The Finnish Multisource National Forest Inventory: Small-Area Estimation and Map Production

Erkki Tomppo[1]

Abstract.—A driving force motivating development of the multisource national forest inventory (MS-NFI) in connection with the Finnish national forest inventory (NFI) was the desire to obtain forest resource information for smaller areas than is possible using field data only without significantly increasing the cost of the inventory. A basic requirement for the method was that it provide applicable information for forestry decision making; e.g., volume estimates—possibly by subclasses such as tree species—timber assortments, and stand-age classes. In an optimal case, the method had to provide all the same estimates for small areas as the field data-based method provides at the national and subnational level. A nonparametric k-NN estimation method fulfills at least part of these requirements. It is simple to apply in its basic form, and the final estimation method is similar to the method that uses field data only. The input data for the Finnish MS-NFI are NFI field data, satellite images, and digital map data of different types, e.g., basic map data and soil data, as well as a digital elevation model. The first operational results were computed in 1990. The method has been modified continuously and new features added. The k-NN method has several advantages but also limitations; e.g., the field plot data should cover the variation of the field variables in the target area. Application of the k-NN estimation method also presumes the selection of estimation parameters. Particularly, predictions of volumes by tree species may be biased if the area of interest is large and covers several different vegetation zones with different tree species compositions. The biases can be reduced if the set of potential nearest neighbors are restricted or the selection of the neighbors directed. A challenging task related to k-NN method is development of an analytic error estimation method for a target area of an arbitrary size. Recent work with the Forest Inventory and Analysis Program of the Forest Service has shown promise in this task. Progress and problems related to the development and application of the k-NN method are discussed, as well as demonstration of the results.

Introduction

The development of the Finnish multisource national forest inventory (MS-NFI) began in 1989; the first operational results were obtained in 1990 (Tomppo 1990, 1991, 1996, 2006). The driving force behind the development of a multisource method was the need to obtain forest resource information for smaller areas than would be possible with field data in an inexpensive manner. Furthermore, new natural resource satellite images provided new possibilities for increasing the efficiency of the inventories at relatively small additional costs.

A basic requirement of the method was that it should be able to provide information applicable to forestry decisionmaking. Thus, methods that are often used in satellite-image-aided approaches but that produce only maps of forest types or land use classes were not considered satisfactory. Methods were sought that would be able to provide area and volume estimates, possibly by subclasses, such as tree species, timber assortments, and stand-age classes. In the optimal case, the method had to be able to provide all the same estimates for small areas as the field-data-based method provides at national and regional levels. The number of variables measured in the field is usually high, typically ranging from 100 to 400. Estimates for additional variables are calculated from these measured ones.

[1] Professor of Forest Inventory, Finnish Forest Research Institute, Vantaa Research Centre, Unioninkatu 40, FIN-00170 Helsinki Finland.
E-mail: erkki.tomppo@metla.fi.

Since the first implementation of the method, it has been modified continuously and new features added (Katila *et al.* 2000; Katila and Tomppo 2001, 2002). The core of the current method is presented in Tomppo and Halme (2004). Any digital land use map or land cover data can be used to improve the accuracy of the predictions (Tomppo 1991, 1996). The list of references of the *k*-Nearest Neighbor (*k*-NN) applications and tests in forest inventories is given in Tomppo (2006). Another nationwide application for Sweden is given in Reese *et al.* (2003).

Application of the *k*-NN estimation method presumes the selection of estimation parameters for each satellite image and for the other data used with the image (Katila and Tomppo 2001). Operational application of the method has also shown that the predictions may be biased if the area of interest is large and covers several vegetation zones with different tree species compositions. Varying imaging conditions within the area of a satellite image can also alter the covariance structure between the field data and the image data. The biases will be reduced if the set of potential nearest neighbors can be restricted to the areas corresponding to the vegetation structure and imaging conditions of the pixel in question. In the first operational applications of the MS-FNFI, a subset of field plots was selected for potential nearest neighbors in the image space for each pixel, usually field plots within a certain geographical horizontal and vertical distance from the pixel in question (Katila and Tomppo 2001).

Tomppo and Halme (2004) presented another method for guiding the selection of field plots that has been in operational use since early 2000. This method uses additional elements in the distance metric vector to guide the selection of nearest neighbors. The elements are variables describing large-area variations in forest characteristics, e.g., mean volumes by tree species, and are map-form predictions of those variables. A relevant, practical variation scale for these variables and their predictions would range between 40 and 60 km. Variation on this scale can be computed from field data only, e.g., from field data acquired in the current or preceding inventory of the area.

The method also uses band transformations in addition to the original image bands because band ratios are assumed to improve the identification of tree species. An optimization method based on a genetic algorithm was developed to find the weight vector, a method that considerably reduces the errors both at the pixel level and in areas of different sizes. The method is called the i*k*-NN method (improved *k*-NN method) in Tomppo and Halme (2004).

One of the open problems related to the *k*-NN method is that until recently, analytical methods for estimating the standard error of any estimate for an area of an arbitrary size have been lacking. However, current progress with a model-based approach has shown promise for error estimation for the *k*-NN method (Kim and Tomppo 2006, McRoberts *et al*. 2007).

Input Data Sets

Field Data

The basic computation unit in image processing is a picture element called a pixel. The pixel size used with Landsat Thematic Mapper (TM) images, for example, is 25 by 25 m. Therefore, it is more convenient to work with volumes per unit area than with volumes of tallied trees. Volumes per ha are estimated for each sample plot by tree species and by timber assortment classes based on the tally tree volumes.

Satellite Images

Images from the Landsat 5 TM or Landsat 7 Enhanced Thematic Mapper Plus (ETM+) sensors have been the most suitable for operational application by virtue of the fairly large coverage area of each image combined with moderate spatial and spectral resolution. These images have been given priority when choosing satellite images to cover an area. If these images are not available, e.g., due to clouds, either Spot 2-4 XS HRV images or IRS-1 C LISS images have been used.

The land area of Finland is 30.4473 million ha, and the total area together with lakes and rivers is 33.8145 million ha. This area was covered by 36 Landsat 5 TM images and 2 Spot 2 XS HRV images in the eighth NFI (NFI8) and its updating in of southern Finland (1990–94), and by 40 Landsat 5 TM or Landsat 7 ETM+ images and 4 IRS-1 C LISS images in the ninth NFI (NFI9) (1996–03).

Areas corresponding to the cloud-free parts of satellite images are used in operational applications. Forests under clouds and in cloud shadows are assumed to be similar on the average to those on the cloud-free part of the same area unit (e.g., municipality). All images are rectified to the national coordinate system using the nearest neighbor resampling method with a pixel size of 25 by 25 m.

Digital Map Data

Digital map data are used to reduce the errors in the estimates. The errors in both the area and total volume estimates can be reduced significantly by the multisource method if the distinguishing forest land from nonforest land can be supported by digital map information in addition to satellite images. The effect of possible map errors on the estimates can be reduced by two alternative statistical methods (Katila *et al.* 2000, Katila and Tomppo 2002). The first is a calibration method using a confusion matrix derived from the land use class distributions on the basis of field plot data and map data, and the second uses stratification of the field plots on the basis of map data. The map information is used to separate forestry land from other land use classes, such as arable land, built-up areas, roads, urban areas, and single houses. In addition, a map is used to stratify the forest land area and corresponding field plots into a mineral soil stratum and a peatland soil stratum (spruce mires, pine mires, open bogs, and fens).

A digital elevation model is used in two ways: for stratification on the basis of elevation data and for correcting the spectral values by reference to the angle between solar illumination and the terrain normal. The latter method is described in detail by Tomppo (1992).

The basic computation unit in the MS-NFI is the municipality. The number of municipalities in the entire country is approximately 500 and their land areas range from around 1,000 ha to some hundreds of thousands of hectares. Digital municipality boundaries are used to delineate the units (Tomppo 1996).

Large-Area Forest Resource Data

The large scale variation of forest variables used in the i*k*-NN method is presented in the form of maps derived from field data using spatial interpolation (Tomppo and Halme 2004). The number of field plots on land in the entire country in NFI9 was 81,249, including 67,264 on forestry land, 62,266 on combined forest land and poorly productive forest land, and 57,457 on forest land alone. All the plots on forest land and poorly productive forest land were used for the large-area maps. The variables were selected in such a way that their values indicate the areas in which the covariance structure between field variables and image variables would be approximately constant. Volumes by tree species (m^3/ha) on forest land and poorly productive forest land were therefore selected as variables.

i*k*-NN Estimation

The Principles of the Method

A nonparametric *k*-NN estimation was used for the MS-NFI calculations during NFI8 and at the beginning of NFI9. The improved method, i*k*-NN, has been applied since 2000 and is described in Tomppo and Halme (2004) and Tomppo (2006). We recall that each field plot has a certain area representativeness, a plot weight, sometimes called a plot expansion factor when forest inventory estimates are calculated from pure field data. This plot weight can be the total land area divided by the number of field plots on land if either systematic or systematic cluster sampling is used (Tomppo 2006). In the MS-NFI, new plot weights (not equal for each plot) are calculated for each plot on an areal unit, e.g., municipality (Tomppo 1996). The weights are calculated for each field plot $i \in F$, where F is the set of field plots on forest land. These plot weights are sums of satellite image pixel weights over the forest land mask pixels.

The pixel weights are in turn calculated by a nonparametric *k*-NN estimation method that uses the distance metric *d*, defined in the feature space of the satellite image data (Tomppo 1991, 1996, 2006). The *k* nearest field plot pixels (in terms of *d*)—i.e., pixels that cover the center of a field plot $i \in F$—are sought for each pixel *p* under the forest land mask of the cloud-free satellite image area. A maximum distance was usually set in the basic *k*-NN method in both a horizontal and a vertical direction to avoid selecting the nearest plots (spectrally similar plots) from a region in which the response of the image variables to the field variables is not similar to that of the pixel under consideration. In the improved i*k*-NN method, only vertical maximum distance is applied. The use of horizontal distance is replaced by the use of large-scale variation of forest variables in equation (2) (Tomppo and Halme 2004, Tomppo 2006). All elements of the distance metric are finally weighted by means of optimization based on a genetic algorithm.

Denote the nearest feasible field plots by $i_1(p),...,i_k(p)$. The weight $w_{i,p}$ of field plot *i* on pixel *p* is defined as

$$w_{i,p} = \begin{cases} \dfrac{1}{d^t_{p_i,p}} \Big/ \displaystyle\sum_{j\varepsilon(i_1(p),...,i_k(p))} \dfrac{1}{d^t_{p_j p}} & \text{if and only if } i\varepsilon\{i_1(p),...,i_k(p)\} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

The power *t* is a real number, usually $t \in (0,2]$. The distance metric *d* in i*k*-NN method is

$$d^2_{p_j,p} = \sum_{l=1}^{n_f} \omega_{l,f}{}^2 (f_{l,p_j} - f_{l,p})^2 + \sum_{l=1}^{n_g} \omega_{l,g}{}^2 (g_{l,p_j} - g_{l,p})^2 \quad (2)$$

where:

f_{l,p_j} = the l^{th} normalised image variable $f_{l,p_j} = f^0{}_{l,p_j} / \cos^r(\alpha)$.

f_{l,p_j} = the original intensity of the spectral band l.

α = the angle between the terrain normal and the solar illumination.

r = the applied power due to non-Lambertian surface.

n_f = the number of spectral features.

$g_{l,p}$ = the large-area prediction for the l^{th} forest variable.

n_g = the number of coarse-scale forest variables.

ω_f and ω_g = the weight vectors for the image features and coarse-scale forest variables, respectively. A pixel size of 1 by 1 km is used in the variables $g_{l,p}$ (Tomppo and Halme 2004, Tomppo 2006).

The values for the elements of the weight vector to be estimated are derived from an optimization using a genetic algorithm, as given below. The first phase of i*k*-NN is to run the optimization algorithm, possibly by strata in the applications, e.g., for the mineral soil stratum and mire and bog stratum separately. The procedure then returns to the basic *k*-NN estimation.

To estimate forest parameters for areal units, the field plot weights for the pixels, $w_{i,p}$, are added for the areal units (e.g., municipalities) in an image analysis process extending over the pixels belonging to each unit. The weight of plot *i* in areal unit *u* is denoted by

$$c_{i,u} = \sum_{p\in u} w_{i,p}. \quad (3)$$

Reduced weight sums $c^r_{i,u}$ are obtained from equation (3) if clouds or their shadows cover part of the areal unit *u*. The real weight sum for plot *i* is estimated by means of the formula

$$c_{i,u} = c^r{}_{i,u} \frac{\hat{A}_{s,u}}{\hat{A}^r_{s,u}}, \quad (4)$$

where $\hat{A}_{s,u}$ is the estimated area of forest land in unit *u,* and $\hat{A}^r_{s,u}$ the estimated area of forest land in unit *u* not covered by the cloud mask. The areas can be taken from digital maps or estimated by means of field plots. It is thus assumed that the forestry land covered by clouds in areal unit *u* is on average similar to the rest of the forest land in that unit with respect to the forest variables (Tomppo and Halme 2004).

Equations (3) and (4) are calculated separately for the mineral soil stratum and peatland stratum within the forest land, and also for other land use classes such as arable land, built-up land, roads and water bodies if a stratification-based map correction method is used (Katila and Tomppo 2002). Alternatively, a statistical calibration and confusion matrix can be used to reduce the effect of map errors on the estimates (Katila *et al.* 2000).

After the final field plot weights on the areal units have been calculated, ratio estimation is used to obtain the estimates (Cochran 1977). In this sense, the estimation procedure is similar to that using field plot data only. Volume estimates, for example, for areal unit *u* and reference unit *s* are calculated in the following way. Mean volumes are estimated by the formula

$$v = \frac{\sum_{i\in I_s} c_{i,u} v_{i,t}}{\sum_{i\in I_s} c_{i,u}}, \quad (5)$$

where $v_{i,t}$ is the estimated volume per hectare of timber assortment (log product) *t* on plot *i* and I_s the set of field plots belonging to stratum *s*. The corresponding total volumes are obtained by replacing the denominator in equation (5) with 1. The forest variable estimators for areal unit *u* thus utilise information from outside unit *u*.

Examples of estimates obtained with MS-NFI are given in table 1. These area and volume estimates are based on the NFI8 inventory field data and satellite images from 1992 for the Kainuu Forestry Centre District (Tomppo *et al.* 1998). Totals for the entire forestry center district are given in two ways, one based on the MS-NFI and the other based on the NFI only. The standard errors for the forestry center totals are based on NFI (e.g., Heikkinen 2006). In addition to table 1, the following other tables were given in MS-NFI8 for all the municipalities in Finland: areas of mineral soil and peatland soils on forest land; poorly productive forest land and unproductive land separately; tree species dominance on forest land and poorly productive forest land, separately; areas of age classes on forest land; mean volumes (m^3/ha) by age classes on forest land; areas of development classes on forest land; mean volumes (m^3/ha) by development classes on forest land; mean and total volumes by tree species; timber assortment classes on forest land and on forest land and poorly productive forest land combined; and some relative distributions for the area and volume estimates.

Predictions of certain (optional) forest variables are distributed in the form of a digital map during the procedure; e.g., the land use classes outside forestry land are transferred to mapform predictions directly from the digital map file. Within forest land, the variables are predicted from the weighted averages of the *k* nearest neighbors (see Tomppo 1991, 1996).

A pixel-level prediction $\hat{m}_p$ of variable *M* for pixel *p* is defined as

$$\hat{m}_p = \sum_{i \in F} w_{i,p} m_i , \qquad (6)$$

where m_i is the value of the variable *M* on plot *i*.

The mode or median value is used instead of the weighted average for categorial variables, i.e., land use class, site fertility class, stand age, mean diameter of stand, mean height of stand, volumes by tree species (e.g., pine, spruce, birch, other broadleaved trees), and by timber assortment class. The total number of maps is thus more than 20.

Optimizing the Variable Weights

The overall aim of the i*k*-NN method is to minimize the errors attached to predictions based on the MS-NFI, both at the pixel level and particularly at higher areal levels (from several tens of thousands of hectares up to several millions of hectares).

Two modifications of the *k*-NN estimation method were introduced:

1. The use of supplementary ancillary variables in addition to spectral data for selecting neighbors.
2. The use of optimal weights for both the image features and the ancillary information.

A vector consisting of these elements is called a vector of explanatory variable weights and denoted by ω. A optimization method based on a genetic algorithm was developed for using ancillary data and finding the optimal explanatory variable weights.

Table 1.—*Mean and total volume of growing stock on forest land and on poorly productive forest land.*

	Forest land			Poorly productive forest land		
	ha	m^3/ha	1000 m^3	ha	m^3/ha	1000 m^3
Hyrynsalmi	113,370	67.4	7,638	14,435	12.0	173
Kajaani	92,850	68.8	6,386	7,798	11.1	87
Kuhmo	388,155	74.9	29,083	49,706	11.4	567
Paltamo	73,713	83.2	6,133	6,429	12.8	82
Puolanko	191,331	69.7	13,341	30,259	13.2	399
Ristijärvi	69,148	68.7	4,748	6613	11.1	73
Sotkamo	222,928	75.7	16,870	16,846	11.5	194
Suomussalmi	389,616	66.1	25,747	65,185	10.4	678
Vaala	88,493	64.0	5,661	18,862	13.3	251
Vuolijoki	52,272	74.6	3,900	8,344	10.2	85
Total, MS-NFI	1,681,876	71.1	119,507	224,477	11.5	2,589
Total, NFI	1,659,701	70.8	117,000	222,675	12.6	2,800
Standard error of NFI	13,895	1.4	2,494	8,969	0.7	198

Volumes by tree species were selected as the variables, because tests with age class distributions, for example, and earlier experiments had shown that the errors in the predictions for other variables are reduced when those attached to volumes by tree species are minimized (Tomppo *et al.* 1998).

Optimization was conducted solely at the pixel level with the expectation and later checking that larger area errors would decrease once the weights were optimized.

A weighted sum of pixel-level biases and root mean square errors (RMSEs) of the predictions was selected as the objective function. The weights are called fitness function weights and denoted by γ (7). The variables used were (1) total volume, (2) volume of pine, (3) volume of spruce, (4) volume of birch, and (5) volume of other broadleaved tree species. These 10 variables have also been used in operational applications of the method. The fitness (objective) function to be minimized with respect to ω is:

$$f(\omega,\gamma,\hat{\delta},\hat{\bar{e}}) = \sum_{j=1}^{n_e} \gamma_j \hat{\sigma}_j(\omega) + \sum_{j=1}^{n_e} \gamma_{j+n_e} \hat{\bar{e}}_j(\omega), \qquad (7)$$

where:

$\gamma > 0$ = user-defined coefficients for the pixel level standard errors $\hat{\sigma}_j$ and biases $\hat{\bar{e}}_j$ in forest variable j (applied in a genetic algorithm).

ω is the weight vector to be estimated (eq. 2).

W = the feasible set of weight vectors.

The pixel-level biases and errors based on cross-validation,

$$\hat{\sigma}_m = \sqrt{\frac{\sum_{i\in F}(\hat{m}_i - m_i)^2}{n_F}} \quad \text{and the bias} \quad \hat{\bar{e}} = \frac{\sum_{i\in F}(\hat{m}_i - m_i)}{n_F},$$

have been applied. Here m_i is the observed value of the variable to be estimated (e.g., total volume), $\hat{m}_i$ its estimate on plot i, and n_F the number of field plots.

The fitness function weights, bias weights, and RMSE weights were experimentally given values and then fixed. This weighted sum was the criterion in the search for good weight vectors for image features and ancillary information.

The main goal in introducing the i*k*-NN method was to improve the accuracy of the predictions of the volumes by tree species, as well as the estimates in different area units. In practical runs of the genetic algorithm (in estimating the weight vector ω), the pixel-level biases are almost totally removed. An example is taken from Tomppo and Halme (2004) for an area in east Finland with the details provided in the original paper. Examples of pixel-level (field-plot-level) biases in predicting volumes by tree species are given for different methods in table 2: the basic *k*-NN prediction, *k*-NN predictions using large-area variables, and i*k*-NN prediction. Leave-one-out cross-validation and field-data-based volume predictions $\hat{V}_F$ as a reference are employed.

Spruce volume was significantly underestimated with the *k*-NN method. The addition of large-area variables to *k*-NN did not alone reduce the biases, but a reduction was noticeable with i*k*-NN, although all predictions were somewhat lower for birch and other broadleaved tree species than for pine and spruce (columns a/b in table 2).

The predictions are validated at the level of groups of municipalities as follows. The area in question is divided into subareas with forest and other wooded land areas, ranging typically between 150,000 ha and 300,000 ha. An example of the predictions for

Table 2.—*Examples of biases attached to k-NN predictions, k-NN predictions with large-area variables (k-NN, la), and ik-NN predictions at the pixel level (field-plot level) on the mineral soil stratum, leave-one-out cross-validation, using field-data-based volume predictions* $\hat{V}_F$ *as a reference. 1,953 field plots, k = 5, upper bounds used.*

		Bias	Standard error of bias	Bias (a)	Standard error of bias	Bias (b)	Standard error of bias	Reduction
		k-NN	k-NN	k-NN, la	k-NN, la	ik-NN	ik-NN	(a) / (b)
Volume	*m³/ha*	*m³/ha*	*m³/ha*	*m³/ha*	*m³/ha*	*m³/ha*	*m³/ha*	%
Pine	63.750	2.430	1.648	2.230	1.570	– 0.002	1.539	99.925
Spruce	38.883	– 3.167	1.304	– 4.725	1.293	– 0.005	1.260	99.891
Birch	15.903	– 0.961	0.684	– 1.571	0.696	– 0.199	0.701	87.346
O. br. l.	3.874	– 0.382	0.376	– 0.430	0.383	– 0.133	0.389	69.057
Total	122.303	– 2.021	1.827	– 4.432	1.764	– 0.259	1.800	94.152

mean volumes by tree species (m^3/ha) is given for a municipality group with an area of forest and poorly productive forest land of 241,200 hectares in table 3. The table also gives standard errors for the field-data-based predictions. The table enables multi-source predictions to be compared with the field data estimates and assessed in terms of the field-data-based standard errors. The i*k*-NN method gave lower deviations from the field-data-based predictions, and thus more accurate predictions. Use of information on large-area variations in forest variables in conjunction with the i*k*-NN method noticeably reduces the problem of distinguishing pine dominant stands from spruce dominant ones, for instance, or of estimating the volumes by tree species.

Conclusions

Since 1990, the Finnish NFI has been using a satellite-image-aided multisource method to obtain results for smaller areas than is possible using field data only. The entire country has been covered more than twice by this method. The method is under continuous refinement. During NFI9, the method was enhanced by using new features: (1) large-area forest variables for directing the selection of nearest neighbors, (2) an optimization method based on a genetic algorithm to weight both large-area forest variables and satellite image variables, and (3) two optional methods to remove the effect of map errors on the estimates. The new i*k*-NN method performs noticeably better than the original *k*-NN method. The use of information on large-area forest variables considerably reduces the problem of distinguishing stands with different tree species, or tree species composition, and reduces the errors entailed in the estimates of volumes by tree species. Any relevant data, such as soil data or vegetation zone data, can be used as ancillary data.

For several reasons, the pixel-level and stand-level errors of the estimates are rather high with current satellite images. The error sources in pixel-level predictions of forest variables have been listed in many papers (e.g., Katila 2004, Tomppo 2006).

Several methods for assessing pixel-level errors exist. Leave-one-out cross-validation has been used in many cases; Kim and Tomppo (2006) applied variogram modelling to the spectral space. The finding of a generally applicable error estimation method for areas larger than a pixel is a challenging task. Because the error in the predictor of a variable depends on the true value of the variable, errors are spatially correlated, and spatial dependences in the image itself make the error structure even more complex. Lappi (2001) presented a different, calibration-type approach to multisource estimation, together with a variogram-based variance estimator, and some other interesting variogram approaches are currently under development (McRoberts *et al.* 2007).

Practical applications of the MS-NFI technique are also currently facing other problems. A serious problem relates to the optical-area images, in particular, the availability of images obtained under cloud-free conditions. The most applicable satellite sensor, Landsat 7 ETM+, has suffered from a scan line corrector failure since 2003. Several correction methods have been introduced, but the quality of the product is not the same as before (USGS 2005). One advantage of the *k*-NN method is that it is applicable to all image material. The precision of the estimates depends on the spectral, spatial, and radiometric resolution of the sensor, however, and some image material may presume the use of other image material as an intermediate step between the field data and the final image data (Tomppo *et al.* 2002). Furthermore, the precision of the estimate will depend

Table 3.—*Estimates of the volume of growing stock (m3/ha) on forest and other wooded land (a) and its standard error (aer) by tree species based on field data and on the k-NN method (b), ik-NN method, (c) and ik-NN method when the resulting large-area weights have been multiplied by 10 (d) for a municipality group with an area of forest land and other wooded land of 241,200 ha. The multisource estimates are compared with the field-data-based estimates.*

Tree species	a	a_{er}	b	b-a	c	c-a	d	d-a
Pine	48.6	2.9	53.8	5.2	47.5	– 1.1	49.7	1.1
Spruce	38.5	2.5	35.6	– 2.9	41.9	3.4	40.3	1.8
Birch	15.7	1.2	15.7	– 0.0	15.8	0.1	15.9	0.2
Other br. 1.	4.3	0.6	3.7	– 0.6	3.6	– 0.7	3.1	– 1.2
Total	107.2	3.3	108.8	1.6	109.0	1.8	108.9	1.7

on how the *k*-NN method is applied, as seen above. Numerous research work has been conducted to analyze the errors and improve the precision of the estimates; the process is ongoing.

Literature Cited

Cochran, W.G. 1977. Sampling techniques. 3rd ed. New York: John Wiley. 428 p.

Heikkinen, J. 2006. Assessment of uncertainty in spatially systematic sampling. In: Kangas, A.; Maltamo, M., eds. Forest inventory, methods and applications. Managing forest ecosystems. Vol. 10. London: Springer Dordrecht: 155-176.

Katila, M. 2004. Controlling the estimation errors in the Finnish multisource National Forest Inventory. The Finnish Forest Research Institute, Research Papers 910. 36 p. Ph.D. dissertation.

Katila, M.; 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.

Katila, M.; Tomppo, E. 2001. Selecting estimation parameters for the Finnish multisource National Forest Inventory. Remote Sensing of Environment. 76: 16-32.

Katila, M.; Tomppo, E. 2002. Stratification by ancillary data in multisource forest inventories using k-nearest-neighbor estimation. Canadian Journal of Forest Research. 32(9): 1548-1561.

Kim, H.J.; Tomppo, E. 2006. Model-based prediction error uncertainty estimation for k-nn method. Remote Sensing of Environment. 104: 257-263.

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.; Tomppo, E.O.; Finley, A.O.; Heikkinen, J. 2007. Estimating areal means and variances of forest attributes using the k-nearest neighbors technique and satellite imagery. Remote Sensing of Environment. 111: 466-480.

Reese, H.; Nilsson, M.; Granqvist Pahlén, T.; Hagner, O.; Joyce, S.; Tingelöf, U.; Egberth, M.; Olsson, H. 2003. Country-wide estimates of forest variables using satellite data and field data from the National Forest Inventory. Ambio. 32: 542-548.

Tomppo, E. 1990. Designing a satellite image-aided national forest survey in Finland. In: The usability of remote sensing for forest inventory and planning. Proceedings, SNS/IUFRO workshop. Umea, Sweden: Swedish University of Agricultural Sciences, Remote Sensing Laboratory, Report 4: 43-47.

Tomppo, E. 1991. Satellite image based National Forest Inventory of Finland. International Archives of Photogrammetry and Remote Sensing. 28(7-1): 419-424.

Tomppo, E. 1992. Satellite image aided forest site fertility estimation for forest income taxation. Acta Forestalia Fennica. 229.

Tomppo, E. 1996. Multi-source National Forest Inventory of Finland. In: Vanclay, R.; Vanclay, J.; Miina, S., eds. New thrusts in forest inventory. Proceedings, IUFRO XX World Congress. Joensuu, Finland: European Forest Institute: 27-41.

Tomppo, E. 2006. The Finnish Multi-Source National Forest Inventory—small area estimation and map production. In: Kangas, A.; Maltamo, M., eds. Forest inventory, methods and applications. Managing forest ecosystems. Vol. 10. London: Springer Dordrecht: 195-224.

Tomppo, E.; Halme, M. 2004. Using coarse scale forest variables as ancillary information and weighting of variables in k-nn estimation: a genetic algorithm approach. Remote Sensing of Environment. 92: 1-20.

Tomppo, E.; Katila, M.; Moilanen, J.; Mäkelä, H.; Peräsaari, J. 1998. Kunnittaiset metsävaratiedot 1990-94. Metsätieteen aikakauskirja. Folia Forestalia. 4B/1998: 619-839. In Finnish.

Tomppo, E.; Korhonen, K.T.; Heikkinen, J.; Yli-Kojola, H. 2001. Multisource inventory of the forests of the Hebei Forestry Bureau, Heilongjiang, China. Silva Fennica. 35: 309-328.

Tomppo, E.; 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.

Topographic Database of Finland. 1998. The National Land Survey of Finland. http://www.nls.fi/kartta/maps/topodb.html. (1 January 2003).

U.S. Geological Survey (USGS). 2005. Landsat Project. http://landsat7.usgs.gov/slc_enhancements/.