

Improving Multi-Source Forest Inventory by Weighting Auxiliary Data Sources

Sakari Tuominen and Simo Poso

Tuominen, S. & Poso, S. 2001. Improving multi-source forest inventory by weighting auxiliary data sources. *Silva Fennica* 35(2): 203–214.

A two-phase sampling design has been applied to forest inventory. First, a large number of first phase sample plots were defined with a square grid in a geographic coordinate system for two study areas of about 1800 and 4500 ha. The first phase sample plots were supplied by auxiliary data of Landsat TM and IRS-1C with principal component transformation for stratification and drawing the second phase sample (field sample). Proportional allocation was used to draw the second phase sample. The number of field sample plots in the two study areas was 300 and 380.

The local estimates of five continuous forest stand variables, mean diameter, mean height, age, basal area, and stem volume, were calculated for each of the first phase sample plots. This was done separately by using one auxiliary data source at a time together with the field sample information. However, if the first phase sample plot for which the stand variables were to be estimated was also a field sample plot, the information of that field sample plot was eliminated according to the cross validation principle. This was because it was then possible to calculate mean square errors of estimates related to a specific auxiliary data source.

The procedure produced as many estimates for each first phase sample plot and forest stand variable as was the number of auxiliary data sources, i.e. seven estimates: These were based on Landsat TM, IRS-1C, digitized aerial photos, ocular stereoscopic interpretation from aerial photographs, data from old forest inventory made by compartments, Landsat TM95–TM89 difference image and IRS96–TM95 difference image. The final estimates were calculated as weighted averages where the weights were inversely proportional to mean square errors. The alternative estimates were calculated by applying simple rules based on knowledge and the outliers were defined. The study shows that this kind of system for finding outliers for elimination and a weighting procedure improves the accuracy of stand variable estimation.

Keywords remote sensing, forest inventory, two-phase sampling, weighting

Authors' addresses *Tuominen*, Finnish Forest Research Institute, Unioninkatu 40 A, FIN-00170 Helsinki, Finland; *Poso*, Department of Forest Resource Management, P.O. Box 27, FIN-00014 University of Helsinki, Finland **E-mail** sakari.tuominen@metla.fi

Received 6 August 1999 **Accepted** 24 January 2001

1 Introduction

The social and economic environment of forestry has undergone rapid change in Finland during the 1990's. Forestry and forest management have been given new and tight requirements regarding the protection of ecologically important biotopes. Forest inventory and planning data are expected to describe the forest characteristics more accurately. Furthermore, the attribute data of forest properties should be georeferenced. Thus the attribute data have to be linked to some geographical coordinate system. In addition to that (or in spite of that) the acquisition of the inventory data should be more efficient and inexpensive, requiring less personnel – a resource, which is already scarce in forest planning organisations.

Based on the above-mentioned requirements, one possibility for attempting to develop and reform forest inventory and management is to use two-phase sampling in combination with auxiliary data sources. This method allows more flexible means for managing territorial constraints, which are caused for example by the requirements of forest and environmental protection legislation, than traditional compartment based management (Holmgren and Thuresson 1995). For sampling based inventory it is appropriate to define forest as a population of sample plots. In two-phase sampling, where the first phase sample is a systematic grid of sample plots, all inventory units are geographically determined, which means that their location is exactly known and those sampling units can be used in successive inventories. Auxiliary data are by definition data that are not accurate enough for the forest management task unaccompanied but is correlated with the true values of forest (stand) characteristics, which the forest managers are actually interested in. The exact values of forest characteristics are measured or they can be derived from variables measured in the second phase (field) plots. Auxiliary data offer forest inventory the means by which to estimate forest characteristics for all first phase sample plots and, through them, to all forest areas (Waite 1991, Poso and Waite 1996).

The aim of this study was to explore possibilities for enhancing the estimation of forest stand variables for any given location with the help of various auxiliary data sources.

2 Study Material

2.1 Study Areas

Two separate study areas were used in this study. The Haukilahti forest area (1768 ha), which is owned by UPM-Kymmene Oyj, is located in the Längelmäki municipality in southern Finland. The Nygård forest area (4522 ha) is owned by Metsä-Serla Oyj and is located in the Kuru municipality in southern Finland.

Both study areas are boreal forest dominated mainly by coniferous trees. Some characteristics of the study areas are described in Table 1. Mainly due to higher average fertility the volume of growing stock per hectare is higher in the Haukilahti study area.

2.2 First Phase Sample Plots

The first phase sample plots were defined as a grid of points, where each point indicated the location of a sample plot centre in the Finnish coordinate system. The distance between points was 20 m in the Haukilahti study area and 25 m in the Kuru study area. The total number of first phase sample plots was 44 204 in the Haukilahti study area and 72 357 in the Kuru study area.

2.3 Field Data

The selection of the second phase sample plots was based on proportional allocation. For this the first phase sample plots were stratified to 40 strata

Table 1. Description of forests of the study areas.

Forest variable	Haukilahti	Kuru
Mean volume total, m ³ /ha	144	113
Pine	42	56
Spruce	87	45
Broad-leaved	14	12
Standard deviation of plot volumes, m ³ /ha	147	100
Mean diameter, cm	18	15
Mean height, m	14	13
Mean age, years	66	51
Mean basal area, m ² /ha	16	15

on the basis the Landsat TM and IRS-1C panchromatic satellite images. In each stratum a number of second phase sample plots was selected to be measured in field in proportion to the number of first phase plots in that stratum. The minimum distance between the field plots was set as 100 m. The number of field plots in the Haukilahti study area was 300 and in the Kuru study area 380. Field data were measured during summer and autumn 1997. The field sample plots were concentric circular plots, from which trees were tallied according to the following rule:

Tree dbh, cm	Circular plot radius, m
5–14	7
14.1–	13

Trees less than 5 cm dbh were counted from two circles (radius 3.99 m) which were located 8 m

from the sample plot centre in the directions 0 and 180 degrees (north and south). The field data were recorded per tree species and tree storey for trees less than 5 cm dbh (Kaukokuvat metsien inventoinnissa 1997).

Field sample plots were plotted on digital aerial ortho-photographs, which were then used for location of plots in the field.

2.4 Auxiliary Data

The auxiliary data sources (7) used in this study are presented in Table 2.

The orientation of satellite images and other raster data operations were carried out with the SMI system, which is a toolkit of computer programs for forest inventory and monitoring. The name SMI is an abbreviation of Satellit-

Table 2. Auxiliary data sources.

Symbol of data source	Auxiliary data source	Date	Number of variables	Description
1	Landsat TM image 190/17	1.6.1995	7	Spectral values of 7 channels, resolution 30 m
2	IRS-1C PAN image 33/23	24.8.1996	2	Average and standard deviation of 5 × 5 pixel window, resolution 5 m
3	Digital aerial ortho-photo, colour-infrared	1995 ^{a)} 1997 ^{b)}	6	Averages and standard deviations of 10 × 10 pixel window from 3 channels (rgb), resolution 2 m ^{e)}
4	Visual aerial photo interpretation	1995 ^{a)} 1997 ^{b)}	7	Visually interpreted variables ^{d)}
5	Old inventory data	1996 ^{a)} 1991–95 ^{b)}	7	Variables taken from old inventory data (transferred to sample plot from the surrounding polygon) ^{e)}
6	Difference image Landsat TM95–Landsat TM89	1989 and 1995	6	Spectral value difference of 6 channels respectively (channels 1, 2, 3, 4, 5 and 7). TM89 image 189/17
7	Difference image IRS1C 96–Landsat TM95	1995 and 1996	2	Spectral value difference of IRS PAN and 2 TM channels (IRS PAN–TM ch2 and IRS PAN–TM ch3)

^{a)} Kuru study area

^{b)} Haukilahti study area

^{c)} Original numerical values were used without any calibration for central projection

^{d)} Variables include 1) development class, 2) dominant tree species, 3) proportion of deciduous trees, 4) site type, 5) additional features of site type, 6) mean tree height and 7) relative density of growing stock (interpretation unit in the Haukilahti study area was sample plot and in Kuru study area inventory compartment delineated from aerial photos)

^{e)} Variables include 1) land cover class, 2) site type, 3) development class, 4) basal area 5) mean tree height, 6) mean diameter and 7) mean tree age

tikuvat Metsien Inventoinnissa (Satellite-based Monitoring and Inventory). The SMI system was developed at the Department of Forest Resource Management of Helsinki University (Wang et al. 1997).

Auxiliary data from raster images were transferred to sample plots as nearest pixel values or mean values and standard deviations of raster windows centered around a sample plot, depending on the resolution of raster data source. Old forest inventory compartment data were transferred to sample plots so that all plots inside a certain polygon received the attribute data of that polygon. Visual interpretation of aerial photos in Kuru study area was carried out per compartments or forest stands. The auxiliary data were transferred to first phase sample plots in a similar way as with the old compartment data. In the Haukilahti study area the visual interpretation was carried out directly per first phase sample plots.

2.5 Difference Channel Values

The difference channel values were calculated for first phase sample plots. Firstly, the spectral values of channels 1, 2, 3, 4, 5 and 7 of the Landsat TM89 image were subtracted from respective channel values of Landsat the TM95 image. Secondly, the spectral values of the Landsat TM95 channels 2 (green, wavelength 520–600 nm) and 3 (red, wavelength 630–690 nm) were subtracted from the IRS-1C panchromatic image (wavelength 500–750 nm) spectral values. Only these two Landsat TM channels were used so that the older and newer images covered roughly similar wavelength areas.

The two study areas were then segmented into homogenous sub-areas (polygons) according to first phase sample plot difference image values. This was done using the Winseg automatic segmentation program. The functions and processes of the segmentation program are described by Wang et al. (1997). The difference values for the polygons were calculated as an average of sample plot values inside the polygon. These values were in turn transferred back to the sample plots so that all sample plots inside the certain polygon received similar difference channel values.

The processing of difference images differed from other raster data sources so that the attribute data were transferred to sample plots as polygon mean values and not as individual pixel values. The reason for this was to avoid disturbances (false changes) caused by recording and location errors and mixed pixels. These disturbances are detrimental especially in pixel-level difference image analysis (Varjo 1997).

3 Estimators

The statistical estimator used for estimating stand variables in this study was the *k*-nearest neighbour method (*k*-nn). In *k*-nn the stand variable estimates of a certain sample plot are determined according to the measured stand variables of *k* number of field plots, which are nearest to the plot to be estimated in the auxiliary data space (feature space). This method has been presented by Kilkki and Päivinen (1987) and applied in the Finnish national forest inventory (Tomppo 1990). For example in Swedish studies it has been recommended that a suitable number for *k* in forest inventory is between 5–10 (Nilsson 1997). In general, when the number of available reference plots is great, the estimation accuracy increases with increasing number of nearest neighbours, when *k* is between 0 and 20 (Nilsson 1997). In this study 5 was chosen as the value of *k* because of the rather small number of field plots.

The nearest neighbours of field sample plots in the feature space of respective auxiliary data source were defined separately for each auxiliary data source (Euclidean distance measures were used). As a result of the process a basic matrix of nearest neighbours was generated. Thus, the maximum number of nearest neighbours for a sample plot in the matrix was 35 (7 auxiliary data sources \times 5 nearest neighbours). Because the nearest neighbours were calculated separately with different auxiliary data, it was possible that the same neighbour appeared several times in the matrix for a certain sample plot.

In the test for estimating accuracy the sample plot to be estimated was removed from the group of potential nearest neighbours (reference plots),

so the estimate was independent from the measured stand variables (ground truth). The stand variable estimates were then compared with stand variables measured in field. This method is statistically known as cross validation. Five stand variables were tested: mean diameter (at breast height), mean height, mean age, basal area and (stem) volume. Sample plot mean variables (diameter, height and age) were calculated as basal area-weighted averages.

3.1 Alternative Estimators for Comparison

- A. Nearest neighbours from the reference plots were determined and estimates were calculated with each individual auxiliary data source separately. The procedure resulted in seven alternative individual estimates for every stand variable.
- B. Estimates were calculated as arithmetic means of alternative estimates of estimator A.
- C. Estimates are calculated as weighted means of the seven alternative estimates of estimator A. The weighting factor for each auxiliary data source

was the inverse value of average mean square error (MSE) from the results of estimator A (Tables 3a and 3b). The weight of best auxiliary data source in estimation of each stand variable was set to 1.000 and others in proportion to it.

- D. Estimates were calculated from the nearest neighbour matrix by weighting each auxiliary data source and each stand variable separately. The weighting factor was the inverse value of MSE of certain stand variable using certain auxiliary data source (Tables 3a and 3b).

It was found that there exists a distinct correlation ($0.45 < r < 0.75$) between the residuals of estimates derived from different auxiliary data sources. As far as different auxiliary data sources do not produce equal sets of sample plots for k nearest neighbours, this correlation probably can be disregarded when giving weights for auxiliary data sources for the final estimates.

- E. Estimates were calculated from the nearest neighbour matrix by weighting each nearest neighbour (reference plot) on the basis of its distance to the plot to be estimated. The weighting factor was calculated as the inverse of the Euclidean distance

Table 3a. Weight factors for auxiliary data sources (Haukilahti study area).

Auxiliary data source	Weight					
	Diameter	Height	Age	Basal area	Volume	Average
1. TM95	0.751	0.709	0.727	0.602	0.609	0.680
2. IRS96	0.792	0.753	0.676	0.623	0.579	0.685
3. Digital aerial photo	0.968	0.980	0.833	0.812	0.739	0.867
4. Visual interpretation	1.000	1.000	1.000	1.000	1.000	1.000
5. Old inventory data	0.923	0.823	0.999	0.638	0.681	0.813
6. TM95–TM89	0.610	0.556	0.622	0.463	0.430	0.536
7. IRS96–TM95	0.724	0.664	0.652	0.596	0.499	0.627

Table 3b. Weight factors for auxiliary data sources (Kuru study area).

Auxiliary data source	Weight					
	Diameter	Height	Age	Basal area	Volume	Average
1. TM95	0.845	0.806	0.798	0.914	0.845	0.842
2. IRS96	0.737	0.747	0.633	1.000	0.909	0.805
3. Digital aerial photo	0.956	0.900	0.852	0.889	0.848	0.889
4. Visual interpretation	0.987	0.982	0.876	0.919	1.000	0.953
5. Old inventory data	1.000	1.000	1.000	0.896	0.932	0.966
6. TM95–TM89	0.623	0.564	0.655	0.596	0.586	0.605
7. IRS96–TM95	0.680	0.688	0.614	0.979	0.896	0.771

to the reference plot in auxiliary data space. For this process the zero distances (value of auxiliary data was exactly the same in the reference plot and in the estimated plot) were set to 0.001.

- F. The difference image IRS96–TM95 was used for selecting sample plots which are likely to have undergone changes other than natural growth (cuttings or natural disturbances) in the time between taking the two images. In estimation of selected plots the auxiliary data sources, which were judged as outdated, were given zero weight. Otherwise weighting was as in estimator C.
- G. Sample plots that were located nearer than 20 meters to the compartment boundary were removed. They were neither estimated nor used as reference plots. Otherwise this estimator was similar to estimator B. This method has been applied only for the Haukilahti study material.

The object of estimator A was to rank different auxiliary data sources on the basis of their usability for estimation and thereby give them different weights. The use of inverse values of error variances as weights is a common method (e.g. Cochran 1963). Weighting auxiliary data sources with inverse values of their mean square errors has been used successfully to combine different auxiliary data sources and to obtain better estimates than with any single auxiliary data source for example in Poso et al. (1999). The inverse value of distance or squared distance in auxiliary data feature space as weight has been applied for example in the multi-source national forest inventory (NFI) of Finland (Tomppo 1990).

3.2 Difference Channel Values in Selection of Changed Sample Plots (F)

Difference channel values were used when searching sample plots which had undergone drastic changes (natural or man-made) between the dates of the two images, and thereby to judge, which auxiliary data sources still were up to date and which were not. The aim was especially to find plots, which had been cut. In these plots the auxiliary data sources that were judged to be outdated would then receive a weight equal to zero. For this purpose a feasible difference image was IRS96–TM95, because most auxiliary data

were dated 1995 or later than that.

The basic hypothesis for this was: cutting or natural disturbance causes stand volume to decrease, which in turn increases the spectral value of the satellite image of the stand. Hereby areas likely to have undergone cuttings or natural disturbances were those, which had high channel difference value (new image–old image). In order to select sample plots, which most likely had undergone changes, a threshold value was calculated. Plots exceeding the threshold value were selected.

The threshold value was calculated in the following way: An average of each satellite image channel (that was used in difference image) was calculated on the basis of all first phase sample plots. The average channel value of old image was subtracted from the average channel value of the new image (IRS96–TM95). The eventual threshold value was calculated by summing the average differences of each channel and the average standard deviation of difference channels. Because the dates of cuttings and natural disturbances that have occurred recently in the study areas were not recorded in the field and neither was any other relevant information available, the above-mentioned somewhat arbitrary method was used in selecting plots that were likely to have changed. The proportion of first phase sample plots exceeding the threshold value was 10.5% in the Haukilahti study area and 12.2% in the Kuru study area.

On plots that were judged to have changed, the auxiliary data sources, which were dated older than the newer image used for difference channel values, were given zero-weight (actually the zero-weight was given to reference plots that were nearest neighbours according to the auxiliary data source).

Auxiliary data sources that were given zero weights in selected plots were:

Haukilahti study area	Kuru study area
– Old inventory data	– Landsat TM image 1995
– Landsat TM image 1995	– Difference image TM95–TM89
– Difference image TM95–TM89	– Digital aerial photo
	– Visual photo interpretation

4 Results

When stand variables were estimated with each auxiliary data source separately, the best individual auxiliary data source proved to be the visual interpretation of aerial photograph in the Haukilahti study area (for all stand variables), and in the Kuru study area the best auxiliary data sources were old inventory data (for diameter, height and age) and visual photo interpretation (for volume). The satellite images were generally the worst performing auxiliary data sources in estimation. The results (RMSE values) of stand variable estimates calculated separately on different auxiliary data sources in each study area are presented in Tables 4a and 4b, which present the comparison of individual auxiliary data sources.

The MSE values (squares of the RMSEs) were the basis of the weighting in estimators C and D. Calculating the estimates as arithmetic means (equal weights) of individual auxiliary data source estimates (estimator B) results lower RMSE value than the best individual auxiliary

data source in the Kuru study area. In the Haukilahti study area the best auxiliary data source (visual aerial photo interpretation) gives better estimates than estimator B. Weighting the auxiliary data sources with the inverse values of their MSEs (estimator C) improves the estimates in both study areas, and calculating the weighting factor separately for each stand variable (estimator D) further improves the estimation slightly. Estimators C and D gave better estimates than the best individual auxiliary data source in both study areas.

The estimates that were weighted on the basis of the Euclidean distances of the nearest neighbours in the feature space (estimator E) are distinctly inferior when compared to other estimators in both study areas (Tables 5a and 5b).

Using the difference image IRS96–TM95 in finding the sample plots, which are likely to have undergone changes, and discarding the outdated auxiliary data sources (estimator F) improves the estimates, when compared to estimator C, in the Haukilahti study area but deteriorates them in the

Table 4a. Root mean square errors of stand variable estimates in the Haukilahti study area calculated with individual auxiliary data sources (estimator A).

Auxiliary data source	Diameter, cm	Height, m	Age, a	Basal area, m ² /ha	Volume, m ³ /ha
1. TM95	9.98	7.09	39.67	9.63	113.01
2. IRS96	9.72	6.88	41.13	9.46	115.89
3. Digital aerial photo	8.79	6.03	37.06	8.29	102.58
4. Visual interpretation	8.65	5.97	33.83	7.47	88.18
5. Old inventory data	9.00	6.58	33.84	9.35	106.82
6. TM95–TM89	11.07	8.01	42.88	10.98	134.42
7. IRS96–TM95	10.17	7.33	41.89	9.68	124.80

Table 4b. Root mean square errors of stand variable estimates in the Kuru study area calculated with individual auxiliary data sources (estimator A).

Auxiliary data source	Diameter, cm	Height, m	Age, a	Basal area, m ² /ha	Volume, m ³ /ha
1. TM95	8.06	6.16	29.51	7.96	82.77
2. IRS96	8.63	6.40	33.15	7.61	79.80
3. Digital aerial photo	7.58	5.83	28.57	8.07	82.62
4. Visual interpretation	7.46	5.58	28.18	7.94	76.08
5. Old inventory data	7.41	5.53	26.37	8.04	78.79
6. TM95–TM89	9.39	7.36	32.59	9.86	99.37
7. IRS96–TM95	8.99	6.67	33.66	7.96	80.38

Table 5a. Root mean square errors (RMSE) for stand variables using different estimators (Haukilahti study area).

Estimator	Diameter, cm	Height, m	RMSE		
			Age, a	Basal area, m ² /ha	Volume, m ³ /ha
B	8.27	5.69	32.52	7.41	88.97
C	8.10	5.55	31.88	7.21	86.12
D	8.12	5.55	31.71	7.17	85.28
E	9.29	6.31	35.89	7.50	89.38
F	7.89	5.39	31.18	7.01	84.57
G	7.96	5.38	31.19	6.90	83.95

Table 5b. Root mean square errors (RMSE) for stand variables using different estimators (Kuru study area).

Estimator	Diameter, cm	Height, m	RMSE		
			Age, a	Basal area, m ² /ha	Volume, m ³ /ha
B	7.16	5.33	26.62	6.86	69.81
C	7.08	5.25	26.33	6.80	69.00
D	7.04	5.23	26.12	6.79	68.91
E	7.35	5.52	27.18	7.34	73.86
F	7.14	5.27	26.52	6.81	68.99

Table 6. Root mean square errors calculated from plots that have been judged to have changed on the basis of difference image IRS96–TM95.

		Diameter, cm	Height, m	RMSE		
				Age, a	Basal area, m ² /ha	Volume, m ³ /ha
Haukilahti	Estimator C	11.15	8.77	40.38	5.98	54.83
	Estimator F	9.91	7.93	40.37	3.67	30.79
Kuru	Estimator C	10.59	7.77	36.62	4.63	38.67
	Estimator F	10.89	7.89	37.60	4.70	38.67

Table 7. Root mean square errors of stand variable estimates calculated with individual auxiliary data sources using estimator G, Haukilahti study area.

Auxiliary data source	Diameter, cm	Height, m	Age, a	Basal area, m ² /ha	Volume, m ³ /ha
TM95	9.68	6.75	39.52	9.10	108.91
IRS96	9.96	7.03	41.31	9.38	116.83
Digital aerial photo	8.34	5.82	36.51	8.48	103.49
Visual interpretation	8.67	5.68	32.26	7.07	83.86
Old inventory data	8.61	6.29	31.42	8.33	96.56
TM95–TM89	11.11	8.03	43.87	11.15	139.42
IRS96–TM95	10.69	7.50	42.73	10.19	129.93

Kuru study area (Tables 5a, 5b and 6). Estimator G, where the sample plots that were located less than 20 meters from nearest compartment border were removed from the estimation, produced the best estimates for stand height, basal area and volume (Table 5a). The proportion of removed plots was about 24%. The effect on the estimates is similar with most of the auxiliary data sources (Table 7). Estimator G was tested only in the Haukilahti study area.

Tables 5a and 5b present RMSE values for estimated stand variables, when auxiliary data sources have been used together to apply different procedures of weighting. Table 6 presents RMSE values for the stand variables including only those sample plots that have been judged to have changed (on the basis of difference image). The results for estimator F are presented in comparison with estimator C.

Table 7 presents RMSE values of stand variables estimated using estimator G, in which plots, whose distance to compartment border is less than 20 meters, are removed (i.e., they are neither estimated nor used as reference plots). These results should be compared with the results for Table 4a.

5 Discussion

When the RMSE values of stand variable estimates were compared in different study areas they show that estimates in the Kuru area are better on average than in the Haukilahti area. Knowing the tendency of underestimating high values and overestimating low values (averaging) in the k-nn method, the following factors are the likely cause for this. In the first place the number of reference plots was greater in the Kuru area, which usually improves the estimation. In the second place the variation was greater in the Haukilahti field material, which means that the distribution of stand variables was wider (Figs. 1a and 1b). Because of these factors, it is likely that, on an average, the variation within k-nn reference plots in the Kuru field material would be smaller in the feature space than in the Haukilahti field material.

The results of different estimation methods

show that weighting auxiliary data sources with the inverse value of their mean square errors (MSE) improved the estimates. The results were slightly better when weights were given individually for each stand variable. This result is consistent in both study areas. Weighting the reference plots with the inverse value of their distance in the feature space did not improve the estimation, on the contrary the results of this method were markedly inferior compared to other estimation methods. This result was also consistent in both study areas. The experiment with estimator E indicated that it is not reasonable to give different weights based on Euclidean distances for individual reference plots in k-nn.

Furthermore, the results demonstrate that application of difference image to select areas that are likely to have changed improved the estimates in the Haukilahti study area but not in the Kuru study area. The improvement was distinguishable in the entire Haukilahti study material but especially evident when only the selected plots were considered. Based on this, it can be assumed that the selection was successful in the Haukilahti study material although this can not be verified on the basis of the field material, because neither cuttings, natural disturbances nor their dates were recorded in the field. The inferior results achieved in the Kuru study material, when this estimation method was used, may be caused by the fact that the estimates, which were judged to be outdated and thus rejected (zero-weighted), were based on visual and digital aerial photo interpretations that otherwise were among the best individual auxiliary data sources (Tables 3b and 4b). Because this method was somewhat arbitrary the results must be considered as mainly demonstrating a general trend. Furthermore a deficiency of this method was the small number of difference channels used (green and red areas). When observing changes it might have been more useful to have used a difference image that covered a wider spectral area including the infra-red wavelength, which usually has significance in interpretation of vegetation (Varjo 1997).

Estimator G, where sample plots located less than 20 meters from compartment borders were removed, also led to a slight improvement in estimation (tested only in the Haukilahti study material). The fact that the number of reference

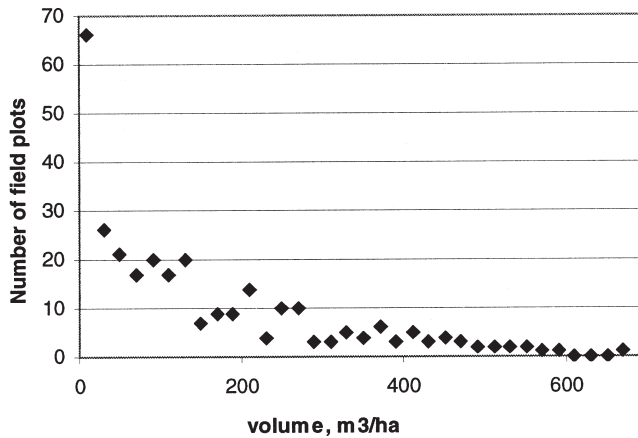


Fig. 1a. Volume distribution of Haukilahti field plots (20 m³/ha volume classes).

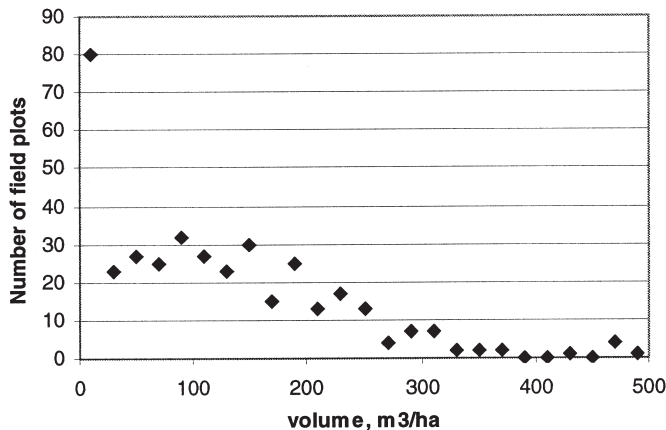


Fig. 1b. Volume distribution of Kuru field plots (20 m³/ha volume classes).

plots was intentionally reduced in this method, which usually decreases the estimation accuracy (Nilsson 1997), makes the improvement somewhat more significant. The basic hypothesis here was that sample plots located near compartment borders are prone to errors caused by positioning errors, tree shadows and mixed pixels, and rejecting them in estimation results in better estimates especially when using digital satellite images and aerial photos. However, when results of individual auxiliary data sources are examined (comparison of Table 4a and Table 7) it can be observed, that

the improvement of estimates was greatest in visual photo interpretation and in old inventory data. Also this method improved estimates for the Landsat TM image.

This result concerning the old inventory data may be caused by the fact that the borders of the old inventory compartments were in many cases erroneously located. These errors could be observed, when the old compartment vector map was printed on ortho-rectified aerial photos, which are regarded to give the best positional accuracy. The delineation of old inventory com-

partments gives reason to presume that compartments have been digitized using analog aerial photos, because the delineation errors are greatest in areas of high elevation variation. As far as visual interpretation of aerial photos is concerned the result is more difficult to explain. The improvement in estimation results, when plots located near compartment borders are removed, gives reason to suspect that some field plots may have been erroneously located. The improvement with Landsat TM image is probably explained by the removal of mixed-pixel plots (plots that receive their spectral value from pixels that are located in two or more different stands). In Landsat TM image these pixels are of great significance because of the low resolution (pixel size $30 \times 30 \text{ m}^2$) of the images. Furthermore the geometry of the TM image was found to be quite poor, when observed visually. When some easily recognisable objects like small lakes were observed visually, it was found that in some channels there was a transition of one pixel-column sideways in their location. The IRS-1C image in turn was found to be geometrically very accurate, when ortho-rectified aerial photo was the standard of comparison.

The estimation method, which was used in this study (k-nearest neighbour method applied individually on different auxiliary data sources) was shown to be technically workable and flexible. The estimation of stand variables was possible also in situations, when some auxiliary data sources were not available for certain plots.

New opportunities to further improve the estimation of plot stand variables might be offered by the use of a digital stereoscopic photo-interpretation unit consisting of specially developed hardware and software. Miettinen (1998) has achieved somewhat better estimation results when using this method, in comparison with traditional stereoscopic interpretation of analog aerial photos, particularly regarding stand height. Developing the use of old inventory data might be another possibility. They can be updated by using growth models and by tracing drastic changes with, for example, multi-temporal remote sensing data or they can be used directly as estimates instead of an auxiliary data source. Furthermore the stand data measured in field plots could have more significance in calibrating local sample plot esti-

mates in their immediate neighbourhood. The immediate neighbourhood could, for example, be defined as a polygon produced by automatic segmentation of digital aerial photos.

The essential results by this study are briefly:

1. The overall accuracy of sample plot estimates, that are achieved using auxiliary data, still can not be considered to be entirely satisfactory.
2. Clear improvements in results are obtained by using more than one auxiliary data source.
3. Also weighting of alternative estimates by inverse value of MSE estimate can be recommended.
4. Implementing rules based on knowledge has been shown to have some potentiality. The procedures call for more research and development effort.

Acknowledgements

The authors wish to express their gratitude to Metsämännut Oyj, Metsämiesten Säätiö, Metsäteho Oy, TEKES and UPM-Kymmene Oyj, whose co-operation and support made this study possible. Further we wish to thank Markku Halinen, Eero Lukkarinen and Janne Soimasuo, whose contribution was vital for accomplishing this study.

References

- Cochran, W. 1963. Sampling techniques. John Wiley & Sons, Inc.
- Holmgren, P. & Thuresson, T. 1995. Avdelningsfritt skogsbruk. Fakta Skog Nr 14 1995.
- Kaukokuvat metsien inventoinnissa. 1997. Tutkimushankkeen maastotyöohje. Helsingin yliopiston metsävarojen käytön laitos. [Field manual for research project: Remote sensing in forest inventory. Department of Forest Resource Management, University of Helsinki].
- Kilki, P. & 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: 210–215.
- Miettinen, J. 1998. Digitoitujen ilmakuvienvisuaalisen tulkinnan soveltaminen metsän koelainventoin-

- tiin. [Visual interpretation of digital aerial photographs for sample plot based forest inventory]. Master's thesis. Department of Forest Resource Management, University of Helsinki.
- Nilsson, M. 1997. Estimation of forest variables using satellite image data and airborne lidar. Doctoral thesis. Acta Universitatis Agriculturae Sueciae 17.
- Poso, S. & Waite, M-L. 1996. Sample based forest inventory and monitoring using remote sensing. In: Remote sensing and computer technology for natural resource assessment. University of Joensuu, Faculty of Forestry, Research Notes 48.
- , Wang, G. & Tuominen, S. 1999. Weighting alternative estimates when using multi-source auxiliary data for forest inventory. *Silva Fennica* 33(1): 41–50.
- Ranta, E., Rita, H. & Kouki, J. 1989. *Biometria. Tilastotiedettä ekologeille*. Yliopistopaino, Helsinki.
- Tomppo, E. 1990. Satellite image -based national forest inventory of Finland. *Photogrammetric Journal of Finland* 12(1).
- Varjo, J. 1997. Change detection and controlling forest information using multi-temporal Landsat TM imagery. *Acta Forestalia Fennica* 258.
- Waite, M-L. 1991. Koealat metsätalouden suunnittelussa. [Sample plots in forestry planning]. In: *Elektroniikka metsänarvioinnissa. Metsän kasvu ja mallittaminen*. Metsäntutkimuslaitoksen tiedonantoja 384. p. 44–48.
- Wang, G., Waite, M-L & Poso, S. 1997. SMI user's guide for forest inventory and monitoring. University of Helsinki, Department of Forest Resource Management, Publications 16.

Total of 13 references