

# Generating a Raster Map Presentation of a Forest Resource by Solving a Transportation Problem

Eero Muinonen

---

**Muinonen, E.** 2005. Generating a raster map presentation of a forest resource by solving a transportation problem. *Silva Fennica* 39(4): 585–598.

Necessary tools for raster map generation, for the approach based on the calibration estimator, were developed and implemented. The allocation of the area weight of each pixel to sample plots was formulated as a transportation problem, using a spectral distance measure as a transportation cost, and solved using the transportation simplex algorithm. Pixel level accuracy was calculated for the methods based on the calibration estimator so that the results could be compared with the results of the nearest neighbour estimation, the reference sample plot method (RSP) at pixel level. Local averaging in a  $3 \times 3$  window was performed for each generated raster map as a postprocessing phase to smooth the map. Test plot results were calculated both for the unfiltered raster map and the filtered raster map. RSP produced the smallest RMSE in the pooled test data. Local averaging with a  $3 \times 3$  filter decreased the pixel level error – and the bias – and the differences between the methods are smaller. Without local averaging, the pixel level errors of the methods based on solving the transportation problem were high. Raster map generation using the methods of this study forms an optional part – followed possibly by the classification of the pixel level results – of the whole computation task, when the area weight computation is based on the calibration estimation. For larger areas than in the present study, such as municipalities, the efficiency of the method based on the transportation model must be improved before it is a usable tool, in practice, for raster map generation. For nearest neighbour methods, the area size is not such a problem, because the inventory area is processed pixel by pixel.

**Keywords** forest inventory, calibration estimator, transportation problem, nonparametric estimation

**Author's address** Finnish Forest Research Institute, Joensuu Research Unit, P.O. Box 68, FI-80101 Joensuu, Finland **E-mail** eero.muinonen@metla.fi

**Received** 26 November 2003 **Revised** 28 February 2005 **Accepted** 26 October 2005

**Available at** <http://www.metla.fi/silvafennica/full/sf39/sf394585.pdf>

---

# 1 Introduction

Forest inventory applications using satellite remote sensing data are needed to achieve necessary input data for forest planning calculations for large areas. Tokola (2000) reviewed and categorised the main methods for forest inventory applications as stratification (or classification) and regression analysis. Further, the applications of nearest neighbour regression have been of great interest – the reference sample plot method (RSP) (Kilikki and Päivinen 1987, Kilikki 1988, Muinonen and Tokola 1990, Tokola et al. 1996, Tokola 2000) and the Finnish multisource forest inventory method (*k*-NN) (Tomppo 1996, Katila and Tomppo 2001, Katila and Tomppo 2002, Tomppo et al. 2002). McRoberts et al. (2002), who used nearest neighbour approach as a part of stratified estimation, gave a comprehensive analysis, with precautions, for applications based on the nearest neighbour technique.

The concept of area weight, or sample plot weight, refers to interpreting weight as the amount of similar forest the plot represents in the inventory area (i.e., the region for which the inventory results are calculated). A method, denoted with SAE in this study, for computing the sample plot weights for small area estimation of forest parameters has been presented by Lappi (2001). Small areas, in this context, refer to areas of a size comparable to a municipality, which is usually 200–1000 km<sup>2</sup> in land area, or smaller ones (see also Lappi 2001). This method is based on the calibration estimation, where satellite data form a basis for the weight calibration acting as calibration variables, because the satellite data are known for every pixel. The initial sample plot weights (later also referred to as prior weights) are calibrated to satisfy the known area level totals or mean values, when computed from the satellite data and weights of the sample plot pixels (i.e., pixels each of which cover a field sample plot). This is performed by formulating calibration equations including the weights of the sample plot pixels and the known totals, or mean values of the calibration variables (see Lappi 2001). The known totals of the calibration variables are computed by analysing all the pixels in the inventory area.

When the sum of the calibrated weights equals

the number of pixels, the estimated mean value of a forest parameter can be calculated as a weighted average of the forest variables obtained from the sample plots. Besides the original satellite data, channel transformations, that is, volume estimates for each pixel based on the sample plot pixels, can be incorporated into the calibration as auxiliary calibration variables. A variance estimator, based on a spatial model, is also presented (Lappi 2001).

The SAE method, based on calibration estimation, includes the computation of the area level totals of the calibration variables (satellite data) for the inventory area, designing a set of initial weights for the sample plots and calculating the calibrated weights. In the methods, based on nearest neighbour approach, the weights for the sample plots are the cumulative weights from all the pixel level estimates of the inventory area.

According to Lappi (2001), there is no guarantee in RSP that the chosen nearest neighbours add up to unbiased or statistically optimal estimates for the whole region. In the SAE approach, the total area that the sample plot represents (the weight of the plot) is estimated, first, starting from a set of initial weights. The initial weights cannot be set according to the inclusion probabilities if the plots outside the inventory area are to be used, hence the set of initial weights becomes a user-defined parameter (Lappi 2001). As a conclusion, the RSP-based weighting of the sample plots is one alternative for having the initial weights for SAE.

Both of the methods, RSP and calibration estimation, offer means to utilise the plot weights in forest planning systems. For forest planning, the input data should be accurate and localised. The methods, based on nearest neighbour analysis, have the capability to produce results for a forest variable of interest in a raster map form. This cannot be carried out directly in the case of SAE which offers only a set of sample plot weights as a computational result. Thus, for generating a raster map of the estimates, the origin of the weight, allocated to each sample plot pixel, should be known at a single pixel level (RSP), or, if only the estimated weights are known (SAE), the allocated weights should be distributed to the pixels of the inventory area in a separate process. Lappi (2001) suggests a formulation of a linear programming

transportation problem to find such a weight allocation, where an overall spectral distance between plot pixels and destination pixels is as small as possible. In the case of using the calibration estimation-based approach, this would combine the small-area calibration estimator and the basic principle of the nearest neighbour method in map production.

In the RSP method, the estimation of forest variables is carried out separately for each pixel in the inventory area, and the raster map can be generated as a by-product, if needed. In the SAE-based methods, the map must be generated in a separate analysis, because the calibration is based on area level means or totals. In the raster map generation, each pixel, in the inventory area, receives weight in such a way that an overall distance is minimised in the whole inventory area. Area level accuracy is not changed in a map generation as the sample plot weighting does not change in distributing the weight over the area. However, the accuracy at pixel level can then be estimated if there is located information on the forest variables available. However, due to the nature of the SAE method, the pixel level results are, then, dependent on the inventory area in which it belongs, and this has to be taken into account when comparing the pixel level results of the method, after map generation, with the results of the RSP method.

The objective of this study is to present and test the feasibility of the approach of solving a transportation problem for the raster map generation, as a part of a forest inventory system based on calibration estimation. For comparison, the results are calculated using a method (RSP) based purely on the nearest neighbour approach. The accuracy of results, at pixel level, is calculated to evaluate the raster map. The calibration estimator-based method, together with initial weights from RSP (i.e., a combined method) is also applied. The framework of the study is to present a map production tool for the SAE-based technique (Lappi 2001), and to present a methodology of using the RSP in the prior weight production.

## 2 Material

The satellite data consist of a Landsat 7 ETM image (WRS 187/17; a floated scene) from eastern Finland, acquired on August 2, 1999. The image was rectified to the Finnish National Coordinate system using first order regression models and 43 control points which were located from the base maps of the area. A 25-m pixel size was produced using nearest neighbour resampling in which the root mean square error in the rectification was 0.4086 pixels.

Digital map data were used both in the rectification of the image and also in the selection of the field plots to the analysis. A numerical map data mask, processed in the NFI remote sensing laboratory, included the following categories: forestry land, agricultural areas, water areas, and other non-forestry land (e.g., roads, buildings). For details of the map data see, for instance, Katila et al. (2000). The map data were used to mask out categories other than forestry land. The selection of the sample plots was based on this map data to mimic the true inventory situation.

Field data consist of the sample plots measured in 1999 for the 9th National Forest Inventory in the forest district of Etelä-Savo, which fall in the forestry land category based on the map data and are located in the satellite image window. The field sample is a systematic cluster sample, where a cluster consists of 14 temporary relascope plots, or 10 permanent plots (every 4th cluster). The distance between the clusters is 6 km in both North-South and East-West directions (Valtakunnan metsien... 1999). The resulting data set comprised 2863 sample plots. The main forest variable of interest is the mean volume of the growing stock ( $\text{m}^3/\text{ha}$ ) (Table 1).

For the pixel level checking and evaluation of the techniques, two  $15 \text{ km} \times 10 \text{ km}$  rectangular inventory areas (test areas TPA and TPB) were generated. The sample plot data were used inside a 12-km inclusion zone around TPA and TPB, respectively. The pixel level accuracy of the estimation was computed for each raster map generation method using randomly selected test data sets of the sample plot pixels. The proportion of the sample plot pixels in TPA and TPB selected to the test data was 75%. The number of outside plots (plots in the inclusion zone) of

**Table 1.** Mean volume of the growing stock (m<sup>3</sup>/ha) in the sample plot data: a) 1999 data (test data not included); b) test area TPA; and c) test area TPB. (n: number of plots; TPAtest, TPBtest: test plots only; TPA&zone, TPB&zone: plots in the test area and in the 12-km zone).

Data set	n	Mean	S.D.	Min	Max
a) 1999 data	2760	130.9	108.1	0.0	657.6
b) TPA	16	141.3	156.1	0.0	657.6
TPAtest	49	109.1	112.8	0.0	483.0
TPA&zone	287	118.9	109.8	0.0	657.6
c) TPB	11	87.1	81.4	0.0	293.6
TPBtest	54	124.0	102.2	0.0	381.0
TPB&zone	298	117.4	101.1	0.0	470.4

**Table 2.** Land use class distribution (number of plots) among the field plots belonging to the map stratum forestry land over 1999 data and the test areas. (Land use class Forestry includes forest land, other wooded land, waste land and other forestry land). a) 1999 data (without test data; n=2760); b) test area TPA; and c) test area TPB. (TPAtest, TPBtest: test plots only; TPA&zone, TPB&zone: plots in the test area and in the 12-km zone).

Data set	Forestry	Arable	Built-up	Traffic, etc.	Water	Total
a) 1999 data	2626	38	49	29	18	2760
b) TPA	16	0	0	0	0	16
TPAtest	46	0	1	2	0	49
TPA&zone	275	3	5	2	2	287
c) TPB	11	0	0	0	0	11
TPBtest	51	2	0	0	1	54
TPB&zone	286	5	2	2	3	298

the areas TPA and TPB was 271 and 287, respectively. Thus, the number of plots available in the estimation for TPA is 287, and there are 298 plots available for TPB. The pixels and sample plots, for the analysis, were selected from the category 'forestry land', based on the digital map data. In TPA, there were 191 895 pixels, and TPB included 193 498 pixels.

The relascope sample trees (tally trees), located on the land use classes, forest land or other wooded land, are measured in the sample plot. In this study, the sample plots on the forestry land map stratum with no trees measured, were given a 0.0 m<sup>3</sup>/ha value of mean volume in the analysis. These plots most usually fall into other land use classes in a field inventory, and, thus, no trees are sampled (Table 2).

## 3 Methods

### 3.1 Nearest Neighbour Estimator

The method applied in the estimation of the forest variable, at pixel level, is the reference sample plot method (RSP) (Kilikki and Päivinen 1987, Kilikki 1988, Muinonen and Tokola 1990, Tokola et al. 1996, Tokola 2000). The spectral distance ( $d_{ij}$ ) between pixels  $i$  and  $j$  is calculated using a distance function

$$d_{ij} = \sqrt{\sum_{h=1}^{n_c} (p_h(b_{ih} - b_{jh}))^2} \quad (1)$$

where  $n_c$  = number of bands,  $i, j$  = pixels,  $b_{ih}$  = the

spectral value of a pixel  $i$  at band  $h$ ;  $p_h$  = empirical constant for band  $h$ .

After computing the distance function values from the target pixel  $i$  (i.e., the pixel for which the forest variables are estimated) to each sample plot pixel  $j$ ,  $m$  spectrally nearest sample plot pixels are sought and the weight ( $w_{ij}$ ) for each neighbour pixel  $j = 1, \dots, m$  of the pixel  $i$  is calculated as

$$w_{ij} = 1 / (1 + d_{ij}) \tag{2}$$

After computing the weights  $w_{ij}$ , the estimator of a forest variable  $y$ , for a pixel  $i$ , can be formulated as in Eq. 3. For a pixel  $i$ , it is a weighted average of the field measured forest variable values for the spectrally  $m$  nearest sample plot pixels:

$$\hat{y}_i = \frac{\sum_{j=1}^m w_{ij} y_{ij}}{\sum_{j=1}^m w_{ij}} \tag{3}$$

where  $y_{ij}$  denotes the forest variable value of the  $j$ th nearest neighbour of the target pixel  $i$ .

The empirical coefficients  $p_h$ , in the distance function, describe the relative importance of the variables. They were selected by testing all the combinations of the values  $p_h \in \{0.1, 5.0, 10.0\}$  for the variables in the distance function. The pixel under estimation was left out from the analysis when selecting the  $m$  spectrally nearest neighbours for the pixel. The analysis was carried out using the so-called leave one out method, that is, a forest variable estimate for each sample plot pixel  $i$  in turn was computed using the other sample plot pixels  $j$  as reference.

### 3.2 Calibration Estimation

Calibration estimation (Eqs. 4 and 5), (see Lappi 2001), can be applied in the estimation of the sample plot weights ( $w_k$ , in Eq. 5),

$$\hat{Y}_d = \sum_{k=1}^n d_k y_k \tag{4}$$

$$\hat{Y} = \sum_{k=1}^n w_k y_k \tag{5}$$

In Eqs. 4 and 5,  $y_k$  denotes a forest variable for a sample unit  $k$ , drawn from a finite population, where each unit is associated with a vector  $\mathbf{x}_k$  of spectral variables or auxiliary variables;  $n$  is the number of sample plots in the analysis. In the case of this study, the sample units are pixels associated with  $y$  variables – forest variables. Eq. 4 describes an estimator (denoted using subscript  $d$ ) for forest variable  $y$ , based on the prior weights  $d_k$ . Eq. 5 is the calibration estimator for the forest variable  $y$  (see Lappi 2001). The calibrated weights  $w_k$  are obtained by calibrating the prior weights (see Eq. 4), respecting the calibration equation:

$$\sum_{k=1}^n w_k \mathbf{x}_k = \mathbf{t}_x \tag{6}$$

where  $\mathbf{t}_x$  is the known population total vector of the known  $x$  variables (i.e., satellite image pixel spectral values, or other known auxiliary variables for each pixel). In the SAE application, sample plot pixels, outside the inventory area, are included in the estimation (Lappi 2001). This means that the prior weights – originally based on the inclusion probabilities – had to be defined in another way in this study. They were defined using a geographical inclusion zone, where the prior weight is constant, inside the inventory area, and decreases linearly, in the inclusion zone, as the distance increases. The other method applied, in the present study, is to use the RSP sample plot weights as prior weights in the calibration.

The computation of the area weight estimation was carried out as in the study of Lappi (2001), (see also Deville and Särndal 1992), incorporating the following distance function:

$$G_k(w_k, d_k) = (x - L) \log\left(\frac{x - L}{1 - L}\right) + (U - x) \log\left(\frac{U - x}{U - 1}\right) \tag{7}$$

where  $x = w_k / d_k$ .

Parameters  $L$  and  $U$  define the range of the weights,  $Ld_k < w_k < Ud_k$ . To have nonnegative weights,  $L$  and  $U$  were given values 0 and 6.5, respectively (see Lappi 2001). The nonlinear optimisation problem, needed in the search for the weights, was solved by using the routines

for the Newton method presented by Press et al. (1992).

In the estimations for this study, the calibration was based on the mean value vector (instead of total values) and  $t_x$  was computed using satellite pixel data in the test area under examination. Pixel values were scaled, by dividing them by their mean value in the data, to obtain a more stable convergence (see Lappi 2001).

### 3.3 Transportation Problem

Pixel level accuracy is calculated for the methods based on the calibration estimator to allow these results to be compared with the results of the nearest neighbour estimator (RSP) at pixel level, also taking into consideration that in SAE, the pixel level results depend on the properties of the whole inventory area it belongs to. The prerequisite for the pixel level accuracy estimation is that the weight allocated to each sample plot has to be assigned to the pixels in the inventory area. This means that a raster map generation technique has to be implemented.

The weight allocation was formulated as a transportation problem which, in general, concerns transporting goods or services from supply centres to multiple demand centres in an optimal manner (see Dykstra 1984). The problem can be viewed as a network problem, where supply and demand centres are represented by nodes, and the transportation system joining them is represented by a set of directed links. Each link has a value of any relevant unit of measure, representing, for example, transportation cost or travel time. (Dykstra 1984). When the problem is presented in a tableau form, the rows illustrate the sources (supply centres), and the columns are the destinations (demand centres).

A usual way to present the problem is to formulate it as a linear programming problem (see also Dykstra 1984):

$$\begin{aligned} \text{Minimize } z &= \sum_{i=1}^M \sum_{j=1}^N C_{ij}x_{ij} \\ \text{subject to } & \\ \sum_{i=1}^M x_{ij} &= D_j, \quad j = 1, \dots, N \\ \sum_{j=1}^N x_{ij} &= S_i, \quad i = 1, \dots, M \\ x_{ij} &\geq 0, \quad i = 1, \dots, M; j = 1, \dots, N \end{aligned} \tag{8}$$

where  $M$  is the number of supply centres,  $N$  is the number of demand centres,  $C_{ij}$  is the relevant link distance, and  $x_{ij}$  is the number of units to be transported from the supply centre  $i$  to the demand centre  $j$ . Total cost of transporting ( $z$ ) is minimised. It follows that in feasible solutions for this kind of a problem:

$$\sum_{i=1}^M S_i = \sum_{j=1}^N D_j \tag{9}$$

Thus, in the transportation model, all constraints are equalities, and the sum of supplies and demands must be equal (Dykstra 1984). For solving this kind of problems, an algorithm that takes advantage of the special structure of the transportation model has been developed. The algorithm, used in this study, has been described in detail with an example case by Dykstra (1984), (see also Taha 1992).

To generate the initial solution for solving the transportation problem, feasible methods, such as the least cost method (LCM) and Vogel's approximation method (VAM), exist (see Taha 1992). In the least cost method, as much weight as possible is iteratively allocated to a variable (tableau cell) with the smallest unit cost in the entire tableau. After each allocation, supply and demand are adjusted accordingly and the satisfied row, or column, is crossed out. VAM is a heuristic method and usually provides a better starting solution than the least cost method and generally yields an optimum, or close to an optimum, starting solution (Taha 1992). For this study, VAM serves an interesting method for map generation in the case that it could produce a suitable map presentation even without solving the transportation problem.

The VAM method is an iterative method, based on computing the penalties – of not choosing the cheapest alternative – for each row and column in a transportation tableau (see Taha 1992). The initial solution is generated by choosing iteratively the row, or column, with the largest penalty and, then, allocating as much weight as possible to the variable (cell) having the least cost.

### 3.4 Applying the Methods in Raster Map Generation

Four different combinations of methods, to produce the raster map of mean volume, were tested and they are summarised in Table 3. For each method, the pixel level accuracy is computed for test sample plot pixels where a true value is known. Test sample plot data are used only in the evaluation part of the analysis. Thus, the test sample plot pixels have not been used as demand centres in the transportation problem either.

Local averaging in a 3×3 window was performed for each generated raster map as a post-processing phase to smooth the map. Test plot results were calculated both for the unfiltered raster map and the filtered raster map. Only the pixels belonging to the forestry land stratum, in the digital map data, were used in the local averaging.

The first area weight estimation was computed using the RSP method. An inventory area  $A$  that contains  $n_A$  pixels in the satellite image is assumed. For each of these  $n_A$  pixels,  $i = 1, \dots, n_A$ , a pixel level estimate is computed using Eq. 3; the weight  $W_{kA}$  which a sample plot  $k$  obtains in the area level estimate for area  $A$  is the sum of the

weights it receives in the pixel level estimates in the inventory area

$$W_{kA} = \sum_{i=1}^{n_A} w_{ik} \tag{10}$$

The weights  $W_{kA}$ ,  $k = 1, \dots, n$ , for the area level estimate for area  $A$ , sum up to  $n_A$ , and are the area weights. If inventory area  $A$  includes only a single pixel, the area level estimate is equal to a single pixel level RSP estimate. Thus, the estimates for an area  $A$ , containing several pixels, would be computed as weighted averages of the plot pixel field measurements, similarly as in Eq. 5. Instead of computing RSP area level estimates, the weights, from this nearest neighbour method, were stored and further used as initial weights for one of the SAE-based methods (Table 3).

In both of the methods, C1 and C2, for the area weight estimates by plots, the calibration estimator was applied, (see Lappi 2001, Kangas and Maltamo 2000). The methods differ only in the mode the initial weights are produced (Table 3). In the C1 method, the initial weights  $d_k$  in Eq. 4 were computed using an approach of a 12-km geographical inclusion zone around the inventory area. Inside the inventory area, the prior weight was set to a constant value, and the weight decreases linearly with respect to the geographical distance to the inventory area. In the C2 method, the sums of weights by plots from the RSP estimation by pixels, were used as prior weights. Thus, the third (C2) area weight estimation was computed using the weights  $W_{kA}$ ,  $k = 1, \dots, n$  from the RSP as prior weights  $d_k$  in Eq. 4. A sample plot is used in the estimation if it has been given a positive prior weight, in the pixel level estima-

**Table 3.** Techniques applied in the raster map generation. Local averaging was tested in postprocessing the generated raster map after each method.

Method	Initial area weights	Area weights	Raster map generation
RSP	Nearest neighbour estimation (5 nearest neighbours)	Initial area weights	Pixel by pixel estimation
C1	Geographical weighting	Calibration estimation	LCM & transportation problem
C2	Nearest neighbour estimation (5 nearest neighbours)	Calibration estimation	LCM & transportation problem
C3	Geographical weighting	Calibration estimation	VAM

tion, for the pixels in the inventory area and, at area level, the prior weight  $d_k$  for sample plot  $k$  (in Eq. 4) is, then, the sum of these weights.

In the methods C1 and C2, quite a large geographical search radius – 60 km from each target pixel, where the spectrally nearest neighbours are sought – was used so as not to limit the usage of the sample plot data, in the surrounding zone of the area, in creating the raster map. The raster map generation was performed using the transportation problem model. For the analysis of this study (see Eqs. 8 and 9), the supply centres are the  $M$  pixels of the area and each pixel  $i$  has a supply  $S_i$  equal to 1. The demand centres are the  $N$  sample plot pixels in the data, each plot pixel  $j$  having a demand  $D_j$  equal to the area weight (noted with  $w_k$  for plot  $k$  in Eq. 5) it has received in the area weight estimation. A transportation cost  $C_{ij}$ , between source  $i$  and destination  $j$ , is a distance function value (Eq. 1) which was calculated from the satellite data, the number of neighbours being 1. As a result, total spectral distance is minimised when transporting the area weight of the pixels to the plots up to a given demand. In the RSP estimation, the raster map is generated as a by-product and no separate computation is needed for this part, as the inventory area is processed pixel by pixel, the number of neighbours being 5.

In the C3 method, VAM was applied in the raster map generation. The spectral distance function value, with 1 nearest neighbour, was used as a transportation cost (Table 3).

### 3.5 Pixel Level Accuracy

Pixel level accuracy was analysed using a cross validation technique. When defining the coefficients for the distance function, the mean volume for each sample plot pixel was estimated using the other plot pixels as a reference. The pixel level accuracy of the estimation was evaluated, first, with

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \tag{11}$$

where  $n$ =number of sample plot pixels in the

estimation; and, second, (see Tokola et al. 1996, Tokola 2000), with

$$R^2 = 1 - \frac{RMSE^2}{\text{var}(y)} \tag{12}$$

In  $R^2$ , the original variation, in the test data under analysis, is taken into account, not only the RMSE of the estimation.

The possible bias of the estimation, at pixel level, was evaluated using

$$\text{bias} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n} \tag{13}$$

Relative RMSE and bias were computed by dividing the RMSE and bias by the mean value of the  $y$  variable in the data set. The statistical significance of bias, at pixel level, was tested by computing

$$t_{\text{bias}} = \frac{\text{bias}}{s / \sqrt{n}} \tag{14}$$

where  $s$ =standard deviation of the differences  $(y_i - \hat{y}_i)$ .

In the evaluation phase, using the test sample plot pixels, the measured value and the estimate from the method under analysis, were drawn for the pixels, and the Eqs. 11–13 were applied.

## 4 Results

The coefficients of the distance function, resulting in a minimum RMSE and an insignificant bias ( $t_{\text{bias}} < 1.96$ ) at pixel level, were sought by performing a grid search. The grid search was made separately for the cases by using 1 nearest neighbour – to be applied in the raster map generation using a transportation model – and 5 nearest neighbours – to be applied in the RSP estimation generating the raster map pixel by pixel.

It can be noted that using one nearest neighbour ( $m=1$  in Eq. 3), in the estimation, produces an RMSE value which is greater than the original standard deviation (Table 4). This means that the overall mean could give a better estimate than using the one nearest neighbour does. Thus, if the

**Table 4.** Coefficients for the distance function, produced by the grid search, and the pixel level accuracy of the nearest neighbour estimates for mean volume (m<sup>3</sup>/ha). Data: sample plots of 1999, test data excluded; n=2760; Mean=130.9 m<sup>3</sup>/ha; S.D.=108.1 m<sup>3</sup>/ha; geographical reference zone=60 km; m is the number of nearest neighbours; TM1...TM5 and TM7 are the satellite image channels 1...5 and 7.

m	TM1	TM2	TM3	TM4	TM5	TM7	RMSE (m <sup>3</sup> /ha)	RMSE/mean (%)	Bias (m <sup>3</sup> /ha)	R <sup>2</sup>
1	5	10	5	5	5	10	116.5	88.97	-1.16	-0.16
5	10	10	5	5	10	5	93.6	71.48	0.38	0.25

Negative R<sup>2</sup> implies that the RMSE is greater than the original standard deviation.

**Table 5.** Pixel level estimation results in the pooled TPA and TPB test data for mean volume; a) without local averaging, and, b) with 3x3 local average filter. Bias is computed in the test data. (Mean 116.89 m<sup>3</sup>/ha; S.D. 107.60 m<sup>3</sup>/ha; n=103).

a)

Method	RMSE (m <sup>3</sup> /ha)	RMSE/mean (%)	Bias (m <sup>3</sup> /ha)	R <sup>2</sup>
RSP	86.0	73.59	-1.9	0.36
C1	108.0	92.37	-1.8	-0.01
C2	109.4	93.60	-2.3	-0.03
C3	94.6	80.95	4.4	0.23

Negative R<sup>2</sup> implies that the RMSE is greater than the original standard deviation

b)

Method	RMSE (m <sup>3</sup> /ha)	RMSE/mean (%)	Bias (m <sup>3</sup> /ha)	R <sup>2</sup>
RSP	79.2	67.74	-1.7	0.46
C1	84.8	72.58	-0.2	0.38
C2	84.5	72.30	0.1	0.38
C3	79.7	68.19	3.7	0.45

number of nearest neighbours can be chosen for the estimation, as in the RSP, a value greater than 1 would be selected. In the raster map generation using RSP, 5 nearest neighbours were used to produce a more proper basis for the comparisons of the usability of the methods.

Pooled pixel level results, in the test areas TPA and TPB, from the raster map generation techniques which were applied, show that the problem of using 1 nearest neighbour in the allocation of

weights causes a poor pixel level quality (Table 5). RSP estimation produced the smallest RMSE in the pooled test data (Table 5). Local averaging with a 3x3 filter decreased the pixel level error – and the bias – in all cases, and the differences between methods are smaller (Table 5). Bias was not found to be significant (*t*<sub>bias</sub> < 1.96).

In the test area TPA, the test plot results show a larger relative pixel level error than in TPB, for the tested methods. Local averaging caused a decrease of RMSE in both areas, and all tested methods. For RSP, local averaging decreased bias in both areas, but for the other techniques, the effect of the local averaging on bias is not so clear in TPA and TPB (Fig. 1). After local averaging, C2 has the smallest bias in the pooled data (Table 5) – as a result of opposite bias values in TPA and TPB test data, but C1 is more stable (Fig. 1). The computationally simple method, C3, has largest bias, after local averaging, in the pooled data, resulting from the bias in TPB test area (Table 5, Fig. 1).

As an example of the raster map presentation of the estimated mean volume, generated by using the C1 and RSP methods, Fig. 2 illustrates the classified output of the applied methods in a selected 1.5 km x 1.5 km sample in the TPA area. The effect of local averaging can be seen, as expected, as a smoother output raster map appearance.

## 5 Discussion

In this study, necessary tools, concerning raster map generation, for the approach based on the

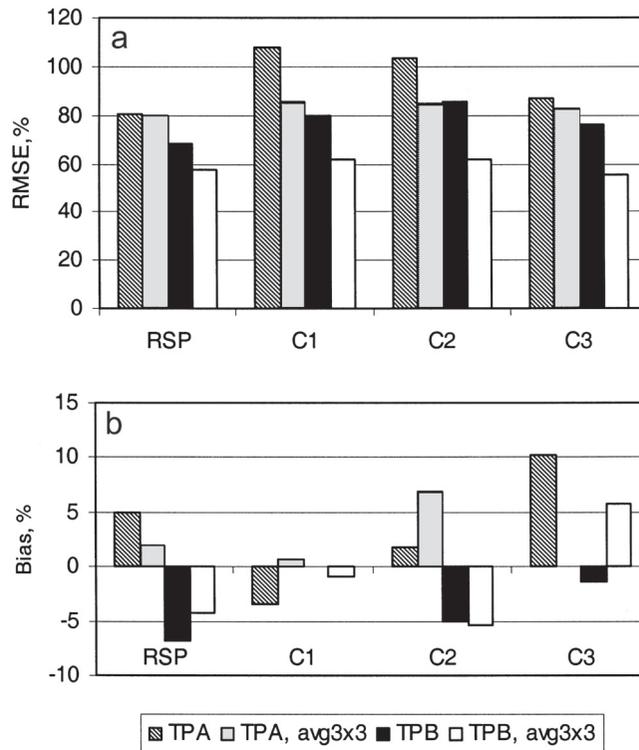


Fig. 1. RMSE (%), a, and bias (%), b, in test data of areas TPA and TPB.

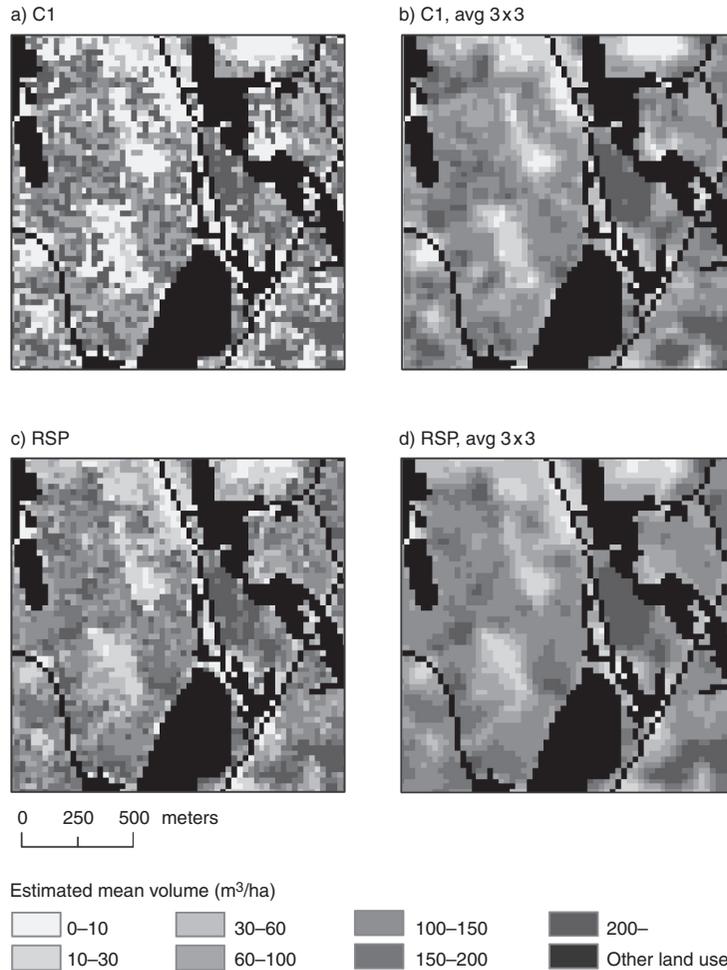
calibration estimator have been implemented. The allocation of the area weight of each pixel to sample plots was formulated as a transportation problem, using a spectral distance measure as a transportation cost, and solved using the transportation simplex algorithm (Dijkstra 1984, Taha 1992).

The raster map generation is necessary for presenting results in a map form in the case of using the method presented by Lappi (2001) – that is, a method based on the calibration estimator – which does not directly produce a map. Another simple way to deal with the sub-area level reporting might be to compute the area weights for each sub-area separately, without a raster map generation, but it may suffer from the possibility of having quite a small number of sample plots available, even if including the plots from a larger geographical inclusion zone. It is clear, however, that the small number of neighbours is problem-

atic in the sense of having accurate pixel level results (Tokola et al. 1996, Tomppo et al. 1998, Katila and Tomppo 2001).

In the raster map generation, the transportation cost was expressed by means of a spectral distance measure. To find a most suitable demand centre (sample plot) for each pixel, the spectral distance was minimised. This leads to an approach similar to a nearest neighbour estimation with the number of neighbours being 1 (C1, C2, C3) or 5 (RSP) and with no constraints for the weights allocated to each plot. Here, the weight to be allocated to a sample plot is constrained to have a maximum value which equals the area weight of the sample plot. This plot level area weighting is obtained from the calibration estimation procedure, where the calibration equations have to be fulfilled.

The selection of the channel weights, in the distance functions, was made by using a grid search method to have a minimum RMSE and an insignif-



**Fig. 2.** Raster map presentations of estimated mean volume (m<sup>3</sup>/ha) classified into 7 classes, in a selected area (1.5 km × 1.5 km). a) C1 estimation, b) C1 estimation and local averaging, c) RSP estimation, d) RSP estimation and local averaging.

icant bias for the mean volume estimation. Other methods, for channel weight search, would be the use of multiobjective optimisation (Haara 2002), or canonical analysis (Moeur and Stage 1995). In these methods, the channel weights could be calculated by taking several forest variables into account. These forest variables could include both age and volume, for example. Another approach, for taking into account several forest variables at area level – and simultaneously also at pixel level – could be incorporating pixel level forest vari-

able estimates into the estimation as calibration variables, as suggested by Lappi (2001).

In the selection of the channel weights, the whole 1999 data were used. However, the models calibrated in the large area, produced bias in the smaller test areas. A larger number of sample plots might be needed, in the smaller test areas, than used in this study; as in the study of Lappi (2001), the zone size could be computed to have a fixed number of sample plots. In this study, also using a part of the plots in the test areas as test

plots decreased the number of the plots available in the estimation.

Channel weights, for the distance function used in the transportation problem, produced large RMSE also in the whole 1999-year data set. The number of nearest neighbours, for this distance function, had to be this small, because the weight of a pixel is transported to one sample plot, in most of the cases. This is the price, in raster map generation, of using the SAE approach – that is, calibration estimation, as used in this study – in computing the area weights. The precaution, reported by McRoberts et al. (2002), that a small number of nearest neighbours may result in an RMSE value greater than the standard deviation of the observations, came true in this study. It is known that with a smaller number of nearest neighbours RMSE increases (Tokola et al. 1996, McRoberts et al. 2002). Increasing the number of neighbours, to a greater value than 5, was not carried out, in order to preserve the estimation variance. When increasing the size of the neighbourhood, the estimation variance decreases and the bias increases (Altman 1992). Hence, the pixel level error of the methods C1–C3, in the study, without averaging was considerable, as could be expected.

Lappi (2001) has suggested local averaging to improve the raster map generation by smoothing the map. This was performed in the present study to see whether this kind of postprocessing would be necessary to improve the pixel level accuracy of the results in the generated raster map. The filtering appeared to be necessary, especially for the methods based on the calibration estimation (C1 and C2), giving a better pixel level accuracy and a smaller bias to the results calculated over the test plots.

The results of RSP, in the test areas TPA and TPB together, show a bias of  $-1.7 \text{ m}^3/\text{ha}$  ( $-1.45\%$ ) also after local averaging (Table 5). The results from RSP, obtained for test plots in TPA and TPB, are biased (ca. 5% in TPA and  $-7\%$  in TPB), although the level of bias in the grid search of the channel weights is not so high (Table 4). The RMSE level of RSP is a result of similar nature as reported when using the Euclidean distance function and RSP (Muinonen and Tokola 1990, Tokola et al. 1996). The estimation may suffer from too few observations in the spectral space,

and the number of sample plots, in the analyses for test areas, was too small, in the study, to have a satisfactory pixel level accuracy. As a result, the calibration of the RSP technique was not good in the test areas, that is, at local level. Further analyses concerning, for instance, the zone size, weights of the covariates, number of neighbours, are not in the scope of this study. A larger inclusion zone could have been a usable choice, especially for RSP estimation.

The transportation simplex algorithm appeared to work as planned in the problem solving. For the purpose, a computer programme was developed due to the size of the problem being large. The problem solving took approximately 5 days with a computer equipped with a 2.4 GHz processor and 992 MB of RAM. Thus, it is clear that for larger areas, there is a need to compress the size of the problem and improve the problem solving, for example, by allowing only a limited set of possible target plots for each pixel. The large number of rows (pixels) is a problem, which also needs to be solved so that the approach could be used in larger scale and in practice. For these reasons, the nonparametric approach has a straightforward capability to produce the raster map output, because the inventory area is processed pixel by pixel. Furthermore, it has to be taken into account that the pixel level estimates of the methods C1–C3 are dependent on the properties of the inventory area. If a different area is selected over a pixel, this pixel can obtain a different estimate, as the area level transportation problem has changed.

In C1 and C2, calibration estimation, after having necessary initial weights, and the transportation model were applied. Geographical initial weighting was the basis of C1, and the initial weighting from the nearest neighbour analysis was the basis of the C2 method. The differences between C1 and C2 are quite small, but local averaging was necessary to lower the pixel level RMSE. (Table 5). The bias level, in the test areas TPA and TPB, is lower in C1 than in C2. Actually, the level of bias calculated over test plots is lowest in C1 than in any other methods tested. For C2, the pooled test results show a small bias, but C2 did not perform well in TPA and TPB test areas, and the average filter increased the bias of C2 (Fig. 1).

The usability of VAM in the map generation can also be discussed by comparing the pixel level results from the VAM solution and transportation model. Satisfactory results from using VAM, especially for larger applications, would offer a way to avoid the computation in solving the transportation problem. For this reason, a raster map was generated and the estimates for the test sample plot pixels were calculated by using the VAM approach, too. It was computationally easier, as the time demand, in the test areas, was only approximately 2 hours for each test area. However, the C3 method had largest bias calculated over test plots. The results, however, indicate that it is still an interesting alternative, when the inventory area increases. It has to be noted that the results, concerning pixel level accuracy and bias, can be the outcome of the non-optimal calibration of the nonparametric estimation. More detailed search of the distance function parameters and variables, might improve the results of each method.

To perform calculations over a larger inventory area, as a municipality, by using the approach of solving the transportation problem, the computational efficiency of the method has to be improved before it is an available tool for raster map generation. This kind of work was not possible in the present research project. For nearest neighbour methods, the area size is not such a problem, as the inventory area is processed pixel by pixel. As a result, the raster map generation approach applied in the study, can be seen as an optional part – followed possibly by the classification of the pixel level results – of the whole computation task, especially when it is based on the calibration estimation, also taking into consideration the properties of the method, and capabilities of satellite data in the estimation of forest variables.

## Acknowledgements

This research work is part of the project “Methods for statistical integration of satellite imagery and GIS data for spatial analysis in forest planning” funded by the Ministry of Agriculture and Forestry of Finland, and coordinated by the Finnish Forest Research Institute, Joensuu Research Unit.

The cooperation with the National Forest Inventory research programme enabled the use of the study material and map data.

## References

- Altman, N.S. 1992. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician* 46(3): 175–185.
- Deville, J-C. & Särndal, C-E. 1992. Calibration estimators in survey sampling. *Journal of the American Statistical Association* 87(418): 376–382.
- Dijkstra, D.P. 1984. *Mathematical programming for natural resource management*. McGraw-Hill Series in Forest Resources. McGraw-Hill, New York. 318 p. ISBN 0-07-018552-2.
- Haara, A. 2002. Kasvuennusteiden luotettavuuden selvittäminen knn-menetelmällä ja monitavoiteoptimoinnilla. *Metsätieteen aikakauskirja* 3/2002: 391–406. (In Finnish).
- Kangas, A. & Maltamo, M. 2000. Calibrating predicted diameter distribution with additional information. *Forest Science* 46(3): 390–396.
- Katila, M. & Tomppo, E. 2001. Selecting estimation parameters for the Finnish multisource National Forest Inventory. *Remote Sensing of Environment* 76: 16–32.
- & Tomppo, E. 2002. Stratification by ancillary data in multisource forest inventories employing k-nearest neighbour estimation. *Canadian Journal of Forest Research* 32: 1548–1561.
- , Heikkinen, J. & Tomppo, E. 2000. Calibration of small-area estimates for map errors in multisource forest inventory. *Canadian Journal of Forest Research* 30: 1329–1339.
- Kilki, P. 1988. Satelliittikuvat valtakunnan metsien inventoinnissa. In: Häme, T., Ihalainen, A. & Kaninen, M. (toim.). *Kaukokartoitus metsätaloudessa, Seminaariesitelmät 7.6.1988 Taksattoriklubi*. Metsäntutkimuslaitos, metsänarvioimisen tutkimusosasto, metsäinventoinnin tutkimussuunta, Tiedonantoja 316: 17–20. ISBN 951-40-1026-4. (In Finnish).
- & Päivinen, R. 1987. Reference sample plots to combine field measurements and satellite data in forest inventory. In: *Remote sensing-aided forest inventory*. Proceedings of seminars organised by SNS and Taksattoriklubi, Hyytiälä, Finland, Dec.

- 10–12, 1986. University of Helsinki, Department of Forest Mensuration and Management, Research Notes 19. p. 209–215. ISBN 951-45-4207-X.
- Lappi, J. 2001. Forest inventory of small areas combining the calibration estimator and a spatial model. *Canadian Journal of Forest Research* 31: 1551–1560.
- McRoberts, R.E., Nelson, M.D. & Wendt, D.G. 2002. Stratified estimation of forest area using satellite imagery, inventory data, and the *k*-Nearest Neighbors technique. *Remote Sensing of Environment* 82: 457–468.
- Moeur, M. & Stage, A.R. 1995. Most similar neighbor: an improved sampling inference procedure for natural resource planning. *Forest Science* 41(2): 337–359.
- Muinonen, E. & Tokola, T. 1990. An application of remote sensing for communal forest inventory. In: *The usability of remote sensing for forest inventory and planning. Proceedings from SNS/IUFRO workshop in Umeå, 26–28 February 1990.* Remote Sensing Laboratory, Swedish University of Agricultural Sciences, Report 4. p. 35–42. ISBN 91-576-4208-7.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P. 1992. *Numerical recipes in C. The art of scientific computing.* Second edition. Cambridge University Press. 994 p. ISBN 0 521 43108 5.
- Taha, H.A. 1992. *Operations research: an introduction.* Fifth edition. MacMillan Publishing Company. New York. 822 p. ISBN 0-02-418975-8.
- Tokola, T. 2000. The influence of field sample data location on growing stock volume estimation in Landsat TM-based forest inventory in eastern Finland. *Remote Sensing of Environment* 74: 422–431.
- , Pitkänen, J., Partinen, S. & Muinonen, E. 1996. Point accuracy of a non-parametric method in estimation of forest characteristics with different satellite materials. *International Journal of Remote Sensing* 17(12): 2333–2351.
- Tomppo, E. 1996. Multisource national forest inventory of Finland. In: Päivinen, R., Vanclay, J., Miina, S. (eds.). *New thrusts in forest inventory. Proceedings of IUFRO XX World Congress, 6–12 Aug. 1995, Tampere, Finland.* EFI Proceedings 7. European Forest Institute, Joensuu, Finland. p. 27–41. ISBN 952-9844-15-8.
- , Katila, M., Moilanen, J., Mäkelä, H. & Peräsaari, J. 1998. *Kunnittaiset metsävaratiedot 1990–94. Metsätieteen aikakauskirja – Folia Forestalia* 4B/1998: 619–839. (In Finnish).
- , Nilsson, M., Rosengren, M., Aalto, P. & Kennedy, P. 2002. Simultaneous use of Landsat-TM and IRS-1C WiFS data in estimating large area tree stem volume and aboveground biomass. *Remote Sensing of Environment* 82: 156–171.
- Valtakunnan metsien 9. inventointi (VMI9). *Maastotyön ohjeet 1999.* Häme-Uusimaa, Pirkanmaa ja Etelä-Savo. Finnish Forest Research Institute, Helsinki. Handout. 143 p. (In Finnish).

*Total of 22 references*