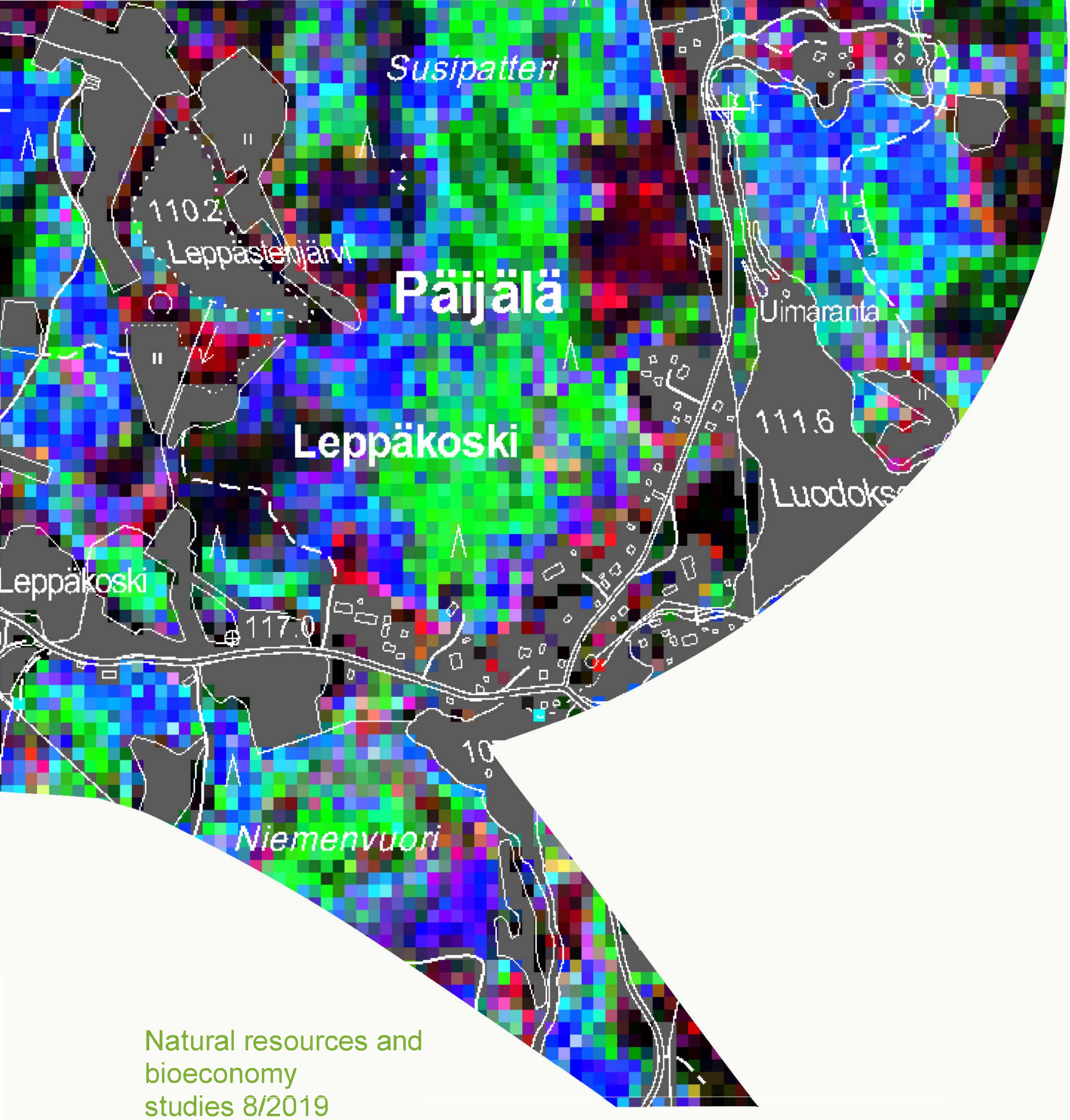


Natural resources and
bioeconomy
studies 8/2019

The Multi-Source National Forest Inventory of Finland – methods and results 2015

Kai Mäkisara, Matti Katila and Jouni Peräsaari

Luke
NATURAL RESOURCES
INSTITUTE FINLAND

Natural resources and bioeconomy studies 8/2019

The Multi-Source National Forest Inventory of Finland – methods and results 2015

Kai Mäkisara, Matti Katila and Jouni Peräsaari

Natural Resources Institute Finland, Helsinki 2019

Luke

Mäkisara, K., Katila, M., and Peräsaari, J. 2019. The Multi-Source national forest inventory of Finland – methods and results 2015. Natural resources and bioeconomy studies 8/2019. Natural Resources Institute Finland, Helsinki. 57 p.

ISBN 978-952-326-710-7 (Print)
ISBN 978-952-326-711-4 (Online)
ISSN 2342-7647 (Print)
ISSN 2342-7639 (Online)
URN http://urn.fi/URN:ISBN:978-952-326-711-4

Authors: Kai Mäkisara, Matti Katila and Jouni Peräsaari
Publisher: Natural Resources Institute Finland (Luke), Helsinki 2019
Year of publication: 2019
Cover photo: MS-NFI volume estimates of birch (red), pine (green) and spruce (blue). Graphic overlay derived from the topographic database from the National Land Survey of Finland (4/2016).
Printing house and publishing sales: PunaMusta Oy, http://luke.juvenesprint.fi

Abstract

Kai Mäkisara[1)], Matti Katila[2)], and Jouni Peräsaari[3)]
[1)]Natural Resource Institute Finland (Luke), Latokartanonkaari 9, FI-00790 Helsinki, kai.makisara@luke.fi
[2)]Natural Resource Institute Finland (Luke), Latokartanonkaari 9, FI-00790 Helsinki, matti.katila@luke.fi
[3)]Natural Resource Institute Finland (Luke), Latokartanonkaari 9, FI-00790 Helsinki, jouni.perasaari@luke.fi

This report presents the methods and results of the Finnish multi-source forest inventory corresponding to year 2015. In addition to field data, satellite images, digital map data and other georeferenced data were used. The main purpose of the article is to make multi-source forest inventory results available for the users and to help the users to understand the principles of the methods and advantages and limitations of the products. The field data originate from the 11th and 12th Finnish National Forest Inventory from years 2012 to 2016. The field data have been computationally updated to the date 31 July, 2015. The satellite images where from years 2015 and 2016 (one frame). The basic features of the improved k-NN, ik-NN, estimation method are described. A new image window based calibration step has been added to processing of some themes.

The results are presented by region (maakunta) and within the regions by municipality, the boundaries as on 1.1.2016. The estimates are given, for example, for land areas, areas of tree species dominance, age, and development classes of stands and often separately for forests available for wood supply. The mean volume and total volume estimates are given in many different ways: by tree species and by timber assortments for forest land, and combined forest land and poorly productive forest land and also for forests available for wood supply, as well as by age and development classes. The biomass estimates are given, in addition to the total biomass estimates, by tree species groups in young thinning stands in which the first commercial thinning was proposed for the first 5-year period, separately for stem and bark and branches and foliage. The biomass estimates of mature forests are presented separately for branches, foliage and stem residuals, and stumps and large roots by tree species groups. These biomass estimates are given separately for land available for wood supply.

In addition to the tabular results, numerical forest resource maps have been computed for 44 themes. The themes include the same variables than in the tables, but the estimation unit is a pixel. Estimates for arbitrary larger units can be computed from the raster maps. Some of the differences between the tabulated results and similar results computed from the maps are discussed in the report.

Keywords: multi-source forest inventory, national forest inventory, remote sensing, satellite images, genetic algorithm, k-nearest neighbours, small-area estimation, stratification, statistical calibration

Tiivistelmä

Kai Mäkisara[1], Matti Katila[2], and Jouni Peräsaari[3]
[1]Luonnonvarakeskus (Luke), Latokartanonkaari 9, 00790 Helsinki, kai.makisara@luke.fi
[2]Luonnonvarakeskus (Luke), Latokartanonkaari 9, 00790 Helsinki, matti.katila@luke.fi
[3]Luonnonvarakeskus (Luke), Latokartanonkaari 9, 00790 Helsinki, jouni.perasaari@luke.fi

Tässä raportissa kuvataan monilähteisen valtakunnan metsien inventoinnin (MVMI) vuoden 2015 menetelmät ja tulokset. VMI:n maastoaineiston lisäksi tulosten laskennassa käytettiin satelliittikuvia, numeerista kartta-aineistoa sekä muita paikkatietoaineistoja. Raportin päätarkoitus on esitellä käyttäjille monilähteisen valtakunnan metsien inventoinnin tulokset ja auttaa käyttäjiä ymmärtämään käytettyjen menetelmien periaatteet sekä tuotettujen aineistojen ominaisuudet ja rajoitteet. Käytetty maastoaineisto on 11. ja 12. valtakunnan metsien inventoinneista kattaen vuodet 2012 – 2016. Käytetyt satelliittikuvat ovat yhtä lukuunottamatta vuodelta 2015 (yksi kuva vuodelta 2016). Raportissa kuvataan käytetty ik-NN-estimointimenetelmä. Uusi kuvakohtainen kalibrointi on lisätty aiemmin käytettyihin menetelmiin joidenkin teemojen kohdalla.

Tulokset on esitetty maakunnittain ja maakuntien sisällä kunnittain käyttäen 1.1.2016 voimassa olleita rajoja. Tuloksia ovat esimerkiksi estimaatit maaluokkien pinta-aloista, eri puulajien vallitsevuuksista, puuston iästä sekä kuvioiden kehitysluokkajakaumista. Useat tulokset on taulukoitu erikseen koko metsätalouden maalle ja puuntuotannon käytössä olevalle alueelle. Keskitilavuusestimaatit on laskettu useille eri tulosositteille: puulajiryhmittäin ja puutavaralajeittain metsämaalle, yhdistetylle metsä- ja kitumaalle ja puuntuotannon käytössä olevalle maalle sekä myös ikäluokittain ja kehitysluokittain.

Biomassan määrä on esitetty puulajiryhmittäin ja puun eri ositteille. Erikseen energiapuuksi sopivien biomassaositteiden määrä on esitetty nuorille kasvatusmetsille, joille on ehdotettu hakkuita ensimmäisellä 5-v. kaudella, ja uudistuskypsille metsiköille.

Taulukkomuotoisten tulosten lisäksi on laskettu numeeriset metsävarakartat 44 eri teemasta. Nämä teemat kuvaavat samoja muuttujia kuin taulukotkin, mutta estimointiyksikkönä on kuva-alkio. Rasterikartoista voidaan laskea tuloksia mielivaltaisille alueille. Raportissa käsitellään joitakin eroavaisuuksia taulukoitujen ja rasterikartoista laskettujen tulosten välillä.

Asiasanat: monilähteinen metsien inventointi, valtakunnan metsien inventointi, kaukokartoitus, satelliittikuvat, geneettiset algoritmit, k-lähimmän naapurin menetelmä, pienalue-estimointi, ositus, tilastollinen kalibrointi

Contents

The tables listed in the Appendix can be retrieved from the URL:
`http://urn.fi/URN:ISBN:978-952-326-712-1`.

List of abbreviations

k-NN	k-Nearest Neighbour method
ik-NN	Improved k-NN method
NFI	National forest inventory
MS-NFI	Multi-source national forest inventory
MS-NFI-2015	Multi-source national forest inventory for 2015
MS-NFI-OA-2015	Open access MS-NFI-2015
Sentinel-2A/B	European remote sensing satellites in the Copernicus program
Sentinel-2 MSI	MultiSpectral Instrument, the optical imager on board the Sentinel-2 satellites
Landsat	Land satellite
Landsat OLI	Operational Land Imager, a high-resolution multispectral imaging system on board the Landsat-8 satellite.
NLS	National Land Survey of Finland
DEM	Digital elevation model
RMSE	Root mean square error
SE	Standard error
FRYL	Forestry land, in MS-NFI covers forest land, poorly productive forest land, and unproductive forest land

1 Introduction

This report presents the municipality levels estimates of forest parameters, and describes the methods and parameters used both when making the year 2015 version of the forest resource maps and the municipality level estimates. This inventory is called MS-NFI-2015 throughout this article. The report is meant to be self-sufficient for understanding this version of the results, but this means that there is a lot of overlap with publications describing the previous versions (e.g., Tomppo et al. 2013, 2014, Mäkisara et al. 2016). A more thorough description of the background is in the book by Tomppo et al. (2008b) describing the results based on the ninth inventory (1996–2003).

The First National Forest inventory (NFI) was carried out in Finland from 1921–1924. The eleventh inventory began in 2009 and the field measurements were completed in 2013. The twelfth inventory started in 2014 and the measurements will be completed in 2018. Field data from both of the two latest inventories has been used in the work reported here.

Based on the information from sample plots, estimates can be made for the entire country, or regions within a country, with a minimum size of about 300 000–500 000 hectares, depending on the forest parameter. The densities of plots are high enough to ensure that the resulting sampling errors are low for core variables, such as areas of land classes and the volume of growing stock. The estimates of forest parameters are currently presented by *regions* (maakunta). Finland is divided into 19 regions (see Fig. 1.1). The forestry land areas of the regions vary from 117 000 (Åland) to 9 million hectares (Lapland) (Korhonen et al. 2017).

High utilisation levels of forests and the changing practices of forestry and forest industries in the 1980's and '90's required more accurate, localised and up-to-date information than previously. The use of field data alone to respond to the increased information needs would have been an expensive alternative. Either a substantially denser field plot grid or some other type of information was required.

The development of the Finnish multi-source national forest inventory (MS-NFI) began in the Forest Research Institute of Finland (Metla) in 1989, and the first operative results were calculated in 1990 (Tomppo 1990, 1991, 1996, 2006b). The MS-NFI was introduced during the 8th rotation of NFI (1986–1994) in the Pohjois-Savo region of the Public Service Unit of the Finnish Forest Centre (see Fig. 1.2). The first results for the entire country were published in 1998 (Tomppo et al. 1998). The second country level results were published in 2008 (Tomppo et al. 2008b), the third results corresponding year 2005 and covering South and Central Finland (Tomppo et al. 2009a), and the fourth results corresponding to year 2007 and covering the entire country except Åland and the most northern Lapland (Tomppo et al. 2012). After this, results have been produced to correspond to a certain year, 2009 (Tomppo et al. 2013) (Northernmost Lapland excluded), 2011 (Tomppo et al. 2014), 2013 (Mäkisara et al. 2016), and now for 2015. In total, eight sets of country level results have been computed.

For MS-NFI, methods were sought that would provide area and volume estimates, possibly broken down into subclasses, such as tree species, timber assortments, and stand-age classes. Since NFI10, the estimates of potential wood energy biomass from forests have been produced. In the optimal case, the MS-NFI method should be able to provide estimates for small areas as accurate as the field based method provides estimates at national and regional levels. Since the first implementation of this method, it has been modified continuously and new features have been added (Katila et al. 2000, Katila and Tomppo 2001, 2002, Tomppo et al. 2009a). The core of the current Finnish MS-NFI method is presented in (Tomppo and Halme 2004) as well as in (Tomppo et al. 2008b).

Somewhat similar methods, which combine field data and satellite images, have been developed and employed or tested in a several other countries like Sweden (Nilsson 1997, Reese et al. 2002, 2003, Hagner and Olsson 2004), USA (Franco-Lopez et al. 2001, McRoberts et al. 2002a,b, McRoberts 2006), Norway (Gjertsen 2005), Austria (Koukal et al. 2005), New Zealand (Tomppo et al. 1999), China (Tomppo et al. 2001), Germany (Diemer et al. 2000) and Italy (Maselli et al. 2005). Tomppo et al. (2008b) gives a list of references. The Swedish k-NN product is used for a multitude of purposes, as well as a basis for post-stratification to produce the official Swedish forest statistics, see also (Tomppo et al. 2008b). McRoberts et al. (2002a), McRoberts et al. (2002b) and Haakana et al. (2018) also applied k-NN products to post-stratified estimation. Other examples of the development work in USA are the studies by McRoberts (2006) and McRoberts et al. (2007), who presented a model-based approach to derive k-NN error estimators for a group of pixels at an arbitrary size, Finlay et al. (2006) and Finlay and McRoberts (2008), who presented two methods of increasing the efficiency of the k-NN search. A review of using remote sensing data in NFIs is presented by McRoberts et al. (2010a) and McRoberts et al. (2010b). An error estimation method based on bootstrapping is presented by McRoberts et al. (2011) and a BRR resampling method by Magnussen et al. (2010).

The progress of the Finnish NFI was changed somewhat for NFI10 (2004–2008). From the fifth to the ninth inventory, the measurements proceeded by region each year. In NFI10, the scheduling was changed so that one-fifth of the plots were measured each year (excluding Åland and Northern Lapland, where measurements were completed within one or two years). At the same time, the inventory rotation was shortened to 5 years, or nearly half of its previous rotation duration.

This change made it possible to compute the basic forest resource estimates annually for the entire country, both from field data and in MS-NFI. In the NFI plans, MS-NFI results were decided to be calculated every second year. The new approach to progress set some additional challenges for the MS-NFI, e.g., field measurements from several years can and must be employed.

To better use the field data from several years, the method for making MS-NFI-2011 was improved and the same basic method has been used since then. Instead of just removing from the ground data the field plots that have changed between imaging and field work, the ground data have been updated to a set date (31 July 2015 in this case). The forest variables at plots that have been cut (or otherwise radically changed) have been updated to the new state. All of the field data have been computationally updated so that the total volume matches the total volume estimated from the field plots for each processing window for year 2015. The method is described in more detail in Section 3.2.3.

Note that this change affects the difference between MS-NFI-2011 or later and MS-NFI-2009 or earlier. With the old method, the mean field data correspond roughly to the midpoint of the field data interval. With the new method, the field data correspond roughly to the end of the field data interval. This means that, for example, the difference between MS-NFI-2009 and MS-NFI-2011 corresponds to a longer time span than two years.

The availability of satellite data has changed from MS-NFI-2013. The Landsat 5 satellite was decommissioned June 5, 2013. The Landsat 8 satellite provided most of the data for this product. The new ESA Sentinel-2A satellite was launched June 23, 2015. Data from one overflight of Sentinel-2A in August 2015 was used in this product. In future, the Sentinel-2A and Sentinel-2B satellites, together with Landsat 8, increase the availability of satellite data.

The methodology was improved from the previous products. The level of several variables are now calibrated for each processing window separately (see Section 3.4.6). This makes the averages computed for large areas from the raster maps match more closely the results computed using only field data (see Subsection 4.2). This is believed to improve the accuracy also for smaller areas.

The ik-NN (improved K-Nearest Neighbour) estimation method used in MS-NFI is very flexible. It can

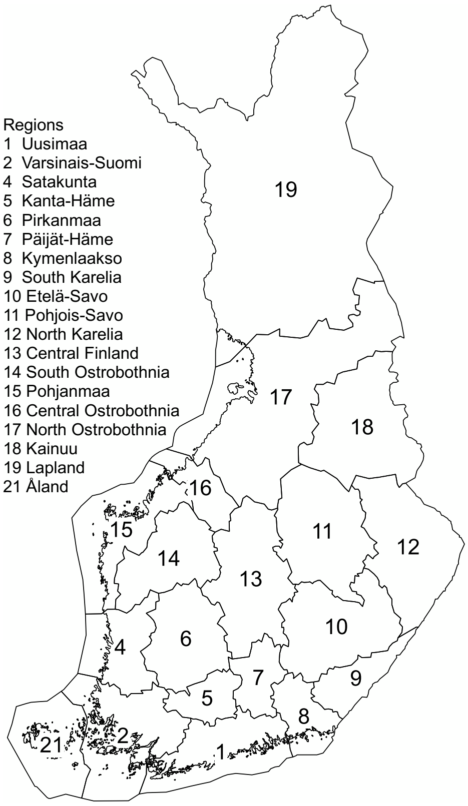


Figure 1.1: The regions (maakunta) of Finland. Digital map data: contains data from the National Land Survey of Finland general map 1:4.5 M 06/2015 and municipal division 1:4.5 M 01/2018.

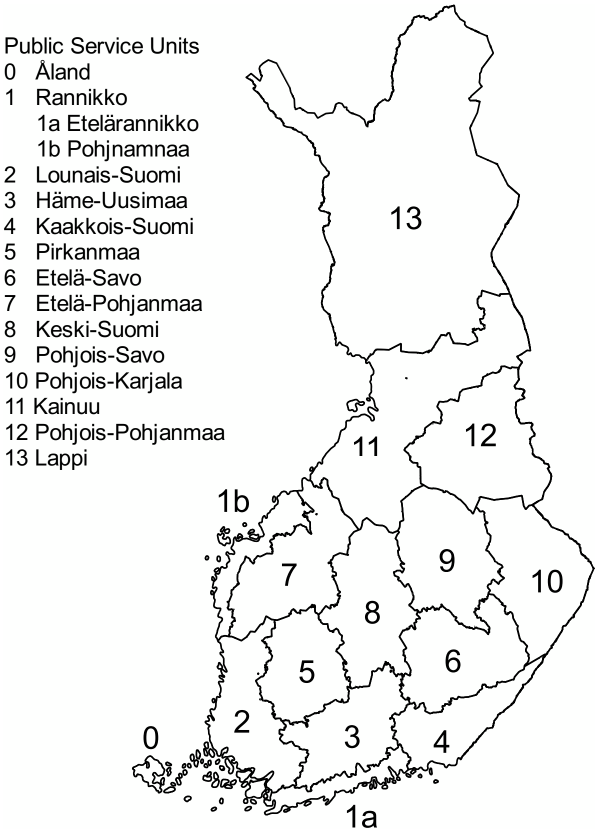


Figure 1.2: The Public Service Units of the Finnish Forest Centre and Åland region 1.1.2013. Digital map data: ©National Land Survey of Finland, licence No. MML/VIR/MYY/328/08.

easily adapt to gradual changes within the image by using constraints in the geographic distances allowed for the neighbours in the feature space. To compute estimates for regions, the weights assigned to the nearest neighbours can be accumulated for the regions. This enables in some cases more accurate estimates for regions of values than using the pixelwise estimates. An example of this is shown in Subsection 4.4.

The main users of the MS-NFI results, municipality level estimates and forest resource maps are the forestry authorities at the Finnish Forest Centre, forest industries and forest environment researchers. More details of the uses are given in Tomppo et al. (2008a,b, 2012, 2013). The number of users has increased after the map form estimates from MS-NFI-2009 were made publicly available in November 2012. The maps from MS-NFI-2015 have been added to the publicly available data sets in December 2017.

The objectives of this article are two-fold, 1) to present the methods and results for the products called multi-source national forest inventory of Finland for year 2015 (MS-NFI-2015), 2) to present briefly the third set of the MS-NFI products used in the Open spatial data sets and interface services available in Finland free of charge at `http://www.paikkatietoikkuna.fi/web/en`, Paikkatietoikkuna (2018) and `http://kartta.luke.fi/index-en.html` (Luke 2018). This data set is called here open access MS-NFI-2015 and abbreviated as MS-NFI-OA-2015. The municipality level results are available in electronic form at the Luke web site `https://www.luke.fi/en/natural-resources/forest/forest-resources-and-forest-planning/forest-resource-maps-and-municipal-statistics/` and as an appendix of this report. In the appendix the municipalities are grouped by region.

2 Materials

2.1 Field data

The field sample plots from NFI11 from years 2012–2013 and from NFI12 from years 2014–2016 were used for the results presented in this article. In total there were 78 312 plots, 66 512 plots on land, 55 029 on forestry land (including forestry roads) and 54 551 plots on the combined forest land, poorly productive forest land and unproductive land.

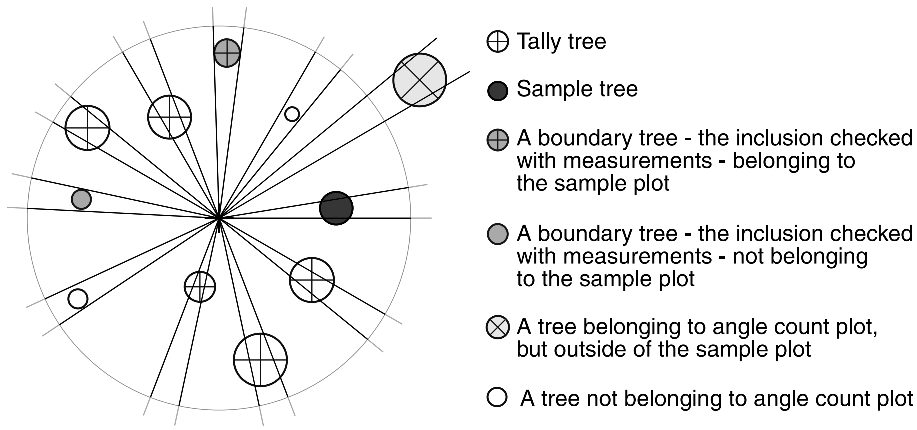


Figure 2.1: A sample plot of NFI11. The maximum distance to trees to be counted is 12.52 m in South Finland (q=2) and 12.45 m in North Finland (sampling regions 4–6 in Fig. 2.2) (q=1.5). Every 7th tree is measured as sampling tree, counted by crews starting from the beginning of the field season.

The Finnish national forest inventory is a sampling based inventory. The sample plots are arranged into clusters. The field measurements and assessments of the NFI are carried out on the field sample plots and on those forest stands that include at least a part of a field plot. The field sample plot is also a unit in the field data based estimation (Tomppo 2006a, Tomppo et al. 2011a, Korhonen et al. 2017).

The field sample plot of NFI has been an angle count plot (Bitterlich) plot since NFI5 and was also in NFI11. A maximum distance from the centre point of the plot to the trees to be included into the plot was introduced during NFI8 in North Finland in 1991 (Fig. 2.1). The maximum distance detracts very little from the reliability of the estimates but decreases the amount of field work and also reduces possible errors caused by unobserved trees (Korhonen et al. 2017).

For NFI12, the field plot type was changed. A fixed radius plot is used, where all trees with diameter $\geq$ 95 mm are measured up to distance of 9 meters. In addition to this, trees with diameter $\geq$ 45 mm are measured up to distance of 5.64 meters. Small trees with diameter $<$ 45 mm are measured using a relascope plot with relascope factor 1.5.

The basic principles of NFI11 and NFI12 design are similar to those of NFI10 and NFI9 (Tomppo 2009a). The country is divided into six sampling regions, shown in Figure 2.2. Åland was further subdivided into 2 regions in NFI11 because ALS-based inventory using NFI sample plots was also performed in the central part (Åland 2, 2.1).

Except in Northern Lapland, the sampling structure can be explained using repeating blocks of clusters. Each block consists of permanent clusters (that can't be moved between inventories) and temporary clusters (that can be moved between inventories). Each cluster consists of a number of sample plots.

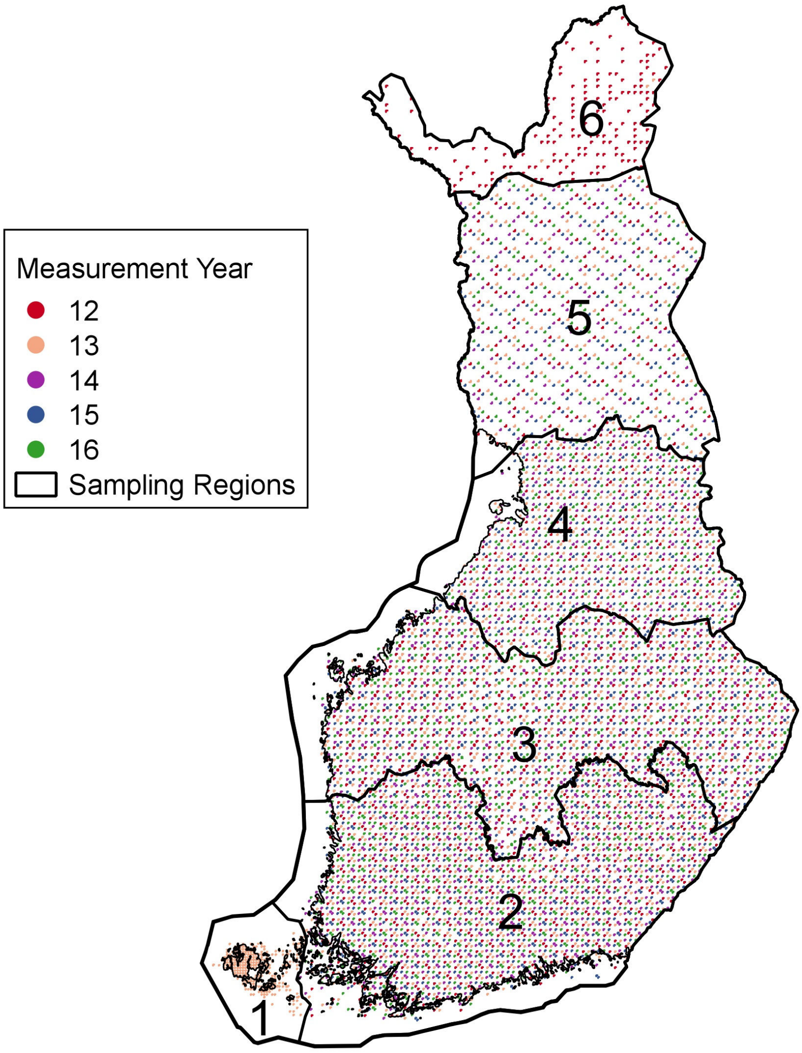


Figure 2.2: NFI layout of clusters and the six geographic regions with different sampling intensities, NFI11 2012, NFI12 2013–2016. Contains map data from the National Land Survey of Finland Municipal Division datasets 03/2014 and 06/2018.

The number of plots in a cluster and the cluster structure should enable measurement of one cluster in a single work day. As an example, the block structure used in Central Finland in NFI11 is shown in Fig. 2.3. The block parameters for the different regions in NFI11 and NFI12 are shown in Table 2.1. These parameters include the nominal representativeness (hectares/plot) that describes the density of the plots in each sampling region.

Stratified sampling was used in Northern Lapland. Based on previous MS-NFI results, the region was divided into six strata. All clusters consisted of nine plots. The number clusters and the locations of the clusters were optimised base on MS-NFI estimates and other data (Korhonen et al. 2017). The representativenesses of plots in each stratum varied from 1048 to 7208 hectares per plot.

The tree level stem volumes on the field sample plots are converted to volumes per hectare in the MS-NFI using the expansion factor. After this, the different plot structure in NFI11 and NFI12 is not visible in the computation. Volumes per hectare are estimated for each sample plot by tree species and by timber assortment classes based on the tally tree volumes. The estimation of volumes and volumes of timber assortments for tally trees from field measurements is described in Tomppo et al. (2011a). Otherwise, the field variables used are similar to those in the NFI calculations that use field data only.

Table 2.1: Sampling parameters for the repeating blocks in the different regions in NFI11.

NFI	region	distance (km)	perm clust		temp clust		hectares / plot
			clust	plots	clust	plots	
11	Åland 1	12	1	10	15	9	99
11	Åland 2	12	1	10	9	9	158
11	South	12	1	10	4	9	313
11	Central	14	1	14	4	11	338
11	Kainuu-PP	14	1	11	4	9	417
11	South Lapland	20	1	11	3	12	851
12	Southern	12	1	10	4	8	343
12	Central	14	1	14	4	9	392
12	Kainuu-PP	14	1	11	4	8	456
12	South Lapland	20	1	11	4	10	784

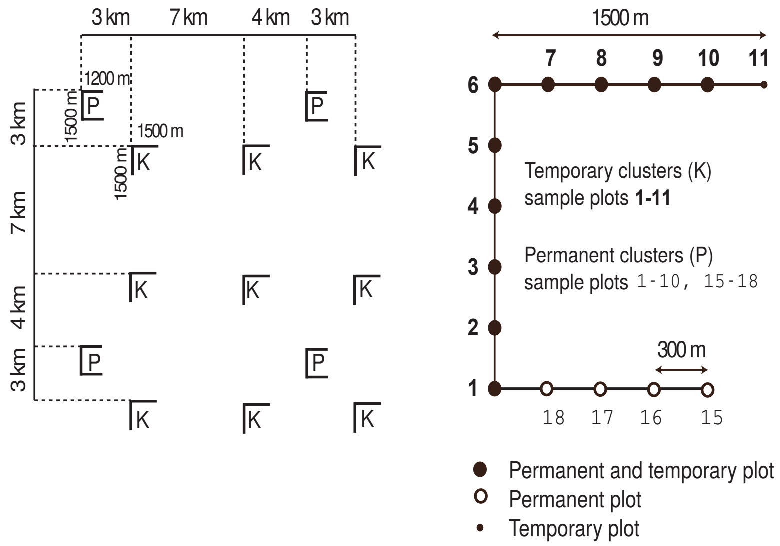


Figure 2.3: Sampling design for the Central Finland for NFI11.

Biomass estimates were calculated for each field plot by tree species groups and by tree compartments. The biomass estimates can be used for assessing carbon balance of forests in small areas, and for energy wood estimation.

2.2 The employed satellite images

High-resolution (about 10–30 metres pixel size) multispectral satellite images were used in the operative application. Large coverage and good availability with reasonable price, or free of charge, were additional selection criteria. Based on these requirements, the Landsat 8 OLI sensor was the most suitable one for this application. High quality Landsat 8 OLI images over Finland were available for summers 2015 and 2016, and these were used for most of the processed area. These images were available from USGS (U.S. Geological Survey) without cost (Landsat 2015). The images were orthorectified to UTM zones 34, 35, and 36, as appropriate.

In addition to the Landsat 8 OLI data, data from the new ESA Sentinel-2A satellite's MSI scanner were used. The data were from one of the first operational overflights over Finland (relative orbit 79). The Sentinel-2 data does not have frame structure similar to Landsat 8. The data is available in 100 by 100

Table 2.2: List of satellite images used in the MS-NFI-2015: image index number in Fig. 2.4, satellite sensor, path/row, acquisition dates of image frames and number of image frames in one image.

Image No. (Fig. 2.4)	Sensor	Path/Row	Date	No. of image frames
1	Landsat 8 OLI	185/16	24082015	1
2	Landsat 8 OLI	186/16to18	15082015	3
3	Landsat 8 OLI	187/14to18	22082015	5
4	Landsat 8 OLI	189/13to15	20082015	4
5	Landsat 8 OLI	189/16	20082015	4
6	Landsat 8 OLI	189/16to18	03072015	3
7	Landsat 8 OLI	189/14to16	07102015	3
8	Landsat 8 OLI	191/11to18	18082015	8
9	Landsat 8 OLI	191/14to15	03092015	2
10	Landsat 8 OLI	192/12to14	10092015	2
11	Landsat 8 OLI	193/11	18082016	1
12	Landsat 8 OLI	193/12to13	16082015	2
13	Landsat 8 OLI	194/12to13	23082015	3
14	Landsat 8 OLI	195/11to12	13072015	2
15	Sentinel-2A MSI	079	17082015	NA

km tiles in the UTM coordinate system along the satellite's path.

The land area of Finland, analysed for MS-NFI-2015, was 30.39 million hectares, and the total area together with the lakes and rivers 33.84 million hectares. For processing, this area was covered using 15 images assembled from 40 Landsat 8 OLI image frames and data from one Sentinel-2A overflight (Table 2.2, Fig. 2.4). Where possible, adjacent satellite image frames from same path and date were combined to increase the number of field plots within image and to simplify processing.

For MS-NFI-2015, the target year of image acquisition was 2015. 39 of the 40 Landsat 8 OLI image frames and the Sentinel-2A overflight were from that year. A suitable imaging season for forest inventory purposes in Finland is from mid-May until the end of August, with the optimal time being from early June until the end of July. The imaging season is rather short and cloud cover is frequent even in the summer, which makes image acquisition problematic. Three images outside the summer season were used to finalise the cover.

All the satellite images needed to be in the ETRS-TM35FIN coordinate system for further processing. The Landsat 8 OLI images were delivered in the UTM projection but the rectification accuracy was not good enough for our purposes. Some of the images were delivered in UTM projection zones 34 and 36. Because of these facts, the OLI images were re-rectified to the ETRS-TM35FIN projection. The location accuracy of the Sentinel-2A data was good enough, and only reprojection to the ETRS-TM35FIN coordinate system with pixel size of 16 by 16 meters was done. See Subsection 3.1 for details.

For processing, the images number 8 and 15 were divided into three parts. The total number of image windows in processing was 19.

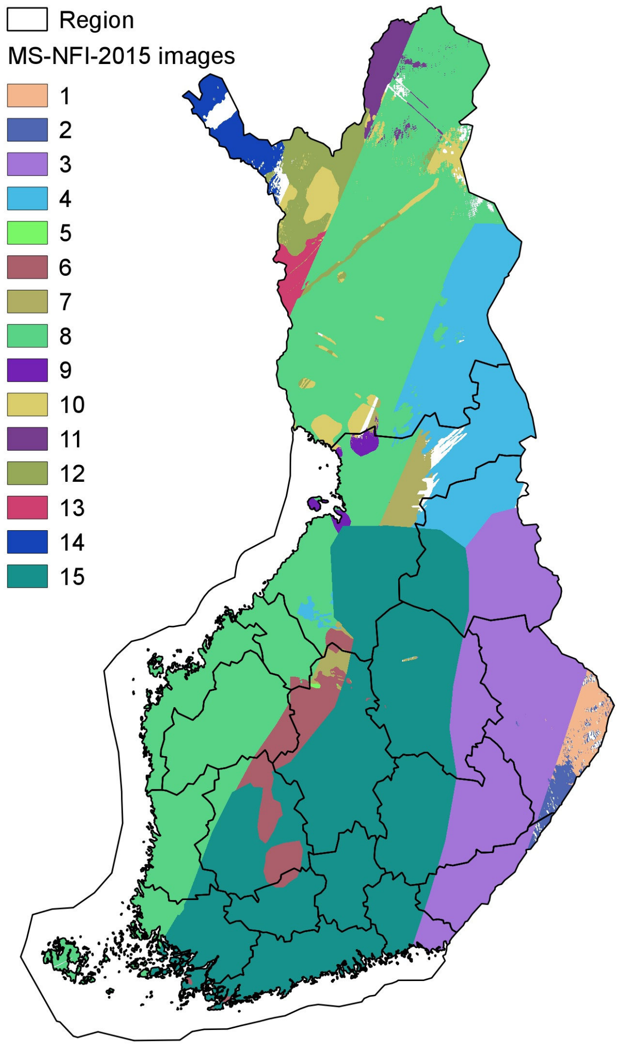


Figure 2.4: The satellite image mosaic used to cover the Finland in MS-NFI-2015 and MS-NFI-OA-2015 and the boundaries of regions. Digital map data: contains data from the National Land Survey of Finland General Map 1:4.5 M 06/2015 and Municipal Division 1:4.5 M 01/2018.

2.3 Digital Map Data

2.3.1 The use of the 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 multi-source method if the differentiation of forestry land from non-forestry land can be supported by digital map information in addition to satellite images (Tomppo 1996). The map information is used to separate forestry land from other land classes, such as arable land, built-up areas, roads, urban areas and single houses. The effect of possible map errors on small-area estimates is reduced by using one of two alternative statistical methods (Katila et al. 2000, Katila and Tomppo 2002, Tomppo et al. 2008b). The first one is a calibration method using a confusion matrix derived from the land class distributions on the basis of field plot data and map data, and the second one employs stratification of the field plots on the basis of map data (see Subsection 3.4.4). In addition, the map data are used to stratify the forestry land area and the corresponding field plots into a mineral soil stratum, a peatland soil stratum and open bog and fen stratum. Digital map data is also

Table 2.3: Sources and quality of the numerical map data.

Map theme(s)	Delivered by	Scale	Date in database	Area covered	Data source
Topographic database; land use classes, peatlands, municipality borders	NLS	1:10 000	4/2016	whole country	Topographic database
Protected forests	SYKE		17.11.2015	whole country	Nature conservation
Nature conservation programmes	SYKE		1978–1996	whole country	Nature conservation databases
Regional land use plans (Maakuntakaava), protected areas 'S'	SYKE	1:250 000	10.5.2016	most of the country	Regional land use GIS-database

used to delineate the computation units in the MS-NFI. Areas of protected forests were used to calculate estimates for forestry land available for wood supply by municipalities.

The mineral soils and different types of organic soils (peatland soils) can have significantly different spectral signatures even when the growing stock is the same (e.g., Katila and Tomppo 2001). In addition, some peatland cannot be separated from mineral soils by means of remote sensing. Therefore, stratification based on digital peatland information is used to decrease the prediction and estimation errors (Tomppo 1996, Katila and Tomppo 2001). The site class definition is vegetation-based in the NFI: the forest stand is considered to be peatland (spruce mires, pine mires, open bogs and fens) if the organic layer covering the mineral soil is peat or if 75 % of the under storey vegetation is peatland vegetation (Lehto and Leikola 1987). A geological definition of peatland is used for the topographic mapping: peatland is covered mainly by peat vegetation and the thickness of peat layer is over 30 cm. Thus, the peatland mask can not be used in a categorical way, but it is used to stratify the forestry land, satellite images and corresponding field plots for subsequent analysis in the estimation phase. Stratification is used to avoid the biases caused by the peatland map that deviates from the peatland used in the NFI, the deviations caused by the different definitions and locational errors in the maps.

Almost all map data were obtained in vector format from the Topographic database of National Land Survey of Finland (NLS) (Table 2.3) (Maanmittauslaitoksen maastotietokohteet 2013). A raster map with 16 m by 16 m pixel size was computed from selected topographic database elements. For the purposes of the calibration method (Subsection 3.4.4), the overlaying of the map elements was done in such a way that it would be possible to form map strata as homogeneous as possible with respect to the NFI field plot based land class distribution (Katila et al. 2000). The main objective was to use the derived digital land use map (Fig. 3.2b) to obtain as precise estimate as possible for the combined forest land, poorly productive forest land and unproductive forest land (denoted by forestry land) when compared to the NFI field data based estimate.

2.3.2 The map data

The elements from the topographic database The version of 2016 of the topographic database was used. The positional accuracy of the topographic database is comparable to maps on scale 1:5 000–1:10 000 (Maanmittauslaitoksen maastotietokohteet 2013). The topographic database consists of map sheets which are updated in 5–10 year periods. However, all roads and almost all the buildings are updated annually, as well as administrative boundaries. Due to this staggered processing, the date of the topographic map elements varies.

All the map elements were rasterised to 16 m by 16 m pixel size. The rasterisation was done separately for each element and suitable widths for line elements (roads, power lines, etc.) and buffer zones for buildings were defined. These widths and buffer zones were determined iteratively by comparing the proportions of the land classes based on the rasterised topographic database and the NFI10 field plot data (Tomppo et al. 2012). The main principle in the rasterisation and generalisation was to keep the total area covered by each map theme as same as that based on the NFI data. The visual appearance of the non-forestry land classes in the MS-NFI output map was considered to be of secondary importance. The selected elements of topographic database, possible width in processing and their priority in overlaying of the elements are described in Table 2.4. The agreement of the resulting map with the NFI field observations has been evaluated in Section 4.1.

The topographic database includes subclasses of open bogs, woody peatland (peat depth $>=$ 30 cm) and paludified lands (peat depth $<$ 30 cm). It was therefore possible to stratify the peatland in the ik-NN estimation into open bogs and woody peatland (Subsection 3.4.1). Paludified peatland correspond most often to mineral soils in NFI field plots and were thus kept as mineral soils in MS-NFI-2015.

Arable land is the third largest land class after forestry land and inland waters with an area of 2.7 million hectares (Finnish Statistical Yearbook of Forestry 2014). The area of the forestry land was 26.2 million hectares and the area of the inland watercourses 3.5 million hectares. Most of the land use changes occur between arable land and other land classes.

Urban areas and other built-up areas (e.g., mineral resources extraction areas, peat production areas, landfill areas, cemeteries, air-fields, parks, sports and recreation areas) were delineated using elements of the topographic database.

Digital boundaries of the computation units The basic computation unit in the multi-source inventory is the municipality, the number being 313 at the beginning of 2016. Their land areas range from around 1000 hectares to some hundreds of thousands of hectares (up to 1.5 million hectares in Inari). Digital municipality boundaries are used to delineate the computation units (Tomppo 1996). The boundary information originates from NLS topographic database and was obtained in vector format. The map data and land areas of the municipalities dating 1.1.2016 were employed to calculate the small area estimates (cf. calibration to official land areas, Subsection 3.4.4).

Digital boundaries of protected forests Some estimates for municipalities were calculated for forestry land available for wood supply (Subsection 4.5, Appendix Tables). Areas of protected forests and nature conservation programmes obtained from the Finnish Environment Institute were used for this purpose (Ympäristöhallinnon paikkatietoaineistot 2018). The protected forests data were obtained in vector format dating 17.11.2015, while the nature conservation programmes delineations originate from the date of founding the programmes (1978–1996). All the map data were rasterised to 16 m by 16 m pixel size. The protected areas included strict nature reserves, national parks, wilderness areas, special protected areas, protected old-growth forest areas, protected herb-rich forest areas, mire conservation areas, nature reserves on private land (protected permanently or temporarily), protected areas established by the Finnish Forest and Park Service and natural habitat types preserved on the basis of Nature Conservation Act. The nature conservation programmes employed are: ”Aarnialue”, areas protected based on decision by the authority responsible of management; mires; herb-rich forests; natural parks and nature reserves developing; avian water areas (’Mikkelin saaret’). The nature conservation programmes digital database has not been updated since its creation. Therefore it contains a) areas where there is not yet final decision of protection made, b) areas with decision of status made. However, among the latter ones there is a minor proportion of areas which have been rejected and thus are erroneously classified as protected forests in our mask.

Table 2.4: The elements of the topographic database selected to the land class map, applied line buffer zone widths in the rasterising. The elements are in the priority order in the table (uppermost first). Forestry land is the area not covered by elements extracted from the topographic database.

MS-NFI code	MS-NFI Map code	Description	Topog. database code(s)	Notes
14	30252	Roads	See roads below	Overl. over water
90	30253	Sea water	36211	
91	30253	Fresh water	36200, 36313	
92	30253	Decomposing reliction area	38300	
13	30251	Built-up area	40200	
16	30252	Railroad	14111, 14112	Buffer 17 m
14	30252	Roads, class Ia - IIIb	12111, 12112	Buffer 16 m
			12121	Buffer 12 m
			12122	Buffer 9 m
			12131, 12132	Buffer 8 m
22	30254	Agricultural field	32611	
19	30250	Quarry, gravel pit	32500, 32111, 32112	
20	30248	Peat production area	32113	
21	30254	Meadow	32800	
12	30252	Airport	32410 – 32418	
3	30250	Graveyard	32200	
4	30250	Landfill	32300	
5	30250	Garden	32612	
6	30250	Park	32900	
7	30250	Earth fill	33000	
8	30250	Sports/recreational area	33100	
11	30250	Built construction	45700	Buffer 5 m
9	30250	Basin	44300	Buffer 5 m
10	30250	Storage area	38900	
15	30252	Other road	12141	Buffer 5 m
17	30250	Power line	22311	Buffer 14 m
			22312	Buffer 5 m
18	30250	Gas pipe	26111	Buffer 20 m
102	prediction	Bare sand	34300	
104	prediction	Exposed bedrock	34100, 34700	
1	30251	Building	42211	Buffer 25 m
			42210, 42212	Buffer 30 m
			42220 – 42222	Buffer 20 m
			42230 – 42232	Buffer 10 m
			42240 – 42242	Buffer 30 m
			42251, 42270	Buffer 30 m
			42260 – 42262	Buffer 5 m
106	prediction	Paludified area, <= 0.25 ha	35300	
107	prediction	Paludified area <= 0.5 ha	35300	
105	prediction	Paludified area	35300	
109	prediction	Forested marsh <= 0.25 ha	35412, 35422	
110	prediction	Forested marsh <= 0.5 ha	35412, 35422	
108	prediction	Forested marsh	35412, 35422	
111	prediction	Open bog	35411, 35421	
112	prediction	Open reliction area	39130	
101	prediction	Forestry land <= 0.25 ha		
102	prediction	Forestry land <= 0.5 ha		
100	prediction	Forestry land		

Thirdly, the protected areas (code ’S’) from the regional land use plans were used to complete the protected forests mask. The data sources for the regional land use plans is called ’Maakuntakaava’ as of 10.5.2016.

It has been noticed that the area of forestry land not available for wood production is an underestimate when compared to the field inventory based estimates from the NFI11. This is due to the multiple-rule based definition of the Metsähallitus non-production areas. Some of these areas were not available in digital formats, e.g., poorly productive forest land and wasteland land areas protected by the decision of the Metsähallitus.

2.4 Digital elevation model

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 angle and the terrain normal (Subsection 3.4.1) (Tomppo 1992, Tomppo et al. 2008b, 2012, 2013). Stratification in this context means using the maximum vertical distance from a pixel to its nearest neighbours. The selection of parameters for stratification and spectral correction has been studied by Katila and Tomppo (2001). The digital elevation model (DEM) employed was a raster file with a horizontal spatial resolution of 10 metres by 10 metres and with a vertical resolution of 0.1 metres (Korkeusmalli 25m 2017). The values used in the 16 by 16 metre grid were obtained by applying a Gaussian filter to the data after resampling. This was done to prevent artefacts in the slope computed from the DEM. The full width of the Gaussian at half maximum (FWHM) was 27 meters.

2.5 Large area forest resource data

The improved ik-NN method was introduced during NFI9. It employs a coarse scale variation of the key forest variables to guide the selection of field plots from which the data are transferred to the pixel to be analysed. The variation is presented in the form of coarse-scale digital forest variable maps (Fig. 2.5), derived either from the current inventory data or from the data of the preceding inventory. For MS-NFI-2015 of the present article, the NFI field plots from 2006–2010 were used. The large area changes in forests are slow and the tree species proportions of the volume of growing stock do not change essentially in a few years. The coarse scale maps made with the field data from the years 2006–2010 are thus relevant also for the MS-NFI products for the year 2015.

There were 72 234 field plots on land across the entire country in the 2006 – 2010 data, of which 59 785 were on forestry land, 54 828 on combined forest land and poorly productive forest land, and 50 492 were on forest land alone. All the plots on forest land and poorly productive forest land were used for the final large-area maps. The principles construction of the maps is described in (Tomppo et al. 2008b). Moving average interpolation was used. The cluster level averages of the volumes by tree species groups (pine, spruce, birch species and other broad-leaved tree species) were first calculated. The averages of these cluster level averages were calculated within a circle of a radius of 30 km and the original cluster levels averages were replaced by these moving averages. The values of these moving averages were predicted for each grid cell of 1 km x 1 km in Finland using 1-NN method. The distance was in the geographical space. Further smoothing was employed using moving averages twice with windows sizes of 11 km x 11 km and 25 km x 25 km.

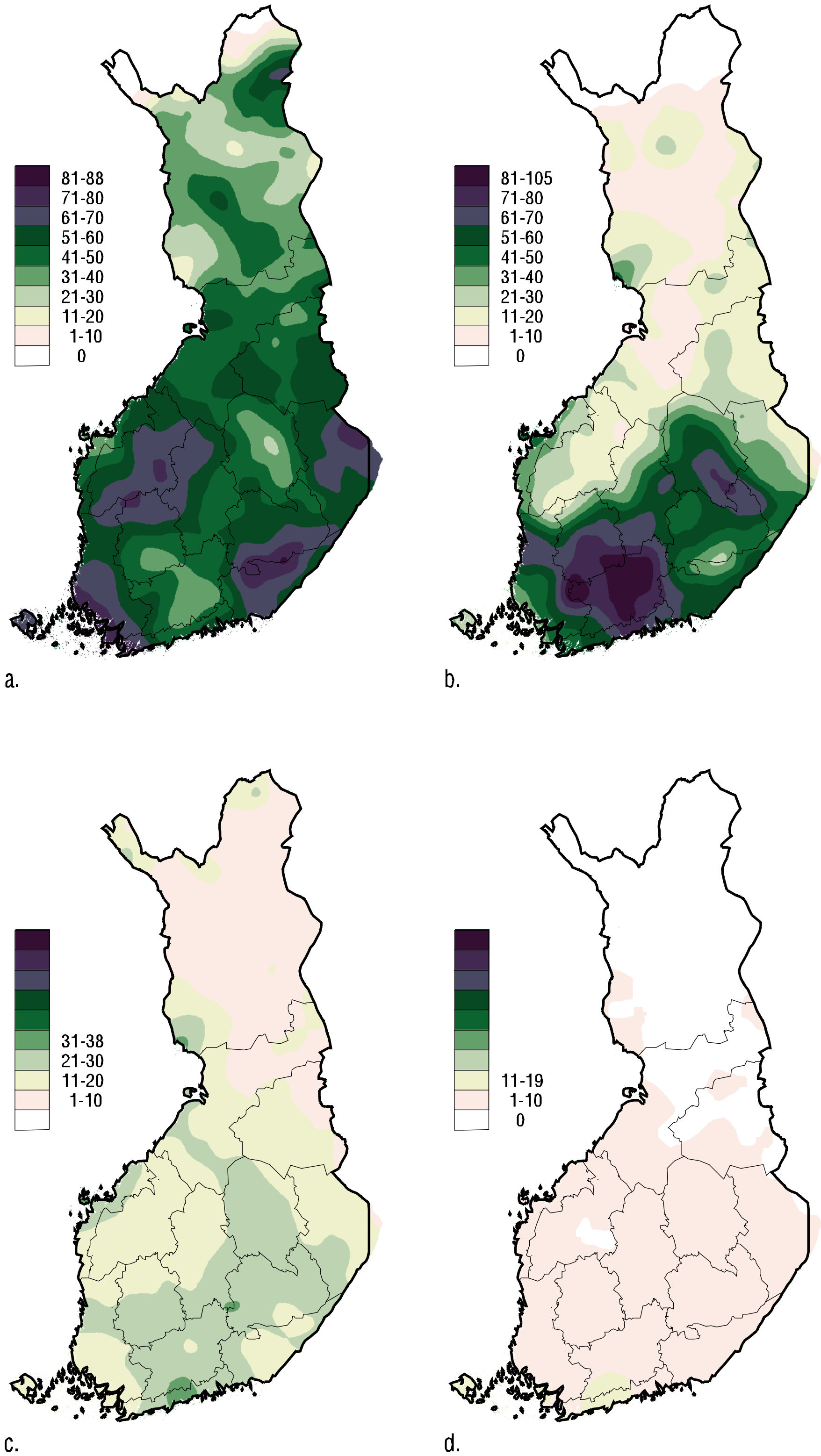


Figure 2.5: Large scale variation of mean volumes (m^3/ha) of pine (a), spruce (b), birch (c), and other tree species (d) on Future Potential Production Forest Land (FPPF) in NFI10 and NFI11 (2006–2010) with boundaries of the Forest Centre regions. Digital map data: ©National Land Survey of Finland, licence MML/VIR/MYY/328/08.

3 Methods

3.1 Image rectification and radiometric correction of the spectral values

The Landsat images were obtained from the USGS archive Landsat (2015) and they were rectified to the UTM coordinate systems using the WGS84 datum. The UTM zone for each frame depended on the location. The images arrived in zone 34, 35, or 36 and the pixel size was 30 meters (15 meters for the panchromatic channel). The MS-NFI processing requires images in the ETRS-TM35FIN coordinate system with pixel size of 16 meters. The rectification accuracy within Finland was not good enough to enable simple reprojection. Because of this, rectification based on control points was used for the Landsat images. The rectification accuracy for the Sentinel images was good enough and no control points were used with that material. The original pixels (10 m and 20 m) were resampled to the grid used in MS-NFI using cubic convolution resampling.

Point-type objects (e.g., small islands) were identified on both the satellite images and the basic maps and a regression model was fitted to their image coordinates and map coordinates. First or second order polynomial regression models were usually employed for this purpose. A typical number of control points has been around 50. The cubic convolution method was applied to re-sampling of the images to pixel size of 16 m by 16 m. No atmospheric correction was performed.

Areas corresponding to the cloud-free parts of satellite images are used in operative applications. A cloud mask is manually made for each satellite image. Automatic methods for this task have been tested, but no satisfactory method has been found so far. Note that cloud masking can be made very coarsely in those parts of the images in the image mosaic that are not needed for the final product (i.e., covered by another, better image).

The slope and aspect of the terrain locally change the illumination conditions of the surface and affect the reflectance from the ground and vegetation, as well as the radiance received by an imaging instrument. A digital elevation model was employed to remove the variation of the spectral values caused by the changes in the slope and aspect of the terrain. The details of the method are given in (Tomppo 1992, Tomppo et al. 2008b). The parameter selection has been studied in Katila and Tomppo (2001).

3.2 Preparation and updating of the field data

3.2.1 Canopy cover

The canopy cover was measured at field plots in NFI10, but the measurement was not continued in NFI11. In order to provide multi-source estimates for canopy cover, it was estimated for NFI11 and NFI12 field plots using NFI10 data. The models presented by Korhonen et al. (2007) were tested using NFI10 field plots. The result was that the models could not reproduce the measured values when the canopy cover was very low or very high. Because of this, a new method for estimating the canopy cover was designed for MS-NFI-2009 and has been used since then.

The design goal was to make a method that can reproduce the values measured in NFI10 as well as possible. Because the NFI10 field data were extensive, the k-NN method was used. The input variables and weights were manually determined. This was done separately for forest land, poorly productive land, and unproductive land because the set of available field measurements was different. For forest land and poorly productive land, the estimation was done separately according to the dominant species (pine, spruce, others).

The canopy cover for deciduous trees was computed from the canopy cover according to the proportion of deciduous trees in the field plot (computed from stem counts in seedling stands and basal area in mature stands).

3.2.2 Overview of updating

The updating method has not changed from MS-NFI-2013. For completeness, the method is explained also in this report.

The satellite images are from a different date than the field work. This means that there may be large differences between the state of the plot at time of field work and at the time of imaging. An example is clear-cut between the dates. In these cases the image data and field data should not be used as such because of the incompatibility of the field data and image data. This incompatibility would increase the estimation and prediction errors.

In the first applications, we solved this problem by omitting these plots from the field data. This, however, changes the distribution of the field data and this tends to reflect in the prediction results. A typical case is clear-cut after the measurement date of the field data and before the imaging date. In this case, a high-volume plot is removed from the field data lowering the average volume of the prediction results.

Another problem is that the field data have been collected during five years wherefore the average date of the field data is about two years before the end of the field data collection. This means that the prediction results reflect in this case (field data one year after the target date was included), on average, the time one year before the images. Within each image in the image mosaic, the spatial distributions of the forest variables depend primarily on the image data and, because of this, correspond primarily to the imaging date.

To solve the problems mentioned above, we decided to use a conservative partial updating controlled by the NFI field data and satellite images. The algorithm was designed to match the total volume after updating to the estimate computed from the field data only (Subsection 3.2.3). The data of individual field plots were modified according to the satellite image data and, in some cases, aerial photographs. If the plot was not cut between the field work and imaging, the growth models were applied to the growing stock variables. If the plot was identified as cut after the field work, the forest variables were modified according to the cut type identified from the image data and the growth models were applied to this modified data. It was not possible to detect thinnings from image data and these cases are included into the growth models. The growth models were controlled so that the total volume after updating matched the target.

The satellite images are from different dates and even different years. This means that the correction for the cuttings is different for different processing windows. Because of this, the updating was performed separately for each processing window.

3.2.3 Updating of the field plot data

The updating target The simplest updating target would be the mean of the total volume of the whole field data set. However, looking at the data from different years within the processing windows shows that, in many cases, this average does not represent well the total volume at the target time of updating. One reason for this may be an increasing or decreasing trend visible in the volumes seen in the field observations within the processing window. The yearly averages within windows fluctuate so that using a single year average would not be a good solution. Because of this, we decided to fit a regression line to the averages of the five years of field data.

The predictions computed with regression were visually compared to the averages computed from the field data for each year. The predictions looked reasonable in all cases in this work. However, the method may need to be refined later if cases will be found where the prediction does not look reasonable.

Large changes The large changes at the field plots between the image data and the field data are mostly due to regeneration cuts, but include also, e.g., severe windfalls. These changes can't be detected with models, but can, in most cases, be identified by examining visually the field plot data and available image data. The field plot data can also be reasonably modified to reflect the state at the image date if the field work has been done before the image date. If the change has occurred after image date but before the field work date, the field data can't be modified. In these cases, the plots were removed from the training material.

The changes where updating the data is possible were handled in the following way. First, the field plots were listed where these kind of changes potentially occur. These plots included advanced thinning stands and mature stands, together with young thinning stands where total volume was at least 100 m^3/ha. The image data at these candidate plots was matched against image data from plots that were cut recently according to field data and the candidate plots were ordered according to decreasing similarity. The candidate plots were then visually checked using the satellite data and, if available, recent aerial photographs. The plots where image data did not visually match what was expected from the field plot data were selected for modification. The selected plots were classified to plots where some trees were left (natural regeneration cut), and plots with no trees (clear cut).

The field plot data were changed according to the status of the field plot in the visual inspection. All of the substands were combined in to one stand. This was because it was not possible in practice to reliably identify different changes for the different substands. The updated forest variables reflected partly the previous state of the centre point stand. The dominant tree species was retained and no other species was assumed to survive. In case of clear cut, the volumes, basal area, mean height, mean diameter, mean age, and tree cover were zeroed. In case of natural regeneration cut, the changes were more complicated. The mean height, mean diameter and mean age were left intact. The total volume was set to the mean of storeys of this type in the field data in this geographic regions (30 m^3/ha in Southern Finland and 20 m^3/ha elsewhere). The other volumes, basal area, canopy cover and biomasses where changed according to the change in total volume.

The main changes to field data were:

development class: temporarily unstocked regeneration stand for plots with no trees, seedling tree or shelter tree stand for plots with some trees, randomised according to the ratio between seed tree and shelter tree cuts in NFI data from 2007–2012

cut type: regeneration cut for artificial regeneration for plots without trees, regeneration cut for natural regeneration for plots with some trees

field work date: midpoint between original field work date and image date, according to the growing season definition (see Subsection Growth models)

dominant tree species: not changed, except when temporarily unstocked stand was changed to young seedling stand

mean age, mean height, mean diameter: set to zero if no trees, otherwise unchanged

total volume: zero for plots without trees, 30 m^3/ha (Southern Finland) or 20 m^3/ha (otherwise) for plots with some trees

second storey volume; zero

other volumes, basal area, canopy cover: zero for plots without trees, otherwise put all to dominant species (canopy cover for broad-leaved trees set only if dominant species is broad-leaved)

biomasses: dominant species modified according to volume change, others set to zero

Growth models Some key plot level (sub-plot level) and stand level variables were updated using growth models, in addition to cutting and natural mortality assessments (see Subsection 3.2.1). The models were applied to each plot part and sub-plot stand separately when a plot intersected several stands (Subsection 3.3). The data were updated to the date 31 July 2015, independently of the date of the image acquisition. Either existing growth models or own models, derived for this purpose, were used to estimate the increment from the date of the field measurements to 31 July 2015. The variables updated with the increments were the plot level mean volumes (m^3/ha) by tree species groups and timber assortments, plot level biomasses by the tree species groups and tree compartments, canopy cover of all trees and separately for broad-leaved trees, as well as stand level variables, mean diameter, mean height and mean age and basal area of trees, as defined in the NFI. The increments of the stand level variables were estimated separately for the dominant tree storey and a possible second storey. The basal area is recorded in NFI also for all tree storeys. The development class of stand was checked and updated based on the changes in the growing stock.

The phases in the increment estimations and models were as follows. The length of the increment period in days was calculated first for each field plot and was defined as the number of the days in the growing season between the date of 31 July 2015 and the date of the field measurement. It was assumed that the growing season starts on 1 May and ends on 10 August. The number of the days in the full season is thus 102. The number of the days in the increment period was changed to the number of growing seasons ($nyear$) by dividing it by 102.

The plot level volume increments were estimated using the stand level models by Nyyssönen and Mielikäinen (1987) for pine and spruce dominated tree layers. The models of pine dominated forests were used also for the broad-leaved dominated tree layers. The models were thus employed by tree storeys. The volumes by tree storeys at plot level were not available in the data. They were estimated as the shares of the total plot level mean volume, the shares proportional to the quantity $i_G i_H$ were i_G is the basal area and i_H the mean height of tree storey i. The increased volumes by tree storeys were combined to the plot level volumes.

The model had been estimated for the natural logarithm of the percentage of the volume increment ($log(p_v)$). The model for a pine dominated stand is

$$log(p_v) = a + b_1 log(T)^2 + b_2 V^{1/V^{0.3}} + b_3 log(D)^8/10000 + b_4 I_{\{sf \leq 4\}}(sf) \quad (3.1)$$

where T and D are the age and mean diameter of the tree storey in question in a stand, V the volume of the tree storey on a plot and $I_{\{sf \leq 4\}}(sf)$ the indicator function of the site fertility class (sf). The values of the parameters are: a=0.7702, b_1=-0.09667, b_2=1.2503, b_3=-0.1796 and b_4=0.1817. If D was zero or missing, the parameters of the model had been estimated without D, and are: a=0.7632, b_1=-0.1181, b_2=1.3516 and b_4=0.09116.

The model for the percentage of the volume increment of a spruce dominated stand is (Nyyssönen and Mielikäinen 1987)

$$log(p_v) = a + b_1 log(T) + b_2 log(V) + b_3(log(T)log(V))^2 + b_4 log(T)V^2/100000+ \\ b_5(log(D))^5 + b_6 I_{\{sf\leq 2\}}(sf) \quad (3.2)$$

The values of the parameters are: a=8.839, b_1=-1.2749, b_2=-0.5948, b_3=0.00309, b_4=-0.1193, b_5=-0.0006095 and b_6=0.1009. If D was zero or missing, the parameters of the model had been estimated without D, and are: a=9.7669, b_1=1.5813, b_2=-0.5730,b_3= 0.003315 and b_4=-0.1177, b_6=0 (site fertility indicator was missing from the model).

The increased volume of the tree storey in the end of the updating period, including the estimated increment over the updating period, was $V_2 = V_1(1 + p_v/100)^{nyear}$ where V_1 and V_2 are the volume of the tree storey in the beginning and in the end of the period, p_v the increment percentage form the model and $nyear$ as above, the number of growing seasons in years. The increased volume V_u for a plot was the sum of the increased volumes of the dominant tree storey, the increased volume of a possible second storey and the original volume of a possible third tree storey. The third tree storey is quite uncommon and the significance of its possible volume to the total volume negligible. The ratio V_u/V_o was used as a factor to calculate the volumes by tree species groups and by timber assortments. Here V_o is the original plot level volume.

The same ratio V_u/V_o was used also to increase the variables canopy cover of trees and canopy cover of broad-leaved trees as well as biomasses by the tree species groups and tree compartments.

For the increment estimates of the other key variables, except the age of a tree storey of a stand, new models were derived using the permanent field plot data of NFI10 and NFI11. The age was updated simply increasing the the assessed age by the number of the growing seasons. The mean diameter D, mean height H and basal area G of a stand were updated also by tree storeys. The model for the logarithm of the relative diameter increment $log(i_D/D)$ was

$$log(i_D/D) = a + b_1 D + b_2 log(T) + b_3 log(T)^2 + b_4 G + b_5 dd + b_6 I_{\{SP=2\}}(SP)+ \\ b_7 I_{\{SP=3\}}(SP) \quad (3.3)$$

where i_D is the annual mean diameter increment of the tree storey in question calculated from the successive measurements of field plot stands of NFI data, G is the basal area of the trees of the tree storey in question, dd the effective temperature sum, $I_{\{SP=i\}}(i)$, the indicator function of tree species groups i, and the other variables as in Eqs. 3.1 and 3.2. The tree species groups were, pine and other coniferous than spruce (1), spruce (2) and broad-leaved species The values of the parameters were a=-0.8152130, b_1=-0.0655599, b_2=-1.0520602, b_3=0.1317468, b_4=-0.0042745, b_5=0.0003396, b_6=0.0091463 and b_7=0.0073079.

The model for the logarithm of the relative height increment $log(i_H/H)$ was

$$log(i_H/H) = a + b_1 H + b_2 log(T) + b_3 log(T)^2 + b_4 dd \quad (3.4)$$

where i_H is the annual mean height increment of the tree storey in question calculated from the successive measurements of field plot stands of NFI data and the other variables as in Eqs. 3.1, 3.2 and 3.3. The values of the parameters were a=-0.6601906, b_1=-0.0101824, b_2=-1.0175697, b_3=0.0890791 and b_4=0.0007121.

The model for the logarithm of the relative increment of the basal area of the trees of the storey in question$log(i_G/G)$ was

$$log(i_G/G) = a + b_1 G + b_2 log(T) + b_3 log(T)^2 + b_4 dd + b_6 I_{\{SP=1\}}(SP) + b_7 I_{\{SP=2\}}(SP) + b_7 I_{\{SP=3\}}(SP) \quad (3.5)$$

where i_G is the annual basal area increment calculated from the successive measurements of field plot stands of NFI data and the other variables as in Eqs. 3.1, 3.2, 3.3 and 3.4. The values of the parameters were a= 1.4523612 b_1=-0.0753081 b_2=-1.8919602 b_3=0.2124715 b_4=0.0009454 b_5=-0.0186978 b_6=-0.0041331 b_7=-0.0129012 The model was applied to the basal areas of the stands by the tree storeys.

Only the dominant tree storey and a possible second tree storey were updated using the increment models of the diameter, height and basal area. A possible third storey was not updated.

The values of variables, D, H and G by tree storeys on 31 July 2015 were calculated in a same way as the volumes, that is, $M_2 = M_1(1 + p_r)^{nyear}$ where M_1 and M_2 are the value of the variable M in the beginning and in the end of the period, p_r the relative increment from the models and $nyear$ as above, the number of growing seasons in years. D and H are given in NFI data only by tree storeys, but G also for all storeys together. The total basal area of the stands was the sum of the basal areas of the tree storeys.

The mean age was adjusted according to the years between the field data date and 2015.

Calibration to the updating target The growth model results were calibrated to the updating target by multiplying the growing time with a multiplier. This multiplier was determined iteratively by computing the total volume (m^3) after the increment prediction and comparing the result to the updating target determined for the corresponding image.

Updating of development class The following possible changes in the development class were considered from the date of the field measurement to the updating date (July 31, 2015):
1) from temporarily unstocked regeneration stand to young seedling stand
2) from shelter tree or seed tree stand to open area or young seedling stand
3) from young seedling stand to advanced seedling stand
4) from advanced seedling stand to young thinning stand
5) from young thinning stand to advanced thinning stand
6) from advanced thinning stand to mature stand.

For updating purposes, mainly to judge the development class of young seedling stand versus open area, we calculated the following four-dimensional distribution F
i) cutting time
ii) development class
iii) the effective temperature sum (three classes, –1049, 1050–1179, 1180–)
iv) site fertility.
Only one effective temperature sum class was used for site fertility class one due to the lack of the data.

The means and standard deviations of the mean diameter of tree storeys of stands by development classes were calculated for making decisions concerning possible transitions.

The transition frequencies from regeneration cutting, both artificial and natural, to the development class young seedling stand were estimated from the NFI data as a function of cutting time, effective temperature sum and site fertility class. The transitions were simulated based on the distribution F.

The possible new development classes, in case of an open and temporarily unstocked regeneration stand for artificial regeneration, were temporarily unstocked regeneration stand and young seedling stand. This rule was used due to the short updating period. The longest updating period was four years.

The distribution of the dominant tree species by site fertility classes in young seedling stands were estimated from the NFI data from the years 2004–2012. The dominant tree species was selected from this distribution in case of transition from a temporarily unstocked regeneration stand to young seedling stand.

For natural regeneration (open area), the possible new development classes were, the one on the date of field measurement (shelter or seedling tree stand), open area and young seedling stand. The dominant tree species for a possible young seedling stand was the one of the shelter tree / seedling tree stand.

Dominant tree species, mean diameter, mean height and age remained / were changed to those corresponding an temporarily unstocked regeneration stand (0) if the result was a temporarily unstocked regeneration stand.

To update case 3), a possible transition from young seedling stand to advanced seedling stand, the updated mean height was first checked using the original height and the height increment (Eq. 3.4). A simple model was derived to estimate the mean diameter as a function of the height.

Similarly, a simple model was estimated for the volume (m^3/ha) as a function of the mean height.

The dominant tree species remained as the same as in the date of the field measurements. The biomass estimates were updated respectively.

A possible change from the development class advanced seedling stand to young thinning stand was done as follows: the development class was 'up-graded' if the updated mean diameter of a stand exceeded the average mean diameter of the young thinning stands by two standard deviations of the mean diameters of those stands. A possible change from the development class young thinning to an advanced thinning stand was done similarly, as well as a possible change from an advanced thinning stand to a mature stand except in the latter case, the standard deviation of the development class mature stand was used, based on practical tests.

3.3 Preparation of the input data sets

In the image analysis (Fig. 3.1), the input data sets were 1) ground truth data, i.e., one record for each plot part and stand corresponding to a centre point of a plot, called here centre point stand and also for stands intersecting other parts of a plot, called here sub-plot stand: 1a) field data and 1b) satellite image data, 1c) digital map data, 1d) and other numeric feature data in text format, 2) a pre-processed satellite image, 3) a digital map of land use classes and mire and open bog mask, 4) a digital elevation model and thereof derived image of the angle between terrain normal and sun illumination, 5) cloud and cloud shadow delineation mask, 6) large-area forest resource data and 7) a map of computation units to calculate small-area estimates (Fig. 3.2).

The land class map was employed to distinguish the combined forest land, poorly productive forest land and unproductive forest land (MS-NFI forestry land) from the other land classes. In this analysis, all the

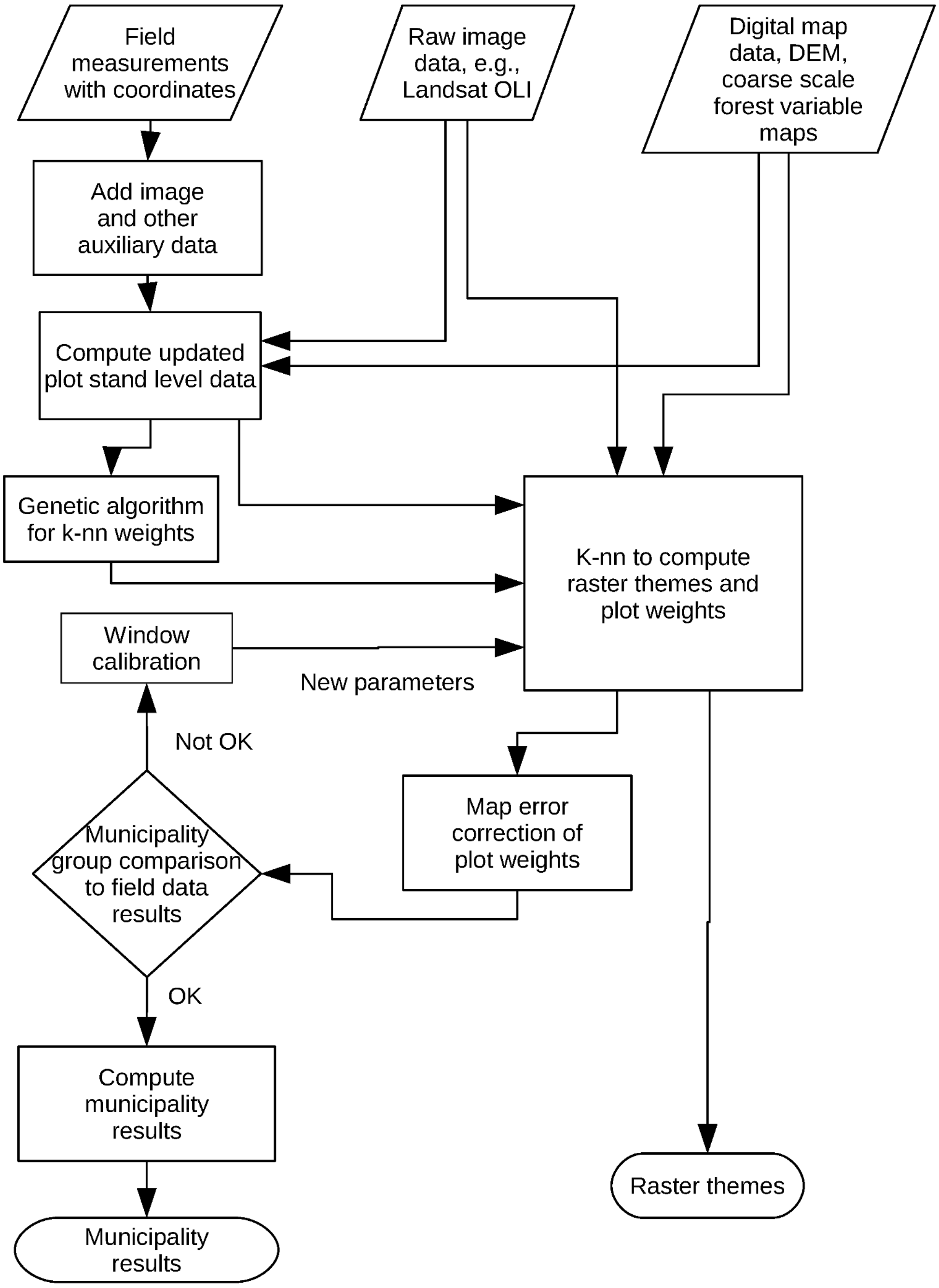
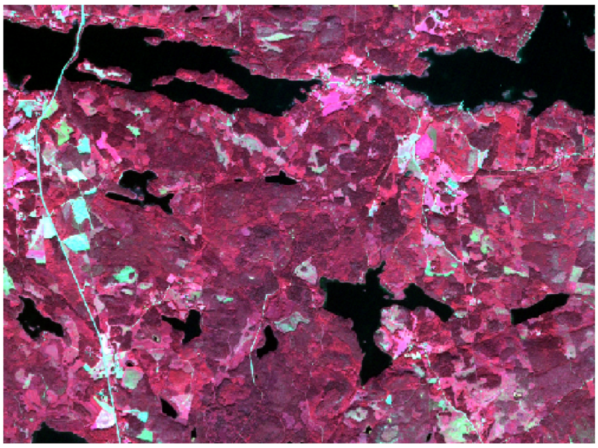


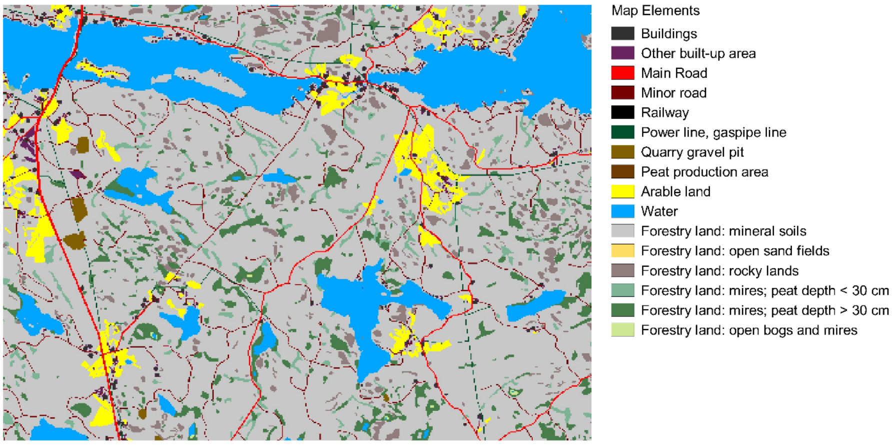
Figure 3.1: Data flow and computational scheme for multisource NFI.

area that was not peat production area, built-up land, arable land, roads or waters in the numerical map was considered forestry land. The numerical map data were not always up-to-date and could contain significant errors. The effect of the map errors on the estimates were corrected the using calibration method (Section 3.3.2).

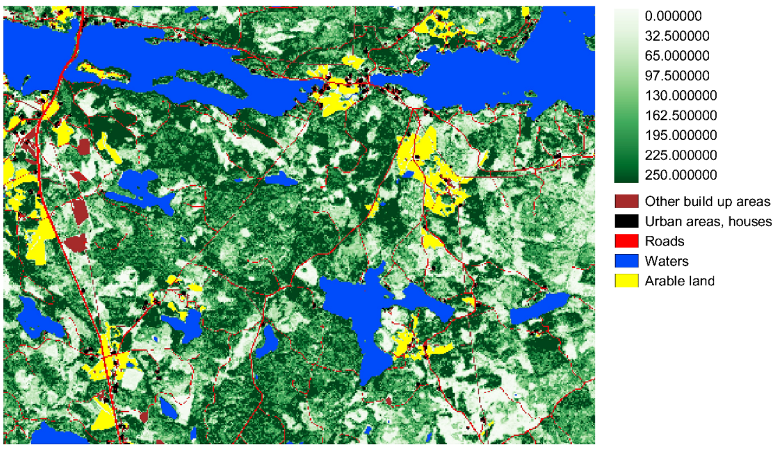
The field data used as training data was sampled, if possible, from a larger area in the image window than is used in the final result mosaic. Field data within the finally included areas was used for testing.



a



b



c

Figure 3.2: Examples of the Sentinel-2A MSI satellite image, multichannel colour composition of channels 2, 3, and 4 (a); the elements of the topographic map database (b); and MS-NFI-2015 map of total volume (m^3/ha) with other land use map data (c). Digital map data: contains data from the National Land Survey of Finland topographic database 4/2016.

3.4 The estimation methods

3.4.1 The improved k-NN method (ik-NN)

The non-parametric k-NN estimation has been employed in the MS-NFI calculation since operational inception in 1990. The method has been improved continuously. The details of the current method employed for this article are given in (Tomppo and Halme 2004, Tomppo et al. 2008b, 2012, 2013). The basic principles are listed here.

With the k-NN method, the plot weights (Eq. 3.9) (not equal for each plot) are computed for each plot by computation units (Tomppo 1996). The computation unit is the unit for which results are computed. It can be a municipality when computing municipality results or a pixel when computing raster maps. The weights are computed for each field sample plot $i \in F$, where F is the set of field plot parts, centre point plot part or sub-plot part, where the centre point of the plot belongs to forestry land. These plot weights are sums of the weights that are computed for the field plots over all satellite image pixels on the forestry land mask of the computation unit. The plot weights corresponding to a single pixel (Eq. 3.7), in turn, are computed by a non-parametric k-NN estimation method (Tomppo 1991, 1996, Tomppo et al. 2008b, 2012, 2013). The method utilises the distance metric d, defined in the current version in the feature space of the satellite image data and coarse scale forest variables. The k nearest field plot pixels p_i, in terms of d, i.e., pixels that cover the centre of a field plot $i \in F$, are sought for each pixel p under the forestry land mask of the cloud free satellite image area. Note that the plot parts belonging to non-FRYL land categories are removed from the data set. The sum of the weights of the rest of the plot parts is scaled to one for each pixel. A maximum geographical distance is employed, if necessary, in order to avoid selecting the nearest plots (spectrally similar plots) from a region in which the response of image variables to field variables is not equal to that of the pixel under consideration. This is due to, e.g., changing atmospheric conditions or a large image frame. The feasible set of nearest neighbours for pixel p is thus

$$\{p_i | d^{(x,y)}_{p,p_i} \leq d^{(x,y)}_{\mathrm{max}}, d^z_{p,p_i} \leq d^z_{max}, R(p_i) = R(p)\} \tag{3.6}$$

where $d^{(x,y)}_{p,p_i}$ is the geographical horizontal distance from pixel p to pixel p_i, d^z is the distance in the vertical direction, $d^{(x,y)}_{max}$ and d^z_{max} are their maximum allowed values, and $R(p)$ is the indicator function of land class on the basis of map data (Tomppo 1990, 1991, 1996, 2006b, Katila et al. 2000, Katila and Tomppo 2001, Tomppo et al. 2008b, 2012, 2013).

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

$$\begin{aligned} w_{i,p} &= \frac{1}{d^t_{p_i,p}} \bigg/ \sum_{j \in \{i_1(p),\ldots,i_k(p)\}} \frac{1}{d^t_{p_j,p}}, \quad \text{if and only if } i \in \{i_1(p), \ldots, i_k(p)\} \\ &= 0 \quad \text{otherwise.} \end{aligned} \tag{3.7}$$

The distance weighting power t is a real number, usually $t \in [0, 2]$. In case the distance is zero for one or more plots, all weight is given to those plots. The distance metric d employed was

$$d^2_{p_j,p} = \sum_{l=1}^{n_f} {\omega_{l,f}}^2 (f_{l,p_j} - f_{l,p})^2 + \sum_{l=1}^{n_g} {\omega_{l,g}}^2 (g_{l,p_j} - g_{l,p})^2 \tag{3.8}$$

where $f_{l,p}$ is the lth feature computed from the spectral bands of the pixel data. The features are normalised using the digital elevation model, when applicable. $f_{l,p_j} = f^0_{l,p_j} / \cos^r(\alpha)$, with α the angle between sun illumination and terrain normal, r the user given power for the cosine correction, $g_{l,p}$ the large area prediction of the lth applied forest variable, n_f the number of image variables (or features), n_g the number of coarse scale forest variables and $\omega_{l,f}$ and $\omega_{l,g}$ the weights for image features and coarse scale forest variables respectively.

The values of the weights $\omega_{l,f}$ and $\omega_{l,g}$ are computed by means of a genetic algorithm (Tomppo and Halme 2004, Tomppo et al. 2008b, 2012, 2013).

A pixel size of 1 km by 1 km is used in the coarse scale forest variable predictions $g_{l,p}$. The first phase of the improved version of k-NN, ik-NN, is to run the optimisation algorithm by strata, e.g., mineral soil stratum and mire and bog stratum. The estimation after that is similar to the basic k-NN estimation.

For computing forest parameter estimates for computation units, sums of field plot weights to pixels, $w_{i,p}$ are calculated by computation units, for example, by municipalities, and by map stratum h over the pixels belonging to the unit u. An example of a stratum could be mineral soil forestry land. The weight of the sub-plot i_l of plot i in forest stratum l and in map stratum h to computation unit u is denoted

$$c_{i_l,h,u} = aq_{i_l} \sum_{p \in u_h} w_{i,p} \tag{3.9}$$

where u_h is the set of the pixels in the map stratum h, a is the pixel size and q_{i_l} is the share of the field plot i belonging to the forest stratum l and map stratum h on forestry land.

Reduced weight sums $c^r_{i_l,h,u}$ are obtained from the formula 3.9, if clouds or their shadows cover a part of the area of the computation unit u. The real weight sum for plot i is obtained expanding the weight (3.9) by the ratio forestry land divided by the forestry land not covered by the clouds in each computation unit.

The weights (3.9) are computed within forestry land separately for mineral soil stratum and peatland strata. The weights are also computed for other land classes, arable land, built-up land, roads and waters using the plots falling in the corresponding stratum if the stratification based map correction method is employed Katila and Tomppo (2002), and plots falling into forestry land map stratum if the calibration method is used (Katila et al. 2000).

After the final field plot weights to computation units ($c_{i_l,h,u}$) have been calculated, the ratio estimation is employed to obtain the small-area estimates (e.g., Cochran (1977)). In this way, the estimation is similar to that using field plot data only (Tomppo 2006a, Tomppo et al. 2008b, 2012, 2013)

$$\hat{y}_{u,l} = \frac{\sum_h \sum_{i \in I_{l,h}} c_{i_l,h,u} y_{i_l}}{\sum_h \sum_{i \in I_{l,h}} c_{i_l,h,u}}, \tag{3.10}$$

where y_{i_l} is the volume per hectare for plot i for the part(s) belonging to field plot sub-class l, and $I_{l,h}$ if the set of field plots belonging to field sub-class l and map stratum h.

Predictions of some forest variables are written in the form of a digital map during the procedure. The land classes outside forestry land are transferred to map form predictions directly from the digital map file. Within forestry land mask, the variables are predicted by the weighted averages of the k nearest neighbours (Tomppo 1991, 1996).

A pixel-level prediction of variable Y for pixel p is defined as

$$\tilde{y}_p = \frac{\sum_{i \in I_h} w_{i,p} \sum_{j \in F_i} a_{i_j} y_{i_j}}{\sum_{i \in I_h} w_{i,p} \sum_{j \in F_i} a_{i_j}}, \tag{3.11}$$

Table 3.1: The ik-NN estimation parameters employed in MS-NFI-2015.

Parameter	Choice
Variables applied in the distance metric	Illumination corrected spectral values for satellite image bands and large area forest variable estimates
Distance metric	Weighted Euclidean distance
Value of k	3–5
Weights attached to the nearest neighbours	Weights proportional to the inverse distance (t=1)
Restrictions for search of nearest neighbours	Large area forest maps are used to direct the NN selection. In addition, the geographic distance between the pixel being processed and the acceptable reference plots was limited.

where y_{i_j} is the value of the forest variable Y on stand j of plot i, I_h the set of the field plots belonging to map stratum h, a_{i_j} the share of the stand j of plot i and F_i the set of the plot parts of plot i where a value has been defined for the variable.

The mode or median value can be used instead of the weighted average for categorical variables. Mode has turned out to work better than median in the practical tests (Tomppo et al. 2009b).

One special detail of the Finnish NFI is that some stand level variables are not recorded in the field for the plot parts not including a centre point of a plot in case there are no tally trees belonging those plot parts. The reason is that the area estimates are based on the numbers of the centre points while volumes are summed up from all tally trees in the stratum in question. The variables not recorded for the sub-plots without a centre point and without any tally tree are for example land class based on the FAO classification, main site class, site fertility class, stand age, mean diameter of stand, mean height of stand, stand basal area, canopy cover of trees and canopy cover of broad-leaved trees.

This fact is taken into account in the municipality level estimates in such a way that the missing value is imputed from the distribution of the variable in question when the distribution is calculated from a similar forest stratum. In pixel level predictions, those plot parts don't have values for these variables, that is, they are not included in set F_i of Eq. 3.11.

The predicted variables in a map form are usually land class, main site class, site fertility class, stand age, mean diameter of stand, mean height of stand, stand basal area, canopy cover of trees, canopy cover of broad-leaved trees and volumes by tree species (pine, spruce, birch, other broad leaved trees) and by timber assortment classes as well as biomass by tree species groups and tree compartments. The total number of the maps in MS-NFI-2015 and MS-NFI-OA-2015 was therefore 44 (Table 4.6).

3.4.2 Taking into account the plot representative areas in ik-NN

For best results with a NN classifier, the number of prototypes for each class should match the a priori probabilities of the classes Davies (1988). Similar argumentation can be applied also to k-NN estimation.

If the auxiliary variables don't provide any information about the target variable, the distribution of the k-NN estimates matches the distribution of the learning data. If the information is perfect, then the distribution of estimates matches the distribution of that variable in the population where estimates are produced. In practice, the situation is somewhere between these extremes.

The distributions matter if, e.g., one is computing an average of a variable over an area. The estimates are computed for each pixel in the area. If the distribution of the estimates matches the distribution of the variable in the area of interest, the resulting average will be unbiased. If the distributions don't match, the average may be biased. This problem is present, e.g., if stratified sampling is used within an inventory region. In northern Lapland (see Section 2.1), the inclusion probabilities of the sample units (plots) of the NFI vary significantly between the strata within the inventory region. In NFI11, in Åland two clearly different sampling densities were used in different areas. Otherwise the differences between sampling densities in adjacent inventory regions are small.

The sampling density can be modelled using the *representative area* of each plot (Tomppo et al. 2011a, Eq. 3.3). This is computed by dividing the sampled area A_c of sampling region c by the number of sample plots n_c. Let's assume that the value of interest in sample plot i is y_i and the representative area of the sample plot is a_{c_i}. The estimate of the average of the variable $\bar{y}$ over an area $A \cup a_i$ can be computed from the samples within the area with:

$$\bar{y} = \frac{\sum_{i \in A} a_{c_i} y_i}{\sum_{i \in A} a_{c_i}} \tag{3.12}$$

Inspired by the previous equation, computation of the weights (Eq. 3.7) is modified to take into account the representative are a_{c_i}:

$$\begin{aligned} w_{i,p} &= \frac{a_{c_i}}{d^t_{p_i,p}} \Bigg/ \sum_{j \in \{i_1(p),\dots,i_k(p)\}} \frac{a_{c_i}}{d^t_{p_j,p}}, \quad \text{if and only if } i \in \{i_1(p),\dots,i_k(p)\} \\ &= 0 \quad \text{otherwise.} \end{aligned} \tag{3.13}$$

This approach does not completely solve the problem because it only modifies the computed results on the condition that the neighbours have been selected without taking into account the representative areas of the sample plots. An example of the effect of this modification is in Subsection 4.3.

3.4.3 Selecting estimation parameters and their values for k-NN

The basic principle of k-NN estimation is straightforward. However, practice has shown that the predictions and estimation errors depend to a large extent on the core estimation parameters of the k-NN algorithm. These are:

1. the variables employed in the distance metric, spectral bands or their transformations, possible correction for variation in illumination angle of the pixel based on elevation variation (slope, aspect) (Tomppo 1996)
2. the distance metric (Tomppo and Halme 2004)
3. the value of k (Katila and Tomppo 2001)
4. the weights to be attached to the nearest neighbours, e.g., even weights or functions of the used distance and powers (negative),
5. the variables employed in restricting the area from which the nearest neighbours are sought for a pixel, e.g., a geographical reference area (Katila and Tomppo 2001). In MS-NFI-2015, the country was divided into 19 processing windows. A geographical distance limit was used in the largest image windows to prevent use of plots from different atmospheric imaging conditions

6. the use of additional information, e.g., large area variation of forest variables in the distance metric (Tomppo and Halme 2004),

7. the use of ancillary data in the estimation, e.g., for stratification.

The parameters and their values in MS-NFI-2015 are given in Table 3.1. The parameters are selected separately for each image (consisting of one or more image frames, see Table 2.2). The criteria are the mean square error and bias of pixel level predictions using leave-one-out cross validation, and particularly, the difference between areal estimates based on i) multi-source inventory and ii) on the field data based estimates and their sampling errors (Tomppo et al. 2008b, 2012, 2013). The differences of the areal estimates are assessed in terms of sampling error based on the field data plots (e.g., Katila and Tomppo 2002, Tomppo and Halme 2004). The values of the parameters usually vary by image depending on, e.g., imaging conditions, number of available field plots and variability of forests. The selections are not independent. A change in one parameter affects the optimal value of the other parameter. A crucial factor concerning the accuracy of the estimates seems to be the performance of the genetic algorithm. It was slightly revised for MS-NFI-2009 as an aim to control the weights of the feature variables coming out from the algorithm.

3.4.4 Area and volume estimates for small areas – correction for map errors

In the multi-source estimation, numerical map data (see Sect. 2.3.2) are employed to decrease estimation errors. If the numerical map data would be error free, the computation unit weights (Eq. 3.9) could be calculated using pixels belonging to forestry land (according the map data) only. However, map data can be out-of-date, include location errors and does not correspond exactly to the definitions of NFI land classes (Table 4.1). Errors can also arise during the post-processing of map data. Two methods have been developed to reduce the effect of map errors on small-area multi-source forest resource estimates: a statistical calibration method (Katila et al. 2000, Katila 2006a) and a stratified k-NN method (Katila and Tomppo 2002).

The calibration method is based on the confusion matrix between land use classes of the field sample plots and corresponding map information. The bias in the land class or other total cover estimates, obtained, e.g., from remote sensing or map data, can be corrected by means of the error probabilities expressed as a confusion matrix (Czaplewski and Catts 1992, Walsh and Burk 1993), assuming that the employed field sample are based on a statistical sampling design (Card 1982).

The employed map strata are defined in such a way that each stratum is reasonably homogeneous with respect to the 'map errors' and the NFI land class distribution. This enables the use of the synthetic small-area estimation method when correcting map errors (Rao 2003). The method utilises the error and land class proportions that have been estimated from a larger region. The large regions of municipalities were formed in such a way that the map errors would be as homogeneous as possible within the regions and within each stratum (Tomppo et al. 2013). Seventeen regions were used for map correction (Fig. 3.3). In the MS-NFI-2015, NFI field plots from years 2012–2016 were used to compute the confusion matrix.

The method given in (Katila et al. 2000) was used to calculate the calibrated field plot weights. The calibration typically increases the mean volume estimates and decreases the FRYL area estimates for small areas, if FRYL is overrepresented on maps. Calibration was carried out by groups of municipalities. Despite the rather simple idea of the calibration, it is quite laborious when implemented in the MS-NFI.

The MS-NFI employs a topographic database for municipality boundaries, while the field inventory employs land and water areas from official statistics of the Finnish Land Survey (Suomen pinta-ala kunnittain

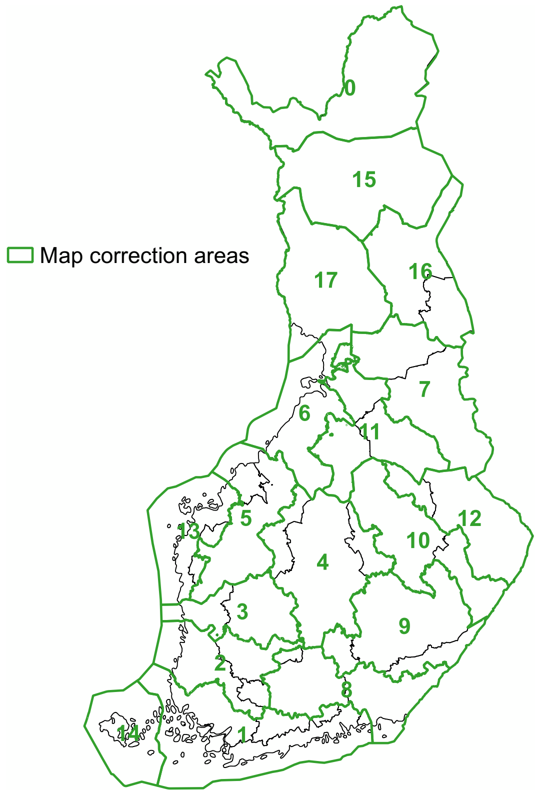


Figure 3.3: The large regions used for calculating the error probabilities between land classes and map strata for MS-NFI-2015, and the boundaries of the Forest Centre regions. Digital map data: National Land Survey of Finland, licence MML/VIR/MYY/328/08.

2011). The area information from the latter data source is more accurate and there are slight differences between the total and land areas of municipalities from these two data sources. Hence, after the correction of map errors, the MS-NFI municipality land areas are calibrated to the official land areas. The calibration coefficient is straightforward $A_{U,Land_{NLS}}/A_{U,Land}$ and this ratio is assumed to also hold for forestry land and the (calibrated or stratified) weights $c_{i,U}$ are multiplied by this coefficient. For the calibrated MS-NFI, the calibrated land area $A_{U,Land}$ must be first estimated (see Tomppo et al. 2008b, 2012, 2013). The calibration to the official land areas is valid only for (random) deviations between the two data sets and not for the case where real and significant boundary changes between municipalities have taken place in either of the two data sources.

3.4.5 Assessing the errors

Deriving an error estimator for an arbitrary group of pixels has proven to be a challenging task. The problem can be divided into the derivation of i) an error estimator for a pixel level prediction and ii) an error estimator for a parameter for an area of interest.

Difficulties arise because:

1. errors depend on the actual value of the variable to be predicted and so pixel-level errors are spatially dependent,
2. the variables measured or observed on the field plots are also spatially dependent,

3. the spectral values of adjacent pixels of a satellite image are dependent due to the atmospheric properties (scattering) and imaging technique.

Furthermore, several error sources make the error estimation complex. Examples of such error sources are given in (Tomppo et al. 2008b, 2012, 2013).

During the data processing phase in the Finnish MS-NFI, the pixel-level root mean square error (RMSE) and the pixel level average bias are calculated using leave-one-out cross-validation using the available field plots. This is also a part of the employed genetic algorithm and the selection of the estimation parameters of k-NN and ik-NN. For a sufficiently large area consisting of a group of pixels, e.g., for areas of 200 000–300 000 ha, the MS-NFI estimates are compared to the estimates and error estimates based solely on field data. Some empirical error estimates are also available for reliability assessments (Katila 2006b, Tomppo et al. 2008a,b, 2012, 2013, 2014). Some recent developments in error estimation, particularly in model-based error estimation are also presented in that publication, see also (Kim and Tomppo 2006, McRoberts and Tomppo 2007, McRoberts et al. 2011, McRoberts et al. 2007, Magnussen et al. 2009, Magnussen 2013, Magnussen and Tomppo 2016, Magnussen et al. 2016, Dash et al. 2015).

3.4.6 Calibration of the results within a processing window

The error assessment usually shows that, in spite of all tuning, the average values computed from MS-NFI do not quite match the values computed from field data. For instance, it is common that there are differences in volumes of tree species classes even if the total volumes are quite close. There are many possible reasons for the differences processing, e.g., the image characteristics, the image area in the mosaic, and the set of sample plots used in the specific processing window.

Because of this, an additional calibration step is added to the method. The purpose is to make the averages computed from several pixels to be closer to the level of field data for each processing window than they would be without this extra calibration step. The effect on the RMS errors is negligible. The method was not applied in three image windows where the cloud-free area was too small for reliable correction.

In MS-NFI-2015, the extra calibration is applied to the volumes, basal area, mean height, mean diameter, age, and the biomasses. The calibration may be applied also to other variables in future.

The correction is basically applied as follows. The average value of variable i in MS-NFI before calibration but including map error correction is denoted by $\overline{v}_{im}$ and the corresponding average from field data by $\overline{v}_{if}$. The calibrated MS-NFI result is the uncalibrated result multiplied by $\overline{b}^c_{if} = \overline{v}_{if}/\overline{v}_{im}$.

A slightly more complicated method is applied to the volumes because the volumes for tree species classes must add up to the total volume for each pixel. Two methods have been used to solve the problem. The recommended method consists the following steps:

1. Apply the multipliers individually to the tree species classes and compute the total volume for each pixel in the processing window as sum of the volumes in the tree species classes.

2. Compute the average total volume from the pixel-wise total volumes.

3. Adjust the multipliers for the tree species classes so that the total volume ratio matches the target.

4. Compute the new field data using the multipliers for the species classes. Compute the total volume for each plot as the sum of the volumes for the species classes.

In the method above, the effective correction factors for total volume vary from one plot to another. In one of the processing windows this effect was problematic. In this case, the adjustment for the multipliers for species classes was computed separately for each plot so that the corrected volumes for tree species classes add up to the total volume. In this method, the effective corrections factors for the species classes tend to be smaller than the original corrections factors.

4 Results

4.1 The accuracy of the MS-NFI-2015 land use map data

The accuracy of the MS-NFI-2015 land use map data is estimated by applying a confusion matrix between the map and field based land classes of the NFI field plot data. The NFI land classes were combined to five classes. The tables (Table 4.1) are presented separately for South Finland and North Finland (sampling regions 4–6). These tables are comparable to those estimated for the MS-NFI9 (Tomppo et al. 2008b, Table 2.11 and 2.12,). The confusion matrices were computed in South Finland using 54 765 sample plots and in North Finland using 23 637 sample plots. The percentage of correctly classified land classes were 93.4 % and 96.8 % for South Finland and North Finland, respectively. The ultimate goal in the MS-NFI is to separate the forestry land from other land use and in this context the percentage of correctly classified NFI plots were 94.5 % and 97.2 %, respectively.

All of these overall accuracy estimates are slightly higher than those in MS-NFI9. However, the overestimation of the forestry land has increased with the latest map data: the forestry land area was overestimated by 4.4 % and 1.7 % in South Finland and North Finland, respectively. The arable land element and the built up land mask underestimated the corresponding land use areas when compared to the NFI field plot data. The arable land in the NFI includes also waste land inside the land use class, small roads and buildings (other than houses) used for agriculture. We assume that the more up-to-date topographic data, the smaller pixel size and improvements in measuring the locations of the sample plots enabled the increase in accuracy compared to the results in Tomppo et al. (2008b).

4.2 Calibration of the results by species classes

The calibration method described in Section 3.4.6 is designed to bring the large area results computed either from the raster images or using the accumulated plot weights closer to the results computed from field data only. The differences between average volumes computed from MS-NFI-2015 and from field data are shown for two administrative regions and for the whole country in Table 4.2. The column significance shows the absolute value of the difference divided by the standard error computed from the field data. Field data from years 2014 - 2016 was used to make the target year 2015 be in the middle of the field data time range.

The difference in total volume and the volumes of the tree species classes are not zero in this comparison because of several reasons. The calibration was done separately for each processing window using the total volume multiplier computed for each window. The calibration is computed from the sampled

Table 4.1: The confusion matrix (percentages) of field based land class against the land use according to topographic database map data on NFI field plots in South Finland (a) and North Finland (b). The number of field plots in South Finland was 54 765 and in North Finland 23 637.

(a)

	Land use map, South Finland					
NFI field plot land class	Forestry land	Arable land	Built up land	Traffic & power lines	Water	Total
Forestry land	61.30	0.24	0.12	0.80	0.23	62.68
Arable land	1.17	10.46	0.12	0.24	0.02	12.01
Built up land	1.76	0.10	2.48	0.33	0.07	4.74
Traffic and power lines	0.80	0.07	0.05	1.42	0.00	2.36
Water	0.40	0.02	0.04	0.01	17.74	18.21
Total	65.44	10.88	2.81	2.81	18.06	100.00

(b)

	Land use map, North Finland					
NFI field plot land class	Forestry land	Arable land	Built up land	Traffic & power lines	Water	Total
Forestry land	85.11	0.10	0.04	0.38	0.16	85.79
Arable land	0.43	2.72	0.02	0.05	0.00	3.22
Built up land	0.66	0.07	0.82	0.05	0.01	1.62
Traffic and power lines	0.67	0.05	0.03	0.81	0.01	1.58
Water	0.40	0.01	0.00	0.00	7.38	7.79
Total	87.27	2.95	0.92	1.30	7.56	100.00

images when doing municipality group validation and this calibration is not completely accurate for the unsampled images. The way the calibration is applied to the tree species classes results in some inaccuracy in the calibration of the tree species classes. The timber assortments were calibrated using the same multiplier that was used for the corresponding tree species class.

Overall the differences after calibration are below twice the standard error of the field results. In two regions, Kainuu (shown in Table 4.2) and South Ostrobothnia, the differences are larger than elsewhere. There were municipality groups in the processing windows where no parameters resulted in completely satisfactory results in validation. This is then reflected in the results computed for these regions.

4.3 Taking into account the stratified sampling

The possible improvement from the algorithm modification described in Section 3.4.2 was studied in an image window covering part of Northernmost Lapland, where stratified sampling was used.

The test was comparison of the estimated fractions of the different land use classes (forest land, poorly productive land, and unproductive forest land) against the fractions computed from the field data. The results are presented in Table 4.3.

The results show that the accuracy is slightly better with the modified method than with the unmodified

Table 4.2: Differences between multi-source results and field data results with or without imagewise calibration. Field data from years 2014 - 2016.

Region	Variable		Without calibration		With calibration	
		field mean	difference	significance	difference	significance
Pirkanmaa	combined	156.1	−2.2	−0.6	1.1	0.3
	pine	55.8	5.8	2.6	2.1	0.9
	pine sawlog	20.5	2.4	2.2	1.0	0.9
	pine pulp	33.4	3.0	2.2	0.8	0.6
	spruce	67.6	−5.1	−1.6	−0.5	−0.2
	birch	25.5	−0.9	−0.9	0.2	0.2
	other	7.2	−2.0	−2.3	−0.7	−0.8
Kainuu	combined	97.3	−1.4	−0.7	−3.2	−1.6
	pine	58.6	1.4	1.0	−1.6	−1.2
	pine sawlog	14.0	−0.6	−0.7	−1.3	−1.6
	pine pulp	41.4	1.8	1.9	−0.4	−0.4
	spruce	21.5	−1.9	−1.5	−1.2	−1.0
	birch	15.7	−0.9	−2.0	−0.5	−0.9
	other	1.5	−0.1	−0.4	0.1	0.7
Whole country	combined	108.9	−0.6	−1.0	−0.4	−0.1
	pine	54.7	2.3	5.2	−0.2	−0.1
	pine sawlog	15.8	0.3	1.4	0.5	0.7
	pine pulp	36.2	1.7	5.6	−0.8	−0.4
	spruce	32.5	−1.9	−4.3	−0.1	0.0
	birch	18.1	−0.5	−2.4	−0.1	−0.1
	other	3.5	−0.5	−4.9	−0.1	0.0

Table 4.3: Comparison of estimated land use class proportions (%) against the proportions computed from the field data.

Method	forest land	poorly productive land	unproductive forest land
Field data	27.7	22.1	50.2
Without representativeness	29.7	22.7	47.6
With representativeness	28.2	22.9	48.9

method. However, the differences are not significant. Pixel-wise examination of the land use classes of the nearest neighbours show that, in most cases, the neighbours are from the same stratum than the pixel being classified. This decreases the importance of taking the representativeness into account in the estimation, as noted also by Davies (1988).

4.4 Comparison of the distribution of estimates and field data

The accuracy of area estimates for ranges of values of a forest variable depend on how well the distribution of the estimates matches the real distribution of the variables. This is demonstrated here with some examples.

The examples cover the central part of processing window 15 (see Table 2.2 and Fig. 2.4). The forest area in the test was 2 460 000 hectares and the mean total volume for forestry land was 149 m^3/ha. The test area is large enough so that reliable results for this area can be computed using the field data.

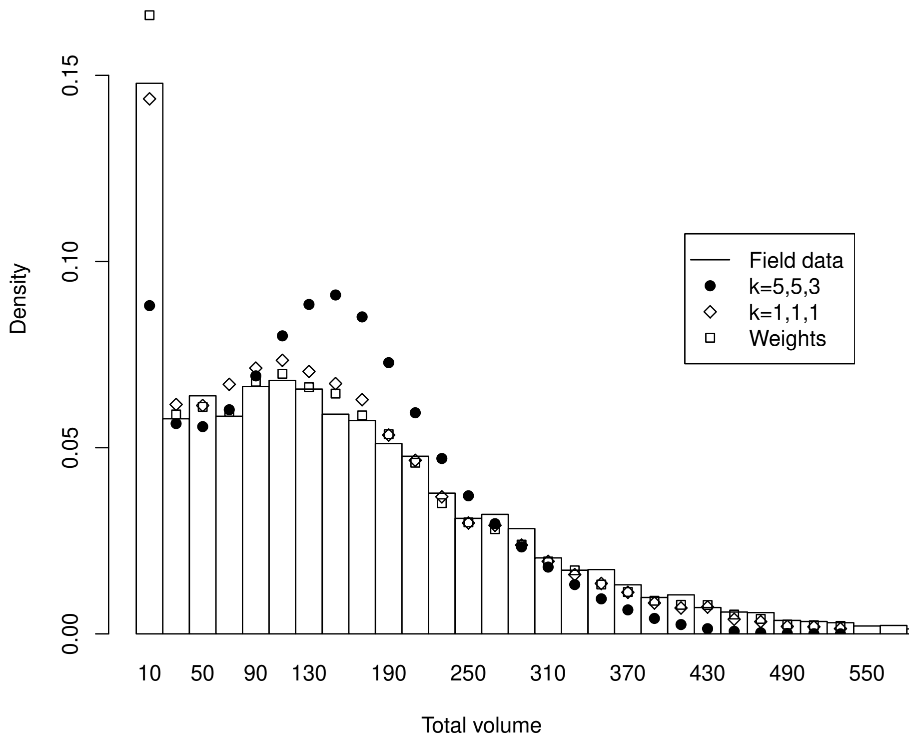


Figure 4.1: Histograms of estimates and field data for the mean volume of growing stock.

Several methods are used in the examples. The ik-NN raster maps were computed using the values of k=5,5,3 (mineral soils, peat land, and open bog) and k=1,1,1. The results presented for accumulated weights were computed using k=5,5,3. Other values of k were also tested but the results did not change noticeably. To compare the ik-NN method for raster maps to another commonly known method, raster maps were computed also using a Generalized Linear Model with logarithmic link function and Gaussian error distribution. The channel weights used with ik-NN were the same weights used in the MS-NFI-2015 maps. Most of the results presented here were computed for the volume of growing stock. For comparison, the cumulative distributions were computed also for some other variables.

Figure 4.1 shows the histograms of the mean volume estimates and the field data for the stem volumes of the sample plots. The bin width was 20 m^3/ha. The field data is shown with bars. The solid dots show the ik-NN result with k=5,5,3. The histogram of estimates using k=1,1,1 is shown using carets. The histogram of the field plot volumes weighted by the accumulated weights is shown using squares. Note that this is not a real histogram of single pixel results, but it shows the distribution the results computed using the accumulated weights are based on.

The figure shows that the distribution of estimates with k=5,5,3 differs significantly from the distribution of field measurements, whereas the result with k=1,1,1 and the (virtual) distribution computed from the accumulated weights are very close to the field data.

The differences between the results are largest at very small volumes. The reason for this is that there are many types of areas with low volume of trees (recent clear cuts, clear cuts with some vegetation, open bogs, etc.). The estimators tend to suppress the extreme values. This is seen also at large volumes,

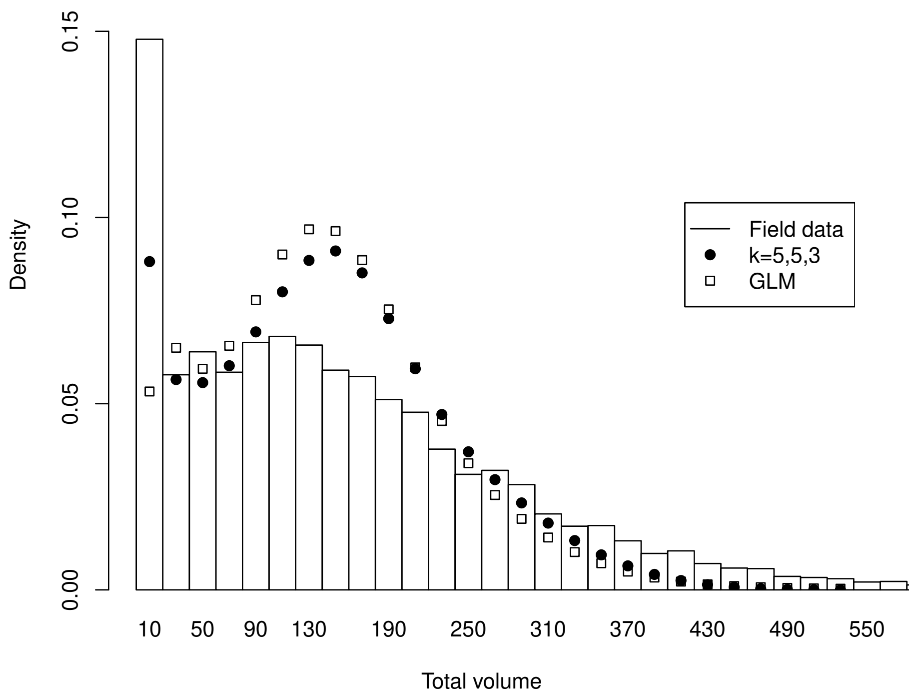


Figure 4.2: Histograms of estimates and field data for the mean volume of trees. Comparison of ik-NN with k=5,5,3 and GLM.

but there the effect seems smaller because the frequencies are small.

Figure 4.2 shows the result with ik-NN (k=5,5,3) and the result with GLM. The distributions of both estimates are narrower than the distribution of the field data. The estimates with ik-NN for small volumes are much closer to the field data for small volumes.

The cumulative distributions can be used to estimate the fraction of the total area in a range of values. Figure 4.3 shows the cumulative distributions of estimates and field data for a set of variables. All of the cumulative frequencies show that the estimates tend to concentrate somewhat to the middle values. The form of the curve depend on the variable because of the different form of the distribution of values. In all cases the estimates computed using accumulated weights are close to the estimates based on field data.

A quantitative example of the narrowing effect is shown in Table 4.4, where the fractions of total area for large volumes are computed from the field data and different estimates. The results show that nearly similar results are obtained with both using the field data and using the accumulated weights. The results computed from the map estimates were smaller for large thresholds when the values of k of (5,5,3) were used. The results with k equal to one were close to the results obtained with field data or accumulated weights.

A serious problem with using small value of k can be seen in the visualisation of the map products. Figure 4.4 shows two total volume estimates for a small window within the test area. The left image is

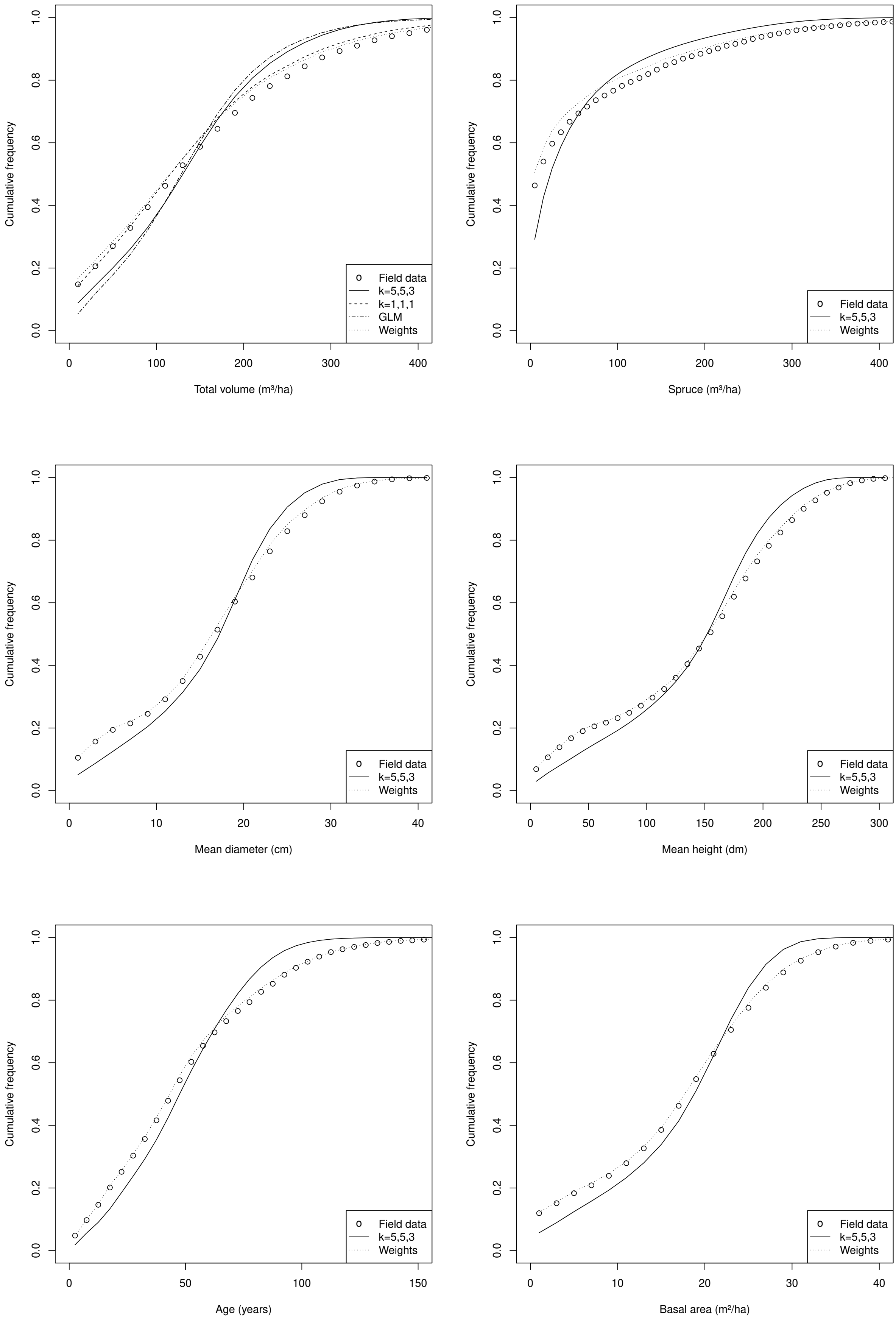


Figure 4.3: The cumulative frequencies for different variables and methods.

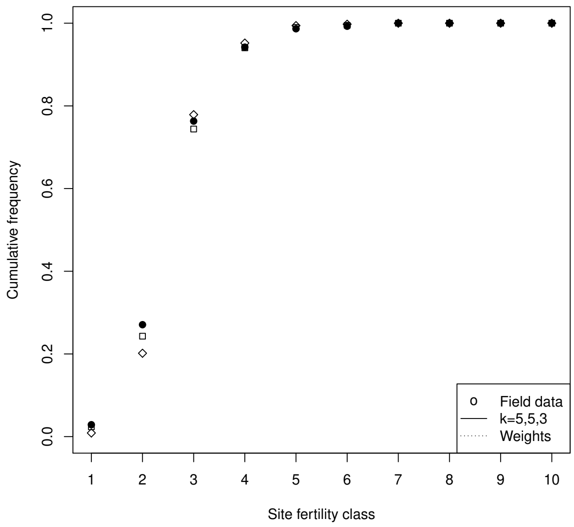


Figure 4.3: The cumulative frequencies for different variables and methods.

Table 4.4: The percentages of the total forest area for forests with large volume.

	Method		Subset		
Class	k	Variation	$V \geq 200$ m^3/ha	$V \geq 300$ m^3/ha	$V \geq 400$ m^3/ha
Field data			29.8	10.9	4.0
ik-NN	1,1,1	weights	27.2	10.9	4.0
ik-NN	5,5,3	weights	27.2	10.9	3.8
ik-NN	1,1,1	map	26.5	10.0	3.3
ik-NN	5,5,3	map	25.3	5.6	0.5
GLM		map	23.0	4.4	0.7

estimated using k equal to one and it looks very grainy. The right image includes much less noise.

This example shows that the choice of k can be used to trade off smoothness of the maps and how much of the statistics of the forest is retained in the estimates. Many studies have shown that the pixelwise accuracy with k = 1 is not as good as when k is larger (see, e.g., Katila and Tomppo (2001)). This, in addition to the visual appearance, suggests that k should be larger than one when making forest resource maps.

4.5 Forest resources by municipalities

The primary results of MS-NFI are the forest resource estimates for municipalities. With the MS-NFI method, it is possible, at least in theory, to estimate for municipalities all the parameters that are usually estimated for the regions using field data only. The estimates are presented in Appendix Tables 1–9 for the parameters whose estimates are considered to be sufficiently precise. These tables include the same parameters as the ones published for MS-NFI-2009 (Tomppo et al. 2013), MS-NFI-2011 (Tomppo et al. 2014) and MS-NFI-2013 (Mäkisara et al. 2016). In the result calculation phase, the MS-NFI estimates for the groups of the municipalities and at region (maakunta) level were compared for the estimates and error estimates based on field data only in order to control possible significant errors, including

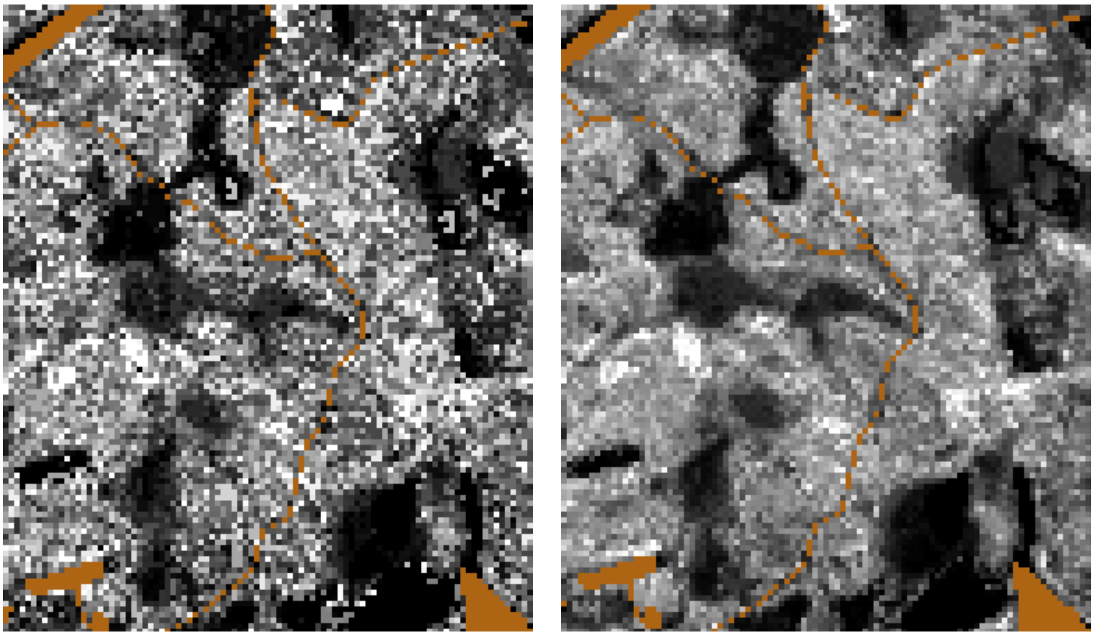


Figure 4.4: Examples of total volume estimates computed using k=(1,1,1) (left) and k=(5,5,3). The brown area is not in the estimated area.

biases. The estimation parameters were changed and additional calibration by key output variables (sect. 3.4.6) was applied until satisfactory estimates were obtained. That is, the MS-NFI estimates of the key parameters, such as areas and volumes by tree species, were within two standard errors of estimates based on the field data only.

The estimates can be divided into area and volume estimates. Some tables present only area estimates, some only volume estimates and some volume estimates for sub-categories of forest land or poorly productive forest land, together with area estimates of the sub-categories.

The estimates of the areas of forest land, poorly productive forest land and unproductive forest land (three forestry land categories) are given in Appendix Table 1a for the entire forestry land and in Appendix Table 1b for forestry land available for wood supply. A national classification is used for land classes, for definitions of these classes and comparison to FAO land classes see Tomppo et al. (2011a). The areas and proportions of forestry land of mineral soils and peatland soils are given in Appendix Tables 2a separately for three forestry land categories, and the similar estimates for forestry land available for wood supply are given in Appendix Tables 2c. The Appendix Tables 2b and 2d show again the areas of forest land and poorly productive forest land on mineral soils and peatland soils, as in Appendix Tables 2a and 2b, and now also the mean volumes of the growing stock for the land categories of the tables.

The dominant tree species by municipalities are presented in Appendix Table 3a for forest land and in Appendix Table 3b for poorly productive forest land. The dominant tree species is defined in the NFI for the field assessment as a stand-level variable. In NFI11 and NFI12, it is the tree species with the highest basal area for the development classes from young thinning stand to mature stand and seed tree and shelterwood stands, and is defined as the tree species with highest number of stems capable of development in young and advanced seedling stands. The proportion of pine dominated forests of forest land is usually high in North Finland (Lappi), often over 80 %, and also in the many municipalities in South, Central and North Ostrobothnia regions. A high proportion of pine mires increases the area and proportion of pine dominated forests in these regions. The proportion is high also in some areas in South Finland in coastal regions and Central Finland in areas where Sub-xeric heath forests are common. Among the areas of the Regions, the proportion of spruce dominated forests on forest land is highest in South Finland in Kanta-Häme and Päijät-Häme Regions being 50 % with the highest municipality level estimates 55 % or more in Pälkäne and Asikkala.

The stand age and the development class of a stand used in MS-NFI are defined in a same way as in the field inventory (Tomppo et al. 2011a). The area estimates for age classes on forest land by municipalities is presented in Appendix Table 4a and for development classes in Appendix Table 5a. The proportion of forest land with a stand age not more than 40 years varies by municipality in South Finland, from about 15 % to about 60 %, the regions level proportions being around 40 %, The proportional area of young forests is high in Eastern and South-Eastern Finland, and also in some municipalities in Central Finland. The proportional area of forest not more than 40 years old is lower in North Finland than in South Finland. One should note that the same age in South Finland and North Finland corresponds to different development class of stand due to slower growth in the North than in the South.

The mean volume and total volume estimates are given in many different ways: mean volumes by tree species and by timber assortments for combined forest land and poorly productive forest land (6a) likewise with total volumes (6b). The similar estimates are given for forest land and poorly productive forest land available for wood supply in Appendix Tables 6c and 6d. Appendix Tables 7a–d present the similar estimates for forest land as Appendix Tables 6a–d for forest land and poorly productive forest land. Note that poorly productive forest land consists either of rocky soils, fjeld forests or less fertile peatland soils, such as oligo-ombrotrophic or ombrotrophic peatland, e.g., Sphagnum fuscum dominated peatland. Note that the water balance of peatland soil also affects the wood production capacity and land class of peatland.

The mean volumes of the growing stock in the municipalities vary significantly by the regions and also within the regions. The mean volume estimates on forest land and forest and poorly productive forest land are given also separately for mineral soils and peatland soils in Appendix Table 2b, and for forest land and forest and poorly productive forest land available for wood supply in Appendix Table 2d.

The mean volume estimates by age classes on forest land are given in Appendix Table 4b and the similar estimates for the forest land available for wood supply in Appendix Table 4d. The corresponding mean volume estimates by development classes on forest land are given in Appendix Tables 5b and 5d.

The mean volume of the growing stock on combined forest land and poorly productive forest land by municipalities varies in Southernmost Finland is typically over 140 or 150 cubic metres per hectare (m^3/ha) except in the municipalities in South coast and Åland region. In continental South Finland (regions 1–16, Fig. 1.1), the mean volume of growing stock by municipalities (Appendix Table 6a and Fig. 4.5) ranged from 93 m^3/ha in Halsua to 195 m^3/ha in Pälkäne and those for pine from 31 m^3/ha in Kärkölä to 86 m^3/ha in Naantali and those for spruce from 9 m^3/ha in Halsua and 107 m^3/ha in Pälkäne and those for birch from 13 m^3/ha in Kauhajoki to 40 m^3/ha in Valkeakoski. In North Finland (Regions 17–19, Fig. 1.1), the respective ranges were from 27 m^3/ha in Utsjoki to 125 m^3/ha Nivala and those for pine from 12 m^3/ha in Utsjoki to 66 m^3/ha in Kalajoki and those for spruce from 0 m^3/ha in Utsjoki to 36 m^3/ha in Nivala and those for birch from 6 m^3/ha in Savukoski to 34 m^3/ha in Hailuoto.

The mean volume of spruce saw log on combined forest land and poorly productive forest land and also on that available for wood supply by municipalities is naturally highest in the region in which spruce volume is highest, that is, in southern part of Central Finland, in regions of Häme-Uusimaa, and the southern regions of Central Finland and Pohjois-Savo. The volume of pine saw log is relatively high, near or over 30 m^3/ha, in some municipalities in South-West (Lounais-Suomi) and South-East (Kaakkois-Suomi) and Etelä-Savo, while the birch saw log, 6–8 m^3/ha in some municipalities in Häme-Uusimaa and Rannikko/Etelä-Rannikko (Appendix Tables 6a and 6c).

The total growing stock on forest land of peatlands available for wood supply are presented in Fig. 4.6.

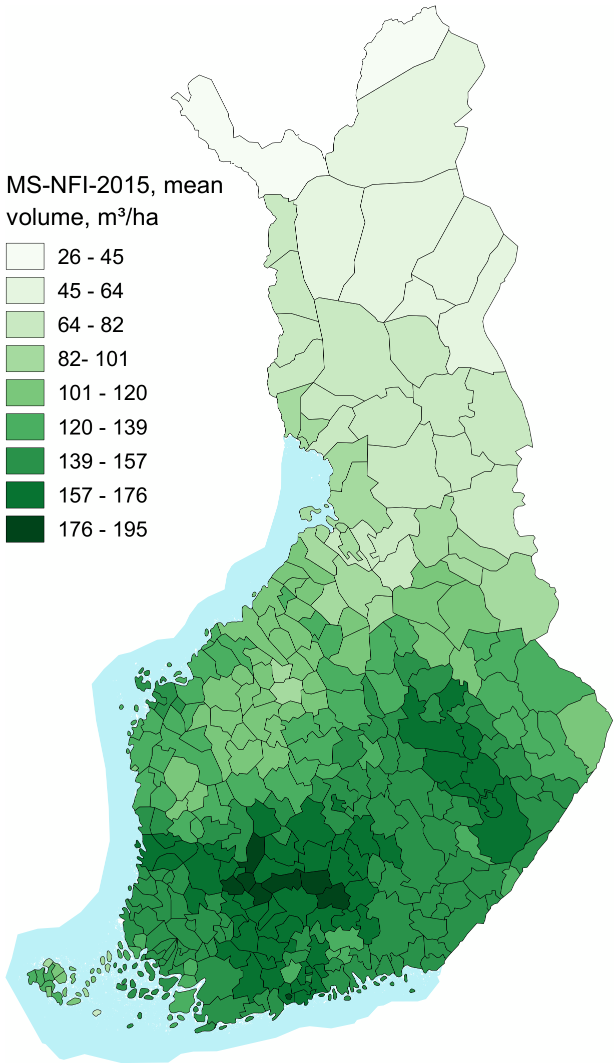


Figure 4.5: The mean volume of growing stock on forest and poorly productive forest land by municipalities. Digital map data: ©National Land Survey of Finland, licence No. MML/VIR/MYY/328/08.

The highest growing stocks are in the Northern Finland and of the municipalities highest in Kuhmo being 14 mill. m^3. One explanation is that the area of the municipalities simply is larger in East and North Finland in addition to the fact that the proportion of peatlands of the forest land are also highest on those areas.

4.5.1 Biomass estimates and available energy wood

The two main potential development classes of wood energy sources of forests in Finland are young thinning stands and mature stands. All tree compartments above stump of those trees to be harvested as energy wood are removed in a young thinning stand, i.e., the entire stem over stump, branches and foliage. Only the residual part of stem, in addition to branches, and optionally stumps and coarse roots (with a theoretical minimum diameter of 2 cm) are removed for energy in the case of the regeneration cutting of a mature stand.

Biomass estimates have been calculated for each field plot and plot part on forest land and poorly productive forest land in the NFI11 and NFI12 data for biomass and energy wood estimation. The biomass

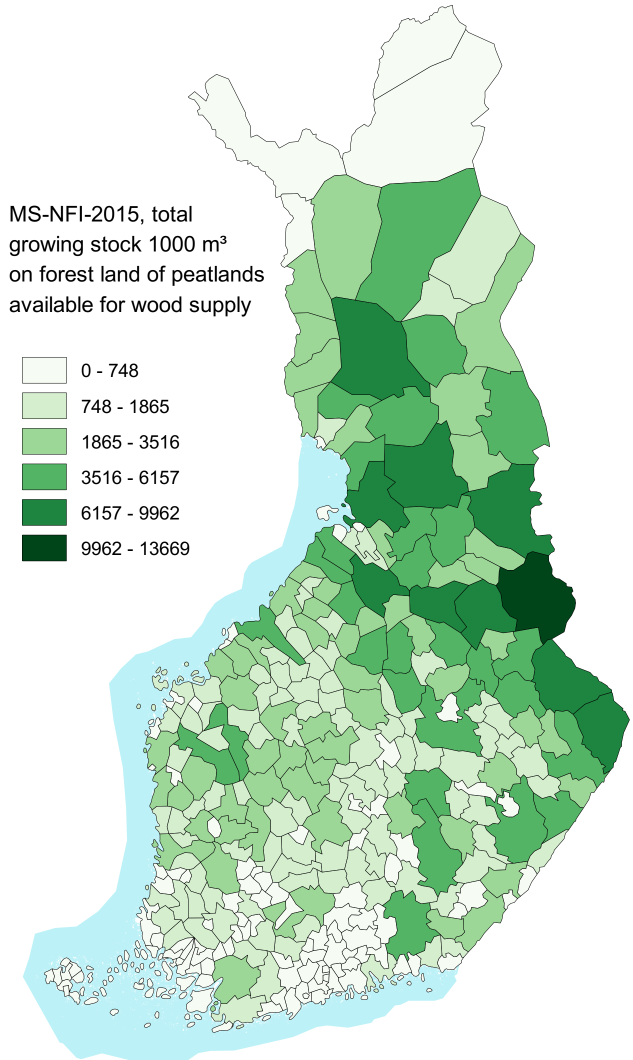


Figure 4.6: The total growing stock on forest land of peatlands available for wood supply. Digital map data: ©National Land Survey of Finland, licence No. MML/VIR/MYY/328/08.

estimates in the field data by tree compartments (Table 4.5) were predicted first for sample trees on forest land and poorly productive forest land of NFI11 and NFI12 and then predicted for tally trees in a similar manner as the volumes. The biomass of the stem (including bark) was calculated from the volume of a stem using stem wood density models by (Repola et al. 2007). The biomass estimates of the other tree compartments were calculated using the models by Repola (2008, 2009) (Table 4.5). Note that the stem residual biomass is included to the stem biomass in the table. Tree level biomass predictions were converted to kilograms per hectare (kg/ha), taking into account the angle count sampling (Bitterlich sampling) basal area factor and the maximum radius of the plot. The biomass estimates by tree species groups in young thinning stands (development class) are presented in Appendix Table 8a with a unit of Gg (10^9 g). The biomasses of stem and bark, branches and foliage were calculated using those field plots on which first commercial thinning was proposed for the first 5-year period or on which pre-commercial thinning was proposed and the treatment was already considered to be delayed in the corresponding stand. The proportion of the field plot biomass capable to be removed was estimated employing the field plot data and the basal area thresholds from the thinning regimes for mineral soils (Hyvän metsänhoidon suositukset 2005) and peatlands (Hyvän metsänhoidon suositukset turvemaille 2007), varying according to region (degree days), dominant tree species and site class. First the dominant height of the field plot stand was estimated from stand mean height. The basal area removed was the the stand basal

Table 4.5: The compartments for tree biomass (Repola et al. 2007, Repola 2008, 2009).

Stem and bark
Living branches
Foliage
Dead branches
Stump
Stem residual (from NFI timber assortment class proportions and stem, and bark biomass)

area minus basal area threshold value after cutting from the particular thinning regime. The basal area removal percentage was converted to volume removal percentage using the relations of these two removals obtained from Motti stand simulator for corresponding regions, dominant tree species and site classes (Hynynen et al. 2002). This percentage was used to estimate the biomass components removal for the selected field plots. Appendix Table 8c presents similar estimates to those in Appendix Table 8a for land available for wood supply. The biomass estimates of mature forests are presented separately for branches, foliage and stem residuals, and stumps and large roots by tree species groups in Appendix Table 8b and for land available for wood supply in Table 8d. In practice, only spruce stumps, branches and stem residuals can be harvested from regeneration cutting areas

The biomass models employed here and in (Tomppo et al. 2012) were different from those used in MS-NFI-2005 (Tomppo et al. 2009a). The effect of the two sets of the models on the biomass estimates and differences is discussed in (Tomppo et al. 2012).

The energy wood estimates represent energy wood potential rather than the energy wood available in practice. The practical constraints in use and harvesting of energy wood, like minimum removal and other cutting operations were not taken into account. For more details of deriving the biomass estimates and the reliability of the estimates, see (Tomppo et al. 2008b).

4.6 Digital thematic output maps

Thematic forest maps in raster format were produced for the most important forest variables: land class, main site class, site fertility class, stand age, mean diameter and height of stands, stand basal area, canopy cover of trees, canopy cover of broad leaved trees and volumes by tree species and timber assortments, for four tree species or species groups (pine, spruce, two birch species combined, and other broad leaved tree species). Twelve volume maps were produced for volumes of saw timber, pulp wood and total volume by tree species groups and one for all species together, without breaking down to saw timber and pulp wood. Twentyone thematic maps show the biomass estimates for three tree species groups (pine, spruce, broad-leaved trees) and for seven tree compartments. Note that the biomass component stem residual is a part of biomass component stem and bark. The maps produced are georeferenced raster layers in ETRS-TM35FIN coordinate system with a spatial resolution of 16 m by 16 m (Table 4.6), and cover Finland as shown in Fig. 2.4. The non-forestry land use cover was obtained from the digital land use map data and overlaid on the satellite image data during the estimation phase. The ik-NN pixel-level predictions were made for the rest of the area. An example of the total volume thematic map is given in Fig. 3.2c. The raster layers can be combined to produce new thematic maps, e.g., dominant tree species and mean volume classes by tree species dominance. More examples of digital thematic maps are given in Tomppo et al. (2008b).

Two kinds of map were made for the open access product: the viewable maps and the complete maps. The background map data has been replaced by a no-data value in both products. The viewable maps are

Table 4.6: The estimated raster themes.

Theme
Biomass, spruce, living branches 2015 (10 kg/ha)
Biomass, spruce, stem residual 2015 (10 kg/ha)
Biomass, spruce, roots, d > 1 cm 2015 (10 kg/ha)
Biomass, spruce, stump 2015 (10 kg/ha)
Biomass, spruce, dead branches 2015 (10 kg/ha)
Biomass, spruce, stem and bark 2015 (10 kg/ha)
Biomass, spruce, foliage 2015 (10 kg/ha)
Biomass, broad-leaved trees, living branches 2015 (10 kg/ha)
Biomass, broad-leaved trees, stem residual 2015 (10 kg/ha)
Biomass, broad-leaved trees, roots, d > 1 cm 2015 (10 kg/ha)
Biomass, broad-leaved trees, stump 2015 (10 kg/ha)
Biomass, broad-leaved trees, dead branches 2015 (10 kg/ha)
Biomass, broad-leaved trees, stem and bark 2015 (10 kg/ha)
Biomass, broad-leaved trees, foliage 2015 (10 kg/ha)
Biomass, pine, living branches 2015 (10 kg/ha)
Biomass, pine, stem residual 2015 (10 kg/ha)
Biomass, pine, roots, d > 1 cm 2015 (10 kg/ha)
Biomass, pine, stump 2015 (10 kg/ha)
Biomass, pine, dead branches 2015 (10 kg/ha)
Biomass, pine, stem and bark 2015 (10 kg/ha)
Biomass, pine, foliage 2015 (10 kg/ha)
Site main class 2015 (1–4)
Site fertility class 2015 (1–8)
Land class 2015 (1–3)
Stand age 2015 (year)
Stand mean diameter of 2015 (cm)
Stand mean height 2015 (dm)
Canopy cover 2015 (%)
Canopy cover of broad-leaved trees 2015 (%)
Stand basal area 2015 (m^2/ha)
Data source index, MS-NFI-2015
Volume, birch 2015 (m^3/ha)
Volume, birch pulpwood 2015 (m^3/ha)
Volume, birch saw timber 2015 (m^3/ha)
Volume, spruce 2015 (m^3/ha)
Volume, spruce pulpwood 2015 (m^3/ha)
Volume, spruce saw timber 2015 (m^3/ha)
Volume, other broad-leaved trees 2015 (m^3/ha)
Volume, other broad-leaved trees pulpwood 2015 (m^3/ha)
Volume, other broad-leaved trees saw timber 2015 (m^3/ha)
Volume, pine 2015 (m^3/ha)
Volume, pine pulpwood 2015 (m^3/ha)
Volume, pine saw timber 2015 (m^3/ha)
Volume, the growing stock 2015 (m^3/ha)

8-bit images where the forest variables have been classified into a small number of classes and colours have been assigned to the classes. These maps are available in viewing services like the Paikkatietoikkuna of the National Land Survey of Finland (`http://www.paikkatietoikkuna.fi/`) (Paikkatietoikkuna 2018). The complete maps retain the full precision of the results in 16-bit raster data. These are available at `http://kartta.luke.fi/index-en.html` (Luke 2018).

For a cover as complete as possible for MS-NFI-OA-2015 from the entire country, the 2015 product has been completed by the data estimates from the recent years. The product thus consists of the following sub-products:

1. The estimates from 2015, the NFI field data from 2012–2016 and the satellite images from 2015–

Source	pixels	1000 ha	%
MS-NFI-2015	1 026 153 536	26 269.5	98.69
MS-NFI-2013	13 254 848	339.3	1.27
MS-NFI-2011	142 489	3.6	0.01
MS-NFI-2009	18	0.0	0.00
2006/2009 Enontekiö	233 289	6.0	0.02
MS-NFI-2007	12 153	0.0	0.00
No data	9 269	0.2	0.00
Total	1 039 805 602	26 619.0	100.0

Table 4.7: The proportions of forest area from the different estimates in MS-NFI-OA-2015.

2016,

2. The estimates from 2013, the NFI field data from 2009–2013 and the satellite images from 2012–2014 (Mäkisara et al. 2016),

3. The estimates from 2011, the NFI field data from 2007–2011 and the satellite images from 2009–2012 (Tomppo et al. 2014),

4. The estimates from 2009, the NFI field data from 2006–2010 and the satellite images from 2009–2010 (MS-NFI-2009) (Tomppo et al. 2013),

5. The estimates for about municipality Enontekiö in North Lapland, the NFI field data from 2003 and the satellite images from 2000,

6. The estimates from 2007, the NFI field data from 2005–2008 and the satellite images from 2005–2008.

Note that the sum and mean values calculated from raster layers could deviate, and also do deviate in most cases, from the area and volume estimates in the Appendix Tables, due to the corrections for map errors. The forestry land area calculated from the maps is greater, and the mean volume estimates smaller, than those in the Appendix Tables in most cases (cf. Subsection 3.4.4). Particularly, if the estimates are calculated for a specific stratum, e.g., the mean or total volume for forests older than 120 years, the small area estimates calculated using the weights (Eq. 3.9) may deviate significantly from the estimates calculated from the map layers (c.f., Table 4.4). The reason is that there is a tendency towards the mean in the map form estimates while original field data are used in the small area estimates. For these reasons, it is recommended to use the small area approach in estimation when estimating strata far from the mean of the variable.

The open access product MS-NFI-OA-2015 includes, besides the latest estimates, estimates from earlier multisource inventories. The proportion of pixels and area included from the different estimates is shown in Table 4.7. The table shows that coverage of the latest year 2015 inventory is very good. Note that the here the total area is the area under the forest mask derived from the topographic database (see Section 2.3.2) and it differs slightly from the area of forestry land computed from the field plot data in NFI11 (26 193 000 ha from Korhonen et al. (2017)).

5 Discussion

The development of the Finnish multi-source inventory (MS-NFI) method began in 1989 in the connection of the 8th National Forest Inventory of Finland (NFI8). The method utilises satellite images, field data of the National Forest Inventory (NFI) and digital map data. The methods and results of the first country-level MS-NFI, MS-NFI8, are presented in (Tomppo et al. 1998). The revised methods and results, MS-NFI9 methods and results, corresponding to the ninth National Forest Inventory (1996–2003), were published in (Tomppo et al. 2008b).

The main purpose of the MS-NFI method is to obtain forest resource information for areas smaller than would be possible using only field data. The adopted k-Nearest Neighbour estimation method (k-NN) meets the requirements set to the method and the results. In addition to the small area estimates, the MS-NFI provides predictions of forest variables in map form.

One of the applications of satellite image-based digital volume maps is to use them to simulate different sampling strategies. For instance, the sampling intensities of Finnish NFIs have been fitted to the spatial variation in forests throughout the whole country, being lower in the north than in the south (Henttonen 1991, Tomppo et al. 2011a).

The method has been improved continuously and new features have been added since its first implementation. Similar development work is being carried out in several other countries.

For MS-NFI-2015 the NFI field plot data from years 2012–2016 were used. Most of the images were from year 2015 (Table 2.2). One frame was from 2016, due to the lack of the images from 2015 for that region. The variables of the field data were projected (updated) to a certain time point, in this case to July 31, 2015. An approach based on stand level models were selected after exploring different alternatives (Subsection 3.2, Tomppo et al. 2014). Existing models were used for volume increments. Models and their parameters for the increments of the other updated quantities were estimated for the MS-NFI purposes. The permanent sample plots of the NFI from years 2004–2011 were used. Big changes on the NFI plots compared to the image data were interpreted. Some of them are difficult to detect. Thus some open questions still remain, in addition to the precisions of the increment models. Examples, in addition to the detection of some big changes, such as regeneration cuttings and heavy storm damages, the detection of light storm damages and thinning cuttings (not tried for this product).

The pixel level prediction error is generally rather high in MS-NFI. Several error sources are listed in (Tomppo et al. 2008b). For this article, the estimates and pixel level predictions were validated when selecting the estimation parameters comparing MS-NFI estimates and error estimates with those based on NFI11 field data using groups of municipalities. In the future, more accurate error estimates for the NFI field data will be available using post-stratification (Haakana et al. 2018).

The ik-NN method can be used to either compute forest resource maps or to directly compute estimates for groups of pixels. The results computed for municipalities in MS-NFI are an example of the latter method. There are significant advantages when computing the results directly compared to using the raster maps as an intermediate step, as exemplified in Subsection 4.4. The examples show how carrying information about the distribution of the estimates to the final aggregation of the results helps in computing the fractions of uncommon values of the forest variables. The information from the distributions is retained in the maps if the value of k is one. Unfortunately this reduces the usability of the maps for applications that are natural to them.

A method to calibrate the estimates with the field data within each processing window was used in MS-

NFI-2015. The effectiveness of the method was tested using regions as target areas (see Section 4.2). The results show that the calibration makes the differences in mean values between the field data and the estimates smaller and reduces the number of significant differences. The users should note that the calibration is based on the estimates after map error correction. When using the raster maps, the map error correction is not available. This means that the differences between averages computed from field data and averages computed from raster estimates are expected to be larger than in Table 4.2.

A method to take into account the differences within sampling density within a processing window was also used in MS-NFI-2015. However, test show that the improvement from this refinement was found to be very small (see Section 4.3).

The topographic map data is used in computing the estimates from remote sensing data. On the other hand, the field plot attributes are used in computing the results from field data. The classification (e.g., land class) of a sample plot is different when computed from the map and when measured in the field, as shown in Subsection 4.1. There are several reasons for these differences (positional error, slightly different definitions of classes, etc.). The effect of these differences can be statistically corrected for large groups of pixels, but it does not correct the errors in the raster estimates.

It was found difficult to make the delineation of the protected areas to match the NFI field inventory estimates. The Metsähallitus multiple step rules for delineation of the non-production land areas were available only at the NFI plot level.

The most serious potential risk in the application of MS-NFI method is the availability of relevant satellite images. An individual satellite image scene should be large enough to cover high enough number of field plots, preferably several thousands, to get satisfactory ground truth data. On the other hand, the pixel size should not be larger than about 30 metres. In addition to the problems caused by clouds, the number of the natural resource satellites with suitable specifications for forest applications has not been large. ESA's Sentinel-2A, launched in 23 June 2015 and Sentinel-2B, launched 7 March 2017, have improved coverage.

This article is one in the series in which the MS-NFI estimates are calculated every second year for the greater part of the country and every fourth year for northernmost Finland. The future method development work will focus, in addition to the decrease of all kinds of estimation errors, to improve the consistency between the subsequent products. Now that there are four openly available products, users are increasingly interested in the trends observed by comparing the products. Up to now, the MS-NFI products have been independent but possibilities to use previous estimates to increase the reliability of current estimates will be investigated.

Acknowledgements

The results presented in this article presume the availability of forest inventory field data, amounting about to 60 000 field plots on forestry land in NFI11 and NFI12 from 2012–2016, as well as functioning computing and data management facilities, all provided by Luke. The manuscript has been read and valuable improvements suggested by Research Prof Tuula Packalen. We thank all the individuals whose support has made this article possible.

References

Card, D.H. 1982. Using known map category marginal frequencies to improve estimates of thematic map accuracy. Photogrammetric Engineering and Remote Sensing 48: 431–439.

Cochran, W.G. 1977. Sampling techniques. Wiley. New York, 428 p.

Czaplewski, R.L. & Catts, G.P. 1992. Calibration of remotely sensed proportion or area estimates for misclassification error. Remote Sensing of Environment 39: 29–43.

Diemer, C. , Lucaschewski, I. , Spelsberg, G. , Tomppo, E. & Pekkarinen, A. 2000. Integration of terrestrial forest sample plot data, map information and satellite data, An operational multisource-inventory concept. In: Ranchin T, Wald L (ed.). Proceedings of the Third Conference ”Fusion of Earth Data: Merging Point Measurements, Raster Maps and Remotely Sensed Images”. Sophia Antipolis, France, January 26–28, 2000, SEE /URISCA, Nice. p. 143–150.

Dash, J. P., Marshall, H. M. & Rawley, B. 2015. Methods for estimating multivariate stand yields and errors using k-NN and aerial laser scanning. Forestry 88:237–244.

Davies, E.R. 1988. Training sets and a priori probabilities with the nearest neighbour method of pattern recognition. Pattern Recognition Letters 8:11–13.

Finley, A.O. & McRoberts, R.E. 2008. Efficient k-nearest neighbor searches for multi-source forest attribute mapping. Remote Sensing of Environment 112: 2203–2211.

Finley, A.O. , McRoberts, R.E., & Ek, A.R. 2006. Applying an efficient k nearest neighbor search to forest attribute imputation. Forest Science 52: 130–135.

Finnish Statistical Yearbook of Forestry 2014. Finnish Forest Research Institute, Vantaa, Finland, 428 p.

Franco-Lopez, H., Ek, A.R. & Bauer, M.E. 2001. Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. Remote Sensing of Environment 77: 251–274.

Gjertsen, A. 2005. Accuracy of forest mapping based on Landsat TM data and a kNN method. In: H. Olsson (ed.). Proceedings of ForestSat 2005. Borås, Sweden, 31 May–3 June 2005. p. 7–11. Available at: http://shop.skogsstyrelsen.se/shop/9098/art92/4646092-23ff72-1740-1.pdf. [Cited 11 Dec 2018].

Haakana, H., Heikkinen, J., Katila, M. & Kangas, A. 2018. Efficiency of post-stratification for a large-scale forest inventory – case Finnish NFI. Manuscript submitted to Annals of Forest Science.

Hagner, O. & Olsson, H. 2004. Normalisation of Within-Scene Op-tical Depth Levels in Multispectral Satellite Imagery Using National Forest Inventory Plot Data. In: Proceedings from the 24th EARSeL Symposium, Workshop on ”Remote sensing of land use and land cover”, Dubrovnik, Croatia, May 28–29, 2004. p. 279–284. Available at: http://www.earsel.org/symposia/2004-symposium-Dubrovnik/pdf/335.pdf. [Cited 11 Dec 2018].

Henttonen, H. 1991. VMI8:n Pohjois-Suomen otanta-asetelmien vertailu satelliittikuvatulkinnan avulla. Comparing of sampling designs through simulation. Finnish Forest Research Institute, 10 p. (In Finnish).

Hynynen, J., Ojansuu R., Hökkä H., Siipilehto J., Salminen H., and Haapala P. 2002. Models for predicting stand development in MELA system. Metsäntutkimuslaitoksen tiedonantoja 835. Vantaa, Finland. (In Finnish).

Hyvän metsänhoidon suositukset 2006. Forestry Development Centre Tapio, Helsinki (In Finnish).

Hyvän metsänhoidon suositukset turvemaille 2007. Forestry Development Centre Tapio, Helsinki (In Finnish).

Katila, M. 2006a. Correcting map errors in forest inventory estimates for small areas. In: Kangas, A. & Maltamo, M, (ed.). Forest Inventory –Methodology and Applications, Managing Forest Ecosystems Vol. 10. Springer, Dodrecht, The Netherlands p. 225–233.

Katila, M. 2006b. Empirical errors of small area estimates from the multi-source National Forest Inventory in Eastern Finland. Silva Fennica 40(4): 729–742.

Katila, M. & Tomppo, E. 2001. Selecting estimation parameters for the Finnish multi-source national forest inventory. Remote Sensing of Environment 76: 16–32.

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

Katila, M. , Heikkinen, J. & Tomppo, E. 2000. Calibration of small-area estimates for map errors in multisource forest inventory. Cana-dian Journal of Forest Research 30:1329–1339.

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

Korkeusmalli 10 m. National Land Survey of Finland. Available at: http://www.maanmittauslaitos.fi/kartat-ja-paikkatieto/asiantuntevalle-kayttajalle/tuotekuvaukset/korkeusmalli-10-m. [Cited 11 Dec 2018]. (In Finnish).

Korhonen, L., Korhonen, K.T., Stenberg, P., Maltamo, M. & Rautiainen, M. 2007. Local models for forest canopy cover with beta regression. Silva Fennica 41(4): 671–685. Available at: http://www.metla.fi/silvafennica/full/sf41/sf414671.pdf [Cited 11 Dec 2018].

Korhonen, K.T., Ihalainen, A., Ahola, A., Heikkinen, J., Henttonen, H.M., Hotanen, J-P., Nevalainen, S., Pitkänen, J., Strandström, M. & Viiri, H. 2017. Suomen metsät 2009-2013 ja niiden kehitys 1921-2013. Natural resources and bioeconomy studies 59/2017, Natural Resources Institute Finland. (In Finnish) 215 s. http://urn.fi/URN:ISBN:978-952-326-467-0. [Cited 11 Dec 2018].

Koukal, T., Suppan, F. & Schneider, W. 2005. The impact of radiometric calibration on kNN predictions of forest attributes. In: Olsson, H. (ed.). Proceedings of ForestSat 2005. Borås, Sweden, 31 May–3 June 2005. p. 17–21. Available at: http://shop.skogsstyrelsen.se/shop/9098/art92/4646092-23ff72-1740-1.pdf. [Cited 11 Dec 2018].

Landsat Missions. USGS web site. http://www.usgs.gov/landsat. [Cited 11 Dec 2018].

Lehto, J. & Leikola, M. 1987. Käytännön metsätyypit. Kirjayhtymä, Helsinki (In Finnish), 96 p.

Luonnonvarakeskus 2018. File service for publicly available data. http://kartta.luke.fi/index-en.html [Cited 11 Dec 2018].

Maanmittauslaitoksen maastotietokohteet (Feature catalog of the topographic database). National Land Survey of Finland, Helsinki, Finland. Available at: http://www.maanmittauslaitos.fi/sites/default/files/Maastotietokohteet_2013.pdf. [Cited 17 Sep 2014].(In Finnish).

Magnussen, S. 2013. An assessment of three variance estimators for the k-nearest neighbour technique. Silva Fennica 47: 1–19.

Magnussen, S. & Tomppo, E. 2016. Model-calibrated k-nearest neighbor estimators. Scandinavian Journal of Forest Research 31: 183–193.

Magnussen, S., McRoberts, R. & Tomppo, E.O. 2009. Model-based mean square error estimators for k-nearest neighbour predictions and applications using remotely sensed data for forest inventories. Remote Sensing of Environment 113: 476–488.

Magnussen, S., McRoberts, R.E. & Tomppo, E.O. 2010. A resampling variance estimator for the k nearest neighbours techniques. Canadian Journal of Forest Research 40: 648–658.

Magnussen, S., Frazer, G. & Penner, M. 2016. Alternative mean-squared error estimators for synthetic estimators of domain means. Journal of Applied Statistics 43, 2550–2573.

Maselli, F., Chirici, G., Bottai, L., Corona, P. & Marchetti, M. 2005. Estimation of Mediterranean forest attributes by the application of k-NN procedures to multitemporal Landsat ETM+ images. International Journal of Remote Sensing 26: 3781–3796.

McRoberts, R.E. 2006. A model-based approach to estimating forest area. Remote Sensing of Environment 103: 56–66.

McRoberts, R.E. & Tomppo, E.O. 2007. Remote sensing support for national forest inventories. Remote Sensing of Environment 110: 412–419.

McRoberts, R.E., Nelson, M.D. & Wendt, D.G. 2002a. Stratified estimation of forest area using satellite imagery, inventory data, and the k-Nearest Neighbors technique. Remote Sensing of Environment 82: 457–468.

McRoberts, R.E., Wendt, D.G., Nelson, M.D. & Hansen, M.H. 2002b. Using a land cover classification based on satellite imagery to improve the precision of forest inventory area estimates. Remote Sensing of Environment 81: 36–44.

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.

McRoberts, R.E., Cohen, W.B., Naesset, E., Stehman, S.V. & Tomppo, E.O. 2010. Using remotely sensed data to construct and assess forest attribute maps and related special products. Scandinavian Journal of Forest Research 25: 340–367.

McRoberts, R.E., Tomppo, E.O.& Naesset, E. 2010. Advances and emerging issues in national forest inventories. Scandinavian Journal of Forest Research 25: 368–381.

McRoberts, R.E., Magnussen, S., Tomppo, E.O. & Chirici, G. 2011. Parametric, bootstrap, and jackknife variance estimators for the k-Nearest Neighbors technique with illustrations using forest inventory and satellite image data. Remote Sensing of Environment 115: 3165–3174.

Mäkisara, K., Katila, M., Peräsaari, J. & Tomppo, E. 2016. The Multi-Source National Forest Inventory of Finland – methods and results 2013. Natural resources and bioeconomy studies 10/2016, Natural Resources Institute Finland. 215 s. http://urn.fi/URN:ISBN:978-952-326-186-0. [Cited 14 Mar 2018].

Nilsson, M. 1997. Estimation of Forest Variables Using Satellite Image Data and Airborne Lidar. Ph.D. thesis, Swedish University of Agricultural Sciences, The Department of Forest Resorce Management and Geomatics, Acta Universitatis Agriculturae Sueciae. Silvestria 17, 84 p.

Nyyssönen, A. & Mielikäinen, K. 1987. Metsikön kasvun arviointi. Summary: Estimation of stand increment. Acta Forestalia Fennica 163. 40 p.

Paikkatietoikkuna.http://www.paikkatietoikkuna.fi/. [Cited 14 Mar 2018].

Rao, J.N.K. 2003. Small Area Estimation. Wiley, New York. 313 p.

Reese, H., Nilsson, M., Sandström, P. & Olsson, H. 2002. Applictions using estimates of forest parameters derived from satellite and forest inventory data. Computers and Electronics in Agriculture 37:37–56.

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

Repola, J. 2008. Biomass equations for birch in Finland. Silva Fennica 42(4): 605–624.

Repola, J. 2009. Biomass equations for Scots pine and Norway spruce in Finland. Silva Fennica 43(4): 625–647.

Repola, J., Ojansuu, R. & Kukkola, M. 2007. Biomass functions forScots pine, Norway spruce and birch in Finland. Metlan työraportteja/Working Papers of the Finnish Forest Research Institute 53. 28 p. Available at: http://urn.fi/URN:ISBN:978-951-40-2046-9. [Cited 14 Mar 2018].

Suomen pinta-ala kunnittain (The area of Finland by municipalities) 1.1.2011 2011. National Land Survey of Finland, Helsinki. Available at: http://www.maanmittauslaitos.fi/sites/default/files/pinta-alatkunnittain_2011.pdf. [Cited 17 Sep 2014]. (In Finnish).

Tomppo, E. 1990. Satellite Image-Based National Forest Inventory of Finland. The Photogrammetric Journal of Finland 12: 115–120.

Tomppo, E. 1991. Satellite Image-Based National Forest Inventory of Finland. In: Proceedings of the symposium on Global and Environmental Monitoring, Techniques and Impacts, September 17-21, 1990 Victoria, British Columbia Canada. International Archives of Photogrammetry and Remote Sensing 28: 419–424.

Tomppo, E. 1992. Satellite image aided forest site fertility estimation for forest income taxation. Acta Forestalia Fennica 229, 70 p.

Tomppo, E. 1996. Multi-source National Forest Inventory of Finland. In: Päivinen, R., Vanclay, J. & Miina, S. (ed.). New Thrusts in Forest Inventory. Proceedings of the subject group S4.02-00 'Forest Resource Inventory and Monitoring' and subject group S4. 12-00 'Remote Sensing Technology', vol. 1. IUFRO XX World Congress, 6–12 Aug. 1995, Tampere, Finland. European Forest Institute, Joensuu, Finland. p. 27–41.

Tomppo, E. 2006a. The Finnish National Forest Inventory. In: Kangas, A. & Maltamo, M. (ed.). Forest inventory. Methodology and applications. Managing Forest Ecosystems Vol 10. Springer, Dordrecht, The Netherlands. p. 179–194.

Tomppo, E. 2006b. The Finnish multi-source national forest inventory - small area estimation and map production. In: Kangas A., Maltamo M (ed.). Forest inventory. Methodology and applications, Managing Forest Ecosystems Vol 10. Springer, Dordrecht, The Netherlands p. 195–224.

Tomppo, E. 2009a. The Finnish National Forest Inventory. In: McRoberts, R.E., Reams, G.A., Van Deusen, P.A. & McWilliams, W.H. (ed.). Proceedings of the Eighth Annual Forest Inventory and Analysis Symposium, Monterey, California, October 16–19, 2006. USDA Forest Service General Technical Report WO-79. p. 39–46.

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. 1998b. Kunnittaiset metsävaratiedot 1990–94. Folia Forestalia 4B/1998: 619–839. (In Finnish).

Tomppo, E., Goulding, C. & Katila, M. 1999. Adapting Finnish multi-source forest inventory techniques to the New Zealand preharvest inventory. Scandinavian Journal of Forest Research 14: 182–192.

Tomppo, E., Korhonen, K.T., Heikkinen, J. & Yli-Kojola, H. 2001. Multi-source inventory of the forests of the Hebei Forestry Bureau, Heilongjiang, China. Silva Fennica 35(3): 309–328.

Tomppo, E., Olsson, H., Ståhl, G., Nilsson, M., Hagner, O. & Katila, M. 2008a. Combining National Forest Inventory Field Plots and Remote Sensing Data for Forest Databases. Remote Sensing of Environment 112: 1982–1999.

Tomppo, E., Haakana, M., Katila, M. & Peräsaari J. 2008b. Multi-Source National Forest Inventory Methods and Applications. Springer. Series: Managing Forest Ecosystems, Vol. 18 2008, XIV. 374 p.

Tomppo, E., Haakana, M., Katila, M., Mäkisara, K. & Peräsaari, J. 2009a. The multi-source National Forest Inventory of Finland - methods and results 2005. Metlan työraportteja/Working Papers of the Finnish Forest Research Institute 111. 277 p. Available at: `http://urn.fi/URN:ISBN:978-951-40-2151-0`. [Cited 14 Mar 2018].

Tomppo, E., Gagliano, C., De Natale, F., Katila, M. & McRoberts, R.E. 2009b. Predicting categorical forest variables using an improved k-Nearest Neighbour esimation and Landsat imagery. Remote Sensing of Environment 113: 500–517.

Tomppo, E., Heikkinen, J., Henttonen, H.M., Ihalainen, A., Katila, M., Mäkelä, H., Tuomainen, T. & Vainikainen, N. 2011a. Designing and conducting a forest inventory - case: 9th National Forest Inventory of Finland. Springer, Managing Forest Ecosystems 21. 207 p.

Tomppo, E., Katila, M., Mäkisara, K. & Peräsaari, J. 2012. The Multi-source National Forest Inventory of Finland –methods and results 2007. Metlan työraportteja / Working Papers of the Finnish Forest Research Institute 227. 233 p.

Tomppo, E., Katila, M., Mäkisara, K. & Peräsaari, J. 2013. The Multi-source National Forest Inventory of Finland –methods and results 2009. Metlan työraportteja / Working Papers of the Finnish Forest Research Institute 273. 216 p.

Tomppo, E., Katila, M., Mäkisara, K. & Peräsaari, J. 2014. The Multi-source National Forest Inventory of Finland –methods and results 2011. Metlan työraportteja / Working Papers of the Finnish Forest Research Institute 319. 224 p.

Walsh, T.A. & Burk, T.E. 1993. Calibration of satellite classifications of land area. Remote Sensing of Environment 46: 281–290.

Ympäristöhallinnon ladattavat paikkatietoaineistot (The downloadable GIS data) 2018. SYKE, Helsinki. Available at: `http://www.syke.fi/avoindata`. [Cited 14 Mar 2018]. (In Finnish).

Appendix

The following tables can be retrieved from the URL `http://urn.fi/URN:ISBN:978-952-326-712-1`.

luke.fi

Natural Resources Institute Finland
Latokartanonkaari 9
FI-00790 Helsinki, Finland
tel. +358 29 532 6000