Simultaneous use of Landsat-TM and IRS-1C WiFS data in estimating large area tree stem volume and aboveground biomass

Erkki Tomppo a,*, Mats Nilsson b,1, Mats Rosengren c,2, Paula Aalto d,3, Pamela Kennedy e,4

a The Finnish Forest Research Institute, Unioninkatu 40 A, FIN-00170 Helsinki, Finland
b Swedish University of Agricultural Sciences, SE-901 83 Umeå, Sweden
c Metria, Box 355, SE-101 27, Sweden
d Stora Enso Forest Consulting Oy Ltd., Talvikkitie C, FIN-01300 Vantaa, Finland
e Sector for Forest and Catchment Studies, Environment and Geo-Information (EGEO) Unit, Space Applications Institute, Joint Research Centre, Commission of the European Communities, I-21020, Ispra, Verona, Italy

Received 11 June 2001; received in revised form 26 February 2002; accepted 4 March 2002

Abstract

A multisource and multiresolution method was developed for estimating large area tree stem volume of growing stock and aboveground biomass of trees. Combined Landsat-TM data and IRS-1C WiFS data, together with field data of National Forest Inventories (NFIs), were applied. Landsat-TM data were used as an intermediate step between the field data and WiFS pixels. A nonparametric \( k \)-nearest neighbour (\( k \)-nn) estimation method was applied with Landsat-TM data and field plot data from the Swedish National Forest Inventory (SNFI). A nonlinear regression analysis was used in deriving models for volume and biomass as a function of WiFS data. The estimates were evaluated by applying independent estimates from the Finnish Multi-source National Forest Inventory (MS-FNFI): The estimates are derived using field plots from the Finnish National Forest Inventory (FNFI) and Landsat-TM images. Mean volume as estimated from the Finnish multisource data for a study area of 447 000 ha was 84.2 m\(^3\) ha\(^{-1}\). This compared with 87.2 m\(^3\) ha\(^{-1}\) as derived from the developed method presented in this paper. The corresponding estimates for aboveground tree biomass were 59.5 and 58.3 tons ha\(^{-1}\), respectively. © 2002 Elsevier Science Inc. All rights reserved.

1. Introduction

The needs for information in forestry are currently increasing because of the awareness of forest health status and the loss of biological diversity, as well as the recognition of the role of forests in reducing effects of global warming. At the same time, the pressure to increase timber consumption is increasing. Forest information is collected by forest inventories. The number of variables measured in the field is usually high, e.g., over 100. New variables are derived from further computations. One of the basic variables has been tree stem volume by tree species, and currently also the total tree biomass. The increasing availability of supplementary data, e.g., remote sensing data (satellite images), has made it possible to increase the efficiency of inventories and to satisfy the new information needs at reasonable costs (Franco-Lopez, Ek, & Bauer, 2001; Nilsson, 1997; Tomppo, 1990). Satellite image data are usually inexpensive but much less accurate than field measurements.

The Finnish Multi-source National Forest Inventory (MS-FNFI), conducted by the Finnish Forest Research Institute, has utilised satellite images and digital map data, in addition to field measurements, in an operative way since 1990. The nonparametric \( k \)-nearest neighbour (\( k \)-nn) method, which is applied, differs from the ordinary estimation and classification methods. An essential property of the method is that all inventory variables can be estimated at the same time. The method also preserves the covariance structure of the variables (Moeur, 1988; Tomppo, 1991). It can be applied in a straightforward way in different forest
conditions and with different remote sensing materials (Eisele, 1997; Franco-Lopez et al., 2001; Nilsson, 1997; Tomppo, Goulding, & Katila, 1999; Tomppo, Korhonen, Heikkinen, & Yli-Kojoja, 2001). Since the \( k \)-nn estimation method is nonparametric, it is important that the total variation of all forest variables in the study area be well represented by the field sample plots. A lack of or a low number of sample plots in certain forest types may lead to high errors.

Results show that \( k \)-nn-based estimates using high-resolution data have high root mean square error (RMSE) at a pixel level (e.g., Tokola, Pitkänen, Partinen, & Muinonen, 1996; Trotter, Dymond, & Goulding, 1997). The reliability increases with the area for which mean values are presented (Nilsson, 1997; Tokola & Heikkilä, 1997; Tomppo, 1996). A study in eastern Finland presented aggregated estimates of stem volume with an estimated standard error of 1.9% for an area of approximately 132,000 ha, and of 2.25% for a 4000-ha area (Muinonen & Tokola, 1990).

The \( k \)-nn method has also been used to estimate stem volume in mature pine (Pinus radiata) stands in New Zealand (Tomppo et al., 1999; Trotter et al., 1997). An RMSE of 46 m\(^3\) ha\(^{-1}\) was obtained for wood volume in mature pine stands of about 40 ha (Trotter et al., 1997). This was concluded to be operationally acceptable for forestry applications in New Zealand. It was also concluded that the \( k \)-nn method produced estimates with similar accuracy as obtained using regression models. In 1999, Tomppo et al. showed that estimates of wood volume in \( P. \) radiata stands could be improved by including additional information, such as age or year after thinning, as obtained from existing stand records. Adding the number of years after thinning lowered the RMSE by 27%. This study also showed the importance of using a sufficiently large number of sample plots in the \( k \)-nn method.

Studies in the Swedish and Finnish boreal forest have shown that high-resolution satellite data and field data from National Forest Inventory (NFI) plots can be used to construct regression models that predict forest variables (e.g., Hagner, 1990; Tomppo, 1987). The expected relative RMSE (i.e., RMSE multiplied by 100 and divided by the estimate) for stem volume per hectare, mean diameter, and age, on a stand level, using regression techniques is about 25%, 15%, and 20%, respectively (Hagner, 1990). This is slightly lower than the RMSEs reported for subjective (ocular) field inventories (Stähl, 1992).

Regression techniques have been used to calibrate low-resolution images such as NOAA-AVHRR with high-resolution data (e.g., Hämé, Salli, Andersson, & Lohi, 1996; Iverson, Cook, & Graham, 1994; Kleinn, Traub, & Dees, 1996; Pitkänen & Päivinen, 1992). In this study by Iverson et al. (1994), forest cover was estimated for two large regions in the eastern United States by combining NOAA-AVHRR and Landsat-TM images using multiple linear regression techniques. The forest cover could be successfully estimated in areas with fairly uniform topography and forest type. Another conclusion was that if there is considerable variation in topography and forest conditions, calibration areas for different ecological zones or forest types should be used.

This study aimed at developing methods for the co-use of high- and medium-resolution Earth Observation (EO) data to estimate the target variables tree stem volume and aboveground biomass of growing stock for forest land (FL) and for other wooded land (OWL) in Boreal forests. The purpose of the medium-resolution data was to test their potential for estimating total tree woody biomass and total tree stem volume over large areas, outside the area from which ground data were used for deriving models. Landsat-5 TM images were used as high-resolution data (HR) and an IRS-1C WiFS image as medium-resolution data (MR).

The field plot data from the Swedish National Forest Inventory (SNFI) were used as ground truth. It was not considered possible to use the SNFI plots directly connected to the 200-m pixels of the corrected IRS-1C WiFS images due to the small size of the ground plots and limited geometric accuracy of the WiFS data. Landsat-TM data based estimates were used as an intermediate step for building regression models for tree stem volume and biomass. The nonparametric \( k \)-nn method was applied with Landsat-TM data and the SNFI plots, and nonlinear regression analysis with IRS-1C WiFS data.

The models were constructed with data from the northern part of central Sweden. The final IRS-1C WiFS estimates were evaluated for an area in west central Finland. Neither volume nor biomass is directly measured in the NFIs in Europe. Both are derived variables. For the purpose of this study, volume is defined as over bark tree stem volume above stump height and measured in cubic metres per hectare. Biomass is defined as aboveground biomass in kilograms or 1000 kg per hectare (Köhli & Päivinen, 1996).

2. Material

The models were constructed for the Swedish site on FL and evaluated on the Finnish site on forestry land (FRYL). In this study, FRYL consists of FL, OWL, and wasteland, excluding forestry roads and depots. Forest land is defined as having a mean annual productivity over the rotation and under the most favourable tree species and growing stock structure of at least 1 m\(^3\) ha\(^{-1}\) yr\(^{-1}\). OWL and wasteland have a productivity of 0.1–1.0 m\(^3\) ha\(^{-1}\) yr\(^{-1}\) and <0.1 m\(^3\) ha\(^{-1}\) yr\(^{-1}\), respectively.

Six types of input data are required for the analysis. These are:

1. Georeferenced ground truth data.
2. Satellite imagery.
3. Digital land use/map data, based, e.g., on the existing digital base maps.
4. Digital terrain model, including information about sun illumination angle.
5. Digital boundaries of computation units, e.g., municipalities.
6. Cloud mask, if required by the meteorological conditions.

As a preprocessing step to the production of estimates, the locations of the ground data is combined with the image data to provide a database of pixel values at each of the plot locations.

2.1. The study areas

The 1.1-million-hectare modelling area is located in the northern part of Sweden (denoted here by Ssite) with coordinates: upper left 64°53’ 26", 18°05’ 10", lower right 63°27’ 50", 21°25’ 03” (Fig. 1). It consists mainly of forests that have been managed for more than one tree generation. About 74% of the land area of the Ssite is forested with the dominant tree species, in order of proportional representation, Scots pine (Pinus sylvestris L.), Norway spruce (Picea abies (L.) Karsten) 28.7%, and birch (Betula pendula Roth and Betula pubensens Ehrh.) 16.6%, and others (mainly Populus tremula L. and Alnus spp.) 3.2% (Table 1). The pine and birch proportions are thus higher and spruce proportions lower than on the Ssite. The areas are representative of the productive forest land of northern Sweden and central Finland.

2.2. Ground data

Three different ground data sets were used in this study:

1. Field plots from the SNFI. Volume and biomass estimates from these field plots were used as ground truth data for predicting volume and biomass from Landsat-TM data and is called here SNFI field data.

2. Field plots from an area of 5000 ha named Brattåker, located in the central part of the test site (64°30’ N, 18°30’ E). Volume and biomass estimates from these field plots were used to evaluate predictions of volume and biomass obtained from the model using Landsat-TM and SNFI field data.

3. Field plots from the Finnish National Forest Inventory (called FNFI) together with Landsat-TM data were used to obtain estimates for the validation of IRS-1C WiFS estimates.

2.2.1. SNFI data

Field data from the SNFI are used as modelling data when estimating aboveground woody biomass and above stump height stem volume. The SNFI is conducted as an annual systematic field sample of circular plots. Plots are aggregated into square-shaped clusters called tracts (Hägglund, 1983, Fig. 3a). The tracts are of two types:

1. Permanent tracts, remeasured within a period of 5–10 years with plot radius 10 m.
2. Temporary tracts, measured only once, plot radius 7 m.

All permanent plots used in this study were located using differential GPS techniques, resulting in a positional accuracy of 5–10 m. The plot location for temporary plots was determined by digitising the locations of the plot centres marked on field maps (scale 1:10 000) during the inventory. Altogether, 1487 SNFI plots located on forest land were available from the time period 1990–1994.

For each plot, digital numbers (DNs) from the TM image were read using cubic convolution interpolation. Since plots from a 5-year period are used, cuttings (thinning or regen-
oration felling) had occurred in the period between the inventory and the image acquisition at some locations. These plots and plots that were poorly matched with the image due to poor plot locations were removed from the data set using a statistical outlier removal method. In this procedure, the relationship between DN for each band and selected forest variables were modelled for each TM band (i) according to the following equation:

\[
\text{DN}_i = \beta_0 + \beta_1 \ln(\text{vol}_{\text{pine}}) + \beta_2 \ln(\text{vol}_{\text{spruce}})
+ \beta_3 \ln(\text{vol}_{\text{deciduous}}) + \beta_4 \ln(\text{age})
+ \beta_5 \ln(\text{vol}_{\text{total}} \times \text{age}) + \beta_6 \ln(m) + \beta_7 \ln(SI_{\text{pine}})
+ \beta_8 v_{\text{eg1}} + \beta_9 v_{\text{eg2}} + \beta_{10} v_{\text{eg3}} + \beta_{11} v_{\text{eg4}} + e \tag{1}
\]

where vol = wood volume, m\(^3\) ha\(^{-1}\); age = basal area weighted mean tree age, years; m = soil moisture class; SI = site index, m; veg1 = 1 if vegetation type 1, 0 if not; veg2 = 1 if vegetation type 2, 0 if not; veg3 = 1 if vegetation type 3, 0 if not; veg4 = 1 if vegetation type 4, 0 if not; \(e \sim N(0, \sigma^2)\) distributed error term.

The difference between the observed DN-vector for a plot and the prediction from Eq. (1) was used as a measure of how unusual the relationship was between the measured forest variables and satellite DNs. The following procedure was used. On the basis of the cutting intensity in the area and estimated location errors of the field plots, it was decided to remove 20% of the original plots. In this region, 2% of the forest land area is cut annually. The number of plots to be removed was also based on the fact that DN values for plots located close to a stand boundary are influenced by the adjacent stand or land use class. The reason for not removing more plots was the risk of not having enough field plots in all forest types. The plots were selected as follows. The standard deviation of each spectral band was calculated. Plots that had an absolute difference between estimated and observed spectral value > \(a \times \text{standard deviation of the band}\) in any of the bands were removed. The value of \(a\) was selected in such a way that 20% of the plots were removed.

All plot data were adjusted through growth models to the date of the image acquisition before the removal of non-representative plots was made. The resulting data set consisted of 1202 SNFI plots. Some statistics for the SNFI plots are given in Table 1.

Stem volume is defined as the volume in cubic metres per hectare (m\(^3\) ha\(^{-1}\)) of all trees on a plot, including bark but excluding branches and stump. Woody biomass is

<table>
<thead>
<tr>
<th>Ground data set</th>
<th>Stem volume (m(^3) ha(^{-1}))</th>
<th>Tree species proportion (% stem volume)</th>
<th>No. of plots</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
<td>Scots pine</td>
</tr>
<tr>
<td>Swedish NFI data</td>
<td>120.4</td>
<td>92.9</td>
<td>49.9</td>
</tr>
<tr>
<td>Finnish NFI data</td>
<td>93.0</td>
<td>80.2</td>
<td>51.9</td>
</tr>
</tbody>
</table>

Fig. 2. Location of the Finnish evaluation area and the 25 municipalities.
defined as total aboveground biomass including stem, bark, branches, needles/leaves, and stump. Volumes and biomass are predicted using models calibrated from the primary variables measured on ground.

2.2.2. FNFI data and multisource estimation data for validation

MS-FNFI estimates based on the field plots of the FNFI and Landsat-TM data from the year 1997 were used for the validation of multiresolution data-based estimates (Figs. 1 and 2). The reason for utilising satellite images in addition to field data on the Fsite was that the errors of field-data-based estimates are high for small areas like municipalities (Tomppo, 1991, 1996). The FNFI field plots are located either on temporary clusters (18 plots per cluster) or permanent clusters (one fourth of the clusters, 14 plots per cluster, Fig. 3b). The field plot is a Bitterlich sampling plot with a relascope factor of 2 and a maximum radius of 12.52 m. A total of 4035 field plots from FNFI was included in the area. The field plots outside the test area, and located on the area covered by the applied Landsat-TM image, were also utilised in the image analysis. The total number of plots used in calculating multisource estimates for municipality level was 6985.

The MS-FNFI also operationally produces estimates and statistics at the municipality level. Statistics from 25 municipalities were used (Fig. 2). The total forestry land area was 685 000 ha and the area utilised in the comparisons was 447 000 ha. The description of the applied field plot data is given in Table 1.

2.3. Satellite images

2.3.1. Landsat-TM data

A Landsat-TM scene, acquired on July 5, 1994 (scene 193/14), was used for the forest parameter estimation in the Ssite. The image was geometrically precision corrected to the Swedish National Grid (RT90) and resampled to 25 m × 25 m pixels using cubic convolution by SSC-Satellitbild.

A Landsat-TM image from June 13, 1997 (scene 119/16) was used for the Fsite. The image was also rectified to a pixel size of 25 m × 25 m to the national base map coordinate system.
2.3.2. IRS-1C WiFS data

2.3.2.1. General characteristics of medium-resolution data. Most of the current operational EO satellites could be categorised as high-spatial-resolution (<100 m pixel size) or low-spatial-resolution (>500 m pixel size) satellites. The interest in medium-resolution data (100 m–500 m pixels) has increased in the last few years. In practice, data from three different medium-resolution sensors are available for users in Europe today, RESURS MSU-SK (Research & Development Center ScanEx, 2000), IRS-1 WiFS (IRS-1C Satellite, 2002), and MODIS (Zhan et al., 2000). All these have a spatial resolution of approximately 180–250 m with scene sizes of 360 000–640 000 km². However, the radiometric resolution and calibration of these data are not as well developed as for Landsat-TM, SPOR-HRV, AVHRR, or SPOT-4 VEGETATION. IRS-1 WiFS data have a ground resolution of 188.3 m. It is two-band data, operating in the red (0.62–0.68 μm) and near-infrared areas (0.77–0.86 μm).

Medium-resolution data are of interest as a compromise between the limiting factors of the high-resolution and low-resolution data. To cover a continent, e.g., Europe, a very large number of Landsat or SPOT-type scenes are needed. Total cost and problems of obtaining a complete coverage due to clouds are the limiting factors. With low-resolution data, the repetition frequency is much higher but the low spatial and sometimes radiometric resolution limit the availability.

The wide field-of-view of the sensors (40–45°) gives a heterogeneous signal over the scene. The factors behind this are the varying solar elevation over the imaged area, variations in the atmospheric composition, variations in the relative sensor viewing/solar illumination geometry, which introduces varying effects caused by atmospheric scattering, and the variation of shadowing effects within the forest (BRDF effects). Atmospheric corrections and possibly BRDF corrections are thus required when considering full scene analysis of medium-resolution data.

2.3.2.2. IRS-1C WiFS data preprocessing. A system-corrected IRS-1C WiFS scene from 18 August 1997 over Sweden and Finland was used. The scene was not perfectly radiometrically corrected, as the image had a difference in grey levels corresponding to a step of 2 DN in band 1 (B1) and 5 DN in band 2 (B2) between the left and right arrays, each covering half the scene. This step was measured by averaging over a 2-pixel-wide vertical column covering the cloud free central area of the scene at both sides of the centre column. A correction was applied by adding these levels to the “lower” half of the scene.

The radiometric variation across the image was studied using the 6S radiative transfer software. The radiometric variations are visualised in Fig. 4, where the multiplicative and additive correction factors over the scene are shown. It illustrates the point that atmospheric effects and variations in illumination must be corrected in order to enable the application of the same estimation models over the full scene. Atmospheric correction was performed and the data converted into ‘at’ ground reflectance. The result was not intended to be a perfect absolute reflectance calibrated data set, but rather to remove radiometric angular variations over the scene in order to test the applicability for extrapolating data (calibrated for one small part of the scene) over the entire scene.

The resulting reflectance images were then geo-corrected against the geocoded TM scene using a second-order poly-
nominal model. The data were resampled to 200 m pixel size using cubic convolution resampling, aligned to the 25-m pixels of the TM data in order to be able to average 8×8 TM pixels corresponding to one WiFS pixel.

2.4. Other digital data

The area of the Ssite was separated into forest land and nonforest land using digital map data at a scale of 1:100000 produced by the Swedish National Land Survey. A forest land mask was created for the test area and the models using Landsat-TM data were used to predict volume and biomass only under this mask. For computing the biomass and volume estimates for the full WiFS scene, a 250-m raster version of a CORINE-like land-cover data set (available from the European Topic Centre, ETC-Land Cover) was used. This was compiled using classifications of Landsat-TM data acquired for Sweden, but was not a part of the original CORINE Land Cover exercise (CORINE Land Cover, 2001).

For the validation of the estimates on the Fsite, forestry land was separated from nonforest land using digital map data provided by National Land Survey of Finland. The errors in statistics caused by errors in maps were corrected as presented in Katila, Heikkinen, and Tomppo (2000). A digital terrain model and information about sun illumination angle were used in determining the angle between sun illumination and terrain normal. A function of the cosine of the angle was used in correcting spectral values of images (Tomppo, 1996). The computation units (municipalities) were delineated using digital boundary data provided by National Land Survey of Finland. The areas covered by clouds or shadows of clouds were delineated using a cloud mask (Tomppo, 1996).

3. Methods

3.1. The general design of the study

The volume and biomass models were first constructed using NFI field data and Landsat-TM data on the Ssite. A nonparametric k-nn estimation method was used to obtain the estimates for units corresponding to IRS-1C WiFS pixels. These estimates together with WiFS pixels were applied in constructing WiFS-based tree stem volume and biomass models. A nonlinear regression analysis was used in estimating the parameters of the models. The models were applied on the Fsite. Landsat-TM and the FNFI field plot data based estimates were computed for pixels and municipalities, using the MS-FNFI with k-nn estimation method, for evaluating these models. The link between the two data sets on the Ssite was performed within an area corresponding to approximately one half of a TM scene. The other half-scene fell over water or was not covered by digital map data.

Only WiFS pixels falling completely within the forest mask on the Ssite were used in deriving the regression models. This was ensured by extracting pixels centred in a 600×600-m zone completely covered by forest pixels on the 25-m map mask utilised. In this way, radiometric influences from nonforested areas, as well as erroneous biomass and volume estimates outside the forested area were reduced.

On the Fsite, the IRS-1C WiFS pixels falling completely on forestry land were used when applying the IRS-1C WiFS-based models. The estimates were computed as averages per hectare (m³ ha⁻¹ and 1000 kg ha⁻¹).

The following flow chart clarifies the steps undertaken.

(1) On Swedish site
   (a) Estimate plot-level volume and biomass using SNFI field data. Denote the method by SNFI method and the estimates by SNFI estimates.
   (b) Using estimates from 1a and the k-nn technique, volume and biomass were predicted for each Landsat-TM pixel. Estimates for each IRS-1C WiFS pixel were then obtained by summing TM pixel predictions for all TM pixels included in the IRS-1C WiFS pixel. Denote the Landsat-TM and SNFI field-plot-based method by TM-SNFI method and the estimates given by the method TM-SNFI estimates.
   (c) Estimate parameters of nonlinear regression models for predicting volume and biomass for IRS-1C WiFS pixels from IRS-1C WiFS reflectance values. The method and the estimates of these models are called WiFS-TM-SNFI method and WiFS-TM-SNFI estimates, respectively.
   (d) Evaluate models.

(2) On Finnish site
   (a) Using models from 1c developed for Swedish site, and Finnish IRS-1C WiFS spectral values, estimate volume and biomass for Finnish IRS-1C WiFS pixels.
   (b) Aggregate pixel-level volume and biomass estimates from 2a for Finnish municipalities.
   (c) Accumulate volume and biomass estimate across all sites to obtain total estimates.
   (d) Compare the municipality level and total area estimates from 2b with those obtained from the Finnish multisource inventory, i.e., those based on FNFI field data and Landsat-TM images from the Finnish site. Denote the method and estimates by MS-FNFI method and MS-FNFI estimates, respectively.

3.2. Volume and biomass functions

Both volume and aboveground biomass are estimated variables in NFIs, as stated previously. The biomass for Scots pine and Norway spruce was estimated using single tree models developed by Marklund (1988). For birch, Marklund’s models were used for stem and branches and the models presented by Kauppi, Tomppo, and Ferm (1995) for leaves. Biomass of stump aboveground was estimated with separate models. The ratio of stump biomass and total
stem biomass was assumed to be the same as the ratio of stump volume and the whole stem volume.

3.3. k-nn estimation

MS-FNIF and TM-SNFI estimates are calculated, respectively, from FNFI and SNFI field plots and Landsat-TM images using a nonparametric k-nn estimation method. Let us first recall the k-nn estimation method. It uses Euclidean distance, \( d_{p,p_i} \), defined in the image feature space. It is computed from pixel \( p \) to be analysed to each pixel \( p_i \), whose ground truth is known (to pixel with sample plot \( i \)). Data from the \( k \) plots, \( i_1(p), \ldots, i_k(p) \), with the shortest distances are utilised in the analysis of pixel \( p \). A maximum distance in the geographical space (usually 50 to 100 km in the horizontal direction) is set from the pixel \( p \) and from the geographical distances were used only on the Fsite. A maximum distance is also set in the vertical direction (in south and central Finland, e.g., usually 50 to 200 m) in order to take into account the vegetation variation caused by elevation variation, provided that a digital terrain model is available. In this study, the geographical distances were used only on the Fsite. The feasible set of nearest neighbours for a pixel \( p \) was \( \{ p_i | d_{p,p_i} \leq d_{\text{max}}^{(x,x)} , d_{y,y}^{(p,p_i)} \leq d_{\text{max}}^{(z,z)} \} \), where \( d_{p,p_i}^{(x,x)} \) is the geographical horizontal and \( d_{y,y}^{(p,p_i)} \) geographical vertical distance from pixel \( p \) to pixel \( p_i \) and \( d_{\text{max}}^{(x,x)} \) and \( d_{\text{max}}^{(z,z)} \) their maximum allowed values.

The weight of the ground data vector of plot \( i \) to pixel \( p \) is then defined by Eq. (2)

\[
w_{i,p} = \frac{1}{d_{p,p_i}^{2}} \sum_{j \in \{i_1(p), \ldots, i_k(p)\}} \frac{1}{d_{j,p}^{2}}, \quad \text{if } i \in \{i_1(p), \ldots, i_k(p)\}
\]

\[
w_{i,p} = 0, \quad \text{otherwise.} \quad (2)
\]

Volume and biomass estimates were written in the form of a digital map, Eq. (3)

\[
\tilde{m}_p = \sum_{j \in \{i_1(p), \ldots, i_k(p)\}} w_{i,p} m_j,
\]

where \( \tilde{m}_p \) is the multisource estimate of the value of variable \( M \) at pixel \( p \), and \( m_j \) is the measured value of variable \( M \) at field plot \( j \).

The land use classes outside forestry land are transferred directly from digital map file (see Tomppo, 1991, 1996).

The sums of weights, \( w_{i,p} \), were calculated by computation units (municipalities) in the estimation process. The weight of plot \( i \) to computation unit \( U \) is then (Eq. (4)):

\[
c_{i,U} = \sum_{p \in U} w_{i,p}.
\]

The mean volume (and biomass) estimates by computation units can be obtained with the formula

\[
v = \frac{\sum_{i \in I_u} c_{i,U} v_{i,U}}{\sum_{i \in I_u} c_{i,U}},
\]

where \( s \) is a computation strata (e.g., forest land), \( I_u \) the set of the sample plots in the strata, \( u \) a computation unit, \( c_{i,U} \), the weight of the sample plot \( i \) in the unit \( u \), and \( v_{i,U} \) the volume (biomass) per hectare of the growing stock on the sample plot \( i \).

The total volume and biomass estimates are obtained by replacing the denominator in formula (5) by 1.

The Landsat-TM image, SNFI field plots, and \( k \)-nn estimation method were used to compute volume and biomass estimates for areas corresponding to IRS-1C WiFS pixels. These estimates (TM-SNFI estimates) were used as ‘ground truth’ in deriving biomass and volume models as a function of the IRS-1C WiFS intensity values.

3.4. Regression models for IRS-1C WiFS-based estimates as function of multisource estimates

Nonlinear regression analyses were used to estimate the parameters of models for predicting volume and biomass for WiFS pixels. The dependent variables were the TM-SNFI estimates of tree stem volume (\( \text{m}^3 \text{ha}^{-1} \)) and aboveground tree biomass (1000 kg ha\(^{-1} \)), computed for the areas of 200\(\times\)200-m corresponding IRS-1C WiFS pixels, and explanatory variables the intensity values of IRS-1C WiFS pixels. Several explanatory variables (functions of IRS-1C WiFS bands) were tested. The modelling was, however, quite straightforward due to a limited number of IRS-1C WiFS bands, i.e., only two.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>NFI estimate</th>
<th>( k=5 )</th>
<th></th>
<th>( k=10 )</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean (( \text{m}^3 \text{ha}^{-1} ))</td>
<td>RMSE (%)</td>
<td>Bias (( \text{m}^3 \text{ha}^{-1} ))</td>
<td>Mean (( \text{m}^3 \text{ha}^{-1} ))</td>
</tr>
<tr>
<td>Stem volume (( \text{m}^3 \text{ha}^{-1} ))</td>
<td>100.4</td>
<td>120.5</td>
<td>61.2</td>
<td>0.1</td>
<td>120.3</td>
</tr>
<tr>
<td>Scots pine</td>
<td>60.0</td>
<td>60.1</td>
<td>103.8</td>
<td>0.1</td>
<td>60.1</td>
</tr>
<tr>
<td>Norway spruce</td>
<td>45.0</td>
<td>44.9</td>
<td>128.5</td>
<td>−0.1</td>
<td>44.9</td>
</tr>
<tr>
<td>Coniferous</td>
<td>105.0</td>
<td>105.0</td>
<td>65.9</td>
<td>0.0</td>
<td>105.0</td>
</tr>
<tr>
<td>Deciduous</td>
<td>15.4</td>
<td>15.5</td>
<td>181.9</td>
<td>0.1</td>
<td>15.3</td>
</tr>
<tr>
<td>Biomass (tons/ha)</td>
<td>70.7</td>
<td>67.2</td>
<td>56.3</td>
<td>−3.5</td>
<td>70.8</td>
</tr>
</tbody>
</table>

Table 2: Root mean square error (RMSE) and bias for forest parameters estimated using \( k \)-nn method and evaluated using cross-validation technique on a pixel level (Sweden)
The final model for both volume and biomass was:

\[
\text{Vol, Biom} = \exp(a_0 + a_1 b_1 + a_2 b_2 + a_3 b_1 b_2 + a_4 b_1 / b_2) + \epsilon
\]  

(6)

where \(b_1\) and \(b_2\) are the corrected WiFS data reflectances of band 1 and band 2, \(a_0, \ldots, a_4\) parameters to be estimated, and \(\epsilon\) an \(N(0, \sigma^2)\) distributed error term (cf. Tomppo, 1987).

4. Results

4.1. The evaluation of TM-based and WiFS-based estimates on the Swedish side

The TM-SNFI estimates of tree stem volume by tree species were evaluated at a pixel level using cross-validation technique and a pixel size of 25 m × 25 m. It was found that the estimation accuracy for stem volume per hectare was rather poor at the pixel level. This has been noticed in the earlier studies as well (Franco-Lopez et al., 2001; Katila & Tomppo, 2001; Nilsson, 1997; Tokola et al., 1996). RMSEs for all estimates at pixel level are smaller with \(k = 10\) than with \(k = 5\) nearest neighbours \((k)\) (Table 2). In a very recent study, the error could have been reduced noticeably with the help of a relocation of plots by means of multicriteria approach (Halme & Tomppo, 2001).

For areas of 4 ha, the relative RMSEs of 50% and 55% were obtained for stem volume and biomass, respectively (Nilsson and Sandström, 2001). The relative RMSE is \(100 \times \text{RMSE}/\bar{m}\). No statistically significant differences were noticed between the RMSEs of the estimates computed with \(k = 5\) and \(k = 10\) (\(F\) test, \(P = .05\)).

4.2. Volume and biomass models for IRS-1C WiFS data

The estimated values of the parameters of the volume model (6), together with asymptotic standard errors and 95% confidence intervals of parameters are given in Table 3 and those for the biomass model (6) in Table 4. The parameters were estimated with NLIN procedure of SAS (SAS Institute, 1989). The variables to be explained were the estimates of volume and biomass for the units corresponding IRS-1C WiFS pixels and estimated with SNFI plots and Landsat-TM data. The final explanatory variables were as given in model (6).

Table 3
Parameter estimates for nonlinear regression of volume from WiFS vs. TM estimates (Ssite)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Asymptotic S.E.</th>
<th>Asymptotic 95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lower</td>
</tr>
<tr>
<td>A0</td>
<td>6.771</td>
<td>0.1701</td>
<td>6.4380</td>
</tr>
<tr>
<td>A1</td>
<td>255.330</td>
<td>10.7900</td>
<td>234.1800</td>
</tr>
<tr>
<td>A2</td>
<td>−8.241</td>
<td>1.3380</td>
<td>−10.8640</td>
</tr>
<tr>
<td>A3</td>
<td>−1176.970</td>
<td>46.4600</td>
<td>−1268.0400</td>
</tr>
<tr>
<td>A4</td>
<td>−16.334</td>
<td>0.7315</td>
<td>−17.7680</td>
</tr>
</tbody>
</table>

Table 4
Parameter estimates for nonlinear regression of biomass from WiFS vs. TM estimates (Ssite)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Asymptotic S.E.</th>
<th>Asymptotic 95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lower</td>
</tr>
<tr>
<td>A0</td>
<td>5.8360</td>
<td>0.1653</td>
<td>5.5110</td>
</tr>
<tr>
<td>A1</td>
<td>224.6500</td>
<td>10.1200</td>
<td>204.8200</td>
</tr>
<tr>
<td>A2</td>
<td>−3.7570</td>
<td>1.2900</td>
<td>−6.2860</td>
</tr>
<tr>
<td>A3</td>
<td>−1088.6000</td>
<td>43.8400</td>
<td>−1174.5700</td>
</tr>
<tr>
<td>A4</td>
<td>−14.1270</td>
<td>0.6882</td>
<td>−15.4770</td>
</tr>
</tbody>
</table>

Fig. 5. TM-SNFI volume estimates against WiFS-TM-SNFI predicted values together with 95% confidence intervals of individual predictions (Ssite).

Fig. 6. TM-SNFI biomass estimates against WiFS-TM-SNFI predicted values together with 95% confidence intervals of individual predictions (Ssite).
The values predicted by the models are shown together with the 95% confidence intervals of the individual predictions in Fig. 5 for volume and in Fig. 6 for biomass. Note that the confidence intervals are for individual predictions and are thus wider than those for the expected values. The residuals against band 2 are shown for biomass in Fig. 7. The residuals for volume and against band 1 are similar. Both the volume model and biomass model explain 89% of the corresponding total variation of the volume or biomass.

4.3. Comparison of the Finnish multisource estimates with IRS-1C WiFS estimates

When applying the regression models (6) across the entire IRS-1C WiFS scene to derive volume and biomass estimates for an area of Finland encompassed in the scene, it was possible to compare the estimates (WiFS-TM-SNFI estimates) to MS-FNFI estimates (Fig. 8). The latter were computed for 25 municipalities (Fig. 2). The calculation of both volume and biomass on the Fsite was done using the operative Finnish MS-FNFI method (i.e., k-nn estimation and map error calibration, Katila et al., 2000; Tomppo, 1991). The total forestry land area of the 25 municipalities was 685,000 ha and the area used in the direct comparisons was 447,000 ha (Table 5). Note that the errors of statistics based only on field plot data are high due to the fact that the number of field plots on forestry land within one municipality varies from 25 to 200 on the Fsite. The standard errors of the estimates based on the field data on the whole area are, however, low; see below.

The evaluation areas were chosen in the following way. The output thematic map from MS-FNFI (with a pixel size of 25 m x 25 m) and the rectified IRS-1C WiFS image were overlaid. The subareas of 200 m² (corresponding IRS-1C WiFS pixels) which were totally on forestry land were

Fig. 7. Residuals of WiFS-TM-SNFI model in predicting TM-SNFI estimate of biomass against IRS-1C WiFS band 2 (Fsite).

Fig. 8. The estimate of above stem woody biomass over Sweden and Finland. The biomass increases from red to green. Blue is water.
selected. Table 5 shows the total forestry land areas and the forestry land areas available for comparisons by municipalities. The MS-FNFI based volume and biomass estimates of those areas were used in comparisons.

### 4.3.1. Mean and total volumes

The MS-FNFI estimates and WiFS-TM-SNFI estimates (model (6)) of mean volumes as well as their differences and relative differences are given in Table 5 and illustrated in Fig. 9. The difference (MS-FNFI–WiFS-TM-SNFI) varies from $-19$ to $26.5$ m$^3$ ha$^{-1}$ and relative differences from $-24.9\%$ to $28.5\%$. The mean difference is $-3.0$ m$^3$ ha$^{-1}$ ($-3.5\%)$. The absolute value of the mean difference is relatively low when taking into account that the modelling data are from another country, and to some extent from a different vegetation zone. The relative differences of the

<table>
<thead>
<tr>
<th>Municipality</th>
<th>Forestry land area (ha)</th>
<th>Area used in comparisons</th>
<th>MS-FNFI mean volume (m$^3$ ha$^{-1}$)</th>
<th>WiFS-TM-SNFI mean volume (m$^3$ ha$^{-1}$)</th>
<th>MS-FNFI–WiFS-TM-SNFI mean volume difference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alahärmä</td>
<td>22341</td>
<td>14544</td>
<td>77.5</td>
<td>69.0</td>
<td>$-8.3$</td>
</tr>
<tr>
<td>Alajärvi</td>
<td>57670</td>
<td>38960</td>
<td>81.9</td>
<td>76.1</td>
<td>$-5.0$</td>
</tr>
<tr>
<td>Evijärvi</td>
<td>27955</td>
<td>18380</td>
<td>88.1</td>
<td>78.5</td>
<td>$-9.6$</td>
</tr>
<tr>
<td>Ilmajoki</td>
<td>37782</td>
<td>25904</td>
<td>100.0</td>
<td>98.5</td>
<td>$-1.5$</td>
</tr>
<tr>
<td>Isoyöyri</td>
<td>21619</td>
<td>12892</td>
<td>82.1</td>
<td>73.5</td>
<td>$-8.6$</td>
</tr>
<tr>
<td>Kauhava</td>
<td>31319</td>
<td>20196</td>
<td>90.0</td>
<td>84.5</td>
<td>$-5.5$</td>
</tr>
<tr>
<td>Kortesjärvi</td>
<td>24694</td>
<td>16100</td>
<td>85.1</td>
<td>76.2</td>
<td>$-8.9$</td>
</tr>
<tr>
<td>Kuortane</td>
<td>34505</td>
<td>24360</td>
<td>98.1</td>
<td>96.4</td>
<td>$-1.7$</td>
</tr>
<tr>
<td>Kurikka</td>
<td>29735</td>
<td>18012</td>
<td>92.0</td>
<td>88.9</td>
<td>$-3.1$</td>
</tr>
<tr>
<td>Lahtiä</td>
<td>36907</td>
<td>26924</td>
<td>86.6</td>
<td>81.8</td>
<td>$-5.8$</td>
</tr>
<tr>
<td>Lappajärvi</td>
<td>32564</td>
<td>25564</td>
<td>89.9</td>
<td>83.5</td>
<td>$-6.4$</td>
</tr>
<tr>
<td>Lapua</td>
<td>48487</td>
<td>34100</td>
<td>96.3</td>
<td>93.1</td>
<td>$-3.2$</td>
</tr>
<tr>
<td>Lehtimäki</td>
<td>21598</td>
<td>12624</td>
<td>81.2</td>
<td>74.5</td>
<td>$-6.7$</td>
</tr>
<tr>
<td>Nurmo</td>
<td>26618</td>
<td>20084</td>
<td>85.5</td>
<td>81.5</td>
<td>$-1.7$</td>
</tr>
<tr>
<td>Oravainen</td>
<td>14820*</td>
<td>9232</td>
<td>85.5</td>
<td>75.2</td>
<td>$-1.0$</td>
</tr>
<tr>
<td>Peräseinäjoki</td>
<td>30879</td>
<td>17840</td>
<td>79.8</td>
<td>74.4</td>
<td>$-5.8$</td>
</tr>
<tr>
<td>Seinäjoki</td>
<td>7111</td>
<td>5196</td>
<td>92.3</td>
<td>89.4</td>
<td>$-4.9$</td>
</tr>
<tr>
<td>Teuva</td>
<td>40650</td>
<td>27756</td>
<td>89.2</td>
<td>88.0</td>
<td>$-3.2$</td>
</tr>
<tr>
<td>Töysä</td>
<td>22609</td>
<td>11836</td>
<td>100.8</td>
<td>95.8</td>
<td>$-8.4$</td>
</tr>
<tr>
<td>Vaasa</td>
<td>13690*</td>
<td>5748</td>
<td>109.1</td>
<td>115.2</td>
<td>$19.5$</td>
</tr>
<tr>
<td>Vimpeli</td>
<td>21798</td>
<td>14152</td>
<td>80.1</td>
<td>71.5</td>
<td>$-13.2$</td>
</tr>
<tr>
<td>Viitani</td>
<td>9572</td>
<td>4548</td>
<td>92.2</td>
<td>92.6</td>
<td>$26.5$</td>
</tr>
<tr>
<td>Vöyri</td>
<td>30890*</td>
<td>19984</td>
<td>93.6</td>
<td>88.7</td>
<td>$9.0$</td>
</tr>
<tr>
<td>Ylhärämä</td>
<td>8276</td>
<td>4040</td>
<td>81.6</td>
<td>65.5</td>
<td>$1.8$</td>
</tr>
<tr>
<td>Ylistaro</td>
<td>30718</td>
<td>17992</td>
<td>94.3</td>
<td>86.2</td>
<td>$4.3$</td>
</tr>
<tr>
<td>Sum*</td>
<td>684807</td>
<td>446968</td>
<td>89.3</td>
<td>84.2</td>
<td>$-3.0$</td>
</tr>
</tbody>
</table>

* Coastal Bothnian, Forestry Centre.

**Table 5** Comparisons of mean volume estimates for the Finnish MS-FNFI and model (6) (WiFS-TM-SNFI) in Southern Bothnian and Coastal Bothnian Forestry Centres area.

![Mean volume comparison](image-url)
total volume estimates are, of course, the same as those of the mean volumes for the areas used in the comparison.

An interesting question is how well the areas used in comparisons represent the total forestry land areas of the municipalities? The estimates for the total forestry land are given in Table 5 as well, column ‘mean volume for total forestry land.’ The differences of the MS-FNFI volume estimates on the total forestry land and WiFS-TM-SNFI estimates on the area used in comparisons are not on the average higher than those when the comparison areas are used also with MS-FNFI estimates. In fact, the average differences seem to be smaller (+2.1 m$^3$ ha$^{-1}$). A possible explanation for this may be the fact that the information of an IRS-1C WiFS pixel comes from a larger area than from that corresponding to the resampled pixel geographically.

4.3.2. Mean biomass

The mean biomass estimates are given in Table 6 and illustrated in Fig. 10. The MS-FNFI mean biomass estimate for the entire test area is 59.5 tons ha$^{-1}$ and the WiFS-TM-SNFI estimate 58.3 tons ha$^{-1}$. The estimates are very close to each other. The relative differences exceed 20% only in two municipalities (Vaasa and Vähäkyrö). The area used for testing is small in these municipalities. The absolute values of relative differences are in most cases less than 10% and in many cases less than 5%. The mean relative difference is 2%.

4.3.3. Comparison with field-data-based estimates

The estimates of total volume and biomass were computed for the whole Fsite using FNFI field plot data only. The standard volume estimation method was applied (Tomppo et al., 2001). The standard errors of the estimates were computed using the grouping method presented by Matérn (1960).

The estimate of the mean volume for forestry land was 91.1 m$^3$ ha$^{-1}$ and its standard error was 3.1 m$^3$ ha$^{-1}$. The WiFS-TM-SNFI estimate (87.2 m$^3$ ha$^{-1}$) is thus clearly within the double standard error of the field-data-based estimate and cannot be regarded as statistically significantly (a=.05) different from the FNFI estimate. The mean biomass estimate, based on FNFI field plots, is 64.4 tons ha$^{-1}$ and its standard error 1.8 tons ha$^{-1}$. The errors of the conversion factors when converting tree stem volume to aboveground tree biomass are not included in the above-mentioned standard error. The difference between the WiFS-TM-SNFI estimate (58.3 tons ha$^{-1}$) and the FNFI field estimate (64.4 tons ha$^{-1}$) is about three times the standard error of the field-data-based estimate. The difference may mainly be explained by the fact that different conversion factors from

<table>
<thead>
<tr>
<th>Municipality</th>
<th>Forestry land area (ha)</th>
<th>Area used in comparisons (ha)</th>
<th>MS-FNFI mean biomass (1000 kg ha$^{-1}$)</th>
<th>WiFS-TM-SNFI mean biomass (1000 kg ha$^{-1}$)</th>
<th>Mean biomass difference (1000 kg ha$^{-1}$)</th>
<th>Mean biomass difference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alahärmä</td>
<td>22341</td>
<td>14544</td>
<td>48.8</td>
<td>50.4</td>
<td>−1.5</td>
<td>−3.1</td>
</tr>
<tr>
<td>Alajärvi</td>
<td>57670</td>
<td>38960</td>
<td>54.0</td>
<td>63.5</td>
<td>−9.6</td>
<td>−17.7</td>
</tr>
<tr>
<td>Evijärvi</td>
<td>27955</td>
<td>18380</td>
<td>55.6</td>
<td>59.6</td>
<td>−4.1</td>
<td>−7.4</td>
</tr>
<tr>
<td>Ilmajoki</td>
<td>37782</td>
<td>25904</td>
<td>69.4</td>
<td>64.6</td>
<td>4.9</td>
<td>7.0</td>
</tr>
<tr>
<td>Isokyrö</td>
<td>21619</td>
<td>12892</td>
<td>51.8</td>
<td>48.2</td>
<td>3.6</td>
<td>7.0</td>
</tr>
<tr>
<td>Kauhava</td>
<td>31319</td>
<td>20196</td>
<td>59.8</td>
<td>56.4</td>
<td>3.4</td>
<td>5.7</td>
</tr>
<tr>
<td>Kortesjärvi</td>
<td>24694</td>
<td>16100</td>
<td>54.0</td>
<td>56.9</td>
<td>−2.9</td>
<td>−5.3</td>
</tr>
<tr>
<td>Kuortane</td>
<td>34505</td>
<td>24360</td>
<td>68.1</td>
<td>65.7</td>
<td>2.4</td>
<td>3.6</td>
</tr>
<tr>
<td>Kurikka</td>
<td>29735</td>
<td>18012</td>
<td>62.6</td>
<td>64.5</td>
<td>−1.9</td>
<td>−3.0</td>
</tr>
<tr>
<td>Laihia</td>
<td>36907</td>
<td>26924</td>
<td>57.6</td>
<td>53.0</td>
<td>4.7</td>
<td>8.1</td>
</tr>
<tr>
<td>Lappajärvi</td>
<td>32564</td>
<td>25564</td>
<td>59.2</td>
<td>60.7</td>
<td>−1.6</td>
<td>−2.7</td>
</tr>
<tr>
<td>Lapua</td>
<td>48487</td>
<td>34100</td>
<td>65.8</td>
<td>57.0</td>
<td>8.8</td>
<td>13.4</td>
</tr>
<tr>
<td>Lehtimäki</td>
<td>21598</td>
<td>12624</td>
<td>52.8</td>
<td>59.5</td>
<td>−6.7</td>
<td>−12.6</td>
</tr>
<tr>
<td>Nurmo</td>
<td>26618</td>
<td>20084</td>
<td>57.7</td>
<td>55.9</td>
<td>1.8</td>
<td>3.0</td>
</tr>
<tr>
<td>Oravainen</td>
<td>14820a</td>
<td>9232</td>
<td>53.0</td>
<td>50.4</td>
<td>2.6</td>
<td>4.9</td>
</tr>
<tr>
<td>Peräseinäjoki</td>
<td>30879</td>
<td>17840</td>
<td>52.8</td>
<td>53.9</td>
<td>−1.1</td>
<td>−2.1</td>
</tr>
<tr>
<td>Seinäjoki</td>
<td>7111</td>
<td>5196</td>
<td>63.2</td>
<td>62.5</td>
<td>0.7</td>
<td>1.1</td>
</tr>
<tr>
<td>Teuva</td>
<td>40650</td>
<td>27756</td>
<td>61.9</td>
<td>60.8</td>
<td>1.1</td>
<td>1.8</td>
</tr>
<tr>
<td>Töyssä</td>
<td>22609</td>
<td>11836</td>
<td>67.8</td>
<td>68.3</td>
<td>−0.5</td>
<td>−0.7</td>
</tr>
<tr>
<td>Vaasa</td>
<td>13690a</td>
<td>5748</td>
<td>80.7</td>
<td>62.8</td>
<td>17.9</td>
<td>22.2</td>
</tr>
<tr>
<td>Vimpeli</td>
<td>21798</td>
<td>14152</td>
<td>50.7</td>
<td>57.2</td>
<td>−6.5</td>
<td>−12.9</td>
</tr>
<tr>
<td>Vähäkyrö</td>
<td>9572</td>
<td>4548</td>
<td>65.0</td>
<td>44.2</td>
<td>20.9</td>
<td>32.1</td>
</tr>
<tr>
<td>Vöyri</td>
<td>30890a</td>
<td>19984</td>
<td>62.4</td>
<td>53.6</td>
<td>8.8</td>
<td>14.2</td>
</tr>
<tr>
<td>Ylihärmä</td>
<td>8276</td>
<td>4040</td>
<td>46.3</td>
<td>43.1</td>
<td>3.2</td>
<td>6.9</td>
</tr>
<tr>
<td>Ylistaro</td>
<td>30718</td>
<td>17992</td>
<td>60.8</td>
<td>55.2</td>
<td>5.6</td>
<td>9.2</td>
</tr>
<tr>
<td>Sum</td>
<td>684807</td>
<td>446968</td>
<td>59.5</td>
<td>58.3</td>
<td>1.2</td>
<td>2.0</td>
</tr>
</tbody>
</table>

* Coastal Bothnian, Forestry Centre.
stump volume to biomass have been applied on the Ssite and the Fsite. The conversion factors from above stump stem volume (cubic meter) to aboveground biomass (1000 kg) on the Fsite were: pine 0.714, spruce 0.684, and broad-leaved trees 0.720 (cf. Kauppi et al., 1995) while Marklund’s (1988) models were applied on the Ssite. Note also that the real standard error is higher than that given above 1.8 tons ha\(^{-1}\).

4.3.4. Factors affecting the errors

One of the largest differences between the Ssite (modeling site) and Fsite (evaluation site) is the much higher peatland proportion of Fsite. Peatland is typically wet with a low volume of growing stock and their reflectance is, however, similar to forests with a high volume of growing stock. The correlation coefficients between municipality

Table 7
Correlation coefficients between peatland % on forest land and on other wooded land with the difference of the volume estimates and with the difference of the biomass estimates. The significance levels are also given.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Peatland % on forest land</th>
<th>Peatland % on other wooded land</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correlation coefficient</td>
<td>Significance level</td>
</tr>
<tr>
<td>Volume (m(^3) ha(^{-1}))</td>
<td>.683</td>
<td>.0002</td>
</tr>
<tr>
<td>Volume (%)</td>
<td>.654</td>
<td>.0004</td>
</tr>
<tr>
<td>Biomass (1000 kg ha(^{-1}))</td>
<td>.694</td>
<td>.0001</td>
</tr>
<tr>
<td>Biomass (%)</td>
<td>.660</td>
<td>.0003</td>
</tr>
</tbody>
</table>

level errors of the estimates and peatland proportion on the forest land and on other wooded land were calculated. The correlation coefficients were calculated between the variable ‘peatland proportion of forest land’ and the two variables ‘(WiFS-TM-SNFI volume estimate−MS-FNFI volume estimate)’ and ‘100×(WiFS-TM-SNFI volume estimate−MS-FNFI volume estimate)/(MS-FNFI volume estimate)’ (Table 7). The similar correlation coefficients were computed when ‘peatland proportion on other wooded’ instead of ‘peatland proportion of forest land’ and biomass instead of volume were used. The coefficients computed with forest land variables varied from −0.65 (relative difference in mean volume, %) to −0.69 (difference in biomass, 1000 kg ha\(^{-1}\)) (Table 7). The correlation coefficients between the peatland proportion of the other wooded land and the differences of two types of estimates were almost as high. This means that the WiFS-TM-SNFI method overestimates volume and biomass and the overestimates are greater with higher peatland proportions. This was as expected, i.e., high value of the volume of growing stock are estimated for wet areas with low volume of growing stock.

The peatland proportion increases on the evaluation area from west to east. The trend in errors can be seen clearly (Fig. 11). The WiFS-TM-SNFI method seem to overestimate both volume and biomass in the eastern parts of the evaluation area and also slightly in the southern part of the area and underestimate in the western part of the region.

Another reason for the clear spatial dependencies, in addition to the variation in the percentage of peatland may be the artefacts of bidirectional reflectance still present in

Fig. 10. Mean biomass (1000 kg ha\(^{-1}\)) comparisons between the MS-FNFI estimates and WiFS-TM-SNFI estimates in Southern Bothnian and Coastal Bothnian Forestry Centres area.
the WiFS data. This phenomenon must be taken into account when deriving models for possible future applications.

5. Discussion

Our results show that models for predicting volume and biomass from Landsat-TM imagery may be used as an intermediate step between ground measurements and 200-m-resolution data corresponding IRS-1C WiFS pixels. In addition, models calibrated using the 200-m-resolution data may be used to extrapolate biomass and volume predictions from WiFS imagery to larger areas than that covered by the ground data. The results are even more promising considering that the modelling data and testing data are from different type of forests and vegetation zones.

The wide field of view of the medium-resolution sensors is a main source of difficulties when processing the full scene for biomass and volume estimation. The key to success is an adequate radiometric (view angle effects) and atmospheric correction of the medium-resolution data. This is particularly important for IRS-1C WiFS data due to the two optics and two CCD cameras. The main effects are in the east–west directions of the scenes, where possibly a geographic limitation for extrapolation could be expected. This limitation might be in the order of 400–500 km in the east–west direction but possibly a larger distance in the north–south direction (along the satellite track). With the support of ground data from two separate regions of an IRS-1C WiFS scene (800×800 km), this problem could be overcome.

The high-resolution data from Landsat-TM and SPOT have been shown to be of use for the estimation of biomass and volume in the boreal region. They are operationally used by the MS-FNFI in Finland (Tomppo, 1996) and are candidates for the next NFI in Sweden (Nilsson, 1997).

Among the potential factors limiting the accuracy of remote-sensing-based methods are:

1. Geometric accuracy of ground plot coordinates.
2. Geometric accuracy of satellite data.
3. Temporal proximity between satellite and ground measurements.
4. Radiometric quality of the data including atmosphere.
5. Topographic effects from shadows, etc.
6. Possible saturation of the signal at moderate biomass levels.
7. The range of biomass in the ground data; small biomass levels must also be included in the regression.
8. Effects from the ground vegetation and soil background.
A method to remove the effects of the factors (1) and (2) is presented by Halme and Tomppo (2001). For an operational scenario, the limiting factors are mainly the need for a certain amount of ground data. For example, the k-nn method with a full Landsat-TM scene requires between at least 1000, preferably 5000, georeferenced field plots.

5.1. Benefits and limitations of medium-resolution data in continent and subcontinent multiphase forest inventory system

When considering the utilisation of remote-sensing-based methods for tree stem volume and/or tree biomass estimation at a continental level, one should identify for what purposes the methods will be used. Many countries already have national- and regional-level inventories producing quite accurate estimates, at least for larger areas. However, these inventories often utilise different parameter definitions, or the frequency of the inventories is not high enough, or they may in other ways be focussed on specific applications that make it difficult to use the information in other contexts. Therefore, the benefits from medium-resolution remote-sensing data might be:

- Possibilities of frequent coverage repetition (e.g., annually).
- Easy to cover larger areas.
- Possibility to extrapolate estimates to areas with no other data.
- Low price per area covered.

The first tests carried out in this study showed promising results but the applicability must be further evaluated before any decisions about the operational utility medium-resolution data, e.g., WiFS, MODIS, or MSU-SK.

In addition to possible accuracy problems, there are also some other questions that should be solved.

(1) The topographic shadowing effects in mountainous areas are a specific problem, which are difficult to solve. The use of VHR data and texture-based methods is a possible solution for specialised treatment of these regions.

(2) Availability of forest masks. In this project, training areas were selected from forest land. Forestry land mask was available for the evaluation area. Digital forest or forestry land masks are available for some countries. In the case that it is not available, remote-sensing-based methods could be utilised to produce such masks (Tomppo et al., 2001).

(3) Availability of training data. There are at least four different alternatives:

(a) Landsat-TM and NFI field-plot-based estimates were applied in this project. The same method could, in principle, also be repeated elsewhere. This presumes that field plots are georeferenced to an acceptable accuracy. High-resolution or very-high-resolution data is necessary because field plots very often are too small to serve directly as training data sets for medium-resolution EO data.

(b) In some cases, stand level or compartment level up-to-date data are available that cover large enough forest patches and can then be used as training data. Examples are management inventory data.

(c) A relevant approach could be the following. Up-to-date field plots, which are available in most countries, will be used together with air photos. Training areas are interpreted from air photos with the support of known field data. A few hundred to a thousand pixels per IRS-1C WiFS scene might be enough for most scenes. These training data pixels would be used in digital analysis of the IRS-1C WiFS image.

(d) The fourth possibility is to FIRS measure a separate training data set. The cost/benefit of this approach depends very much on the application of the system and whether the measured data can be utilised for other purposes as well.

(4) Cost of the system. The total cost to cover the whole of Europe, e.g., has been estimated to about 5 million euros in FMERS-II (1999) presuming that field data are available and only computer processing is needed, see also Kennedy et al. (2000).

The most unreliable part in the cost calculation is the price of the training data, whose costs depend on the plot design, their distribution and the number of parameters to be gathered. A considerable amount of training data could possibly be made available through the NFIs.

References


