Silva Fennica 1992, Vol. 26 N:o 4: 231–239

Calibration of upper diameter models in large-scale forest inventory

Kari T. Korhonen

TIIVISTELMÄ: YLÄLÄPIMITTAMALLIEN KALIBROINTI SUURALUEEN METSÄNINVENTOINNISSA

Korhonen, K.T. 1992. Calibration of upper diameter models in large-scale forest inventory. Tiivistelmä: Yläläpimittamallien kalibrointi suuralueen metsäninventoinnissa. Silva Fennica 26(4): 231–239.

Models for estimating the upper diameter of trees were constructed using sample tree data measured in the 7th National Forest Inventory in Finland. Calibration of the models was tested with data from the 8th National Forest Inventory. The results show that using mixed estimation for combining the two data sets improves the reliability of the models. Models and methods used in this study can be recommended for use in forest inventories.

Tutkimuksessa esitetään kapenemismallit yläläpimitan estimoimiseksi männylle, kuuselle ja koivulle. Laadinta-aineistona käytettiin valtakunnan metsien 7. inventoinnin koepuita. Mallien kalibroimista tutkittiin 8. inventoinnin koepuuaineistolla. Tutkimuksessa havaittiin, että koepuiden käyttöä voidaan tehostaa käyttämällä otantateoreettiseen lähestymistapaan perustuvaa menetelmää vanhan ja uuden koepuutiedon yhdistämiseen.

Keywords: forest inventories, diameter, models, simulation, estimation.
FDC 52

Author's address: The Finnish Forest Research Institute, Joensuu Research Station. P.O. Box 68, SF-80101 Joensuu, Finland.

Accepted March 17, 1993

Notation

d = diameter at breast height (1.3 meters from ground level) [cm]
h = height of the tree from ground level to the top of the tree, [m]
t = age of the tree at breast height [years]
d_6 = upper diameter of the tree measured at a height of 6 meters from ground level [cm]
$\hat{d}_{6L}$ = upper diameter estimate obtained using the taper curve models of Laasasenaho (1982) and measured d and h [cm]
G = basal area of the growing stock [m^2/ha]
Y = p-coordinate of the plot (distance from the Equator) [km]
YC = $(Y - 6620)/1000$
X = i-coordinate of the plot (distance from the Greenwich meridian) [km]
XC = $(X - 60)/1000$

1 Introduction

In Finnish forest inventory systems the volume of trees or logs is usually estimated using the volume or taper curve models presented by Laasasenaho (1982). These models estimate the volume or taper curve either as a function of diameter at breast height (d in this text) and height (h in this text) or as a function of d, diameter at the height of six meters (d_6 in this text), and h. The use of d_6 as a regressor improves the reliability of volume estimates (Laasasenaho 1982, Kilkki 1983). In small sub-populations, volume estimates obtained using d and h only can be markedly biased.

Measuring d_6, however, is time-consuming and expensive (Kilkki 1983). Therefore Päivinen (1978) and Vähäsaari (1989), for example, have studied estimation of upper diameter with regression models using stand and tree variables as regressors. In both studies, statistically significant models were found.

A correctly formulated regression model gives unbiased results for the population used for constructing the model. If the population changes, e.g. as a function of time, calibration must be used to combine the existing information (e.g. regression model for upper diameter) with new measurements. The main methods for using prior information can be grouped in three classes (Kangas et al. 1990):

1. Bayes estimators (see e.g. Green and Strawderman 1985)
2. Methods based on the sampling theoretical approach (Teräsvirta 1981, Burk and Ek 1982, Pekkonen 1983)
3. Random parameter models (Lappi 1986, Lappi and Bailey 1988)

The use of upper diameter models for estimating the volume of sample trees in the National Forest Inventory of Finland was investigated in this paper. A sampling theoretical approach (mixed estimation) was used to combine sample tree information from two sources. At the first stage of the study upper diameter models were constructed for pine (*Pinus sylvestris*), spruce (*Picea abies*), silver birch (*Betula pendula*) and white birch (*Betula pubescens*) using the data from the 7th National Forest Inventory of Finland in southern Finland (called NFI7 in this text). Models and their use in volume estimation were tested in the data used for constructing the models.

At the second stage of the study the use of the models was investigated with data measured in the 8th National Forest Inventory of Finland (NFI8). The upper diameter models constructed at Stage 1 are re-estimated for each National Forestry Board district. Use of the models for estimating volume was studied with the help of a simulation.

2 Materials and methods

2.1 Study materials

Data measured in NFI7 were used as first level data in this study. The sample tree data from NFI7 consisted of 19559 pines, 19181 spruces, 1861 silver birches and 5150 white birches with measured upper diameter. From each sample tree the following measured dimensions were used in this (recording units in brackets):

- diameter at breast height [cm]
- height of the tree from ground level to top of the tree [dm]
- upper diameter at the height of 6 meters from the ground [cm]
- age of the tree [a]

Sample trees were selected with a relascope (basal area factor 2). For each plot, several characteristics describing the site and growing stock were recorded (Valtakunnan metsien... 1977). Data were measured during 1977–1983. Data used in this study covered National Forestry Board Districts 0 to 17 (see Fig. 1).

Sample tree data from NFI8 measured during 1986–1990 were used as second level data. Se-

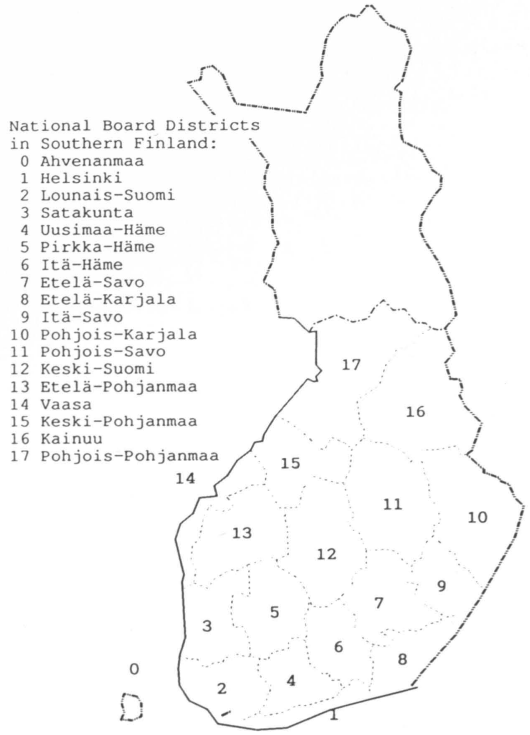

Fig. 1. The study area (National Board districts 1–17).

lection of trees and measured characteristics were, for the most part, the same as in NFI7 (Valtakunnan metsien... 1989).

2.2 Estimation of model parameters

Upper diameter models were constructed using ordinary least squares (OLS) technique. This method gives the best linear unbiased estimates for the parameters under assumptions:

1. The residuals of the model are not correlated
2. Variance of the residuals is constant

Assumption 1 probably does not hold in this case. The data consist of observations grouped in plots, and it is evident that residuals are correlated within plots. Even though Assumption 1 does not hold OLS gives unbiased estimates of the parameters.

To fulfill Assumption 2, the logarithm of d_6 was used as the dependent variable.

2.3 Calibration of models

The problem of calibrating models is the same as the problem of combining information from two levels: first-level information (= existing models or old sample tree measurements) and second-level information (= 'new' sample tree measurements). Let us denote the model to be estimated by Formula (1).

$$y = X\beta \tag{1}$$

where y = the vector containing the values of the dependent variable
X = matrix containing the values of the independent variables, and
β = a parameter vector

Let us denote the first-level data on y and X by r and R, respectively. Correspondingly, the second-level data on y and X are denoted by s and S.

To obtain the second-level estimates for parameter vector β, the first-level and second-level data are combined using Formula (2) (Theil and Goldberger 1961, Teräsvirta 1981).

$$\beta = (S'S + kR'R) - 1 (S's + kR'r) \tag{2}$$

where k = weight of the prior information

If the second-level information is used for estimating only some of the parameters in β, the model (1) must be re-written as follows:

$$r = R_1\beta_1 + R_2\beta_2 \tag{3}$$

$$s = S_1\beta_1 + S_2\beta_2 \tag{4}$$

where R_1 and S_1 contain regressors whose parameters are estimated using only the first-level information; R_2 and S_2 contain regressors whose final parameter estimates are obtained using both first and second-level information; and β_1 and β_2 are the respective parameter vectors.

The reason for dividing the regressors into two components can be, e.g., that the second-level data do not contain valid information for estimating the parameters in vector β_1.

Under this notation the β_1-vector is estimated using, e.g., OLS and the first-level data. The second-level estimate for β_2-vector is obtained with formula (5) using both first-level and second-level data (Lappi, J., The Finnish Forest

Research Institute, Suonenjoki Research Station, pers. comm. 1992).

$$\beta_2 = (S_2'S_2 + kR_2'R_2)^{-1} (S_2'u + kR_2'v) \quad (5)$$

where $u = s - S_1\beta_1$,
$v = r - R_1\beta_1$, and
β_1 = first level estimate for β_1

Parameter k in Formula (5) determines the weight given to the first-level data. Naturally, the weight of the first-level data also depends on the number of observations in the $(R_2'R_2)$ and $(S_2'S_2)$ matrices. In this study the data used as first-level information covered all southern Finland. Because the models were calibrated districtwise, the second-level data consisted of the sample trees measured in one district. Therefore, the effect of two data sets was scaled to equality by multiplying the k parameter by m/n, where m is the number of observations in the second-level data and n is the number of observations in the data used as prior information. The final formula for estimation of β_2 is presented in Formula (6).

$$\beta_2 = (S_2'S_2 + k\ m/n\ R_2'R_2)^{-1} (S_2'u + k\ m/n\ R_2'v) \quad (6)$$

2.4 Sampling simulations

Calibration of the upper diameter models was studied by simulating sampling of sample trees in the NFI8 data. Because inventories are carried out districtwise, models need to be calibrated independently for each National Forestry Board district (see Fig. 1). The $(R_2'R_2)$ and $(R_2'v)$ matrices (see Formula (6)) formed from the NFI7 data were used as prior information. In the simulations, upper diameter was assumed to be measured from varying number of trees. From these sample tree measurements, $(S_2'S_2)$ and $(S_2'u)$ matrices were formed and combined with prior information to estimate the parameter vector β_2 (see Formula (6)). Volumes of trees were estimated using measured d and h, and $\hat{d}_6$ obtained with the calibrated model. 'True' volumes of the trees were estimated using measured d, d_6, and h.

Sample trees for the calibration were selected using systematic sampling with a random starting point. The starting point was selected independently for each plot with the help of random numbers. To study the effect of number of sample trees simulation was done taking every 3rd, 5th, 10th, 20th, and 30th tree as a sample tree. Values 0, 0.5, 1, and 2 were tested for k in Formula (6).

Each sampling simulation (= combination of k and number of sample trees) was repeated 100 times with different random numbers used in selecting sample trees. In each simulation, the bias and the relative error variance of treewise volume estimates were calculated. 'True' and estimated mean volumes (m^3/ha) for different National Forestry Board districts (see Fig. 2) were also calculated in each simulation. Mean and standard deviation of the difference between true and estimated mean volume were calculated to describe the effect of sampling error in selecting calibration trees. The trees selected as sample trees were included in the calculation of biases and error variances.

3 Results

3.1 Upper diameter models

As the dependent variable in his models Vähäsaari (1989) used the logarithm of upper diameter. The independent variables were different transformations of the following variables:

- diameter at breast height
- height
- total age of the tree
- $\hat{d}_{6L}$ = upper diameter estimated with the taper curve models of Laasasenaho (1982) using d and h as regressors
- location of the plot
- height of the plot above sea level
- effective temperature sum (degree days) of the plot
- basal area of the growing stock

The model of Vähäsaari (1989) was used as a basis for constructing the model for this study. It was found that the parameter estimates of Vähäsaari's model are very unstable – their values are highly dependent on the data used for estimating the parameters. The reason for the unstability of the parameter estimates is the high correlation between regressors.

Table 1. Parameter estimates and their t-values, RMSE and R^2 values of the upper diameter models by tree species.

	Estimate	t-value	Estimate	t-value
	Pine		Spruce	
Constant	–0.002509	0.289	–0.015615	1.460
d^2	–0.000038	12.015	0.000038	11.029
h^2	0.000179	23.894	–0.000040	4.185
t	0.000204	5.547	0.000393	9.287
d/t	–0.052523	7.706	–0.039257	5.909
$\hat{d}_{6L}$	0.976351	323.956	0.979847	306.627
ln(G)	–0.003111	2.566	0.010456	5.357
YC	0.072682	4.059	0.043968	2.073
YC^2	0.098399	3.433	0.035077	1.077
XC	0.070963	2.840	0.025102	0.827
XC^2	0.060902	1.572	0.013995	0.325
YC*XC	–0.207773	4.507	–0.10750	1.987
RMSE	0.083		0.090	
R^2	0.971		0.968	
	Silver birch		White birch	
Constant	0.149516	10.604	0.108766	11.582
h^2	0.000106	4.995	0.000173	5.915
t	–0.000353	2.331	0.000235	2.155
d/t	–0.105945	5.521	–0.112499	6.363
$\hat{d}_{6L}$	0.950042	106.387	0.947809	141.241
RMSE	0.099		0.137	
R^2	0.967		0.951	

Therefore, a simplified version of Vähäsaari's model (1989) was used in this study. For pine and spruce, a satisfactory model was found using diameter, height, $\hat{d}_{6L}$, and age of the tree, basal area of the growing stock, and location of the plot as regressors. The model is presented in Equation (7) (note that in Vähäsaari's model the total age of the tree was used; in this study the age at breast height was used).

$$\ln(d_6) = b_0 + b_1{*}d^2 + b_2{*}h^2 + b_3{*}t + b_4{*}d/t + b_5{*}\ln(\hat{d}_{6L}) + b_6{*}\ln(G) + b_7{*}YC + b_8{*}YC^2 + b_9XC + b_{10}{*}XC^2 + b_{11}{*}YC{*}XC \quad (7)$$

where $b_0,\ldots,b_{11}$ are parameters, and $\ln(d_6)$ is the natural logarithm of d_6.

In the models of birches the coordinates, basal area and diameter of the tree were not significant regressors. Thus, for birches only h^2, t, d/t, and $\hat{d}_{6L}$ were used as regressors.

Parameter estimates and their t-values for each tree species are presented in Table 1. The root mean square errors (RMSE) and degree of determination values (R^2) are also presented. The parameters in Table 1 were estimated using the NFI7 data.

3.2 Use of the upper diameter models in volume estimation

3.2.1 Effect of residual error on the volume estimates

When volume is estimated as a function of $\hat{d}_6$ and $\hat{d}_6$ contains random (estimation) error, volume estimates are biased if the relation between volume and d_6 is non-linear (see e.g. Kilkki 1979). The effect of the estimation error of d_6 on the

Table 2. Mean and standard deviation (Sd) of residuals when volumes are estimated with d, h, and $\hat{d}_6$. Pine.

Relative, %		Absolute, dm^3	
Mean	Sd	Mean	Sd
0.11	7.45	0.04	38.46

volume estimates was studied in the data of NFI7 when model (7) was used for estimating d_6. Volumes estimated with measured d, h, and d_6 were used as 'true' volumes (see Table 2 for results). The relative mean and standard deviation were calculated for variable $(v - \hat{v}) / \hat{v}$, where v is the true volume and $\hat{v}$ is the estimated volume.

Results show that the bias caused by residual error in $\hat{d}_6$ is neglible. There are two reasons for this:

1. The relation between d_6 and volume is close to linear (when d and h are fixed)
2. The residual error of the upper diameter model (7) is relatively small (see Table 1)

3.2.2 Calibration of models

The reliability of the treewise volume estimates and the mean volume estimates by tree species were studied in different National Forestry Board districts using the data of NFI8. In this paper are presented the results for Districts 3 and 10 (see Fig. 1).

In the simulations the upper diameter model (7) was used in three different ways:

1. No prior information was used, i.e. only sample trees measured in the simulation were used to estimate the parameters of the model
2. Both sample trees and prior information were used and values 0.5, 1, and 2 were used for k
3. Only prior information was used, i.e. no sample trees were measured for calibrating the upper diameter models

Relative bias and RMSE of volume estimates at different values of k and different number of sample trees are presented in Figs. 2 and 3 for Districts 3 and 10, respectively. Value 0 for k refers to case 1 in the list above – no prior information is used. Value 0 for number of sample trees refers to case 3 in the list above – prior information is used without calibration.

In District 3, prior information improves the reliability of treewise upper diameter estimates in all cases studied. In District 10, prior information improves the results only when the number of sample trees is less than 460 (if k = 0.5). If more than 460 sample trees are measured, the

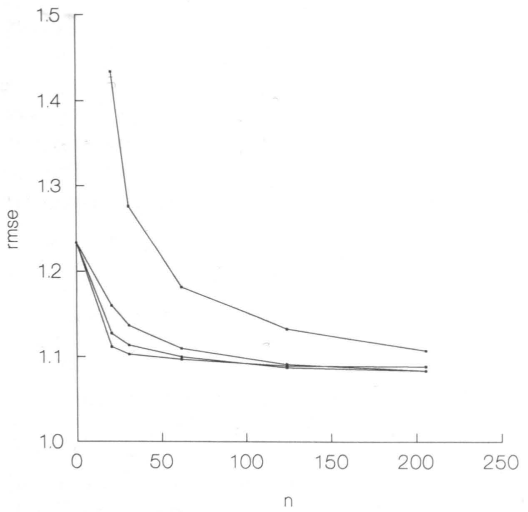

Fig. 2. Root mean square error (RMSE) of treewise volume estimates as a function of number of sample trees with different k-values. Pine, District 3.

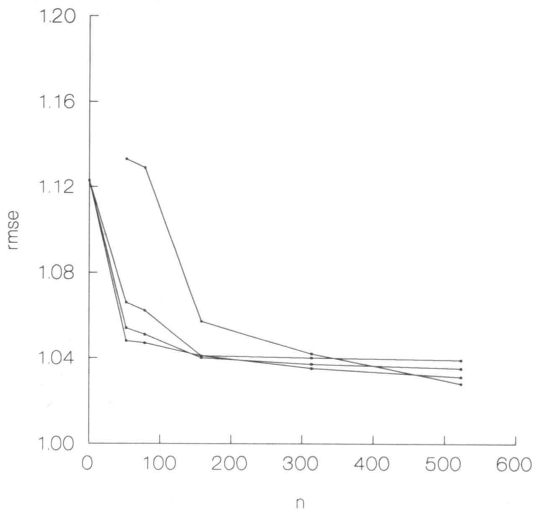

Fig. 3. Root mean square error (RMSE) of treewise volume estimates as a function of number of sample trees with different k-values. Pine, District 10.

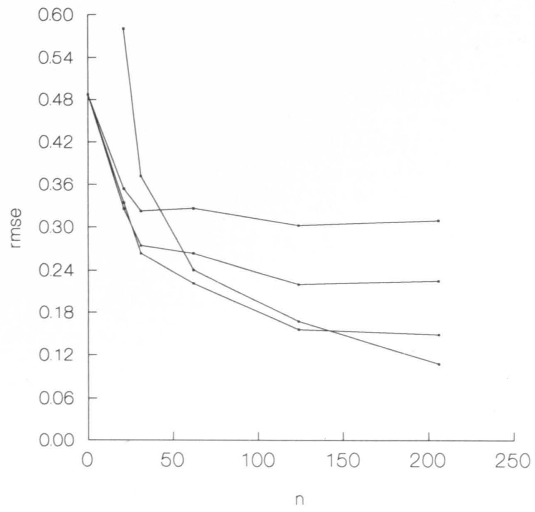

Fig. 4. Root mean square error (RMSE) of mean volume estimate of District 3 as a function of number of sample trees with different k-values. Pine.

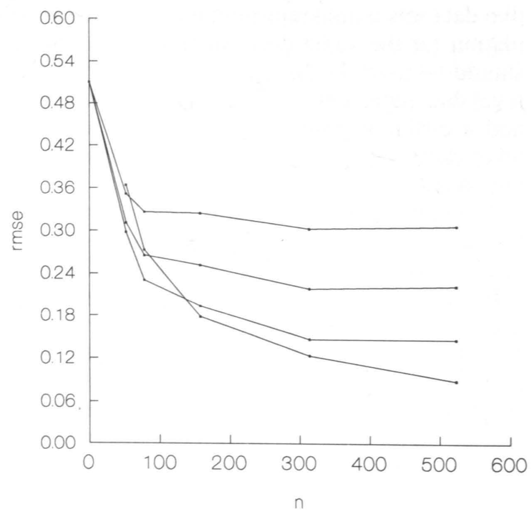

Fig. 5. Root mean square error (RMSE) of mean volume estimate of District 10 as a function of number of sample trees with different k-values. Pine.

best way to estimate the parameters is estimation without prior information. The reason for this is that the model used as prior information was found to be somewhat biased for the data set from NFI8. Tests in this data showed that model (7) overestimates upper diameters by 0.28 cm in District 3 and by 0.23 cm in District 10.

Figs. 4 and 5 show the RMSE estimates of the mean volume estimates in 100 replications in Districts 3 and 10, respectively. When only a few sample trees are measured, the use of prior information markedly improves the reliability of mean volume estimates. Because the prior information is biased, it is not advantageous to use it when the number of sample trees is large.

4 Conclusions and discussion

In the upper diameter model for pine YC, YC^2, XC, and YC*XC of the variables describing geographical location were significant regressors (at the 5 % level of probability). In the model for spruce, only YC and YC*XC were significant. The non-significant variables describing location were included in the model only to keep the trend surface quadratic (see Ripley 1981). For birch, none of the coordinate variables were significant.

To improve accuracy of the numerical analysis, Ripley (1981, page 30) recommends rescaling coordinates to limits from –1 to +1. In this paper y- and x-coordinates were rescaled to vary approximately within limits 0 and 0.7.

The use of prior information was evaluated in this study using two criteria – reliability of treewise volume estimates and reliability of mean volume estimates for whole calculation units. In general, the reliability of mean volume estimates is of more interest. It should be noted, however, that reliable treewise estimates also guarantee reliable results for small sub-units, e.g. different age classes. Therefore, reliability of the mean volume estimate for the whole calculation unit should not be the only criterion when different methods are compared.

Both tree and area results show that using prior information with small weight (k = 0.5) is the safest choice. If prior information is used, 50–100 tapering sample trees per district and tree species seems to be adequate.

In this study, factor m/n was used for scaling the sizes of the first- and second-level data. If the two data sets would represent exactly same population (at the same point in time), no scaling should be used. In this case, however, the first-level data represented a larger geographical area and a different point in time than the second-level data. Therefore, scaling was regarded as necessary.

The upper diameter models used as prior information seemed to be slightly biased for both districts studied. Both prior information data and calibration data were samples from two large populations that overlap geographically but represent different points in time. Both data sets contain sampling errors. Therefore, it can be regarded even justified to use estimators that are 'biased' in the direction of the prior information. Naturally, this is not the case when the goal of the inventory is to discover changes, in which case prior information must be replaced with measurements.

In this study large scale trends of the residual of the upper diameter model were estimated using ordinary least squares technique. It would also be possible to use Kriging-methods to calibrate the models for small areas (see e.g. Henttonen 1991). The main problem in using Kriging-methods in this context is the estimation of the covariance functions. It is probable that the form of the functions varies in different parts of the country. Furthermore, the spatial correlation is probably not isotropic, especially near the coast. Using Kriging for smoothing residuals requires inversion of the covariance matrix of the observations (Henttonen 1991). In a large set of data, such as the NFI-data, this is technically difficult. Despite these drawbacks, research on using Kriging-methods e.g. in establishing local volume functions from inventory data, could lead to applicable solutions.

Acknowledgements: Dr. Helena Henttonen, Lic.Sc. Annika Kangas, Dr. Juha Lappi, and Dr. Erkki Tomppo have given valuable comments on the manuscript. Dr. Joan von Weissenberg has revised the English text. The author wishes to thank all these people for their help.

References

Green, E.J. & Strawderman, W.E. 1985. The use of Bayes/empirical Bayes estimation in individual tree volume equation development. Forest Science 31: 975–990.

Burk, T.E. & Ek, A.R. 1982. Application of empirical Bayes/James Stein procedures to simultaneous estimation problems in forest inventory. Forest Science 28: 753–771.

Henttonen, H. 1991. Kriging in interpolating July mean temperatures and precipitation sums. Reports from the Department of Statistics, University of Jyväskylä. 42 p.

Kangas, A., Kangas, J., Korhonen, K.T., Maltamo, M. & Päivinen, R. 1990. Metsää kuvaavat mallit. [Modelling forests. In Finnish.] Silva Carelica 17. 143 p.

Kilkki, P. 1979. An outline for a data processing system in forest mensuration. Silva Fennica 13(4): 368–384.

— 1983. Sample trees in timber volume estimation. Folia Forestalia 182. 35 p.

Laasasenaho, J. 1982. Taper curve and volume functions for pine, spruce and birch. Communicationes Instituti Forestalis Fenniae 108. 74 p.

Lappi, J. 1986. Mixed linear models for analyzing and predicting stem form variation of Scots pine. Communicationes Instituti Forestalis Fenniae 134. 69 p.

— & Bailey, R.L. 1988. A height prediction model with random stand and tree parameters: an alternative to traditional site index methods. Forest Science 34(4): 907–927.

Päivinen, R. 1978. Kapenemis- ja kuorimallit männylle, kuuselle ja koivulle. Summary: Taper and bark thickness models for pine, spruce and birch. Folia Forestalia 353. 31 p.

Pekkonen, T. 1983. Leimikon puuston tilavuuden arviointi regressioennustinta käyttäen. [Regression estimators in predicting the volume of the stand marked for cutting. In Finnish]. Metsäntutkimuslaitoksen tiedonantoja 86. 63 p.

Ripley, B.D. 1981. Spatial Statistics. John Wiley & Sons Inc., New York. 252 p.

Teräsvirta, T. 1981. Some results on improving the least squares estimation of linear models by mixed estimation. Scandinavian Journal of Statistics 8: 33–38.

Theil, H. & Goldberger, A.S. 1961. On pure and mixed statistical estimation in economics. International Econom. Rev. 2: 65–78.

Vähäsaari, H. 1989. Yläläpimitan estimointi kahden tunnuksen runkokäyrän ja metsikkötunnusten avulla. [Estimation of upper diameter with taper curve model and standwise regressors. Unpublished manuscript in Finnish.] The Finnish Forest Research Institute, Helsinki. 23 p.

Valtakunnan metsien inventoinnin kenttätyöohjeet. Yleinen osa. 1977. [Field instructions for National Forest Inventory of Finland. In Finnish.] The Finnish Forest Research Institute. Helsinki. 59 p.

Valtakunnan metsien 8. inventointi. Kenttätyön ohjeet. Pohjois-Karjalan versio. 1989. [Field instructions for the 8th National Forest Inventory of Finland. In Finnish.] The Finnish Forest Research Institute, Helsinki. 96 p.

Total of 17 references