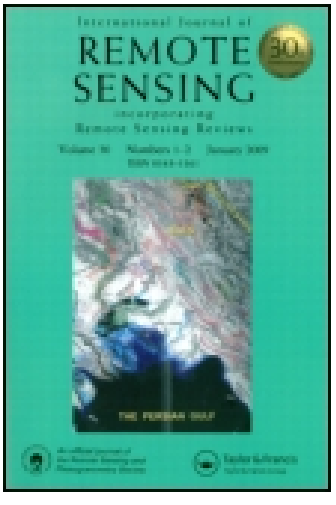

International Journal of Remote Sensing

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tres20

Forest attribute imputation using machine-learning methods and ASTER data: comparison of k-NN, SVR and random forest regression algorithms

Shaban Shataee [a], Syavash Kalbi [b], Asghar Fallah [b] & Dieter Pelz [c]

[a] Department of Forestry, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran

[b] Department of Forestry, Sari University of Agricultural Sciences and Natural Resources, Sari, Iran

[c] Department of Forest Biometry, Freiburg University, Freiburg, Germany

Published online: 29 May 2012.

To cite this article: Shaban Shataee, Syavash Kalbi, Asghar Fallah & Dieter Pelz (2012) Forest attribute imputation using machine-learning methods and ASTER data: comparison of k-NN, SVR and random forest regression algorithms, International Journal of Remote Sensing, 33:19, 6254-6280, DOI: 10.1080/01431161.2012.682661

To link to this article: http://dx.doi.org/10.1080/01431161.2012.682661

International Journal of Remote Sensing
Vol. 33, No. 19, 10 October 2012, 6254–6280

Taylor & Francis
Taylor & Francis Group

Forest attribute imputation using machine-learning methods and ASTER data: comparison of *k*-NN, SVR and random forest regression algorithms

SHABAN SHATAEE*†, SYAVASH KALBI‡, ASGHAR FALLAH‡ and DIETER PELZ§

†Department of Forestry, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran

‡Department of Forestry, Sari University of Agricultural Sciences and Natural Resources, Sari, Iran

§Department of Forest Biometry, Freiburg University, Freiburg, Germany

(*Received 17 September 2010; in final form 24 September 2011*)

In a small case study of mixed hardwood Hyrcanian forests of Iran, three non-parametric methods, namely *k*-nearest neighbour (*k*-NN), support vector machine regression (SVR) and tree regression based on random forest (RF), were used in plot-level estimation of volume/ha, basal area/ha and stems/ha using field inventory and Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) data. Relevant pre-processing and processing steps were applied to the ASTER data for geometric and atmospheric correction and for enhancing quantitative forest parameters. After collecting terrestrial information on trees in the 101 samples, the volume, basal area and tree number per hectare were calculated in each plot. In the *k*-NN implementation using different distance measures and *k*, the cross-validation method was used to find the best distance measure and optimal *k*. In SVR, the best regularized parameters of four kernel types were obtained using leave-one-out cross-validation. RF was implemented using a bootstrap learning method with regularized parameters for decision tree model and stopping. The validity of performances was examined using unused test samples by absolute and relative root mean square error (RMSE) and bias metrics. In volume/ha estimation, the results showed that all the three algorithms had similar performances. However, SVR and RF produced better results than *k*-NN with relative RMSE values of 28.54, 25.86 and 26.86 (m^3 ha^{-1}), respectively, using *k*-NN, SVR and RF algorithms, but RF could generate unbiased estimation. In basal area/ha and stems/ha estimation, the implementation results of RF showed that RF was slightly superior in relative RMSE (18.39, 20.64) to SVR (19.35, 22.09) and *k*-NN (20.20, 21.53), but *k*-NN could generate unbiased estimation compared with the other two algorithms used.

1. Introduction

The forests of Iran, with an area of about 12.4 million hectares, comprise 7.4% of the country's area (FAO 2002). Of the five large vegetation regions in Iran, the most

*Corresponding author. Email: Shataee@yahoo.com

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online © 2012 Taylor & Francis
http://www.tandfonline.com
http://dx.doi.org/10.1080/01431161.2012.682661

important according to forest density, canopy cover and diversity is the Hyrcanian (Caspian) region. The region covers an area of 1 925 125 ha, extending throughout the south coast of the Caspian Sea in the northern part of the country. Quantitative and qualitative attributes in these forests are updated in 10 year cycles in forest management plans. These collections are time-consuming and cost-intensive. On the other hand, remote-sensing imagery has considerable potential for monitoring and continuously predicting forest attributes. Combining field data with satellite data can be useful for producing a continuous map of forest attributes through classification or regression. So far, various remote-sensing data in terms of spatial resolutions from very high (airborne or spaceborne systems such as digital aerial photographs or QuickBird and IKONOS imagery) to medium spatial resolution imagery have been used in the estimation of forest attributes. Although the use of high and very high spatial resolution imagery can result in high-precision imputation, especially for species- or tree- and plot-level estimation, medium spatial resolution imagery is the most popular for use in both stand- and plot-level estimation. So far, various types of medium-resolution imagery such as Landsat-TM (Franco-Lopez *et al*. 2001, Holmstrom 2002, Makela and Pekkarinen 2004, Huiyan *et al*. 2006, Gjertsen 2007, McRobert 2009, Latifi *et al*. 2010), Landsat-ETM+ (Lister *et al*. 2004, Kajisa *et al*. 2008) and SPOT (Holmström 2002, Reese *et al*. 2002) have been used for the prediction of forest attributes. Following the ETM+ sensor, new sensors were launched into space for improving Earth observation. In 1999, NASA launched the Terra satellite in a cooperative Earth Observing System (EOS) project. Terra is equipped with five instruments including the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER). These instruments provide different images for different goals in terms of spectral, spatial and temporal resolutions. The ASTER instrument provides higher spatial resolution images compared with the other Terra systems. It contains three subsystems operating in a wide spectrum of wavelengths, including visible and near-infrared (VNIR) with a spatial resolution of 15 m in three bands, shortwave infrared (SWIR) or middle infrared with a spatial resolution of 30 m in six wavelength bands and thermal infrared (TIR) with a spatial resolution of 90 m in five bands. Owing to these variants and high geometric and radiometric accuracies, the ASTER data offer new capabilities for investigation in different applications. The VNIR and SWIR images have more potential to expose biophysical characteristics of vegetation compared to other popular medium-resolution images such as Landsat-TM or ETM+.

ASTER data have been used for different applications. In forest applications, Muukkonen and Heiskanen (2005) estimated the biomass of forest stands using ASTER images and stand-wise forest inventory data by non-linear regression and neural networks. Heiskanen (2005) modelled statistical relationships between field measurements and ASTER spectral reflectance values using canonical correlation analysis (CCA) and reduced major axis (RMA) regression to demonstrate the potential of ASTER data for estimating biomass and the leaf area index (LAI) in the mountainous birch forests in northern Finland. Falkowski *et al*. (2004) tested the capability of ASTER data for mapping and characterizing canopy closure and crown bulk density using field data and spectral values in northern Idaho. ASTER data have also been used for the estimation of forest attributes using the *k*-nearest neighbour (*k*-NN) algorithm, for example by Fuchs *et al*. (2009) for estimating the aboveground carbon in the catchments of the Siberian forest tundra and by Kutzer (2008) for monitoring the off-reserve forest resources in Ghana. In all of these studies, estimation was

done in softwood forests or, rarely, in mixed softwood and hardwood forests with low species richness.

In the Hyrcanian region, forests are composed of mixed hardwood tree species with a high species richness. Our literature review on these forests showed that there are only a few remote-sensing-based studies on the estimation of forest attributes. In these studies (Khorrami 2004, Mohammadi *et al*. 2010), estimation was done using only ETM+ imagery because of its low price and high coverage. Khorrami (2004) and Mohammadi *et al*. (2010) investigated the relations between spectral values of ETM+ imagery and volume values of field-based inventory data using parametric multiple linear regressions to predict the plot-based volume of forest trees and modelled forest volume and tree density at the plot level. They suggested that using other parametric or non-parametric algorithms might improve the estimation. The literature review indicated that ASTER data have not been used for estimation of forest attributes such as volume, basal area and tree density in the Hyrcanian forests.

In the past two decades, parametric algorithms such as multiple linear and non-linear regressions have been popular methods for estimating forest attributes using remote-sensing data (Hall *et al*. 2006, Huiyan *et al*. 2006, Sivanpillai *et al*. 2006, Gebreslasie *et al*. 2008, Mohammadi *et al*. 2010). Recently, non-parametric algorithms have been explored for the estimation of forest attributes because of advantages such as their flexibility and ability to describe non-linear dependencies compared to parametric algorithms (Franco-Lopez *et al*. 2001, Makela and Pekkarinen 2004, Hyvönen 2007, McRobert *et al*. 2007, Sironen *et al*. 2010). One of the greatest advantages of non-parametric algorithms is that they are free from assumptions of any given probability distribution and observations are assumed to be independent of each other (Sironen *et al*. 2010). Machine-learning algorithms are groups of data mining and non-parametric-based algorithms that use numerous independent variables in classification and regression applications. *k*-NN, support vector machine regression (SVR) and random forest (RF) are the three popular and most commonly used machine-learning algorithms for the estimation and imputation of forest attributes.

1.1 k-*nearest neighbour*

The *k*-NN method is the simplest machine-learning algorithm used for both classification and regression. For regression, *k*-NN simply assigns the property value for the object to be the average of the values of its *k*-NNs. *k*-NN is widely used for the estimation of forest attributes using various remote-sensing data (Franco-Lopez *et al*. 2001, Katila and Tomppo 2001, Ohmann and Gregory 2002, Makela and Pekkarinen 2004, Finley *et al*. 2006, McRobert *et al*. 2007, Tatjana *et al*. 2007).

In *k*-NN implementations, three parameters should be determined including the number of *k*, the type of distance measure and weights for nearest neighbours. For continuous response variables, the *k*-NN prediction for the *i*th target set element is

$$\overline{y_i} = \frac{1}{w_i} \sum_{j=1}^{k} w_{ij}\, y_j^i, \tag{1}$$

where $\{y_{ij}, j = 1, 2, \ldots, k\}$ is the set of response variable observations for k reference set elements that are nearest to the ith target set element in feature space with respect

to a distance measure, d, and w_{ij} is the weight assigned to the jth nearest neighbour with

$$w_i = \sum_{j=1}^{k} w_{ij}. \tag{2}$$

The optimal k depends on the data and can be determined by the user. A smaller k may lead to higher variance (less stable) and a larger k often leads to higher bias (less precise) (Kozma 2008). In distance-weighted k-NN, each of k's neighbours is weighted according to its distance to a reference unit and gives greater weight to closer neighbours:

$$w = \frac{1}{d\,(x_{\mathrm{q}}, x_{\mathrm{i}})^2}, \tag{3}$$

where x_{q} and x_{i} are the target and reference units, respectively, and d is the distance between them.

There are many distance measures for calculating the distance between neighbours. In Statistica software (Statistica 2010), the distance measures are the Euclidean, squared Euclidean, city-block (Manhattan) and Chebychev distances. The Euclidean distance is the most commonly used distance measure. It is a simple geometric distance in multidimensional space (Statistica 2010):

$$D(x, p) = \sqrt{(x - p)^2}\,. \tag{4}$$

In squared Euclidean, the distance between target and reference units would be squared in order to give progressively greater weight to objects that are closer or more similar:

$$D(x, p) = (x - p)^2\,. \tag{5}$$

In Manhattan, absolute value distances are considered. However, the effect of single large differences (outliers) is dampened because they are not squared (Statistica 2010):

$$D(x, p) = \mathrm{abs}\,(x - p)\,. \tag{6}$$

The Chebychev distance, which is also called maximum value distance, examines the absolute magnitude of differences between coordinates of a pair of objects. This distance can be used for both ordinal and quantitative variables and may be appropriated when one wants to define two objects as 'different' if they are different on any one of the dimensions:

$$D(x, p) = \max\,(|x - p|)\,. \tag{7}$$

In all formulae, D is the distance between target and reference units, x is the target unit and p is the reference unit.

Among the four distance measures mentioned above, the squared Euclidean distance is the most commonly used distance measure (Franco-Lopez *et al*. 2001, Reese

et al. 2002, Sironen *et al*. 2010). However, in some studies, the other distance measures such as Manhattan (Sironen *et al*. 2001), weighted Euclidean (Eskelson *et al*. 2009) and non-weighted Euclidean (McRobert 2009) were used for the estimation of different objects.

1.2 *Support vector machine regression*

The SVM algorithm is a family of classification and regression techniques based on statistical learning theory (Walton 2008). It is a non-linear generalization of the generalized portrait algorithm developed by Vapnik (1998). Generally, SVMs focus on the boundary between classes and map the input space created by independent variables using a non-linear transformation according to a kernel function. Linear, polynomial, radial basis function (RBF) and sigmoid are the four most commonly used kernel types. The RBF is the most popular kernel, which is used in SVMs (Cortez and Morais 2007, Durbha *et al*. 2007). In this mapped high-dimensional space, an optimal linear separator or hyperplane is found that maximizes the margin between classes. By maximizing the margin, the solution is generalized, and over-fitting will be reduced (Walton 2008). The performance of SVR is affected by three parameters, the capacity (C), which presents a trade-off between the model complexity and the amount up to which deviations larger than C are tolerated, epsilon (ε), which controls the width of the ε-insensitive zone used to fit the training data, and gamma (γ), which is a kernel function parameter (Cortez and Morais 2007). The value of epsilon can affect the number of support vectors (SVs) used to construct the regression function. A larger epsilon will result in fewer selected SVs, less complex regression estimates and less training time (Wang *et al*. 2005). The kernel parameter can be selected by some means such as prior knowledge or user experiences or it can be determined by fixing a parameter and controlling other parameters. In SVR, the algorithm is trying to find a hyperplane that can accurately predict the distribution of information, but not the plane on how to classify the data (Wang *et al*. 2009). SVR has theoretical advantages over NNs, such as absence of local minima in the model optimization phase (Cortez and Morais 2007). In some studies (Zhang and Ma 2008, Shafri and Ramle 2009, Ostapowicz *et al*. 2010), the SVM has been used for forest classification. According to our literature reviews, SVR has been used for the estimation of forest biophysical variables in a few studies. Wang and Brenner (2009) used SVR for forest canopy cover mapping using Landsat ETM+ and demonstrated that SVR is an efficient algorithm that can be used as an alternative to linear regression algorithms. Walton (2008) in a study of urban forest canopy cover mapping using ETM+ compared the performance of SVR, RF and cubist regression methods and concluded that implementing SVR using tasselled cap bands led to the best estimation, whereas the cubist method using a combination of reflectance and tasselled cap bands was the best estimator when predicting impervious surface cover. Dalponte *et al*. (2008) used SVR for tree-level estimation of biomass in alpine forest areas using lidar data and demonstrated that SVR was effective for the estimation of stem diameters and tree biomass. Durbha *et al*. (2007) used SVR for retrieval of LAI from a multiangle imaging spectroradiometer.

1.3 *Random forest*

RF is an extension of the classification and regression tree (CART) methods (Breiman 2001). RF can be used for regression-type problems (to predict a continuous dependent variable) and classification problems (to predict a categorical dependent

variable). In regression problems, RF is an arbitrary number (ensemble) of simple trees (subset of independent variables), which are used to vote their responses to be combined (averaged) to obtain an estimate of the dependent variable. The data and variables can be randomly sampled in an iteratively bagging bootstrap sampling to generate a forest of regression trees. The mean square error for an RF is calculated as

$$\text{random forest predictions} = \frac{1}{k}\sum_{k=1}^{k} k_{\text{th}}\ \text{tree response}, \tag{8}$$

where the index k runs over the individual trees in the forest.

The predictions of RF are taken to be the average of the predictions of the trees:

$$\text{mean error} = (\text{observed-tree response})^2. \tag{9}$$

The distance between target and reference units is calculated as one minus the proportion of trees (terminal nodes) from all regression trees, where a target sample is in the same terminal node as a reference sample (for more details, see Breiman 2001). Implementation of RF is dependent on the regularization of a decision tree and stopping parameters. The decision tree model parameters include the maximum number of trees that must be grown in the forest and the number of variables (k predictor or independent variables in each node for predicting dependent values) that are randomly selected in each node (Statistica 2010). Alternatively, choosing a small number of predictor variables may downgrade prediction performance, because this can exclude variables that may account for most of the variability and trend in the data (Statistica 2010). The stopping parameters or control parameters are used to stop running the algorithm when satisfactory results have been achieved.

1.4 *Objectives*

The Hyrcanian forests are mixed and mountainous, with unevenly aged structures that are topographically complex and comprise different hardwood species. Our literature review showed that no study has been carried out on the application of machine-learning algorithms (k-NN, SVR or RF algorithms with their different options) using ASTER data for the estimation of volume, basal area or tree density in the Hyrcanian forest. Therefore, in this study, three of the most commonly used non-parametric machine-learning methods were applied for the first time for the estimation of forest attributes using ASTER data. Consequently, the main objective of this study was examination and comparison of the three most commonly used machine-learning methods (k-NN, SVR and RF) in plot-level imputations of volume, basal area and tree number per hectare using ASTER data in the Hyrcanian forests.

2. Study materials

2.1 *Study area*

The study area is located in the Hyrcanian forests, district 1 of Darabkola's forests, northern Iran (figures 1(*a*)–(*c*)). Darabkola's forest, with an area of about 2500 ha, is a natural and mature forest with uneven-aged and dense to semidense stands, comprising mixed hardwood types including *Fagus* dominant, mixed *Fagus*, *Carpinus–Fagus*,

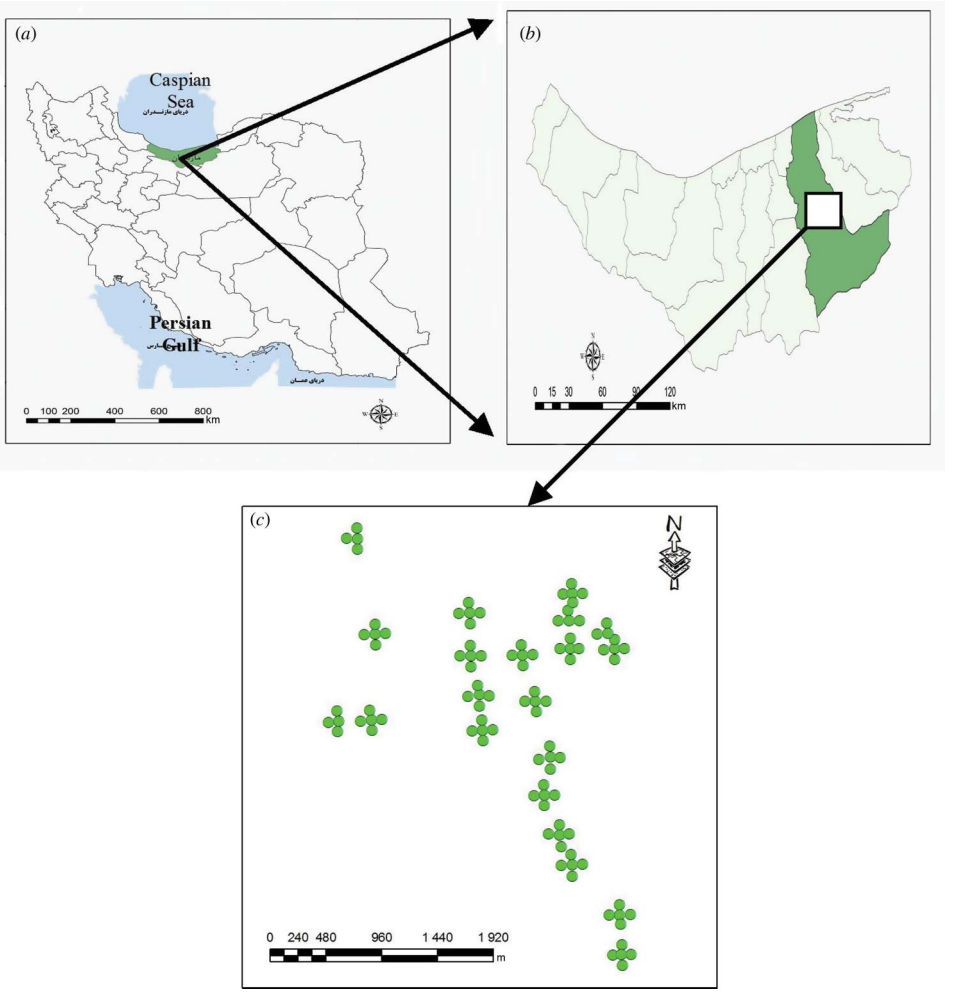

Figure 1. Location of plots (*a–c*) in Darabkola's forestry district, Mazandaran province, northern Iran.

Fagus–Carpinus, *Carpinus–Tilia* and so on. The elevation ranges from 140 to 920 m from free sea level and the general aspect of the study area is northern, but with some fine different slope aspects. However, the sample plots were selected on the northern slopes and only in the *Fagus* stands. The logging is done using a method of selecting individual trees, so as to be close to a natural silviculture. The investigation was carried out in only a part of Darabkola's district 1, about 1224 ha, where Persian beech (*Fagus orientalis*) is a dominant species (*Fagus* dominant type). The volume of *Fagus* stands based on our plot measurements ranged from 131 to 643 m^3 ha^{-1} (see also table 5).

2.2 *ASTER data*

For this study, the images of the VNIR and SWIR subsystems were used from the ASTER system, acquired on 10 November 2007. This date roughly marks the end of the summer season in the studied area; the forest crown cover is completely full and in

a stable condition in terms of vegetation characteristics. The VNIR imagery contains three bands (green, red and near-infrared) with a spatial resolution of 15 m. The SWIR imagery has a spatial resolution of 30 m in six middle infrared wavelength bands from 1.6 to 2.43 μm. Having six bands in the middle infrared spectrum in ASTER imagery is superior compared with Landsat imagery and is expected to have better results in similar studies.

2.3 *Field inventory data*

Operational application of the *k*-NN and other machine-learning methods has also shown that the predictions may be biased if the area of interest is large and covers several vegetation zones with different tree species compositions (Tomppo 2006). In addition, varying imaging conditions within the area of a satellite image can also alter the covariance structure between the field data and the image data. Therefore, the biases will be reduced if the set of potential nearest neighbours can be restricted to areas corresponding to the vegetation structure and imaging conditions of the pixel in question (Tomppo 2006). In this study, for reasons mentioned above, the plots were selected on the forest type that was relatively homogeneous, i.e. only on the *Fagus* dominant forest type. For reducing the effect of topography on illumination, the plots were randomly selected only on the northern slope. The plot sizes of 60 × 60 m were chosen due to the misregistration of ASTER imagery (i.e. two times the SWIR pixel size) and prior knowledge of the position accuracy of the global positioning system (GPS). In 101 plots, the kind of species, diameter at breast height (DBH) for all trees with DBH > 7.5 cm and height of three trees including the nearest and farthest trees to the centre and the biggest tree in diameter were measured. The GPS measurements in plot locations were averaged over the whole time spent for the forest plot measurements in order to enhance accuracy. The volume per hectare in each plot was calculated based on tariff tables for the Darabkola forest. The basal area and tree number per hectare were also computed for each plot based on the field data.

3. Methods

3.1 *Image processing*

The VNIR and SWIR images were geometrically orthorectified using a 10 m resolution digital elevation model (DEM) and some control points obtained by a handheld GPS (3–5 m precision). The obtained root mean square errors (RMSEs) of imagery were less than 10 m for both VNIR and SWIR images. The geometric accuracy of orthorectified images was confirmed using some independent ground control points. Some commonly used vegetation indices (table 1) for quantifying the vegetation attributes and enhancing the biophysical characteristics were generated using the VNIR and SWIR imagery. In addition, a standardized principal component analysis (PCA) was individually applied on both SWIR and VNIR imagery. The tasselled cap components (brightness, greenness and wetness), which are demonstrably related to biophysical properties of vegetation (Cohen *et al*. 1995, Mohammadi *et al*. 2010), were obtained by multiplying coefficients prepared by Yarbrough *et al*. (2005) on the VNIR and SWIR imagery. Some related texture indices including a dominance index (Hmax-H), centre versus neighbours (CVN) and binary comparison matrix (BCM) using both 3 × 3 and 5 × 5 kernels were individually applied on the near-infrared

Table 1. The vegetation indices.

Index	Formula	References
SVI	NIR/Red	Birth and McVey (1968)
TVI	$(NDVI + 0.5)^{1/2}$	Deering *et al*. (1975)
NDVI	NIR – Red/NIR + Red	Rose *et al*. (1973)
RVI	Red/NIR	Richardson and Wiegand (1977)
DVI	NIR – Red	Tucker (1979)
RDVI	$(NDVI \times DVI)^{1/2}$	Roujean and Breon (1995)
NRVI	(RVI – 1)/(RVI + 1)	Baret and Guyot (1991)
TTVI	$\sqrt{}$absolute (NDVI + 0.5)	Thiam (1997)
CTVI	NDVI + 0.5/absolute (NDVI + 0.5) × $\sqrt{}$absolute (NDVI + 0.5)	Perry and Lautenschlagei (1984)
NDVISWIR1–2	SWIR2 – SWIR1/SWIR2 + SWIR1	Mayer and Scribner (2002)
NDVISWIR1–3	SWIR3 – SWIR1/SWIR3 + SWIR1	Mayer and Scribner (2002)
NDVISWIR1–4	SWIR4 – SWIR1/SWIR4 + SWIR1	Mayer and Scribner (2002)
NDVISWIR1–5	SWIR5 – SWIR1/SWIR5 + SWIR1	Mayer and Scribner (2002)
NDVISWIR1–6	SWIR6 – SWIR1/SWIR6 + SWIR1	Mayer and Scribner (2002)
GNDVI	NIR – Green/NIR + Green	
SWNDVI	SWIR1 – Red/SWIR1 + Red	
SWGNDVI	SWIR1 – Green/SWIR1 + Green	
SWIR2NDVI	SWIR2 – Red/SWIR2 + Red	

Note: SVI, simple ratio vegetation index; TVI, transformed vegetation index; NDVI, normalized difference vegetation index; RVI, ratio vegetation index; DVI, differential vegetation index; RDVI, renormalized difference vegetation index; NRVI, normalized ratio vegetation index; TTVI, Thiam's transformed vegetation index; CTVI, corrected transformed vegetation index.

band, which has a spatial resolution of 15 m. In total, 40 independent variables and predictors were used in the analysis for predicting the forest attributes.

3.2 *Extraction of spectral features*

In order to match the pixel dimensions with plot size, the main and processed VNIR (15 m) and SWIR (30 m) images were resized to a 60 m resolution using size factor coefficients of 4 and 2, respectively, for VNIR and SWIR, and then their spectral values were averaged by resampling. A point vector file containing the position of the pixels' centres was generated by the raster to points function in ArcGIS (ESRI, Redlands, CA, USA). Spectral values of pixels in all the images used corresponding with inventoried plots as reference units and other pixels as target units were extracted by a point vector file.

3.3 *Randomly stratified sample splitting method*

In all statistical analyses that need to have prototype samples, i.e. training, test and validation data sets, the sample sets should be uniform and representative of all the data (Keller 2010). Therefore, the plots were first stratified based on their frequency

Table 2. Frequency of samples in each volume category, training and validation.

Volume category (m^3 ha^{-1})	100–200	200–300	300–400	400–500	500–600	600–700	Total
Training (80%)	9	11	27	15	16	2	80
Validation (20%)	2	3	7	4	4	1	21
Sum	11	14	34	19	20	3	101

Table 3. Frequency of samples in each basal area category, training and validation.

Basal area category (m^2 ha^{-1})	15–20	21–25	26–30	31–35	36–40	41–45	46–50	Total
Training (80%)	3	7	18	19	19	10	4	80
Validation (20%)	1	2	5	5	5	2	1	21
Sum	4	9	23	24	24	12	5	101

Table 4. Frequency of samples in each tree number category, training and validation.

Tree number category (n ha^{-1})	150–200	201–250	251–300	301–350	351–400	401–450	Total
Training (80%)	8	11	22	18	15	6	80
Validation (20%)	2	3	6	4	4	2	21
Sum	10	14	28	22	19	8	101

distributions in different internal classes and were then randomly divided into two training and validation sets, so that they had a sufficient distribution range in terms of their frequency in whole classes. This sample splitting method can be called *randomly stratified sample splitting*. For doing this, the values of plots based on volume, basal area or tree number were stratified in categories that are currently used for classification or mapping of discrete variables in forest management, and the frequency of plots was computed in each category. The training and validation samples were randomly selected as 80% for training and 20% for validation in each category (tables 2–4).

3.4 *Implementation of machine-learning methods*

3.4.1 k-nearest neighbour. In k-NN implementations, the number of k-NNs, the type of distance measure and the weighting for nearest neighbours are three important parameters. Determination of k would be very important in terms of calculation time and producing unbiased results. For finding the optimal k, the v-fold (dividing training samples into two v-fold parts, i.e. $v - 1$ for prototype and onefold for validation) and leave-one-out (LOO) cross-validation selection methods were used based on a k search range (1–20). The LOO is an accepted method for generating unbiased estimates of expected classification or estimation error for the k-NN estimator (Gong 1986). In this method, after importing the given range of k, the algorithm will calculate the RMSE values in validation sample sets for each k value and would then find the best k that has produced the lowest RMSE. Consideration of the 1–20 search range is based on McRobert (2009), who has reported that higher k values may lead to a reduced effect of noise in calculations but would require more samples. Most importantly, in the

case of a higher k, the pixel-level results will average towards the mean (Nilsson 1997), which leads to a higher bias and less precision. In most of the studies, the optimal k was reported to be between 5 and 10 (Reese *et al*. 2002, Huiyan *et al*. 2006, Kutzer 2008). In some studies, such as Finley *et al*. (2006) with k between 1 and 35 and McRobert (2009) with k between 1 and 50, higher k values tested better in reducing the bias. A smaller k often leads to a higher variance and less stable results (Kozma 2008). The optimal k is dependent on the data and goals of the estimation (Hagner *et al*. 2002). Kajisa *et al*. (2008), in a study on the estimation of stand volume of a forest in Kyushu, Japan, reported that the RMSE and relative RMSE rapidly decreased as the number of nearest neighbours (k) increased up to 5 and then the rate of decline decreased, with the conclusion that the numbers of nearest neighbours >5 were the most suitable numbers for stand volume estimation. For making better decisions on finding the optimal k, a range of 1–5 was also applied in k-NN implementations. In addition, for an efficient comparison of distance measures in k-NN implementation, the four prepared distance measures in Statistica software (StatSoft. Inc., Tulsa, OK, USA), including Euclidean, squared Euclidean, city block (Manhattan) and Chebychev, as both weighted and non-weighted, were individually used and their results were compared with each other. In all implementations, the independent variables were standardized through a simple transformation of between 0 and 1. This transformation has a wide variety of applications because it makes the distributions of values easy to compare across variables and/or subsets (Statistica 2010).

3.4.2 Support vector machine regression. The prerequisite for SVR to achieve better results is an appropriate determination of the parameters that play key roles in achieving high accuracy and better performance (Wang *et al*. 2009). The specified grid search using v-fold cross-validation (Durbha *et al*. 2007) is the most commonly used method to find suitable parameters, i.e. epsilon (ε) and capacity (C) with fixed gamma, that would produce high-accuracy results. A brief description of the proposed methods is summarized in Durbha *et al*. (2007).

In this study, three different kernels, including the RBF, polynomial and sigmoid, were examined in a fixed 0.025 of gamma that is calculated based on 1/number of independent variables (Hsu *et al*. 2010). For selecting the best parameters, 10-fold cross-validation with 1000 iterations for minimizing the error function (Schölkopf *et al*. 1998) was used through a specified grid search method (Hsu *et al*. 2010) to determine the best capacity and epsilon rates. The specified grid search included a range of capacity from 1 to 40, which is equal to the range of input variables (Mattera and Haykin 1999), and epsilon values from 0.1 to 0.5.

3.4.3 Random forest. For a high quality of RF regression performance, the decision tree model and stopping parameters should be regularized. In this study, we examined different conditions of RF options by importing different model parameters. For determination of the optimal tree number, 200 initial trees were used to produce a graph, showing the average squared error rates against each number of trees for training and test samples. This is a powerful analytical tool for exploration of data and verification of the optimal number of trees in RFs. By interpretation of the graph, the optimal number of trees is found based on a tree number that produces a stable error. Then, the RF implementation was again repeated using this optimal number of trees and other fixed parameters. For applying the bootstrap training and learning,

we examined subsamples – proportions of 50%, 60%, 70%, 80% and 90% of training samples in a bagging bootstrap sampling.

One of the main parameters in RF implementations is the k predictor (independent variables) in each node for predicting dependent values (response). The simplest way for determining k is the calculation of the square root of total independent variables ($k \leq \sqrt{m}$, where m is the number of input variables). For further investigation of the effect of k on the results, we examined a range of 3–9 (square root of total variables ± 3) for the fixed last parameters.

The best parameters were selected based on the assessment of their results using validation indices (RMSE and bias) on the unused test samples to evaluate the validity of the model. In addition, default rates of stopping and splitting parameters or conditions were used to stop the processes of growing the trees when stopping conditions were reached. The stopping parameters for all estimations included a minimum of one child node and a maximum of 100 nodes to stop growing the trees in 10 iterations to calculate the mean error and a 5% decrease in the training error.

3.5 *Validation and quality performance assessment*

The validity of performances was examined using regression diagnostics metrics, i.e. RMSE, relative RMSE, bias and relative bias, and using the independent and unused 21 samples. In addition, some common graphical diagnostic tools (McRobert 2009) were used to illustrate the quality of performances.

$$\mathrm{RMSE} = \frac{\sqrt{\sum_{i=1}^{m} (e_i - o_i)}}{m}, \tag{10}$$

$$\mathrm{bias} = \frac{\sqrt{\sum_{i=1}^{m} (e_i - o_i)}}{m}, \tag{11}$$

$$\mathrm{RMSE}\% = \frac{\sqrt{\sum_{i=1}^{m} (e_i - o_i)/m}}{\sum_{i=1}^{m} (o_i)/m} \times 100 \tag{12}$$

and

$$\mathrm{bias}\% = \frac{\sqrt{\sum_{i=1}^{m} (e_i - o_i)/m}}{\sum_{i=1}^{m} (o_i)/m} \times 100, \tag{13}$$

where e is the estimation values from the implementation of algorithms in m validation samples, o is the observation values and m is the number of validation samples.

4. Results and discussion

A descriptive analysis of training and validation samples selected by randomly stratified sample splitting showed that there is no significant difference between validation and training sample sets (table 5). It means that the test samples were representative of training samples, and the reliability and performance of validation results were good.

Table 5. Descriptive analysis results for validation and training subset samples selected by a randomly stratified sample splitting method.

	Volume/ha			Basal area/ha			Stems/ha		
Descriptive	Validation	Training	Total	Validation	Training	Total	Validation	Training	Total
Mean	389.34	376.14	378.88	33.71	33.75	33.74	310.28	301.14	303.04
Standard error	30.27	14.61	13.11	1.67	0.823	0.73	16.11	7.92	7.1
Median	380.70	359.35	364.8	34.00	33.6	34	300.3	294.4	297.2
Standard deviation	138.73	130.67	131.77	7.64	7.29	7.33	73.85	70.91	71.26
Sample variance	19246.20	17075.27	17367.71	58.44	53.18	53.70	5454.15	5029.43	5077.99
Kurtosis	−0.88	−0.80	−0.84	−0.24	−0.30	−0.34	−0.64	−0.71	−0.71
Skewness	−0.02	0.06	0.05	0.14	−0.03	0.006	0.068	−0.041	−0.015
Range	476.10	512.2	512.2	31.00	32.5	32.5	277.8	285.5	294.7
Minimum	140.00	131.3	131.3	19.00	17.5	17.5	172.2	155.3	155.3
Maximum	616.10	643.5	643.5	50.00	50	50	450	440.8	450
Coefficient of variation	0.356	0.347	0.348	0.226	0.216	0.216	0.238	0.235	0.235

4.1 k-*nearest neighbour*

An accurate performance of *k*-NN in the estimation of forest attributes using field inventory and remote-sensing data is generally dependent on the application of optimal *k* and distance measure. The cross-validation results showed that using more *k* in the different implementations, and with weighted or non-weighted distance measures, could not produce significant improvements in terms of relative RMSE (approximately between 1 and 2% for volume and basal area and about 1% for tree number). For more details, see tables 6–8.

Using a *k* of value more than 5 in all implementations with different distance measures caused positive influences on prediction accuracies for all three parameters: volume/ha, basal area/ha and stems/ha estimations. These results are in agreement with the results of McRobert *et al*. (2007) and Breidenbach *et al*. (2010), who suggested $k \geq 5$ for *k*-NN predictions. For forest attribute imputations in small areas using medium-resolution satellite data such as that from ASTER, whose correlation between spectral responses and forest structural attributes may be weak, selecting *k* from 5 to 20 seems to have a positive effect on the good predictions of forest attributes as has been seen in our results. Our results are in agreement with the results of Kajisa *et al*. (2008) for stand volume estimation using ETM+ data in Japan, which revealed that the RMSE and relative RMSE of the volume estimates rapidly decreased as the number of nearest neighbours (*k*) increased to 5 and then the rate of decline decreased and the best results were found for $k = 10$.

Some studies (Nilsson 1997, Kozma 2008) have shown that applying more *k* may lead to biased results due to averaging of the values of more pixels, but in our study on

Table 6. Results of different *k*-NN implementations using a cross-validation method for volume/ha estimation.

Range of *k*	Weight	Measure	RMSE (m^3 ha^{-1})	RMSE (%)	Bias (m^3 ha^{-1})	Bias (%)	Optimal *k*
1–5	*	Euclidean	116.47	29.91	−19.25	−5.28	5
1–5	–	Euclidean	115.79	29.79	−19.25	−5.20	5
1–20	*	Euclidean	111.61	28.67	−12.77	−3.39	20
1–20	–	Euclidean	113.85	29.24	−13.74	−3.66	7
1–5	*	Squared Euclidean	117.10	30.07	−19.44	−5.26	5
1–5	–	Squared Euclidean	115.99	29.79	−19.25	−5.20	5
1–20	*	Squared Euclidean	113.26	29.09	−12.04	−3.19	20
1–20	–	Squared Euclidean	113.85	29.24	−13.74	−3.66	7
1–5	*	Manhattan	129.15	33.17	−16.44	−4.41	5
1–5	–	Manhattan	121.17	31.12	−17.68	−4.76	5
1–20	*	Manhattan	118.67	30.48	−9.73	−2.56	19
1–20	–	Manhattan	114.36	29.37	−12.60	−3.34	7
1–5	*	Chebychev	121.36	31.17	−6.27	−1.64	5
1–5	–	Chebychev	120.91	31.05	−10.67	−2.82	4
1–20	*	Chebychev	112.19	28.81	−11.75	−3.11	20
1–20	–	Chebychev	111.12	28.54	−13.24	−3.52	17

Note: *Means weighting used.

Table 7. Results of different *k*-NN implementations using a cross-validation method for basal area/ha estimation.

Range of *k*	Weight	Measure	RMSE (m^2 ha^{-1})	RMSE (%)	Bias (m^2 ha^{-1})	Bias (%)	Optimal *k*
1–5	*	Euclidean	7.51	22.29	−0.32	−0.96	5
1–5	–	Euclidean	7.50	22.25	−0.41	−1.23	5
1–20	*	Euclidean	6.94	20.60	0.11	0.33	20
1–20	–	Euclidean	6.88	20.40	0.12	0.37	20
1–5	*	Squared Euclidean	7.58	22.49	−0.22	−0.66	5
1–5	–	Squared Euclidean	7.50	22.25	−0.41	−1.23	5
1–20	*	Squared Euclidean	7.08	21.02	0.14	0.44	20
1–20	–	Squared Euclidean	6.88	20.40	0.12	0.37	20
1–5	*	Manhattan	7.96	23.61	0.49	1.44	5
1–5	–	Manhattan	7.87	23.35	−0.41	1.20	5
1–20	*	Manhattan	7.37	21.84	0.32	0.94	20
1–20	–	Manhattan	6.81	20.20	−0.08	−0.24	14
1–5	*	Chebychev	7.16	21.25	−1.12	−3.44	5
1–5	–	Chebychev	7.15	21.23	−1.14	−3.51	5
1–20	*	Chebychev	7.42	22.01	−0.44	−1.33	20
1–20	–	Chebychev	7.58	22.50	−0.44	−1.34	18

Note: *Means weighting used.

Table 8. Results of different *k*-NN implementations using cross-validation method for stems/ha estimation.

Range of *k*	Weight	Measure	RMSE (n ha^{-1})	RMSE (%)	Bias (n ha^{-1})	Bias (%)	Optimal *k*
1–5	*	Euclidean	68.11	21.95	7.79	2.44	5
1–5	–	Euclidean	67.79	21.85	7.52	2.36	5
1–20	*	Euclidean	66.82	21.53	−1.48	−0.48	19
1–20	–	Euclidean	67.96	21.90	7.71	2.42	6
1–5	*	Squared Euclidean	68.43	22.06	7.91	2.48	5
1–5	–	Squared Euclidean	67.79	21.85	7.52	2.36	5
1–20	*	Squared Euclidean	67.13	21.63	−1.73	−0.56	19
1–20	–	Squared Euclidean	67.96	21.90	7.71	2.42	6
1–5	*	Manhattan	67.87	21.87	2.91	0.93	9
1–5	–	Manhattan	72.75	23.44	0.19	0.06	5
1–20	*	Manhattan	67.14	21.64	0.68	0.22	18
1–20	–	Manhattan	72.75	23.44	0.19	0.06	5
1–5	*	Chebychev	64.13	20.66	2.16	0.69	5
1–5	–	Chebychev	64.04	20.64	2.06	0.66	5
1–20	*	Chebychev	69.25	22.31	−0.11	−0.04	20
1–20	–	Chebychev	68.91	22.21	−3.31	−1.07	9

Note: *Means weighting used.

volume/ha estimations, applying $k = 20$ led to a lower bias compared to applying $k = 5$ in all distance measures, except for Chebychev (table 6). As Hagner *et al*. (2002) and Kozma (2008) contended, the optimal number of k was absolutely dependent on the data and goals of the estimation; our opposite results may be due to the local structure of the study area and remote-sensing data, i.e. presence of noisy or irrelevant features. In addition, using a larger k will reduce the effect of noise (Mäkelä and Pekkarinen 2004). Gong (1986) reported that using LOO v-fold cross-validation with numerous iterations in different partitions led to stable results. These less-biased results come through averaging v-fold error predictions and yield a single measure (model error). Lower optimal k was found when using non-weighted distance measures when ranges of 1–20 for k were applied in v-fold cross-validation. It means that applying weighted distances could improve estimations using more k-NNs, so that using more k may reduce relative RMSE and bias.

In addition, in volume estimation, considering the weight for distances did not significantly improve the results, but applying the Chebychev distance measure produced results with a slightly lower relative RMSE and bias compared with the other three distance measures. However, a lower RMSE and bias were also obtained using the weighted Euclidean and Chebychev distance measures when a k range of 1–50 was used. These results confirmed the outcomes of Kajisa *et al*. (2008), who reported that k-NN implementations with the Euclidean distance had consistently smaller RMSE and relative RMSE than those with the Mahalanobis distance.

In basal area estimations (table 7), applying more k caused slight improvements in both relative RMSE and bias. As table 6 shows, the implementations using weighted or non-weighted Euclidean and squared Euclidean distance measures have produced approximately equal results in terms of relative RMSE and bias. The use of more k (1–20) could produce little improvement (about 2% in relative RMSE and a small reduction in relative bias). Basal area estimations have low bias and relative bias in comparison with volume and tree number estimations because of the lowering of the coefficient of variations (CVs) in basal area values in both training and test samples (see also table 5). Generally, regarding relative RMSE and bias, in more k implementations, the Manhattan distance measure produced a better performance than other distance measures.

In stems/ha estimations, the results were more or less the same as those obtained for volume and basal area (table 8). The relative RMSEs of estimations were equal when the Euclidean and squared Euclidean distance measures were used but using the weighted measure could improve the results about 2%. An increase of k could decrease the relative and absolute RMSEs and bias when using both the distance measures. The use of Manhattan and Chebychev distance measures produced acceptable relative bias in both weighted and non-weighted forms as well as in different k values. The k-NN implementation using both weighted and non-weighted Chebychev distance measure with a small k ($k = 5$) generated lower relative RMSE compared with the other measures.

In the k-NN implementation using a weighted Chebychev distance measure, the results showed that with an increase in k, the relative RMSE increased by about 2% but with a low bias. This revealed that with an increase of k, more reference units would participate in averaging and consequently bias would be decreased. Table 9 shows that the best results for volume, basal area and stems per hectare were obtained from k-NN implementations with different k range and distance measures. The table also shows that there is no significant superiority in applying weighted or non-weighted distance

Table 9. The best k-NN performance results for estimations of forest attributes.

Dependent variable	Weight	Measure	RMSE	RMSE (%)	Bias	Bias (%)	Optimal k
Volume (m^3 ha^{-1})	*	Euclidean	111.61	28.67	−12.77	−3.39	20
Basal area (m^2 ha^{-1})	–	Manhattan	6.81	20.20	−0.08	−0.24	14
Tree number (n ha^{-1})	–	Chebychev	64.04	20.64	2.06	0.66	5

Note: *Means weighting used.

measures with different k compared to using other distance measures and k, but lower errors were obtained using these parameters.

4.2 *Support vector machine regression*

The SVR algorithm is another machine-learning method whose result is completely dependent on the regularization of model parameters. The selection of the best parameters that maximize differences between target and reference units in a hyperplane feature space plays an important role in generating accurate models. Therefore, the potential of SVR models is strongly dependent on the optimal selection of the SVR parameters, that is C and ε. Applying the cross-validation method (Schölkopf *et al*. 1998) and grid searches for regularizing the SVR models presented the best parameters that produced low relative RMSE and bias. The results of SVR implementations with optimal regularized model parameters for volume/ha, basal area/ha and stems/ha estimations are given in tables 10–12, respectively.

Table 10. Results of SVR implementations with the best parameters determined by cross-validation and the grid search method for volume/ha estimation.

Kernel	Capacity (C)	Epsilon (ε)	Gamma (γ)	RMSE (m^3 ha^{-1})	RMSE (%)	Bias (m^3 ha^{-1})	Bias (%)
RBF	40	0.5	0.025	124.62	32	16.86	4.15
Polynomial (degree 3)	34	0.32	0.025	100.72	25.86	2.80	0.71
Sigmoid	40	0.4	0.025	101.01	25.94	2.89	0.74
Linear	39	0.16	–	129.68	33.30	24.26	5.86

Table 11. Results of SVR implementations with the best parameters determined by cross-validation and the grid search method for basal area/ha estimation.

Kernel	Capacity (C)	Epsilon (ε)	Gamma (γ)	RMSE (m^2 ha^{-1})	RMSE (%)	Bias (m^2 ha^{-1})	Bias (%)
RBF	40	0.41	0.025	7.01	20.80	0.51	1.50
Polynomial (degree 3)	40	0.45	0.025	6.52	19.35	0.38	1.11
Sigmoid	40	0.38	0.025	7.81	23.19	0.56	1.64
Linear	35	0.17	–	8.45	25.07	−0.01	−0.04

Table 12. Results of SVR implementations with the best parameters determined by cross-validation and the grid search method for stems/ha estimation.

Kernel	Capacity (C)	Epsilon (ε)	Gamma (γ)	RMSE (n ha^{-1})	RMSE (%)	Bias (n ha^{-1})	Bias (%)
RBF	40	0.1	0.025	79.62	25.66	1.71	0.55
Polynomial (degree 3)	18	0.26	0.025	68.55	22.09	−4.18	−1.36
Sigmoid	39	0.1	0.025	85.35	27.5	1.26	0.40
Linear	30	0.35	–	77.28	24.90	−5.80	−1.90

The results of volume estimation showed that lower relative RMSE and bias were found using a three-degree polynomial kernel type, although similar results were obtained using a sigmoid kernel type in terms of relative RMSE and bias. Considerably different RMSE, relative RMSE, bias and relative bias come from RBF and linear kernels compared with polynomial and sigmoid kernels (table 10). The best hyperplane was fitted with high values of C and ε in all the four kernel types using cross-validation and grid searches.

In basal area/ha estimation using SVR, the best results based on relative RMSE and bias were obtained using a polynomial kernel type (table 11). A low CV within the samples (in both training and test samples) compared with other dependent variables may be one of the main reasons for these results.

In tree number estimation, the SVR implementation using a polynomial kernel produced better results in terms of relative RMSE, but with a slightly higher bias and relative bias, compared to using the RBF and sigmoid kernels (table 12). Although the grid search and gamma were fixed for all four SVR kernel-based regressions, different capacity and epsilon parameters were fitted for hyperplanes.

The results also showed that applying a three-degree polynomial kernel could result in a better design for the hyperplane, leading to more accurate estimates with lower relative RMSE compared with other kernels. However, in some studies (Shafri and Ramle 2009, Hsu *et al*. 2010), the RBF kernel produced better results. There are cases where the RBF is not suitable, particularly when the number of features is very large (Hsu *et al*. 2010).

4.3 *Random forest*

As noted before, the performance of RF implementation depends on determining the number of trees and number of predictors in each node for producing a good response. In figure 2, for instance in the implementation of volume estimation, the graph shows average squared error rates against the number of trees. In this graph, the improvement in accuracy is slow after about 50 trees and 200 trees is a good estimate of an optimum number to use and shows the lowest error and a stable response.

Table 13 shows the results of RF performances using the default number of predictors (6) in each node and the optimal k number of trees in RFs for estimating volume/ha, basal area/ha and stems/ha. In volume estimation using small trees, the results show a low RMSE compared with volume estimation using large trees, but the rate of bias is very high, and if large trees were used, the bias would decrease to about 0 (unbiased results).

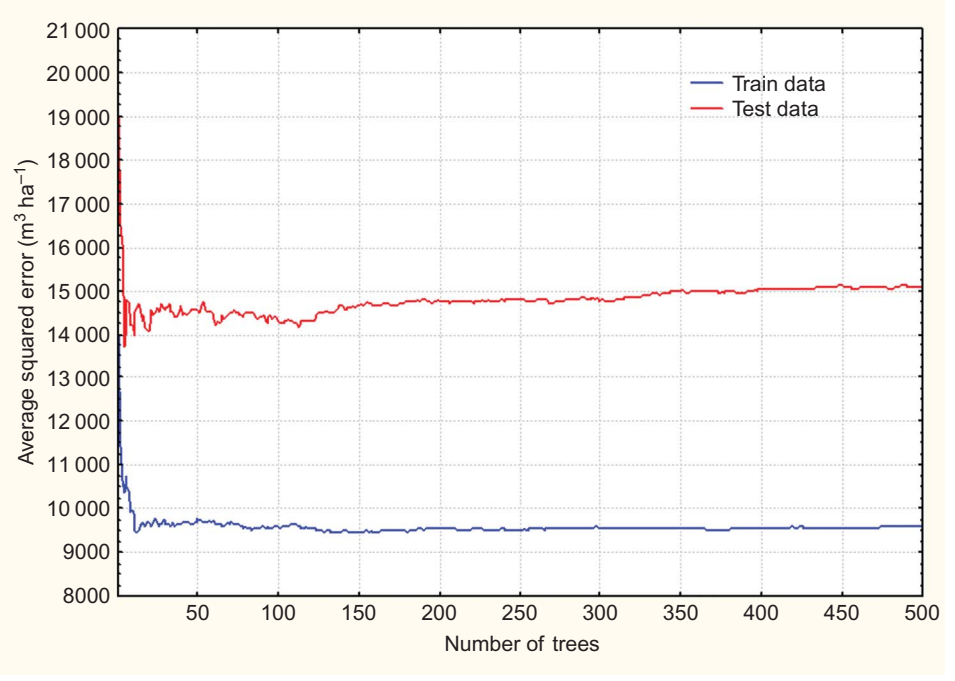

Figure 2. Random forest implementation graph. Graph of average square error against number of trees in training and test data in the implementation of volume estimation.

Table 13. Validity results of RF implementations for volume/ha, basal area/ha and tree number/ha attributes.

Variable	*k* predictor	Optimal *k* tree	RMSE	RMSE (%)	Bias	Bias (%)
Volume (m^3 ha^{-1})	6	180	104.61	26.86	0.09	0.02
Basal area (m^2 ha^{-1})	6	25	6.39	18.39	0.28	0.83
Tree number (n ha^{-1})	6	200	64.05	20.64	−3.14	−1.02

As shown in table 14, applying a different rate of *k* predictors, which can be grouped to form a node, did not have a marked effect on the performance of RF in any of the estimations. As can be observed in table 14, except for volume/ha estimation using a rate of six predictors (obtained from computing the square root of number of variables) in each node resulted in a slight improvement in relative RMSE and bias in two other forest variables (the bold values). Improvements did not occur when fewer or more *k* predictors were used. These results confirmed that the use of the square root of the number of variables would be a suitable way to define the number of predictors in each node of trees.

4.4 *Comparison of three machine-learning methods*

In this study, a comparison of the performance of the three most commonly used machine-learning algorithms, namely *k*-NN, SVR and RF, was made for imputation of forest volume/ha, basal area/ha and stems/ha using field inventory and ASTER

Table 14. Effect of different *k* predictors on the results of random forest performances.

	Optimal number of tree	*k* predictor	RMSE	RMSE (%)	Bias	Bias (%)
Volume/ha	180	3	106.12	27.25	−1.86	−0.48
		4	104.56	26.85	−0.60	−0.15
		5	103.31	26.53	−0.28	−0.07
		Default (6)	104.61	26.86	0.09	0.02
		7	104.90	26.94	1.76	0.45
		8	102.20	26.25	1.27	0.32
		9	105.44	27.08	−1.00	−0.25
Basal area/ha	25	3	6.64	19.17	0.30	0.88
		4	6.44	19.10	0.19	0.56
		5	6.57	19.49	0.45	1.32
		Default (6)	**6.20**	**18.39**	**0.28**	**0.83**
		7	6.43	19.09	0.74	2.15
		8	6.61	19.61	0.61	1.79
		9	6.72	19.93	0.52	1.52
		3	64.93	20.92	−5.05	−1.65
		4	65.53	21.12	−2.86	−0.93
		5	65.10	20.98	−3.68	−1.20
Stems/ha	200	**Default (6)**	**64.05**	**20.64**	**−3.11**	**−1.02**
		7	63.94	20.61	−3.73	−1.22
		8	64.32	20.73	−3.68	−1.20
		9	64.27	20.71	−3.49	−1.14

data. Table 15 shows the results from the best model implementation for each of the three algorithms and forest variables. Although the results from the three algorithms were generally similar, they did have some differences in the estimation of three attributes. In the estimation of two attributes (i.e. basal area/ha and stems/ha), the RF algorithm produced slightly better results in terms of RMSE and relative RMSE. However, in volume/ha estimation, the SVR algorithm better imputed the target pixels than the other algorithms.

The results of volume estimation in this study are in line with the results of Hudak *et al*. (2008), who reported an unbiased estimation for RF. In our study, the RF algorithm produced less biased estimation compared with SVR and *k*-NN algorithms. For volume/ha imputations, more biased estimation came from the implementation of a *k*-NN algorithm as well as in terms of RMSE compared with SVR and RF. However, Eskelson *et al*. (2009) reported that volume/ha imputation using climate, topography and satellite data by RF was biased compared with the most similar neighbour (MSN) and gradient nearest neighbour (GNN) as two variants of *k*-NN. In addition, Latifi *et al*. (2010) in a plot-level estimation of volume/ha found that RF results were slightly biased compared with *k*-NN, but the best unbiased volume estimations were obtained using MSN.

For basal area/ha imputation, RF showed a slightly better performance in terms of RMSE compared with *k*-NN and SVR. Our results are in contrast to the findings of McInerney and Nieuwenhuis (2009), who reported that RMSE was lower for *k*-NN than for RF in plot-level estimation of stand volume and basal area, showing that the RF algorithm could produce better results compared with *k*-NN in terms of RMSE with slightly higher biased results.

Table 15. Comparison of the performance of the best models obtained from the three machine-learning methods in the study area.

	Volume/ha				Basal area/ha				Stems/ha			
Method	RMSE	RMSE (%)	Bias	Bias (%)	RMSE	RMSE (%)	Bias	Bias (%)	RMSE	RMSE (%)	Bias	Bias (%)
k-NN	111.12	28.54	−13.24	−3.52	6.81	20.20	−0.08	−0.24	66.82	21.53	−1.48	−0.48
SVR	100.72	25.86	2.80	0.71	6.52	19.35	0.38	1.11	68.55	22.09	−4.18	−1.36
RF	104.61	26.86	0.09	0.02	6.20	18.39	0.28	0.83	64.05	20.64	−3.14	−1.02

In stems/ha imputation, the results showed that high performances were, respectively, achieved using RF, *k*-NN and SVR in terms of RMSE. However, in terms of bias results, the best results were obtained using *k*-NN, RF and SVR, respectively.

5. Conclusion

Three machine-learning methods (*k*-NN, SVR and RF) were compared for the estimation of forest attributes (volume, basal area and tree number) at the plot level using ASTER remote-sensing and field inventory data. Estimation of forest attributes and wall-to-wall imputation using auxiliary data such as remote-sensing data for whole forest areas can be done at three levels: species, plot and stand levels. Species-level imputation is generally performed in the places where the crown of trees can be distinguished using high-resolution imagery. Therefore, most studies of the species level have been accomplished in monocultures or even-aged stand structures using high-resolution spatial data. However, because of the general needs of forest managers as well as the conditions of forest structure, most research studies by machine learning for forest attribute imputation have been done at the plot level using medium-resolution remote-sensing data. This is because medium-resolution imagery (Landsat, SPOT and ASTER) is more accessible, has good coverage over time and is of lower cost in addition to being more applicable to plot-level estimation in forest management plans compared with stand level, which is often used for general planning. Unfortunately, for our study area, high-resolution imagery was not available, but there was medium-resolution imagery such as Landsat ETM+, TM and ASTER data. The choice of ASTER images over ETM+ and TM was due to their greater capabilities in terms of spatial (VNIR with 15 m resolution) and spectral resolution superiorities (six SWIR bands in middle infrared compared with three bands in ETM+ and TM).

Under similar study conditions, our results using nonparametric algorithms are not in line with the results of Mohammadi *et al*. (2010), who used parametric linear regression. They predicted volume/ha with an RMSE of 97.4 m^3 ha^{-1} and a bias of 28.08 m^3 ha^{-1} but we could estimate volume/ha with an RMSE of 100.72 m^3 ha^{-1} and with roughly unbiased estimates (2.80 m^3 ha^{-1}) using the SVR algorithm. Our unbiased volume estimation has a high superiority in contrast to a slightly lower RMSE difference. This superiority is due to the use of non-parametric machine-learning algorithms together with LOO cross-validation and due to the use of ASTER data and its capabilities against ETM+ data. In addition, Mohammadi *et al*. (2010) predicted stems/ha with RMSE = 170.13 n ha^{-1} and with bias = 61.475 n ha^{-1}, whereas we imputed stems/ha with RMSE = 64.05 n ha^{-1} and bias about 3 n ha^{-1}. It has to be noticed that a large difference between RMSE and bias in our study and those obtained by Mohammadi *et al*. (2010) may be due to the different conditions of stand structure in the two study areas: the minimum, maximum and mean of tree numbers per hectare measured about 183, 1136 and 514, respectively, in the study area of Mohammadi *et al*. (2010), whereas in our study area, the minimum, maximum and mean of tree numbers per hectare measured about 155, 450 and 303, respectively. These values show that our study area in comparison with the Mohammadi *et al*. (2010) study area is more homogeneous in terms of tree numbers.

In this study, a randomly stratified sample splitting method was also used for choosing the training and test samples. The total samples were classified based on some currently used classes, and their frequency in each class was computed. The training and test samples were randomly selected as 80% and 20%, respectively. This method

helps us to have sufficient number of reliable frequency distributed samples for training, and takes into consideration all conditions of dependent variables in the study area. In addition, as the validation sample sets have been selected from the whole range of data sets, the results of validation assessment are reliable.

Generally, the studies showed that RF has the capability of providing reasonable imputations for forest attributes using remote-sensing data (LeMay and Temesgen 2005, Hudak *et al*. 2008, Eskelson *et al*. 2009, Breidenbach *et al*. 2010) and for mapping forest composition and structure (Ohmann and Gregory 2002). The RF method produced low-bias results in comparison with other machine-learning methods (Eskelson *et al*. 2009). Vauhkonen *et al*. (2010) found that RF proved to be a flexible method with an ability to handle numerous predictors with no need for their reduction.

In conclusion, our study showed that tree-based regression methods, i.e. RF, could be more useful for an accurate estimation of forest attributes. Consequently, the use of other tree regression variants such as boosting tree regression may improve the results and can be a subject of future work. In addition, in some studies (Tuominen 2007, Latifi *et al*. 2010), it has been reported that the accuracy of the machine-learning algorithms can be severely degraded by the presence of noisy or irrelevant features. The use of robust feature selection strategies that select the relevant independent variables may improve the estimations and can be another important subject in future forest attribute imputations using machine-learning algorithms.

References

Baret, F. and Guyot, G., 1991, Potential and limits of vegetation indices for LAI and AFAR assessment. *Remote Sensing of Environment*, **35**, pp. 161–173.

Birth, G.S. and McVey, G., 1968, Measuring the color of growing turf with a reflectance spectrophotometer. *Agronomy Journal*, **60**, pp. 640–643.

Breidenbach, J., Nothdurft, A. and Kandler, G., 2010, Comparison of nearest neighbor approaches for small area estimation of tree species-specific forest inventory attributes in central Europe using airborne laser scanner data. *European Journal of Forest Research*, **129**, pp. 833–845.

Breiman, L., 2001, Random forests. *Machine Learning*, **45**, pp. 5–32.

Cohen, W.B., Spies, T.A. and Fiorella, M., 1995, Estimating the age and structure of forests in a multi-ownership landscape of western Oregon, U.S.A. *International Journal of Remote Sensing*, **16**, pp. 721–746.

Cortez, P. and Morais, A., 2007, A data mining approach to predict forest fires using meteorological data. In *Proceedings of the EPIA 2007 – Portuguese Conference on Artificial Intelligence*, December 2007, J. Neves, M.F. Santos and J.M. Machado (Eds.), Guimarães, Portugal (Heidelberg: Springer), pp. 512–523.

Dalponte, M., Bruzzone, L. and Gianelle, D., 2008, Estimation of tree biomass volume in alpine forest areas using multi return lidar data and support vector regression. *Proceedings of the SPIE*, **7109**, pp. 71090C-1–71090C-12.

Deering, D.W., Rose, J.W., Mass, R.H. and Schell, J.A., 1975, Measuring forage production of grazing units from Landsat MSS data. In *Proceedings of the 10th International Symposium on Remote Sensing of Environment*, 6–10 October 1975, Ann Arbor, MI, pp. 1169–1178.

Durbha, S.S., King, R.L. and Younan, N.H., 2007, Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer. *Remote Sensing of Environment*, **107**, pp. 348–361.

Eskelson, B.N.I., Hailemariam, T. and Barrett, T.M., 2009, Estimating current forest attributes from paneled inventory data using plot-level imputation: a study from the pacific northwest. *Forest Science*, **55**, pp. 64–71.

Falkowski, M.J., Gessler, P., Morgan, P., Hudak, A.T. and Smith, A.M.S., 2004, Evaluating the ASTER sensor for mapping and characterizing forest fire fuels in Northern Idaho. In *Proceedings of Tenth Biennial USDA Forest Service Remote Sensing Applications Conference*, April 2004, Salt Lake City, UT, 11pp.

FAO, 2002, Forests and the forestry sector. Available online at: www.fao.org/forestry/site/23747/en/iran (accessed 21 December 2009).

Finley, A.O., McRobert, R.E. and Ek, A.R., 2006, Applying an efficient *k*-nearest neighbor search to forest attribute imputation. *Forest Science*, **52**, pp. 130–135.

Franco-Lopez, H., Ek, A.R. and 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**, pp. 251–274.

Fuchs, H., Magdon, P., Kleinn, C. and Flessa, H., 2009, Estimating aboveground carbon in a catchment of the Siberian forest tundra: combining satellite imagery and field inventory. *Remote Sensing of Environment*, **113**, pp. 518–531.

Gebreslasie, M.T., Ahmed, F.B. and Aardt, J., 2008, Estimating plot-level forest structural attributes using high spectral resolution ASTER satellite data in even-aged Eucalyptus plantations in southern KwaZulu-Natal, South Africa. *Southern Forests*, **70**, pp. 1–10.

Gjertsen, A.K., 2007, Accuracy of forest mapping based on Landsat TM data and a *k*NN-based method. *Remote Sensing of Environment*, **110**, pp. 420–430.

Gong, G., 1986, Cross-validation, the jackknife, and the bootstrap: excess error estimation in forward logistic regression. *Journal of American Statistics Association*, **81**, pp. 108–113.

Hagner, O., Nilsson, M., Joyce, S. and Olsson, H., 2002, *K*-nearest neighbor methods and artificial neural networks for estimation of forest parameters from satellite imagery.

Hall, R.J., Skakun, R.S., Arsenault, E.J. and Case, B.S., 2006, Modeling forest stand structure attributes using Landsat ETM+ data: application to mapping of aboveground biomass and stand volume. *Forest Ecology and Management*, **225**, pp. 378–390.

Heiskanen, J., 2005, Remote sensing of mountain birch forest biomass and leaf area index using ASTER data. In *Proceedings of the 31st International Symposium on Remote Sensing of Environment*, 20–24 June 2005, St. Petersburg, Russia.

Holmström, H., 2002, Estimation of single tree characteristics using *k*NN method and plot wise aerial photograph interpretations. *Forest Ecology and Management*, **167**, pp. 303–314.

Hsu, C.W., Chang, C.C. and Lin, C.J., 2010, *A Practical Guide to Support Vector Classification* (Taipei: Department of Computer Science, National Taiwan University). Available online at: http://www.csie.ntu.edu.tw/~cjlin

Hudak, A.T., Crookston, N.L., Evans, J.S., Hall, D.E. and Falkowski, M.J., 2008, Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. *Remote Sensing of Environment*, **112**, pp. 289–290.

Huiyan, G., Limin, D., Gang, W., Dong, X., Shunzhong, W. and Hui, W., 2006, Estimation of forest volumes by integrating Landsat TM imagery and forest inventory data. *Science in China, Series E: Technological Sciences*, **49**, pp. 54–62.

Hyvönen, P., 2007, The updating of forest resource data for management planning for privately owned forests in Finland. Academic dissertation, Faculty of Forestry, University of Joensuu, Joensuu, Finland, 40 p.

Kajisa, T., Murakami, T., Mizoue, N., Kitahara, F. and Yoshida, S., 2008, Estimation of stand volumes using the *k*-nearest neighbors method in Kyushu, Japan. *Journal of Forest Research*, **13**, pp. 249–254.

Katila, M. and Tomppo, E., 2001, Selecting estimation parameters for the Finnish multisource national forest inventory. *Remote Sensing of Environment*, **76**, pp. 16–32.

Keller, F., 2010, Evaluation, connectionist and statistical language processing. Online lecture, Computerlinguistik, Universitäat des Saarlandes. Available online at: http://www.coli.uni-saarland.de/~crocker/Teaching/Connectionist/lecture11_4up.pdf (accessed April 2010).

Khorrami, K.R., 2004, Investigation of the potential of Landsat ETM+ Data in volume estimating of beech forest stands (case study: Sangdeh area in north of Iran). MSc thesis, Faculty of Natural Resources, University of Tehran, Tehran, Iran, 80 p. [in Persian].

Kozma, L., 2008, *k* Nearest neighbor algorithm (*k*NN). Helsinki University of Technology, Special course in computer and information science. Available online at: www.lkozma.net/knn2.pdf

Kutzer, C., 2008, Potential of the *k*NN method for estimation and monitoring off-reserve forest resources in Ghana. PhD thesis, Faculty of Forest and Environmental Sciences, Albert-Ludwigs-Universität, Freiburg im Breisgau, Germany, 145 p.

Latifi, H., Nothdurfet, A. and Koch, B., 2010, Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: optimization of variable selection on optical/Lidar-derived predictors. *Forestry Journal*, **83**, pp. 395–407.

Lemay, V. and Temesgen, H., 2005, Comparison of nearest neighbor methods for estimating basal area and stems per hectare using aerial auxiliary variables. *Forest Science*, **51**, pp. 109–199.

Lister, A., Hoppus, M. and Czaplewski, R.L., 2004, *K*-nearest neighbor imputation of forest inventory variables in New Hampshire. In *Remote Sensing for Field Users, Proceedings of the Tenth Forest Service Remote Sensing Applications Conference*, 5–9 April 2004, Salt Lake City, UT.

Makela, H. and Pekkarinen, A., 2004, Estimation of forest stands volumes by Landsat TM imagery and stand-level field-inventory data. *Forest Ecology and Management*, **196**, pp. 245–255.

Mattera, D. and Haykin, S., 1999, Support vector machines for dynamic reconstruction of a chaotic system. In *Advances in Kernel Methods: Support Vector Learning*, B. Schölkopf, C.J.C. Burges and A.J. Smola (Eds.), pp. 211−242 (Cambridge, MA: MIT Press).

Mayer, R.R. and Scribner, D.A., 2002, Extending the normalized difference vegetation index (NDVI) to short-wave infrared radiation (SWIR) (1- to 2.5-μm). In *Proceedings of the Imaging Spectrometry VIII Conference*, 8 July 2002, Seattle, WA.

McInerney, D.O. and Nieuwenhuis, M., 2009, A comparative analysis of *k*NN and decision tree methods for the Irish National Forest Inventory. *International Journal of Remote Sensing*, **30**, pp. 4937–4955.

McRobert, R., Tomppo, E., Finley, A. and Heikkinen, J., 2007, Estimating aerial means and variances of forest attributes using the *k*-nearest neighbors technique and satellite imagery. *Remote Sensing of Environment*, **111**, pp. 466–480.

McRobert, R.E., 2009, Diagnostic tools for nearest neighbors techniques when used with satellite imagery. *Remote Sensing of Environment*, **113**, pp. 489–499.

Mohammadi, J., Shataee, Sh., Yaghmaee, F. and Mahiny, A., 2010, Modeling forest stand volume and tree density using Landsat ETM+ data. *International Journal of Remote Sensing*, **31**, pp. 2959–2975.

Muukkonen, P. and Heiskanen, J., 2005, Estimating biomass for boreal forests using ASTER satellite data combined with stand wise forest inventory data. *Remote Sensing of Environment*, **99**, pp. 434–447.

Nilsson, M., 1997, Estimation of forest variables using satellite image data and airborne lidar. PhD thesis, Acta Universitatis Agriculturae Suecia, Silvestria 17. Department of Forest Resource Management and Geomatics, Swedish University of Agricultural Sciences, Umeå, Sweden.

Ohmann, J.L. and 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**, pp. 725–741.

Ostapowicz, K., Lakes, T. and Kozak, J., 2010, Modeling of land cover change using support vector machine. In *13th AGILE International Conference on Geographic Information Science*, 10–14 May 2010, Guimarães, Portugal.

Perry, C.R. and Lautenschlagei, L.F., 1984, Functional equivalence of vegetation indices. *Remote Sensing and Environment*, **14**, pp. 169–182.

Reese, H., Nilsson, M., Sandström, P. and Olsson, H., 2002, Applications using estimates of forest parameters derived from satellite and forest inventory data. *Computers and Electronics in Agriculture*, **37**, pp. 37–55.

Richardson, A.J. and Wiegand, C.L., 1977, Distinguishing vegetation from soil background information. *Photogrammetric Engineering and Remote Sensing*, **43**, pp. 1541–1552.

Rose, J.W., Mass, R.H., Schell, J.A. and Deering, D.W., 1973, Monitoring vegetation systems in the Great Plains with ERTS. In *Earth Resource Technology Satellite-1 Symposium*, 10–14 December 1973, Goddard Space Flight Center, Washington, DC, pp. 309–317.

Roujean, J.L. and Breon, F.M., 1995, Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sensing of Environment*, **51**, pp. 375–384.

Schölkopf, B., Smola, A. and Müller, K.R., 1998, Nonlinear component analysis as a kernel eigen value problem. *Neural Computation*, **10**, pp. 1299−1319.

Shafri, H.Z.M. and Ramle, F.S.H., 2009, A comparison of support vector machine and decision tree classifications using satellite data of Langkawi islands. *Information Technology Journal*, **8**, pp. 64–70.

Sironen, S., Kangas, A. and Maltamo, M., 2010, Comparison of different non-parametric growth imputation methods in the presence of correlated observations. *Forestry*, **83**, pp. 39–51.

Sironen, S., Kangas, A., Maltamo, M. and Kangas, J., 2001, Estimating individual tree growth with the *k*-nearest neighbor and k-most similar neighbor methods. *Silva Fenninica*, **35**, pp. 453–467.

Sivanpillai, R., Smith, C.T., Srinivasan, R., Messina, M.G. and Ben Wu, X., 2006, Estimation of managed loblolly pine stands age and density with Landsat ETM+ data. *Forest Ecology and Management*, **223**, pp. 247–254.

Statistica, 2010, Electronic text book, Stat Soft Inc. Available online at: www.Statsoft.com

Tatjana, K., Suppan, F. and Schneider, W., 2007, The impact of relative radiometric calibration on the accuracy of *k*NN-predictions of forest attributes. *Remote Sensing of Environment*, **110**, pp. 431–437.

Thiam, A.K., 1997, Geographic information system and remote sensing methods for assessing and monitoring land degradation in the shale: the case of Southern Mauritania. Doctoral dissertation, Darks University.

Tomppo, E., 2006, The Finnish multi-source national forest inventory, small area estimation and map production. In *Forest Inventory – Methodology and Applications*, A. Kangas and M. Maltamo (Eds.), pp. 195–224 (The Netherlands: Springer).

Tucker, F., 1979, Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, **8**, pp. 127–150.

Tuominen, S., 2007, Estimation of local forest attributes, utilizing two-phase sampling and auxiliary data. Academic dissertation, Finnish Forest Research Institute, University of Helsinki, Helsinki, Finland, 45 p.

Vapnik, V., 1998, *Statistical Learning Theory* (New York: John Wiley & Sons).

Vauhkonen, J., Korpela, I., Maltamo, M. and Tokola, T., 2010, Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics. *Remote Sensing of Environment*, **114**, pp. 1263–1276.

Walton, J.T., 2008, Sub pixel urban land cover estimation: comparing cubist, random forests, and support vector regression. *Photogrammetric Engineering & Remote Sensing*, **74**, pp. 1213–1222.

Wang, X., Yang, C., Qin, B. and Gui, W., 2005, Parameter selection of support vector regression based on hybrid optimization algorithm and its application. *Journal of Control Theory and Applications*, **4**, pp. 371–376.

Wang, Y., Wang, J., Du, W., Wang, C., Liang, Y., Zhou, C. and Huang, L., 2009, Immune particle swarm optimization for support vector regression on forest fire prediction. In *Advances in Neural Networks*, W. Yu, H. He and N. Zhang (Eds.), ISNN 2009, Part II, LNCS 5552, pp. 382–390 (Berlin: Springer).

Wang, Z. and Brenner, A., 2009, An integrated method for forest canopy cover mapping using Landsat ETM+ imagery. In *ASPRS/MAPPS Fall Conference*, 16–19 November 2009, San Antonio, TX.

Yarbrough, L.D., Easson, G. and Kuszmaul, J.S., 2005, Using at-sensor radiance and reflectance tasseled cap transforms applied to change detection for the ASTER sensor. In *IEEE Third International Workshop on the Analysis of Multi-temporal Remote Sensing Images*, 16–18 May 2005, Biloxi, MS.

Zhang, R. and Ma, J., 2008, An improved SVM method P-SVM for classification of remotely sensed data. *International Journal of Remote Sensing*, **29**, pp. 6029–6036.