Combining Remotely Sensed Optical and Radar Data in *k*NN-Estimation of Forest Variables

Hampus Holmström and Johan E.S. Fransson

ABSTRACT. The use of optical and radar data for estimation of forest variables has been investigated and evaluated by employing the *k* nearest neighbor (*k*NN) method. The investigation was performed at a test site located in the south of Sweden consisting mainly of Norway spruce and Scots pine forests with standwise stem volume in the range of 0–430 m^3 ha^{-1}. The *k*NN method imputes weighted reference plot variables to areas to be estimated (target areas), facilitating further use of data in forestry planning models. Remotely sensed multispectral optical data from the SPOT-4 XS satellite and radar data from the airborne CARABAS-II VHF SAR sensor were used, separately and combined, to define weights in the *k*NN algorithm. The weights were inversely proportional to the image feature distance between the reference plot and the target area. The distance metric was defined using regression models based on the image data sources. Positive impact on the accuracies of stem volume and age estimates was found by combining the two image data sources. Stem volume, at stand level, was estimated with a *RMSE* of 37 m^3 ha^{-1} (22% of the true mean value) using the combination of optical and radar data, compared to 50 m^3 ha^{-1} (30%) for the best single-sensor case in this study. In conclusion, the results indicate that the accuracv of forest variable estimations was substantially improved by using multisensor data.

Key Words: Data assessment, forest inventory, imputation, remote sensing.

DATA WITH HIGH ACCURACY and high spatial resolution acquired by forest inventory methods are crucial in forest management planning. In Scandinavia, the traditional methods for collecting standwise forest data over large areas are based on manual intervention and interpretation. As a consequence, the cost tends to be high, and the data accuracy depends on subjective judgments (Ståhl 1992). The forest data collection can be performed in different ways; however, it is desirable that the cost of capturing data should be gained when decisions are made based on the information at hand (e.g., Schreuder et al. 1993, p. 62–106). Decisions will seldom be made from only a single forest variable but from several. Even though stem volume is the most important variable, according to major Swedish forest companies (Walter 1998), planning models need additional input data, such as age and tree species composition. There is also a need for global mapping of forests in support of international climate change treaties such as the Kyoto protocol. New, cost-effective inventory methods need to be developed to meet the requirements of forest companies and environmental agencies.

Field measurements in combination with remotely sensed data are of interest for forest inventory, offering possibilities to use accurate field data together with full image coverage from high-resolution satellite and/or airborne sensors. This allows for the development of forest applications, based on objective and cost-effective methods.

Much work is still left to be done before optical satellite data can be used for forest variable estimation at stand level with accuracy suitable for operational forest mapping and management planning (Holmgren and Thuresson 1998). However, improvements of the estimation accuracies are possible by combining information from different sensors (e.g., Cohen and Spies 1992) and by adding some ancillary data source, such as tree height information (e.g., Nilsson 1997) and existing stand record information (e.g., Tomppo et al. 1999, Holmgren et al. 2000).

The use of radar remote sensing for forest inventory has been an area of active research since the 1960s (e.g., Leckie and Ranson 1998, p. 435–510.). Providing its own illumination source, the radar signal is capable of penetrating to the vegetation regardless of sunlight and cloud cover. This is a great advantage compared with optical sensors, which depend on passive reflected solar radiation and relatively clear skies. However, the potential of using data from microwave remote sensing systems (operating at wavelengths of decimeters) for biomass retrieval is severely limited mainly due to signal saturation (Imhoff 1995). The saturation problem also occurs in optical images when the forest canopy is fully closed, corresponding to a stem volume of about 250 m^3 ha^{-1} in managed forests.

In recent years, promising results have been shown relating forest variables to radar data from the Swedish airborne CARABAS-II VHF SAR (Synthetic Aperture Radar) sensor (Hellsten et al. 1996), operating at very long wavelengths. The results show that the radar response is linearly related to standwise stem volume and that the accuracy of estimates from denser forests, located on relatively horizontal ground, are comparable to the accuracy from subjective ground-based forest inventory accuracies (Fransson et al. 2000b). The linear relationship between radar response and stem volume has also been shown by physical modeling (Smith and Ulander 2000). In contrast to radar data from microwave systems and optical sensors, no saturation of the radar response is found up to about 1,000 m^3 ha^{-1} (Ulander et al. 2000). However, the radar response from stands with stem volume less than about 80 m^3 ha^{-1} is close to the system noise level, which constitutes a limitation when using CARABAS-II radar data. Hence, remotely sensed data from optical and radar sensors contain complementary information. By combining data from the two sources, an improvement in estimation of forest variables may be expected.

A novel way of integrating multisource data is to use the *k* nearest neighbor (*k*NN) estimation method. This method has been used in the Finnish national forest inventory (NFI) since 1990 (Tomppo 1990), where field and optical satellite data are integrated. Areas known only by their spectral signatures in the satellite image are allotted field data values as weighted mean values of variables from the *k* nearest field plots; nearness is measured in a feature space defined by the different spectral wavelength bands of the satellite image. Other types of information can also be applied and combined with remotely sensed data for use in *k*NN-estimations, such as mean values from existing stand registers (e.g., age, site index, and silvicultural history data), terrain data (e.g., slope and aspect), and geographical distances. The *k*NN method simultaneously provides estimates for the entire suite of variables available at the reference plots. The natural covariance structure among these plot level variables is hence preserved, a valuable property if the estimates are to be used as input data in a planning system (e.g., Moeur and Stage 1995). Another advantage of the *k*NN method is the simplicity with which new sources of information can be used to strengthen the association between reference plots and areas to be allotted forest data. Information used in the phase of imputation (allotment) of plot data is referred to here as carrier data; as to “carry” the data available at the reference plots to the areas to be estimated. The carrier data information is available at both reference plots (whose true forest states are assumed to be known) and target plots. Here, a target plot denotes the area to which reference data are to be imputed, i.e., where associated variables are to be estimated. Although plot data will be imputed rather than inventoried in the field, they will appear in a format that readily can be used in a traditional planning system.

The objective of this study is to evaluate the accuracy of *k*NN-estimation of forest variables, at plot and stand level, when remotely sensed optical and radar data are used both separately and combined.

Material and Methods

Test Site

The test site, Remningstorp, is located in the southwest of Sweden (58°30’N, 13°40’E) with a ground elevation moderately varying between 120 and 145 m above sea level. The estate covers 1,200 ha of forestland (potential annual productivity ≥1 m^3 ha^{-1}) and consists of 340 forest stands. About 10% of the area is forested peatland. The dominant tree species are Norway spruce (*Picea abies*), Scots pine (*Pinus sylvestris*), and birch (*Betula* spp.). A few stands dominated by oak (*Quercus robur*) and beech (*Fagus silvatica*), and some minor areas with field experiments, were excluded from the study. The dominant soil type is till with a field layer consisting of blueberry (*Vaccinium myrtillus*) and cowberry (*Vaccinium vitis-idaea*). Further descriptions of the Remningstorp estate are presented by Ahlberg and Kardell (1997). The Forestry Society’s Estate Management Company provided a stand register with associated forest map covering the test site.

Field Data

Two objective field inventories, performed at the Remningstorp estate, provided reference and evaluation data. The inventories were conducted during 1997/1998 and 1998/1999 and included: (1) 230 plots systematically sampled in a grid covering parts of the test area, with 100 m between plots along lines and 200 m between the lines (all during 1997/1998), and (2) 541 plots systematically sampled in 47 stands (155 plots in 13 stands during 1997/1998 and 386 plots in 34 stands during 1998/1999). The stands were sampled with probability proportional to size (PPS) given in the stand register. Originally, 814 plots were inventoried. Plots affected by forestry activities (clearcutting, thinning, etc.), carried out between field data collection and acquisition of remote sensing data, were removed from the dataset. Thus, data from a total of 771 plots formed the reference material (Table 1). The 47 stands, used as evaluation objects (i.e., target stands in the *k*NN-estimations), were inventoried with 6–14 plots per stand (on average 11.5), systematically sampled in a grid with square lattice (randomly positioned). The target stands are also described in Table 1. The field inventories were made according to the instructions of the Forest Management Planning Package, FMPP (Jonsson et al. 1993), using a plot radius of 10 m. Tree species and diameter at breast height were recorded for all trees on the plots, and additional measurements of tree height, age, etc., were made on a subsample of the trees (chosen with probability proportional to basal area). Other plot data were also collected, such as site index and soil and vegetation type. From the field inventory data, standing stem volumes at plot and stand level were calculated using single-tree models (Söderberg 1986), as parts of the FMPP.

All plot centers were positioned using the Global Positioning System (GPS). An internal base station and a six-channel field receiver, both observing course/acquisition (C/A) code and carrier phase, were used. GPS data were continuously collected at each plot for 10–15 minutes, corresponding approximately to the time of field data collection. To estimate plot center coordinates, GPS data were postprocessed using the carrier phase measurements. Submeter accuracies in the horizontal plane were achieved for most of the field plot coordinates (Holmström et al. 2001).

Optical Data

The test site was covered by one multispectral optical SPOT-4 XS satellite image, acquired July 10, 1999. The onboard HRVIR (High Resolution Visible and Infrared) passive sensor gives spectral signatures in 8-bit data format (represented by 0–255 digital numbers, *DN*s) in four different wavelength bands; 0.50–0.59 μm (green), 0.61–0.68 μm (red), 0.79–0.89 μm (near-infrared), and 1.58–1.75 μm (mid-infrared). In forests, the registered spectral signatures are mainly due to sunlight reflected by the upper forest canopy. The image was geometrically corrected (utilizing orbital parameters, ground control points, and a digital elevation model, DEM) to the Swedish National Grid by the Swedish Space Corporation. The ground resolution (and pixel size) is 20×20 m. The local image geometry within the test site was evaluated using digital ortho-photographs. No deviations larger than half a pixel could be identified, thus no further corrections were performed. In relating image data to field plot data, the plot center coordinates were used to extract *DN*s by utilizing cubic convolution. In Figure 1, plotwise log-transforms of the digital number in the four spectral bands are plotted versus the logarithm of stem volume, rendering correlation coefficients, $\rho = -0.752$, -0.671, -0.709, and -0.754, respectively.

Radar Data

The CARABAS-II VHF SAR flight campaign was conducted in June 1999. The SAR sensor operates in the VHF-band, with wavelengths between 3.3 and 15 m (Hellsten et al. 1996). The long wavelengths make the CARABAS-II sensor unique among SAR sensors worldwide and capable of penetrating into dense forests. The side-looking active radar sensor transmits and receives electromagnetic energy pulses, traveling at the speed of light. The backscattered energy (radar echo) from the forest is recorded onboard the aircraft, with a capacity of 2 km^2 s^{-1}, and later processed to SAR images. Characteristics that affect the radar backscattering from a forested area are the geometrical (shape, size, density, orientation, roughness), dielectric (permitivity), and sensor (wavelength, polarization, incidence angle) properties. In this study, radar data were acquired at an altitude of 3,600 m and processed to a SAR image with a pixel size of 1×1 m (the ground resolution is about 3×3 m). The extraction of radar data was made from the plots (radius = 10 m) and the stands (using the forest map and excluding a 10 m buffer zone from the stand borders) by using a fully automatic geocoding and extraction algorithm, which utilizes the recorded flight parameters and a DEM (Walter et al. 1999). Disturbing nonforest objects, such as power lines, buildings, and fences were visually identified and excluded from the analysis. The radar backscattering amplitude, s°, for each plot/stand was calculated from the radar image according to Fransson et al. (2000b). In Figure 2, plotwise radar backscattering amplitude are plotted versus stem volume, rendering a correlation coefficient, $\rho = 0.711$.

Table 1. Descriptive statistics of reference plot data, *n* = 771, and target stand data, *n* = 47.

	Reference plots			Target stands		
Variable	Ave	Min	Max	Ave	Min	Max
Area (ha)	—	—	—	5.0	0.6	19.3
Stem volume (m^3 ha^{-1})	171.0	0.0	743.2	172.4	0.0	426.0
Age (yr)	44.5	0.0	149.0	44.8	0.0	107.0
Tree species composition (%)						
Pine	18.3	0.0	100.0	17.8	0.0	98.2
Spruce	72.8	0.0	100.0	73.4	0.0	100.0
Broad-leaved	8.9	0.0	100.0	8.8	0.0	38.2

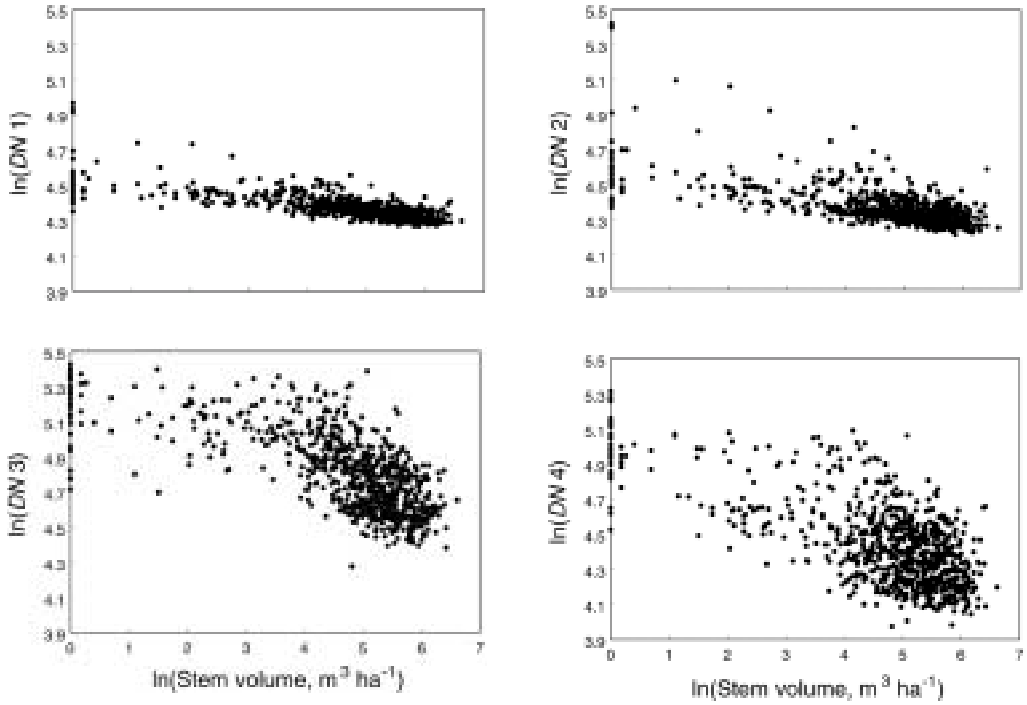

Figure 1. Plotwise logarithms of digital number (*DN*) from SPOT-4 XS in four spectral bands plotted versus logarithms of stem volume (m^3 ha^{-1}), *n* = 771.

Estimations

The choice of method was made to enable extended use of estimates in planning models requiring high-resolution input data (at individual tree level). The estimations were obtained using the *k* nearest neighbor (*k*NN) method (e.g., Muinonen and Tokola 1990, Tomppo 1990) as weighted means of the *k* nearest reference sample plot values, where nearness was measured in the feature space defined by the carrier data. Plotwise, multispectral optical data and radar backscattering amplitude were used as carrier data, both separately and combined. In this application, *k* was set to 10. The balance between preserving all natural variability in the reference data ($k = 1$) and minimizing the errors in the estimates ($k >> 1$) led to the choice of $k = 10$, supported by results in other studies (e.g., Nilsson 1997). Since entire reference plots actually were imputed to each target plot, all variables available at the reference plots (obtained from field measurements) were estimated simultaneously. In the *k*NN-estimations, the *k* nearest reference plots (indexed by *r*) were closest in the feature space to the target plot. A variable value, *v*, for target plot *t* was estimated as:

$$\hat{v}_t = \sum_{r=1}^{k} w_{t,r} v_r \tag{1}$$

where the weights, *w*, were set proportional to the inverse distance, *a*, between target plot *t* and reference plot *r* (cf. Isaaks and Srivastava 1989, p. 257–259):

$$w_{t,r} = \frac{\dfrac{1}{a_{t,r}}}{\displaystyle\sum_{r=1}^{k} \frac{1}{a_{t,r}}} \tag{2}$$

The distance *a* was derived using the carrier data in regression models (e.g., Tokola et al. 1996). The different carrier data variables were hence weighted according to the additional information supplied by a certain carrier data variable to explain the variable of interest. Such property becomes important when using several carrier data variables of different qualities in combination. Regression models were used to estimate forest variables, *Y's*, for all plots and the regression-based distance, $a_{t,r}$, between target plot *t* and reference plot *r* was defined as:

$$a_{t,r} = \left| \hat{Y}_t - \hat{Y}_r \right| \tag{3}$$

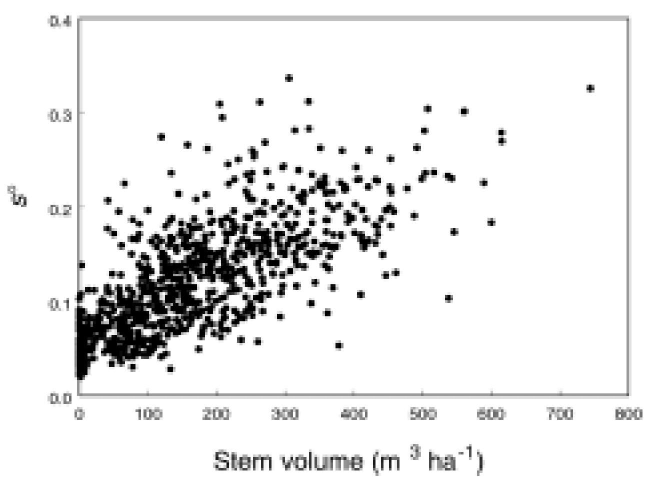

Figure 2. Plotwise radar backscattering amplitude, *s*°, from CARABAS-II radar data plotted versus stem volume (m^3 ha^{-1}), *n* = 771.

With Y set to stem volume (m^3 ha^{-1}), age (yr), and proportion of conifers (%), respectively, regression models were developed using the plot data and the m carrier data variables, c, as independent variables, i.e., $\hat{Y} = f\left(c_1, c_2, ..., c_m\right)$. The different regression-based distances (with different Y's) were used in correspondence with the variable to be evaluated. The following general multiplicative regression model was used:

$$Y = e^{\alpha} \cdot c_1^{\beta_1} \cdot c_2^{\beta_2} \cdot \ldots \cdot c_m^{\beta_m} \cdot \varepsilon \tag{4}$$

where linearity is obtained by logarithmic transformation:

$$\ln Y = \alpha + \beta_1 \cdot \ln c_1 + \beta_2 \cdot \ln c_2 + \ldots + \beta_m \cdot \ln c_m + \ln \varepsilon \tag{5}$$

This was valid for all cases, except when using only radar data, where the following simple linear model was used:

$$Y = \alpha + \beta \cdot c + \varepsilon \tag{6}$$

The choice of multiplicative models when using optical data was supported by results in earlier studies (e.g., Hagner 1990, Ripple et al. 1991, Ardö 1992, Cohen and Spies 1992, Nilsson 1997) and analyses of residual plots in the present study. The choice of an additive model when using radar data follows from Smith and Ulander (2000). Coefficients for the regression models were estimated using ordinary least squares (OLS). The estimates and the coefficients of determination, R^2, are presented in Table 2.

Evaluations

Imputation of plot data with the kNN method enables simultaneous estimation of all variables available at the reference plots. The evaluation focuses on estimates of three important variables, from a forest management planning point of view: stem volume (m^3 ha^{-1}), age (yr), and tree species composition (i.e., proportion of conifers, %). Evaluations were made both at plot and stand level. The variables of the 541 target plots belonging to the 47 target stands were estimated. When estimating variable values for a specific plot, all other plots except those belonging to the same stand as the target plot were used as reference plots. Estimations at the stand level were made by calculating the arithmetic mean values of the target plot estimates for the specific stand. The difference between estimated and true variable values (calculated from the field inventory data), $\Delta_i = \hat{v}_i - v_i$, for plot/stand i was used to calculate the standard deviation, Std_Δ, and the average error (i.e., empirical bias), $\overline{\Delta}$:

$$Std_\Delta = \sqrt{\frac{\sum_{i=1}^{n}\left(\Delta_i - \overline{\Delta}\right)^2}{n-1}} \tag{7}$$

$$\overline{\Delta} = \frac{\sum_{i=1}^{n} \Delta_i}{n} \tag{8}$$

where n is the number of plots/stands. The root mean square error, $RMSE$, was calculated as:

$$RMSE = \sqrt{Std_\Delta^2 + \overline{\Delta}^2} \tag{9}$$

Relative errors (given in percentage, %) were calculated in relation to the true mean values, obtained from the field measurements. Evaluations were made for all target plots,

Table 2. Estimated regression coefficients and coefficients of determination, R^2, for the models estimating stem volume, age, and proportion of conifers, using SPOT-4 XS optical data, *DN*, in four spectral bands, CARABAS-II radar backscattering amplitude, S°, and both optical and radar data, n = 771. Level of significance is denoted by: *P = <0.05, **P = <0.01, ***P = <0.001.

	Coefficient		
Parameter	Optical	Radar	Optical and radar
Y = stem volume (m^3 ha^{-1})			
Constant	62.22***	–23.85**	48.23***
ln(*DN* 1)	–10.22***	—	–6.512**
ln(*DN* 2)	1.566	—	0.648
ln(*DN* 3)	–2.140***	—	–1.519***
ln(*DN* 4)	–2.157***	—	–2.589***
s°, ln(s°)	—	1,548***	6.055***
R^2 (%)	69.2	50.6	71.1
Y = age (yr)			
Constant	26.12***	10.66***	21.81***
ln(*DN* 1)	–1.540	—	–0.396
ln(*DN* 2)	–1.779**	—	–2.062**
ln(*DN* 3)	–2.341***	—	–2.149***
ln(*DN* 4)	0.732***	—	0.599***
s°, ln(s°)	—	269.0***	1.867***
R^2 (%)	65.8	37.1	66.7
Y = conifer proportion (%)			
Constant	16.68***	57.10***	15.00**
ln(*DN* 1)	3.282	—	3.727
ln(*DN* 2)	–1.664	—	–1.774
ln(*DN* 3)	–1.373***	—	–1.299***
ln(*DN* 4)	–2.950***	—	–3.002***
s°, ln(s°)	—	177.3***	0.726
R^2 (%)	47.1	11.0	47.1

target stands, and for two subdivisions of the target stand dataset: (1) class 1 representing young forests (n = 23) with average stem volume equal to 73.6 m^3 ha^{-1} (min = 0, max = 164.2), and (2) class 2 representing old forests (n = 24) with average stem volume equal to 268.3 m^3 ha^{-1} (min = 165.2, max = 426.0).

As described above, ten reference plots (k = 10) were imputed to each field inventoried target plot within a specific stand, and on average, 11.5 plots were inventoried in each target stand. Both optical and radar data were available, giving complete coverage of the target stands. It should be noted that a field inventory plot covers an area of 314 m^2. The 47 stands have a total area of 235 ha, and the 541 target plots have a total area of approximately 17 ha. Thus, only about 7% of the available remotely sensed information in the target stands was used in the procedure described above. This scheme enables plotwise evaluations and comparisons with results from previous studies (e.g., when using plotwise aerial photograph interpretations as carrier data, see Holmström et al. 2001). As a following step, all available optical and radar information within the target stands were used. The area within each stand was split into cells with a 20 × 20 m grid. Optical and radar data were extracted for each cell in the same manner as earlier. Also, the kNN-estimation procedure was the same, except now it imputed the ten nearest reference plots to each cell within each target stand. Standwise mean values were then estimated by calculating the arithmetic mean value of the cellwise estimates within each stand. Variables for a total of 4,368 cells within 46 stands were estimated (one of the stands partly outside the radar image was removed from the target dataset), with an average of 95 cells per stand (min = 7, max = 408). Here, it was legitimate to compensate for sampling errors in true data (the field inventories), in contrast to the case when imputing reference data to the same plots as those used to calculate true states. Since the target stands were inventoried using a systematic plot sample design, half of the sampling variances were subtracted from the calculated, empirical variances (e.g., Lindgren 1984). In general, this compensation had only a minor impact on the estimation accuracies, decreasing the Std_Δ with 0.2–1.8%.

In a final evaluation, estimates of stem volume for a subset of 31 target stands (from the original 47) were further studied. These stands were characterized by covering at least 50 cells of 20 × 20 m (i.e., a minimum area of 2 ha). On average, the area of the stands was 5.1 ha, with an average stem volume of 147.4 m^3 ha^{-1} (min = 0, max = 426.0). The stem volume of each cell was estimated using the kNN method described in previous sections. The standwise stem volume was then estimated by averaging a specific number of cell-estimates in each target stand. The sample size n_s (number of cellwise estimates used) was varied from 1 up to 50 for each stand (randomly sampled with replacement). In a practical application using carrier data with full coverage for the target stands, it is natural to use all information available. However, by increasing the sampling intensity within the stands and observing the errors at different levels, a rough measure of the spatial correlation in the errors for optical and radar data can be given. Independence between cellwise stem volume estimates would decrease the Std_Δ at the stand level with a factor $n_s^{-0.5}$ when increasing the sampling intensity (Thompson 1992, p. 11–25).

Results

Estimations of Forest Variables Using Optical Data

The estimations were made with Y as stem volume (m^3 ha^{-1}), age (yr), and proportion of conifers (%), respectively, when calculating the regression-based distances. Results for the entire target dataset (plot and stand level) as well as for subdivisions of the dataset (stand level) are given in Table 3. The errors in the stem volume estimates at stand level were approximately 50% smaller than the errors at plot level, while age and conifer proportion estimates decreased by about one third. At stand level, the *RMSE*s in the stem volume, age, and conifer proportion estimates were found to be about 58 m^3 ha^{-1} (34%), 14 yrs (30%), and 17% (21%), respectively. Substantial systematic errors were found in the stem volume estimates, where volumes in young forests were overestimated with about 33 m^3– ha^{-1} and a corresponding underestimation in old forests.

Table 3. Standard deviation, Std_Δ, average error, $\bar{\Delta}$, and root mean square error, *RMSE*, in absolute terms and relative terms (in percentage within parentheses), from *k*NN-estimations of forest variables using SPOT-4 XS optical data, *n* = 541 (plot level), 47 (stand level), 23 (stand level, class 1), and 24 (stand level, class 2).

Forest variable	Std_Δ	$\bar{\Delta}$	*RMSE*
Total target dataset			
Stem volume (m^3 ha^{-1}), plot level	107.7 (64.3%)	0.9 (0.5%)	107.7 (64.3%)
Stem volume (m^3 ha^{-1}), stand level	58.0 (33.5%)	–1.0 (–0.6%)	58.0 (33.5%)
Age (yr), plot level	18.9 (43.0%)	–1.2 (–2.8%)	19.0 (43.1%)
Age (yr), stand level	13.4 (30.0%)	–1.5 (–3.3%)	13.5 (30.2%)
Conifer proportion (%), plot level	26.1 (33.4%)	0.6 (0.8%)	26.2 (33.4%)
Conifer proportion (%), stand level	17.2 (21.2%)	–1.3 (–1.5%)	17.2 (21.3%)
Class 1 (young forests)			
Stem volume (m^3 ha^{-1}), stand level	30.5 (41.4%)	32.9 (44.7%)	44.9 (61.0%)
Age (yr), stand level	14.9 (48.6%)	–0.4 (–1.3%)	14.9 (48.6%)
Conifer proportion (%), stand level	23.1 (33.3%)	–0.5 (–0.7%)	23.1 (33.3%)
Class 2 (old forests)			
Stem volume (m^3 ha^{-1}), stand level	59.4 (22.1%)	–33.5 (–12.5%)	68.2 (25.4%)
Age (yr), stand level	11.8 (20.2%)	–2.5 (–4.3%)	12.1 (20.7%)
Conifer proportion (%), stand level	7.8 (8.5%)	–2.0 (–2.2 %)	8.1 (8.8%)

Table 4. Standard deviation, Std_Δ, average error, $\overline{\Delta}$, and root mean square error, *RMSE*, in absolute terms and relative terms (in percentage within parentheses), from *k*NN-estimations of forest variables using CARABAS-II radar data, n = 541 (plot level), 47 (stand level), 23 (stand level, class 1), and 24 (stand level, class 2).

Forest variable	Std_Δ	$\overline{\Delta}$	*RMSE*
Total target dataset			
Stem volume (m^3 ha^{-1}), plot level	109.3 (65.3%)	2.6 (1.6%)	109.3 (65.3%)
Stem volume (m^3 ha^{-1}), stand level	72.6 (41.9%)	–1.5 (–0.9%)	72.6 (41.9%)
Age (yr), plot level	23.2 (52.7%)	0.1 (0.2%)	23.2 (52.7%)
Age (yr), stand level	16.9 (37.7%)	0.0 (0.0%)	16.9 (37.7%)
Conifer proportion (%), plot level	35.8 (45.8%)	2.1 (2.7%)	35.9 (45.9%)
Conifer proportion (%), stand level	26.9 (33.2%)	–0.1 (–0.2%)	26.9 (33.2%)
Class 1 (young forests)			
Stem volume (m^3 ha^{-1}), stand level	51.9 (70.4%)	38.2 (51.9%)	64.4 (87.5%)
Age (yr), stand level	19.5 (63.3%)	1.9 (6.2%)	19.5 (63.9%)
Conifer proportion (%), stand level	36.0 (51.8%)	3.8 (5.4%)	36.2 (52.1%)
Class 2 (old forests)			
Stem volume (m^3 ha^{-1}), stand level	69.1 (25.7%)	–39.6 (–14.8%)	79.6 (29.7%)
Age (yr), stand level	13.7 (23.5%)	–1.9 (–3.2%)	13.8 (23.7%)
Conifer proportion (%), stand level	12.0 (13.1%)	–3.9 (–4.2%)	12.6 (13.7%)

Estimations of Forest Variables Using Radar Data

The estimation procedure is as in the previous section with Y as stem volume (m^3 ha^{-1}), age (yr), and proportion of conifers (%), respectively. Results for the entire target dataset (plot and stand level) as well as for subdivisions of the dataset (stand level) are given in Table 4. The errors in the stem volume estimates at stand level were here only 36% smaller than the errors at plot level. At stand level, the *RMSE*s in the stem volume, age, and conifer proportion estimates were found to be about 73 m^3 ha^{-1} (42%), 17 yr (38%), and 27% (33%), respectively. Compared with the case when using only optical data, the systematic errors were slightly more pronounced in the classwise stem volume estimates (overestimation in young forests and underestimation in old forests).

Estimations of Forest Variables Using Optical and Radar Data in Combination

The estimations follow the procedure in previous sections, and results are given in Table 5. Here, the estimates are substantially more accurate than the results presented in both Table 3 and Table 4, where the remote sensing data sources were used separately. The *RMSE*s in the stem volume estimates, at stand level, decreased by 19% (11 m^3 ha^{-1}) and 36% (26 m^3 ha^{-1}) when optical and radar data were used together as carrier data in the *k*NN-estimations, in comparison with using the sources separately. Also the age estimates were improved by combining data from the two sources, showing relative *RMSE*s of 25%. However, the conifer proportion estimates when using only optical data was not improved by adding radar data, except for a minor improvement in old forests. Although still showing systematic errors for the stem volume estimates in the two forest classes, the combination of optical and radar data made these errors less pronounced.

Estimates for all 20 × 20 m cells within the target stands, using the total coverage of remote sensing data, are presented in Table 6. The errors decreased, compared to the case when a sample of plots within the stands was used, but only to a minor extent. However, in the stem volume estimates using optical and radar data combined, the *RMSE* decreased from 47 m^3 ha^{-1} to 37 m^3 ha^{-1} (22% of the true mean value). Overall, the systematic errors were relatively small.

The impact of increasing numbers of sampled observations within target stands is exemplified in Figure 3. As expected, the random errors of the standwise stem volume estimates did not decrease by a factor $n_s^{-0.5}$ after averaging n_s cellwise estimations. The spatial correlation of the errors in both optical and radar data here becomes obvious.

Table 5. Standard deviation, Std_Δ, average error, $\overline{\Delta}$, and root mean square error, *RMSE*, in absolute terms and relative terms (in percentage within parentheses), from *k*NN-estimations of forest variables using SPOT-4 XS optical and CARABAS-II radar data, n = 541 (plot level), 47 (stand level), 23 (stand level, class 1), and 24 (stand level, class 2).

Forest variable	Std_Δ	$\overline{\Delta}$	*RMSE*
Total target dataset			
Stem volume (m^3 ha^{-1}), plot level	88.3 (52.7%)	–0.4 (–0.2%)	88.3 (52.7%)
Stem volume (m^3 ha^{-1}), stand level	46.7 (27.0%)	–2.8 (–1.6%)	46.8 (27.1%)
Age (yr), plot level	16.4 (37.1%)	–0.7 (–1.5%)	16.4 (37.2%)
Age (yr), stand level	11.2 (25.1%)	–1.2 (–2.6%)	11.3 (25.2%)
Conifer proportion (%), plot level	25.1 (32.0%)	0.9 (1.2%)	25.1 (32.0%)
Conifer proportion (%), stand level	17.6 (21.8%)	–1.0 (–1.3%)	17.6 (21.8%)
Class 1 (young forests)			
Stem volume (m^3 ha^{-1}), stand level	29.0 (39.4%)	17.9 (24.4%)	34.1 (46.3%)
Age (yr), stand level	13.7 (45.0%)	0.1 (0.2%)	13.7 (45.0%)
Conifer proportion (%), stand level	24.4 (35.1%)	–1.1 (–1.5%)	24.4 (35.1%)
Class 2 (old forests)			
Stem volume (m^3 ha^{-1}), stand level	51.7 (19.3%)	–22.6 (–8.4%)	56.4 (21.0%)
Age (yr), stand level	7.9 (13.6%)	–2.4 (–4.1%)	8.3 (14.2%)
Conifer proportion (%), stand level	6.2 (6.7%)	–1.0 (–1.1%)	6.2 (6.8%)

Table 6. **Standard deviation, Std_Δ, average error, $\bar{\Delta}$, and root mean square error, *RMSE*, in absolute terms and relative terms (in percentage within parentheses), from *k*NN-estimations of forest variables using all available SPOT-4 XS optical or/and CARABAS-II radar data, n = 46 (stand level).**

Forest variable	Std_Δ	$\bar{\Delta}$	*RMSE*
Optical data			
Stem volume ($m^3\ ha^{-1}$), stand level	50.1 (29.6%)	6.7 (4.0%)	50.5 (29.9%)
Age (yr), stand level	13.3 (29.6%)	–0.7 (–1.6%)	13.3 (29.6%)
Conifer proportion (%), stand level	16.2 (20.2%)	–1.5 (–1.9%)	16.3 (20.2%)
Radar data			
Stem volume ($m^3\ ha^{-1}$), stand level	64.6 (38.2%)	2.2 (1.3%)	64.6 (38.2%)
Age (yr), stand level	16.9 (37.5%)	–0.9 (–2.0%)	16.9 (37.6%)
Conifer proportion (%), stand level	25.8 (32.1%)	–1.1 (–1.4%)	25.8 (32.1%)
Optical and radar data			
Stem volume ($m^3\ ha^{-1}$), stand level	36.9 (21.8%)	2.8 (1.7%)	37.0 (21.9%)
Age (yr), stand level	11.2 (24.9%)	–0.5 (–1.2%)	11.2 (24.9%)
Conifer proportion (%), stand level	16.3 (20.3%)	–1.2 (–1.5%)	16.3 (20.3%)

Figure 3 shows, graphs of the relative standard deviation plotted versus the sample size, are augmented with corresponding fitted regression functions. These functions indicate an exponent for n_s of approximately –0.1, rather than –0.5.

Discussion

The *k*NN method, sometimes called the reference sample plot (RSP) method, utilizing multisource data is applicable for estimation of forest variables in support of forest management planning. A valuable property of the method is that the natural covariance relationship between the estimated forest variables is preserved. In a study by Moeur and Stage (1995), the advantages of using a nearest neighbor method (with k = 1) were described, and the results showed the method's ability to closely reproduce the covariance structure between different variables. In studies by Nilsson (1997), the estimation errors were shown to decrease with increasing k. However, more than ten reference plots imputed only improved the estimates marginally. According to Tomppo et al. (1999), too many nearest neighbors reduce the natural spatial variation of the estimates and results in too smoothed descriptions. The ability to estimate forest characteristics with the full range of variation and correlations among between variables might, in some planning situations, be as important as the accuracy of the variable estimates. To keep the covariance structure is of special importance when data are used in growth models based on real single-tree observations, such as in models dealing with biodiversity-related and uneven-aged management, i.e., selective cutting (Wikström 2000). From the above reasoning, k was set equal to ten in this study.

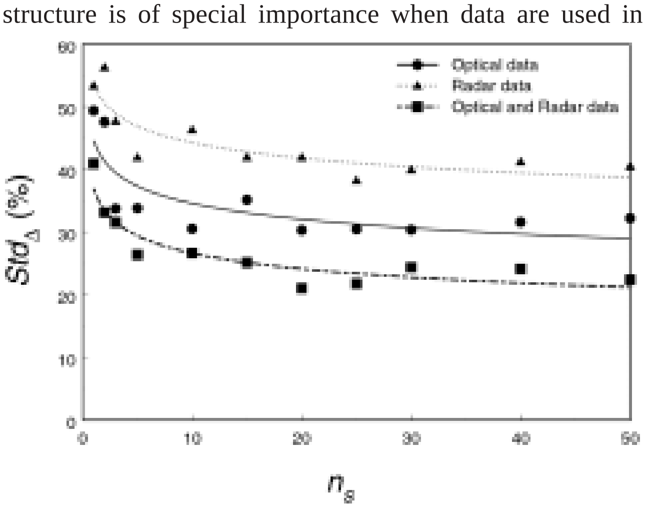

Figure 3. Relative standard deviation, Std_Δ, in standwise stem volume estimates plotted versus the sample size, n_s, with corresponding fitted regression functions, n = 31.

In previous studies, the *k*NN method has been used mainly in applications combining field and optical satellite data. At plot level, the errors in estimation of forest variables are relatively large, but decrease considerably at stand level. Poso et al. (1999) used satellite data together with aerial photography and old stand information, obtaining stem volume estimates for compartments (1–2 ha in size) with estimation errors of about 38%. In Holmgren et al. (2000), the standard error of standwise stem volume estimates were improved from 36% to 17% when ancillary data (site index, age, tree height) were added.

In cases where the objective is to obtain mean stand level estimates only, e.g., regression analysis might give equal or more accurate results compared to using *k*NN with the same information. Fransson et al. (2001) used regression analysis, and reported relative *RMSE*s of 24–38% for stem volume in the range of 0–305 $m^3\ ha^{-1}$ using SPOT XS data. In the study by Fransson et al. (2001), stem volume was also estimated using regression analysis with CARABAS-II VHF SAR data, resulting in a *RMSE* of 19% corresponding to 66 $m^3\ ha^{-1}$ in absolute terms (80–625 $m^3\ ha^{-1}$).

In Scandinavia, the traditional subjective methods used at stand level to acquire forest data yield standard errors of about 15–25% in comparison with the objective ground-based methods (seldom used for full coverage inventories) giving errors of about 10% in stem volume estimates (Ståhl 1992). However, the accuracy of these methods strongly depends on the number of sample plots used, the homogeneity of the forest stand, and surveyors' skill.

In this *k*NN study, the accuracy in terms of *RMSE* for stem volume estimates was found to be 37 $m^3\ ha^{-1}$ (22%) at stand level when combining SPOT-4 XS and CARABAS-II VHF SAR data. The results were substantially improved compared to the best single-sensor case, where only multispectral optical data were used (*RMSE* = 50 $m^3\ ha^{-1}$, 30%). Positive impact was also shown for the age estimates when using both image data sources as carrier data. In general, a strong correlation between age and stem volume is expected in managed forests. However, no improvement was obtained in the estimates of tree species composition (except for a minor improvement in old forests). This is explained by the main backscattering mechanism in coniferous forests at VHF-band, which is the ground-trunk scattering (Smith and Ulander 2000). Coniferous tree species variations do not seem to have any effect on the radar backscattering (Fransson 2000a). Using texture information derived from high-resolution image data has shown to enhance the separation between tree species (e.g., Franklin et al. 2000). However, in this study, no texture features were investigated due to the size of the target plots (radius = 10 m) in relation to the spatial resolution in SPOT-4 XS data (20×20 m). The CARABAS-II sensor captured the structure of larger objects (i.e., the tree trunks) and hence, texture analysis was assumed to be less informative in this case. In the analysis, carrier data were weighted proportionally to each source's ability to predict the variable of interest. As expected, the two image data sources contain complementary information.

When applying the *k*NN method, it is important that the reference plots cover the complete range of values in feature space (here represented by *DN*s and radar backscattering amplitude), ground conditions, and associated forest variables (Tomppo et al. 1999). In this study, the range of variable values in the reference data was not fully adequate for old forests. Systematic errors were observed in this study; e.g., stem volume was underestimated in old forests and overestimated in younger. These effects are caused by the *k*NN method employed, of necessity imputing the nearest available reference plot values (without any possibilities to extrapolate). To improve the results, the reference plot data should be augmented. Additional reference material might be obtained by gathering plot data from earlier inventories in the vicinity of the target area. Further improvement can be made by adjusting the forest variables of the reference plots to the year of image acquisition. At most, the time difference between image acquisition and field data collection was 1.5 yr. Another error source possibly affecting the results is the georeferencing of the image datasets to the sample plots/stands. In the case of the radar image, the displacement is on the order of 3–5 m (Walter 1999) and for the optical image, the error was visually determined to be half a pixel (10 m). It should be noted that the analysis was based on relatively well-positioned (submeter accuracy) objectively inventoried field plots. Furthermore, no corrections have been made due to ground slope (moderate at the test site) and moisture induced radar backscattering variation (Fransson et al. 2000a), which is especially pronounced at plot level.

Conclusion

Data from a considerable number of new remote sensing sensors, both satellite and airborne, are becoming commercially available at reasonable prices. In the *k*NN application presented here, data acquired by the optical SPOT-4 XS satellite and the airborne CARABAS-II VHF SAR system were easily integrated. The two image data sources co-vary with the forest characteristics, and the relationship between the reference plots and target area was substantially strengthened when using the two image sources in combination. Since the *k*NN method imputes reference plot data in their original format (and regardless of the format of the carrier data; spectral signatures, radar backscattering, etc.), traditional planning models demanding input data from field inventories can directly use the estimations.

By combining SPOT-4 XS optical data and CARABAS-II radar data, the two remotely sensed data sources improved the estimation accuracies of stem volume and age compared to when optical and radar data were used separately. The results are satisfactory and comparable to subjective inventories for forest stands. In conclusion, the results imply that further development of the *k*NN method, such as using a combination of data sources, has potential for operational use in forest management planning.

Literature Cited

Ahlberg, O., and L. Kardell. 1997. Remningstorp—från herresäte till skogslaboratorium. Västergötlands Tryckeri AB, Skara, Sweden. 184 p. (in Swedish).

Ardö, J. 1992. Volume quantification of coniferous forest compartments using spectral radiance recorded by Landsat Thematic Mapper. Int. J. Rem. Sens. 13:1779–1786.

Cohen, W.B., and T.A. Spies. 1992. Estimating structural attributes of Douglas-fir/western Hemlock forest stands from Landsat and SPOT imagery. Rem. Sens. Environ. 41:1–17.

Franklin, S.E., R.J. Hall, L.M. Moskal, A.J. Maudie, and M.B. Lavigne. 2000. Incorporating texture into classification of forest species composition from airborne multispectral images. Int. J. Rem. Sens. 21:61–79.

Fransson, J.E.S., A. Gustavsson, L.M.H. Ulander, and F. Walter. 2000a. Towards an operational use of VHF SAR data for forest mapping and forest management. P. 399–401 in Proc. of IGARSS 2000, Stein, T.I. (ed.). IEEE, Piscataway, NJ.

Fransson, J.E.S., F. Walter, and L.M.H. Ulander. 2001. Estimation of forest parameters using CARABAS-II VHF SAR data. IEEE Trans. Geosci. Rem. Sens. 38:720–727.

Fransson, J.E.S., G. Smith, J. Askne, and H. Olsson. 2001. Stem volume estimation in boreal forests using ERS-1/2 coherence and SPOT XS optical data. Int. J. Rem. Sens. 22:2777–2791.

Hagner, O. 1990. Computer aided forest stand delineation and inventory based on satellite remote sensing. P. 94–105 in Proc. from SNS/IUFRO workshop on The usability of remote sensing for forest inventory and planning, Swedish Univ. of Agric. Sci., Rem. Sens. Lab., Rep. 4. Umeå, Sweden.

Hellsten, H., L.M.H. Ulander, A. Gustavsson, and B. Larsson. 1996. Development of VHF CARABAS II SAR. P. 48–60 in Proc. of Radar Sensor Technology, vol. 2747. SPIE, Bellingham, WA.

Holmgren, J., S. Joyce, M. Nilsson, and H. Olsson. 2000. Estimating stem volume and basal area in forest compartments by combining satellite image data with field data. Scand. J. For. Res. 15:103–111.

Holmgren, P., and T. Thuresson. 1998. Satellite remote sensing for forestry planning—a review. Scand. J. For. Res. 13:90–110.

Holmström, H., M. Nilsson, and G. Ståhl. 2001. Simultaneous estimations of forest parameters using aerial photograph interpreted data and the *k* nearest neighbour method. Scand. J. For. Res. 16:67–78.

Imhoff, M.L. 1995. Radar backscatter and biomass saturation: ramifications for global biomass inventory. IEEE Trans. Geosci. Rem. Sens. 33:511–518.

Isaaks, E.H., and R.M. Srivastava. 1989. An introduction to applied geostatistics. Oxford University Press, New York. 561 p.

Jonsson, B., J. Jacobsson, and H. Kallur. 1993. The forest management planning package. Theory and application. Studia For. Suec. 189. 56 p.

Leckie, D.G., and K.J. Ranson. 1998. Forestry applications using imaging radar. P. 435-510 in Principles and applications of imaging radar, Manual of remote sensing. Ed. 3. Vol. 2. Henderson, F.M., and A.J. Lewis (eds.). Wiley, New York.

Lindgren, O. 1984. A study on circular plot sampling of Swedish forest compartments. Swedish Univ. of Agric. Sci., Dep. of Biometry and For. Manage., Rep. 11. Umeå, Sweden. 153 p.

Moeur, M., and A.R. Stage. 1995. Most similar neighbor—an improved sampling inference procedure for natural resource planning. For. Sci. 41:337–359.

Muinonen, E., and T. Tokola. 1990. An application of remote sensing for communal forest inventory. P. 35–42 in Proc. from SNS/IUFRO workshop on The usability of remote sensing for forest inventory and planning, Swedish Univ. of Agric. Sci., Rem. Sens. Lab., Rep. 4. Umeå, Sweden.

Nilsson, M. 1997. Estimation of forest variables using satellite image data and airborne lidar. Ph.D. thesis, Swedish Univ. of Agric. Sci., Silvestria 17, Umeå, Sweden. 31 p.

Poso, S., G. Wang, and S. Tuominen. 1999. Weighting alternative estimates when using multi-source auxiliary data for forest inventory. Silva Fenn. 33:41–50.

Ripple, W.J., S. Wang, D.L. Isaacson, and D.P. Paine. 1991. A preliminary comparison of Landsat Thematic Mapper and SPOT-1 HVR multispectral data for estimating coniferous forest volume. Int. J. Rem. Sens. 12:1971–1977.

Schreuder, H.T., T.G. Gregoire, and G.B. Wood. 1993. Sampling methods for multisource forest inventory. Wiley, New York. 446 p.

Smith, G., and L.M.H. Ulander. 2000. A model relating VHF-band backscatter to stem volume of coniferous boreal forest. IEEE Trans. Geosci. Rem. Sens. 38:728–740.

Söderberg, U. 1986. Functions for forecasting of timber yields—increment and form height for individual trees of native species in Sweden. Swedish Univ. of Agric. Sci., Section of For. Mensur. and Manage., Rep. 14. Umeå, Sweden. 251 p. (in Swedish with English summary).

Ståhl, G. 1992. A study on the quality of compartmentwise forest data acquired by subjective inventory methods. Swedish Univ. of Agric. Sci., Dep. of Biometry and For. Manage., Rep. 24. Umeå, Sweden. 179 p. (in Swedish with English summary).

Thompson, S.K. 1992. Sampling. Wiley, New York. 343 p.

Tokola, T., J. Pitkänen, S. Partinen, and E. Muinonen. 1996. Point accuracy of a non-parametric method in estimation of forest characteristics with different satellite materials. Int. J. Rem. Sens. 17:2333–2351.

Tomppo, E. 1990. Designing a satellite image-aided national forest survey in Finland. P. 43–47 in Proc. from SNS/IUFRO workshop on The usability of remote sensing for forest inventory and planning, Swedish Univ. of Agric. Sci., Rem. Sens. Lab., Rep. 4. Umeå, Sweden.

Tomppo, E., C. Goulding, and M. Katila. 1999. Adapting Finnish multi-source forest inventory techniques to the New Zealand preharvest inventory. Scand. J. For. Res. 14:182–192.

Ulander, L.M.H., J. Askne, J.E.S. Fransson, A. Gustavsson, T. Le Toan, T. Manninen, J.-M. Martinez, P. Melon, G. Smith, and F. Walter. 2000. Retrieval of stem volume in coniferous forest from low VHF-band SAR. P. 441–443 *in* Proc. of IGARSS 2000, Stein, T.I. (ed.). IEEE, Piscataway, NJ.

Walter, F. 1998. Remote-sensing for forestry planning. For. Res. Inst. of Sweden. Rep. 9. Uppsala, Sweden. 37 p. (in Swedish with English summary).

Walter, F., J.E.S. Fransson, and P.-O. Frölind. 1999. Fully automatic geo-coding of CARABAS-II VHF SAR images. P. 569–573 in Proc. of IGARSS'99, Stein, T.I. (ed.). IEEE, Piscataway, NJ.

Wikström, P. 2000. Solving stand level planning problems that involve multiple criteria and single-tree growth model. Ph.D. thesis, Swedish Univ. of Agric. Sci., Silvestria 167, Umeå, Sweden. 26 p.