 Received 18.3.2010, Accepted 01.06.2010

SITE-TYPE ESTIMATION USING AIRBORNE LASER SCANNING AND STAND REGISTER DATA

Markus Holopainen[1], Mikko Vastaranta[1], Reija Haapanen[2], Xiaowei Yu[3], Juha Hyyppä[3], Harri Kaartinen[3], Risto Viitala[4], and Hannu Hyyppä[5]

[1] University of Helsinki, Dept. of Forest Sciences, Latokartanonkaari 7, 00014 Helsinki, Finland

[2] Haapanen Forest Consulting, Finland
[3] Finnish Geodetic Institute, Finland
[4] HAMK University of Applied Sciences, Finland
[5] Aalto University, Finland

markus.holopainen@helsinki.fi, mikko.vastaranta@helsinki.fi, reija.haapanen@haapanenforestconsulting.fi, yu.xiaowei@fgi.fi, juha.hyyppa@fgi.fi, harri.kaartinen@fgi.fi, risto.viitala@hamk.fi, hannu.hyyppa@tkk.fi

ABSTRACT

The objective of this study was to determine the suitability of airborne laser-scanning (ALS) and existing stand register data in the estimation of forest site types via dominant height and stand age. Dominant height was estimated with the nonparametric k-nearest neighbour (k-NN) method, using ALS height metrics and field reference. For comparison, dominant height was also derived directly from the ALS height metrics. The site types were then estimated, using site-indexing models for artificially or naturally regenerated stands and conversion tables between site index and site type. Furthermore, the sensitivity of site-type estimation to errors in stand age and dominant height was tested. The study material consisted of 103 stands in which the dominant tree species was either Scots pine or Norway spruce. The study area was located in Evo, southern Finland. The site types varied from very poor (ClT) to groves (OMaT). Using ALS data and k-NN, the dominant height could be estimated with a root-mean squared error (RMSE) of 9.7%, while the direct use of heights derived from the ALS pulse height distribution resulted in an RMSE of 11.4%. In this case, the height at which 95% of the first pulse hits were accumulated was used as a proxy for the dominant height. The forest site types (five classes) were estimated with an overall accuracy of 70.9% (kappa 0.6), while site index measurement in the field resulted in an overall accuracy of 78.6% (kappa 0.7) for site-type classification. Site-type classification was more successful in Scots pine-dominated stands than in Norway spruce-dominated stands. Site index models created for artificially regenerated stands outperformed those for the naturally regenerated stands.

1. INTRODUCTION

Site classification is needed for describing the production potential of forest stands, selecting optimal harvesting strategies, as well as determination of nature protection and recreational values. Site productivity in turn is a decisive factor in determining the yield values of forest estates: differences in productivity are emphasized when the yield values are simulated over several decades.

Forest and mire site-type classification in Finland is based on the theoretical work of Cajander (1909; 1913; 1926; 1949). Cajander (1909) showed that the interaction of the dominant fertility factors of soil and climate is reflected in the ground vegetation. Site types are thus classified with an indirect method, based on the vegetation composition, assuming that similar ground vegetation cover is found at sites that are equal in tree productivity. Differences in vegetation community structures between site types are most distinctive in mature, closed, fully stocked stands. Classification is problematic, e.g. in seedling stands or clear-cut areas where the composition and phenotypes of ground vegetation differ from those of closed forest stands.

In forest site-type classification there are two directions, of which one takes into account the vegetation zonation (depends on climate) and another, more local, that reflects mainly the soil properties. This means that several forest site types can be distinguished, depending on the classification accuracy and climate. For practical forestry, classification has been simplified by removing the climate-driven zonation and describing the influence of soil with six site fertility types: OMaT (groves), OMT (Oxalis-Myrtillus type, rich mineral soil forests), MT (Myrtillus type, medium-rich mineral soil forests), VT (Vaccinium type, medium-poor mineral soil forests), CT (Calluna type, poor mineral soil forests) and ClT (Cladonia type, extemely poor mineral soil forests) (Ojansuu, 2005).

One way to measure the productivity of the site is the site index, which is expressed as the dominant height (H_{dom}) of the stand at a certain age. In the field, H_{dom} is calculated as the arithmetic mean height (H_{mean}) of the 100 thickest trees per hectare. In Finland, the index age is 100 years with conifers (H_{100}) and 50 years with deciduous species (H_{50}). The site index is a more direct method for determining the productivity than forest or mire site types (Ojansuu, 2005). The site index is determined via models that describe the development of H_{dom}. Development of the H_{dom} of each tree species is presented as a set of curves that show the height development according to the corresponding site index. In Finland, site index curves have been created separately for coniferous forests of natural origin (Gustavsen, 1980), artificially regenerated coniferous forests (Vuokila and Väliaho, 1980) and artificially regenerated silver birch (*Betula pendula* Roth) forests (Oikarinen, 1983). The site index method is usually applied to single-tree species, comprising evenly aged stands that have undergone undisturbed development, have been treated with low thinning and are free of damage (Ojansuu, 2005).

Major uncertainties in the site index method include disturbances occurring at early stages of stand development. The significance, however, levels off with increasing stand age. Site indexes cannot further take into account naturally differing rhythms in development, which cause noise, especially at young stages (Ojansuu, 2005). Thus, the site index method is recommended for use in forests in advanced developmental stages. The effects of fertilization and drainage should also be considered. The accuracy of the site index method is dependent on errors in measuring stand mean age and H_{mean}. Generally, the site index is based on the biological age of the dominant trees, but applying breast height would reduce the uncertainty caused by disturbances at young age (Ojansuu, 2005).

A clear relationship exists between site index and site type; however, this relationship varies, since the same site index can be found in several forest site types (Ilvessalo, 1920; Nyyssönen, 1954; Vuokila, 1956; Koivisto, 1959; Vuokila and Väliaho, 1980; Oikarinen, 1983; Tamminen, 1993; Karlsson, 1996). There are controversial results from applying the site index method to mixed-species forests. Lappi-Seppälä (1930), Jonsson (1962) and Hägglund (1975) observed that species composition affects the height development of stands. However, Mielikäinen (1980; 1985a; 1985b) concluded that site fertility can be based on the site index method even in mixed-species forests, because the height development of Scots pine (*Pinus sylvestris* L.), Norway spruce (*Picea abies* (L.) H. Karst.), and birch (*Betula* L.) is not sensitive to the surrounding mixture of tree species.

Airborne laser scanning (ALS)-based inventories open new opportunities for determining stand productivity, because both area-based (e.g. Næsset 1997; 2002; 2004) and single-tree-based (e.g. Hyyppä and Inkinen, 1999; Leckie et al., 2003; Popescu et al., 2003; Maltamo et al., 2004) applications of ALS function best at estimating height parameters. There is always variation in productivity within a certain forest site type. Compared with traditional site-type classification, site indexing with ALS-based height information can give information on the intrastand variation, because the variation in stand height is acquirable and most of the stands are evenly aged. Furthermore the site index, being a continuous variable, could mitigate the problems arising from forcing the stands into distinct site-type classes. However, with regard to determining the site index, the problem of measuring the stand age remains.

Few results are available on ALS-based site-type classification. Gatziolis (2007) estimated the H_{dom} and site types with a single-tree-based ALS method in the coastal Pacific Northwest of the USA. The ALS measurements were carried out in two campaigns: leaves on and off, with a pulse density of 8.7-9.8/m^2. Single trees were detected with ALS, while stand age was derived from the forest management plan. The accuracy of age estimates was controlled with field sample plots. The coefficient of determination (R^2) between the site indexes derived from field inventory or ALS measurements was 0.42. Wide variation in topography, as well as stand density, significantly affected the results, and far better results (R^2=0.88) were obtained when the data were filtered to include average slopes and stand densities only.

Vehmas et al. (2008) estimated mineral soil forest site types (five classes) with area-based ALS-inventory and a nonparametric k-nearest neighbour (k-NN) method in a nature protection area in Finland. They hypothesized that different forest site types would produce different vertical distributions of laser pulses, due to increasing numbers of deciduous trees in better site types. The best overall accuracy was 58%, and the best percentage correct for a single class was 73%. Vehmas et al. (2008) concluded that one source of error is the subjective determination of forest site type, resulting in larger errors in ground truth than in the actual estimation with ALS data. Using a similar method, Vehmas et al. (2009) identified herb-rich forest stands from less fertile site types, with an overall classification accuracy of 88.9%. However, Vehmas et al. (2008; 2009) did not carry out any ALS-based site indexing, but estimated the forest site types directly.

Vega and St-Onge (2008) introduced a remote-sensing (RS) method for site index classification with promising results. The method was based on ALS and time series of aerial photographs. In their study, the average bias of site index and age were 0.76 m and 1.86 years, respectively. Future site indexing could be based on multitemporal ALS.

The objective of this study was to determine the suitability of low-pulse density ALS and stand register data in the estimation of site types via H_{dom} and age-based site indexing. H_{dom} was estimated with the nonparametric k-NN method and, for comparison, was derived directly from the distribution of ALS pulses. The site indexes were then estimated, using models for artificially or naturally regenerated stands and converted to site types. Furthermore, the sensitivity of site-indexing-based site-type classification to errors in determining stand age was tested. Finally, the ALS-based site indexes were compared with site indexes derived using field measurement data. The study aimed at developing methods suitable for operational forest management planning.

2. MATERIALS AND METHODS

2.1 Study area and field measurements

The study area is located at Evo, Finland (61.19ºN, 25.11ºE, Fig. 1) and comprises approximately 2000 ha of managed boreal forest. The average stand size in the area was slightly less than 1 ha. Field measurement data from 333 fixed-radius (9.77 m) field plots were collected from the study area in summer 2007 and in winter 2008. Sampling of the field plots was based on prestratification of existing stand inventory data. There was a 1-year gap between acquisition of the RS data and field data measurements; only plots that had remained untreated during the year were measured and the latest growth in height was subtracted. The plots were located, using Trimble's GEOXM 2005 Global Positioning System (GPS) device (Trimble Navigation Ltd., Sunnyvale, CA, USA), and the locations were postprocessed with local base station data, resulting in an average error of approximately 0.6 m.

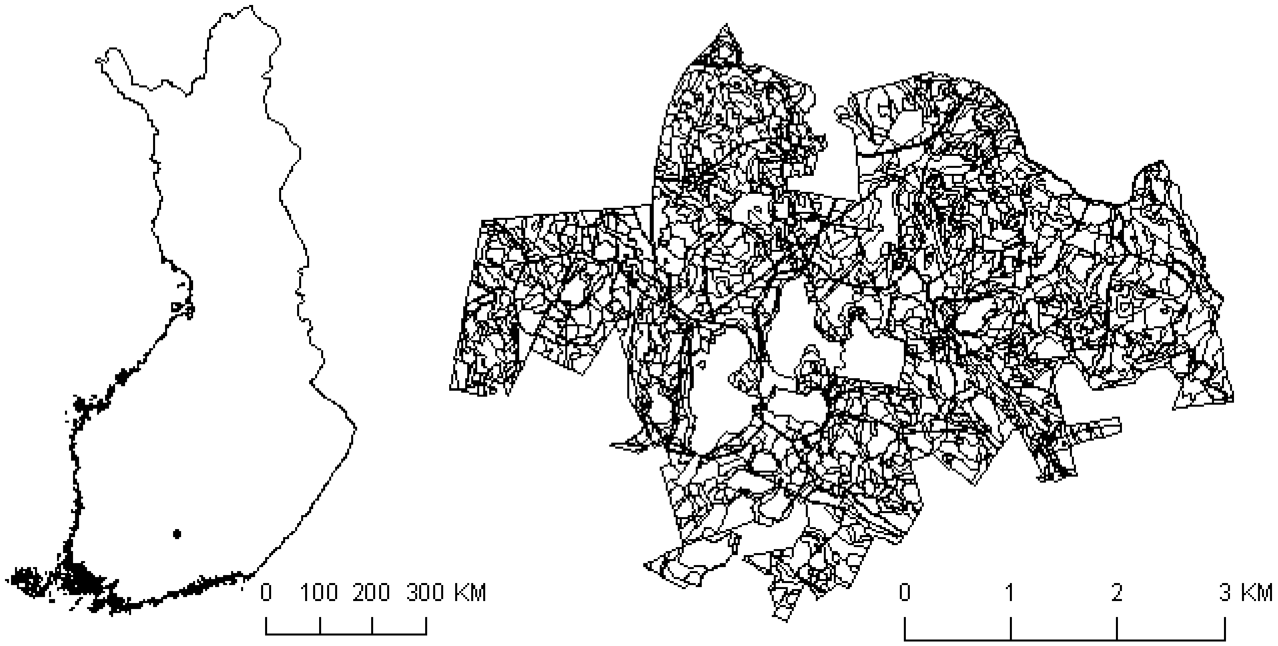

Figure 1. Location of the Evo study area in southern Finland and the stands within it.

Field plots (n = 103) fulfilling the following criteria were selected for this study: the dominant tree species was either Scots pine or Norway spruce, and the stand age 50-100 years. The forest site types were determined according to the theory of Cajander (1949). Age information obtained from the stand databases was checked in the field. If there were visually detected errors in stand age compared with plot-level age, the ages were determined by boring with an increment borer. The site types ranged from groves (OMaT) to very poor sites (CT) (Fig. 2). According to the stand database of the forest management plan, the ages of the field plots were 25-109 years. The H_{dom} varied between 7.5 m and 34.5 m (Fig. 3). Tree-level measurements were carried out at the field plots. In these measurements, tree species, diameter-at-breast height (dbh) and tree height were determined from every tree having a dbh of over 5 cm. The values for dbh were obtained, using steel callipers. The tree heights were measured using a Haglöf Vertex clinometer (Haglöf Sweden AB, Långsele, Sweden). Stand variables were derived from the tree measurements. The characteristics needed for site index classification based on the field data are presented in Figure 3.

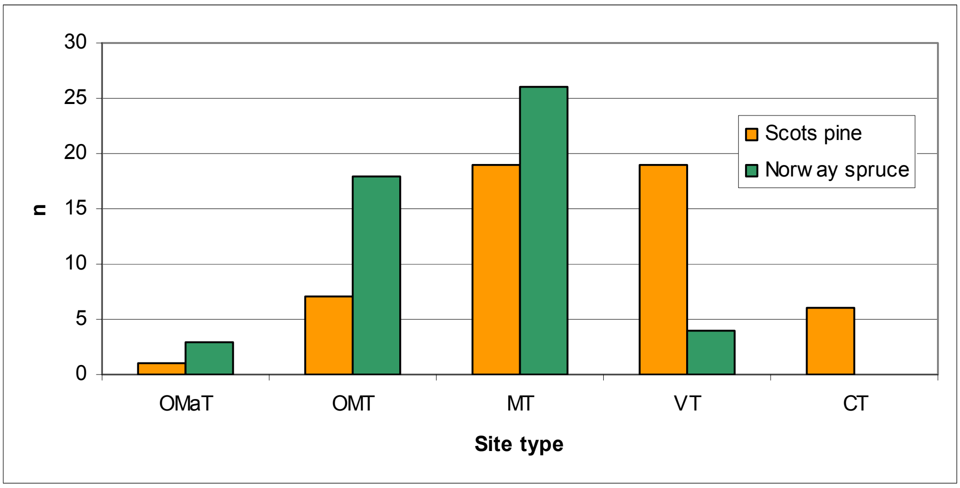

Figure 2. Site types according to vegetation-based field inventory.

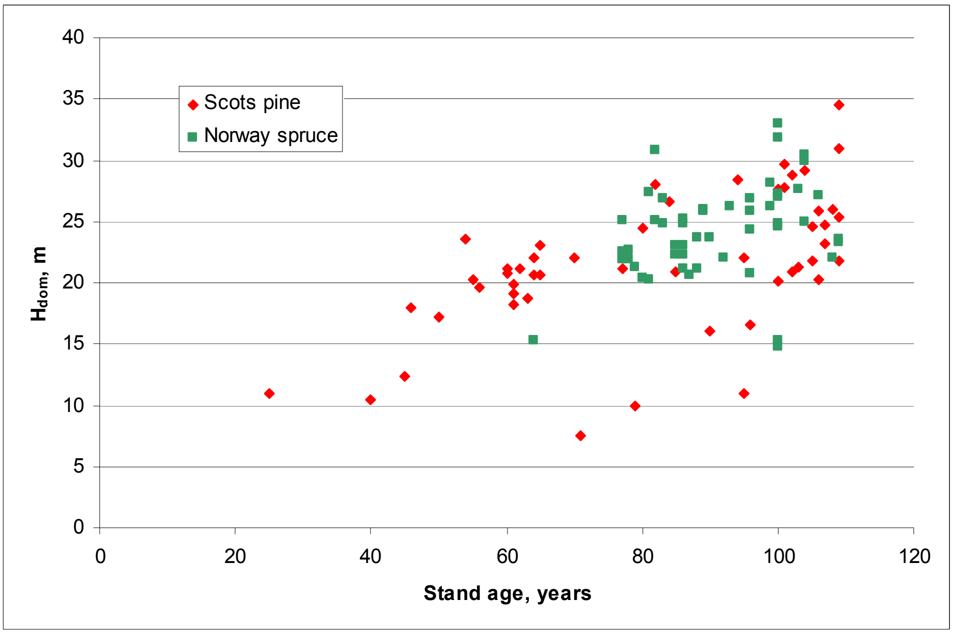

Figure 3. Field-measured dominant heights and stand ages from the stand register for the field plots.

2.2 ALS measurements and aerial photographs

The ALS data were acquired in midsummer 2006, using an Optech ALTM3100C-EA system (Optech Inc., Vaughan, Ontario, Canada). The flying altitude was 1900 m at a speed of 75 m/s, a half-angle of 14 degrees, a pulse rate of 70 kHz and a footprint of 1.14 m^2. The density of the pulses echoes returned within the plots was 1.8/m^2 (only, first (F), intermediate or last (L); 1.3/m2 if only the first returned pulses were considered). A digital terrain model (DTM) and, consequently, heights above ground level, were computed by the data provider. Same-date aerial photographs were taken with a Vexcel Ultracam digital camera (Vexcel Corporation, Boulder, CO, USA). The photographs were orthorectified, resampled to a pixel size of 0.5 m and mosaiced to a single image covering the entire area. The near-infrared (NIR), red (R) and green (G) bands were available.

2.2.1 Feature extraction and selection

Several statistical and textural features were extracted from the ALS data and aerial photographs for Hdom estimations. Firstly, the means and standard deviations of aerial photograph spectral values and ALS height and intensity were calculated. Secondly, Haralick textural features (Haralick et al., 1973; Haralick, 1979) were derived from the spectral values and ALS height and intensity. The Haralick textural features were computed for four directions: 0˚, 45˚, 90˚ and 135˚. For the above mentioned features, the extraction window was 20 x 20 m, corresponding approximately to the size of the reference field plots. Thirdly, 'standard texture' features referring to a set of averages and standard deviations of spectral values, ALS heights (first returns) and intensities were calculated for a 32 x 32-pixel window from images rasterized to 0.5-m pixel size. Finally, height statistics for the first (F) and last (L) pulses were calculated: H_{mean} and maximum height (H_{max}), standard deviation and coefficient of variation of height (hstd, hcv), heights at which certain relative amounts of laser points had accumulated (h05-h95) as well as proportion of pulse returns below various relative heights (p05-p95). Only pulses exceeding a 2-m height limit were included to remove hits to ground vegetation and bushes. Percentages of points over 2 m in height were also added (penetration). For the height statistics, the calculation unit was the actual circular field plot area. The means and standard deviations of ALS height were included only once in the final dataset. All features were standardized to a mean of 0 and standard deviation of 1. Feature selection was based on the genetic algorithm method presented by Holopainen et al. (2008). A reduced set of features (11 features, see Holopainen et al. 2008) was used in the estimation.

2.2.2 Estimation algorithm of H_{dom}

The estimation method for H_{dom} was nonparametric k-NN, which has long been used in RS-aided forest inventory applications in Finland (e.g. Kilkki and Päivinen, 1987; Muinonen and Tokola, 1990; Tomppo, 1991). The nearest neighbours were determined by calculating the Euclidean distances between the observations in the *n*-dimensional feature space, which in this case consist of ALS metrics and features extracted from the aerial photographs. The nearest plots were weighted with inverse-squared distances. The number of nearest neighbours was set at five.

For comparison, the H_{dom} values were derived directly from the ALS height metrics. Heights at which 100%, 95% or 90% of the first pulses had accumulated (Fp100, Fp95 and Fp90) were tested as proxies for H_{dom} without any modelling.

2.3 Determining the site index and site type

Site index determination was based on the H_{dom} and age of the stand. The site type in turn was determined, using H_{dom} increment models for artificially regenerated Scots pine and Norway spruce stands by Vuokila and Väliaho (1980, VV) (equations 1, 2), and site index model for naturally regenerated coniferous stands by Gustavsen (1980, G) (equation 3). In Figure 4, site index curves for Scots pine stands, derived from the dominant height increment models (Vuokila and Väliaho, 1980) are shown. In the figure the development of H_{dom} in different forest site types can also be seen.

The curves in Figure 4 were developed, using equation (1) (Vuokila and Väliaho, 1980):

$$P_{\text{Hdom(Scots pine)}} = \frac{1}{\left(H_{dom}^{0.4} * T^{1.1}\right)} + \frac{Y}{\left(H_{dom}^{0.4} * T^{1.1}\right)^2} + C \qquad (1)$$

where P is the annual increment of the H_{dom} during the next 5-year period, H_{dom} the dominant height, T the biological age of the stand, Y the length of the forecast period, and C a species-specific constant.

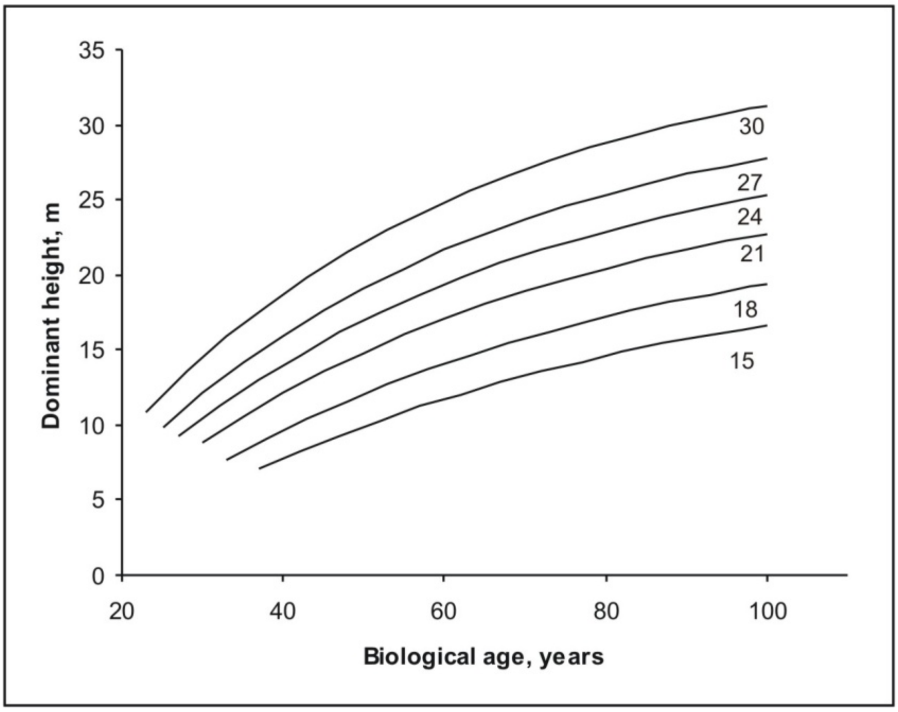

Figure 4. Site indexes for pine stands based on dominant height over age (Vuokila and Väliaho, 1980).

Corresponding curves can be generated for Norway spruce stands, using equation (2)

$$P_{Hdom(Norway\ sruce)} = \frac{1}{\left(H_{dom}^{0.55} * T^{1.05}\right)} + \frac{1}{\left(H_{dom}^{0.55} * T^{1.05}\right)^2} + C \qquad (2)$$

For naturally regenerated coniferous forests, the site index can be directly derived, using equation (3):

$$H_{100} = d \cdot \exp\left[\frac{\ln H_{dom} - \ln d}{\exp\left(\frac{b_1}{T^{a-1}} - \frac{a-1}{100^{a-1}}\right)}\right] \quad \text{(Gustavsen 1980)} \qquad (3)$$

where d, a-1 and b_1 vary according to the species.

Each reference site type was determined in the field, based on the understorey vegetation (Cajander, 1949). The site types applied were OMaT, OMT, MT, VT, CT and ClT. The site indexes can be converted to site types using tables (Table 1). The H_{dom} classes of Vuokila and Väliaho (1980) or Gustavsen (1980) do not fully conform with the site types used in practical forest management planning: one forest site type may be spread over more than one H_{dom} class. In this study the nearest site type was used.

Table 1. Approximate locations of various forest site types in the height-over-age site classification system.

H_{100}	Vuokila &Väliaho	Gustavsen
33		
30	OMT (OMat)	OMT
27	MT (OMT)	OMT-, MT+ (OMat, OMT+)
24	VT (MT)	MT, VT+ (OMT)
21	VT-, CT+ (VT)	VT (MT)
18	CT	CT (VT)
15	CT-, Clt+	CT-

2.4 Determining the forest site-type estimation accuracy

The forest site-type classification accuracy was determined by computing the site index for each field plot with the models developed by Vuokila and Väliaho (1980, VV) and Gustavsen (1980, G). The site indexes were determined separately for the field inventory and ALS approaches. In both cases, the stand age was taken from an existing forest management plan. The site indexes were converted to site types and compared with those assessed in the field. The success of classification was evaluated, using overall accuracies and kappa values.

3. RESULTS

Two methods were used for the estimation of H_{dom} for the field plots from the ALS data: nonparametric k-NN and ALS height metrics. In the latter, the heights in which 100% (Fp100), 95% (Fp95) or 90% (Fp90) of the first pulses had accumulated were tested as proxies for H_{dom} without any modelling. The H_{dom} estimation accuracies obtained, using these methods, are presented in Table 2.

Table 2. Bias (m) / (Bias%) and RMSE (m) / (RMSE%) of dominant height, when dominant height was estimated using either k-NN or directly from ALS height metrics.

	BIAS	RMSE
k-NN	-0.2 (0.7)	2.2 (9.7)
FP100	1.2 (5.2)	3.0 (13.3)
FP95	-0.6 (2.5)	2.6 (11.4)
FP90	-2.3 (10.2)	3.5 (15.7)

Based on the results in Table 2, we conclude that proxies of H_{dom} can be derived from ALS height metrics. The height at which 95% of the first pulses accumulated (Fp95) was the best-performing of the percentage points tested, giving a bias of -0.6 m (2.5%) and root-mean squared error (RMSE) of 2.6 m (11.4%). Using the highest point (Fp100) resulted in a significant overestimation of H_{dom} (bias 1.2 m) and, correspondingly, using the point at which 90% of the laser heights accumulated (Fp90) led to an underestimate (bias -2.3 m). An unbiased estimate may be found somewhere between Fp95 and Fp100. Using this point, one may result in lower RMSE than the one obtained with Fp95. Of the methods tested, k-NN performed best, resulting in H_{dom} estimates with a bias of -0.2 m (0.7%) and RMSE of 2.2 m (9.7%). Note that the material comprised stands in advanced developmental stages, meaning that despite the large absolute errors in height, the relative errors were quite low and similar to those in previous studies (Næsset, 1997, 2002, 2004, Næsset et al., 2004). The better of the methods tested for H_{dom} estimation, k-NN (later referred to as ALS_{VV} or ALS_G, depending on the site index model used), was chosen for further comparisons with a field measurement-based method, i.e. the forest site types were estimated via site indexes, using these approaches. Models for artificially and naturally regenerated stands were tested for both approaches and the error matrices and corresponding percentages correct are presented in Tables 3 and 4 (ALS_{VV}, ALS_G, $FIELD_{VV}$ and $FIELD_G$).

Table 3. Error matrices for forest site types (SC), estimated using site index models by Vuokila and Väliaho (1980, VV) and Gustavsen (1980, G) (see Table 2).

SC	SC H100 $FIELD_{VV}$				
	OMaT	OMT	MT	VT	CT
OMat	3	0	1	0	0
OMT	5	19	1	0	0
MT	1	6	35	3	0
VT	0	0	2	20	1
CT	0	0	0	2	4

SC	SC H100 ALS_{VV}				
	OMaT	OMT	MT	VT	CT
OMat	2	0	2	0	0
OMT	4	20	1	0	0
MT	1	11	28	5	0
VT	0	0	4	18	1
CT	0	0	0	1	5

SC	SC H100 $FIELD_G$				
	OMaT	OMT	MT	VT	CT
OMat	2	1	1	0	0
OMT	17	6	2	0	0
MT	6	17	22	0	0
VT	0	1	11	9	2
CT	0	0	0	1	5

SC	SC H100 ALS_G				
	OMaT	OMT	MT	VT	CT
OMat	2	1	1	0	0
OMT	17	6	2	0	0
MT	8	19	16	2	0
VT	0	0	10	12	1
CT	0	0	0	0	6

Table 4. Percentages correct for site types (SC) acquired via site index models by Vuokila and Väliaho (1980, VV) and Gustavsen (1980, G).

SC	$FIELD_{VV}$	ALS_{VV}	$FIELD_G$	ALS_G
OMaT	75.0	50.0	50.0	50.0
OMT	76.0	80.0	24.0	24.0
MT	77.8	62.2	48.9	35.6
VT	87.0	78.3	39.1	52.2
CT	66.7	83.3	83.3	100.0
Total	78.6	70.9	42.7	40.8
Kappa	0.70	0.60	0.24	0.23

Table 4 shows that the site index models for artificially regenerated stands (Vuokila and Väliaho, 1980, VV) performed better in almost all forest site types than the models for naturally regenerated stands (Gustavsen, 1980, G). The overall accuracies for models VV were 70.9% (ALS_{VV}) and 78.9% ($FIELD_{VV}$), and for models G 40.8% (ALS_G) and 42.7% ($FIELD_G$). As expected, the field measurement-based approach performed slightly better than the ALS-based approach, the percentages correct for single forest site-type classes being 66.7-87% for the former and 50-83.3% for the latter. The forest site-type estimation accuracies for the 52 Scots pine-dominated stands and 51 Norway spruce-dominated stands are presented in Tables 5-7.

Table 5. Error matrix for forest site types (SC) estimated for Scots pine-dominated stands (models VV and G).

SC	SC H100 FIELD$_{VV}$				
	OMaT	OMT	MT	VT	CT
OMat	1	0	0	0	0
OMT	1	6	0	0	0
MT	0	1	17	1	0
VT	0	0	2	16	1
CT	0	0	0	2	4

SC	SC H100 ALS$_{VV}$				
	OMaT	OMT	MT	VT	CT
OMat	0	0	1	0	0
OMT	0	7	0	0	0
MT	0	3	14	2	0
VT	0	0	4	14	1
CT	0	0	0	1	5

SC	SC H100 FIELD$_{G}$				
	OMaT	OMT	MT	VT	CT
OMat	0	1	0	0	0
OMT	0	4	3	0	0
MT	0	2	17	0	0
VT	0	1	8	7	3
CT	0	0	0	1	5

SC	SC H100 ALS$_{G}$				
	OMaT	OMT	MT	VT	CT
OMat	0	0	1	0	0
OMT	0	4	3	0	0
MT	0	5	13	1	0
VT	0	0	8	9	2
CT	0	0	0	0	6

Table 6. Error matrix for forest site types estimated for Norway spruce-dominated stands (models VV and G).

SC	SC H100 FIELD$_{VV}$				
	OMaT	OMT	MT	VT	CT
OMat	2	0	1	0	0
OMT	4	13	1	0	0
MT	1	5	18	2	0
VT	0	0	0	4	0
CT	0	0	0	0	0

SC	SC H100 ALS$_{VV}$				
	OMaT	OMT	MT	VT	CT
OMat	2	0	1	0	0
OMT	4	13	1	0	0
MT	1	8	14	3	0
VT	0	0	0	4	0
CT	0	0	0	0	0

SC	SC H100 FIELD$_{G}$				
	OMaT	OMT	MT	VT	CT
OMat	2	0	1	0	0
OMT	17	1	0	0	0
MT	6	13	7	0	0
VT	0	0	2	2	0
CT	0	0	0	0	0

SC	SC H100 ALS$_{G}$				
	OMaT	OMT	MT	VT	CT
OMat	2	1	0	0	0
OMT	15	3	0	0	0
MT	6	14	4	2	0
VT	0	0	2	2	0
CT	0	0	0	0	0

Table 7. Percentages of correct site-type classification by dominant tree species, using different approaches. Values without parentheses are for Scots pine-dominated stands and those in parentheses for Norway spruce-dominated stands.

SC	FIELD$_{VV}$	ALS$_{VV}$	FIELD$_{G}$	ALS$_{G}$
OMaT	100.0 (66.7)	0 (66.7)	0.0 (66.7)	0 (66.7)
OMT	85.7 (72.2)	100 (72.2)	57.1 (5.6)	57.1 (16.7)
MT	89.5 (69.2)	73.7 (53.8)	89.5 (26.9)	68.4 (15.4)
VT	84.2 (100.0)	73.7 (100.0)	36.8 (50.0)	47.4 (50.0)
CT	66.7 (*)	83.3 (*)	83.3 (*)	100 (*)
Total	84.6 (72.5)	76.9 (64.7)	65.4 (19.6)	63.5 (17.6)
Kappa	0.78 (0.58)	0.67 (0.48)	0.48 (0.01)	0.46 (0.00)

In almost all forest site types, the site-type estimation succeeded better in Scots pine-dominated stands than in Norway spruce-dominated plots in both H_{dom} determination approaches. When field measurement-based H_{dom} determination was used, site index model VV (for artificially regenerated stands) resulted in 84.6% accuracy for site-type estimation in pine stands and 72.5% in spruce stands. For the ALS/ k-NN-based approach, the corresponding percentages correct were 76.9% and 64.7%. The difference between these tree species was even greater when site index models for naturally regenerated forests (G) were applied: the percentages correct were 63.5% for pine and 17.6% for spruce.

Figure 5 shows the effect of errors in stand age and Figure 6 in H_{dom} on forest site-type estimation accuracy. Correct stand age and H_{dom} information is especially important when site index models are used for artificially regenerated stands (VV): a 5-year error in age or 1-m error in H_{dom} lowers the overall accuracy of site-type estimation by app. 5-15%. On the other hand, site index models for naturally regenerated stands are not as sensitive to errors in age information: a 15-year error in age is allowed before the accuracy lowers by 10% (Fig. 5).

Figure 5. Effect of error in the determination of stand age on forest site-type estimation accuracy; field measurement and ALS/k-NN -based approaches, site index models VV and G.

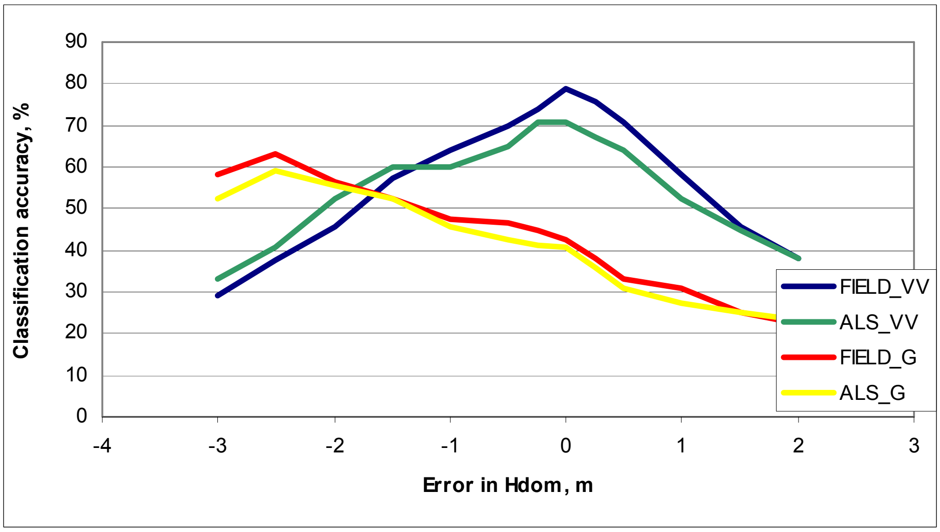

Figure 6. Effect of error in the determination of dominant height on forest site-type estimation accuracy; field measurement and ALS/k-NN -based approaches, site index models VV and G.

Further information can be obtained from Figures 5 and 6 on why the site index models developed for naturally regenerated forests clearly performed more poorly than those for artificially regenerated forests. Probably most of the stands in the study area originated in artificial regeneration programmes, because both field measurement-based and ALS site-indexing-based methods resulted in best classification accuracies when the age or height information was correct. When site index models for naturally regenerated stands are used, the best results are obtained when the stand age is overestimated by app. 20 years or H_{dom} underestimated by app. 3 m. The poor success in estimating forest site types for Norway spruce-dominated stands (Table 7) when the site index model for naturally regenerated stands is used is explained by this phenomenon.

4. DISCUSSION

In the present study we tested the suitability of site indexes derived from ALS- and stand register data for the estimation of forest site types. The ALS-based site types and site indexes were estimated via the nonparametric k-NN method, using features extracted for the areas of field plots as predictors. The results were compared with site types determined in the field. The age information was obtained from a forest management plan, but checked in the field.

The simplest way to determine the H_{dom} from ALS data is to derive it from the distribution of heights of the first pulses. When the height is used at which 95% (Fp95) of the first pulse hits were accumulated as a proxy for the H_{dom}, a minimum RMSE of 11.4% was obtained, while in the k-NN -based estimation the RMSE was 9.7%. We tested only the percentage points of 90%, 95% and 100%. We assumed that testing other percentages between 90% and 100%, and selecting the one giving the least biased estimate would have lowered the accuracy to the level obtained via k-NN. The advantage of the k-NN method is that the H_{dom} estimation can be seamlessly added to the estimation of other standard forest attributes in operational forest management planning. On the other hand, the advantage of using an ALS-based height distribution is that it requires field measurements only for calibration. Thus, we recommend this method if there is no need for data other than site index.

The ALS-based estimation of forest site types succeeded relatively well. When site index models for artificially regenerated stands were used, the overall accuracy in site-type classification was 70.9% (kappa 0.6). The field measurement-based inventory, which was used to benchmark the ALS-based method, resulted in an overall accuracy of 78.9% (kappa 0.7). This result indicates that the same site index can be found in several forest site types.

The percentage correct of a single forest site type ranged from 50% to 83.3%, the lowest accuracy occurring in groves (OMaT) and the best at very poor sites (ClT). The groves were, however, rare: only four of the field plots belonged to this class. In the error matrix (Table 4), the greatest uncertainties in the ALS estimates are found at medium-rich mineral soil forest sites (MT; class 3), which are often confused with poorer (VT: dryish site) or richer (OMT) site types.

In boreal forests, it is important to determine the most fertile sites (OMat and OMT), since they are the most productive from the standpoint of economy, and usually the key biotopes are also valuable for biodiversity. When we combined these two forest site types, we obtained a percentage correct of 81.5% for the combined class, suggesting that the method applied could be used for detecting fertile sites.

Our results are similar to those obtained by Vehmas et al. (2008, 2009) in eastern Finland, using the k-NN estimation method and area-based extraction of ALS features. Vehmas et al. (2008) obtained an overall accuracy of 58% for five forest site-type classes (corresponding to our classes). The most fertile site types (OMaT, OMT) were separated from other site types, with an overall accuracy of 88.9% (Vehmas et al. 2009). Note that Vehmas et al. (2008, 2009) did not carry out any ALS-based site indexing, but estimated the forest site types directly.

When the results are interpreted, it must be remembered that the determination of forest site type is prone to error in the field, as well: the indicator plants may be lacking or there may be wide variation within the stand. In other words, the site index and site type will never be perfect matches, even if determined in the field. This suggests that the overall accuracy obtained via benchmarking the field inventory-based method (78.9%) can be considered as an upper bound to the accuracies achievable within these data.

In the ALS-based site index approach, accurate determination of stand age is challenging. When applied in practice, the age information must be taken from an existing forest management plan. If the stand has been regenerated artificially, an accurate age may be found in the registers, but in naturally regenerated stands there may be significant errors in age information, because the trees are not generally bored into when field data for a forest management plan are collected. In the present study age information was controlled with borings. This procedure was used where needed, because stand-level age information was generalized to the plot level. There were some variations in plot-level tree species and age compared with the stand level. This procedure is not needed in practice, when site indexing is done at the stand level. Height-based characteristics can be estimated with relatively good accuracy with ALS methods (e.g. Næsset 1997, 2002, 2004). The accuracy of H_{dom} estimation when area-based ALS inventory is used is 1-2 m, depending on the height of the trees and intrastand variability.

Site index models for both artificially regenerated stands (Vuokila and Väliaho 1980) and naturally regenerated stands (Gustavsen 1980) were tested in this study. Most of our study area consisted of artificially regenerated forests and, consequently, the models developed specifically for these outperformed those developed for natural forests. On the other hand, the models for artificially regenerated forests were much more demanding in determining the stand age. In Figure 5 it can be seen that even an error of 5 years in the stand age results in a 10% drop in overall accuracy of site-type classification when site index models for artificially regenerated stands are applied, while an error of 15 years is needed to develop a similar reduction in accuracy when models for naturally regenerated stands are used. Based on this result, we conclude that the target area should be stratified according to the regeneration method whenever possible, after which site indexes should be estimated separately for artificially and naturally regenerated stands. Furthermore, when ALS data-based methods become more popular and the data more widespread, there will be a niche for site index models that are less sensitive to stand age than those currently used.

In future studies we will continue to search for ways to employ low-pulse density ALS data for the determination of H_{dom} in stands. In addition to the estimation methods tested here (k-NN- and height distribution-based methods) the potential use of ALS-based single-tree detection was tested (Holopainen et al. 2009). Although low-pulse density ALS data have performed poorly in the determination of the diameter distribution of stands (the smallest trees are not found), the situation is significantly better with regard to the larger dominant trees, suggesting that single-tree-based H_{dom} determination should be successful in these types of data, as well. For single-tree detection it would be better if the definition for H_{dom} would be based on the top tree heights instead of the heights of the thickest trees. Problems in determining H_{dom} from the site indexing standpoint is earlier studied in Mailly et al. (2004) and Holopainen et al. (2009).

Our results are promising and useful in developing a practical site index method. It must be emphasized that the results are somewhat data-dependent and cannot be directly generalized to other forests, e.g. with respect to overall accuracies of separate forest site types. Furthermore, the H_{dom}-based site index method is suitable mainly for stands with clearly dominant tree species in advanced developmental stages, and the sensitivity to errors in stand age limit the usability of this method.

ALS-based site indexes could be used in addition to the traditional site types (based on ground vegetation) in advanced, single-tree species stands. Wide variation may occur in traditionally determined site types within forest management-planning units. With ALS-based determination of the site index, it would be possible to obtain objective information on site productivity and its spatial variation. The ALS-based inventory of forest stands is interesting with regard to pricing of forest estates, as well, due to the strong association between site productivity and economic value of stands.

Inventory data for forest planning in Finland will be collected mainly via ALS measurements, suggesting that all commercial forests in Finland will be covered with low-pulse laser-scanning data within the next 10-15 years. An area-based ALS inventory produces results, e.g. for grids of certain resolution or substands derived via automated segmentation. For these inventory units it is then possible to estimate site indexes, giving information on the intrastand variation in productivity. If collecting of accurate age information were taken into the forest-planning system, it would be possible to produce a nationwide site index map with field resolution corresponding to that used in ALS inventory (e.g. 16 m x 16 m). Furthermore, this map would not be dependent on the delineation of forest stands.

5. CONCLUSIONS

Site-type estimation via site indexes provides a useful method for the determination of stand productivity. ALS-based forest inventory will open new opportunities for the implementation of site indexing in practice: operative forest management planning, estimating the value of forest estates and mapping of ecologically important habitats. The results of this study suggest that forest site type and site index can be estimated nearly as well with ALS-based estimation of H_{dom} as with field measurements involving single trees. However, further investigations are needed to develop methods for determining stand age and functioning of site index models.

6. ACKNOWLEDGEMENTS

This study was made possible by financial aid from the Finnish Academy project Improving Forest Supply Chain by Means of Advanced Laser Measurements (L-IMPACT).

7. REFERENCES

Cajander, A.K., 1909. Über Waldtypen. Acta Forestalia Fennica, 1, pp. 1-175.

Cajander, A.K., 1913. Studien über die Moore Finnlands. Acta Forestalia Fennica, 2, pp. 1-208.

Cajander, A.K., 1926. The theory of forest types. Acta Forestalia Fennica, 56, 71 p.

Cajander, A.K., 1949. Forest types and their significance. Acta Forestalia Fennica, 56, pp. 1-71.

Gatziolis, D., 2007. Lidar-derived site index in the U.S. Pacific Northwest—challenges and opportunities. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Espoo, Finland, vol. XXXVI, Part 3/W52, pp. 136–143.

Gustavsen, H., 1980. Site index curves for conifer stands in Finland, Folia Forest, 454, pp. 1-31.

Haralick, R. M., Shanmugan, K., and Dinstein, I., 1973. Textural features for image classification. IEEE Transactions on Systems, Man and Cybernetics, 3(6), pp. 610-621.

Haralick, R., 1979. Statistical and structural approaches to texture. Proceedings of the IEEE, 67(5), pp. 786-804.

Holopainen, M., Haapanen, R., Tuominen, S., and Viitala, R., 2008. Performance of airborne laser scanning- and aerial photograph-based statistical and textural features in forest variable estimation. In Hill, R., Rossette, J. and Suárez, J. (Eds.) Silvilaser 2008 proceedings, pp. 105-112.

Holopainen, M., Vastaranta, M., Yu, X., Hyyppä, J., Kaartinen, H., and Hyyppä, H., 2009. Forest site type classification using single tree level Airborne Laser Scanning. In: Popescu, S., Nelson, R., Zhao, K. & Neuemschwander, A. (Eds.). Silvilaser 2009 proceedings, pp. 72-80.

Hyyppä, J. and Inkinen, M., 1999. Detecting and estimating attributes for single trees using laser scanner. The Photogrammetric Journal of Finland, 16(2), pp. 27-42.

Hägglund, B., 1975. Övre höjdens utveckling i blandbestånd av tall och gran - en orientereande studie [Development of dominant height in artificially regenerated pine- and spruce stands]. Skogshögskola. Institutionen för skogsproduktion. Intern rapport, 22 p.

Ilvessalo, Y., 1920. Kasvu- ja tuottotaulukot Suomen eteläpuoliskon mänty-, kuusi- ja koivumetsille [Growth- and yield tables for pine- spruce and birch stands in southern Finland]. Acta Forestalia Fennica, 15.4, pp. 1-96.

Jonsson, B., 1962. Om barrblandsskogens volymproduktion. Tallens och granens höjd och diameter tillväxt i orörda bestånd med olika grader av trädslagsblalandning i Norrland, Kopparbergs och Värmlands län. Summary: Yield of mixed coniferous forests. Height and diameter growth of Scots pine and Norway spruce in virgin stands at various proportions of mixture in northern Sweden and in the provinces of Kopparberg and Värmland. Meddelande från statens skogsforsknings institut 50(8), 143 p.

Karlsson, K., 1996. Kasvupaikkojen puuntuotoskyvyn ja puuston kasvun alueellinen vaihtelu Pohjanmaan rannikolta sisämaahan [Regional variation in growth and yield from coastal to inland in Ostrobothnia]. Folia Forestalia, 2, pp. 113-132.

Kilkki, P. and Päivinen, R., 1987. Reference sample plots to combine field measurements and satellite data in forest inventory. Department of Forest Mensuration and Management, University of Helsinki. Research notes, 19, pp. 210-215.

Koivisto, P., 1959. Kasvu- ja tuottotaulukoista [About growth and yield tables]. Communicationes Instituti Forestalis Fenniae, 51.8, 49 p.

Lappi-Seppälä, M., 1930. Die entwicklung gleichaltriger mischbestände aus kiefer und birke [Studies about development of mixed pine-birch stand]. Metsätieteellisen tutkimuslaitoksen julkaisuja 15(2), 241 p.

Leckie, D., Gougeon, F., Hill, D., Quinn, R., Armstrong, L., and Shreenan, R., 2003. Combined high-density lidar and multispectral imagery for individual tree crown analysis. Canadian Journal of Remote Sensing, 29, pp. 633-649.

Mailly, T., Turbis, S., Auger, I., and Pothier, D., 2004. The influence of site tree selection method on site index determination and yield prediction in black spruce stands in northeastern Québec. The Forestry Chronicle, vol 80/1, pp. 134-140.

Maltamo, M., Eerikäinen, K., Pitkänen, J., Hyyppä, J., and Vehmas, M., 2004. Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions. Remote Sensing of Environment, 90(3), pp. 319-330.

Mielikäinen, K., 1980. Mänty-koivusekametsiköiden rakenne ja kehitys. Summary: Structure and development of mixed pine and birch stands. Communicationes Instituti Forestales Fenniae, 99(3), 82 p.

Mielikäinen, K., 1985a. The structure and development of pine and spruce stands with birch mixture. In: Hägglund, B. and Peterson, G. (Eds.). Broadleaves in boreal silviculture - an obstacle or an asset? Sveriges lantbruksuniversitet. Institutuonen för skogsskötsel. Rapporter, 14, pp. 189-206.

Mielikäinen, K., 1985b. Koivusekoituksen vaikutus kuusikon rakenteeseen ja kehitykseen. Summary: Effect of an admixture of birch on the structure and development of Norway spruce stands. Communicationes Instituti Forestalis Fenniae, 133, 79 p.

Muinonen, E. and Tokola, T., 1990. An application of remote sensing for communal forest inventory. Proceedings from SNS/IUFRO workshop: The usability of remote sensing for forest inventory and planning, 26-28 February 1990, Umeå, Sweden. Remote Sensing Laboratory, Swedish University of Agricultural Sciences, Report 4, pp. 35-42.

Næsset, E., 1997. Determination of mean tree height of forest stands using airborne laser scanner data. ISPRS Journal of Photogrammetry and Remote Sensing, 52(2), pp. 49-56.

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), pp. 88-99.

Næsset, E., 2004. Accuracy of forest inventory using airborne laser-scanning: Evaluating the first Nordic full-scale operational project. Scandinavian Journal of Forest Research, 19(6), pp. 554-557.

Næsset, E., Gobakken, T., Holmgren, J., Hyyppä, H., Hyyppä, J., Maltamo, M., Nilsson, M., Olsson, H., Persson, Å., and Söderman, U., 2004. Laser scanning of forest resources: the Nordic experience. Scandinavian Journal of Forest Research, 19(6), pp. 482-499.

Nyyssönen, A., 1954. Hakkauksilla käsiteltyjen männiköiden rakenteesta ja kehityksestä [About stucture and development of managed pine stands]. Acta Forestalia Fennica, 60, 194 p.
Oikarinen, M., 1983. Etelä-Suomen viljeltyjen rauduskoivikoiden kasvatusmallit. Summary: growth and yield models for silver birch (Betula pendula) plantations in southern Finland. Communicationes Instituti Forestalis Fenniae, 113, pp. 1-75. (In Finnish with English summary)

Ojansuu, R., 2005. Kasvupaikka ja puuntuotoskyky [Site type, growth and yield]. In Hynynen, J., Valkonen, S., and Rantala, S. (edit.). Tuottava metsänkasvatus. Metsäntutkimuslaitos ja Metsäkustannus Oy, pp. 49-62. (In Finnish)

Popescu, S., Wynne, R., and Nelson, R., 2003. Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. Canadian Journal of Forest Research, 29(5), pp. 564-577.

Tamminen, P., 1993. Pituusboniteetin ennustaminen kasvupaikan ominaisuuksien avulla Etelä-Suomen kangasmetsissä [Estimation of site index in mineral soils using site type in southern Finland]. Folia Forestalia, 819, 26 p.

Tomppo, E., 1991. Satellite image-based national forest inventory of Finland. International Archives of Photogrammetry and Remote Sensing, 28, pp. 419-424.

Vega, C. and St-Onge, B., 2008. Mapping site index and age by linking a time series of canopy height models with growth curves. Forest Ecology & Management, 257(3), pp. 951-959.

Vehmas, M., Eerikäinen, K., Peuhkurinen, J., Packalén, P. and Maltamo, M., 2008. Airborne laser scanning for the identification of boreal forest site types. In Hill, R., Rossette, J. and Suárez, J. 2008. Silvilaser 2008 proceedings, pp. 58-65.

Vehmas, M., Eerikäinen, K., Peuhkurinen, J., Packalén, P. and Maltamo, M., 2009. Identification of boreal forest stands with high herbaceous plant diversity using airborne laser scanning. Forest Ecology and Management 257(1), pp. 46-53.

Vuokila, Y., 1956. Etelä-Suomen hoidettujen kuusikoiden kehityksestä [About development of managed spruce stands in southern Finland]. Communicationes instituti Forestalis Fenniae, 48.1., 138 p.

Vuokila, Y. and Väliaho, H., 1980. Growth and yield models for conifer cultures in Finland, Commun. Inst. Forest Fenn., 99.2, pp. 1-271.