Remote Sensing of Environment 113 (2009) 500–517

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Predicting categorical forest variables using an improved k-Nearest Neighbour estimator and Landsat imagery

Erkki O. Tomppo [a,*,1], Caterina Gagliano [b,2], Flora De Natale [b,3], Matti Katila [a,4], Ronald E. McRoberts [c,5]

[a] Metla – Finnish Forest Research Institute, Unioninkatu 40 A, FIN-00170 Helsinki Finland
[b] CRA – MPF – National Council for Agricultural Research – Forest Monitoring and Planning Research Unit, Piazza Nicolini 6, 38100 Villazzano Trento – Italy
[c] Northern Research Station, U.S. Forest Service, 1992 Folwell Avenue, Saint Paul, Minnesota 55108 USA

ARTICLE INFO

Article history:
Received 21 December 2007
Received in revised form 22 May 2008
Accepted 24 May 2008

Keywords:
National Forest Inventory
ik-NN
Categorical variables
Genetic algorithm

ABSTRACT

The k-Nearest Neighbour (k-NN) estimation and prediction technique is widely used to produce pixel-level predictions and areal estimates of continuous forest variables such as area and volume, often by sub-categories such as species. An advantage of k-NN is that the same parameters (e.g., k-value, distance metric, weight vector for the feature space variables) can be used for all variables, whether continuous or categorical. An obvious question is the degree to which accuracy can be improved if the k-NN estimation parameters are tailored for specific variable groups such as volumes by tree species or categorical variables. We investigated prediction of categorical forest attribute variables from satellite image spectral data using k-NN with optimisation of the weight vector for the ancillary variables obtained using a genetic algorithm. We tested several genetic algorithm fitness functions, all derived from well-known accuracy measures. For a Finnish test site, the categorical forest attribute variables were site fertility and tree species dominance, and for an Italian test site, the variables were forest type and conifer/broad-leaved dominance. The results for both test sites were validated using independent data sets. Our results indicate that use of the genetic algorithm to optimize the weight vector for prediction of a single forest attribute variable had a slight positive effect on the prediction accuracies for other variables. Errors can be further decreased if the optimisation is done by variable groups.

1. Introduction

Current, accurate, and detailed assessments of forest resources are crucial for forest management, international reporting, and sustainability assessments. The national forest inventories (NFI) conducted in Europe, North America, and elsewhere are regarded as important sources of comprehensive information for these purposes. Because complete, enumerative inventories are prohibitively expensive, NFIs have traditionally sampled populations of interest and calculated estimates of forest resources using probability-based estimators. Satellite imagery has been demonstrated to be a particularly useful source of ancillary data that can be used to decrease NFI sampling intensities, increase the precision of the probability-based estimates, and complement tabular estimates with maps depicting the spatial distribution of forest resources (McRoberts, 2006; McRoberts et al., 2002, 2005; Reese et al., 2003; Tomppo, 1991; Tomppo & Halme, 2004).

Forest inventory programs observe a combination of continuous variables such as volume, basal area, and tree density and categorical variables such as forest/non-forest, forest type, and site fertility. Mapping applications typically entail constructing statistical models of relationships between forest attribute variables and ancillary variables including satellite image spectral variables, and then predicting values for the attribute variables for image pixels. To preserve consistency among predictions of diverse inventory variables, multivariate statistical modelling approaches are necessary. Unfortunately, parametric multivariate statistical methods for constructing models generally assume that values of the forest attribute variables follow Gaussian distributions, an assumption that is violated for many inventory variables. Thus, multivariate, non-parametric, nearest neighbour techniques have enjoyed increasing popularity for forest inventory mapping and estimation applications (McRoberts et al., 2007; Muinonen & Tokola, 1990; Tomppo, 1990).

For nearest neighbours techniques, the set of image pixels that contain centres of inventory plots for which forest attribute variables have been observed is designated the *reference set*; the set of image pixels for which predictions of these variables are desired is

* Corresponding author.
E-mail addresses: erkki.tomppo@metla.fi (E.O. Tomppo), caterina.gagliano@entecra.it (C. Gagliano), flora.denatale@entecra.it (F. De Natale), matti.katila@metla.fi (M. Katila), rmcroberts@fs.fed.us (R.E. McRoberts).
[1] Tel.: + 358 10 211 2170.
[2] Tel.: + 39 0461 381139.
[3] Tel.: + 39 06 61571037.
[4] Tel.: + 358 10 211 2159.
[5] Tel.: + 1 649 5174.

0034-4257/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2008.05.021

designated the *target set*; and the space defined by the ancillary variables is designated the *feature space*. Observations of the feature space variables are available for all reference set and all target set pixels. Feature space variables are often derived from satellite image spectral data, although topographic, climatic, and soil variables are also used. Predictions for forest attribute variables for target set pixels are calculated as linear combinations of values for the reference set pixels that are nearest or most similar in the feature space with respect to a selected distance metric. A variety of distance metrics have been used including Euclidean or weighted Euclidean distance (Franco-Lopez et al., 2001; McRoberts et al., 2007; Reese et al., 2003; Tokola et al., 1996; Tomppo, 1996; Tomppo & Halme, 2004), Mahalanobis distance, and metrics based on canonical correlation analysis (LeMay & Temesgen, 2005; Temesgen et al., 2003) and canonical correspondence analysis (Ohmann & Gregory, 2002).

Nearest neighbours techniques have been used for a wide variety of forest inventory applications including imputing values for missing observations in inventory databases (LeMay & Temesgen, 2005; Moeur & Stage, 1995; Temesgen et al., 2003), mapping forest attributes using the spectral values of satellite imagery (Franco-Lopez et al., 2001; McRoberts et al., 2007), increasing the precision of probability-based forest inventory estimates (Baffetta et al., 2007; McRoberts et al., 2002; Tomppo, 1991, 1996), and model-based inference (Magnussen et al., 2009-this issue; McRoberts et al., 2007). The majority of studies using forest inventory data have focused on predicting continuous variables: volume by species, age, height, basal area, mean diameter in Finland (Tomppo, 2006; Tomppo & Halme, 2004); age, height, and volume in Sweden (Holmström & Fransson, 2003; Nilsson, 1997; Reese et al., 2003); volume, stocking, diameter, height, and basal area in Ireland (McInerney et al., 2005); proportion forest area, basal area, volume, and tree density in the USA (McRoberts et al., 2007); area and volume by stand age, diameter, and tree species dominance classes in China (Tomppo et al., 2001), volume and volume by species in Germany (Diemer et al., 2000), volumes of log products in New Zealand (Tomppo et al., 1999); and basal area, leaf area index, and volume by species in Italy (Bertini et al., 2007; Chiavetta et al., 2007; Maselli et al., 2005). Nearest neighbours techniques have been used operationally by the Finnish NFI since 1990 (Katila & Tomppo, 2001; Tomppo, 1990). The Swedish NFI has been constructing statistical and map products using nearest neighbours techniques on a 5-year basis since 2001 and is currently developing an annual product. Swedish government agencies such as the Swedish National Forestry Board, County Administrative Boards, and the Environmental Protection Agency use the products for small area estimation, identifying valuable areas in forest reserves, and wildlife habitat evaluation and modelling (Reese et al., 2003; Tomppo et al., 2008).

Although reports of predicting categorical forest inventory variables using nearest neighbours techniques are fewer, some have been published in recent years: forest type and land cover in the USA (Franco-Lopez et al., 2001; Haapanen et al., 2004), forest type in Austria (Koukal et al., 2007), species groups in Norway (Gjertsen, 2005), and classes of site variables in Finland (Tomppo, 2006).

Use of nearest neighbours techniques has not been restricted to forest inventory or even natural resources applications. However, applications in other disciplines frequently emphasize prediction of categorical variables and combine nearest neighbour techniques with genetic algorithms (GA) for selecting weights for the feature space variables (e.g., He et al., 2000; Romualdi et al., 2003). GAs mimic Darwinian natural selection and evolution (Goldberg, 1989; Holland, 1975). They begin with a large, initial population of randomly generated sets of weights. In each generation, individuals are evaluated with respect to their fitness, and those with greater fitness are more likely to be selected for the next generation. In subsequent generations, individuals are combined and modified to produce offspring which are then evaluated for fitness. Typically, the individuals in the population converge to a near-optimal set of weights over multiple generations. Applications of combinations of nearest neighbours techniques and GAs range over a wide diversity of disciplines: e.g., prediction of classes of water molecules with respect to protein–water interactions (Raymer et al., 1997), prediction of classes of soil samples with respect to agricultural growth environments (Punch et al., 1993), prediction of classes of children with developmental delay into individual syndromes based on facial characteristics (Boehringer et al., 2006), and recognition of handwritten digit strings (Oliveira et al., 2003).

Tomppo and Halme (2004) used a k-Nearest Neighbour (k-NN) technique with a GA to optimise selection of weights for spectral feature space variables to predict volume by species for Landsat Thematic Mapper (TM) pixels. They characterized k-NN with feature space variables optimised using a GA as *improved k-NN* (ik-NN). Variables in the optimisation function included mean square errors and mean biases for volume predictions by species for reference set pixels. Although implementation of the GA was computationally intensive, the resulting feature space weights reduced mean biases by factors ranging from 0.69 to nearly 1.00. The optimal weight vector was used to predict all forest attribute variables and to calculate plot expansion factors.

Categorical inventory variables have received increased recent attention for forest inventory applications. The Ministerial Conference on the Protection of Forests in Europe (MCPFE) (MCPFE, 1997, 2003a,b) includes area by forest type as an indicator for a criterion related to maintaining forest resources, and the Montréal Process (MPCI, 2007) includes the same indicator for a criterion related to maintaining ecosystem biodiversity. Forest type has been widely used as an inventory variable in North America and in Mediterranean and central European countries; it is used less in Nordic countries, although it was added recently to accommodate international reporting requirements. In addition, Action E43 (Harmonisation of the national forest inventories of Europe) (COST E43, 2007) of the Cooperation in the fields of Science and Technology in Europe programme has selected forest type as a core variable for biodiversity assessment. Finally, commercial forest enterprises are frequently interested in estimates of volume by tree species composition, forest type, and site fertility classes to support forest management planning activities such as planning regeneration to sustain wood production.

The overall objectives of the study were technical in nature: (1) to investigate distance metrics for the k-NN technique for use in prediction of the classes of land characteristics for Landsat TM and ETM+ pixels, (2) to investigate use of ik-NN, and (3) to compare the results obtained for the Finnish and Italian test sites using different options. Two categorical variables were selected for each country: dominant tree species and site fertility class for Finland and forest type and conifer/broad-leaved dominance for Italy.

2. Material

2.1. Finnish data

2.1.1. Satellite imagery

The cloud-free portions of two adjacent Landsat 7 ETM+ images, path 186, rows 16 and 17, from June 10, 2000, were used. Both images were rectified to the national coordinate system by fitting a second order regression to the coordinates of 50 control points identified from both satellite images and base maps. The nearest neighbour method was used to re-sample images to a pixel size of 25-m×25-m. The absolute values of the residuals in the model used for geo-rectification ranged from 0.3 pixels to 0.6 pixels. For predicting forest attribute variables, data for spectral bands 1–5, 7–9 and all ratios of these bands were used. Landsat 7 ETM+ imagery includes the 60-m×60-m resolution thermal band 8 and the 15-m×15-m panchromatic band 9 that are not included with Landsat 5 TM imagery. ETM+ band 6 includes the same information as band 8 but is calibrated for water and, therefore, was not used for this study. The thermal band can be expected to provide information on site fertility, and thus also on tree species composition. Band weights obtained using

GA optimisation are expected to accommodate discrepancies between field plot and pixel information caused by the coarser spatial resolution of this band. Such discrepancies can be interpreted as measurement error. The method reported by Tomppo (1992) was used with a digital elevation model to remove the effects of the variation of the angle between sun illumination and terrain normal (see Section 3.1).

2.1.2. Field data from 9th Finnish NFI

The study area is located in the eastern part of Central Finland and covers parts of the North Karelia and South Savo forestry centres, approximately between latitudes 61°10′ N, 63°95′ N and longitudes 27°20′ E, 31°10′E (Fig. 1). The study area corresponded to the cloud-free and non-shadow parts of the satellite images and included total land area of 2.22 million hectares of which 1.97 million hectares were forestry land. The remaining area was arable land and built-up areas. The cumulative daytime temperature of the growing season ranges between 1000 °C and 1300 °C. Water covers 23% of the study area. The main tree species are Scots pine (*Pinus sylvestris* L.), Norway spruce (*Picea abies* (L.) Karst.) and birch (*Betula* spp.), with some aspen (*Populus tremula* L.) and alder (*Alnus* spp.).

Field data from the 9th Finnish NFI (NFI9) acquired between June 1 and August 31, 2000 from a systematic cluster sampling design were used. In North Karelia, a cluster includes either 18 temporary or 14 permanent field plots located along a rectangular tract with a spacing of 300 m (Fig. 2). In South Savo, the field plot cluster was a half rectangle and included either 14 temporary or 10 permanent plots with a spacing of 250 m. Using Finnish NFI definitions, which deviate somewhat from those of the United Nations Food and Agriculture Organisation (FAO), forestry land is divided into forest land (FL), other wooded land (OWL), forestry roads and waste land. The combination of FL and OWL is denoted FOWL. The two image scenes included 10,860 NFI9 field plots of which 8494 were on land; 7541 were on forestry land; 7322 were on FOWL; 7143 were on FL; 179 were on OWL; and 49 were on forestry roads. Mean growing stock volume on FOWL is 102.0 m^3/ha in North Karelia, 136.2 m^3/ha in South Savo, and ranges from 0 to 530 m^3/ha on field plots.

We used only plots entirely on forestry land that were also at least 20 m from the nearest stand boundary, thus excluding plots located partially on non-forestry land (Tables 1 and 2). This practice is used in the operational Multi-Source NFI (MS-NFI) (Tomppo, 2006) when using the GA to calculate spectral variable weights. The field plots were geolocated using global positioning system (GPS) receivers. Trees located on plots on FOWL were measured using PPS-sampling with the inclusion probability of a tree proportional to its cross-sectional area at a height of 1.3 m for basal area factor 2. A maximum distance of 13.45 m is, however, used with basal area factor 2 to avoid measuring trees far away from the centre point. Sampling of plot trees and differences in plot and satellite image pixel shapes and sizes means that field plot and spectral data are not in exact correspondence. These effects can be considered a form of measurement error when estimating volume. For this study, the predicted variables are assessed at stand-level with stand sizes varying from 0.5 to a few hectares. Field plots whose ground attributes had changed between field measurement and image acquisition dates were identified by graphing volume against spectral values and removed from the data sets. The number of such plots was small because the satellite image and field measurements were from the same season.

Information from a coarse scale map of Finland was also used for ancillary information. The map was constructed in three steps: (1) the mean of observed forest attribute variables for all plots in the same NFI cluster was calculated; (2) the cluster mean for the geographically closest NFI plot was imputed to each 1-km×1-km cell; and (3) moving average filters with window sizes of 11-km×11-km, 20-km×20-km, and 25-km×25-km were applied in succession to the 1-km×1-km cells. The variables used were the same as in the operative NFI, i.e., total volume and volumes by tree species in four tree species groups on FOWL (Tomppo & Halme, 2004). In operative applications of the Finnish MS-NFI, a topographic map of the country obtained from the National Land Survey of Finland is used to stratify all image pixels and plots on forestry land into mineral soils and peatland soils (spruce mires, pine mires, open bogs, fens). We used only field plots for this study, and stratification was done on the basis of field plot data.

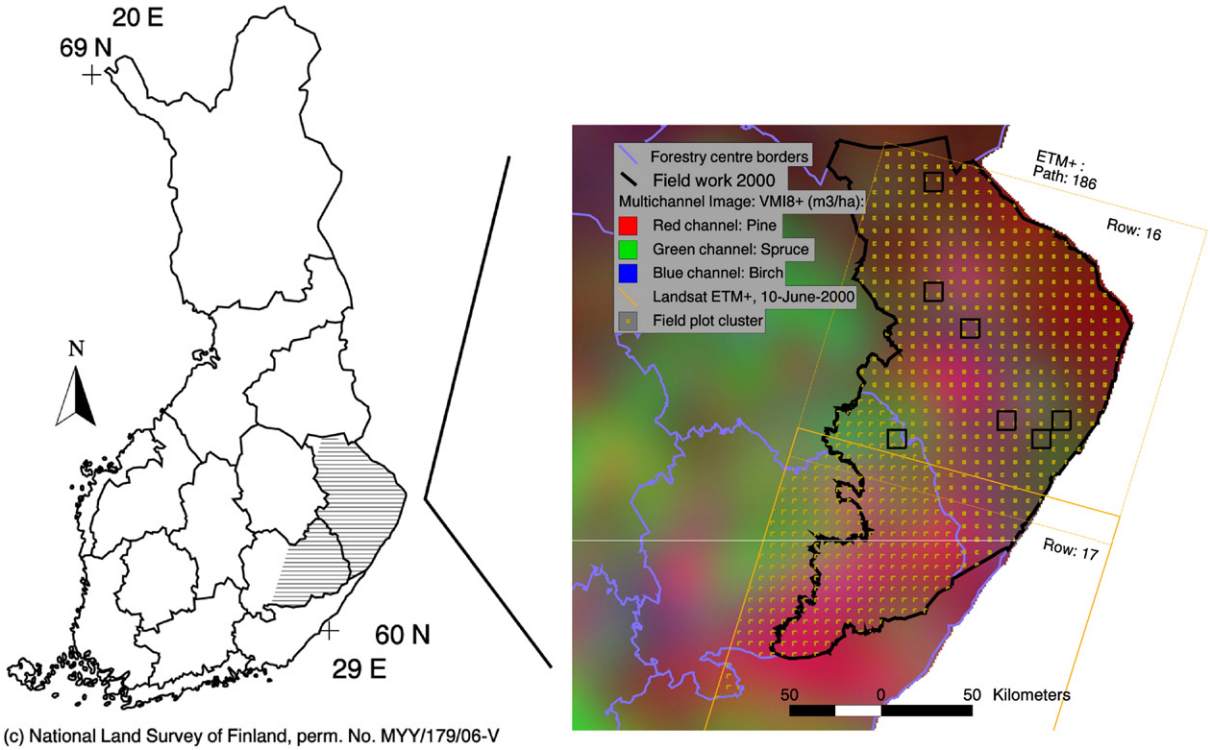

Fig. 1. Finnish study area with NFI field plot clusters displayed on the coarse scale volume maps, seven test units of independent field plot data, and areas covered by the two Landsat 7 ETM+ images.

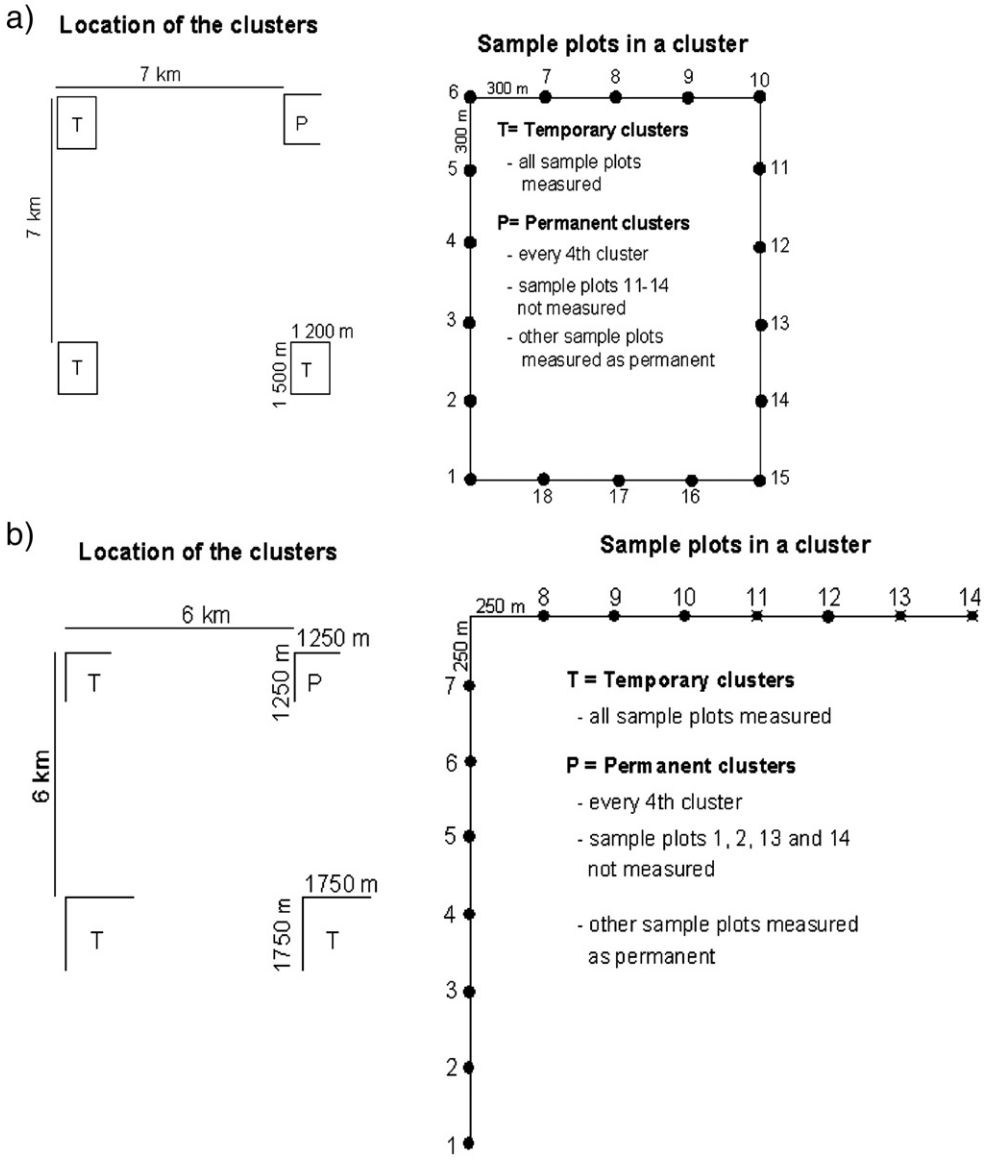

Fig. 2. Sampling designs for Finnish NFI in the study area, a) North Karelia, b) South Savo.

2.1.3. Independent field data

Independent field plot data were used to evaluate the pixel-level predictions. The data were obtained for seven test units, each approximately 100 km^2 in size, that were widely dispersed within one of the ETM+ scenes (Fig. 1). In each test unit, field plots with centres on a 400-m × 300-m systematic design had been measured in year 2000. For the seven areas, 4892 plots were on FOWL (Katila & Tomppo, 2006), but of these only the 1411 plots on mineral soil and the 387 plots on peatland which were at least 20 m from the nearest stand boundary were selected for the study. The mean volume estimates on FOWL were in the range 72–150 m^3/ha for the seven test units. The proportions of the selected study plots assigned to particular fertility classes and the proportions of tree species represented on these plots are shown in Tables 1 and 2.

Table 1
Distributions of Finnish reference and independent target plots with respect to site fertility class

Site fertility class	Soil stratum							
	Mineral soil				Peatland			
	Reference		Target		Reference		Target	
	Count	%	Count	%	Count	%	Count	%
HRS	57	3.0	12	0.8	16	2.3	4	1.1
HRHF	370	19.5	223	15.7	64	9.3	22	5.7
MF	985	51.8	870	61.2	165	23.9	113	29.2
SX	441	23.2	285	20.1	192	27.8	130	33.6
X	35	1.8	28	2.0	205	29.7	107	27.6
B	0	0.0	1	0.1	48	7.0	11	2.8
R	13	0.7	2	0.1	0	0.0	0	0.0
Total	1901	100.0	1421	100.0	690	100.0	387	100.0

Table 2
Distributions of Finnish reference and independent target plots with respect to tree species groups

Species group	Soil stratum							
	Mineral soil				Peatland			
	Reference		Target		Reference		Target	
	Count	%	Count	%	Count	%	Count	%
Open	25	1.3	24	1.7	103	14.9	30	7.8
Pine	1,138	59.9	814	57.3	447	64.8	273	70.5
Spruce	412	21.7	479	33.7	61	8.8	55	14.2
Birch	252	13.3	98	6.9	79	11.5	29	7.5
Other broadleaved	74	3.9	6	0.4	0	0.0	0	0.0
Total	1901	100.0	1421	100.0	690	100.0	387	100.0

A detailed species list is given Table 3.

2.1.4. Forest attribute variables

For Finland, we selected two stand-level categorical forest attribute variables, site fertility class and dominant tree species. Both are core variables in forest management planning and are included in the NFI. Site fertility describes wood production potential, and its determination as used in the NFI is based on the composition of ground vegetation on the site, the species composition of the growing stock of the site, and the assumed vegetation composition at the mature development state. The classification was created in the beginning of the 20th century and further developed for forestry and NFI purposes (Cajander, 1926; Lehto & Leikola, 1987). The site fertility classes are as follows:

1. *Herb rich sites (HRS)*: mineral soils and eutrophic swamps, fens and corresponding drained peatlands.
2. *Herb rich heath forests (HRHF)*: mineral soils and mesotrophic swamps, fens and corresponding drained peatland forests.
3. *Mesic heath forest (MF)*: mineral soils and meso-oligotrophic natural and drained peatlands.
4. *Sub-xeric heath forests (SX)*: mineral soil forests and oligotrophic natural and drained peatlands.
5. *Xeric heath forests (X)*: mineral soil forests and oligo-ombrotrophic natural and drained peatlands.
6. *Barren heath forests (B)*: mineral soil forests and *Sphagnum fuscum* dominated (ombrotrophic) natural and drained peatlands.
7. *Rocky forests (R)*: rocky and sandy soils and alluvial lands which are typical in coastal regions.
8. *Summit forests (S)*: summit (hilltops) and fjeld (hill) forest, only in North Finland.

Dominant tree species is assessed by tree storeys, although generally not for more than two storeys. The first phase is to assess whether a storey is coniferous or broad-leaved tree species dominated. In cases of equal proportions, the assessment is based on future treatments and predicted favoured tree species. For a coniferous storey, the dominant tree species is a coniferous species; similarly for a broad-leaved tree species dominated storey. The assessment criteria for tree species dominance varies by the development class of a storey. For the young thinning class and older development classes, tree species proportion is assessed on the basis of basal area. For the seedling development classes, tree species proportion is assessed on the basis of the number of stems capable of further development. The dominant tree species in coniferous dominant storeys is the coniferous species with the highest proportion of basal area, except for seedling stands where the assessment is based on the number of seedlings capable of development. Similar determinations are made for broad-leaved tree species dominant storeys. The dominant species of the dominant storey is also the dominant species of the stand.

The three most important storeys are Dominant, Over-storey and Under-storey. Tree species proportion by stand is assessed in greater detail for the first two storeys and for the third only when it is capable of further development. For all development classes other than Non-stocked regeneration area, the dominant and second tree species and their proportions are always reported. For the Non-stocked regeneration area class, tree species is not determined.

The tree species coding used in the initial analyses with the reference data to investigate the technique was the same as for the operational MS-NFI and was designated Species Grouping 1 (Table 3). Because the areas of stands dominated by some species were very small in Species Grouping 1, Species Grouping 2 was used for the validation analyses with the independent data (Table 3).

Table 3
Species groups in Finland

Species grouping 1	Species grouping 2
Open regeneration area, or otherwise treeless area such as non-productive forest land, open bog or rock	Open regeneration area, or otherwise treeless area such as non-productive forest land, open bog or rock
Scots pine *(Pinus sylvestris)*	Scots pine *(Pinus sylvestris)* and other coniferous species
Other coniferous species such as:	
Lodgepole pine *(Pinus contorta)*,	
Stone pine *(Pinus sembra)*,	
Larch *(Larix sp.)*,	
Common juniper *(Juniperus communis)*	
Silver birch *(Petula pendula)*	Silver birch *(Petula pendula)* and
Downy birch *(Petula pubescens)*	Downy birch *(Petula pubescens)*
Norway spruce *(Picea abies)*	Norway spruce *(Picea abies)*
European aspen *(Populus tremula)*	Other broad-leaved tree species
Grey alder *(Alnus incana)*	
Black alder *(Alnus glutinosa)*	
Ash *(Sorbus aucuparia)*	
Goat willow *(Salix caprea)*	
Other broad-leaved tree species such as:	
Bay willow *(Salix pentandra)*,	
Fluttering elm *(Ulmus laevis)*,	
Wych elm, *(Ulmus glabra)*,	
Small-leaved lime *(Tilia cordata)*,	
Poplar *(Populus sp.)*,	
European ash *(Fraxinus excelsior)*,	
Pedunculate Oak *(Quercus robur)*,	
Bird cherry *(Prunus padus*,	
Norway maple *(Acer platanoides)*	

2.2. Italian data

2.2.1. Satellite imagery

A Landsat 5 TM image for path 192, row 28, dated 19 June 2000, was used as the source of spectral data. Only 2.4% of the image area was covered by the clouds. The image was geo-referenced using a first order nearest neighbour method and geometrically rectified using a digital elevation model derived from elevation curves at 10-m intervals. A pixel size of 30 m×30 m was used, and a positional error of less than one pixel was obtained using 50 ground control points and a first degree polynomial. The first order nearest neighbour method was chosen to preserve the original spectral values (Manual of Remote Sensing, 1983). The image was topographically corrected by using a digital terrain model and a cosine function in a similar manner as was done for the Finnish spectral data (Tomppo, 1992). No atmospheric correction was made. Forest attribute variables were predicted using data for spectral bands 1–5 and 7 and all possible band ratios. Although the geographical extent of the study area is smaller than that of the Finnish site, variation in vegetation types is high,

2.2.2. Data from the 2nd Italian NFI

The Italian study area is located in the Eastern Italian Alps and corresponds to the Province of Trento (46°04′ N, 11°08′ E) (Fig. 3). The study area has complex mountainous topography with elevations ranging between 100 m and 3500 m above sea level. The climate is typically Alpine but with some Mediterranean features in the southern part of the study area. Average annual temperature is 10–11 °C, and average annual precipitation is approximately 1200 mm. Total land area is approximately 620,000 ha with 65% covered by FOWL. Using FRA-2000 definitions (FAO, 2000), the most recent NFI results (INFC, 2007) indicate 375,402 ha of FL and 32,129 ha of OWL. High forest, which reproduces from seeds, prevails, although coppices are found on approximately 14% of FL. Coniferous stands are dominant with spruce stands covering more than 36% of forestry land. Larch (*Larix decidua*), black pine *(Pinus nigra)*, Scots pine *(Pinus sylvestris)* and fir (*Abies alba*) are also very widespread. Broad-leaved species usually are found in mixed stands in the southern part of the study area or in valleys with spruce-fir-beech and hornbeam-Scots pine stands being the most common. Mixed stands are well represented due to the promotion of tree species diversity by forest management for the last 40 years.

The field data were obtained from the 2nd Italian National Inventory of Forests and Forest Carbon Sinks (INFC) and were collected in 2004 and 2005. The 2nd INFC adopted a sampling design featuring three-phase sampling for stratification which is intended to produce reliable statistics for each of the 21 Italian administrative districts. In the study area, which corresponds to one of the 21 districts, 6200 first-

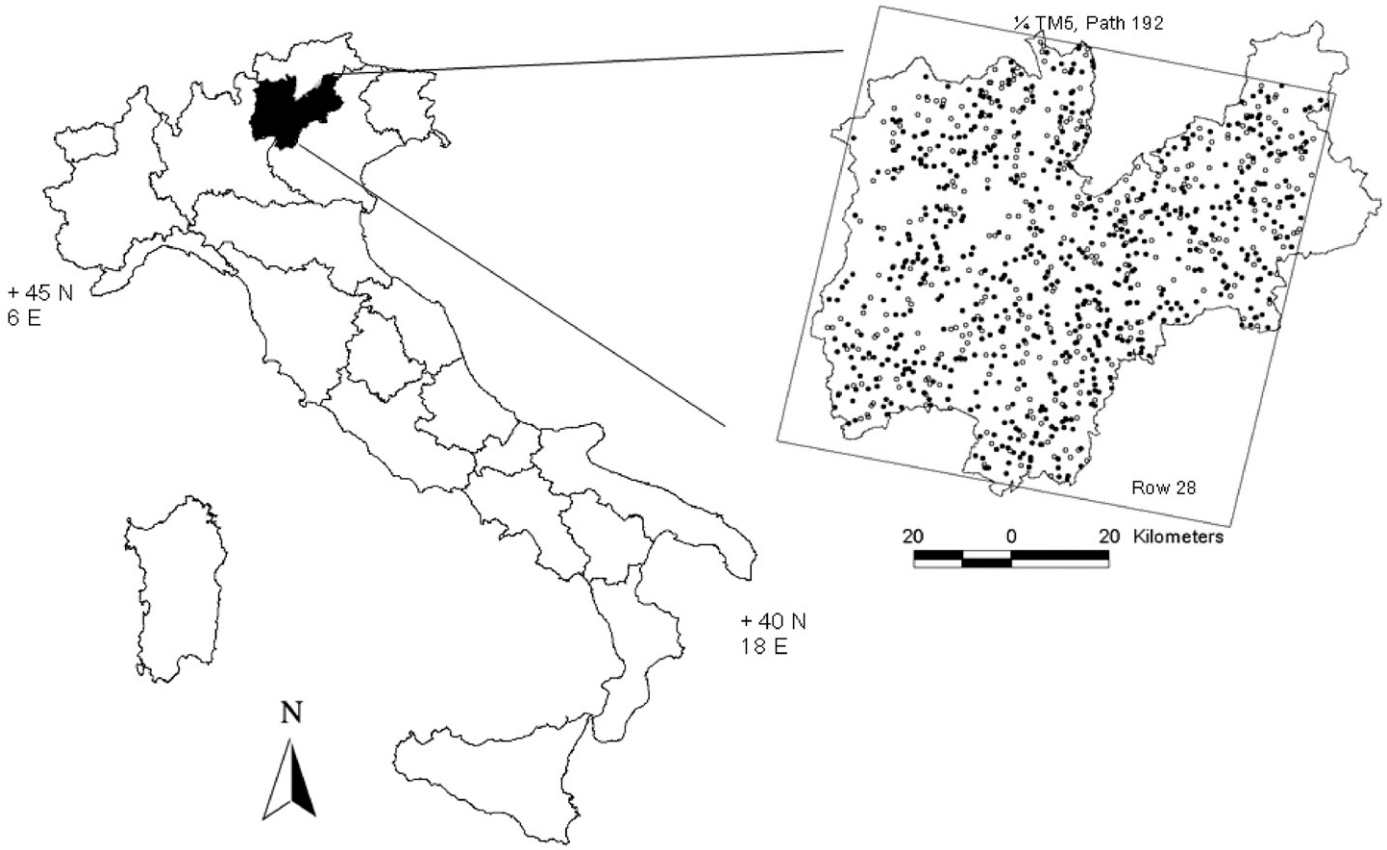

Fig. 3. Italian study area covered by a quarter scene of Landsat TM5 image with reference plots (dark and closed points) and target plots (light and open points).

phase sampling points were distributed using systematic unaligned sampling characterized by random selection of a point in each cell of a 1-km × 1-km grid. All points were photo interpreted and assigned to broad land use/land cover classes with FL and OWL included in the same class and with an *uncertain* class. In the second phase, a sub-sample of approximately 1100 points from the first-phase sample was randomly selected from FOWL and *uncertain* classes according to their proportions of the total district area. The points were located in the field using GPS receivers and served as centres for circular plots of radius 25 m. Field crews observed the plots to distinguish FL and OWL and to assess forest type and other categorical stand-level variables (Tabacchi et al., 2007). In the third phase, the field crews made quantitative measurements of tree and stand attributes on a subsample of approximately 30% of the second phase sampling points (Fattorini et al., 2006).

Forest attribute data used for the study were obtained from the second-phase assessments of 956 sampling points. Some points were excluded because of missing data, and some were excluded because they fell on non-forest land as indicated by a mask provided by the Province of Trento. Conifer/broad-leaved dominance was assigned on the basis of dominant crown cover using three classes, conifer, broad-leaved, and mixed, and the variable was denoted CBM. Forest type was also assigned on the basis of the tree species or species group with the greatest proportion of crown cover. In the study area, 13 forest types were observed.

The time interval between image acquisition in 2000 and NFI field measurements in 2004–2005 was not expected to produce serious discrepancies between field and spectral data. Forest area and species composition changes occur only slowly in temperate regions, so that intervals of five years are not expected to produce substantial ground level changes, except following large disturbances such as widespread fires, avalanches, landslides, or extensive cutting or planting. During the period between image acquisition and NFI data collection, none of these events occurred in the study area. Also, the second phase field survey conducted in 2004–2005 was a sub-sample of the first phase points which were classified using 1999 orthophotos. The second phase field survey confirmed the land use class assigned by photo interpretation for 99% of the second phase points.

The classification scheme adopted for forest types is intended for all Italian forest conditions and is based on visual recognition of prevailing tree species without regard to site characteristics, stand structure or under-storey composition. However, this scheme is too detailed for satellite image-based prediction because some forest type classes typically occurring together are similar with respect to both ecological requirements and spectral properties, and are therefore difficult to distinguish using satellite imagery. Therefore, the 13 classes were aggregated into seven classes on the basis of their ecological similarities (Table 4). The new classification scheme represents a compromise between accuracy gain and information loss.

Table 4
Forest type classes in Italy

13 forest types (NFI data sets)	7 forest types
Larch and stone pine forests (LSP)	Larch and stone pine forests (LSP)
Norway spruce forests (NS)	Norway spruce and fir forests (NSF)
Fir forests (F)	
Scots pine and Mountain pine (SP)	Pine forests (PF)
Black pines (BP)	
Beech forests (B)	Beech forests (B)
Hornbeam forests (H)	Hornbeam forests (H)
Oak forests (OF)	Other broadleaved forests (OB)
Chestnut forests (C)	
Hygrophilous forests (HY)	
Mixed deciduous broadleaved forests (MB)	
Holm oak forests (HO)	
Subalpine shrubland (SS)	Subalpine shrubland (SS)

No independent Italian validation data were available as was the case for the Finnish analyses. Therefore, for the initial Italian analyses which focused on investigating the k-NN technique and the GA, the entire reference set was used with leave-one-out cross validation to determine values of k and the GA operational parameters. However, for the validation analyses, an independent validation target data set was constructed by randomly splitting the reference set into a reduced reference set consisting of two-thirds of the sample data and a target set consisting of one-third of the sample data.

3. Methods

This section consists of four subsections: (3.1) a description of the k-NN technique with attention given to a distance metric used by the Finnish NFI, (3.2) a description of the ik-NN technique with attention given to the formulation and use of a GA for optimizing the weights for feature space variables, (3.3) a description of several accuracy measures with discussion of how they are used in the GA, and (3.4) a description of a discriminant analysis approach to prediction whose results that can be compared to those obtained using the ik-NN technique.

3.1. k-Nearest Neighbours prediction

Forestry applications often require simultaneous and consistent estimates for all core forest variables. Traditional image classification methods cannot easily provide this type of information. The non-parametric k-NN technique mimics forest inventory estimation in that estimates for all variables are obtained simultaneously using the same underlying field data. With this technique, field plot weights, also called plot expansion factors, are calculated as sums of satellite image pixel weights over forestry land or forestry land mask pixels.

With the k-NN technique, the prediction, $\hat{m}_p$, for pixel p for continuous variable, M, is calculated as,

$$\hat{m}_p = \sum_{i=1}^{k} w_{i,p} m_i, \tag{1}$$

where m_i is the value of variable M on reference plot i. The pixel weights, $w_{i,\,p}$, are calculated as,

$$w_{i,p} = \begin{cases} d_{p_i,p}^{-t} \Big/ \sum_{j \in \{i_1(p),\ldots,i_k(p)\}} d_{p_j,p}^{-t}, & \text{if and only if } i \in \{i_1(p), \ldots, i_k(p)\} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where, i is a reference set field plot, p is a target set pixel, p_j is the pixel corresponding to field plot j, and $\{i_1(p),\ldots,i_k(p)\}$ is the set of the k field plots nearest to pixel p in feature space when using distance metric, d, and $t \in [0, 2]$. Both Franco-Lopez et al. (2001) and Katila and Tomppo (2001) reported the effects of different t-values on pixel-level prediction errors.

Tomppo and Halme (2004) developed a distance metric that is now used operationally by the Finnish MS-NFI,

$$d_{p_j,p}^2 = \sum_{l=1}^{n_f} \omega_{l,f}{}^2 \left(f_{l,p_j} - f_{l,p}\right)^2 + \sum_{l=1}^{n_g} \omega_{l,g}{}^2 \left(g_{l,p_j} - g_{l,p}\right)^2, \tag{3a}$$

where

f_{l,p_j}^0 is the original intensity of the spectral band l,
f_{l,p_j} $= f_{l,p_j}^0 / \cos^r(\alpha)$ is the normalised intensity value of the spectral band l,
α is the angle between terrain normal and sun illumination,
r is the power used due to non-Lambertian surface (Tomppo, 1992),
n_f is the number of image variables,
$g_{l,p}$ is the coarse scale prediction of the lth forest variable for pixel p,
n_g is the number of large area forest variables,
p is the target pixel,
p_i is a field plot pixel, and
$\omega_{l,f}$ and $\omega_{l,g}$ are the weights for the lth feature and coarse-scale variable respectively.

The power of cosine, r, is usually estimated from the data, often using a trial and error method, (e.g., Katila & Tomppo, 2001; Tomppo, 1992). Continuous variables predicted in the operative Finnish MS-NFI, include stand age, mean stand diameter, mean stand height, and volumes by tree species and by timber assortment class, as well as some categorical variables such as land classes within forestry land and site fertility.

For this study, the distance metric described by Eq. (3a) was used for categorical variables for the Finnish site (see Subsection 2.1.2), but for the Italian site information on coarse scale variables was not available, so the distance metric was simply,

$$d_{p_j,p}^2 = \sum_{l=1}^{n_f} \omega_{l,f}{}^2 \left(f_{l,p_j} - f_{l,p}\right)^2. \tag{3b}$$

For categorical variables, the mode or median of the predicted classes for the nearest neighbours can be used as a prediction instead of a weighted average as is used for continuous variables. For this study, the mode was selected after investigations of both the mode and median. The predicted class has the greatest sum of the weights, $w_{i,p}$, when added by class over the k nearest neighbours. In theory, equal sums are very rare when weights are used; in fact, the probability is zero if rounding is not considered. In cases of equal sums for two or more classes, one class is selected randomly from among those with the greatest sum. Categorical variables whose classes can be predicted using this approach include land use, site fertility and forest type.

3.2. Improved k-Nearest Neighbours prediction (ik-NN)

We recall first the general ik-NN technique introduced in Tomppo and Halme (2004) and then present the ik-NN technique tailored specifically for categorical variables in Section 3.3. The ik-NN technique is identical to the k-NN technique except the weights, $\boldsymbol{\omega}$, for the feature space variables used in Eqs. (3a) and (3b) are optimised using a 4-step GA. For typical k-NN applications, equal weights are used for all variables.

Step 1. Initialisation. An initial population of weight vectors of size n_{pop} is constructed by randomly generating weight vectors of length $n_f + n_g$. For each weight vector, the value of the fitness function is calculated for Step 2 (Eq. (12), Subsection 3.3).

Step 2. Selection. Pairs of weight vectors in the population are chosen and compared with respect to their fitness. The weight vector with greater fitness is selected for a medipopulation with probability, p_{t1}; both vectors are selected with probability p_{t2}; and the vector with lesser fitness is selected with probability $1 - p_{t1} - p_{t2}$. Weight vectors can be chosen for comparison systematically, randomly, or by other means. Pair-wise comparisons are continued until a medipopulation of size n_{pop} has been selected.

Step 3. Crossover. A new population is formed from the medipopulation by using two successive vectors, designated parents, of the medipopulation to produce two offspring. For the first offspring, the kth element of its weight vector is selected from the first parent with probability p_u and from the second parent with probability $1 - p_u$. The elements of the weight vector for the second offspring are selected in a similar manner, although independently of the selections for the first offspring. From among the two parent and the two offspring weight vectors, the most fit is selected for the next population.

Step 4. Mutation. Each vector of the new population is subject to the possibility of mutation. Each element of each weight vector in the population is selected for mutation with probability p_m. If an element is selected for mutation, it experiences either radical or non-radical mutation. Radical mutation entails subtracting the weight vector element from 1.0 and occurs with probability p_{rm}. Non-radical mutation entails multiplying the weight vector element by 0.8 with probability $0.5(1-p_{rm})$ or 1.2 with probability $0.5(1-p_{rm})$. If the fitness of the mutated vector is better than that of the original, the mutated vector replaces the original in the population. If the fitness of the mutated vector is less than that of the original, the mutated vector replaces the original with probability p_c.

Following Step 4, a new generation of the population has been constructed. The GA then repeats Steps 2–4 until n_{gen} generations have been constructed. Typically, the individuals in the population converge to a single, near-optimal set of weights over multiple generations. For this study, the GA reported by Tomppo and Halme (2004) was used with the same parameter values for both the Finnish and Italian study areas:

n_{gen}	number of generations (50)
n_{pop}	number of weight vectors in one population and number of vectors in the medipopulation (does not have to be the same) (50, 50)
p_u	probability used in uniform crossover (0.75)
p_c	probability of accepting an inferior solution created by mutation (0.5)
p_m	mutation probability (0.05)
p_{rm}	radical mutation probability (0.35)
p_{t1}	probability 1 in selection (0.95)
p_{t2}	probability 2 in selection (0.03).

The GA parameters control how fast and likely a possible global optimum is found and depend very little on the particular feature space variables. Tomppo and Halme (2004) obtained optimized weights $\boldsymbol{\omega}$ used in Eq. (3a) for predicting continuous forest attribute variables using the fitness function,

$$f\left(\boldsymbol{\omega},\boldsymbol{\gamma},\hat{\sigma},\hat{\bar{e}}\right)=\sum_{j=1}^{n_e}\gamma_j\hat{\sigma}_j(\boldsymbol{\omega})+\sum_{j=1}^{n_e}\gamma_{j+n_e}\hat{\bar{e}}_j(\boldsymbol{\omega}), \tag{4}$$

where the γs are user-defined coefficients corresponding to pixel-level mean square errors, $\hat{\sigma}_j$, and mean biases, $\hat{\bar{e}}_j$, for selected forest attribute variables indexed by j, and n_e is the number of forest attribute variables. The selected variables were total volume, volume of pine, volume of spruce, volume of birch, and volume of other broad-leaved tree species for predicting all variables, both continuous and categorical. The pixel-level mean square errors and mean biases were estimated using field plots and leave-one-out cross-validation as $\hat{\sigma}_M=\sqrt{\frac{\sum_{i\in F}\left(\hat{m}_i-m_i\right)^2}{n_F}}$ and $\hat{\bar{e}}_M=\frac{\sum_{i\in F}\left(\hat{m}_i-m_i\right)}{n_F}$ where m_i is the observed value of the variable to be predicted on plot i (e.g. total volume), $\hat{m}_i$ is the prediction for plot i, F designates the set of field plots of which there are n_F. The rationale for selecting measures of prediction errors for these forest attribute variables in the fitness functions was that optimisation with respect to these measures was expected to reduce prediction errors and errors in areal estimates for other forest attribute variables or parameters derived from them. The values of the elements of the coefficient weight vector, γ, were originally determined using an iterative trial and error method when optimizing the values of the elements of the vector, ω, and were constant over the images (Tomppo and Halme, 2004).

This GA approach to obtaining weights $\boldsymbol{\omega}$ for image feature space variables is tailored for prediction of continuous variables by minimising a linear combination of RMSEs and biases of the target variables, particularly volumes and volumes by tree species. A relevant question relates to the degree to which prediction errors for a variable or a group of variables can be reduced if the weights, $\boldsymbol{\omega}$, are tailored for that specific variable or variable group, and in our case for categorical variables which are forest attribute variables for both sites. A key issue is selection of the fitness function to be optimized; in particular, what is the most relevant accuracy measure for categorical response variables, and should optimisation for these response variables be conducted separately or jointly?

3.3. ik-NN tailored for categorical variables and accuracy assessment

Categorical variables require different pixel-level accuracy measures than standard deviation, bias or root mean square error as are used to assess the accuracy of predictions of continuous variables. Therefore, for this study, the fitness function (4) used by Tomppo and Halme (2004) must be modified to include accuracy measures appropriate for categorical variables. Before introducing the modified fitness functions, we first report the accuracy measures used for this study. Many accuracy measures are derived from the error matrix, $\boldsymbol{X}$, with elements, x_{ij}. In our case, the columns of $\boldsymbol{X}$ represent observed classes of the categorical variables, and the rows represent predicted classes. The number of classes is r; the row totals are $x_{i+}=\sum_{j=1}^{r}x_{ij}$; the column totals are $x_{+j}=\sum_{i=1}^{r}x_{ij}$ and $N=\sum_{i=1}^{r}x_{+i}$ is the total number of reference set samples. Widely used accuracy measures for categorical variables based on $\boldsymbol{X}$ are overall accuracy, $OA(\mathbf{X})$,

$$OA(\boldsymbol{X})=\frac{1}{N}\sum_{i=1}^{r}x_{ii}, \tag{5}$$

and the average proportion over all classes of correct pixel predictions, $C(\mathbf{X})$,

$$C(\boldsymbol{X})=\frac{1}{r}\sum_{i=1}^{r}x_{ii}/x_{+i}, \tag{6}$$

Other obvious accuracy measures include omission or commission errors by categories. Another commonly used measure of agreement based on error matrices is the Kappa measure, $\boldsymbol{\kappa(X)}$, (Cohen, 1960; Congalton & Green, 1999),

$$\kappa(\boldsymbol{X})=\frac{\sum_{i=1}^{r}\left(x_{ii}-x_{+i}\,x_{i+}/N\right)}{N-\sum_{i=1}^{r}x_{+i}\,x_{i+}/N} \tag{7}$$

Nishii and Tanaka (1999) reported two new measures, denoted $J_{uni}(\mathbf{X})$ and $J_{pro}(\mathbf{X})$, derived from the Kullback–Leibler information measure as,

$$J_{\mathrm{uni}}(\boldsymbol{X})\equiv\prod_{i=1}^{r}\left(\frac{x_{ii}+0.5}{x_{+i}+0.5}\right)^{1/r}, \tag{8}$$

and

$$J_{\mathrm{pro}}(\boldsymbol{X})\equiv\prod_{i=1}^{r}\left(\frac{x_{ii}+0.5}{x_{+i}+0.5}\right)^{x_{+i}/N}. \tag{9}$$

Nishii and Tanaka (1999) demonstrate the properties of $J_{uni}(\mathbf{X})$ and $J_{pro}(\mathbf{X})$ with simple examples. For example, $J_{uni}(\mathbf{X})$ is sensitive to poor prediction for a single category while J_{pro} is a weighted variant using a priori information.

The four accuracy measures, $OA(\mathbf{X})$ (Eq. (5)), $\boldsymbol{\kappa}(\mathbf{X})$ (Eq. (7)), $J_{uni}(\mathbf{X})$ (Eq. (8)) and $J_{pro}(\mathbf{X})$ (Eq. (9)), were used for two purposes. First, they were tested for use in the fitness functions when applying the GA algorithm to find optimal weights for feature space variables. Second, they were used to assess the accuracy of predictions. Because the same measures were used both for the fitness functions and for evaluating prediction accuracies, as a means of avoiding confusion a subscript f

(e.g., Kappa_f or κ_f) denoting *fitness* is added when a measure is used for the fitness function.

A statistical test of significance is the appropriate method for comparing two values of the same accuracy measure for two different classifications; e.g., a k-NN classification and a ik-NN classification, or two ik-NN classifications using different accuracy measures in the GA fitness function. However, traditional parametric tests cannot be used for this study because the classifications were not based on independent samples. McKenzie et al. (1996) and Foody (2004) recommend and reported two non-parametric resampling tests for dependent samples for testing the statistical significance of the difference between two κ values. The random, or Monte Carlo, permutation test was used to take random samples without replacement from all the possible permutations. For example, let κ_1^o and κ_2^o denote the κ values obtained from classifications of the same data using two different accuracy measures in the fitness function. For each reference plot, the observed class of the response variable was swapped with the observed class for a second randomly selected reference plot. This set of permuted observations was paired separately with the class predictions obtained using the two accuracy measures in the fitness function, and the resulting κ values were denoted κ_1^p and κ_2^p. For each pair of accuracy measures used in the fitness function, this procedure was repeated 1000 times, and the number of times that $\kappa_1^p - \kappa_2^p$ equalled or exceeded $\kappa_1^o - \kappa_2^o$ was counted. If the ratio of the count plus 1 and the number of permutations plus 1 was less than or equal to 0.05, then the null hypothesis of no statistically significant difference between the two κ coefficients at $\alpha = 0.05$ was rejected. Approximately 1000 permutations were considered adequate for significance testing at the $\alpha = 0.05$ level (McKenzie et al., 1996). A similar test procedure was used for the other three accuracy measures, also.

Let B be a common coefficient vector denoting accuracy measures based on $OA(\mathbf{X})$, $\kappa(\mathbf{X})$, $J_{\text{uni}}(\mathbf{X})$ and $J_{\text{pro}}(\mathbf{X})$. The new fitness function for use with categorical response variables and to be substituted for Eq. (4) and minimized with respect to $\boldsymbol{\omega}$ is proposed to be,

$$f[\boldsymbol{\omega}, \gamma, B(\mathbf{X})] = \sum_{j=1}^{n_m} \gamma_j \left[1 - B_j(\mathbf{X}_j)\right], \tag{12}$$

where $\gamma_j > 0$ is a user defined coefficient, B_j is the accuracy measure with response variable j whose classes are to be predicted, n_m is the number of response variables to be considered in the optimisation procedure and $\boldsymbol{\omega}$ is the weight vector to be optimized (Eqs. (3a) and (3b)).

3.4. Discriminant analysis as a prediction method

Discriminant analysis was used as an optional method for obtaining predictions that could be compared to those obtained using ik-NN with the new fitness functions. Canonical variables derived with canonical discriminant analysis were used to obtain uncorrelated feature space variables. The purpose of a GA is somewhat similar to canonical analysis; i.e., to find variables that summarize between class variations. The non-correlation property is not necessarily guaranteed with coefficients obtained with GA. The generalized squared distance from a vector of feature space variables, $\boldsymbol{f}$, to the estimated class mean $\hat{\boldsymbol{\mu}}_c$, of class c was,

$$D_c^2(\boldsymbol{f}) = (\boldsymbol{f} - \hat{\boldsymbol{\mu}}_c)' \hat{\mathbf{S}}_c^{-1} (\boldsymbol{f} - \hat{\boldsymbol{\mu}}_c) + \ln|\hat{\mathbf{S}}_c| - 2\ln(q_c), \tag{13}$$

where $\hat{\mathbf{S}}_c$ is the estimate of the within class covariance matrix and q_c is the prior probability of class c (e.g., Gonzalez & Woods, 2002; Tomppo, 1992). Equal priors were used in this study, and the feature space variables were calculated from the canonical variables. Priors obtained from the field data could also have been used and would have been acceptable if the distribution of the categories of the response variable had been the same in the target and reference sets. However, because the field plots had been sampled, the distributions could not be guaranteed to be the same. The posterior probability of an observation with a feature vector $\boldsymbol{f}$ belonging to class c under the multinormal assumption is (e.g., Gonzalez & Woods, 2002; Tomppo, 1992),

$$p(c|\boldsymbol{f}) = \frac{\exp\left(-0.5 D_c^2(\boldsymbol{f})\right)}{\sum_{i=1}^{r} \exp\left(-0.5 D_i^2(\boldsymbol{f})\right)}. \tag{14}$$

4. Results

Experiments were conducted to compare results obtained with different weights for the accuracy measures (Eqs. (5), (7)–(9) and (12)) and for the k-NN estimation parameters such as the number (k) of nearest neighbours and the value of the illumination correction due to the elevation variation. Results are reported for only one set of parameters because the objective was to investigate accuracy gains achieved when using a GA. The choice of the value of k, for example, depends on the technical objective; i.e., low pixel level RMSE or low bias for areas of interest. A thorough analysis of this issue is beyond the scope of this study. Values of $k = 5$ and $r = 0.1$, the power of cosine in normalising intensity values (Eq. (3a)), were selected after experimentation. For both the Finnish and Italian test sites, the results obtained using the different fitness functions in the GA optimisation were first analysed using the entire Finnish NFI9 and Italian INFC reference data sets using leave-one-out cross-validation with the four accuracy measures. A second set of analyses was conducted using the independent validation data sets.

Several iterations of the GA were conducted for each experiment and set of parameters to ensure that a near global optimum, instead of a local optimum, was found (Tomppo & Halme, 2004). For the final results, five iterations were used.

Results are reported first for the Finnish site (Section 4.1) and second for the Italian site (Section 4.2). For the Finnish site, four analyses are reported: (4.1.1) a comparison of the ranges of spectral values for the reference and target sets, (4.1.2) accuracy assessment using the reference set data only, (4.1.3) results for site fertility with the independent data set, and (4.1.4) results for dominant tree species with the independent data set. For the Finnish site, all results are reported separately for peatland and mineral soils. For the Italian site, three analyses are reported: (4.2.1) a comparison of the ranges of spectral values for the reference and target sets, (4.2.2) accuracy assessment using the reference set data only, and (4.2.3) results for both CBM and forest type when splitting the reference set into independent reduced reference and target sets.

The weights for feature space variables were optimized both separately and jointly for the two response variables for both the Finnish and Italian test sites. The accuracy statistics were somewhat greater for the Finnish data when the weights were estimated separately (Tables 5–12), whereas for the Italian data, the accuracy statistics were similar for the separate and joint optimisations. Therefore, only the results for the joint optimisation are presented for the Italian data (Tables 13–16). A possible explanation for the different results for the Finnish and Italian data is that the Italian response variables were highly dependent whereas dependence was not as great for the Finnish variables.

4.1. Finnish site

4.1.1. Spectral distributions of reference data and target data by categories

The spectral distributions of the original ETM+ data were compared for the reference and independent test target data to determine if the reference data covered the spectral variation in the target area, a prerequisite for successful k-NN prediction and

Table 5
Accuracy statistics for 1901 Finnish mineral soil reference set plots using leave-one-out cross validation and five different fitness functions, best of five iterations

Prediction technique		Forest attribute variable							
		Site fertility				Dominant tree species			
		Accuracy measures				Accuracy measures			
		OA	J_{uni}	J_{pro}	Kappa	OA	J_{uni}	J_{pro}	Kappa
k-NN		0.584	0.279	0.534	0.311	0.657	0.161	0.557	0.362
ik-NN	OA_f[a]	0.628	0.299	0.580	0.379	0.684	0.233	0.614	0.413
	$J_{uni,f}$[a]	0.594	0.372	0.553	0.328	0.673	0.229	0.600	0.396
	$J_{pro,f}$[a]	0.621	0.300	0.580	0.374	0.689	0.232	0.621	0.425
	$Kappa_f$[a]	0.629	0.288	0.584	0.389	0.690	0.205	0.610	0.431
	Volume[b]	0.569	0.273	0.525	0.293	0.680	0.300	0.627	0.412
Discriminant analysis		0.547	0.633	0.540	0.339	0.611	0.294	0.542	0.393

[a] Accuracy measure used in fitness function for genetic algorithm.
[b] Fitness function based on RMSEs and biases used for genetic algorithm.

estimation. In the Finnish case, we may also draw more general conclusions as to whether the NFI field plot data can be expected to cover spectral variations of typical small forested areas of size 100 km^2 within an area the size of a Landsat image. The means, standard deviations and ranges of the spectral data were compared for each site fertility class and for the five aggregated tree species classes (Species Grouping 2). The analyses were conducted separately for the mineral soil and peatland soil field plots using only plots at least 20 m from the nearest stand boundary.

On mineral soils, within site fertility classes and tree species groups, the means of the digital numbers (DN) for the reference and independent target data were similar, and the standard deviations of the DNs for the independent target data were less than or equal to the standard deviations for the reference data (Table 17). For the somewhat rare site fertility class HRS, the standard deviation for the reference data for band 5 was less than the standard deviation for the independent target data. However, the minimum and maximum values for the independent target data were within the range for the reference data. For the open land class on mineral soils (n=24), the standard deviations for the independent target data for bands 3, 5 and 8 were larger than for the reference data, and the minimum values for the independent target data were beyond the range for the reference data (Table 18). These findings suggest the possibility of poor prediction accuracy for the class (cf. Tables 11 and 12). However, for the spruce class (n=479), the standard deviations of the reference and independent data were more similar (Table 18). For band 5, the ranges were quite similar despite the differences in the standard deviations.

On peatland soils, the means and standard deviations of the DNs were similar for the reference and independent target data. For site fertility class HRHF (n=22), the standard deviation of the reference data for band 5 was smaller than for the independent target data (Table 19). In this case, the range of the band values for the reference data was slightly larger than for independent target data. For tree species, the largest differences were for band 5 for the spruce class, but the ranges of the DNs were almost equal for the reference and independent target data (Table 20).

Table 6
Accuracy measures for 690 Finnish peatland soil reference set plots using leave-one-out cross validation and five different fitness functions, best of five iterations

Prediction technique		Forest attribute variable							
		Site fertility				Dominant tree species			
		Accuracy measures				Accuracy measures			
		OA	J_{uni}	J_{pro}	Kappa	OA	J_{uni}	J_{pro}	Kappa
k-NN		0.486	0.291	0.462	0.314	0.799	0.378	0.753	0.583
ik-NN	OA_f[a]	0.529	0.313	0.501	0.370	0.812	0.397	0.773	0.612
	$J_{uni,f}$[a]	0.513	0.430	0.507	0.356	0.810	0.558	0.786	0.618
	$J_{pro,f}$[a]	0.539	0.441	0.522	0.388	0.813	0.530	0.780	0.615
	$Kappa_f$[a]	0.370	0.314	0.500	0.371	0.829	0.488	0.797	0.648
	Volume[b]	0.464	0.276	0.435	0.286	0.797	0.565	0.757	0.589
Discriminant analysis		0.504	0.502	0.496	0.358	0.770	0.692	0.768	0.616

[a] Accuracy measure used in fitness function for genetic algorithm.
[b] Fitness function based on RMSEs and biases used for genetic algorithm.

Table 7
Accuracy measures for 1411 Finnish mineral soil independent target set plots and 1901 reference set plots and five different fitness functions, best of five iterations

Prediction technique		Forest attribute variable							
		Site fertility				Dominant tree species			
		Accuracy measure				Accuracy measure			
		OA	J_{uni}	J_{pro}	Kappa	OA	J_{uni}	J_{pro}	Kappa
k-NN		0.605	0.180	0.547	0.269	0.653	0.196	0.612	0.395
ik-NN	OA_f[a]	0.592	0.229	0.559	0.263	0.660	0.307	0.628	0.400
	$J_{uni,f}$[a]	0.583	0.208	0.534	0.247	0.654	0.374	0.627	0.388
	$J_{pro,f}$[a]	0.599	0.208	0.545	0.252	0.646	0.242	0.614	0.380
	$Kappa_f$[a]	0.601	0.213	0.560	0.282	0.651	0.302	0.621	0.386
	Volume[b]	0.579	0.201	0.519	0.210	0.627	0.188	0.588	0.353
	No csv[c]	0.588	0.177	0.531	0.258	0.651	0.337	0.618	0.377
Discriminant analysis		0.552	0.461	0.546	0.282	0.613	0.360	0.605	0.382

[a] Accuracy measure used in fitness function for genetic algorithm.
[b] Fitness function based on RMSEs and biases used for genetic algorithm.
[c] OA_f, No coarse scale variables.

Table 8
Accuracy measures for 387 Finnish peatland soil independent target set plots using 690 reference set plots and five different fitness functions, best of five iterations

Prediction technique		Forest attribute variable							
		Site fertility				Dominant tree species			
		Accuracy measure				Accuracy measure			
		OA	J_{uni}	J_{pro}	Kappa	OA	J_{uni}	J_{pro}	Kappa
k-NN		0.450	0.357	0.439	0.240	0.788	0.516	0.741	0.471
ik-NN	OA_f[a]	0.450	0.341	0.440	0.235	0.801	0.542	0.763	0.516
	$J_{uni,f}$[a]	0.494	0.441	0.488	0.301	0.796	0.522	0.750	0.492
	$J_{pro,f}$[a]	0.463	0.387	0.443	0.252	0.793	0.551	0.759	0.499
	$Kappa_f$[a]	0.473	0.268	0.420	0.264	0.804	0.533	0.761	0.511
	Volume[b]	0.493	0.427	0.487	0.298	0.765	0.507	0.717	0.425
	No csv[c]	0.475	0.386	0.469	0.279	0.804	0.606	0.781	0.547
Discriminant analysis		0.465	0.402	0.453	0.272	0.765	0.729	0.765	0.556

[a] Accuracy measure used in fitness function for genetic algorithm.
[b] Fitness function based on RMSEs and biases used for genetic algorithm.
[c] OA_f, No coarse scale variables.

4.1.2. Accuracy assessment using reference data

Results obtained using the reference data with Species Grouping 1 were obtained using leave-one-out cross-validation. Accuracies for the four accuracy measures without and with GA optimisation for the

Table 9
Error matrix for Finnish site fertility for independent peatland soil target set without optimization

Predicted site fertility class[a]	Observed site fertility class[a]							
	HRS	HRHF	MF	SX	X	B	Total	User's accuracy (%)
HRS	1	0	2	0	0	0	3	33.3
HRHF	1	3	13	3	0	1	21	14.3
MF	0	12	54	21	6	0	93	58.1
EX	2	5	32	71	48	2	160	44.4
SX	0	2	11	28	41	4	86	47.7
X	0	0	1	7	12	4	24	16.7
Total	4	22	113	130	107	11	387	
Producer's accuracy (%)	25.0	13.6	47.8	54.6	38.3	36.4	45.0	

[a] See Section 2.1.4 for site fertility class definitions.

Table 10
Error matrix for Finnish site fertility for independent peatland soil target set with optimization using $J_{uni,f}$ in fitness function

Predicted site fertility class[a]	Observed site fertility class[a]							
	HRS	HRHF	MF	SX	X	B	Total	User's accuracy (%)
HRS	2	1	0	1	0	0	4	50.0
HRHF	0	4	11	2	0	0	17	23.5
MF	0	14	59	23	7	0	103	57.3
SX	1	1	31	62	32	1	128	48.4
X	1	1	10	40	59	5	116	50.9
B	0	1	2	2	9	5	19	26.3
Total	4	22	113	130	107	11	387	
Producer's accuracy (%)	50.0	18.2	52.2	47.7	55.1	45.5	49.4	

[a] See Section 2.1.4 for site fertility class definitions.

Table 12
Error matrix for Finnish tree species grouping for independent mineral soil target set with optimization using OA in fitness function

Predicted tree species groups	Observed tree species groups					Total	User's accuracy (%)
	Open	Pine	Spruce	Birch	Other broad-leaved		
Open	1	7	1	0	0	9	11.1
Pine	9	641	161	31	0	842	76.1
Spruce	8	79	241	14	1	343	70.3
Birch	6	71	61	47	4	189	24.9
Other broad-leaved	0	11	10	6	1	28	3.6
Total	24	809	474	98	6	1411	
Producer's accuracy (%)	4.2	79.2	50.8	48.0	16.7	66.0	

different fitness functions are reported for both mineral soil (Table 5) and peatland soil (Table 6). For comparison purposes, the results with a fitness function based on minimising the RMSEs and biases for prediction of a continuous variable, volume (Eq. (4)), as used by Tomppo and Halme (2004), rather than categorical variables, are also presented in Tables 5 and 6 (row ik-NN, Volume). Accuracies obtained using canonical variables with the discriminant analysis are also reported. Results are reported separately for site fertility and dominant tree species in Tables 5 and 6.

Accuracies for the different accuracy measures vary substantially. The values for OA and J_{pro} are similar as should be expected based on their definitions. The lower value for J_{uni} is due to its sensitivity to errors in individual classes independently of the number of the observations in the class. Some site fertility classes were very rare in the test site and, therefore, yielded low class frequencies (Table 1). A similar result holds for tree species dominances (Table 2). Pixel-level OA is slightly greater than 60% for site fertility and approximately 70% for dominant tree species.

Because the accuracy measures are dependent, optimisation with respect to one measure generally increases the values of the other measures without optimisation. No clear differences were found among the different fitness functions. When used as a fitness function, $Kappa_f$ gave the greatest value for OA for site fertility with mineral soil plots; and $J_{pro,f}$ and $Kappa_f$ gave the greatest values for dominant tree species for OA. GA optimisation increased agreements in all cases. The different nature of J_{uni} both as fitness function and in evaluation can also be seen in that its greatest value is obtained using discriminant analysis (Table 5).

GA optimisation increased accuracies for predictions of the categorical variables for peatland soils, also. Accuracies were greater for all measures, except for J_{uni}, when using ik-NN than for discriminant analysis. Pixel-level OA was approximately 54% for site fertility in the best cases and slightly greater than 80% for dominant tree species. Satellite image-based predictions are vulnerable to pixel-level errors on peatland soils due to moisture variation which is not necessarily linked to the variation of the variables of interest. Of the different fitness functions, $J_{pro,f}$ produced the greatest values for all accuracy measures for site fertility; $Kappa_f$ produced the greatest values for OA; and $J_{pro,f}$ and $Kappa_f$ produced the greatest accuracies for dominant tree species for OA. A somewhat unexpected result is that $Kappa_f$ gave a low OA value for site fertility on peatland soil. This result could be attributed to difficulty in site fertility prediction, particularly on peatlands, a random effect, or both. Results with discriminant analysis were similar to those achieved for mineral soils. These results indicate no clear difference among fitness functions except perhaps that use of $J_{uni,f}$ as a fitness function often produced the greatest value for J_{uni}. This result could lead to a recommendation to use $J_{uni,f}$ when the technical objective is high prediction accuracies for all classes, although it could lead to compromising the highest possible overall accuracy or Kappa value.

An interesting question pertains to the gain in prediction accuracy when using fitness functions based on maximization of the accuracies for the categorical variables (Eq. (12)) instead of fitness functions based on volumes (Eq. (4)). Accuracy measures are somewhat higher for site fertility when the optimization is based on Eq. (12) rather than Eq. (4). A similar increase was not achieved when predicting tree species. This result is attributed to volumes by tree species better reflecting tree species dominance than site fertility.

Table 11
Error matrix for Finnish tree species grouping for independent mineral soil target set without optimization

Predicted tree species groups	Observed tree species groups					Total	User's accuracy (%)
	Open	Pine	Spruce	Birch	Other broad-leaved		
Open	0	1	0	2	0	3	0.0
Pine	11	627	142	36	0	816	76.8
Spruce	8	85	251	12	1	357	70.3
Birch	5	91	76	44	5	221	19.9
Other broad-leaved	0	5	5	4	0	14	0.0
Total	24	809	474	98	6	1411	
Producer's accuracy (%)	0.0	77.5	53.0	44.9	0.0	65.3	

4.1.3. Results for site fertility for mineral and peatland soils using independent data set

Because none of the accuracy measures was clearly superior using only the reference data, all measures were retained for the analyses with the independent target data.

The overall increases in accuracy were smaller for the independent target data (Tables 7 and 8) than for the leave-one-out analyses with the reference data only (Tables 5 and 6). Despite the large number of plots in the reference data set, this result is consistent with our

Table 13
Accuracy measures for 955 Italian NFI plots using leave-one-out cross validation for best of five iterations

Prediction technique		Forest attribute variable							
		Conifer/broad-leaved/mixed dominance				Forest type			
		Accuracy measure				Accuracy measure			
		OA	J_{uni}	J_{pro}	Kappa	OA	J_{uni}	J_{pro}	Kappa
k-NN		0.663	0.570	0.634	0.442	0.471	0.272	0.407	0.305
ik-NN	OA_f[a]	0.694	0.597	0.664	0.494	0.534	0.319	0.476	0.386
	$J_{uni,f}$[a]	0.668	0.573	0.639	0.454	0.506	0.381	0.478	0.354
	$J_{pro,f}$[a]	0.683	0.588	0.654	0.477	0.519	0.372	0.486	0.372
	$Kappa_f$[a]	0.693	0.612	0.671	0.494	0.525	0.346	0.477	0.373
Discriminant analysis		0.661	0.623	0.656	0.463	0.507	0.468	0.496	0.387

[a] Accuracy measure used in fitness function for genetic algorithm.

Table 14
Accuracy measures for 318 Italian independent target set plots using 637 reference set plots for best of five iterations

Prediction technique		Forest attribute variable							
		Conifer/broad-leaved/mixed dominance				Forest type			
		Accuracy measure				Accuracy measure			
		OA	J_{uni}	J_{pro}	Kappa	OA	J_{uni}	J_{pro}	Kappa
k-NN		0.632	0.564	0.608	0.401	0.465	0.335	0.435	0.298
ik-NN	OA_f[a]	0.620	0.530	0.582	0.377	0.491	0.330	0.441	0.331
	$J_{uni,f}$[a]	0.645	0.548	0.602	0.421	0.525	0.384	0.494	0.373
	$J_{pro,f}$[a]	0.616	0.510	0.567	0.367	0.443	0.306	0.404	0.276
	$Kappa_f$[a]	0.632	0.529	0.587	0.392	0.509	0.345	0.466	0.351
Discriminant analysis		0.619	0.560	0.604	0.386	0.453	0.352	0.418	0.325

[a] Accuracy measure used in fitness function for genetic algorithm.

Table 16
Error matrix for Italian forest type classification with optimisation using $J_{uni,f}$ in fitness function

Predicted forest type classes[a]	Observed forest type classes[a]								
	LSP	NSF	SP	B	OB	H	SS	Total	User's accuracy (%)
LSP	20	11	1	6	2	0	5	45	44.4
NSF	20	93	7	11	6	3	1	141	66.0
SP	1	4	8	1	1	0	1	16	50.0
B	2	8	3	24	8	10	1	56	42.9
OB	0	2	1	5	4	6	1	19	21.1
H	3	3	1	5	2	15	0	29	51.7
SS	3	3	2	1	0	0	3	12	25.0
Total	49	124	23	53	23	34	12	318	
Producer's accuracy (%)	40.8	75.0	45.0	45.3	17.4	44.1	25.0	52.5	

[a] See Table 4 for forest type class definitions.

experiences from the operative Finnish MS-NFI; the errors for the independent target data at the plot level are usually greater than for the leave-one-out cross-validation errors for the reference data (Tomppo, 2006). A second possible reason for these results is that the areas from which the Finnish independent target data were selected represent extremes of variation for the forest attribute variables. Although ik-NN produced slightly better results than discriminant analysis, none of the fitness functions produced increases in OA for predictions of site fertility for mineral soils (Table 7). Optimisation using all four fitness functions increased J_{uni}. When using OA_f for the fitness function, J_{uni} increased by 27%; when using $Kappa_f$ and OA_f for the fitness functions, J_{pro} increased slightly; and using $Kappa_f$ as the fitness function produced the greatest value for Kappa.

Accuracy increases obtained with GA optimisation for the fitness functions were greater for site fertility on peatland soils than for mineral soils. The increases ranged from 10–25% (Table 8). Optimisation using the fitness function based on $J_{uni,f}$ produced the greatest values for all accuracy measures for site fertility on peatland. In addition, these values were all greater than those obtained when using discriminant analysis.

Accuracies obtained with fitness functions based on RMSEs and biases of volumes (Eq. (4)) for mineral soils are, on average, slightly lower than those based on the fitness functions using the new accuracy measures (Table 7). On peatland soils, the differences seem to be more attributable to random variation (Table 8). A reason for these results could be the previously mentioned difficulty in predicting site fertility.

The fitness function based on Eq. (12) was also tested without using coarse scale forest variables in the distance metric of Eq. (3a), i.e., using the distance metric of Eq. (3b). Only OA_f was used in these tests (Tables 7 and 8). Slightly lower accuracies, on the average, were obtained in predicting site fertility for mineral soils without the coarse-scale forest variables than with them. The differences were, however, small and could be attributed to random variation. For peatland soils, the opposite results were obtained. This could confirm the random nature of this test result, the previously mentioned difficulty in predicting site fertility on peatland soil, or both.

In general, pixel-level prediction of site fertility using satellite imagery is difficult. On mineral soils, tree species proportions are not necessarily good indicators of site fertility after intensive forest management regimes. In these cases, regeneration favours pine on site fertility class MF and sometimes for class HRHF.

Examples of error matrices for predictions of site fertility with and without GA optimisation are shown in Tables 9 and 10 for peatland soils. These results were obtained for the independent target data using a fitness function based on $J_{uni,f}$. Regardless of the small increases in the accuracies, both user's and producer's accuracies increased in the major classes, except users' accuracy for the MF class and producer's accuracy for the SX class. The site fertility classes can be considered to be on an ordinal scale so that an error of one category (i.e., erroneously predicting an adjacent class) is not as serious as an error of several categories. The accuracy measures are not sensitive to the observation that most incorrect predictions are only of one category.

Table 15
Error matrix for Italian forest type classification without optimisation

Predicted forest type classes[a]	Observed forest type classes[a]								
	LSP	NSF	SP	B	OB	H	SS	Total	User's accuracy (%)
LSP	18	16	2	7	3	3	3	52	34.6
NSF	17	86	7	10	5	7	3	135	63.7
SP	2	7	6	2	1	1	0	19	31.6
B	4	6	6	21	8	10	2	57	36.8
OB	1	3	0	6	4	2	1	17	23.5
H	3	4	0	5	2	10	0	24	41.7
SS	4	2	2	2	0	1	3	14	21.4
Total	49	124	23	53	23	34	12	318	
Producer's accuracy (%)	36.7	69.4	26.1	39.6	17.4	29.4	25.0	46.5	

[a] See Table 4 for forest type class definitions.

4.1.4. Dominant tree species for mineral and peatland soils using independent data sets

Accurate prediction of classes with rare species for Species Grouping 1 was difficult using only reference data. Because accuracies would be even less when using independent target data, prediction with Species Grouping 2 was also tested. Tree species dominance results in this section are given for Species Grouping 2 only. The results presented in Table 7 for mineral soils and in Table 8 for peatland soils are obtained using the independent reference and target data sets.

Improvement in accuracy using ik-NN over that obtained with k-NN is less with the independent test data than with the reference data only. In some cases k-NN gives even slightly better accuracy than ik-NN. However, differences are small and can be attributed to sampling and random variation. GA optimisation seems to increase accuracies slightly more for tree species dominance than for site fertility, except for the OA measure. Pixel-level OA is approximately 65% for mineral soils and 80% for peatland soils. The increase in accuracies that can be attributed to GA optimisation varies from 1–90% for mineral soils and 2–10% for peatland soils. With GA optimisation, k-NN produced greater accuracies than discriminant analysis with only a few exceptions; e.g., using J_{uni} for mineral and peatland soils.

One reason for greater OA, as well as greater accuracies for tree species dominance (Species Grouping 2) on peatland soils than on mineral soils may be attributed to the fact that 57% of the independent data plots are pine dominant on mineral soils (Table 2) while the percentage on peatland soils is 71%.

Table 17
Spectral statistics for reference and independent target sets for Finnish site fertility classes on mineral soils

ETM+ band	Measure	Site fertility class[a]													
		HRS		HRHF		MF		SX		X		B		R	
		Ref	Tar	Ref	Tar	Ref	Tar	Ref	Tar	Ref	Tar	Ref	Tar	Ref	Tar
1	Mean	61.4	60.1	61.6	61.0	62.1	62.2	62.6	62.8	64.4	63.0	64.1		66.4	62.1
	St dev	3.3	1.9	3.5	3.7	3.9	4.3	3.7	3.6	4.1	3.2	.		4.2	3.9
	Min	57.4	58.2	55.6	54.4	54.3	55.1	55.3	54.7	72.9	57.5	64.1		61.4	59.3
	Max	75.9	64.3	74.7	73.3	76.8	79.9	77.4	73.4	58.5	69.8	64.1		73.3	64.9
2	Mean	46.8	46.7	46.6	45.8	46.9	47.4	46.9	47.4	48.8	48.3	49.6		49.4	46.2
	St dev	4.5	3.9	5.1	5.1	5.0	5.8	4.0	4.4	3.9	3.1	.		3.8	1.7
	Min	40.1	43.1	37.3	37.3	38.9	37.9	39.9	34.0	58.4	43.1	49.6		44.0	45.0
	Max	63.6	57.2	64.4	60.8	63.3	66.5	61.5	62.6	42.1	55.5	49.6		56.7	47.4
3	Mean	34.6	33.7	35.2	35.0	37.2	38.4	39.1	40.1	44.3	40.8	39.3		41.8	37.5
	St dev	7.8	3.4	7.4	7.3	8.0	9.0	6.8	7.6	7.0	4.8	.		8.6	8.2
	Min	27.8	28.6	25.3	26.9	26.7	26.8	29.7	26.8	60.6	32.9	39.3		28.6	31.7
	Max	71.8	40.9	68.2	65.6	66.7	71.6	67.8	73.0	33.9	53.4	39.3		60.9	43.3
4	Mean	79.5	76.8	65.1	62.8	58.9	58.4	53.8	54.7	52.2	54.8	50.6		55.6	48.2
	St dev	12.8	16.4	14.7	16.3	10.6	11.5	6.1	7.4	4.7	6.7	.		8.5	1.7
	Min	47.0	62.1	24.8	38.1	22.7	26.7	39.1	23.7	64.5	48.1	50.6		45.9	47.0
	Max	106.7	119.5	120.2	132.3	90.9	94.8	72.3	82.8	45.2	73.9	50.6		79.1	49.4
5	Mean	61.4	64.0	57.3	54.7	58.2	59.9	60.3	63.3	68.0	65.6	62.0		65.8	58.5
	St dev	12.3	11.8	20.7	20.6	20.4	23.0	15.9	17.8	14.5	11.0	.		14.8	7.6
	Min	38.8	49.0	28.9	28.2	18.6	24.6	35.0	23.7	95.7	44.2	62.0		49.0	53.2
	Max	95.5	91.9	123.3	111.1	120.8	122.1	110.6	113.1	49.3	88.3	62.0		94.9	63.9
7	Mean	130.4	129.1	131.3	130.5	132.6	132.5	135.0	135.3	139.1	137.1	131.3		135.0	131.9
	St dev	3.6	2.8	5.9	5.4	6.8	6.9	7.0	7.2	6.5	7.1	.		6.2	8.7
	Min	124.5	124.6	118.0	123.2	117.5	120.4	120.5	121.2	155.3	124.3	131.3		128.4	125.8
	Max	140.5	135.8	161.0	154.9	164.9	158.5	157.4	160.5	128.6	149.9	131.3		145.6	138.0
8	Mean	30.4	30.2	30.6	29.7	32.4	33.7	34.7	36.9	40.4	37.8	35.1		38.1	32.9
	St dev	6.7	6.1	12.5	12.3	12.8	14.2	11.0	12.6	10.9	8.0	.		9.9	3.1
	Min	21.4	22.5	15.3	16.6	13.4	11.3	21.5	14.4	65.7	24.6	35.1		27.6	30.7
	Max	51.4	42.9	87.5	77.9	85.0	83.5	77.8	76.6	25.7	56.5	35.1		57.8	35.0
9	Mean	54.5	55.3	47.3	46.0	44.7	44.6	42.2	42.9	42.8	42.5	41.3		42.6	37.0
	St dev	7.1	8.6	9.2	10.2	7.5	8.3	4.6	5.4	3.2	4.5	.		6.2	3.1
	Min	36.8	46.0	28.3	29.6	28.8	22.6	29.8	27.9	50.3	32.9	41.3		29.3	34.8
	Max	69.7	75.6	73.9	77.0	74.8	67.0	59.7	60.5	37.0	54.4	41.3		55.5	39.1

[a] See Section 2.1.4 for site fertility class definitions.

Accuracies for Species Grouping 2 on mineral soils obtained with fitness functions based on RMSEs and biases of volumes (Eq. (4)) are also, on average, slightly smaller than accuracies obtained using fitness functions based on the new accuracy measures (Table 7). Results for site fertility were similar to those for Species Grouping 2. On peatland soils, the prediction results with fitness functions based on Eq. (4) were similar (Table 8), although differences are small in both cases.

Accuracies for Species Grouping 2 for OA_f with and without use of the coarse scale forest variables in the distance metric (Eq. (3a)) were almost the same. On peatland soils slightly greater accuracies were obtained without coarse scale variables than with them. Again, this result could reflect the difficulties for prediction for peatland forests, or a lacking explanatory power of coarse scale variables in predicting Species Grouping 2 for peatland forests.

Error matrices for tree species dominance predictions (Species Grouping 2) without GA optimisation are reported in Table 11 and with GA optimisation in Table 12. GA optimisation produced only slight increases in accuracies, none of which were statistically significant. A somewhat greater increase in J_{uni} (Table 7, rows k-NN and OA_f) is caused by a very small increase in the proportion of correct predictions for birch and other broad-leaved trees.

The statistical significance of differences in prediction accuracies with and without GA was tested in two cases using the Monte Carlo permutation test to gain insights for improvements in prediction using GA. The significance of the difference between accuracies from two dependent predictions was estimated for independent target data in Finland. The k-NN accuracies for site fertility for peatland soils (Table 9) were compared to the ik-NN $J_{uni,f}$ optimized accuracies (Table 10), and the k-NN accuracies for tree species groups (Species Grouping 2) on mineral soils (Table 11) were compared to the ik-NN OA optimized accuracies (Table 12). Accuracies for the four measures for these predictions are reported in the Tables 7 and 8. For site fertility on peatland soils, the null hypothesis of no significant difference between the accuracy measures was only rejected for the J_{uni} at the α=0.05 level of significance (p-value 0.007). The other measures were significantly different at the α=0.10 (p-values 0.074, 0.066 and 0.053 for Kappa, OA and J_{pro}, respectively). For Species Grouping 2 on mineral soils, the difference between two accuracy measures was significant at the α=0.05 level only for J_{uni} (p-values 0.754, 0.528, 0.001 and 0.149 for κ, OA, J_{uni} and J_{pro}, respectively).

4.2. Italian site

4.2.1. Spectral distributions of reference data and target data by categories

The original spectral values by classes were analysed to compare spectral variability in the reference and target sets (Table 21). For classes consisting mostly of conifers (LSP, NFS, and SS), means, variances and particularly ranges of the spectral values of the reference and target sets are very different. This phenomenon, which is more evident in the visible bands than in the infrared bands, is attributed to two effects: first, random selection of target plots from the reference set, a relatively small sample size compared to the variability of the forests and spectral values, and second, the presence of anomalous spectral values far from the mean.

When analysed with respect to prediction accuracies, the differences in the distributions of spectral values by forest type classes apparently did not directly influence producer's accuracy (Tables 15 and 16). For example, NSF, the largest of the seven forest type classes, was predicted well despite substantial differences between the reference and target set with respect to maximums and standard deviations of spectral values. However, for OB, despite similarity between the reference and target

Table 18
Spectral statistics for reference and independent target sets for Finnish dominant tree species groups on mineral soils

ETM+ band	Measure	Tree species groups									
		Open		Pine		Spruce		Birch		Other broad-leaved	
		Ref	Tar	Ref	Tar	Ref	Tar	Ref	Tar	Ref	Tar
1	Mean	69.1	64.6	62.0	62.5	60.8	61.4	63.5	62.2	65.7	59.8
	St dev	4.2	6.5	3.4	3.8	3.4	4.3	4.0	3.9	3.8	1.8
	Min	58.1	56.3	54.3	54.7	54.9	54.4	54.6	56.3	59.4	58.2
	Max	75.8	79.9	77.4	76.0	75.9	76.7	74.6	73.6	76.8	62.6
2	Mean	54.5	49.6	46.2	47.4	45.1	46.2	50.2	48.9	53.5	46.3
	St dev	3.8	7.5	3.8	4.8	4.6	6.1	5.0	5.3	5.4	2.9
	Min	42.5	39.9	39.1	34.0	38.9	37.3	39.0	41.1	37.3	41.6
	Max	59.3	62.5	61.5	62.6	64.4	66.5	63.3	61.4	62.6	50.3
3	Mean	55.9	45.6	37.0	39.1	34.2	36.3	40.3	38.1	44.5	33.7
	St dev	9.4	14	6.7	8.0	6.7	8.7	8.9	9.0	8.8	2.5
	Min	33.2	29.4	26.7	26.8	25.6	26.8	25.3	27.6	25.9	31.2
	Max	66.7	68.0	67.8	73.0	71.8	71.6	66.4	71.5	66.5	37.9
4	Mean	59.5	55.9	56.2	57.3	56.2	56.8	74.7	75.1	75.5	79.4
	St dev	10.2	8.8	8.0	9.2	11.1	13.3	12.4	13.2	11.6	18.3
	Min	46.7	42.9	22.7	23.7	35.7	32.8	32.9	50.3	24.8	43.7
	Max	88.5	73.8	94.2	91.3	91.3	98.9	120.2	132.3	95.4	91.4
5	Mean	95.6	74.2	56.4	62.0	48.5	53.5	75.2	70.0	85.8	62.9
	St dev	16.1	32.7	15.1	19.4	17.7	23.2	18.6	18.4	16.3	16.9
	Min	41.5	36.0	18.6	23.7	28.9	27.9	35.1	41.9	33.1	33.3
	Max	120.8	120.3	114.4	91.3	123.3	122.1	115.8	115.4	114.7	84.1
7	Mean	146.6	137.9	132.9	133.9	130.4	130.8	135.0	132.5	137.2	129.5
	St dev	8.1	10.7	6.5	6.9	5.7	6.4	6.9	6.2	6.8	2.0
	Min	131.9	124.8	117.5	120.9	118.0	120.4	123.4	122.6	125.3	127.7
	Max	164.9	155.4	164.1	160.5	157.4	158.5	159.2	156.3	152.6	133.3
8	Mean	59.6	44.7	31.7	35.3	27.1	30.1	39.9	36.6	46.6	30.2
	St dev	13.7	21.1	10.3	13.0	11.2	13.8	12.8	11.4	12.2	6.8
	Min	24.9	19.6	13.4	11.3	15.3	16.4	19.5	22.6	17.6	19.8
	Max	85.0	72	77.9	79.6	87.5	83.5	69.8	72.5	78.8	41.0
9	Mean	49.0	44	43.0	44.1	42.2	43.1	54.3	54.5	56.3	53.9
	St dev	5.4	8.5	5.4	6.4	7.5	9.4	6.8	7.5	5.9	11.3
	Min	37.3	33.5	28.9	25.6	28.3	22.6	37.0	39.0	42.8	32.3
	Max	61.4	63.6	74.8	66.4	65.4	67.0	73.9	77.0	68.7	62.6

sets with respect to distributions of spectral values, misclassification was high, possibly because of the low frequency in the reference set relative to other classes such as NSF.

Overall, accuracy was possibly influenced by the number of observations for each class which supports the assertion that successful application of nearest neighbour techniques requires a sample of the population of interest that is representative for both feature space and response variables (Moeur, 1988).

4.2.2. Accuracy assessment using reference data

Results for the four accuracy measures for the Italian test site using leave-one-out cross-validation with and without GA optimisation for the different fitness functions are reported in Table 13. Accuracies for the discriminant analyses are also reported. GA optimisation produced slight increases in accuracy as for the Finnish site. OA increased from 66.3% to 69.4% for CBM and from 47.1% to 53.4% for forest type using OA_f in the fitness function.

As for the Finnish site, the different fitness functions did not produce large differences in accuracies. The effect of the number of the observations in the classes can be seen in the values of J_{uni} which are greater for CBM than for forest type. These results can be attributed to the smaller number of classes and the larger number of observations per class for CBM than for forest type which results in smaller probabilities of large prediction errors for the CBM class.

GA optimisation with k-NN usually produced greater accuracies for all measures compared to k-NN without optimisation, except for J_{uni} as was the case for the Finnish site (Table 13). These results together with the results from the Finnish test sites could indicate the limited capability of a non-parametric k-NN method to predict classes with very few observations, particularly when variation within classes in the feature space variables is large.

4.2.3. Results with the independent reference and target data

When predicting categories for independent target sets, the optimal weights for feature space variables were determined using the reference data only (Subsection 2.2.2). GA optimisation produced smaller accuracy increases for the independent target data than for the reference data using leave-one-out cross-validation (Table 14). GA optimisation using $J_{uni,f}$ in the fitness function generally produced the greatest accuracies. The increase in OA for CBM was only approximately 2% and for Kappa approximately 5%. J_{uni} and J_{pro} were greatest without optimisation.

When using $J_{uni,f}$ as a fitness function, the increase in the accuracy of forest type predictions was 13% for OA, 15%, for J_{uni}, 14% for J_{pro} and 25% for Kappa. Despite the small increases in accuracies, both user's and producer's accuracies increased for nearly all categories when GA optimisation was used; for example, forest type optimised with J_{uni} (Tables 15 and 16), except user's accuracy in the other broad-leaved category for which the number of the target observations is small.

For the Italian analyses, GA optimisation produced greater accuracy gains for forest type than for CBM. These results may be attributed to either the greater weight assigned to forest type in the GA optimisation or the greater opportunity for improvement for forest type because accuracies for CBM were already relatively high without GA optimisation.

5. Discussion and conclusions

We have reported methods and new fitness functions for predicting classes of categorical variables using both the k-NN and ik-NN techniques. In most cases, GA optimization produced slight increases in the values of accuracy measures relative to the ones obtained without GA optimisation. An underlying assumption was that field classification was correct.

Table 19
Spectral statistics for reference and independent target sets for Finnish site fertility classes on peatland soils

ETM+ band	Measure	Site fertility class[a]											
		HRS		HRHF		MF		SX		X		B	
		Ref	Tar	Ref	Tar	Ref	Tar	Ref	Tar	Ref	Tar	Ref	Tar
1	Mean	61.3	61.4	62.0	61.3	61.4	61.1	61.8	61.7	63.5	63.3	64.8	65.2
	St dev	2.7	3.7	3.3	3.4	3.4	3.1	3.0	2.5	3.6	3.0	2.7	2.3
	Min	57.5	58.6	54.5	56.4	54.4	56.4	55.5	56.4	56.5	58.5	60.6	61.4
	Max	66.8	66.9	72.9	70.9	76.0	70.9	72.1	70.8	75.0	72.7	72.9	69.8
2	Mean	46.1	48.1	47.6	46.3	45.8	46.2	46.4	46.5	48.8	48.0	51.3	51.1
	St dev	2.9	3.6	4.7	5.1	4.4	3.9	3.5	3.0	5.1	4.6	3.3	3.5
	Min	42.1	44.2	34.9	39.0	31.8	38.9	41.1	41.0	41.0	41.0	45.2	47.2
	Max	52.4	52.5	60.6	59.6	61.6	57.5	61.5	57.6	65.8	62.6	62.6	59.6
3	Mean	33.5	35.7	37.4	37.2	35.9	35.8	38.1	38.1	44.2	42.9	49.9	47.8
	St dev	4.1	5.4	7.0	8.9	7.0	6.4	6.8	5.9	9.2	8.4	5.6	5.9
	Min	27.7	30.9	27.7	28.8	21.6	27.7	29.7	28.8	29.7	30.8	38.0	40.1
	Max	40.1	43.2	61.6	63.7	68.8	58.5	64.6	58.5	72.0	70.9	65.7	61.6
4	Mean	70.2	74.6	67.4	63.9	57.9	59.5	56.7	56.6	59.2	56.4	68.7	65.8
	St dev	14.8	23.2	14.0	17.6	11.3	10.7	7.1	6.2	10.8	8.4	12.0	14.3
	Min	50.3	51.4	22.6	32.9	11.3	39.9	46.2	41.0	45.1	36.9	51.4	48.4
	Max	95.4	98.9	105.2	105.8	87.0	85.5	86.1	70.8	91.4	81.1	93.5	90.2
5	Mean	58.0	63.2	64.6	59.2	55.1	55.2	56.2	58.0	62.0	62.4	67.7	72.3
	St dev	12.0	9.4	15.2	21.4	17.2	14.9	9.2	10.2	10.2	10.6	6.8	6.0
	Min	38.0	52.3	32.8	31.8	9.2	31.8	39.0	31.8	35.9	21.6	48.3	65.6
	Max	83.6	75.2	97.6	101.7	106.8	93.5	89.2	87.3	101.8	83.1	80.1	87.3
7	Mean	131.1	131.3	133.1	131.8	131.9	131.1	133.7	133.4	138.1	136.6	144.3	141.6
	St dev	4.3	4.3	5.4	5.4	5.3	4.1	5.3	4.9	6.2	5.3	4.9	5.1
	Min	125.8	125.5	123.6	123.2	121.2	123.4	123.1	125.3	126.1	126.1	133.5	130.2
	Max	140.6	134.7	145.6	142.8	148.3	144.2	149.9	147.9	153.0	149.7	152.0	149.7
8	Mean	30.5	32.1	34.7	31.7	30.2	30.4	31.6	32.4	35.7	36.0	38.3	41.3
	St dev	7.8	3.7	9.5	11.4	10.0	8.9	5.8	6.2	6.6	6.7	4.0	4.4
	Min	20.5	28.8	18.5	18.5	10.3	17.4	22.5	20.5	21.6	18.5	27.7	34.9
	Max	51.4	37.1	60.6	63.7	66.8	57.5	53.4	50.4	65.8	53.4	46.2	51.4
9	Mean	48.3	52.2	49.5	46.2	43.8	44.2	43.6	43.3	46.3	44.5	52.7	50.8
	St dev	7.5	9.6	7.7	9.3	7.0	6.9	6.0	4.7	8.1	7.1	7.8	9.7
	Min	35.9	43.1	30.8	27.7	16.4	30.7	30.8	31.8	30.8	28.7	40.0	37.1
	Max	63.6	62.8	66.6	68.9	61.7	58.7	68.5	56.5	72.0	65.7	65.7	67.8

[a] See Section 2.1.4 for site fertility class definitions.

Classes of categorical forest attribute variables such as site fertility and tree species dominance in Finland and conifer/broad-leaved dominance in Italy are difficult to predict using satellite imagery and result in large prediction errors. This difficulty may be partially attributed to the fact that the assessment of these variables, particularly site fertility, is sometimes difficult even in the field, and that the forest attribute variables were assessed at stand-level. This latter fact could have a positive effect on prediction accuracy if forest stands are homogeneous. However, within stand variation for attributes such as proportions for dominant tree species, may lead to differences between stand-level field data and spectral data, particularly for small pixel sizes. Under such conditions, opportunities for increased accuracies using GA optimisation of weights for feature space variables are limited. These response variables were used, despite prediction difficulties, because they are often selected as key indicator variables in international assessments. Overall accuracy (OA) for site fertility for the Finnish test site with the independent reference and target data was approximately 60% for mineral soils and 50% for peatland soils. OA for tree species dominance was 65% on mineral soils and 80% on peatland soils. For the Italian site, when using GA optimisation and an independent target set, OA for forest type increased from 47.1% to 53.4%, and OA for CBM increased from 63.2% to 64.5%.

Increases in classical user's and producer's accuracies are somewhat greater than the OA measure reveals. For the ordinal scale site fertility variable in Finland, most incorrect predictions were not incorrect by more than one class. For operational applications, a classification error of one class is less severe than a larger error. Neither OA nor the producer's and user's accuracies by classes are sensitive to the magnitude of the incorrect predictions for ordinal scale variables. This problem is discussed in Tomppo (1992) and is a topic for future studies.

For all analyses, $k=5$ was used. Franco-Lopez et al. (2001) reported that increasing k beyond $k=1$ increased OA but decreased producer's accuracy for less populated classes in k-NN estimation of forest cover type and basal area classes. There seemed to be a tradeoff between OA and mean class accuracy when choosing the k-value. Inclusion of the k-value in the GA optimization remains a subject for future study.

For the Finnish analyses, we also tested GA optimization using fitness functions based on RMSEs and biases of volumes as in Tomppo and Halme (2004), and GA optimisation without using coarse scale forest variables in the distance metric. The overall result was that on mineral soils, accuracies were slightly greater when predicting site fertility and dominant tree species (Species Grouping 2) when using the new fitness functions than the fitness functions based on RMSEs and biases of volumes. On peatland soils, the accuracies did not increase in a same way, almost vice versa. This result reflects the difficulties in prediction for peatland forests where spectral values are affected by moisture variation and not directly linked to the variation of forest variables.

One of the new fitness functions (OA_f) was tested without use of the coarse scale forest variables in the distance metric for the Finnish analyses. For predicting site fertility, accuracies were only slightly less without coarse-scale forest variables than with them, or in some cases equal. The differences were small and could be caused by random variation. For peatland soils, the result was opposite, as well as for dominant tree species. Thus, the role and selection of coarse scale forest variables for predicting categorical forest variables require further investigation. For predicting volume, the positive effects of using of coarse scale variables in the distance metrics have been shown to be significant (Tomppo & Halme, 2004).

Table 20
Spectral statistics for reference and independent target sets for Finnish dominant tree species groups on peatland soils

ETM+ band	Measure	Tree species groups							
		Open		Pine		Spruce		Birch	
		Ref	Tar	Ref	Tar	Ref	Tar	Ref	Tar
1	Mean	66.3	65.8	61.8	62.1	60.5	60.0	62.7	60.9
	St dev	3.9	3.8	2.8	2.6	2.5	2.7	3.5	2.9
	Min	54.4	58.4	55.5	56.4	54.5	56.4	56.6	57.4
	Max	76.0	70.9	72.0	72.7	68.8	67.8	73.0	66.9
2	Mean	52.9	52.8	46.3	46.8	44.5	44.4	48.7	47.6
	St dev	6.0	6.1	3.1	3.2	3.7	3.6	4.2	3.1
	Min	31.8	38.9	39.0	41.0	39.0	39.0	41.1	42.1
	Max	65.8	62.6	63.6	59.5	57.4	54.5	60.6	55.5
3	Mean	52.8	52.7	38.3	38.9	33.1	33.3	38.6	35.8
	St dev	9.2	10.0	6.1	6.0	4.6	5.2	7.4	5.3
	Min	21.6	28.7	25.7	27.7	26.6	27.7	28.8	28.8
	Max	72.0	70.9	66.7	61.4	48.3	53.4	58.5	46.2
4	Mean	68.0	63.8	56.4	56.7	55.4	54.9	72.1	73.7
	St dev	16.4	13.6	6.7	7.1	9.7	11.5	11.5	11.3
	Min	11.3	32.9	44.2	36.9	40.9	41.0	44.2	51.4
	Max	93.5	90.2	87.0	81.3	93.2	98.9	105.2	105.8
5	Mean	61.2	64.0	58.4	60.1	47.2	47.0	70.8	65.4
	St dev	16.5	19.4	9.8	10.5	12.1	14.4	14.7	11.0
	Min	9.2	21.6	40.1	40.0	32.8	31.8	43.1	52.3
	Max	106.8	101.7	94.5	90.7	97.3	97.6	105.3	93.5
7	Mean	144.4	140.4	133.9	133.9	129.6	129.8	134.7	133.0
	St dev	5.4	6.4	5.0	4.9	4.0	4.2	6.3	4.4
	Min	128.4	123.9	121.2	125.3	123.1	123.2	124.3	125.5
	Max	153.0	149.7	149.9	149.7	144.8	139.9	149.9	143.8
8	Mean	36.7	38.8	32.7	33.8	25.7	26.1	37.5	33.0
	St dev	9.8	11.1	6.2	6.8	6.7	7.8	10.2	6.1
	Min	10.3	18.5	20.5	20.5	18.4	17.4	22.6	24.6
	Max	66.8	63.7	56.5	54.3	51.2	53.4	62.4	47.2
9	Mean	53.4	50.2	43.4	43.6	41.7	41.0	52.0	52.1
	St dev	10.7	10.5	5.0	5.3	6.4	6.5	6.1	5.7
	Min	16.4	27.7	30.8	28.7	30.8	30.8	38.1	39.0
	Max	72.0	67.8	68.7	60.6	66.6	62.8	64.8	69.8

Increases in accuracy were somewhat greater when using leave-one-out cross-validation for reference data than for independent target data, although not for all variables. These results could possibly be attributed to the fact the GA is trained with reference data and also, to some extent, problems in using leave-one-out RMSE as an error estimate (Kim & Tomppo, 2006), which causes underestimation of prediction error (Hammond & Verbyla, 1996). Also, prediction errors are overestimated as a result of field plot location errors (Halme & Tomppo, 2001; Verbyla & Hammond, 1995). Our tests confirm that leave-one-out approaches do not necessarily produce realistic error estimates, particularly when dealing with error estimates by strata such as mineral and peatland soils. Therefore the pixel-level results should often be considered in a relative manner rather than in terms of absolute measures of accuracy.

Prediction errors were also analysed with respect to the spectral distributions of reference and target data by classes of the categorical forest attribute variables. A portion of the prediction errors can be explained by differences in the ranges of spectral feature space variables for the reference and target data sets. A thorough diagnostic analysis, as well as a large reference set is an important precondition for successful use of nearest neighbour techniques (McRoberts, 2009-this issue).

Several questions should be addressed in future studies:

(a) What are relevant accuracy statistics for assessing the quality of predictions of categorical variables?
(b) Is it possible to derive feasible accuracy measures from the ones discussed in this study, e.g., weighting the measures by classes or taking into account the severity of pixel-level prediction errors?
(c) To what degree can accuracy be increased if the weights for the feature space variables are determined separately for individual response variables compared to estimating the weights simultaneously for all variables or for a group of variables?
(d) What are the effects of separate and joint estimation of the feature space weights on the covariances of the predictions of response variables relative to the covariances of observations of the variables?
(e) What are the effects of greater accuracy of field assessments and elimination of subjective factors on the increase in classification accuracy that can be attributed to GA

Table 21
Spectral statistics for reference and target sets for Italian forest type classes

TM band	Measure	Forest type class[a]													
		LSP		NSF		PF		B		OB		H		SS	
		Ref	Tar	Ref	Tar	Ref	Tar	Ref	Tar	Ref	Tar	Ref	Tar	Ref	Tar
1	Mean	53.4	60.7	50.2	52.5	53.9	56.7	53.0	52.9	57.0	55.2	57.0	57.1	66.6	59.7
	St dev	11.7	39.4	3.9	19.3	5.1	6.2	3.7	3.4	7.1	5.1	4.7	5.5	42.0	10.6
	Min	44	45	41	43	47	49	46	47	47	49	49	47	47	46
	Max	137	255	71	255	79	77	69	66	82	67	77	73	255	88
2	Mean	23.3	27.8	20.5	22.1	21.7	24.0	21.8	21.6	24.3	23.5	23.8	24.5	31.7	27.8
	St dev	7.7	22.2	3.0	13.6	3.7	4.6	2.7	2.3	4.7	3.7	3.4	4.1	25.2	6.2
	Min	15	19	13	17	18	19	17	16	16	18	18	18	18	19
	Max	73	144	35	166	41	39	36	27	39	32	37	36	144	42
3	Mean	20.6	26.1	17.0	18.9	19.1	21.6	17.9	17.9	21.0	19.9	20.0	20.6	32.2	27.4
	St dev	10.4	28.3	3.3	17.5	5.2	6.4	3.6	2.8	6.4	4.6	4.4	4.3	32.6	9.3
	Min	12	14	11	12	14	16	13	14	13	14	14	15	16	17
	Max	92	173	30	205	46	42	43	30	42	32	38	36	177	46
4	Mean	74.1	82.7	69.0	69.2	72.1	78.5	101.5	95.3	100.3	100.0	99.5	101.3	82.5	79.3
	St dev	23.0	29.2	24.1	25.7	22.8	24.7	28.2	26.1	27.6	31.7	21.0	27.5	27.7	28.1
	Min	10	40	13	35	33	50	41	35	41	43	50	46	44	36
	Max	143	162	144	180	124	130	164	157	145	161	146	149	171	145
5	Mean	56.3	66.3	42.3	43.8	51.7	59.1	66.0	62.8	68.7	69.7	72.5	73.5	79.1	74.1
	St dev	24.0	34.1	18.6	25.0	17.6	20.8	18.8	16.4	18.6	20.9	17.2	19.7	38.5	26.7
	Min	11	29	8	16	19	33	18	24	24	34	33	26	32	25
	Max	177	214	109	233	103	99	128	92	119	107	102	100	222	124
7	Mean	20.0	25.3	14.1	15.9	17.1	20.5	20.2	18.8	22.0	21.9	22.8	22.9	32.9	30.3
	St dev	10.4	20.8	6.1	13.6	6.5	8.5	6.0	5.1	9.1	7.5	6.4	6.1	24.6	14.7
	Min	4	10	3	6	6	11	8	9	7	10	11	10	13	9
	Max	78	132	49	140	45	49	52	35	61	41	39	36	131	61

[a] See Table 4 for forest type class definitions.

optimisation? Also, what are the effects on the results for different accuracy measures?

In this study, the effects of estimating weights for feature space variables separately or jointly for the two response variables of interest were evaluated for both test sites. Only small differences in accuracies were found for the Italian data, whereas somewhat greater accuracies were found for the Finnish data when estimating the weights separately. A possible explanation is that the Italian response variables were highly dependent whereas the dependence was less for the Finnish variables.

One outcome from the study is that GA optimisation could be a tool to weight the spectral variables to tailor a classification algorithm for a specific technical objective such as maximising overall accuracy, maximising the greatest user's or producer's accuracy, or minimising the greatest class error. Although results for different fitness functions and different accuracy measures were somewhat mixed and further investigations are necessary before a final recommendation can be made, a clear conclusion is that selection of the fitness functions and accuracy measures depend on user needs.

Acknowledgements

We thank three anonymous reviewers for a thorough review and many valuable comments and suggestions for improvements. We also thank the Italian National Forest Service (CFS) for allowing us to use the INFC data.

References

Baffetta, F., Fattorini, L., Franceschi, S., & Corona, P. (2009). Design-based approach to k-NN technique for coupling field and remotely sensed data in forest surveys. *Remote Sensing of Environment*, *113*, 463–475 (this issue).

Bertini, R., Chirici, G., Corona, P., & Travaglini, D. (2007). Comparison between parametric and non-parametric methods for the spazialization of forest standing volume by integrating field measures, remote sensing data and ancillary data.*For-est@*, *4*, 110–117. Online URL: http://www.sisef.it/forest@/

Boehringer, S., Vollman, T., Tasse, C., Wortz, R. P., Gillessen-Kaebach, G., & Horst, B. (2006). Syndrome identification based on 2D analysis software. *European Journal of Human Genetics*, *14*, 1082–1089.

Cajander, A. K. (1926). The theory of forest types.*Acta Forestalia Fennica*, *29.3* 108. pp.

Chiavetta, U., Chirici, G., Corona, P., Marchetti, M., & Ottaviano, M. (2007). Confronto tra diverse tecniche di spazializzazione dei dati inventariali dell'Alto Molise. Seminario: Telerilevamento e spazializzazione nel monitoraggio forestale e ambientale per la difesa degli ecosistemi. Mercoledi 24 gennaio 2007 – Facoltà di Scienze MMFFNN, Pesche (IS).

Cohen, J. (1960). *A coefficient of agreement for nominal scalesEduc. Psych. Measure*, *Vol 20*. (pp. 37–46).

Congalton, R. G., & Green, K. (1999). *Assessing the accuracy of remotely sensed data: Principles and practices.*Boca Raton: Lewis Publisher 137 pp.

COST E43 (2007). *Harmonization of the national forest inventories of Europe.* http://www.metla.fi/eu/cost/e43/

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. (eds.). 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.

FAO – Global Forest Resources Assessment. (2000). Main report. FAO Forestry Paper: 140. www.fao.org/forestry/fo/fra/main/index.jsp

Fattorini, L., Marcheselli, M., & Pisani, C. (2006). A three-phase sampling strategy for large-scale multiresource forest inventories. *Journal of Agricultural, Biological, and Environmental Statistics*, *11*(3), 1–21.

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.

Foody, G. M. (2004). Thematic map comparison: evaluating the statistical significance of differences in classification accuracy. *Photogramm Engineering & Remote Sensing*, *70*, 627–633.

Gjertsen, A. (2005). Accuracy of forest mapping based on Landsat TM data and a kNN-based method. *Remote Sensing of Environment*, *110*, 420–430.

Goldberg, D. (1989). *Genetic algorithms in search, optimization, and machine learning.* San Mateo, California: Addison-Weslery.

Gonzalez, R. C., & Woods, R. E. (2002). *Digital image processing*, 2nd ed. Upper Saddle River, NJ: Prentice Hall.

Haapanen, R., Ek, A. R., Bauer, M. E., & Finley, A. O. (2004). Delineation of forest/nonforest land use classes using nearest neighbor methods. *Remote Sensing of Environment*, *89*, 265–271.

Halme, M., & Tomppo, E. (2001). Improving the accuracy of multisource forest inventory estimates by reducing plot location error – A multi-criteria approach. *Remote Sensing of Environment*, *78*, 321–327.

Hammond, T. O., & Verbyla, D. L. (1996). Optimistic bias in classification accuracy assessment. *International Journal of Remote Sensing*, *17*, 1261–1266.

He, H., Hawkins, S., Graco, W., & Yao, X. (2000). Application of genetic algorithm and k-nearest neighbour method in real world medical fraud detection problem. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, *4*, 130–137.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems.* Ann Arbor, Michigan: University of Michigan Press.

Holmström, H., & Fransson, J. E. S. (2003). Combining remotely sensed optical and radar data in k-NN estimation of forest variables. *Forest Science*, *4*, 409–418.

INFC (2007). Le stime di superficie 2005 – Prima parte. In G. Tabacchi, F. De Natale, L. Di Cosmo, A. Floris, C. Gagliano, P. Gasparini, L. Genchi, G. Scrinzi, & V. Tosi (Eds.), *Inventario Nazionale delle Foreste e dei Serbatoi Forestali di Carbonio. MiPAF – Ispettorato Generale Corpo Forestale dello Stato, CRA – ISAFA, Trento* [on line] URL: http://www.infc.it – [on digital support].

Katila, M., & Tomppo, E. (2001). Selecting estimation parameters for the Finnish multisource National Forest Inventory. *Remote Sensing of Environment*, *76*, 16–32.

Katila, M., & Tomppo, E. (2006). Sampling simulation on multi-source output forest maps – An application for small areas. In M. Caetano, & M. Painho (Eds.), *Proceedings of 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, 5–7 July 2006, Lissabon, Portugal* (pp. 614–623). Lissabon: Instituto Geogràfico Português.

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

Koukal, T., Suppan, F., & Schneider, W. (2007). The impact of radiometric calibration on the accuracy of kNN-predictions of forest attributes. *Remote Sensing of Environment*, *110*, 431–437.

Lehto, J. & Leikola, M. (1987). Käytännön metsätyypit. Kirjayhtymä. Helsinki. (in Finnish).

LeMay, V., & Temesgen, H. (2005). Comparison of nearest neighbor methods for estimating basal area and stems per ha using aerial auxiliary variables. *Forest Science*, *51*(2), 109–119.

Magnussen, S., McRoberts, R. E., & Tomppo, E. O. (2009). A model-based estimator of the mean square error of k-nearest neighbour predictions with remotely-sensed ancillary variables.*Remote Sensing of Environment*, *113*, 476–488 (this issue).

Manual of Remote Sensing (1983). *American society of photogrammetry.* Falls Church, Virginia.

Maselli, F., Chirici, G., Bottai, L., Corona, P., & Marchetti, M. (2005). Estimation of Mediterranean forest attributes by the application of k-NN procedure to multitemporal Landsat ETM+ images. *International Journal of Remote Sensing*, *17*, 3781–3796.

McInerney, D., Pekkarinen, A., & Haakana, M. (2005). Combining Landsat ETM+ with field data for Ireland's National Forest inventory – A pilot study for Co. Clare. In H. Olsson (Ed.), *Proceedings of Forest Sat 2005: Operational tools in forestry using remote sensing techniques. 31 May – 1 June 2005* (pp. 12–16). Sweden: Borås.

McKenzie, D. P., Mackinnon, A. J., Péladeau, N., Onghena, P., Bruce, P. C., & Clarke, D. M. (1996). Comparing correlated kappas by resampling: Is one level of agreement significantly different from another? *Journal of Psychiatric Research*, *30*, 483–492.

MCPFE (1997). Work-programme on the conservation and enhancement of biological and landscape biodiversity in forest ecosystems. *Adopted at Expert-Level by the Third Meeting of the Executive Bureau of the Pan-European Biological and Landscape Diversity Strategy 20–21 November, 1997, Geneva, Switzerland.*

MCPFE (2003). *Vienna Declaration and Vienna Resolutions adopted at the Fourth Ministerial Conference on the Protection of Forests in Europe, 28–30 April 2003, Vienna Austria.*

MCPFE (2003). *Improved Pan-European Indicators for Sustainable Forest Management as adopted by the MCPFE Expert Level Meeting 7–8 October 2002, Vienna, Austria.*

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

McRoberts, R. E. (2009). Workplace violence: Diagnostic tools for nearest neighbors techniques when used with satellite imagery. *Remote Sensing of Environment*, *113*, 489–499 (this issue).

McRoberts, R. E., Holden, G. R., Nelson, M. D., Liknes, G. C., & Gormanson, D. D. (2005). Using satellite imagery as ancillary data for increasing the precision of estimates for the Forest Inventory and Analysis program of the USDA Forest Service. *Canadian Journal of Forest Research*, *36*, 2968–2980.

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

McRoberts, R. E., Tomppo, E. O., Finley, A. O., & Heikkinen, J. (2007). Estimating areal means and variances of forest attributes using the k-Nearest Neighbor technique and satellite imagery. *Remote Sensing of Environment*, *111*, 466–480.

Moeur, M. (1988). Nearest neighbor inference for correlated multivariate attributes. In A. R. Ek, S. R. Shifley, & T. E. Burk (Eds.), *Proceedings, IUFRO conference; 1987 August 23–27; Minneapolis, Minnesota, USA: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment StationForest growth and modeling*, *Vol. 2*. (pp. 716–723).

Moeur, M., & Stage, A. R. (1995). Most similar neighbor — An improved sampling inference procedure for natural resource planning. *Forest Science*, *41*(2), 337–359.

MPCI (2007). *The Montréal Process.* http://www.rinya.maff.go.jp/mpci/

Muinonen, E., & Tokola, T. (1990). An application of remote sensing for communal forest inventory. In: The usability of remote sensing for forest inventory and planning. *Proceedings from SNS/IUFRO workshop* (pp. 35–42). Umea°, Sweden: Remote Sensing Laboratory, Swedish University of Agricultural Sciences (Report 4).

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 Resource Management and Geomatics. Acta Universitatis Agriculturae Sueciae. Silvestria 17.

Nishii, R., & Tanaka, S. (1999). Accuracy and inaccuracy assessments in land-cover classification. *IEEE Transactions on Geosciences and Remote Sensing*, *37*(1), 491–498.

Ohmann, J. L., & Gregory, M. J. (2002). Predictive mapping of forest composition and structure with direct gradient analysis and nearest neighbor imputation in coastal Oregon, U.S.A. *Canadian Journal of Forest Research*, *32*, 725–741.

Oliveira, L. S., Sabourin, R., Bortolozzi, F., & Suen, C. Y. (2003). A methodology for feature selection using multi-objective genetic algorithms for handwritten digit string recognition. *International Journal of Pattern Recognition and Artificial Intelligence*, *17*, 903–930.

Punch, W. F., Goodman, E. D., Pei, M., Chia-Shun, L., Hovland, P., & Enbody, R. (1993). In S. Forrest (Ed.), *Proceedings of the 5th International Conference on Genetic Algorithms, ICGA-93* San Mateo, CA: Morgan Kaufmann Publishers.

Raymer, M. L., Sanschagrin, P. C., Punch, W. F., Venkataraman, S., Goodman, E. D., & Kuhn, L. A. (1997). Predicting conserved water-mediated and polar ligand interactions in proteins using a K-nearest-neighbors genetic algorithm. *Journal of Molecular Biology*, *265*, 445–464.

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

Romualdi, C., Campanaro, S., Campagna, D., Celegato, B., Cannata, N., Toppo, S., Valle, G., & Lanfranchi, G. (2003). Pattern recognition in gene expression profiling using DNA array: A comparative study of different statistical methods applied to cancer classification. *Human Molecular Genetics*, *2002*, 823–836.

Tabacchi, G., De Natale, F., Floris, A., Gasparini, P., Gagliano, C., Scrinzi, G., & Tosi, V., eds. (2007). Italian national forest inventory: methods, state of the project and future developments. In McRoberts, R.E., Reams, G.A., Van Deusen, P.C., & McWilliams, W.H. (eds.). Proceedings of the seventh annual forest inventory and analysis symposium; 2005 October 3–6; Portland, ME. Gen. Tech. Report WO-77. Washington, DC: U.S. Department of Agriculture Forest Service. 319 p.

Temesgen, H., LeMay, V. M., Froese, K. L., & Marshall, P. L. (2003). Imputing tree-lists from aerial attributes for complex stands of south-eastern British Columbia. *Forest Ecology and Management*, *177*, 277–285.

Tokola, T., Pitkanen, J., Partinen, S., & Muinonen, E. (1996). Point accuracy of a non-parametric method in estimation of forest characteristics with different satellite materials. *International Journal of Remote Sensing*, *17*, 2333–2351.

Tomppo, E. (1990). Designing a satellite image-aided national forest survey in Finland. *The Usability of Remote Sensing For Forest Inventory and Planning, Proceedings from SNS/IUFRO workshop in Umeå 26 – 28 February 1990* (pp. 43–47). Remote Sensing Laboratory: Swedish University of Agricultural Sciences Report 4.

Tomppo, E. (1991). Satellite image-based national forest inventory of Finland. *Proceedings of the symposium on Global and Environmental Monitoring, Techniques and Impacts, September 17–21, 1990 Victoria, British Columbia CanadaInternational Archives of Photogrammetry and Remote Sensing*, *Vol. 28*. (pp. 419–424) Part 7–1.

Tomppo, E. (1992). Satellite image aided forest site fertility estimation for forest income taxation. Tiivistelmä: Satellittikuva-avusteinen metsien kasvupaikkaluokitus metsäverotusta varten.*Acta Forestalia Fennica*, *229* 70 pp.

Tomppo, E. (1996). Multi-source National Forest Inventory of Finland. In: R. Vanclay, J. Vanclay, & S. Miina (Eds.), 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, (pp. 27–41) Tampere, Finland. EFI Proceedings, 7. European Forest Institute. Joensuu. Finland.

Tomppo, E. (2006). The Finnish multi-source national forest inventory – Small area estimation and map production. *Forest inventory – methodology and applications* (pp. 195–224). Dordrecht, The Netherlands: Springer.

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., & Halme, M. (2004). Using coarse scale forest variables as ancillary information and weighting of variable sin k-NN estimation: A genetic algorithm approach. *Remote Sensing of Environment*, *92*, 1–20.

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

Tomppo, E., Olsson, H., Ståhl, G., Nilsson, M., Hagner, O., & Katila, M. (2008). Combining national forest inventory field plots and remote sensing data for forest databases. *Remote Sensing of Environment*, *112*, 1982–1999.

Verbyla, D. L., & Hammond, T. O. (1995). Conservative bias in classification accuracy assessment due to pixel-by-pixel comparison of classified images with reference grids. *International Journal of Remote Sensing*, *16*, 581–587.