#### M. LANG ET AL.

# Assessment of Different Estimation Algorithms and Remote Sensing Data Sources for Regional Level Wood Volume Mapping in Hemiboreal Mixed Forests

#### MAIT LANG\*1,3, LINDA GULBE<sup>2</sup>, AGRIS TRAŠKOVS<sup>2</sup> AND ARTŪRS STEPČENKO<sup>2</sup>

<sup>1</sup> Tartu Observatory, 61602 Tõravere, Tartumaa, Estonia, mait.lang@to.ee

<sup>2</sup> Engineering Research institute "Ventspils International Radio Astronomy Centre", Ventspils University College, Inzenieru 101, LV-3601, Ventspils, Latvia

<sup>3</sup> Institute of Forestry and Rural Engineering, Estonian University of Life Sciences, Kreutzwaldi 5, 51014 Tartu, Estonia

Lang, M., Gulbe, L., Traškovs, A. and Stepčenko, A. 2016. Assessment of Different Estimation Algorithms and Remote Sensing Data Sources for Regional Level Wood Volume Mapping in Hemiboreal Mixed Forests. *Baltic Forestry* 22(2): 283-296.

#### Abstract

Remote sensing data provide opportunity to estimate wood volume in vast areas with lower financial expenses compared to field measurements. In this study, we tested wood volume mapping of hemiboreal mixed forests at stand-level and regionally using forest management inventory data as a reference set, various remote sensing data sources (Landsat-5 TM, Landsat-7 ETM+, SPOT-4 HRVIR, ALOS PALSAR, airborne laser scanner data) and three nonparametric estimation algorithms (k-nearest neighbours (k-NN), general regression neural network, regression tree). The experiment in Kurzeme region, Latvia, was organized as case studies regarding some aspects of the estimation procedure: impact of randomness in reference set sample on the k-NN volume estimation, assessment of the influence of the image and training plot combination on the k-NN volume estimations, comparison of the estimation algorithms and comparison of multisource and multitemporal data fusion. All the estimators performed quite similarly due to the complex relationships between forest inventory data and remote sensing data. The smallest RMSE=60 m3/ha was achieved in the special study site in Slitere National Park by combining five feature variables that included the 70th percentile of the ALS point cloud height distribution, green band from the Landsat and SPOT image, and NIR and SWIR bands from the Landsat image. When spectral feature variables and reference samples from full-size satellite scenes were used, the RMSE of wood volume estimates ranged from 72 m<sup>3</sup>/ha to 129 m<sup>3</sup>/ha for forest in the scenes. Higher estimation accuracy was obtained with mid-growing season Landsat images and then with SPOT images from the snow-covered period. Case studies indicated that the estimation accuracy depends on a particular image, but the randomness in the reference set does not impact accuracy substantially when there is a sufficient number of reference sample plots. The combined influence of a particular image and reference samples for the image was detectable and the RMSE of the stand-level wood volume estimates in the image overlap areas ranged between 17 m<sup>3</sup>/ha and 42 m<sup>3</sup>/ha, and mean error of estimate ranged from - 26 m3/ha to 21 m3/ha.

Keywords: wood volume, multisource remote sensing data, nonparametric estimators, forest management inventory.

#### Introduction

Regularly updated regional data about forest resources are required for sustainable forest management planning and assessment of available biomass for energy production. Two forest inventory methods are used in Latvia: the first is sampling based National Forest Inventory (NFI) and the second is regular stand wise forest inventory (RFI or forest management inventory). The forest management inventory in Latvia ensures information about forest management units (stands) which are mapped and inventoried according to specific rules (Forest 2000) using aerial photos and field inspection. According to legal provisions in Latvia, the data are updated with 20year interval for each stand if no management activities are done by the forest owner (Forest 2000). Construction of wood volume (V) maps using remote sensing data and supervised estimation algorithms offers additional input to forest inventories such as the RFI for overall analysis of the current state of the forests and for change detection (Wulder and Franklin 2003). Multispectral satellite images provide data over vast areas and give an opportunity to produce fast wall-to-wall estimates at lower financial expenses compared to field measurements required for the RFI (Lu 2006). However, the practical use of the remote sensing image data is limited by the relationships between

ASSESSMENT OF DIFFERENT ESTIMATION ALGORITHMS AND REMOTE SENSING /.../ M. LANG ET AL.

extracted feature variables and the forest inventory variables of interest and the properties of processing algorithm. Lu (2006) concluded that major approaches to the estimation of biomass are regression models; algorithms based on k-nearest neighbours and neural networks. The k-nearest neighbour estimator (k-NN) has attracted a lot of interest for forestry applications (Tomppo 1991, Franco-Lopez et al. 2001, McRoberts and Tomppo 2007, Gjertsen 2007). The effectiveness of k-NN is determined by the simplicity of the method: convenient inclusion of new data sources, ability to estimate many response variables at once including both numerical and discrete variables, and the ability to preserve natural variations in sample data (Holmström and Fransson 2003, McRoberts 2012). However, the drawback of k-NN is high computational complexity in time and space motivating comparison of k-NN and other estimators to achieve equivalent or higher accuracy with a decreased computational cost.

The accuracy of wood volume estimation depends on the type of remote sensing data and spatial resolution is one of the most important characteristics of images (Maltamo et al. 2004). The most recent and popular data source for wood volume and biomass estimation is airborne lidar data (Li et al. 2008, Zhao et al. 2012). However, the use of lidar data is limited by high acquisition costs (Koch 2010). Hyyppä et al. (2000) investigated optical spectral data from Landsat-5 TM, SPOT PAN and XS, and radar data from ERS-1/2 for the retrieval of forest stand attributes with the main focus on stem wood volume and concluded that optical data are more informative than radar data. Particularly, SPOT 20 m and 10 m spatial resolution images were found to be better (standard error 78.9 m3/ha) than 30 m Landsat images (standard error 87.5 m<sup>3</sup>/ha). Similar results for Finnish forests were achieved by Mäkela and Pekkarinen (2004) using Landsat-5 TM data (RMSE 71.3-80.5 m<sup>3</sup>/ha). Maltamo et al. (2004) emphasized the saturation of forest reflectance and wood volume relationship and pointed out that after canopy closure the observed spectral values do not correlate well with wood volume. Magnusson et al. (2007) evaluated ALOS PALSAR Fine Beam Single Polarization mode data and achieved an RMSE of 30% for the best case. There have also been studies to evaluate the combined processing of different sensor data and multitemporal data (Maltamo et al. 2004, Maltamo et al. 2006, Popescu et al. 2004, Nelson et al. 2007, Cartus et al. 2011). Since each specific data source has its limitations, the multisource approach in the general case may improve estimation accuracy (Koch 2010).

The aim of this study was stem wood volume (m<sup>3</sup>/ha) mapping in mixed hemiboreal forests at stand-level and for a larger region using a reference set drawn from a forest management inventory database, various remote sensing data sources and three non-parametric estimation al-

gorithms. This study is presented as case studies devoted to some particular aspects of the procedure e.g. choice of algorithm or impact of the image acquisition season on the estimation accuracy. The following case studies were performed:

• evaluation of the performance of SPOT-4 HRVIR, Landsat-5 TM and Landsat-5 ETM+ satellite images for wood volume mapping using k-NN;

· analysis of the influence of reference set randomness on the wood volume estimation accuracy;

· assessment of the combined influence of image characteristics and reference set observations on the k-NN estimated wood volume;

multisource and multitemporal data fusion;

· comparison of three nonparametric estimators: k-NN, general regression neural network (GRNN) and regression tree (RT).

#### **Material and Methods**

#### Study site

Kurzeme planning region is located in the western part of Latvia between 56° and 58° North and 21° and 23.5° East with a total area of 13,607 km<sup>2</sup> (Figure 1). The climate in Kurzeme is usual to the temperate climate zone with substantial maritime features due to the vicinity of the Baltic Sea (Rutkis 1960). The frequent cloud cover limits the availability of multispectral satellite images. The mean cover of low clouds is smaller in northern and western parts of Kurzeme (Avotniece et al. 2015). If information on a specific period of time is required, then it could be necessary to use satellite images of different seasons or years to cover the whole region. Approximately half of Kurzeme area is covered by forests. In the landscape of Kurzeme forest patches form heterogeneous mosaics with agricultural land, waters and urban areas (Saliņš 2002). Latvia is situated in the hemiboreal forest zone where evergreen coniferous tree species and broadleaved deciduous tree species can be found in the landscape (Znotina 2002).

Coniferous trees are represented by four species and deciduous trees by 17 tree species in Kurzeme (Bušs 1987). While the coniferous trees often form pure stands, the deciduous trees are found almost exclusively in mixed stands. The dominant coniferous tree species in Kurzeme region are due to climatic and economic factors Scots pine (43.1% of the total area, Pinus sylvestris L.) and Norway spruce (15.3%, Picea abies (L.) Karst.). The most common deciduous tree species are birch (21.5%, Betula pendula Roth and Betula pubescens Ehrh.), black alder (1.7%, Alnus glutinosa (L.) Gaertn.), aspen (2.0%, Populus tremula L.) and gray alder (3.1%, Alnus incana (L.) Moench). The percentage shows the proportion of the area covered by specific dominant tree species as recorded

M. LANG ET AL.

ASSESSMENT OF DIFFERENT ESTIMATION ALGORITHMS AND REMOTE SENSING /.../



**Figure 1.** The study region is indicated by pattern. The triangle shows location of the special test site in Slitere National Park

in the forest inventory data base. In contrast to the small number of tree species, the forest structure in Kurzeme is diverse. Tree species composition, stand density, age and volume vary substantially even within small geographical regions. The mean forest stand size is approximately 1.3 hectares. Our smaller and detailed study site where lidar data were acquired is located in Slitere National Park in the northern part of Kurzeme. Slitere National Park was founded in 1921 and is one of the most important Natura 2000 sites in Latvia. Tree species composition in this study site is typical of forests in Kurzeme, but the forests are natural and less fragmented with limited management activities performed.

#### Data sets and data pre-processing

In this paper we use the terminology proposed by McRoberts (2012). The feature variables for wood volume estimation experiments were extracted from four ancillary data sets: 1) Landsat-5 Thematic Mapper (TM) and Landsat-7 Enhanced Thematic Mapper Plus (ETM+) images; 2) SPOT-4 High Resolution Visible and Infrared (HRVIR) images; 3) ALOS Phased Array type L-band Synthetic Aperture Radar (PALSAR) data; and 4) airborne discrete return laser scanner data. The ancillary data (image data) were obtained for the whole area of Kurzeme, except the airborne lidar dataset which was available only for 63.3 km<sup>2</sup> in Slitere National Park. The list of all employed satellite images and derived feature variables with their short labels is given in Table 1. If a specific band of a multispectral image is discussed in the text, then bX, where X is band number, is added to the image label: e.g. the first band from Landsat-5 TM image from 28.06.2010 will be referred to as imL14b1.

SPOT-4 images were from different seasons: leaveson autumn (imS11), leaves-off spring (imS12) and winter (imS1, imS2, imS4-imS10) to investigate temporal effects on the wood volume estimation. The SPOT-4 winter time image set covered the whole Kurzeme area. Satellite images from the time of the year when snow covers the ground surface have been used by Peterson et al. (2004) for forest mapping and the fraction of shaded bright snow in a forest is probably correlated with wood volume and biomass of the forest (Leboeuf et al. 2007). The SPOT-4 images (imS1-imS12) were projected into the Latvian state coordinate system LKS92 and ortorectified using ground control points (40 points per scene) from a topographic map (scale 1:50000) with the average root mean square error of 20 m. The pixel size was set to 20 m. The second order polynomial was used to calculate rectified coordinates and nearest neighbour resampling method was used for pixel values.

The Landsat images (imL13-imL18) were selected and downloaded from the USGS image archive to obtain the maximum continuous cover over the Kurzeme test site. Some of the images had regions with haze and clouds and the Landsat-7 ETM+ image (imL13) had missing scanlines due to operating in the SLC-off mode since the year 2003. All the Landsat images (imL13-imL18) were from the phenology situation when trees in Kurzeme region have full foliage. Two of the Landsat images (imL17 and imL18), however, were from the beginning of September when the illumination conditions are not the best due to the low Solar angle at the time of image acquisition. The Landsat images were projected into the LKS92 coordinate system by using nearest neighbour sampling and 25 m pixels. The areas with clouds and cloud shadows were digitized manually and assigned a missing data value. Since spectral bands in the Landsat images cover slightly different areas, additional masks for each image were created to determine the common area. Only the multi-spectral channels from the optical spectral range (400-2500 nm) were used from the Landsat images.

If images from different dates are used for map construction with a k-NN estimator, then it seems feasible to join the individual images as a mosaic. Hence, some normalization may be required, since spectral radiance of the images is influenced by the differences in illumination, by the atmospheric conditions, sensor decay and also by the actual natural changes of forests. Olsson (1993) proposed a simple, regression-based method for relative calibration of satellite images using closed canopy coniferous stands older than 40 years as the stable reflectance objects. Koukal et al. (2007) showed that if the phenological differences are not large then regression models can be applied to create image mosaics to increase the number of sample plots per image field for the k-NN. We tested the relative calibration of the Landsat images (imL13-imL18) using middle-aged coniferous stands and linear regression according to Olsson (1993) to create a single image for Kurzeme region. However, validation in the image overlap areas showed significant differences between the

#### ASSESSMENT OF DIFFERENT ESTIMATION ALGORITHMS AND REMOTE SENSING /.../ M. LANG ET AL.

estimated radiance values, and therefore no radiometric calibration was applied to the images. On the other hand, the number of training samples in a reference set for each Landsat image was rather large, ranging between 2,879 and 8,979 forest stands per image and the estimation of response variables was carried out using each individual scene separately.

The ALOS PALSAR data provides information in the L-band with a spatial resolution of 20 m. The radar data were ortorectified and resampled into the LKS92 coordinate system. The ALOS PALSAR data were calibrated for antenna profile offset and finally HV polarization band sigma values were filtered with a 3x3 median filter. Sigma is a measure of the reflective strength of the radar target.

The airborne laser scanning (ALS) data for the special study site in Slitere National Park was acquired with the ALTM Gemini scanner with the following parameters: the laser pulse repetition rate was 125 kHz, scan frequency 50 Hz, scanning angle  $\pm 17$  degrees and the average flight height 700 m resulting in the mean point density of 6.28 points/m<sup>2</sup>. Lidar data processing was done in FUSION/ LDV environment (McGaughey 2014). A digital terrain model (DTM) was created after extracting points for the ground surface and a digital surface model (DSM) was created by filtering the highest points from the ALS data. Finally, a canopy height model (CHM) was obtained by subtracting the DTM from the DSM. The spatial resolution of the CHM was 1m. The following feature variables were calculated from the CHM for each forest stand (averaged over each forest stand polygon): mean height, the 70<sup>th</sup> percentile of pulse return height distribution, which had the strongest correlation with forest height, and canopy cover. Canopy cover was calculated as the proportion of canopy hits above the 1.5 m threshold.

A forest management inventory data base obtained from the JSC "Latvijas valsts meži" was used to select forest stands for k-NN experiments. The RFI database contained tree species composition data, site fertility class, forest age (A), wood volume in cubic meters per hectare (V), and other common inventory variables. First, a random subset consisting of 69,038 forest land records was selected from the RFI database as possible training and validation stands. The information in the records was collected from the year 2005 and onwards and the size of the forests was more than 2.1 ha. The records were exclusively assigned to a training data set (reference set) or validation data set. Since the remote sensing data were from 2010-2012, all objects (forest stands) in the reference set and in the validation set were tested for 1) variation of the spectral radiance with their polygons and 2) deviation from forest age or wood volume to the radiance relationship. The radiance variation coefficient  $(L_{var})$  for each band of the Landsat images (imL13-imL18) within each forest stand polygon was calculated as the ratio be-

Next, to detect disturbed stands in the training or validation dataset we searched outliers based on the relationship between forest age or wood volume and forest stand spectral radiance (L). The relationships were fitted using the model (1) (Nilson and Peterson 1994) for each band of the images imL13-imL18

$$L = a_1 + (1 - a_2 \exp(a_3 x)), \tag{1}$$

where  $a_1, a_2$  and  $a_3$  are the estimated parameters, exp is the exponent symbol and x stands for wood volume or stand age.

For forest age dependent spectral radiance regression model the L from each image band was calculated as the mean value of pixels within each forest stand polygon. When the independent variable in the model (1) was forest age, then two model residual standard errors were used for outlier detection. The filter was not applied for the stands that were younger than five years due to the natural rapid spectral reflectance change in this particular period (Nilson and Peterson 1994). For wood volume dependent spectral radiance regression model the spectral radiance for each forest stand was extracted from the forest polygon centroid pixel, since our k-NN implementations were designed for small sample plots represented by their nearest pixel in raster images. Outlier detection in the wood volume-based radiance regression model (1) was based on the 2.5 model residual standard error rule. The filter was not applied to the stands that had a wood volume less than 20 m<sup>3</sup>/ha. After the filtering the validation dataset contained 18,195 stands and the training dataset contained exclusively 18,816 stands.

#### Nonparametric estimation methods

Three of the most frequently employed choices for wood volume and biomass estimators are k-nearest neighbours (k-NN), multiple regression analysis and neural networks (Lu 2006). The basic principle of k-NN is simple – it is a supervised classifier which compares the target pixel to be classified with all the sample pixels in the reference set and assigns a response variable value e.g. wood volume to the pixel in the target set by taking into account the field measurement values of k nearest reference observations according to the feature variables (Franco-Lopez et al. 2001, McRoberts 2012). Computational complexity is high in time and space, but the k-NN can handle numeric response variables (e.g. wood volume) and discrete classes (e.g. tree species). Performance of the algorithm depends on the internal parameter k – the number of the

ASSESSMENT OF DIFFERENT ESTIMATION ALGORITHMS AND REMOTE SENSING /.../ M. LANG ET AL.



Figure 2. Distribution of the reference set and validation set observations according to wood volume and main species (colours are available in the electronic version)

Table 1. The list of satellite images and other feature variables employed in the study. For the Landsat images the path and row according to the Worldwide Reference System is given. Abbreviations:  $h_{Sun}$  – solar elevation angle,  $S_{azimuth}$  – Sun azimuth, degrees; VNA - View nadir angle, degrees; ALS - airborne laser scanner

Label	Sensor	Date	Short description
imS1	SPOT-4 HRVIR2	28.02.2011	h <sub>sun</sub> =22.6, S <sub>azimuth</sub> =162.4, VNA=+6.9
imS2	SPOT-4 HRVIR2	28.02.2011	h <sub>sun</sub> =23.6, S <sub>azimuth</sub> =162.0, VNA=+6.9
imS3	SPOT-4 HRVIR1	02.07.2012	h <sub>sun</sub> =50.6, S <sub>azimuth</sub> =139.0, VNA=-17.9
imS4	SPOT-4 HRVIR2	28.02.2011	h <sub>sun</sub> =23.4, S <sub>azimuth</sub> =161.7, VNA=+6.9
imS5	SPOT-4 HRVIR2	28.02.2011	h <sub>sun</sub> =22.6, S <sub>azimuth</sub> =161.4, VNA=+9.9
imS6	SPOT-4 HRVIR1	28.02.2011	h <sub>sun</sub> =22.6, S <sub>azimuth</sub> =163.2, VNA=+9.9
imS7	SPOT-4 HRVIR1	28.02.2011	h <sub>sun</sub> =23.6, S <sub>azimuth</sub> =162.9, VNA=+9.9
imS8	SPOT-4 HRVIR1	28.02.2011	h <sub>sun</sub> =23.4, S <sub>azimuth</sub> =162.5, VNA=+9.9
imS9	SPOT-4 HRVIR2	17.02.2011	h <sub>sun</sub> =20.3, S <sub>azimuth</sub> =165.6, VNA=+21.9
imS10	SPOT-4 HRVIR1	23.02.2011	h <sub>sun</sub> =21.9, S <sub>azimuth</sub> =162.2, VNA=+9.3
imS11	SPOT-4 HRVIR1	30.09.2011	h <sub>sun</sub> =29.1, S <sub>azimuth</sub> =161.5, VNA=-3.6
imS12	SPOT-4 HRVIR1	31.03.2011	h <sub>sun</sub> =35.5, S <sub>azimuth</sub> =164.7, VNA=+17.5
imL13	Landsat-7 ETM+	07.06.2011	h <sub>sun</sub> =53.7, S <sub>azimuth</sub> = 155.7, path 188, row 020, SLC-off
imL14	Landsat-5 TM	28.06.2010	h <sub>sun</sub> =53.8, S <sub>azimuth</sub> =152.7, path 188, row 020
imL15	Landsat-5 TM	29.06.2011	h <sub>sun</sub> =53.6, S <sub>azimuth</sub> =152.1, path 190, row 020
imL16	Landsat-5 TM	18.08.2011	<i>h<sub>sun</sub></i> =43.8, <i>S<sub>azimuth</sub></i> =155.6, path 188, row 020
imL17	Landsat-5 TM	07.09.2010	h <sub>sun</sub> =37.2, S <sub>azimuth</sub> =159.9, path 189, row 020
imL18	Landsat-5 TM	07.09.2010	h <sub>sun</sub> =38.4, S <sub>azimuth</sub> =158.7, path 189, row 021
imR19	ALOS PALSAR	28.06.2010	Incidence angle: 38.721 degrees
datLCH	ALTM Gemini	24.08.2011	ALS-based canopy height model
datLMH	ALTM Gemini	24.08.2011	ALS-based mean canopy height
datL70	ALTM Gemini	24.08.2011	70 <sup>th</sup> -percentile of ALS pulse return height distribution

2016, Vol. 22, No. 2 (43)

ASSESSMENT OF DIFFERENT ESTIMATION ALGORITHMS AND REMOTE SENSING /.../

nearest neighbours employed. Most recent k-NN implementations use also optimization routines (McRoberts et al. 2015), but our implementations had just the basic functionality described above.

The general regression neural network (GRNN) is a logical extension of k-NN and a type of radial basis neural network. The GRNN can be easily trained by setting the spread of radial basis functions which determine the impact of sample points on the estimation according to their distance from the pixel to be classified (Specht 1991). The spread is the only internal parameter to be fitted for estimation.

Multivariate regression aims to find relations between dependent (numeric forest inventory variables) and independent (remote sensing data values) variables. From the wide choice of regression techniques, regression trees (RT) were chosen. An RT predicts the response variable by following decisions in the tree from the root node to the leaf nodes (Moisen and Frescino 2002). The RT is binary, meaning that each step in a prediction checks the value of one predictor. The algorithm is very fast, but the main drawback is the limited number of estimated values compared to k-NN and GRNN which theoretically allow estimating a variable in a continuous interval of values. Similarly to neural networks, the RT can be overfitted to training data resulting in weak performance on validation. Regression tree size refers to a number of decisions included in the tree structure. More decision nodes in the regression tree can describe smaller variations in the data set, but there is a risk of overfitting. Pruning of the tree reduces the number of decision nodes and the overfitting risk retaining the basic relationships between variables in the data set. Hence, the most important internal parameter of the RT to be fitted is the pruning level, meaning the reduction in tree size and removal of leaves which correspond to small variations in remote sensing data. We selected k-NN as the base algorithm in our case studies while the GRNN and RT were tested in some case studies for comparison with the k-NN.

All our implementations of the algorithms worked at pixel-level. For all the case studies except study No. 5, centroid pixels of forest stands were selected to obtain values from raster images for feature variables. For the fifth study stand-level mean values were used via a separate query module. The mean value of pixels from within each stand polygon was used for validation of the constructed wood volume maps. For the case studies 2, 3 and 4 the k-NN implementation tknn from Tartu Observatory (T. Lükk, personal contacts) was used. For the other case studies the programs written in Ventspils University College were used (including the k-NN implementation vknn). Both k-NN implementations are based on Franco-Lopez et al. (2001) using the Euclidean distance and they differed only by the t-parameter value (see Eq. 1 in Franco-Lopez et al. (2001)). The t=2 was used in tknn and t=1 was used in vknn. We compared the two implementations and the difference in the overall RMSE of the estimated wood volume was marginal.

#### Methodology of validation

The accuracy assessment of the estimated wood volume  $\hat{V}$  was carried out on the validation data set using the root mean square error (RMSE):

$$RMSE = \sqrt{\frac{\sum (\hat{y}_i - y_i)^2}{n}}, \qquad (2)$$

where  $y_i$  is the observed wood volume,  $\hat{y}_i$  is the estimated wood volume and *n* is the number of observations. The mean error of estimate (MEE) was calculated as:

$$MEE = \frac{\sum (\hat{y}_i - y_i)}{n}.$$
 (3)

The equations (2) and (3) were also used to compare the wood volume estimates extracted for forest stands located in overlapping areas of the Landsat images.

#### Case study 1: Wood volume estimation with k-NN and SPOT-4 HRVIR images

Feature variables for wood volume estimation were extracted from the image set imS. All of the SPOT-4 HR-VIR bands were employed with equal weights in the estimation procedure. The number of the nearest neighbours k=21 to achieve the highest accuracy was evaluated in case study 6. The training accuracy at pixel-level was assessed by using the RMSE of leave-one-out cross-validation for each image to investigate the impact of image acquisition season and the reference set randomness. The estimation accuracy was calculated by using validation stands.

#### Case study 2: Wood volume estimation with k-NN, Landsat-5 TM and Landsat-7 ETM+ images

Feature variables for wood volume estimation were extracted from the image set imL. The blue band (TM b1, ETM+ b1) and the second shortwave infrared band (SWIR: TM b7, ETM+ b7) were excluded due to strong atmospheric path radiance in the blue band and strong correlation between the two SWIR bands of the scanners. The bands b2, b3, b4, b5 of the images in the imL set were used with equal weights. The number of neighbours k=5was estimated from Figure 2 in the paper by McRoberts (2012) as a compromise. The k=5 was also used by Fazakas et al (1999) who comment that using  $k \ge 5-10$  does not substantially increase estimation accuracy but introduces artificially improved correlation between the estimated values of response variables. The chosen k was much smaller than in case study 1, but using a sufficiently small

#### ASSESSMENT OF DIFFERENT ESTIMATION ALGORITHMS AND REMOTE SENSING /.../

M. LANG ET AL.

number of reference observations in the k-NN preserves more natural variation which may be beneficial for standlevel estimates at extreme values of response variables. The estimation accuracy was calculated by using validation stands.

### Case study 3: Impact of randomness in the reference set sample on the k-NN volume estimations

Case study 3 used the same feature variables and algorithms as case study 2. For each image in the imL set, the k-NN was run ten times by using 33% of randomly drawn reference observations from the image area. The number of selected reference observations ranged between 809 and 2,985 for individual images, but did not vary much for any particular image. For each forest stand in the validation dataset the mean estimated wood volume was calculated after each k-NN run using the pixels located within the stand border. The dependence of the standard error of the ten estimates on the mean estimated value and on the V from forest inventory data was analysed.

#### Case study 4: Assessment of the influence of image and training plot combination on the k-NN volume estimations

Since the Landsat images in the imL set were from different dates and years (Table 1) we tested relative radio-

metric calibration first, but the results were unsatisfactory. An assumption was made that the number of training samples in the reference set was sufficiently large to describe the variation for the k-NN experiments. The combined effect of the image and the available reference set samples was assessed by comparing the wood volume estimates of the validation stands in the image overlap areas.

#### Case study 5: Selection of multisource data variables in Slitere National Park

Figure 3 shows some examples of the relationships between V and remote sensing data-based feature variables (average value per stand polygon). The relationships for multispectral satellite images are nonlinear and for a standing wood volume of larger than 110 m<sup>3</sup>/ha multispectral feature variables tend to saturate but the variance remains high. The canopy height estimate from the ALS did not have such a saturating relationship with V, and ALOS PALSAR radar data are of interest because the data acquisition is almost independent from clouds.

The aim of this test was to find an optimal multisource/multitemporal set of feature variables for the standing wood volume estimation. A total of 1,620 stands were used in this case study. The experiment of data fusion was performed as follows. For each forest stand, feature variables were obtained as the mean value of pixels



**Figure 3.** Relationships between standing wood volume and a) imS6b1, b) imL14b2, c) imR19, d) the 70<sup>th</sup> percentile of the ALS height over ground. The feature variables are described in Table 1.

2016, Vol. 22, No. 2 (43)	ISSN 2029-9230
, , , , , ,	

ASSESSMENT OF DIFFERENT ESTIMATION ALGORITHMS AND REMOTE SENSING /.../

within stands from each layer in the imS6, imL14, imR19, datLCH, datLMH and datL70 data sets. All the feature variables were standardized by using z-scores. Then the k-NN was run for all possible combinations of the variables (a total of 16,383 different feature variable combinations) and the RMSE of leave-one-out cross-validation was calculated.

#### **Case study 6: Comparison of estimators**

The case study was based on the image imS8. Three different estimation methods (k-NN, GRNN, RT) were applied in the test with varying internal parameter values (*k* for k-NN, pruning level for RT and the spread of basis function for the GRNN) to perform the internal parameter optimization. The internal parameter value resulting in the smallest RMSE was selected as the optimal value. For the study, 3,705 centroid pixels of the forest stand polygons were available. One thousand pixels were randomly drawn as a training set, and the rest (2,705) were used as a validation set. The variability in the accuracy depending on the training set selection was checked by selecting 100 different training (1,000 pixels) and validation sets (2,705 pixels) and evaluating the RMSE.

#### Results

### Case study 1: Wood volume estimation with k-NN and SPOT-4 HRVIR images

Table 2 summarizes the the leave-one-out crossvalidation (LOOC) statistics and the results show some dependence of the estimation accuracy on the individual images, but there was no evident impact of the image acquisition season. The LOOC RMSE values are slightly more optimistic than the separate validation set based RMSE. A similar conclusion about the LOOC was made by Tomppo et al. (2009). Wood volume was estimated with RMSE=115 m3/ha according to validation of the independent set of forest stands when feature variables from the SPOT-4 HRVIR image set imS were used in the k-NN. When the difference between the estimated wood volume  $\hat{V}$  and inventoried wood volume V was looked as a function of V, then there was an evident lack of fit of the estimates relative to the measured values. Wood volume in the youngest stands was estimated with a small positive error, however, with increase in V the overestimation rapidly increased reaching the maximum at about  $V \cong 100$ m<sup>3</sup>/ha. After the maximum point of overestimation the difference between the  $\hat{V}$  and V started to decrease almost linearly reaching the zero value at about the overall mean of V. With the further increase in V the V was always underestimated. At  $V \cong 450 \text{ m}^3/\text{ha}$ , the underestimation was substantial. However, the overall mean wood volume estimate had only a marginal MEE, since with the aggrega-

<b>Table 2.</b> The RMSE and MEE of wood volume $(V, m^3/ha)$ es-
timates using feature variables from multi-spectral images. For
SPOT-4 images (imS) the results are based on the reference set
pixels using leave-one out cross-validation. For the Landsat
images (imL) the results are based on validation stands. The
image descriptions are in Table 1

Image	Reference observations	MEE (m³/ha)	RMSE (m³/ha)
imS1	2,219	-4.8	94
imS2	4,382	-1.3	91
imS3	5,391	-1.4	89
imS4	2,562	-10.6	103
imS5	2,377	-2.3	104
imS6	494	0.9	92
imS7	5,666	7.0	96
imS8	3,621	-4.5	87
imS9	3,621	0.6	106
imS10	487	3.9	129
imS11	1,230	-2.4	113
imL13	7,434	-5.9	74
imL14	8,979	-5.8	72
imL15	5,639	-5.1	75
imL16	2,828	-1.7	77
imL17	6,518	-0.7	80
imL18	4,288	0.0	81

tion of  $\hat{V}$  the estimation error is cancelled out. Some of the validation stands were located in overlapping areas of the SPOT images. The difference in the estimated wood volume for the stands showed a mean RMSE=30 m<sup>3</sup>/ha when the estimates from different images were compared to each other.

#### Case study 2: Wood volume estimation with k-NN, Landsat-5 TM and Landsat-7 ETM+ images

The mean wood volume based on the Landsat images was marginally (5 m<sup>3</sup>/ha) underestimated when the images from the first half of the growing season were used. The RMSE ranged from 72 m<sup>3</sup>/ha to 81 m<sup>3</sup>/ha (Table 2). An interesting phenomenon was that the RMSE was smaller for the images from the beginning or middle of the growing season and increased towards the end of the vegetation period. This can be related to the overall scene illumination and the range of pixel values in the images. Surprisingly, the RMSE of the wood volume estimations based on the Landsat images were less compared to the estimates based on the mid-growing season SPOT-4 HR-VIR image (imS3). The reason for the lower accuracy of *V* estimates can be related to the smaller instantaneous field

ISSN 2029-9230

**BALTIC FORESTRY** 

ASSESSMENT OF DIFFERENT ESTIMATION ALGORITHMS AND REMOTE SENSING /.../

of view of the HRVIR sensor compared to the TM and ETM+ sensors, since spectral signatures for the reference set in the k-NN were obtained from the centroid pixel of each stand. As a contrast to sample plot-based reference observations, the measured wood volume V for the reference set observations was obtained from the RFI database and was the mean value for the whole stand area. Both of the used k-NN implementations were developed for sample plot-based reference observations and for each observation the closest pixel from the maps of feature variables is attached. In this study, however, the response variables were obtained as forest stand-level estimates and feature variable data were extracted from the stand polygon centroid pixel. For an experiment we replaced the original centroid pixel data in the images imL13, imL14 and iml17 with stand mean values and ran the k-NN again to predict V. The constructed maps were visually slightly different from centroid pixel-based results. However, based on the validation set, the relationship between the two estimates was linear and there was hardly any difference in the lack of fit of the estimated wood volume between the two methods used to calculate the feature variable values.

### Case study 3: Impact of randomness in the reference set sample on the k-NN volume estimations

The standard error of the estimated stand-level wood volume was calculated from ten reference set subsamplebased k-NN estimates and it was usually less than 4 m3/ha and only in some younger stands reached 8 m3/ha. This indicates that the number of observations in the reference sets was sufficient. The results from the training plot subsampling test prove also that the detection of disturbed stands in the reference set was reliable. There was a notable dependence of the standard error on the estimated wood volume  $\hat{V}$  and the inventoried wood volume V. The standard error started to increase from almost zero value with  $\hat{V}$ , then reached its maximum at about  $\hat{V} \cong 100 \text{ m}^3/\text{ha}$ and then started to decrease again. However, the ratio of the standard error and  $\hat{V}$  was the largest at small wood volumes (4-10%) and constantly decreased with an increase in  $\dot{V}$ . Similar results were obtained with the standard error relative to the inventoried wood volume V but the scatter was substantially larger and in young stands the relative error was sometimes close to 30% of the measured value. This indicates that the estimated values are rather stable when the number of reference observations for a Landsat TM image is large (the smallest sample in this study was 809 stands). However, the estimated wood volume may have relatively large random errors compared to the measured wood volume at stand-level in younger than middleaged forests.

#### Case study 4: Assessment of the influence of image and training plot combination on the k-NN volume estimations

The k-NN wood volume estimates were dependent on the individual Landsat images (imL image set). The comparison of the wood volume estimates in image overlap areas revealed notable discrepancies that were dependent on the estimated value itself (Figure 4). In some cases the estimates were more scattered but unbiased and in some cases the scatter was small, but the estimates were biased. The RMSE of the stand-level wood volume estimates in the image overlap areas ranged between  $17 \text{ m}^3$ / ha and 42 m<sup>3</sup>/ha, and MEE ranged from -26 m<sup>3</sup>/ha to 21 m<sup>3</sup>/ha. The wood volume estimates were larger when feature variables from late summer and early autumn images were used (Figure 4a-e, Table 1). There was an interesting pair of Landsat-5 TM images (imL17 and imL18) taken from the same orbit on the same day. The wood volume of stands was, on average, overestimated by 17 m<sup>3</sup>/ha when reference observations from the southern image (imL18) were used (Figure 4c) compared to the northern image (imL17). This indicates clearly that there is a combined influence of reference set and properties of the particular image on the estimates of response variables in k-NN.

## Case study 5: Selection of multisource data variables in Slitere National Park

In this test one feature set included ALS-based feature variables in addition to spectral and radar data. The inclusion of ALS data increased the V estimation accuracy by 10 m<sup>3</sup>/ha while radar data-based feature variables showed no clear influence (Table 3). Starting with the two most informative feature variables (the 70th percentile of the ALS point cloud height distribution and NIR band from the Landsat image) the RMSE of the estimated wood volume was 63.2 m<sup>3</sup>/ha. By adding more feature variables, the RMSE first decreased and then started to increase again. Similar observations were made on the feature set where ALS data was excluded. The minimum RMSE was achieved by five or six feature variables that included the 70th percentile of the ALS point cloud height distribution, the green band from the Landsat and SPOT image, and NIR and SWIR bands from the Landsat image. However, the difference in the RMSE was small when the number of feature variables changed. Increasing the number of features over 6 also led to an increase in the RMSE due to the "curse of dimensionality" (Duda 2012). The most useful spectral bands for volume estimation were green (both winter and summer - imS6b1, imL14b2), near infrared and mid-infrared bands (imS6b4, imL14b4, imL14b7).

ASSESSMENT OF DIFFERENT ESTIMATION ALGORITHMS AND REMOTE SENSING /.../

ISSN 2029-9230



Figure 4. The k-NN estimated wood volume of validation set forest stands located in the overlapping areas of the Landsat images. The images are described in Table 1

Table 3. Search for optimal feature sets for the wood volume (V, m<sup>3</sup>/ha) estimation. Feature variable names are composed from the data set label (Table 1) and specific variable (e.g. band)

No. of features	Feature sets with ALS	RMSE (m³/ha)	Feature sets without ALS	RMSE (m³/ha)
2	datL70, imL14b4	63.20	imS6b3, imL14b2	77.37
3	datL70, imL14b4, imLb7	60.76	imS6b1, imS6b4, imL14b4	73.81
4	datL70, imL14b2, imL14b4, imL14b7	60.24	imS6b1, imS6b4, imL14b2, imL14b4	71.51
5	datL70, imS6b1, imL14b2, imL14b4, imL14b7	59.97	imS6b1, imS6b4, imL14b2, imL14b4, imL14b7	69.34
6	datL70, imS6b3, imS6b4, imL14b2, imL14b4, imL14b7	59.97	imS6b1, imS6b2, imS6b4, imL14b2, imL14b4, imL14b7	69.15
7	datL70, imS6b1, imS6b3, imS6b4, imL14b2, imL14b4, imL14b7	60.01	imS6b1, imS6b3, imS6b4, imL14b2, imL14b4, imL14b5, imL14b7	69.49
8	datL70, imS6b1, imS6b3, imS6b4, imL14b2, imL14b4, imL14b3, imL14b7	60.01	imS6b1, imS6b2, imS6b4, imL14b2, imL14b3, imL14b4, imL14b5, imL14b7	69.63
9	datL70, imS6b1, imS6b2, imS6b3, imS6b4, imL14b2, imL14b3, imL14b4, imL14b7	60.09	imS6b1, imS6b2, imS6b3, imS6b4, imL14b2, imL14b3, imL14b4, imL14b5, imL14b7	69.89
10	datL70, imS6b1, imS6b2, imS6b3, imS6b4, imL14b2, imL14b3, imL14b4, imL14b5, imL14b7	60.18	imR19, imS6b1, imS6b2, imS6b3, imS6b4, imL14b2, imL14b3, imL14b4, imL14b5, imL14b7	70.46
11	imR19, datLCH, datLMH, datL70, imS6b1, imS6b4, imL14b2, imL14b3, imL14b4, imL14b5, imL14b7	60.39	imR19, imS6b1, imS6b2, imS6b3, imS6b4, imL14b1, imL14b2, imL14b3, imL14b4, imL14b5, imL14b7	70.73
12	imR19, datLMH, datL70, imS6b1, imS6b2, imS6b3, imS6b4, imL14b2, imL14b3, imL14b4, imL14b5, imL14b7	60.50		
13	imR19, datLCH, datLMH, datL70, imS6b1, imS6b2, imS6b3, imS6b4, imL14b2, imL14b3, imL14b4, imL14b5, imL14b7	60.65		
>14	All features	61.16		

ASSESSMENT OF DIFFERENT ESTIMATION ALGORITHMS AND REMOTE SENSING /.../

M. LANG ET AL.



Figure 5. The lack of fit of the estimated standing wood volume in case of a) k-NN and b) GRNN

Table 4. Comparison of wood volume estimation methods: the k-NN, RT and GRNN

Droporty or statistic	Algorithm			
Property of statistic	k-NN RT		GRNN	
Internal parameters	Number of nearest neighbours k=21	LOOC based best pruning level: 125 Number of responses: 5	Spread of radial basis function: 2	
Best RMSE achieved for initial test set	79.8	77.7	77.9	
RMSE for initial training set	83.8* (LOOC)	74.3	62.3	
RMSE variability (100 training and test sets)	max(RMSE)=83 min(RMSE)=80 std(RMSE)=0.9 mean(RMSE)=81.4	max(RMSE)=80 min(RMSE)=77 std(RMSE)=1.1 mean(RMSE)=79.2	max(RMSE)=81 min(RMSE)=76 std(RMSE)=0.9 mean(RMSE)=78.3	
Computational time for test set	estimation 6.60 s	training 0.08 s, estimation 0.0085 s	training 0.048 s, estimation 0.4267 s	

\* k-NN leave-one-out cross-validation (LOOC) statistics are not fully equivalent to the training set evaluation for the RT and GRNN.

#### **Case study 6: Comparison of estimators**

The results of the estimator comparison are shown in Table 4 and the lack of fit for the k-NN and GRNN is shown in Figure 5. In the presence of high variance and nonlinear asymptotically saturating relationships between wood volume and spectral feature variables, all methods performed similarly with differences in accuracy of approximately 2% (for the k-NN and GRNN). The computational time can be reduced by choosing the GRNN or RT instead of the k-NN. Since the computational time depends on implementation and hardware properties, it is included only as additional information. However, when an estimator has to be optimized using e.g. genetic algorithms or bootstrapping for a large reference set the computational speed is important.

#### **Discussion and Conclusions**

There are few studies that incorporate stand-wise forest management inventory data for a regional wallto-wall estimation of forest structure variables using 2000, Mäkela and Pekkarinen 2004, Maltamo et al. 2006, McRoberts 2008, Tamm and Remm 2009, Lang et al. 2014). In fact, sample plot data from the strategic National Forest Inventory (NFI) are preferred due to instrumental measurements and high accuracy of the estimated plot-level data. The RFI data, on the other hand, can be have systematic errors, may have large random errors and the records are updated at an interval of ten or more years. The update interval of NFI data is usually five years (Tomppo et. al. 2010). However, the advantages of RFI data over NFI data include much better spatial coverage and an option to have more stable signatures when averaging the pixel values under the forest stand polygon. An NFI sample plot with a radius of 7 m to 10 m covers approximately one or less than one pixel in a medium spatial resolution satellite image. Errors occur in the NFI plot location coordinates and there is an inevitable up to half pixel location error in image data due to the gridding process (Liang 2004) during image construction. Hence, a lot of randomness exists in the relationships between

k-NN or machine learning algorithms (Holmgren et al.

ISSN 2029-9230

#### ASSESSMENT OF DIFFERENT ESTIMATION ALGORITHMS AND REMOTE SENSING /.../ M. LANG ET AL.

NFI variables and remote sensing variables. On the other hand, the estimates of feature variables calculated using RFI stand polygons can have large variance for several reasons and outliers have to be removed before using the data with k-NN. We used the relationships between spectral radiance and stand age or wood volume to clean the reference set and validation set for outliers. The procedure was successful as indicated by the RMSE of estimates which was 36...40% of the measured mean value when using the reference variables extracted from the Landsat images. The results are well comparable to Holmgren et al. (2000) or Mäkela and Pekkarinen (2004) who carried experiments out in much simpler boreal forests and in a smaller study area.

In forest management the planning of thinnings and other operations requires stand-level data and average statistics for a larger area i.e. a region are not so much of interest. When k-NN based wood volume estimates are aggregated over a larger area (e.g. 100 ha) then the mean values have a smaller relative error (Fazakas et al 1999), but for an individual stand the estimation error may be large. As found earlier by other authors (Fazakas et al 1999, Holmgren et al. 2000) there was a characteristic lack of fit in the k-NN estimated wood volume values also in our results, expressed by the overestimation of small values, almost correct estimates for the stands with the mean Vand underestimation for the stands which had a larger than mean volume of wood. Gilichinsky et al. (2012) propose a procedure based on the histogram matching of k-NN wood volume estimates with regional statistics, however, it is not known how such general adjustments are applicable, if e.g. species composition is to be estimated at the same time (Lang et al. 2014). The reasons for the lack of fit are probably related to nonlinearities between response and feature variables, noise in the reference data set and feature variable redundancy. The feature variable redundancy effect on the wood volume estimates was found also in our feature variable selection experiment in the special study site in Slitere National Park. Different optimization routines are proposed to select informative feature variables and sample plots (McRoberts et al. 2015) and our experiments confirmed that the k-NN or GRNN implementations without optimization are sensitive to the nonlinear relationship, noise and the variable redundancy. The ALS data-based feature variables were superior to the spectral data and radar data for wood volume estimation. Although the ALS data provided highest accuracy, the employment of multispectral satellite images is still of interest to ensure regional coverage with a short update interval at lower financial expenses since free public access was provided to medium spatial resolution Landsat and Sentinel data.

When sample plots are used to create a reference dataset for the k-NN, higher spatial resolution of feature images is preferred. However, the instantaneous field of view of scanners is inversely related to the swath of the image and individual high spatial resolution images cover a smaller area and more images are required to cover a certain region. This decreases the number of reference observations per image and requires the use of images from different growing seasons, view nadir angles, atmospheric conditions or sensors. We tested the impact of the reference set randomness effect and the combined effect of the reference set and feature variables extracted from different images on the k-NN estimated wood volume. Different samples of reference observations may be available for multitemporal multispectral satellite images due to clouds and cloud shadows or sensor malfunctioning. Our case studies indicated that the estimation accuracy depends on a particular image, but the randomness in the reference set does not impact accuracy substantially. While the reference observation randomness caused a marginal 4 m<sup>3</sup>/ ha standard error of the estimated values, the standard error was up to 30% when viewed relative to the measured wood volumes particularly in younger stands. So, the randomness of the reference sample is not a major issue if the number of reference observations is large. However, even a large number of reference observations does not remove the lack of fit of estimates occurring near to the minimum and maximum of the measured wood volume.

The impact of the image acquisition season was detected in the Landsat image set where higher accuracy was achieved with images from early summer when the scene is better illuminated. The estimations based on the SPOT HRVIR images were highly dependent on the specific image. Our hypothesis that images from late winter when the ground is covered with snow are useful for wood volume estimation in Kurzeme region was partially proven. The estimates had a bigger RMSE compared to the midgrowing season Landsat images. The reasons may be related to the 8-bit radiometric resolution of SPOT HRVIR data and possible contamination of the winter images by cirrus clouds which are hard to detect and possible snow on branches. However, new sensors e.g. the Operational Land Imager on-board Landsat-8 have special bands for detecting cirrus clouds and atmospheric haze.

We tested different k-NN implementations and carried out also tests with the GRNN and RT. All the methods performed similarly due to complex relationships between forest inventory data and remote sensing data. We found that the computational time can be reduced by using the GRNN, however, future work is required to compare different implementations of the k-NN, options for optimization to decrease the lack of fit, and to investigate

ASSESSMENT OF DIFFERENT ESTIMATION ALGORITHMS AND REMOTE SENSING /.../

the impact of the algorithm structure on the estimation of forest structure variables.

Finally, it can be concluded that multispectral satellite data similar to Landsat-5 TM or Landsat-7 ETM+ images can be well used for regular regional wood volume estimations. Even with the relatively high RMSE these estimates complete the conventional forest management inventory with additional information.

#### Acknowledgements

The study was carried out for the Central Baltic Interreg IV A programme 2007–2013 project No. CB56. Data analysis and manuscript writing was partially supported by the Estonian Ministry of Education and Research grant IUT21-4. ALS data and forest inventory data were released by the JSC "Latvijas valsts meži", Forest State Service and the foundation "Institute for Environmental Solutions". Landsat images were made available for public use by the USGS. We also thank two anonymous reviewers for their helpful comments.

#### References

- Avotniece, Z., Klavins, M. and Lizuma, L. 2015. Fog climatology in Latvia. Theoretical and Applied Climatology 122:97-109.
- Bušs, M. 1987. Latvijas meži [Forests of Latvia]. Avots, (in Latvian).
- Cartus, O., Santoro, M., Schmullius, C. and Li, Z. 2011. Large area forest stem volume mapping in the boreal zone using synergy of ERS-1/2 tandem coherence and MODIS vegetation continuous fields. Remote Sensing of Environment 115: 931-943
- Duda, R.O., Hart, P.E. and Stork, D.G. 2012. Pattern classification. John Wiley & Sons.
- Fazakas, Z., Nilsson, M. and Olsson, H. 1999. Regional forest biomass and wood volume estimation using satellite data and ancillary data. Agricultural and Forest Meteorology 98/99: 417-425.
- Forest. 2000. Forest Law, accepted 24.02.2000, in effect from 17.03.2000. Latvijas Vēstnesis 98/99 (2009/2010), 16.03.2000.
- 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: 251-274.
- Gilichinsky, M., Heiskanen, J., Barth, A., Wallerman, J., Egberth, M. and Nilsson, M. 2012. Histogram matching for the calibration of kNN stem volume estimates. International Journal of Remote Sensing 33: 7117-7131.
- Gjertsen, A.K. 2007. Accuracy of forest mapping based on Landsat TM data and a kNN-based method. Remote Sensing of Environment 110: 420-430.
- Holmgren, J., Joyce, S., Nilsson, M. and Olsson, H. 2000. Estimating stem volume and basal area in forest compartments by combining satellite image data with field data. Scandinavian Journal of Forest Research 15: 103-111.
- Holmström, H. and Fransson, J.E. 2003. Combining remotely sensed optical and radar data in kNN-estimation of forest variables. Forest Science 49: 409-418.
- Hyyppä, J., Hyyppä, H., Inkinen, M., Engdahl, M., Linko, S. and Zhu, Y.H. 2000. Accuracy comparison of various remote

sensing data sources in the retrieval of forest stand attributes. Forest Ecology and Management 128: 109-120.

- Koch, B. 2010. Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment. ISPRS Journal of Photogrammetry and Remote Sensing 65: 581-590.
- Koukal, T., Suppan, F. and Schneider, W. 2007. The impact of relative radiometric calibration on the accuracy of kNN-predictions of forest attributes. Remote Sensing of Environment 110.431-437
- Lang, M., Arumäe, T., Lükk, T. and Sims, A. 2014. Estimation of standing wood volume and species composition in managed nemoral multi-layer mixed forests by using nearest neighbour classifier, multispectral satellite images and airborne lidar data. Forestry Studies | Metsanduslikud Uurimused 61: 47-68.
- Leboeuf, A., Beaudoin, A., Fournier, R.A., Guindon, L., Luther, J.E., and Lambert, M.C. 2007. A shadow fraction method for mapping biomass of northern boreal black spruce forests using QuickBird imagery. Remote Sensing of Environment 110: 488-500.
- Li, Y., Andersen, H.E. and McGaughey, R. 2008. A comparison of statistical methods for estimating forest biomass from light detection and ranging data. Western Journal of Applied Forestry 23: 223-231.
- Liang, S. 2004. Quantitative remote sensing of land surfaces. Hoboken, New Jersey. John Wiley & Sons, Inc.
- Lu, D. 2006. The potential and challenge of remote sensing based biomass estimation. International journal of remote sensing 27: 1297-1328.
- Magnusson, M., Fransson, J. E., Eriksson, L.E., Sandberg, G., Smith-Jonforsen, G. and Ulander, L.M. 2007. Estimation of forest stem volume using ALOS PALSAR satellite images. In Geoscience and Remote Sensing Symposium. IGARSS 2007. IEEE International, 2007: 4343-4346.
- Maltamo, M., Eerikäinen, K., Pitkänen, J., Hyyppä, J. and Vehmas, M. 2004. Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions. Remote Sensing of Environment 90: 319-330.
- Maltamo, M., Malinen, J., Packalén, P., Suvanto, A. and Kangas, J. 2006. Nonparametric estimation of stem volume using airborne laser scanning, aerial photography, and stand-register data. Canadian Journal of Forest Research 36: 426-436.
- McGaughey, R.J. 2014. FUSION/LDV: Software for LIDAR Data Analysis and visualization. March 2014 - FUSION, Version 3.42. United States Department of Agriculture Forest Service Pacific Northwest Research Station.
- McRoberts, R.E. 2008. Using satellite imagery and the k-nearest neighbors technique as a bridge between strategic and management forest inventories. Remote Sensing of Environment 112: 2212-2221.
- McRoberts, R.E. 2012. Estimating forest attribute parameters for small areas using nearest neighbour techniques. Forest Ecology and Management 272: 3-12.
- McRoberts, R.E., and Tomppo, E.O. 2007. Remote sensing support for national forest inventories. Remote Sensing of Environment 110: 412-419.
- McRoberts, R.E., Næsset, E. and Gobakken, T. 2015. Optimizing the k-Nearest Neighbors technique for estimating forest aboveground biomass using airborne laser scanning data. Remote Sensing of Environment 163: 13-22.
- Moisen, G.G. and Frescino, T.S. 2002. Comparing five modelling techniques for predicting forest characteristics. Ecological modelling 157: 209-225.
- Mäkelä, H. and Pekkarinen, A. 2004. Estimation of forest stand volumes by Landsat TM imagery and stand-level field-inventory data. Forest Ecology and Management 196: 245-255.

ISSN 2029-9230

ASSESSMENT OF DIFFERENT ESTIMATION ALGORITHMS AND REMOTE SENSING /.../ M. LANG ET AL.

- Nelson, R.F., Hyde, P., Johnson, P., Emessiene, B., Imhoff, M.L., Campbell, R. and Edwards, W. 2007. Investigating Ra-DAR-LiDAR synergy in a North Carolina pine forest. Remote Sensing of Environment 110: 98-108.
- Nilson, T. and Peterson, U. 1994. Age dependence of forest reflectance: Analysis of main driving factors. Remote Sensing of Environment 48: 319-331.
- Olsson, H. 1993. Regression functions for multitemporal relative calibration of Thematic Mapper data over boreal forest. Remote Sensing of Environment 46: 89-102.
- Peterson, U., Püssa, K. and Liira, J. 2004. Issues related to delineation of forest boundaries on Landsat Thematic Mapper winter images. International Journal of Remote Sensing 25: 5617-5628.
- Popescu, S.C., Wynne, R.H. and Scrivani, J.A. 2004. Fusion of small-footprint lidar and multispectral data to estimate plotlevel volume and biomass in deciduous and pine forests in Virginia, USA. Forest Science 50: 551-565.
- Rutkis, J. 1960. Latvijas ģeogrāfija [Geography of Latvia]. Stockholm, Apgads Zemgale, p. 794 (in Latvian).
- Saliņš, Z. 2002. Mežs Latvijas nacionālā bagātība [Forest national treasure of Latvia]. Jelgava, p. 248 (in Latvian).
- Specht, D.F. 1991. A general regression neural network. IEEE Transactions on Neural Networks 2: 568-576.
- Tamm, T. and Remm, K. 2009. Estimating the parameters of forest inventory using machine learning and the reduction of remote

sensing features. International Journal of Applied Earth Observation and Geoinformation 11: 290-297.

- Tomppo, E. 1991. Satellite image-based national forest inventory of Finland. International Archives of Photogrammetry and Remote Sensing 28: 419-424.
- Tomppo, E., Gagliano, C., De Natale, F., Katila, M. and McRoberts, R.E. 2009. Predicting categorical forest variables using an improved k-Nearest Neighbour estimator and Landsat imagery. Remote Sensing of Environment 113: 500-517.
- Tomppo, E., Gschwantner, Th., Lawrence, M. and McRoberts, R.E. 2010. National Forest Inventories. Pathways for Common Reporting. Springer, p. 612.
- Wulder, M.A. and Franklin, S.E. 2003. Remote sensing of forest environments: concepts and case studies. Springer.
- Zhao, F., Guo, Q. and Kelly, M. 2012. Allometric equation choice impacts lidar-based forest biomass estimates: A case study from the Sierra National Forest, CA. Agricultural and Forest Meteorology 165: 64-72.
- Znotina, V. 2002. Latvijas mežu raksturojums [Characterization] of forests in Latvia]. http://latvijas.daba.lv/biotopi/mezi.shtmlv151, (in Latvian).

Received 22 January 2016 Accepted 30 September 2016