., 1989). Compartment-wise estimations of wood volume based on data from the Landsat and SPOT satellites have shown to be comparable with estimations carried out using subjective inventories (Hagner, 1990; Poso et al., 1987). The k Nearest Neighbour (kNN) method has been used operationally in the Finnish NFI since 1990 (Tomppo, 1990). In this method, field data and satellite image data are combined so each reference plot receives the corresponding...
pixel value of the image. For each pixel in the inventory area, the distance in feature space is calculated to each reference plot. The estimation for each pixel in the inventory area is determined by choosing the attribute values for the $k$ closest reference plots (neighbours) in the feature space and weighting them with the inverse square distances to the estimation pixel. The method is also referred to as the Reference sample plot method (Tokola et al., 1996).

Digital techniques have made it easier to integrate data from different sources using a Geographical Information System (GIS). There was a tendency to treat remote sensing in isolation in the early days of environmental remote sensing. During the 1980s a major development occurred. One could achieve more effective information extracted from the remotely sensed data by feeding additional data sets into the image processing systems (Leckie, 1990a; Jackson and Mason, 1986). Also, there is a potential synergism of combining different sensors (e.g. Leckie, 1990b).

The objectives of this study are to:

- Evaluate wood volume and basal area estimations by combining NFI data and satellite image data on a compartment level using the $k$NN method.
- Investigate how the use of different distance functions effects the estimation accuracy for wood volume and basal area in the $k$NN method.
- Investigate the trend of overestimating low volumes and underestimating high volumes in the $k$NN method.
- Investigate if using ancillary data will improve the estimation accuracy of wood volume and basal area.
MATERIAL AND METHODS

Study area
The study area was located in western Sweden, adjacent to the Norwegian border, including most of the forest owned by STORA Forest and Timber in the county of Värmland (Figure 1). The dominating tree species were Norway spruce (*Picea abies*), Scots pine (*Pinus sylvestris*) and Birch (*Betula spp.*). The topography was extraordinary for the country, with altitudes from 44 to 723 m. The landscape had several long depressions, forming rivers and lake-systems. These formations were elongated in a Northwest to Southwest direction, with rivers and creeks flowing from Norway to Lake Vänern.

![Figure 1. Location of the study area.](image)

Satellite image data
A Landsat 5 TM quarter scene covering the study area was acquired July 6, 1995, path 196 and frame 18. The image was geometrically precision corrected to the Swedish National Grid and the pixel size was resampled to 25 m using cubic convolution interpolation by SCC Satellitbild in Kiruna. Furthermore, altitude induced displacement was corrected using a Digital Elevation Model (DEM) with a resolution of 50 m. The spectral bands 2-5 and 7 were used for estimations of wood volume and basal area.

Sampling spectral values
The polygons representing compartments and roads were converted to grids matching the image pixel. A grid of compartments without any roads was created. This layer was then shrunk, removing the boundary cells for each compartment. Spectral values from the image covered by this layer were sampled. Spectral values for the NFI plots were sampled using cubic convolution.
Reference data
Field data from the Swedish NFI were used as reference data for the estimations. The NFI is conducted as a yearly systematic field sample, consisting of both temporary and permanent plots. The plots are positioned along the sides of sample clusters called tracts (Ranneby et al., 1987). Permanent tracts contain 8 plots with a radius of 10 m and temporary tracts have 12 plots with a radius of 7 m. In the study area, tract side length is 1000 m for permanent and 1500 m for temporary plots (Hägglund, 1983). Altogether, 2209 NFI plots, inventoried during the period 1992-1996, were located within the study area. Positions of plots were determined by digitising their location from maps used by the survey team. The attribute data, such as wood volume, species composition and basal area were forecasted using growth functions (Söderberg, 1986) to the year of the Landsat image acquisition. The selected plots were located on forested land with an annual wood production greater than 1 m³/ha. Some reference plots were affected by major changes between the time of data collection and the acquisition of the TM image. For example, some plots were harvested, thinned or affected by natural disturbances. Plots that had non-representative spectral values due to incorrect plot locations were also to be removed from the data set. Therefore, plots suspected to be non-representative were discarded from the NFI data. The Digital Numbers (DN) were estimated for each TM band using a regression function of wood volume, and plots with the greatest residuals were subsequently removed. A number of 1703 NFI plots remained after the outlier reduction. Statistics for the reduced and original data set are presented in Table 1. Correlation between variables are listed in Table 5.

Table 1. The NFI plots, which were distributed over the Landsat TM quarter scene and used as reference data, for the whole and reduced material

<table>
<thead>
<tr>
<th>Variable</th>
<th>Average</th>
<th>Std. dev.</th>
<th>Max</th>
<th>Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site Index (m)</td>
<td>23 (23)</td>
<td>4 (4)</td>
<td>32 (32)</td>
<td>8 (0)</td>
</tr>
<tr>
<td>Age (year)</td>
<td>53 (52)</td>
<td>39 (40)</td>
<td>180 (180)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Tree height (m)</td>
<td>12 (12)</td>
<td>8 (8)</td>
<td>33 (33)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Wood volume (m³/ha)</td>
<td>125 (125)</td>
<td>120 (122)</td>
<td>699 (699)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Wood volume pine (m³/ha)</td>
<td>48 (48)</td>
<td>69 (70)</td>
<td>572 (572)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Wood volume spruce (m³/ha)</td>
<td>65 (65)</td>
<td>99 (100)</td>
<td>659 (659)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Wood volume deciduous (m³/ha)</td>
<td>12 (12)</td>
<td>32 (32)</td>
<td>387 (387)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Basal area (m²/ha)</td>
<td>17 (17)</td>
<td>12 (13)</td>
<td>71 (71)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

* Numbers within brackets represents unreduced material.

Ancillary data
Estimations were made with different information levels, realistic for forest management performed by a forest company in Sweden. The four different levels used are described with following paragraphs:

0. Satellite data,
1. Satellite data together with a DEM,
2. Satellite data, DEM, site quality and age for the compartments, and
3. Satellite data, DEM, site quality, age and mean tree height for the compartments.
Validation data
STORA Forest and Timber inventoried the compartments, used for assessing the estimations, during the field season 1997. The inventory was committed in order to support strategic planning. Circular plots and a square lattice were used for the systematic sampling. Simulation studies have shown the average standard error of systematic sampling to be 80% of the error for simple random sampling (Lindgren, 1984). By average, eight to nine sample plots were distributed in each compartment. More than 600 compartments were inventoried. Because many compartments were divided into smaller areas during the field inventory, and thus had several sets of attribute data for each compartment polygon, only 296 could be used for the evaluation (Table 2 and 3). Several compartments had a rather irregular shape and consisted of more than one polygon.

Table 2. The validation compartments, a total number of 296, distributed over the Landsat TM quarter scene

<table>
<thead>
<tr>
<th>Variable</th>
<th>Average</th>
<th>Std. dev.</th>
<th>Max</th>
<th>Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site Index (m)</td>
<td>24</td>
<td>3</td>
<td>33</td>
<td>15</td>
</tr>
<tr>
<td>Age (year)</td>
<td>62</td>
<td>34</td>
<td>153</td>
<td>0</td>
</tr>
<tr>
<td>Tree height (m)</td>
<td>14</td>
<td>5</td>
<td>29</td>
<td>0</td>
</tr>
<tr>
<td>Wood volume (m3/ha)</td>
<td>156</td>
<td>97</td>
<td>479</td>
<td>0</td>
</tr>
<tr>
<td>Wood volume pine (m3/ha)</td>
<td>61</td>
<td>57</td>
<td>266</td>
<td>0</td>
</tr>
<tr>
<td>Wood volume spruce (m3/ha)</td>
<td>81</td>
<td>82</td>
<td>474</td>
<td>0</td>
</tr>
<tr>
<td>Wood volume deciduous (m3/ha)</td>
<td>14</td>
<td>24</td>
<td>216</td>
<td>0</td>
</tr>
<tr>
<td>Basal area (m2/ha)</td>
<td>20</td>
<td>9</td>
<td>42</td>
<td>0</td>
</tr>
<tr>
<td>Compartment size (ha)</td>
<td>19</td>
<td>16</td>
<td>85</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3. Means and standard errors from the field inventory of the validation compartments, for different volume classes and tree species

<table>
<thead>
<tr>
<th>Volume Class (m3/ha)</th>
<th>Average (m3/ha)</th>
<th>Std. dev. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 100</td>
<td>46</td>
<td>26</td>
</tr>
<tr>
<td>100 - 200</td>
<td>147</td>
<td>16</td>
</tr>
<tr>
<td>200 - 300</td>
<td>46</td>
<td>14</td>
</tr>
<tr>
<td>&gt;300</td>
<td>349</td>
<td>14</td>
</tr>
<tr>
<td>All forest</td>
<td>156</td>
<td>19</td>
</tr>
<tr>
<td>All forest pine</td>
<td>61</td>
<td>30</td>
</tr>
<tr>
<td>All forest spruce</td>
<td>81</td>
<td>32</td>
</tr>
<tr>
<td>All forest deciduous</td>
<td>14</td>
<td>130</td>
</tr>
</tbody>
</table>

Topography correction
It was assumed that topography could have an impact on the estimations. A model described by Teillet et al. (1982) was applied. Linear regression functions were derived for each spectral band using the positions of the reference plots

\[ DN_r = m \times \cos(i) + b \]  

where

\[ DN_r = \text{DN observed for sloped terrain}, \ m \text{ and } b \text{ are constants, and} \]

\[ i = \text{sun slope normal angle.} \]
Using the DEM together with information about sun position, the \( \cos(i) \) was derived for each pixel

\[
\cos(i) = \cos e \cos z + \sin e \sin z \cos(\Phi_s - \Phi_n)
\]

where
- \( e \) = surface normal zenith angle or terrain slope,
- \( z \) = solar zenith angle,
- \( \Phi_s \) = solar azimuth angle, and
- \( \Phi_n \) = surface aspect of the slope angle. 

Equation (1) was used for computing a parameter \( c \) (Table 4) for modelling the diffuse sky radiation

\[
c = b / m
\]  

Equation (2)

This parameter was added to the frequently used cosine correction

\[
DN_H = DN_r \left[ \frac{\cos(z) + c}{\cos(i) + c} \right]
\]

where
- \( DN_H \) = DN observed for a horizontal surface.

\( c = b / m \) 

Equation (3)

Equation (4)

Equation (5)

Table 4. \textit{Variables in the equation } \( DN_r = m \times \cos(i) + b \), where \( DN_r \) is the DN observed for sloped terrain, and \( i \) is the sun slope normal angle. \( R^2 \) is for using the model on the reference data

<table>
<thead>
<tr>
<th>Band</th>
<th>( b )</th>
<th>( m )</th>
<th>( c )</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>16.31</td>
<td>6.22</td>
<td>2.62</td>
<td>0.030</td>
</tr>
<tr>
<td>3</td>
<td>12.98</td>
<td>6.78</td>
<td>1.92</td>
<td>0.015</td>
</tr>
<tr>
<td>4</td>
<td>22.46</td>
<td>35.47</td>
<td>0.63</td>
<td>0.037</td>
</tr>
<tr>
<td>5</td>
<td>7.48</td>
<td>44.19</td>
<td>0.17</td>
<td>0.034</td>
</tr>
<tr>
<td>7</td>
<td>3.07</td>
<td>13.47</td>
<td>0.23</td>
<td>0.020</td>
</tr>
</tbody>
</table>

\textbf{Temperature sum}

The Temperature Sum (TS) was included as a predicting variable. It is defined as the sum of all daily mean values exceeding a chosen threshold value (Odin et al., 1983). It was derived as a function of \textit{latitude} and \textit{altitude}. The function is for a threshold of 5 degrees Celsius and based on data collected by the Swedish Meteorological and Hydrological Institute (SMHI) during a 30 year period from 513 stations in Sweden (Moren and Perttu, 1994).

\[
TS = 4922 - 60.4 \times \text{latitude} - 0.837 \times \text{altitude}
\]
Table 5. Correlation coefficients between predicting variables and wood volume on a pixel level

<table>
<thead>
<tr>
<th></th>
<th>Band 2</th>
<th>Band 3</th>
<th>Band 4</th>
<th>Band 5</th>
<th>Band 7</th>
<th>Age</th>
<th>Height</th>
<th>SIS</th>
<th>Tsum</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Band 2</td>
<td>1.00</td>
<td>0.89</td>
<td>0.70</td>
<td>0.87</td>
<td>0.84</td>
<td>-0.42</td>
<td>-0.67</td>
<td>-0.20</td>
<td>-0.28</td>
<td>-0.56</td>
</tr>
<tr>
<td>Band 3</td>
<td>0.89</td>
<td>1.00</td>
<td>0.52</td>
<td>0.88</td>
<td>0.91</td>
<td>-0.34</td>
<td>-0.60</td>
<td>-0.28</td>
<td>-0.30</td>
<td>-0.52</td>
</tr>
<tr>
<td>Band 4</td>
<td>0.70</td>
<td>0.52</td>
<td>1.00</td>
<td>0.72</td>
<td>0.56</td>
<td>-0.55</td>
<td>-0.66</td>
<td>0.04</td>
<td>-0.14</td>
<td>-0.54</td>
</tr>
<tr>
<td>Band 5</td>
<td>0.87</td>
<td>0.88</td>
<td>0.72</td>
<td>1.00</td>
<td>0.95</td>
<td>-0.42</td>
<td>-0.67</td>
<td>-0.21</td>
<td>-0.29</td>
<td>-0.57</td>
</tr>
<tr>
<td>Band 7</td>
<td>0.84</td>
<td>0.91</td>
<td>0.56</td>
<td>0.95</td>
<td>1.00</td>
<td>-0.36</td>
<td>-0.61</td>
<td>-0.24</td>
<td>-0.29</td>
<td>-0.51</td>
</tr>
<tr>
<td>Age</td>
<td>-0.42</td>
<td>-0.34</td>
<td>-0.55</td>
<td>-0.42</td>
<td>-0.36</td>
<td>1.00</td>
<td>0.75</td>
<td>-0.15</td>
<td>-0.06</td>
<td>0.60</td>
</tr>
<tr>
<td>Height</td>
<td>-0.67</td>
<td>-0.60</td>
<td>-0.66</td>
<td>-0.67</td>
<td>-0.61</td>
<td>0.75</td>
<td>1.00</td>
<td>0.21</td>
<td>0.24</td>
<td>0.83</td>
</tr>
<tr>
<td>SIS</td>
<td>-0.20</td>
<td>-0.28</td>
<td>0.04</td>
<td>-0.21</td>
<td>-0.24</td>
<td>-0.15</td>
<td>0.21</td>
<td>1.00</td>
<td>0.50</td>
<td>0.26</td>
</tr>
<tr>
<td>Tsum</td>
<td>-0.28</td>
<td>-0.30</td>
<td>-0.14</td>
<td>-0.29</td>
<td>-0.29</td>
<td>-0.06</td>
<td>0.24</td>
<td>0.50</td>
<td>1.00</td>
<td>0.22</td>
</tr>
<tr>
<td>Volume</td>
<td>-0.56</td>
<td>-0.52</td>
<td>-0.54</td>
<td>-0.57</td>
<td>-0.51</td>
<td>0.60</td>
<td>0.83</td>
<td>0.26</td>
<td>0.22</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Estimation algorithms

The forest variables were estimated using the $k$ Nearest Neighbour method (e.g. Tomppo, 1990; Tokola et al., 1996). The values for pixels were estimated as weighted mean values of the $k$ nearest samples in feature space. The $k$ nearest samples are defined as the $k$ reference plots in the feature space having the shortest distance to the pixel that is to be estimated. The Mahalanobis distance ($dm$) was used in the feature space of predicting variables and can be written for the distance between plot $r$ and $s$ as

$$dm^2_{rs} = (x_r - x_s)^T C^{-1} (x_r - x_s)$$  \hspace{1cm} (6)

where $C$ is the covariance matrix. Another distance, referred to as the prediction difference ($dp$) was used. This is a distance in feature space of predicted value, $\hat{y}_{pr}$ for variable $p$, plot $r$, and is, for the distance between plot $r$ and $s$, defined as

$$dp^2_{rs} = \sum_{p=1}^{P} (\hat{y}_{rp} - \hat{y}_{sp})^2$$  \hspace{1cm} (7)

A general linear model was used with independent variable $x_j$, where $j = 1,2,...,m$ for estimating $\hat{y}_{p}$. The model can be written as

$$y_p = \beta_{p,0} + x_1\beta_{p,1} + ... + x_m\beta_{p,m} + \epsilon_p$$  \hspace{1cm} (8)

where $\beta_0, \beta_1,...,\beta_m$ are unknown parameters and $\epsilon_p$ a random error with zero expectation. Ordinary Least Square (OSL) estimators where used for $\beta$,

$$\hat{\beta}_p = (X^T X)^{-1} X^T y_p$$  \hspace{1cm} (9)

Equation (9) was then inserted into Equation (7) and the distance $dp$ could be computed.
A number of various options to weight the chosen neighbours have been proposed (Tokola et al., 1996). In this study, the weights were chosen proportional to the inverse squared distance measured in the feature space

\[ dp^2 = \sum_{p=1}^{P} (x_r - x_s) \hat{\beta}_p \hat{\beta}_p^T (x_r - x_s)^T \]  

(10)

Studies have shown that at least 5 to 10 neighbours should be used in order to achieve as high estimation accuracy as possible (Nilsson, 1997; Tokola et al., 1996). High numbers of neighbours make the correlation stronger between estimated parameters than obtained for field measurements (Nilsson, 1997). These two facts were taken into consideration and it was decided that the five \((k=5)\) closest pixels in the feature space should be used.

**Validation procedure**

Estimations were made for each pixel in the compartments, and then the average of the estimation was derived, resulting in a mean value for each compartment. Estimations on pixel level were made for wood volume, basal area, and volume portion of pine, spruce and deciduous trees. Wood volume and basal area were aggregated to a compartment level but wood volume for different species were not evaluated due to a high standard error in the validation data.

\[ Var(D) = \frac{1}{N} \sum_{r=1}^{N} (\hat{y}_r - y_r)^2 - bias^2 \]  

(12)

where

\[ \hat{y}_r = \text{estimated mean value for compartment } r, \]

\[ y_r = \text{field measured mean value for compartment } r, \]

\[ D = \text{difference between estimated and field measured mean value for compartment.} \]

In order to make comparisons with different subsets of the material and forest variables, relative standard errors were calculated by dividing the standard error with the mean value from the objective field inventory. The bias was calculated as the mean of the
difference between the estimated mean value and the mean value derived from the field inventory.

If no consideration is taken for the random error in the ground truth data the result from the evaluation will be deceptive. A model with the assumption that bias is the same for different true values was used for the whole material and volume classes. The estimated values can be described as in Equation (13) (Ståhl, 1992).

\[ y_r = c + t_r + \varepsilon \]  

(13)

where \( c \) is a constant, \( t_r \) true value, and \( \varepsilon \) the random error. The relation between the true value \( t_r \) and the field measured value \( y_r \) is

\[ y_r = t_r + \delta_r \]  

(14)

where \( \delta_r \) is the random error from the objective field inventory and has an expected value of zero. \( D \) for the \( r \)th compartment is given as

\[ D_r = c - \delta_r + \varepsilon \]  

(15)

Because satellite image and validation data are independent, the variance of \( D \) can be written as

\[ Var(D) = Var(c - \delta_r + \varepsilon) = Var(\delta) + Var(\varepsilon) \]  

(16)

The variance for the random error can be derived as

\[ Var(\varepsilon) = Var(D) - Var(\delta) \]  

(17)

Thus, the variance of the satellite estimated random error could be derived as the variance of the differences between satellite and field values, minus the variance for the sigma term. Sigma was derived as the mean of the estimated variances for compartment means estimated with the field inventory. The standard error was then derived as the square root of \( Var(\varepsilon) \).

Low volumes were presumed to be overestimated and high volumes underestimated with the \( k \)NN method. Therefore, a simple model with a systematic error with trend was used for evaluating wood volume for the whole material. The model (Ståhl, 1992) can be described as

\[ y_r = a + b \times t_r + \varepsilon \]  

(18)

If the true values had been known, the parameter could have been estimated with regression of \( \hat{y} \) on \( t \). However, the relation between field measured and true values is described in Equation (14) and a correction factor can be derived as
\[ b = B \times \left[ 1 + \frac{\text{Var}(\delta)}{\text{Var}(y) - \text{Var}(\delta)} \right] \]  

(19)

\[ a = \bar{y} - b \times \bar{y} \]  

(20)

A value of \( b \) lower than 1 indicates that the inventory method overestimates low volumes and underestimates high volumes.

\[ \text{Var}(\varepsilon) = \text{Var}(\varepsilon') \times \frac{\left(1 - \frac{r^2 \times \text{Var}(y)}{\text{Var}(y) - \text{Var}(\delta)}\right)}{1 - r^2} \]  

(21)

Equation (21) gives the variance of the random error and should not be used for small materials because estimations of variance and correlation may be unsure. \( \text{Var}(\varepsilon') \) is the variance of the residuals and \( r \) the correlation between the satellite estimated values and the values from the objective field inventory. Then presenting the results, the difference between standard errors obtained with the two different models (Equation 13 and 18) is referred to as systematic error.
RESULT

Neither topography correction of illumination nor incorporating the TS had any effect on the systematic or standard error. Age and site index decreased the sum of standard and systematic error (RMSE) by 8% regardless of the distance used (Figure 2). Adding mean tree height lowered the RMSE by another 10%, but only if the prediction difference was used. The trend was the same for basal area (Figure 3).

Figure 2. The standard and systematic errors (%) of volume estimations for Level 0 (spectral values), Level 1 (spectral values and TS), Level 2 (spectral values, TS, age of forest, and site index), Level 3 (spectral value, TS, age of forest, site index, and mean tree height) with the prediction difference (dp) and the Mahalanobis distance (dm). The standard error was derived with a model with trend $\hat{y}_r = a + b \times t_r + \varepsilon$. The systematic error was derived as the difference between the standard error calculated by a model with trend and the standard error derived if a model with no trend $\hat{y}_r = c + t_r + \varepsilon$ would have been used. In the models, $\hat{y}_r$ is estimated value, $t_r$ true value and $\varepsilon$ random error. The symbols $a$, $b$ and $c$ are constants.

Figure 3. The standard and systematic errors (%) of basal area estimations. Definitions are the same as in Figure 2.
The \( a \) and \( b \) values in the model with a trend \( \hat{y}_r = a + b \times t_r + \epsilon \), where \( \hat{y}_r \) is the estimated value, \( t_r \) the true value and \( \epsilon \) the random error, are presented for both wood volume and basal area (Table 6).

Table 6. Values for constants \( a \) and \( b \) for volume and basal area (BA) estimations. *Symbols are the same as in Figure 2*

<table>
<thead>
<tr>
<th>Level / Distance</th>
<th>Volume a</th>
<th>Volume b</th>
<th>BA a</th>
<th>BA b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 0 dm</td>
<td>67</td>
<td>0.54</td>
<td>7.35</td>
<td>0.63</td>
</tr>
<tr>
<td>Level 1 dm</td>
<td>68</td>
<td>0.52</td>
<td>7.89</td>
<td>0.59</td>
</tr>
<tr>
<td>Level 2 dm</td>
<td>40</td>
<td>0.72</td>
<td>5.62</td>
<td>0.72</td>
</tr>
<tr>
<td>Level 3 dm</td>
<td>40</td>
<td>0.68</td>
<td>6.33</td>
<td>0.68</td>
</tr>
<tr>
<td>Level 0 dp</td>
<td>62</td>
<td>0.59</td>
<td>6.57</td>
<td>0.67</td>
</tr>
<tr>
<td>Level 1 dp</td>
<td>60</td>
<td>0.59</td>
<td>6.68</td>
<td>0.66</td>
</tr>
<tr>
<td>Level 2 dp</td>
<td>34</td>
<td>0.81</td>
<td>4.15</td>
<td>0.83</td>
</tr>
<tr>
<td>Level 3 dp</td>
<td>20</td>
<td>0.81</td>
<td>4.22</td>
<td>0.81</td>
</tr>
</tbody>
</table>

To facilitate interpretation, the \( a \) and \( b \) values were inserted into Equation (18). Not only the level of information, but also the different distance functions yielded different bias. Note that the orders of lines are different, dependent on what true value is estimated (Figure 4).

Figure 4. The bias for volume estimations as a function of true value using the model \( \hat{y}_r = a + b \times t_r + \epsilon \), where \( \hat{y}_r \) is the estimated value, \( t_r \) the true value and \( \epsilon \) the random error. *Definitions are the same as in Figure 2.*
Standard error and bias were calculated with the assumption that bias is constant in each volume class (Figure 5). The standard error was lower with satellite data, TS, site index, age and tree height (Level 3) compared with other levels in the middle class and the class with least volume. Using satellite data, DEM, site index, and age (Level 2) gave a lower standard error in the lowest and highest volume class compared with using only satellite data and DEM (Level 1). In the middle volume class (100-300 m$^3$/ha), standard error decreased from 21%, with Level 2, to 12% when tree height was also added (Level3).

![Figure 5. The Standard Error of volume estimations using prediction difference presented for volume classes (validation data). Definitions are the same as in Figure 2.](image)

The bias was small for volumes between 100 and 300 m$^3$/ha (Figure 6). Adding site index and age (Level 2) improved the estimations, both for low and high volumes, compared with only using spectral values and TS (Level 1). Also, adding tree height (Level 3) improved the estimations for low volumes compared with using spectral values, temperature sum, site index and age (Level 2).

![Figure 6. The bias for volume estimations using prediction difference presented for volume classes (validation data). Definitions are the same as in Figure 2.](image)
DISCUSSION

Standard errors in this study were comparable to what has been achieved in northern Sweden. Ståhl (1992) presents a standard error of 20-25% for volume estimations with a satellite method. The satellite method and material is the same as Hagner (1990) describes. Most of the forest in those studies has a volume between 100 and 300 m³/ha, corresponding to the middle class in Figure 4 and 5. Ståhl (1992) also investigated several subjective methods: Pure ocular estimation, a relascope method, a circular plot method, aerial photo interpretation, and aerial photo interpretation combined with field control. The satellite method yielded a high standard error compared with other methods. In this study, the standard error was about 12% when all ancillary data (temperature sum, site index, age and mean tree height) were added. This corresponds to the results from aerial photo interpretation combined with field control presented by Ståhl (1992).

Compared to northern Sweden, there is a greater proportion of dense forest in the study area. Wood volume is difficult to predict in dense forest. There is also a higher variation in altitude than found in most parts of the country. The effect of this is at least threefold: First, the topography causes differences in illumination in the image. Second, there is a high local variation in site quality due to steep and long slopes. Third, the overall variation in site quality is high due to the differences in altitude. The Landsat-5 TM sensor has operated since 1984, much longer than the planned 5 years nominal life (Kramer, 1986). Therefore, the performance of the optics in the sensor is doubtful. New sensors, which have a better spatial resolution, will soon be available (Konecny, 1996). The locations of the reference plots were determined by interpretation of maps by the survey team. It is likely that accuracy would increase if GPS had been available during the inventory. GPS has been used in the Swedish NFI since 1996.

One reason for high standard errors of low volumes when only using spectral data together with temperature sum, could be because canopy is not closed at this stage. Another reason could be the great variation of tree height in young forest. The improvements of the estimations when site index, age and tree height were added indicate this to be a correct assumption. Also, the growth functions used were not as good for young forest compared to older forest. One reason for high standard error in the upper volume class is the weak correlation between spectral value and wood volume in a closed canopy (Franklin, 1986). Weak correlation causes overestimation of low wood volumes and underestimation of high volumes. This is because the probability is higher that a reference plot with high volume is chosen when estimating low volumes and vice versa.

The topographic correction for differences in illumination did not have an effect on the wood volume estimation accuracy. There are at least two reasons; the number of extreme steep slopes in the validation data is not considerable and a simple model was used. If some extreme slopes were to be evaluated in this study, the correction might have indicated a significant improvement in the estimation accuracy. Itten et al. (1993) stated that a good DEM is needed and used a raster size identical to the TM data. In this study was the DEM resolution 50 m. Different forest structures reflect the light different for different sun slope normal angles. Therefore, it would be useful to take forest structure
into consideration. It would also have been desirable to model the atmosphere (e.g. Hill et al., 1994).

Kilkki and Pävinen (1987) stress the importance of the pixel value being independent of its location. It was assumed that a pixel on a high altitude would have a different spectral value than one on a low altitude because of difference in productivity. One reason for not improving the estimation accuracy when adding a climate index as a predicting variable could be that climate doesn’t vary much locally but all other predicting variables vary on a pixel or compartment level. It is perhaps more appropriate to use the TS for stratifying the material or weight the distance in feature space differently.

The fact that bias is lower if ancillary data are added indicates how important it is to include these to make the method more reliable for all volume classes. The distance function used proved important when variables other than spectral values were added that are differently correlated to the predicted variables. The method is less sensitive for nonsense variables if the prediction difference is used. However, there is seldom a simple linear relation between predicting and predicted variables. Furthermore, it could be useful to take the variance of predicting variables into consideration. The ancillary data can be collected in different ways, always choosing the most cost-efficient option. Variables that are known over time, such as site index and age, can be collected with field methods and stored in a database. Site index can also be estimated using the DEM together with a soil map (Holmgren, 1990). Other variables, that are more difficult to control over time, such as tree height, can be collected with airborne lidar (Aldred and Bonnor, 1985), microwave radar (Hyppä et al., 1992) or aerial photography (Åge, 1985).

The forest policy determines what inventory method that is suitable. The sizes of the compartments are determined by the intensity of forestry. The accuracy of volume estimations with the kNN method is better for large areas, as compared to small areas (Nilsson, 1997). The average size of the compartments in this study was 19 hectares. It is also evident that the kNN method performs differently in different kinds of forest. The results show that it is easiest to estimate volume in the middle volume class (100-300 m³/ha). Compartments in this volume class will usually soon be harvested (thinning or final felling) making reliable decision support information important. Continued inventory is of more importance in this volume class because of high growth-rate.

The kNN method might be used in different ways on a compartment level in forest management planning. First, it might be combined with a simple field control, where site index, age and tree height could be detected together with some other information for updating the database. Second, if the information is poor in the database, the estimations with the method could be used for finding forest that is to be harvested. Third, the method might be used for controlling the quality of information in the database. Another possibility is to estimate wood volume and basal area for larger areas than compartments, for example a small forest holding.
REFERENCES


Hägglund B., 1983, En ny svensk riksskogstaxering, Swedish University of Agricultural Sciences, Department of Forest Survey, the Swedish University of Agricultural Studies, Umeå.


Lindgren O., 1984, A study of circular plot sampling of Swedish forest compartments, Swedish University of Agricultural Sciences Section of Forest Mensuration and Management, Report no. 11.


Ståhl G., 1992, A study on the quality of compartment-wise forest data acquired by subjective inventory methods, Swedish University of Agricultural Sciences, Department of Forest Mensuration and Management, Report no. 24.


Serien Arbetsrapporter utges i första hand för institutionens eget behov av viss dokumentation. Rapporterna är indelade i följande grupper: Riksskogstaxeringen, Planering och inventering, Biometri, Fjärranalyser, Kompendier och undervisningsmaterial, Examensarbeten samt Internationellt. Författarna svarar själva för rapporternas vetenskapliga innehåll.

**Riksskogstaxeringen:**

1995 1  Kempe, G. Hjälpmedel för bestämning av slutheten i plant- och ungskog. ISRN SLU-SRG-AR--1--SE

2  Riksskogstaxeringen och Ståndortskarteringen vid regional miljöövervakning. - metoder för att förbättra upplösningen vid inventering i skogliga avrinningsområden. ISRN SLU-SRG-AR--2--SE.


24  Fridman, J. & Walheim, M. Död ved i Sverige. - Statistik från Riksskogstaxeringen. ISRN SLU-SRG-AR--24--SE.


34  Löfgren, P. Skogsmark, samt träd- och buskmark inom fjällområdet. En skattning av arealer enligt internationella ägoslagsdefinitioner. ISRN SLU-SRG-AR--34--SE.

37  Odell, G. & Ståhl, G. Vegetationsförändringar i svensk skogsmark mellan 1980- och 90-talet. -En studie grundad på Ståndortskarteringen. ISRN SLU-SRG-AR--37--SE.

38  Lind, T. Quantifying the area of edge zones in Swedish forest to assess the impact of nature conservation on timber yields. ISRN SLU-SRG-AR--38--SE.

**Planering och inventering:**


4  Ståhl, G. The Transect Relascope - An Instrument for the Quantification of Coarse Woody Debris. ISRN SLU-SRG-AR--4--SE.

1996 15  van Kerkvoorde, M. A sequential approach in mathematical programming to include spatial aspects of biodiversity in long range forest management planning. ISRN SLU-SRG-AR--15--SE.


Lämås, T. & Ståhl, G. Skattning av tillstånd och förändringar genom inventerings simulering - En handledning till programpaketet "NVSIM". ISRN SLU-SRG-AR--25--SE

Lämås, T. & Ståhl, G. Om dektektering av förändringar av populationer i begränsade områden. ISRN SLU-SRG-AR--26--SE

Biometri:


Fjärranalys:


1997 29. Hagner, O. Textur i flygbilder för skattning av beståndsegenskaper. ISRN SLU-SRG-AR--29--SE.


43 Wallerman, J. Brattåkerinventeringen. ISRN SLU-SRG-AR--43--SE.

47 Holmgren, J. Estimating Wood Volume and Basal Area in Forest Compartments by Combining Satellite Image Data with Field Data. Examens arbete i ämnet fjärranalys. ISRN SLU-SRG-AR--47--SE.

Kompendier och undervisningsmaterial:

1996 14 Holm, S. & Thuresson, T. samt jägm.studenter kurs 92/96. En analys av skogstillståndet samt några alternativa avverkningsberäkningar för en del av Östads säteri. ISRN SLU-SRG-AR--14--SE.

21 Holm, S. & Thuresson, T. samt jägm.studenter kurs 93/97. En analys av skogstillståndet samt några alternativa avverkningsberäkningar för en stor del av Östads säteri. ISRN SLU-SRG-AR--21--SE.

1998 42 Holm, S. & Lämås, T. samt jägm.studenter kurs 93/97. En analysis of the state of the forest and of some management alternatives for the Östad estate. ISRN SLU-SRG-AR--42--SE.

Examensarbeten:

1995 5 Törnquist, K. Ekologisk landskapsplanering i svensk skogsbruk - hur började det?. Examensarbete i ämnet skogsutskattning och skogsindelning. ISRN SLU-SRG-AR--5--SE.


8 Ranvald, C. Sortimentsinriktad avverkning. Examensarbete i ämnet skogsuppskattning och skogsindelning. ISRN SLU-SRG-AR--8--SE.


11 Djurberg, H. Den skogliga informationens roll i ett kundanpassat virkesflöde. - En bakgrundsstudie samt simulering av inventeringsmetoders inverkan på noggrannhet i leveransprognoser till sågverk. Examensarbete i ämnet skogsuppskattning och skogsindelning. ISRN SLU-SRG-AR--11--SE.

12 Bredberg, J. Skattning av ålder och andra beståndsvärder - en fallstudie baserad på MoDo:s indelningsrutiner. Examensarbete i ämnet skogsuppskattning och skogsindelning. ISRN SLU-SRG-AR--12--SE.

13 Gunnarsson, F. On the potential of Kriging for forest management planning. Examensarbete i ämnet skogsuppskattning och skogsindelning. ISRN SLU-SRG-AR--13--SE.

16 Tormalm, K. Implementering av FSC-certifiering för mindre enskilda markägares skogsbruk. Examensarbete i ämnet skogsuppskattning och skogsindelning. ISRN SLU-SRG-AR--16--SE.

1997 17 Engberg, M. Naturvärden i skog lämnad vid slutavverkning. - En inventering av upp till 35 år gamla förnyningsytor på Sundsvalls arbetsområde, SCA. Examensarbete i ämnet skogsuppskattning och skogsindelning. ISRN-SRG-AR--17--SE.

20 Cedervind, J. GPS under krontak i skog. Examensarbete i ämnet skogsuppskattning och skogsindelning. ISRN SLU-SRG-AR--20--SE.

27 Karlsson, A. En studie av tre inventeringsmetoder i slutavverkningsbestånd. Examensarbete. ISRN SLU-SRG-AR--27--SE.


33 Jonsson, Ö. Trädskikt och ståndortsförhållanden i strandskog. - En studie av tre bäckar
35 Claesson, S. Thinning response functions for single trees of Common oak (Quercus Robur L.) Examensarbete. ISRN SLU-SRG-AR--35--SE.

36 Lindskog, M. New legal minimum ages for final felling. Consequences and forest owner attitudes in the county of Västerbotten. Examensarbete. ISRN SLU-SRG-AR--36--SE.


41 Eriksson, F. Markbaserade sensorer för insamling av skogliga data - en förstudie. Examensarbete. ISRN SLU-SRG-AR--41--SE.

45 Gessler, C. Impedimentens potentiella betydelse för biologisk mångfald. - En studie av myr- och bergimpediment i ett skogslandskap i Västerbotten. Examensarbete ISRN SLU-SRG-AR--45--SE.

46 Gustafsson, K. Långsiktssplanering med geografiska hänsyn - en studie på Bräcke arbetsområde, SCA Forest and Timber. Examensarbete. ISRN SLU-SRG-AR--46--SE.

Internationellt

1998 39 Sandewall, M., Ohlsson, B & Sandewall, R.K. People’s options on forest land use. - a research study of land use dynamics and socio-economic conditions in a historical perspective in the Upper Nam Nan Water Catchment Area, Lao PDR. ISRN SLU-SRG-AR--39--SE.