

L- and P-band backscatter intensity for biomass retrieval in hemiboreal forest

Johan Fransson


Remote Sensing of Environment

Cite this paper

Downloaded from [Academia.edu](#) 

[Get the citation in MLA, APA, or Chicago styles](#)

Related papers

[Download a PDF Pack](#) of the best related papers 



[Airborne S-Band SAR for Forest Biophysical Retrieval in Temperate Mixed Forests of the UK](#)

Heiko Balzter

[Impact of spatial variability of tropical forest structure on radar estimation of aboveground biomass](#)

Robin Chazdon

[Accuracy of biomass and structure estimates from radar and lidar](#)

Razi Ahmed



L- and P-band backscatter intensity for biomass retrieval in hemiboreal forest

G. Sandberg^{a,*}, L.M.H. Ulander^{a,b}, J.E.S. Fransson^c, J. Holmgren^c, T. Le Toan^d

^a Chalmers University of Technology, Göteborg, Sweden

^b Swedish Defence Research Agency, Linköping, Sweden

^c Swedish University of Agricultural Sciences, Umeå, Sweden

^d Centre d'Etudes Spatiales de la Biosphère, Toulouse, France

ARTICLE INFO

Article history:

Received 26 March 2009

Received in revised form 8 February 2010

Accepted 12 March 2010

Available online 13 May 2011

Keywords:

SAR

Forestry

Biomass estimation

Linear regression

Model selection

Polarimetry

P-band

L-band

ABSTRACT

At present, the greatest source of uncertainty in the global carbon cycle is in the terrestrial ecosystems. In order to reduce these uncertainties it is necessary to provide consistent and accurate global estimates of the world forest biomass. One of the most promising methods for obtaining such estimates is through polarimetric SAR backscatter measurements at low frequencies. In this paper, the relation between polarimetric SAR backscatter at L- and P-bands and forest biomass is investigated using data acquired within the BioSAR-I campaign in southern Sweden during 2007. Methods for estimating biomass on stand level using these data are developed and evaluated, and the results for the two frequency bands are compared. For L-band data, the best results were obtained using HV-polarized backscatter only, giving estimation errors in terms of root mean square errors (RMSE) between 31% and 46% of the mean biomass for stands with biomass ranging from 10 to 290 t/ha, and an (adjusted) coefficient of determination (R^2) between 0.4 and 0.6. For P-band data, the results are better than for L-band. Models using HV- or HH-polarized P-band backscatter give similar results, as does a model including both HV and HH. The RMSEs were between 18 and 27%, and the R^2 values were between 0.7 and 0.8.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

In view of the continued threat of global warming, the need of understanding and modeling the climate is becoming more urgent. It is of special importance to understand the global carbon cycle, and therefore accurate and consistent measurements of carbon in its various forms are essential. At present, the largest single source of uncertainty in the global carbon budget lies in the terrestrial ecosystems (IPPC, 2007; Le Toan et al., this issue). In order to reduce this uncertainty, it is necessary to improve the estimates of the distribution and changes in forest biomass on a global scale. One of the most promising methods for accurate measurements of these essential parameters is by means of L- or P-band Synthetic Aperture Radar (SAR). Radar sensors using L-band frequencies have already been successfully placed into orbit, most recently through the Phased Array L-band type Synthetic Aperture Radar (PALSAR) onboard the Japanese Advanced Land Observing Satellite (ALOS) (Rosenqvist et al., 2007; Shimada et al., 2009). The proposed ESA Earth Explorer Mission BIOMASS is the prime candidate to be the first P-band SAR satellite (Bezy et al., 2007).

Several studies on the relationship between biomass and L- and P-band backscatter have been published, e.g. (Le Toan et al., 1992; Baker et al., 1994; Beaudoin et al., 1994; Israelsson et al., 1994; Ranson & Sun,

1994; Rauste et al., 1994; Imhoff, 1995; Rignot et al., 1995; Fransson & Israelsson, 1999; Hoekman & Quiriones, 2000; Saatchi et al., 2007). In all of these studies, HV- and HH polarized backscatter have been found to be dependent on biomass. The studies also agree that P-band backscatter shows stronger dependence on biomass than L-band backscatter. However, while the models for biomass retrieval presented in these papers are similar, there is no consensus on which model to use for biomass estimation. Moreover, the number of test sites and measurement campaigns for which the biomass-backscatter relation has been analyzed is limited. Thus, there is a continued need of both new experimental data and a further improvement of the existing models for biomass estimation from SAR data.

In this paper, L- and P-band SAR backscatter data are compared with respect to their usefulness for estimation of forest biomass. A method for biomass estimation based on regression analysis is developed and evaluated for both frequency bands. The SAR data were acquired within the BioSAR-I experiment conducted between March and May 2007 (Hajnsek et al., 2008). *In-situ* measurements and high density laser scanning data are used to obtain reference biomass estimates at stand level. An analysis of the temporal stability of the backscatter and estimation methods is also performed. Similar temporal analysis has previously been made for L-band data by e.g. Fransson and Israelsson (1999).

The paper is organized as follows. In Section 2, the experiment and test site are described, followed by a description of the available field, laser scanning and SAR data (Section 3). In Section 4, the processing of

* Corresponding author.

E-mail addresses: gustaf.sandberg@chalmers.se, gustaf.sandberg@gmail.com (G. Sandberg).

field and laser scanning data into stand level estimates of biomass is presented, including error analyzes. The extraction of backscatter estimates at stand level as well as errors in these estimates is also described. The methodology used to estimate biomass from the SAR data is described in Section 5. Then, an analysis of the temporal stability of the backscatter is presented (Section 6), followed by a qualitative analysis of the backscatter–biomass relation based on plots of backscatter versus biomass at stand level (Section 7). The key results, the evaluation of biomass retrieval using L- and P-band SAR backscatter, are presented in Section 8. Finally, the results are summarized and conclusions are drawn.

2. Experiment and test site

2.1. The BioSAR-I experiment

During the spring of 2007 the BioSAR-I experiment was conducted in Sweden. The aim of the experiment was to provide new data for the ongoing studies related to the proposed ESA Earth Explorer Mission BIOMASS. L- and P-band SAR data were collected by the airborne Experimental-SAR (E-SAR) system, owned and operated by the German Aerospace Center (DLR). Data were collected on three separate occasions over a period of about two months, allowing studies of temporal stability of backscatter measurements over time periods comparable to the repeat-pass time between image acquisitions for a radar satellite. *In-situ* data on forest properties and soil moisture, weather data and high resolution helicopter borne laser scanning data were also collected. Further details on the experiment setup and initial evaluation can be found in Hajnsek et al. (2008). The collected data (field, laser scanning and SAR data) and processing thereof are further described in Sections 3 and 4, respectively.

2.2. Test site

The test site, Remningstorp, is located in the south of Sweden (58° 30' N, 13° 40' E), and covers about 1200 ha of productive forest land (Fig. 1). The forest is classified as hemiboreal (Ahti et al., 1968). The dominant tree species are Norway spruce (*Picea abies* (L.) Karst.), Scots pine (*Pinus sylvestris* L.) and birch (*Betula* spp.). The dominant soil type

is till with a field layer consisting of different herbs, blueberry (*Vaccinium myrtillus* L.) and narrow-leaved grass (*Deschampsia flexuosa* (L.) Trin.). In denser old spruce stands the field layer is absent. The area is fairly flat, with ground elevations between 120 and 145 m above sea level.

3. Data description

3.1. Field data

Two field data sets were used in this paper. The first data set was collected during 2004 and 2005, and consisted of 849 field plots (radius 10 m), distributed in a systematic grid with a spacing of about 40 m. For these plots, field data were collected using the routines developed for the Forest Management Planning Package (FMPP) as described in Jonsson et al. (1993). The field measurements include calliper of all trees on each plot with diameter at breast height (i.e. 1.3 m above ground level) greater than 0.05 m and measurements of tree height for a subset (about 10%) of the trees. The field plots were too small to be useful in the backscatter–biomass analysis (the field plots are only 0.03 ha, i.e. smaller than the 0.1 ha areas discarded later in the paper, see further details in Section 4.4). Instead, they were used in combination with laser scanning data to estimate stem volume and above-ground dry biomass for 58 stands with sizes between 0.5 and 9.4 ha. These forest stands were used to develop a regression model describing the relationship between biomass and SAR backscatter, and will be referred to as training stands throughout the paper. Key variables for the 58 training stands are summarized in Table 3.

The second data set consisted of field measurements made during 2006 and 2007. For 17 stands measurements were made for every single tree with a diameter at breast height greater than 0.05 m. Due to a storm event in January 2007, stands inventoried during 2006 were revisited at the end of 2007 to assess storm damages (all trees were checked). The size of the stands were either 0.1 ha (7 stands) or 0.7 ha (10 stands). The measurements on tree level include position, diameter at breast height, species and height. Tree height was measured for all trees in the stands that were inventoried in 2007 (11 stands) and on a subset of the trees for the rest of the stands. Allometric equations based on measurements of diameter and height were used to obtain tree level estimates of stem volume (including bark, but excluding stump and roots) and biomass (including bark, branches and leaves/needles, but excluding stump and roots) (Näslund, 1947; Marklund, 1988). Separate equations for stem and branch (including leaves/needles) biomass were used. Species-dependent equations are available for pine, spruce and birch, and for any other deciduous species the same equations as for birch was used. For trees without height measurements secondary regression functions based solely on diameter were used. The secondary regression functions were developed using data from trees for which height measurements had been made. The tree level estimates of biomass and stem volume were then aggregated to stand level. As with the field plots in the first data set, the 0.1 ha large stands were found to be too small to be useful in the regression analysis relating biomass to SAR backscatter and were therefore discarded (see further details in Section 4.4). The remaining 10 forest stands (0.7 ha large) were used to validate the developed biomass regression models and will be referred to as validation stands throughout the paper. Key variables for the 10 validation stands are summarized in Table 1. The extent of the stands are shown in Fig. 2.

None of the stands are characterized by large ground slopes, although some within stand topographic variations is present. Field visits were made in connection to all image acquisitions, and overall weather and moisture conditions were recorded. No strong winds or precipitation were observed in Remningstorp on any of the image acquisition dates. Measurements of soil moisture were available from two sensors, one



Fig. 1. The location of Remningstorp test site, shown on a map of the Nordic countries.

Table 1

Summary of key variables for the 10 validation stands. Note that the errors associated with the biomass estimates are limited to a few percent of the biomass. Six of the stands are dominated by spruce, two stands are dominated by pine and the remaining two stands are dominated by birch.

Stand variable	Mean	Min	Max
Area [ha]	0.68	0.66	0.69
Stem volume [m ³ /ha]	347	70	527
Biomass [t/ha]	181	52	267
Tree no. density	562	412	815
Maximum tree height [m]	28	20	36

located inside a forest and one located in a field. The volumetric moisture content inside the forest were about 22%, 16–17% and 14% on 9 March, 31 March to 2 April and 2 May 2007, respectively. Outside the forest the measured differences between the acquisition dates were larger and on 9 March 2007 the measurement device was close to being saturated. It is important to note that local variations in moisture conditions exist.

3.2. Laser scanning data and digital aerial images

The laser scanning was performed on 24 April 2007 at an altitude of 130 m above ground level using the laser scanning system TopEye S/N 425 mounted on a SE-JTF helicopter. The laser pulse density was on average 30 pulses per square meter. The main area covered by laser scanning data is 4 km by 1 km large in northing and easting directions, respectively. In addition, some smaller areas to the east and west of the main area were covered in order to have a complete coverage over the entire area that was field inventoried (Fig. 2). With support from the 849

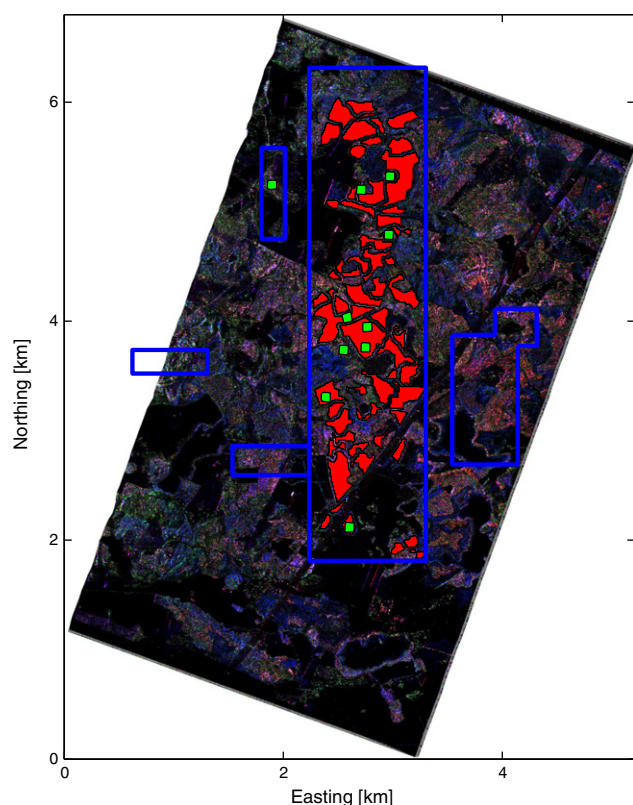


Fig. 2. The image is an RGB composite (HH = red, HV = green, and VV = blue) of a P-band image acquired on 2 May 2007. Overlaid is the extent of training (red) and validation (green) stands. The blue lines show the approximate area covered by laser scanning data. Note that only laser scanning data from the main area was used for stem volume and biomass estimation.

Table 2

Summary of available E-SAR images. The number of images corresponds to the number of fully polarimetric images. Note that one of the two available L-band images from 9 March 2007 was discarded due to noted system warmup effects.

Acquisition date	Number of L-band images	Number of P-band images
9 March 2007	1 (2)	2
31 March 2007	2	0
2 April 2007	0	2
2 May 2007	2	2

field plots, laser scanning data from the main area were processed into a stem volume map, which was then used to obtain biomass estimates for the 58 training stands. Digital near infrared images from the airborne Z/I DMC camera were also available and used to estimate stem volume proportions for different tree species classes at stand level. The laser scanning data and near infrared images were available in the Swedish National Grid (RT90) coordinate system with height in the RH70 system, but were projected into the Universal Transverse Mercator (UTM) coordinate system in order to geographically match the SAR data.

3.3. SAR data

SAR data were collected using the airborne E-SAR system within the BIOSAR-I experiment, which is described in Section 2.1. For the purpose of this study, the SAR data consisted of fully polarimetric geocoded intensity images with full bandwidth (94 MHz for L-band, 70 MHz for P-band). The center frequency was 1.3 GHz for L-band and 340 MHz for P-band. The acquisition and processing of SAR images were made by DLR. The images were geocoded to the UTM coordinate system using the WGS 1984 ellipsoid and a digital elevation model (DEM) from the Swedish National Land Survey. For the primary heading (200° true north), six P-band and six L-band geocoded images were available, acquired on three time periods: 9 March, 31 March to 2 April, and 2 May 2007. P-band data from a secondary heading (179° true north) were also available, but since no L-band data were available for this heading these data have not been investigated in this paper. The investigated forest stands (58 training stands and 10 validation stands) were located in the incidence angle range of 28–50°. Three corner reflectors were deployed in Remningstorp, two of which were covered by the images from the primary heading. The corner reflectors were not used for radiometric calibration, but for evaluation of the radiometric calibration. The reflectors also served as references for the geocoding process.

The analysis of the radiometric calibration in Hajnsek et al. (2008) revealed a reduced radiometric accuracy for the L-band data acquired on 9 March 2007, compared to the rest of the SAR data. The measured radar cross sections (RCSs) were slightly more than 1 dB from the expected value for both of the geocoded images from this date, which is more than the specified tolerance. One of these images was discarded, as suggested by Hajnsek et al. (2008), since detailed analysis showed system warmup effects. The measured RCSs for all other SAR images (both L- and P-bands) were within the specified tolerance of ± 1 dB. The available SAR data from the different acquisition dates are summarized in Table 2.

Table 3

Summary of key variables for 58 training stands. The biomass and stem volume estimates are based on laser scanning data. The error estimates are based on a comparison to biomass and stem volume estimated from *in-situ* data.

Stand variable	Mean	Min	Max	Estimated error (approximate)
Area [ha]	2.0	0.50	9.4	–
Stem volume [m ³ /ha]	241	12	556	45
Biomass [t/ha]	129	9	287	25

4. Data processing and error analysis

4.1. Field data

The stand level estimates of biomass for the 17 stands (measured *in-situ*) were obtained by the summation of single tree biomass estimated using allometric equations (Marklund, 1988). It is of importance to quantify the error associated with these stand level estimates. In Marklund (1988) errors (standard deviation of the residuals) for the allometric equations are given. For a single tree the error in the total biomass estimate depends on the tree species and on the correlation between the errors of the stem and branch biomass estimates. Assuming a worst case scenario (deciduous trees and highly correlated errors) the error in the biomass estimates can be 30% at tree level. At stand level the relative error will be smaller. The ten 0.7 ha large validation stands contains at least 250 trees, giving an error less than 2% for the stand biomass if errors between trees are uncorrelated. Since it is likely that some correlation exists between the errors for trees within the same stand, the error for the stand biomass is probably larger than 2%. The increased error due to correlation between trees is hard to quantify, but we estimate the error in the stand level biomass estimates to be a few percent. Note also that the 30% error was a worst case estimate.

The arguments above holds for stands for which diameter and height were measured for all trees. For stands without height measurements for all trees, secondary biomass functions were used as described above. Biomass estimates obtained using the secondary biomass functions were in good agreement with the biomass estimates obtained using Marklund (1988). Therefore, the errors in the biomass estimates at stand level obtained using the secondary biomass functions were likely similar to the errors obtained using Marklund (1988). Again, we estimate the error to be only a few percent.

4.2. Estimation of stem volume using laser scanning data

Field data in combination with variables derived from laser scanning data were used to estimate stem volume (Lefsky et al., 1999; Means et al., 2000; Næsset, 2002; Holmgren, 2004; Magnusson et al., 2007). The field data used consist of 849 field plots and are described in Section 3.1. Because of storm events that occurred between field inventory and laser scanning data acquisition, efforts were made to remove field plots that could be affected. This was done using a buffer zone based on GPS measured positions of wind-thrown trees. In total, 603 plots remained after the spatial buffer zone operation. The laser scanning data in terms of laser returns were classified as ground or vegetation returns using a progressive triangular irregular network densification method (Axelsson, 1999, 2000) implemented in the TerraScan software as described in Soininen (2004). A DEM was derived using the laser returns classified as ground returns. The height value of a laser return (referred to as laser height) was computed as the difference between the z-value of the laser return and the z-value of the DEM at the laser return x and y positions. A number of variables were derived from the laser scanning data for the field plots, based on the height distribution of laser returns in the canopy. A height threshold of 10% of the maximum laser height and laser height ≥ 1.0 m was applied on each field plot in order to separate canopy returns from returns of e.g. stones and low vegetation. Laser height percentiles, h_x , where x is an index for 10, 20, ..., 90, 95 and 100%, were calculated from the canopy returns. A vegetation-ratio (D_v) was calculated as the ratio between number of laser returns above the height threshold and total number of returns on the plot. A regression model (Eq. (1)) was developed to estimate stem volume (V) as measured with the FMPP inventory routines (Jonsson et al., 1993):

$$\ln(V) = \beta_0 + \beta_1 \ln(h_x) + \beta_2 \ln(D_v) + \beta_3 \ln(r_x) + \epsilon \quad (1)$$

where r_x is the ratio between a low and a high laser height percentile, and ϵ is the random error assumed to be normally distributed with

zero expectation. The field plots were assigned to five different strata based on the field measurements: Scots pine dominated forest, pure Norway spruce forest, Norway spruce dominated forest, deciduous dominated forest, and low forest (h_{95} less than 8.0 m). For each stratum, the stem volume regression model (Eq. (1)) was tested and the laser height percentile and the ratio of percentiles which yielded the lowest residual sum of squares were selected. The regression coefficients β_0 , β_1 , β_2 and β_3 were estimated for each stratum by means of ordinary least squares (OLS). Variables with non-significant coefficients (p-value larger than 0.05) were excluded from the model and the residual plots were studied. Some outliers could be observed, most probably caused by remaining wind-thrown trees on some plots after the spatial buffer zone operation. These outliers were removed to improve the stem volume estimation. From the 603 plots remaining after the buffer zone operation 28 plots were subjectively removed by observing the residuals.

The same variables as derived from laser scanning data for the field plots were derived for each cell (10 m by 10 m) in a raster (referred to as laser raster). Delineation of the forest into homogeneous patches (*i.e.* stands) and estimation of stem volume proportion of Scots pine, Norway spruce, and deciduous trees were done using stereo aerial photo-interpretation of digital color near infrared images from the Z/I DMC camera. The stand polygons were converted to a raster, with the same geometry and cell size as for the laser raster, where all raster cells within a stand was assigned to one and the same stratum (out of the five predefined strata). This raster is in the following referred to as stratum raster. Stem volume was then estimated for each cell in a raster using the laser raster and stratum raster as input data to the stratum specific regression model (Eq. (1)). The stem volume estimates on log scale were re-transformed to linear scale. To correct for logarithmic bias, these re-transformed values were multiplied by the ratio between the means of *in-situ* stem volume and re-transformed values. After this correction, unbiased estimates of stem volume were obtained. Since the cells in the stem volume raster were too small (10 m by 10 m) for a useful comparison with the SAR data, the raster cell estimates were aggregated to stand level. The extent of each stand was defined based on the stratum raster. Only stands completely covered by the stem volume raster were considered. To minimize border effects and errors due to geographical positioning, the raster cells closest to the border of each stand were removed. Stands consisting of disjoint areas after the border removal were treated as multiple stands. Stands smaller than 0.5 ha were rejected. A total of 59 stands were found which met these criteria. One of these stands, however, was removed since it bordered to a metallic fence strongly visible in the SAR data, leaving 58 forest stands for the analysis.

4.3. Conversion from stem volume to biomass

For each of the 58 stands the stem volume, estimated using laser scanning data, was converted to biomass. This was done by using a Biomass Expansion Factor (BEF) defined as $W = BEF \cdot V$, where W is the above-ground dry biomass in tons/ha and V is the stem volume in m^3/ha . Using the 17 stands measured *in-situ*, tree species dependent (local) BEFs were estimated for the Remningstorp test site. For each of these stands the ratio between biomass and stem volume (both estimated using allometric equations) was calculated. Each stand was then assigned to the dominating species class (pine, spruce or birch), and the BEF for each species was then estimated as the arithmetic mean of BEFs for stands which were assigned to the same species class. Thus, the stands were considered to be homogeneous in terms of tree species. The estimated BEFs were 0.49, 0.52 and 0.71 t/m^3 for pine, spruce and birch, respectively.

Among the 17 stands only two were dominated by deciduous species, and both these stands were dominated by birch. For the 58 training stands, the BEF for birch was used for all stands that were classified as deciduous.

Four of the 58 training stands had a stem volume less than 40 m³/ha. These stands were all younger than 16 years old and dominated by spruce. In Jalkanen et al. (2005) it is found that the BEF for spruce is highly dependent on stand age, whereas the age dependence for pine and birch is weak. Since all of the spruce dominated stands (among the 17 stands) were mature stands, the BEF estimate for spruce calculated using the method described above is not accurate for these young spruce stands with low stem volume. To study the behavior of small trees, all spruce trees with a field measured tree height less than 8 m were extracted, and the ratio between the biomass and stem volume was calculated for each tree. The ratio was found to vary between about 0.70 and 1.16 t/m³, with a clear tendency to increase with decreasing tree height. The mean ratio for the spruce trees with a tree height less than 8 m was 0.85 t/m³. This corresponds well to the BEF for young spruce stands in southern Sweden found in Jalkanen et al. (2005). Thus, the BEF for spruce in the four stands described above was taken to be 0.85.

The estimated BEFs given above were used to convert stem volume to biomass for the 58 training stands. Estimated stem volume proportions for Scots pine, Norway spruce and deciduous trees from the aerial photo-interpretation were used in the conversion. Table 3 summarizes key variables of the 58 training stands.

Errors in estimates of stem volume and biomass using laser scanning data were evaluated using the 10 validation stands. One of these stands was not covered by the stem volume raster (only data from the main area covered by laser scanning data were used to estimate stem volume), leaving 9 stands for evaluation. Estimates of biomass and stem volume derived from laser scanning data is plotted against the corresponding *in-situ* estimates in Fig. 3. The conversion from stem volume to biomass was done in the same way as for the 58 training stands. For most stands, the laser derived estimates correspond well to the *in-situ* estimates. One stand has a lower stem volume than the other stands, and for this stand the estimation error is larger. This is the only stand dominated by birch. The RMSE for the stem volume is 44 m³/ha, while the RMSE for biomass is 26 t/ha. A reasonable estimate of the error in the biomass derived from laser scanning data is therefore 25 t/ha. However, there is a considerable amount of uncertainty in the error estimate. None of 9 validation stands were young spruce stands. Thus, Fig. 3 cannot be used to evaluate errors for the 4 young spruce stands (among the 58 training stands). However, residual studies using the field plots give an error estimate of 50% of the stem volume for young stands. Based on the BEFs for individual trees (see above), the error in the BEF for young spruce is estimated to be at most 50%. Since the estimated biomass for the young spruce stands is small (less than 26 t/ha), the 25 t/ha error estimate can be used also for the young spruce stands.

4.4. Radar data

The radar backscattering coefficient normalized for incidence angle (γ_0), was estimated for each image and forest stand according to Ulaby et al. (1982):

$$\sigma_{0,i,n,m} = (P_{i,n,m} + K)^2 \cdot C \cdot \sin(\theta_{i,n,m}) \quad (2)$$

$$\gamma_{0,i,m} = \frac{1}{N_m} \sum_{n \in \Omega_m} \frac{\sigma_{0,i,n,m}}{\cos(\theta_{i,n,m})} \quad (3)$$

where $P_{i,n,m}$ is the pixel value in image i , for pixel n in stand m and $\theta_{i,n,m}$ is the corresponding local incidence angle. K and C are calibration constants with the values 32,769 and 10^{-6} , respectively (Hajnsek et al., 2008). The image identifier i corresponds to an image with a specific frequency, polarization and acquisition time. N_m is the number of pixels in stand m and Ω_m is the set of pixels belonging to stand m . The geometric projection factor $\sin(\theta_{i,n,m})$ is an approximation, the correct projection factor is $\cos\psi$, where ψ is the angle between the surface normal and the image plane normal (Ulander, 1996). It is shown in

Ulander (1996) that this approximation is valid as long as there are no significant azimuthal slopes.

The differences in the backscattering coefficient at stand level between images with the same frequency, polarization and heading acquired on the same day were smaller than 0.7 dB for all such image pairs, which is comparable to the measured radiometric calibration uncertainty. In order to simplify the analysis and to reduce the effects of noise, images from such image pairs were combined using simple averaging.

A noise level was not subtracted in Eq. (2) for three reasons. Firstly, the SAR images only contained a few areas where the noise dominates over the signal (e.g. lakes and fields). Therefore, reliable noise estimates based on the images were difficult to obtain. Secondly, the backscatter levels measured over two such areas indicated that the noise level is low, less than -25 dB for HH- and VV-polarized backscatter (L- and P-bands), -30 dB for HV-polarized L-band backscatter and -35 dB for HV-polarized P-band backscatter. This is low enough to be neglected for the purpose of this study. Finally, the subtraction of a (small) constant noise level will not result in any significant changes in the backscatter–biomass relationship. It was concluded that the noise in the SAR images could be ignored without a significant impact on the analysis in this paper.

The SAR images used in the analysis were multi-look images with about 2.5 looks. A resolution cell in the multi-looked images was about 3 m (azimuth) by 2 m (slant range) and 4 m (azimuth) by 2 m (slant range) for L- and P-bands, respectively. For steep incidence angles, this gives resolution cells in the geocoded images of about 12 m² and 16 m² for L- and P-bands, respectively. The number of looks for the average backscatter for an area of size A can be calculated as $1/2 \cdot A/A_{res} \cdot N$, where A_{res} is the area of a resolution cell, N is the number of looks in the image, and the factor two has been included to correct for correlation between resolution cells. For the smallest stands used in the analysis (0.5 ha) this gives about 520 and 390 looks for L- and P-bands, respectively. Thus, the variability due to speckle for the stand level estimates of backscatter is about 0.2 dB for both frequency bands. For the discarded 0.1 ha stands, the corresponding variability due to speckle is more doubled, i.e. about 0.4–0.5 dB.

In addition to forest parameters, thermal noise and speckle, the estimated backscatter for a forest stand can be affected by geolocation errors, layover of tree canopy, border effects, as well as topography and moisture conditions (e.g. soil moisture). The impact of these factors is difficult to quantify. In this context, the spatial scale is of great importance. For small areas (e.g. the seven discarded 0.1 ha stands) the combined effect of the error sources discussed above is likely to have large impacts on the measured backscatter, and therefore we choose 0.5 ha as a minimum stand size. As discussed above, the speckle variations for 0.5 ha large stands is about 0.2 dB. Border effects and geolocation errors should also be small. Moreover, stands of this size are large enough so that estimates of forest properties (biomass, tree species, etc.) are likely to be stable given small variations in e.g. stand delineation.

If the area size is increased beyond 0.5 ha, the accuracy of spatial averages of e.g. radar backscatter is increased. However, at some point the area becomes so large that these averages are no longer sufficient to describe the underlying area. This occurs if several stands with different forest properties are contained in the same areal unit. In this paper spatial averages are always made within stands, inside which the forest is fairly homogeneous.

As mentioned above, topography and soil moisture are factors influencing the backscatter. Since the topographic variations within the test site are small, topography should only have a small impact on the measured backscatter. However, local topography may affect the SAR data, especially for the HH-polarization at P-band (Hallberg et al., 2008). As stated in Section 3.1, no strong variations in soil moisture within the forest have been observed, although the moisture content is somewhat higher for 9 March 2007 than for the other image acquisition dates. As will be shown in this paper, the L-band data have

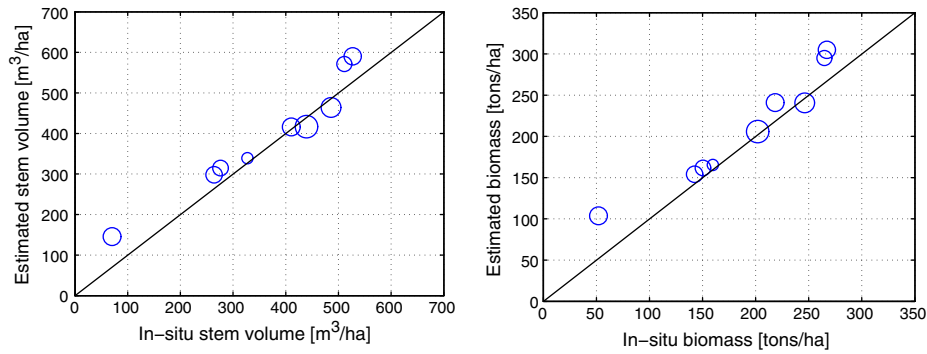


Fig. 3. Stem volume (left) and biomass (right) estimated using laser scanning data plotted against the corresponding quantity estimated from *in-situ* data for 9 of the 10 validation stands. The marker radius is inversely proportional to the number of trees in the stand. Note that the stand with the lowest stem volume is dominated by birch, whereas the other stands are dominated by pine (2 stands) or spruce (6 stands).

the highest variability in backscatter for the acquisition corresponding to the highest moisture content.

5. Biomass retrieval method

The main objectives of this study are to assess the accuracy of biomass retrieval in boreal forests from L- and P-band data, and to investigate the difference between the frequency bands. For this purpose a regression model was introduced, which was able to describe the data while still being linear in all regression coefficients (Draper & Smith, 1998). For linear models, estimation of regression coefficients is straightforward, and the statistical properties of the estimates are well understood. The predictor variables (variables used for biomass prediction) were chosen based on previous studies on biomass retrieval using L- and P-band data, e.g. Beaudoin et al. (1994); Ranson and Sun (1994); Rignot et al. (1995); Fransson and Israelsson (1999); Saatchi et al. (2007). The regression model was selected to be:

$$W^\lambda = \beta_0 + \beta_1 \cdot \gamma_0^{dB} + \beta_2 \gamma_0^{1/2} + \beta_3 \gamma_0 + \beta_4 \gamma_0^{3/2} + \beta_5 \gamma_0^2 + \epsilon. \quad (4)$$

In Eq. (4), W is the biomass (tons/ha), and λ is a parameter to be estimated from the data. To improve inversion stability and simplify the interpretation of the estimation models, λ was restricted to half integers between 0 and 2. For $\lambda=0$, W^λ is interpreted as $\ln(W)$. The transformation W^λ is called the Box–Cox transformation. In Draper and Smith (1998) it is proposed as an often useful transformation, especially if the range of values for W is large, or if the error variance (*i.e.* the variance of the residuals) is non-constant over the range of W , as is the case in this paper. The right hand side of Eq. (4) contains a constant and a linear combination of predictors γ_0^q , where q is half integers ranging from 0 to 2. For $q=0$, is γ_0^q interpreted as the logarithm of γ_0 . By convention, the backscatter is expressed in decibels rather than the natural logarithm. The notation $\gamma_0^{dB} = 10 \log(\gamma_0)$, where \log is the logarithm with base 10, is used throughout the paper. Finally, ϵ is a random error assumed to be normally distributed with zero mean and constant variance for the “true” value for λ . A selection procedure, described below, was used to select the best predictors for the model.

5.1. Optimizing λ

The first step in the regression analysis is to estimate the parameter λ . This was done using the Box–Cox method, as described in Draper and Smith (1998). The idea is to choose a λ , perform a regression analysis using this particular λ , and calculate the sum of squared residuals for the model. This is then repeated for several λ within the specified range, and the parameter value which yields the minimum residual sum of squares is selected. This value is the maximum likelihood estimate of λ under the

assumption that the error is normally distributed with zero expectation and constant variance for this parameter value. To overcome the problem that W^λ converges to 1 as λ goes to 0, W^λ is not used directly in the optimization procedure, but rather the transformed variable (Draper & Smith, 1998):

$$V = \begin{cases} (W^\lambda - 1) / (\lambda \dot{W}^{\lambda-1}) & \text{for } \lambda \neq 0 \\ \dot{W} \ln(W) & \text{for } \lambda = 0 \end{cases}$$

\dot{W} is the geometric mean of the biomass at stand level. In a first step λ was allowed to take any value between 0 and 2. The Box–Cox method also gives a confidence interval for λ . Once an optimal parameter value has been obtained, the nearest half integer value was selected, provided that it lies within the estimated confidence interval for λ . Since data from several dates exist for each combination of frequency and polarization, the results for all dates must be considered before selecting which λ to use. As a first step, all predictors in Eq. (4) were used in the regression analysis. At a later stage, when the best predictors had been found, the procedure was repeated using only these predictors. The resultant estimates of λ were compared with the ones found using all predictors.

5.2. Model selection method

For a given value of the transformation parameter λ the “best” set of predictor variables should be selected. The problem is to keep the number of predictor variables low while still selecting a model able to describe the data. Stepwise regression, as described in Draper and Smith (1998), was used to find which predictors to retain in the final model. The principle of the method is as follows:

1. The predictor variable with the highest correlation to the response variable is included in the model, and the residual sum of squares (RSS) is calculated for this model.
2. All predictor variables not included in the current model are added one at a time, and the RSS for the models are calculated. The best model (lowest RSS) is compared to the first model using a statistical test called the F-test. The F-test is based on the fact that the difference between the two RSSs, with proper normalization for degrees of freedom (number of parameters) in the models, follows an F-distribution (Draper & Smith, 1998). If the F-test is significant at a 5% level (*i.e.* the p-value for the test is less than 0.05), there is enough evidence to reject the null hypothesis that the improvement in the RSS due to the inclusion of the additional predictor variables is explained by chance. The second (larger) model is then used as a new baseline model, and the second step is

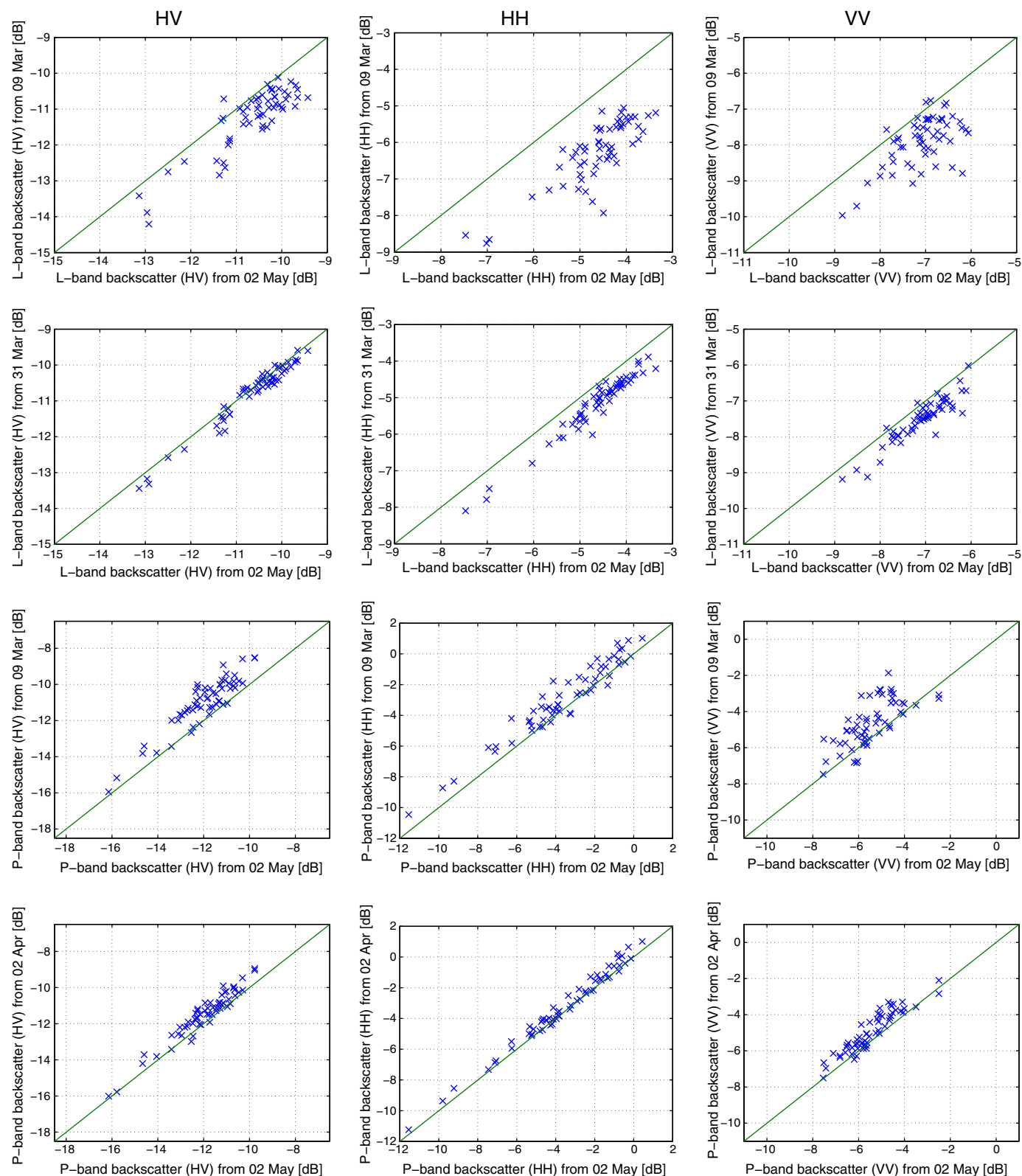


Fig. 4. Plots of the temporal stability of backscatter. The data points correspond to the stand level average backscattering coefficient (γ_0) for the 58 training stands. The first column shows the results for HV-, the second for HH- and the last for VV-polarized backscatter. The data from 2 May 2007 are used as reference in all plots (shown on the horizontal axis). On the vertical axis L-band data from 9 March 2007, L-band data from 31 March 2007, P-band data from 9 March 2007 and P-band data from 2 April 2007 are shown on rows 1–4, respectively.

repeated. When no more terms are added to the model, the third step of the model selection procedure is initiated.

- Each of the predictor variables included in the model after step 2 are again compared to the model where the predictor variable has

been excluded. If any of the F-tests turn out not to be significant at 10% level, the predictor variable associated with the highest p-value is excluded from the model, and the model selection procedure is started again from the second step. When no predictor

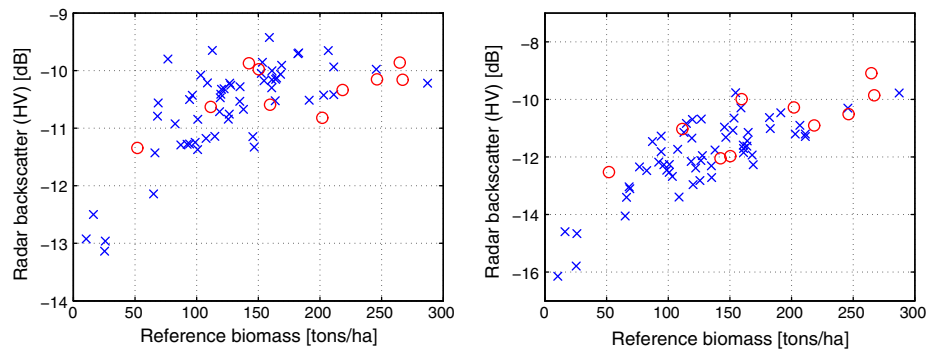


Fig. 5. HV-polarized backscatter from 2 May 2007 plotted against reference biomass for L- (left) and P-band (right). The crosses correspond to 58 training stands for which the biomass has been estimated using laser scanning data, and the circles correspond to 10 validation stands. For clarity no error bars have been included in these plots. Note that the scaling of the vertical axis differs between the plots.

variables are removed from the model, the final model has been obtained.

In cases where the selected models varied between acquisition dates, RMSEs and coefficients of determination were studied in order to select a unique model.

For each frequency, inversion models using multiple polarizations were also investigated. If all predictors in Eq. (4) were to be used in this analysis it would lead to a set of up to 15 predictors, which was deemed to be too many. Instead, only the predictors retained in the single-polarized regression models were used. These predictors were used as input to the stepwise regression procedure explained above.

5.3. Model validation

To validate the biomass regression models the 10 validation stands were used. For these stands, RMSE was calculated using the model with regression coefficients estimated on the training stands. The validation using these 10 stands cannot be considered to be fully independent, since there is some overlap between the training stands and the validation stands (see Fig. 2). There is also some overlap between the systematic-grid plots and the 10 validation stands. Nevertheless, these dependencies can be considered to be of limited importance, and thus the 10 validation stands were used for validation.

6. Temporal stability of backscatter

In this section the temporal stability of the backscatter intensity is analyzed. The backscatter from the 58 training stands for different dates are plotted against each other in Fig. 4. The temporal stability is generally

good, except for the plots on the first row in Fig. 4, which compares L-band data from 2 May 2007 with L-band data from 9 March 2007. Especially for VV-polarization the correlation between the backscatter from these dates is low. The reduced correlation for L-band data is likely caused by the fact that the soil moisture content is somewhat higher for 9 March 2007 than for the other acquisition dates. Note also that only one L-band image was used for this data, while for all other dates the average of two images were used (see Section 3.3).

In contrast to the first row in Fig. 4, the second row shows that the L-band backscatter from 2 May 2007 and 31 March 2007 are well correlated. The difference between the mean backscatter levels (mean over all training stands) is at most 0.5 dB for all polarizations, which is less than the radiometric calibration uncertainty. On the third row in Fig. 4, the P-band backscatter from 2 May 2007 and 9 March 2007 are compared. The difference between the means is at most 1 dB, which is within the tolerance limits for the radiometric calibration. The comparison between the data from 2 May 2007 and the 2 April 2007 is shown on the last row. The difference between the mean P-band backscatter levels is less than 0.6 dB. In summary, the variations in the mean backscatter levels over time are within the radiometric calibration uncertainty, except for the L-band data from 9 March 2007 which show deviating behavior. At single stand level the variations are larger, but the backscatter from the different dates is well correlated. It was concluded that the SAR data in general show good temporal stability, especially for the P-band HV- and HH-polarizations.

7. Qualitative analysis of the backscatter–biomass relation

In this section, a qualitative analysis of the backscatter–biomass relation is presented. The analysis is based on plots of the measured

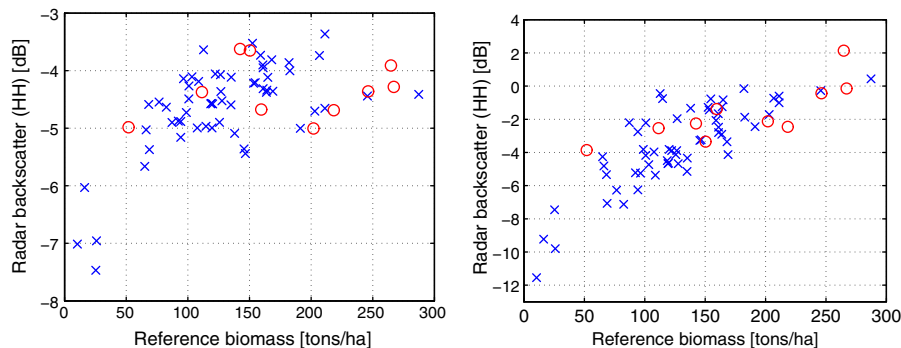


Fig. 6. HH-polarized backscatter from 2 May 2007 plotted against reference biomass for L- (left) and P-band (right). The crosses correspond to 58 training stands for which the biomass has been estimated using laser scanning data, and the circles correspond to 10 validation stands. For clarity no error bars have been included in these plots. Note that the scaling of the vertical axis differs between the plots.

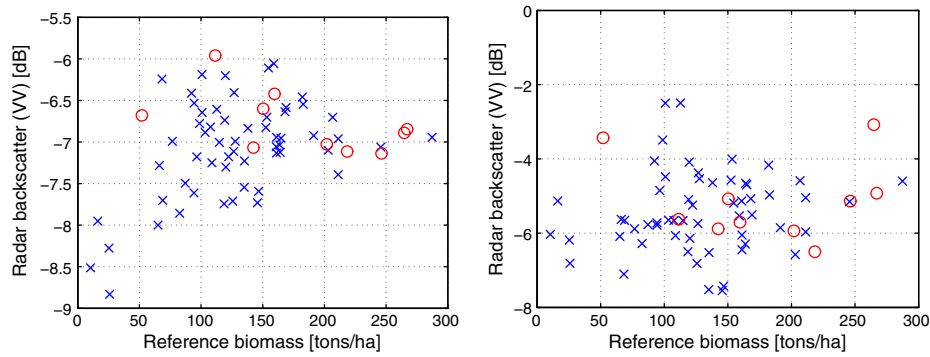


Fig. 7. VV-polarized backscatter from 2 May 2007 plotted against reference biomass for L- (left) and P-band (right). The crosses correspond to 58 training stands for which the biomass has been estimated using laser scanning data, and the circles correspond to 10 validation stands. For clarity no error bars have been included in these plots. Note that the scaling of the vertical axis differs between the plots.

backscatter and reference biomass for the 58 training stands and the 10 validation stands. The plots are based on data from 2 May 2007, since this is the only single acquisition date with the same number of images available for both L- and P-bands (see Table 2). The discussion of the results, however, is based on the results for all acquisition dates.

The scattering from forests can be assumed reciprocal, *i.e.* the scattering for HV- and VH-polarizations are equal. The SAR data in this study were collected using an airborne system, and thus no ionospheric effects causes differences between HV- and VH-polarizations. Therefore, any differences between the two cross polarized channels should be caused by system distortions (cross-talk and channel gain imbalance) and noise. A comparison between HV and VH has been made, and only small differences were found. For simplicity results from the VH channel will be omitted in this paper.

In Fig. 5, the HV-polarized backscatter from 2 May 2007 is plotted against biomass for the two frequency bands. Both plots show that the backscatter is highly dependent on biomass. For L-band, the backscatter increases with biomass up to some level around 150 t/ha, after which no further increase is seen. For P-band, the backscatter increases with biomass over the full range of biomass, *i.e.* up to 290 t/ha. There is a slight tendency of saturation for high biomass, but much less than for L-band. Since saturation effects are present for both frequency bands, the relation between the backscatter and biomass cannot be described as linear. A notable difference between the plots is that P-band has a larger dynamic range than L-band, 7 dB compared to 4 dB (on 2 May 2007).

In Fig. 6, the HH-polarized backscatter from 2 May 2007 is plotted against biomass for both L- and P-band data. For L-band, the plot is very similar to the corresponding plot for HV-polarized backscatter. The backscatter increases with biomass for biomass less than about 150 t/ha.

For higher biomass no further increase is seen. For P-band HH-polarized backscatter, the general behavior is similar to that of the HV-polarized backscatter. There is an increase of backscatter with biomass over the full range of biomass. As for HV-polarized backscatter, the dynamic range is larger for P-band than for L-band, 14 dB compared to 4 dB (2 May 2007).

For VV-polarized backscatter, there was less evidence of strong correlation with biomass. In Fig. 7, results for the data from 2 May 2007 are shown. For L-band, the behavior shows some similarities with the other polarizations, but the backscatter dependence on biomass is weaker. The spread caused by features other than biomass (*e.g.* ground topography, vertical branch distribution, and soil moisture) is so large that there is little reason to hope that biomass can be estimated using VV-polarized L-band data. However, in combination with other polarizations these data might improve the biomass estimates, and therefore VV-polarized L-band data were included in the multi-polarized regression model. For P-band, the VV-polarized backscatter shows no dependence on biomass. Due to the lack of correlation between backscatter and biomass, no biomass regression models based on VV-polarized P-band backscatter are presented in this paper.

The result that VV-polarized backscatter and biomass are uncorrelated differs from previously published studies (Beaudoin et al., 1994; Le Toan et al., 1992; Saatchi et al., 2007). The reasons for the lack of correlation observed in this study are unclear, and further studies are needed to fully understand the behavior for VV-polarized data, especially at P-band.

8. Results of biomass retrieval

In this section, the results of the regression analysis are presented. As in the previous section, the scatter plots are based on data from 2 May

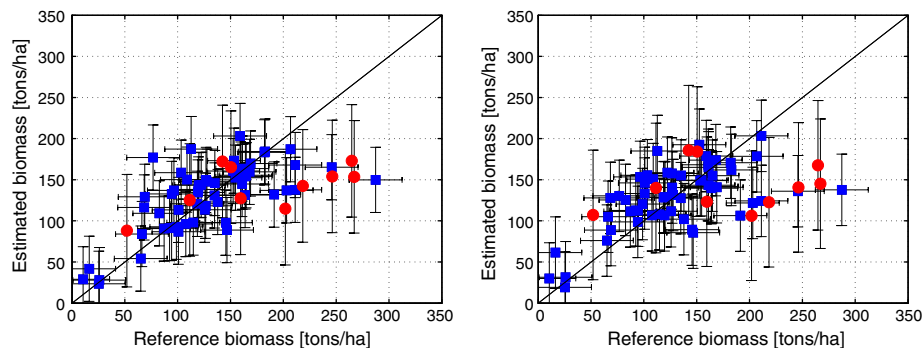


Fig. 8. Result of biomass estimation using L-band data from 2 May 2007 based on the regression model given in Eq. (5). To the left the results for HV- and to the right the results for HH-polarization are shown. The squares and circles correspond to the 58 training and 10 validation stands, respectively, and the solid line is the one-to-one line. The length of the horizontal error bars equals the estimated error in the biomass estimates for the training stands (see Section 4.3). The corresponding error for the validation stands is small and is therefore not shown. The length of the vertical error bars equals $RMSE_c$ and $RMSE_v$ for the training and validation stands, respectively. The biomass estimates were corrected for transformation bias. Estimated regression coefficients are found in Table 5.

2007, but the results and discussions are based on data from all acquisition dates. The results from the single-polarized regression models are presented for each frequency, followed by results obtained using a combination of polarizations.

The parameters of the regression model (Eq. (4)) were estimated using the methods described in Section 5. First the parameter λ was estimated and confidence intervals for the parameter were found. In this step all predictor variables were included in the model. The estimate for λ was then rounded to a half integer value. For all frequencies and polarizations, $\lambda = 1/2$ was included in the confidence interval for λ . Moreover, $\lambda = 1/2$ was in most cases the half integer value closest to the estimated value.

Using $\lambda = 1/2$, stepwise regression was used to find which predictor variables to include in the regression model. Except for L-band data from 9 March 2007, the stepwise regression procedures resulted in a single predictor variable in all models, namely γ_0^{dB} (i.e. the backscatter in dB). Using only γ_0^{dB} as predictor variable, the estimation procedure for λ was repeated. Again $\lambda = 1/2$ was the only half integer value included in all confidence intervals for λ . It was concluded that the best predictor for the HV- and HH-polarized channels is the backscatter in dB for both L- and P-bands. Therefore, Eq. (4) was reduced to

$$W^{1/2} = \beta_0 + \beta_1 \gamma_0^{dB}. \quad (5)$$

8.1. Single-polarized models: L-band

In Fig. 8, the results of biomass estimation using L-band data from 2 May 2007 are shown. Since the regression coefficients for Eq. (5) were estimated to minimize the error in \sqrt{W} for the training stands, some estimation bias was present after transformation to W (without the square root). Therefore, the biomass estimates were corrected for this transformation bias by applying a correction factor to make sure that the estimates were unbiased for the training stands (compare to Section 4.2). It can be seen in Fig. 8 that the results are good for low biomass, but there are a few stands with high biomass which are significantly underestimated. Table 4 shows RMSE calculated using leave-out-one cross-validation of the training (RMSE_c) and validation stands (RMSE_v). The coefficients of determination R^2 (adjusted for degrees of freedom, calculated in the transformed space), for the training stands, are also shown in Table 4.

RMSE_c varied between 40 and 48 t/ha (31–37% of mean biomass) for HV- and 43–50 t/ha (33–39% of mean biomass) for HH-polarization. A large portion of the error stems from the significantly underestimated stands with high biomass. The largest errors were obtained for L-band data from 9 March 2007. It should be noted that 1) some of the SAR data from this date were subject to calibration problems (see Section 3.3) and 2) the measured soil moisture content was higher on this date than on any of the other acquisition dates.

For the validation stands the errors were larger. RMSE_v was estimated to be 68–83 t/ha (38–46% of mean biomass) for HV- and 75–

Table 4

Results of biomass regression models using L-band (Eq. (5)). R^2 is the (adjusted) coefficient of determination and RMSE_c is the root mean square error calculated using cross-validation of the training stands (58 stands). RMSE_v is the error when applying the regression model to the validation stands (10 stands). It should be noted that the mean biomass of the validation stands is higher than the corresponding mean for the training stands (see Tables 1 and 3).

Polarization	Date	R^2	RMSE _c [t/ha]	RMSE _v [t/ha]
HV	9 March	0.39	48	83
	31 March	0.56	41	71
	2 May	0.61	40	68
HH	9 March	0.31	50	83
	31 March	0.52	43	75
	2 May	0.51	43	79

Table 5

Estimated regression coefficients for biomass regression model (Eq. (5)) using L-band data. The limits of the coefficients are estimated 95% confidence intervals. The correction factor for transformation bias is 1.02 for the HV-polarized model from 31 March and 2 May 2007 and 1.03 for all other models.

Polarization	Date	β_0	β_1
HV	9 March	32.0 ± 6.9	1.9 ± 0.6
	31 March	35.6 ± 5.8	2.3 ± 0.5
	2 May	37.8 ± 5.7	2.5 ± 0.5
HH	9 March	21.6 ± 4.1	1.7 ± 0.6
	31 March	22.6 ± 3.0	2.2 ± 0.6
	2 May	22.1 ± 2.9	2.4 ± 0.6

83 t/ha (41–46% of mean biomass) for HH-polarization. Again, the largest errors were obtained for the data from 9 March 2007. The increase in relative terms is not as large, since the validation stands have a higher mean biomass than the stands used for training. The stands with biomass above 150 t/ha are consistently underestimated, leading to an overall bias of more than 40 t/ha for both polarizations. Comparing the HV- and HH-polarizations, it is seen that both RMSE and R^2 values indicate that HV- is better suited for biomass estimation than the HH-polarization.

Estimated coefficients for the regression models are found in Table 5. The temporal variations in the regression coefficients are comparable to or less than the estimated confidence intervals for the coefficients.

8.2. Single-polarized models: P-band

In Fig. 9, the results of biomass estimation using P-band data from 2 May 2007 are shown. In contrast to the estimations using L-band data, the estimation errors for stands with high biomass values are not significantly larger than for stands with low biomass. This has a strong positive effect on the RMSE values. RMSE_c is 33–34 t/ha (26–27% of mean biomass) for HV- and 30–32 t/ha (23–25% of mean biomass) for HH-polarization. R^2 varies between 0.69–0.70 and 0.72–0.75 for HV- and HH-polarizations, respectively. These results are consistently better than for L-band data.

Fig. 9 also shows the results of biomass estimation for the validation stands. The errors are of the same order as for the training stands, and there is no clear increase in the estimation error for high biomass as there were for L-band data. RMSE_v is 32–42 t/ha (18–23% of mean biomass) for HV- and 41–45 t/ha (23–25% of mean biomass) for HH-polarization. The bias is much lower than for L-band, and is at most –9 t/ha for HV- and –13 t/ha for HH-polarization. Comparing the two polarizations, it is seen that RMSE_c is slightly better for HH- than for HV-polarization, while RMSE_v is lower for HV- than HH-polarization. It is therefore difficult to judge which of the two polarizations that is most suitable for biomass prediction (Table 6).

Estimated regression coefficients and their corresponding confidence intervals are shown in Table 7. The temporal variations in the regression coefficients are comparable to or less than the estimated confidence intervals for the coefficients.

8.3. Multi-polarized models

From the analysis using one polarization at a time it was decided that the multi-polarized regression model should include HV- and HH-polarized backscatter in dB. For L-band, VV-polarized backscatter in dB was also included, as well as the square root of the backscatter (in m²/m²), as these predictor variables were found to give some information on biomass. The multi-polarized regression model was selected to be

$$W^{1/2} = \beta_0 + \beta_1 \gamma_{0,HV}^{dB} + \beta_2 \gamma_{0,HH}^{dB} + I_k \beta_3 \gamma_{0,VV}^{dB} + I_k \beta_4 \gamma_{0,VV}^{1/2} \quad (6)$$

where I_k is an indicator variable that is one for L-band and zero for P-band. An analysis of the regression model utilizing several polarizations

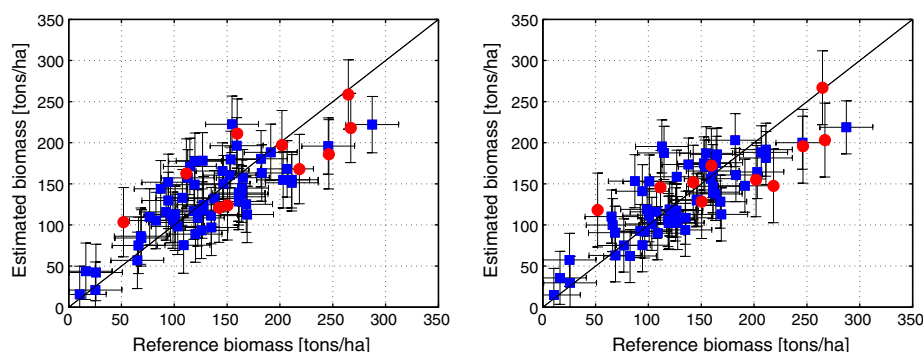


Fig. 9. Result of biomass estimation using P-band data from 2 May 2007 based on the regression model given in Eq. (5). To the left the results for HV- and to the right the results for HH-polarization are shown. The squares and circles correspond to the 58 training and 10 validation stands, respectively, and the solid line is the one-to-one line. The length of the horizontal error bars equals the estimated error in the biomass estimates for the training stands (see Section 4.3). The corresponding error for the validation stands is small and is therefore not shown. The length of the vertical error bars equals $RMSE_c$ and $RMSE_v$ for the training and validation stands, respectively. The biomass estimates were corrected for transformation bias. Estimated regression coefficients are found in Table 7.

simultaneously was performed using Eq. (6) and stepwise regression. For L-band data, the best predictor variable was $\gamma_{0,HV}^{dB}$. After the inclusion of this predictor variable in the model, no significant improvement in the estimation errors were obtained by adding other predictor variables to the model. The exception from these results was for 9 March 2007 where the square root of VV-polarized backscatter gave a significant improvement of the model. The fact that HH-polarized L-band backscatter did not significantly improve the model might seem surprising. However, for L-band HV- and HH-polarizations are strongly correlated, which means that little is gained by including both polarizations in the model. Thus, it was concluded that the best inversion model for L-band, among the models investigated in this paper, was the single-polarized regression model using only $\gamma_{0,HV}^{dB}$ as predictor variable. Results obtained using this model have already been presented in Section 8.1.

For P-band data, the regression model using both HV- and HH-polarized backscatter was found to be significantly better than any of the single-polarized models. The results of the inversion for 2 May 2007 are shown in Fig. 10. The estimation error seems to be fairly constant over the full range of biomass for both the training and validation stands. No clear signs of underestimation for high biomass values are seen. Values for R^2 and RMSE are shown in Table 8. The $RMSE_c$ is consistently better than for the single-polarized models, and varies between 29 and 30 t/ha (22–24% of mean biomass). R^2 is 0.77–0.79, which is also better than what was obtained using the models utilizing only one polarization. $RMSE_v$ varies between 35 and 40 t/ha (19–22% of mean biomass), which is comparable to the HV-polarized model and slightly lower than for the HH-polarized model. The overall bias for the training stands was found to be at most -7 t/ha. No significant temporal variations in the estimated regression coefficients were seen.

Table 6

Results of biomass regression models using P-band (Eq. (5)). R^2 is the (adjusted) coefficient of determination and $RMSE_c$ is the root mean square error calculated using cross-validation of the training stands (58 stands). $RMSE_v$ is the error when applying the regression model to the validation stands (10 stands). It should be noted that the mean biomass of the validation stands is higher than the corresponding mean for the training stands (see Tables 1 and 3).

Polarization	Date	R^2	$RMSE_c$ [tons/ha]	$RMSE_v$ [tons/ha]
HV	9 March	0.69	33	32
	31 March	0.69	34	35
	2 May	0.70	34	42
HH	9 March	0.75	30	43
	31 March	0.74	31	41
	2 May	0.72	32	45

9. Summary and conclusions

In this paper, a method for estimating forest biomass on stand level using SAR backscatter data for L- and P-band frequencies has been developed and evaluated. Two sets of forest stands were used. One set of stands consisted of 58 stands (training stands) larger than 0.5 ha, for which biomass estimates were obtained using laser scanning data. The error in the biomass estimates for these stands was estimated to be about 25 t/ha. A smaller error could probably have been obtained if the laser scanning data were used to estimate biomass directly, instead of estimating stem volume and then converting this estimate to biomass. These 58 training stands were used to estimate regression coefficients for the biomass estimation models. The other set consisted of 10 stands (validation stands), all about 0.7 ha large, for which *in-situ* measurements were made on every single tree. The error in the biomass estimates for these 10 stands was limited to only a few percent of the biomass, and is therefore negligible. These stands were used as validation stands to test the biomass regression models.

9.1. Temporal stability

Temporal variations in the backscatter for the 58 training stands were analyzed. For L-band, data acquired on 31 March 2007 and 2 May 2007 were well correlated, and the difference in mean backscatter level (mean over all training stands) was at most 0.5 dB. The correlation between L-band data acquired 9 March 2007 and the rest of the L-band data was lower. The reason for this is likely that the soil moisture content was somewhat higher on this date than for the other acquisition dates. For P-band, the backscatter from the different acquisition dates were well correlated, and the maximum difference in mean backscatter level (mean over all stands) was 1 dB, less than the specified calibration tolerance (Hajnsek et al., 2008). Considering the large time span between the first and last image acquisitions, this is an important result.

Table 7

Estimated regression coefficients for biomass regression model (Eq. (5)) for P-band data. The limits of the coefficients are estimated 95% confidence intervals. The correction factor for transformation bias is 1.01 for the HH-polarized models and 1.02 for the HV-polarized models.

Polarization	Date	β_0	β_1
HV	9 March	27.5 ± 2.9	1.5 ± 0.3
	31 March	29.4 ± 3.2	1.6 ± 0.3
	2 May	31.5 ± 3.5	1.7 ± 0.3
HH	9 March	13.8 ± 0.6	1.0 ± 0.1
	31 March	14.0 ± 0.6	0.9 ± 0.1
	2 May	14.3 ± 0.6	0.9 ± 0.1

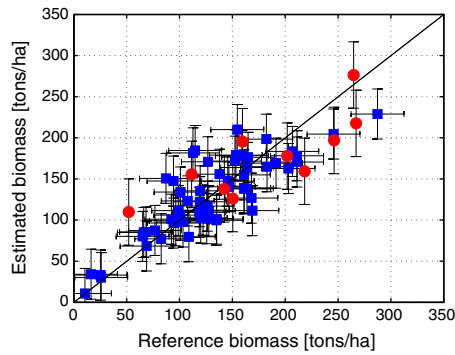


Fig. 10. Result of biomass estimation using P-band data from 2 May 2007 based on the multi-polarized regression model: $W^{1/2} = \beta_0 + \beta_1 \gamma_{HV}^{dB} + \beta_2 \gamma_{HH}^{dB}$. The squares and circles correspond to the 58 training and 10 validation stands, respectively, and the solid line is the one-one line. The SAR data is from 2 May 2007. Error bars have the same meaning as in Fig. 9. Estimated regression coefficients are presented in Table 9.

The temporal variations in the estimated regression coefficients were comparable to or less than the corresponding confidence intervals.

9.2. Estimation of biomass

A qualitative analysis of the SAR backscatter dependency on biomass for the different frequencies and polarizations was made. The analysis was based on scatter plots of the measured backscatter and estimated biomass for the investigated forest stands (58 training stands and 10 validation stands). For L-band it was found that the dependence on biomass was similar for HV- and HH-polarized backscatter. The backscatter in general increased with increasing biomass up to about 150 t/ha, after which a decreased sensitivity to biomass was observed. The VV-polarized backscatter showed a somewhat similar behavior, but the dependence on biomass was weaker. For P-band, the HV- and HH-polarized backscatter both showed a strong dependence on biomass over the full range of biomass (10–290 t/ha). Slight signs of saturation were seen, but much less pronounced than for L-band. The VV-polarized backscatter, on the other hand, showed virtually no dependence on biomass. A notable difference between the frequency bands was that P-band had a larger dynamic range than L-band. The dynamic range for HV- and HH-polarized L-band backscatter was about 4 dB, while for P-band the dynamic range was 7 and 14 dB for HV- and HH-polarized backscatter, respectively.

Models for biomass estimation was developed by means of regression analysis for L- and P-band HV- and HH-polarized backscatter data. VV-polarized data were not used for regression, since the dependence on biomass was weak. A biomass regression model was proposed, and predictor variables were selected using stepwise regression. The resulting regression model is given in Eq. (5).

For L-band HV-polarized data, R^2 (adjusted) was between 0.4 and 0.6, $RMSE_c$ (based on cross-validation of the training stands) was 40–48 t/ha (31–37% of mean biomass) and $RMSE_v$ (based on validation

Table 8

Results of biomass regression model (Eq. (6)) for multiple polarizations using P-band data. R^2 is the (adjusted) coefficient of determination and $RMSE_c$ is the root mean square error calculated using cross-validation of the training stands (58 stands). $RMSE_v$ is the error when applying the regression model to the validation stands (10 stands). It should be noted that the mean biomass of the validation stands is higher than the corresponding mean for the training stands (see Tables 1 and 3).

Date	R^2	$RMSE_c$ [t/ha]	$RMSE_v$ [t/ha]
9 March	0.79	29	35
31 March	0.79	29	35
2 May	0.77	30	40

Table 9

Estimated regression coefficients for biomass regression model (Eq. (6)) using HV- and HH-polarized P-band data. The limits of the coefficients are estimated 95% confidence intervals. The correction factor for transformation bias is 1.01 for all models.

Date	β_0	β_1	β_2
9 March	20.3 ± 3.7	0.68 ± 0.38	0.61 ± 0.24
31 March	21.3 ± 4.2	0.73 ± 0.42	0.58 ± 0.23
2 May	23.2 ± 5.0	0.86 ± 0.48	0.53 ± 0.25

stands) was 68–83 t/ha (38–46% of mean biomass). The largest estimation errors were obtained for L-band data from 9 March 2007. The results were better for low biomass, and stands with biomass larger than about 150 t/ha were underestimated. For HH-polarization, the results were consistently slightly worse than for HV.

For P-band HV- and HH-polarized data, R^2 (adjusted) was in the range of 0.69–0.70 and 0.72–0.75, respectively. $RMSE_c$ was 33–34 t/ha (26–27% of mean biomass) for HV- and 30–32 t/ha (23–25% of mean biomass) for HH-polarized backscatter. In contrast to L-band data, the estimation error was not larger for high biomass. Consequently, the difference between $RMSE_c$ and $RMSE_v$ was not as large for P-band as it was for L-band data. $RMSE_v$ varied between 18% and 25% of the mean biomass, and was higher for HH- than HV-polarized data.

Multi-polarized models were investigated. For L-band the stepwise regression procedure showed that no significant improvement was found by combining multiple polarizations (except for 9 March 2007 where the square root of VV-polarized backscatter also was included in the model). This result was caused by two facts: 1) there is a large variability in the backscatter-biomass relation and 2) the different polarizations are highly correlated. In combination, these two facts resulted in that the reduction in estimation error obtained by the inclusion of multiple polarizations was not statistically significant. For P-band, the model combining HV- and HH-polarized data was found to be better than the single polarized models. The multi polarized model had an R^2 (adjusted) of about 0.8, $RMSE_c$ of 29–30 t/ha (22–24% of mean biomass) and $RMSE_v$ of 35–40 t/ha (19–22% of mean biomass).

In conclusion, P-band polarized backscatter has been found to be better than L-band for biomass estimation. For L-band, the HV-polarization gave better results than HH-polarization, while for P-band the difference between the polarizations was small. The RMSE on stand level for P-band was about 18–27% of the mean biomass for stands with biomass ranging from 10 to 290 t/ha, and R^2 (adjusted) was found to be 0.7–0.8.

9.3. Validity of results

The results obtained in this paper hold for data acquired under similar conditions. The investigated forest is classified as hemiboreal. The general conclusions are likely to hold for forest types similar to this, *i.e.* boreal and possibly temperate forest dominated by coniferous species. However, extrapolation to tropical forest should not be done. The investigated test site is also fairly flat. For more topographic areas the difference between HV- and HH-polarizations is expected to be larger, since HH-polarization is more sensitive to topography than HV-polarization (Beaudoin et al., 1994; Hallberg et al., 2008). The data were acquired during the spring, and the soil moisture conditions were neither very dry or wet. For other weather conditions, especially during winter time with frozen conditions, the conclusions from this paper might be invalid. In terms of spatial scale, this paper focused on stand level measurements for stands larger than 0.5 ha. On smaller scales the results were found to be unstable. For areas much larger than typical for a stand, averaging must be done across stand borders, which might also influence the results. However, since the proposed regression models are linear, it should be possible to extrapolate the results to larger homogeneous areas.

Acknowledgments

The authors would like to thank the Swedish National Space Board for the main funding of this research, ESA for funding the BioSAR-I campaign and DLR for providing calibrated and geocoded SAR images. We would also like to thank Associate Professor Sören Holm and Associate Professor Hans Petersson at SLU in Umeå for input regarding biomass estimation using *in-situ* data and errors in these estimates.

References

- Ahti, T., Hämet-Ahti, L., & Jalas, J. (1968). Vegetation zones and their sections in northwestern Europe. *Annales Botanici Fennici*, 5, 169–211.
- Axelsson, P. (1999). Processing of laser scanner data — Algorithms and applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54, 138–147.
- Axelsson, P. (2000). DEM generation from laser scanner data using adaptive TIN models. *International Archives of Photogrammetry and Remote Sensing*, 33, 110–117.
- Baker, J. R., Mitchell, P. L., Cordey, R. A., Groom, G., Settle, J. J., & Stileman, M. R. (1994). Relationships between physical characteristics and polarimetric radar backscatter for Corsian pine stands in Thetford forest, U.K. *International Journal of Remote Sensing*, 15(14), 2827–2849.
- Beaudoin, A., Toan, T. L., Goze, S., Nezry, E., Lopes, A., Mougin, E., et al. (1994). Retrieval of forest biomass from SAR data. *International Journal of Remote Sensing*, 15, 2777–2796.
- Bezy, J. -L., Bensi, P., Lin, C., Durand, Y., Heliere, F., Regan, A., et al. (2007, July 23–27). ESA future earth observation explorer missions. *Proc. IEEE international geoscience and remote sensing symposium IGARSS, sensing and understanding our planet. Barcelona, Spain* (pp. 212–215).
- Draper, N. R., & Smith, H. (1998). *Applied regression analysis third edition*. John Wiley & Sons Inc.
- Fransson, J. E. S., & Israelsson, H. (1999). Estimation of stem volume in boreal forest using ERS-1 C- and JERS-1 L-band SAR data. *International Journal of Remote Sensing*, 20, 123–137.
- Hajsek, I., Scheiber, R., Ulander, L., Gustavsson, A., Sandberg, G., Tebaldini, S., et al. (2008). *BioSAR 2007 technical assistance for the development of airborne SAR and geophysical measurements during the BioSAR 2007 experiment: Final report without synthesis*. Contract no.: 20755/07/nl/cb. European Space Agency.
- Hallberg, B., Smith-Jonforsen, G., Ulander, L., & Sandberg, G. (2008). A physical-optics model for double-bounce scattering from tree stems standing on an undulating ground surface. *IEEE Transactions on Geoscience and Remote Sensing*, 46(9), 2607–2621.
- Hoekman, D., & Quiriones, M. (2000). Land cover type and biomass classification using AirSAR data for evaluation of monitoring scenarios in the Colombian Amazon. *IEEE Transactions on Geoscience and Remote Sensing*, 38(2), 685–696.
- Holmgren, J. (2004). Prediction of tree height, basal area and stem volume in forest stands using airborne laser scanning. *Scandinavian Journal of Forest Research*, 19(6), 543–553.
- Imhoff, M. (1995). Radar backscatter and biomass saturation: ramifications for global biomass inventory. *IEEE Transactions on Geoscience and Remote Sensing*, 33(2), 511–518.
- IPPC (2007). *Climate change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.
- Israelsson, H., Askne, J., & Sylvander, R. (1994). Potential of SAR for forest bole volume estimation. *International Journal of Remote Sensing*, 15(14), 2809–2826.
- Jalkanen, A., Mäkipää, R., Ståhl, G., Lehtonen, A., & Petersson, H. (2005). Estimation of the biomass stock of trees in Sweden: Comparison of biomass equations and age-dependent biomass expansion factors. *Annals of Forest Science*, 62, 845–851.
- Jonsson, B., Jacobsson, J., & Kallur, H. (1993). The forest management planning package. Theory and application. *Studia Forestalia Suecica*, 189, 1–56.
- Le Toan, T., Beaudoin, A., Riou, J., & Guyon, D. (1992). Relating forest biomass to SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 30(2), 403–411.
- Le Toan, T., Quegan, S., Davidson, M., Balzer, H., Paillou, P., Papathanassiou, K., et al. (this issue, March). The BIOMASS Mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. Remote sensing of environment.
- Lefsky, M. A., Cohen, W. B., Acker, S. A., Parker, G. G., Spies, T. A., & Harding, D. (1999). Lidar remote sensing of the canopy structure and biophysical properties of Douglas-fir Western hemlock forest. *Remote Sensing of Environment*, 70, 339–361.
- Magnusson, M., Fransson, J. E. S., & Holmgren, J. (2007). Effects on estimation accuracy of forest variables using different pulse density of laser data. *Forest Science*, 53(6), 619–626.
- Marklund, L. G. (1988). *Biomassfunktioner för tall, gran och björk i Sverige. Rapport 45*. Umeå, Sweden: Institutionen för skogstaxering, Sveriges lantbruksuniversitet in Swedish.
- Means, J. E., Acker, S. A., Fitt, B. J., Renslow, M., Emerson, L., & Hendrix, C. J. (2000). Predicting forest stand characteristics with airborne scanning lidar. *Photogrammetric Engineering and Remote Sensing*, 66(11), 1367–1371.
- Næsset, E. (2002). Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment*, 80(1), 88–99.
- Näslund, M. (1947). Funktioner och tabeller för kubering av stående träd. Tall, gran och björk i södra Sverige samt hela landet. *Meddelande från Statens skogsforskningsinstitut*, 36(3) in Swedish.
- Ranson, K., & Sun, G. (1994). Mapping biomass of a northern forest using multifrequency SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 32(2), 388–396.
- Rauste, Y., Häma, T., ans K., Heiska, J. P., & Hallikainen, M. (1994). Radar-based forest biomass estimation. *International Journal of Remote Sensing*, 15, 1797–2808.
- Rignot, E., Zimmermann, R., & van Zyl, J. (1995). Spaceborne applications of P band imaging radars for measuring forest biomass. *IEEE Transactions on Geoscience and Remote Sensing*, 33(5), 1162–1169.
- Rosengqvist, A., Shimada, M., Ito, N., & Watanabe, M. (2007). ALOS PALSAR: A pathfinder mission for global-scale monitoring of the environment. *IEEE Transactions on Geoscience and Remote Sensing*, 45(11), 3307–3316.
- Saatchi, S., Halligan, K., Despain, D., & Crabtree, R. (2007). Estimation of forest fuel load from radar remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 45(6), 1726–1740.
- Shimada, M., Isoguchi, O., Tadono, T., & Isono, K. (2009). PALSAR radiometric and geometric calibration. *IEEE Transactions on Geoscience and Remote Sensing*, 47(12), 3915–3932.
- Soininen, A. (2004). *TerraScan for microstation. User's guide*. Jyväskylä, Finland: Terrasolid Ltd.
- Ulaby, F. T., Moore, R. K., & Fung, A. K. (1982). *Microwave remote sensing — Active and passive, volume II*, Addison-Wesley publishing company.
- Ulander, L. M. H. (1996). Radiometric slope correction of synthetic-aperture radar images. *IEEE Transactions on Geoscience and Remote Sensing*, 34(5), 1115–1122.