

Research paper

Remote-sensing support for the Estonian National Forest Inventory, facilitating the construction of maps for forest height, standing-wood volume, and tree species composition

Mait Lang^{1,2}, Allan Sims³, Kalev Pärna⁴, Raul Kangro⁴, Märt Möls⁴, Marta Mõistus¹, Andres Kiviste², Mati Tee⁵, Toivo Vajakas⁵ and Mattias Rennel⁵

Lang, M., Sims, A., Pärna, K., Kangro, R., Möls, M., Mõistus, M., Kiviste, A., Tee, M., Vajakas, T., Rennel, M. 2020. Remote-sensing support for the Estonian National Forest Inventory, facilitating the construction of maps for forest height, standing-wood volume, and tree species composition. – *Forestry Studies | Metsanduslikud Uurimused* 73, 77–97, ISSN 1406-9954. Journal homepage: <http://mi.emu.ee/forestry.studies>

Abstract. Since 1999, Estonia has conducted the National Forest Inventory (NFI) on the basis of sample plots. This paper presents a new module, incorporating remote-sensing feature variables from airborne laser scanning (ALS) and from multispectral satellite images, for the construction of maps of forest height, standing-wood volume, and tree species composition for the entire country. The models for sparse ALS point clouds yield coefficients of determination of 89.5–94.8% for stand height and 84.2–91.7% for wood volume. For the tree species prediction, the models yield Cohen’s kappa values (taking 95% confidence intervals) of 0.69–0.72 upon comparing model results against a previous map, and values of 0.51–0.54 upon comparing model results against NFI sample plots. This paper additionally examines the influence of foliage phenology on the predictions and discusses options for further enhancement of the system.

Key words: National Forest Inventory, lidar, multispectral satellite images, sample plots.

Author’s addresses: ¹Tartu Observatory, University of Tartu, 61602 Tõravere, Tartumaa, Estonia; ²Institute of Forestry and Rural Engineering, Estonian University of Life Sciences, Kreutzwaldi 5, 51014 Tartu, Estonia; ³Forest Department, Estonian Environment Agency, Mustamäe tee 33, 10616 Tallinn, Estonia; ⁴Institute of Mathematics and Statistics, University of Tartu, Narva mnt 18, 51009 Tartu, Estonia; ⁵Reach-U, Riia 24, 51010 Tartu, Estonia; *e-mail: lang@to.ee

Introduction

Over past centuries, land use in Estonia has had a substantial impact on land cover. While the share of forest land has been as small as 20–25%, it has gradually increased as a consequence of land drainage, a de-

crease in the number of households managing small farms, and the abandonment of arable land (Kremser, 1998; Etverk, 2003; Jõgiste *et al.*, 2018). Resource assessment is a prerequisite for decision-making and planning. To facilitate assessment, field survey and monitoring results are stored

DOI: 10.2478/fsmu-2020-0016



© 2020 by the authors. Licensee Estonian University of Life Sciences, Tartu, Estonia. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

as data tables and maps, in formats supporting numeric and spatial queries and analyses. Up-to-date data on land cover and land use form the basis for policy implementation in forestry, agriculture, and rural development.

Forest inventory provides the woodland owner with data required for forest management planning. Forest land is divided into patches based on the properties of soil and plant cover, enabling the forest stand on each generated unit to be characterized by mean values of inventory variables (tree species composition, forest height, wood volume, age of stand elements, etc.) with a sufficiently small variability within each unit (Krigul, 1972; Vaus, 2005). Such forest management inventories (FMIs) have been conducted in Estonia over the past 150 years, with a typical FMI repetition cycle of 10 years (Meikar, 1998). Forest land-use summaries have been published on the basis of FMI data (Aru & Okas, 1959; Aru *et al.*, 1975; Tappo, 1982; Polli & Viilup, 1989; Viilup, 1995).

The FMI would give a complete overview of Estonian forests if all data were accessible. Natural though full accessibility is in the case of state-owned forests, however, it is not feasible for forests in private ownership. The records from FMI databases are updated only when the forest owner considers it necessary. To overcome these problems, many countries have a second type of forest inventory, a National Forest Inventory (NFI) based on a regularly spaced sample of field plots 7–15 m in radius, instrumentally measured (NFIEUROPE, 1997; Tomppo *et al.*, 2010). The first known Estonian forest-inventory tests covering a substantial area and based on sample plots were conducted by Prof. A. Nilson in the Järvselja Experimental and Training Forest District at the beginning of the 1970s. P. Kohava from the Estonian Forest Survey Centre tested an NFI design on the Estonian island of Hiiumaa in 1997 (Kohava, 1998). An independent test at the end of the 1990s was carried out by

Metsaekspert OÜ (Metsainventuur, 2000a; 2000b). The basic principles of the current Estonian NFI design have been published by Adermann (2010). The latest complete and exhaustive FMI- and NFI-based overview of forests in Estonia can be found in the yearbook series *Forest* (Forest, 2020; first issued as Forest, 1995). However, the statistics are not presented at a finer level of detail than the individual county, because there are insufficient field observations for smaller land units, such as the individual municipality.

Forest land surveys have used remote sensing since 1920, when half a million acres of forest land was inventoried in Canada on the basis of preliminary aerial photography (Howard, 1991). Aerial photography has been used for FMI in Estonia since 1961 when monochrome aerial photos were taken on emulsions sensitive to the visible and near-infrared part of the spectrum over Järvselja forests in southeast Estonia, as a basis for the construction of 1:10,000 forest-stand maps.

The automated processing of remote-sensing data to support the construction of maps of forest-inventory variables was proposed for the particular case of Finland by Poso *et al.* (1990). In many countries, the task of combining NFI sample plots with feature variables from remote sensing has been addressed either with the k -nearest neighbour algorithm (k -NN) (Fazakas *et al.*, 1999; Franco-Lopez *et al.*, 2001) or with the k -MSN algorithm (Packalén & Maltamo, 2007). Tamm & Remm (2009) tested an application of machine learning to the estimation of standing-wood volume for forest stands in northeastern Estonia, using multispectral satellite images. Peterson & Aunap (1998) showed that multitemporal Landsat MSS images can be used in Estonia to map changes of agricultural land use. Lang *et al.* (2014) used k -NN to predict wood volume and tree species composition in Estonia's Laeva test site, on the basis of data from the Landsat-8 Operational Land Imager (OLI)

and from airborne laser scanning (ALS). In many countries, airborne lidar data have been used extensively during the last decade for the estimation of forest variables, because the three-dimensional point clouds contain information that strongly correlates with forest height and wood volume (Næsset, 1997; Næsset *et al.*, 2004; Næsset, 2005; Lang *et al.*, 2012; Kotivuori *et al.*, 2016).

Although many countries have already incorporated multispectral satellite images and airborne laser scanning data into their NFI systems to construct maps of forest variables (McRoberts & Tomppo, 2007; Barrett *et al.*, 2016), and even though Estonia completed its first nationwide sampling campaign as early as 1999 (Kohava, 2000), such a module has until now been missing from the Estonian NFI. Through a synthesis of previous research, the present study offers a module at a high level of technical readiness, enabling the Estonian NFI to incorporate remote-sensing data into its construction of maps of principal forest-inventory variables (forest height, wood volume, and tree species composition). We present data processing principles, describe models, and report test results for our remote-sensing support module. Additionally, we discuss basic principles for constructing a map of errors, and propose methods for the annual updating, in the light of both forest growth and disturbances, of the predicted forest height and standing-wood volume maps.

Material and Methods

Forests in Estonia

The main forest-forming tree species in Estonia are European aspen (*Populus tremula* L.), silver birch (*Betula pendula* Roth), downy birch (*B. pubescens* Ehrh.), black alder (*Alnus glutinosa* (L.) Gaertn.), grey alder (*A. incana* (L.) Moench), Norway spruce (*Picea abies* (L.) Karst.), Scots pine (*Pinus sylvestris* L.), and common ash (*Fraxinus*

excelsior L.). Depending on soil fertility and forest management, these species occur either in a variety of mixtures or as pure stands. Norway spruce forms a mid-storey in many stands growing on fertile soils. Forest land (23,308 km²) accounts for more than half (53.6%) of the Estonian terrestrial territory (Forest, 2020). About 47% of the forest land is owned by the state. The rest belongs largely to private owners, with a small share owned by municipalities and other public institutions. According to the forest register, more than 31% of forest land is drained to improve tree growth conditions. On 24.6% of Estonian forest land, use is restricted under a variety of nature-protection regulations (Forest, 2020). Wood is harvested through thinning and through shelter-wood cuttings, and also through the creation of small clearfellings (to a maximum area of 7 ha) with retention trees (Forestman, 2007; Forestact, 2016). More than 50% of the forest land use is classified as retention forestry (Gustafsson *et al.*, 2012). Since the year 2000, the annual regeneration fellings and annual maintenance fellings have been in the respective area ranges of 17.1–38.8 thousand ha and 24.6–42.9 thousand ha (Forest, 2020). The average size of a forest stand is 1.25 ha. Forests are seminatural, with all the above-mentioned tree species regenerating naturally, and with Norway spruce and Scots pine occurring both under cultivation and with natural regeneration. The share of forest land has been increasing since the end of World War II, due to the abandonment of agricultural land. Leading the process of abandonment are those land parcels which are smaller and more remote from roads (Mandel *et al.*, 2019).

Sampling design of the Estonian NFI

The first Estonian-wide NFI with a sample plot layout similar to the current NFI was launched in 1999 (Kohava, 2000). With a rather modest budget and equipment, the NFI was able to give a fairly accurate assessment of the forest area, resources, and

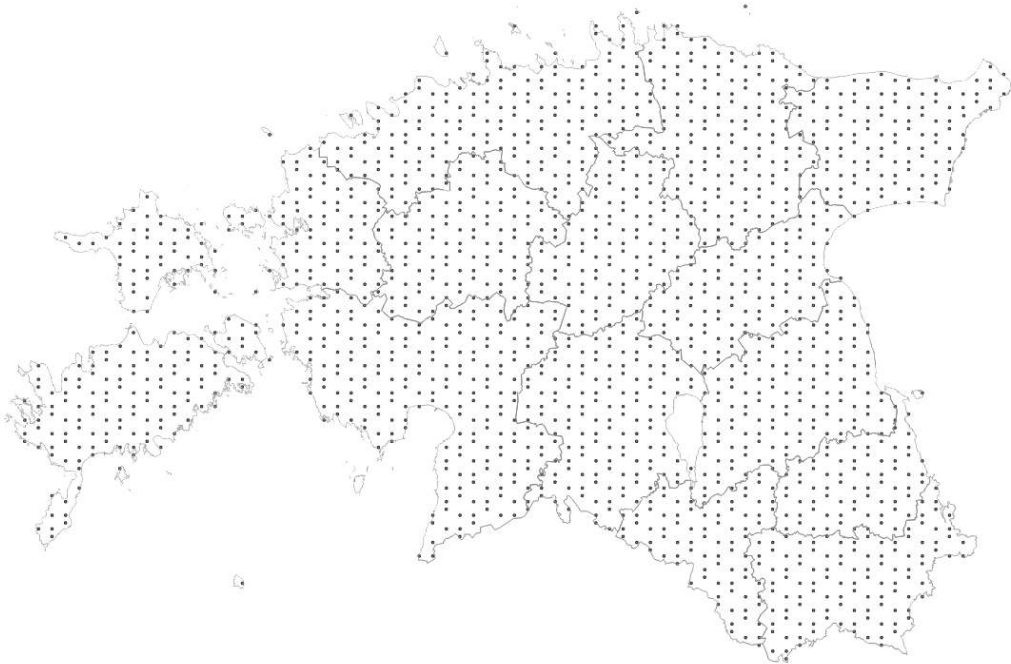


Figure 1. Cluster network of the Estonian NFI (2014–2018).

Joonis 1. Statistilise metsainventuuri proovitükkide klastrite paiknemine (2014–2018).



Figure 2. Number of measured permanent (P), temporary (T), and new permanent (NP) sample plots per year.

Joonis 2. Aastas mõõdetav alaliste (P), ajutiste (T) ja uute alaliste (NP) proovitükkide arv.

cutting volume. The main initial objective of the NFI was the estimation of major characteristics of forest stands. At present, however, the NFI has a wider scope, reporting also such data as the distribution of land by land-use categories and the afforestation and growing stock of non-forest land. The smallest target unit for reporting is the individual county.

Methodologically, the NFI is designed as an annual research effort, which, using optimal methods, must ensure continuous updating of its assembled information, including the forest database. Since 2014, sampling has been conducted with increased frequency, on a network of sample plots covering the entire country (Figure 1). Under this procedure, in each given year approximately 375 clusters (i.e. 20%) from the entire ensemble of sample plots, is measured (Figure 2). This ensures re-measurement of permanent plots once in every 5 years. Point estimates of parameters are calculated with data from the sample plots, as a basis for inferences to the entire population.

The Estonian NFI is designed as a sys-

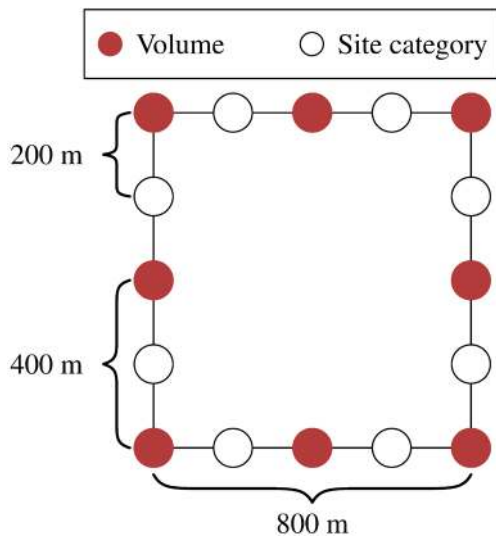


Figure 3. Estonian NFI cluster design and sample plot arrangement.

Joonis 3. Eesti statistilise metsainventuuri proovitükkide paiknemine klastris.

tematic sample without pre-stratification. The sampling grid is established to meet accuracy requirements at the national level. About 5,500 sample plots are measured per year, with the sampling intensity the same throughout the country. The sample (cluster) distribution is based on the national 5×5 km quadrangle grid, determined by the Estonian base-map coordinates system (EPSG:3301).

The observation unit is the individual field plot, determined by its centre coordinates. The method of sampling with partial replacement is used. The sample plot area is subdivided if its area overlaps with different types of forest stands or parcels with different land cover. To enhance the efficiency of the survey, sample plots are concentrated into clusters, defined as 800×800 m squares (Figure 3). Some sample-plot clusters are deemed permanent, others temporary. The radius of a given sample plot depends on the variables selected for assessment, and additionally on variable values (with, e.g., smaller sample plots used for lower or regrowth layers). The radii of the principal sample plots are 10 m and 7 m. For land-use category determination, plots of other radii are taken, although without modification in the scheme for selecting centre coordinates. For the construction of models for stand height and wood volume, we used the sample plots in which a tree layer was present and which were located entirely inside a forest stand.

Remote-sensing data and other spatial data

The remote-sensing support system of the Estonian NFI is based on data from the European Union Copernicus programme, from the NASA/USGS Landsat programme, and from the airborne photography and laser scanning programme of the Estonian Land Board. The medium spatial resolution (10–30 m) multispectral images from Sentinel-2 MSI (ESA, 2015) and Landsat-8 OLI (USGS, 2019) sensors and SAR

images from Sentinel-1 (ESA, 2020) are used for the prediction of tree species composition and for the detection of changes. The data are available from the regional data centre ESTHub (2016).

Aerial photography and airborne laser scanning are conducted by the Estonian Land Board under a repetition schema that provides, for any given location, either summer or springtime data in each second year, and additionally produces measurements from each similar growth season after every four years (Figure 4) (Maa-amet, 2020). The point density of the archived ALS data ranges from 0.15 m^{-2} to 2 m^{-2} . The ALS pulse footprint diameter at canopy level is about 0.5 m, with a scanning angle that does not exceed 30° from nadir. Data are distributed according to a 1 km^2 map sheet system (Maa-amet, 2019). Before the year 2017, the Estonian Land Board used the Leica ALS50-II airborne scanning system. From 2017 onward, measurements have instead been taken with the Riegl VQ-1560i. In the Estonian NFI, orthopho-

tos (in the RGB + NIR bands) are currently used only for visual interpretation, in the estimation of land cover type during the preparation of fieldwork. In the remote-sensing module offered in this paper for the Estonian NFI, ALS data are used for the prediction of forest height and wood volume. Ancillary data sources for our offered module are a 1:10,000 base map and soil map of Estonia, the FMI database with its stand-level forest-management inventory data, and a digital terrain module (DTM) provided by the Estonian Land Board.

Data processing and models

The first two variables to be predicted are basal-area-weighted forest height (Lorey's mean height) H and wood volume M . The prediction of H and M is based on ALS data. The data processing system is designed in such a way that the user can develop and apply any models that take as their inputs ALS point-cloud metrics in combination with ancillary variables. However, for Estonian forests a simple linear regression model can be used with sufficient predictive power (Lang *et al.*, 2012) for forest height:

$$H_{ALS} = a + b \cdot H_{Px} + \epsilon. \quad (1)$$

Here a and b are model parameters, H_{Px} is one of the upper percentiles of the point-cloud height distribution, and ϵ is the model residual error. In the calculation of point-cloud height distribution percentiles, points near the ground are excluded. Wood volume M can be predicted with a model that is based on a relationship involving basal area G , forest height H , and stand form factor F ($M = G \cdot H \cdot F$):

$$M_{ALS} = (a \cdot H_{P80}^b + c \cdot H_{P25}) \cdot K^d + \epsilon, \quad (2)$$

where a , b , c , and d are the model parameters; H_{P80} is the 80th percentile; H_{P25} is the lower quartile of the point-cloud height distribution; and $K = C_{above}/C$ is a proxy for canopy cover calculated using the ALS



Figure 4. Location of target areas for fitting forest height and standing-wood volume models, corresponding to the Estonian Land Board ALS measurement schema. The areas are partially overlapping.

Joonis 4. Maa-ameti aerofotomöödistuse objektid, mille piires lähendatakse puistu kõrguse ja tüvemahu prognoosimudelid. Alad on mõningase ülekattega.

pulse return count C_{above} above a height threshold (1–2 m from the ground) and total count of returns C per target pixel or sample plot (Lang *et al.*, 2012). In calculating the height percentiles of point clouds, returns from near the ground are excluded, and K is calculated using all returns recorded for each emitted pulse. The FUSION toolbox (McGaughey, 2018) is used for ALS data processing in this study, with ALS metrics calculated for 10 m target pixels. Point clouds for sample plots were extracted using a circle with a radius of 15 m.

The parameters for height and wood volume models are estimated using NFI sample-plot field measurement data for each particular flight campaign, with each campaign covering roughly one-quarter of Estonia's area (Figure 4). No correction is currently made for springtime phenology and the consequent increased gap fraction in the canopy, and tree species are not currently distinguished for H and M models. The current species-independent modelling also ignores differences in the interaction of laser pulse with tree canopy which may arise from variations in scanner settings and from variations in the summertime structure of the canopy. The species-independent model was selected because it was considered that the number of field measurements would become insufficient if the set of field measurements were partitioned into subsets according to dominant species. In the current study, the wood volume corresponds to the upper canopy layer.

Prediction of tree species composition for target units with a size of 10–30 m is a challenging task, since most forests in Estonia are of mixed composition, since the spectral signatures of broadleaf deciduous trees are quite similar to each other, and since variables that are related to stand age and structure have a substantial influence on forest spectral signature (Nilson & Peterson, 1994) as calculated from Landsat-8 OLI and Sentinel-2 MSI images. The proposed solution for the Estonian NFI is

based on machine learning, using a large number of samples from the FMI database and data from a 1:10,000 soil map. Lang *et al.* (2018) found the *random forest* algorithm implementation in the GRASS GIS to be suitable for the construction of tree-species composition maps.

The machine-learning procedure starts with the selection of a potential training set of stands from the FMI database. The next step after the cloud masking of image data is the removal of outliers, based on relationships of spectral radiance with forest age and wood volume in red, near infrared, and shortwave infrared bands (Lang *et al.*, 2016). Additional pairwise comparison of images is applied to remove stands with detectable disturbances that may have occurred during the period covered by the selection of satellite images. Experience has shown that this winnowing procedure yields a final set of about 100,000 samples, appropriate for machine learning for the whole of Estonia. The target area is divided into 8 overlapping subareas, to account for regional characteristics of forests in Estonia and to exclude training samples which have been measured over a long distance (Lang *et al.*, 2018). Predictions for each subarea are made using either single images or combinations of images that cover more than half of the subarea, with a maximum of 3 images allowed in any one combination.

The model fitting on empirical data starts with the selection of informative features. The next step is model hyper-parameter optimization. Finally, a prediction map is constructed. Although the variable being predicted is the dominant species code, additional useful information is obtained from the entire vector of class probabilities (Lang *et al.*, 2018) for each target pixel. Finally, the predicted class probabilities (where each class corresponds either to one of the possible dominant forest species or to a classification as non-forest) are averaged for each pixel, and dominant species information is extracted according to the estimated proportions.

For the construction of a tree species composition map, we used Landsat-8 OLI and Sentinel-2 images from April 2018 to September 2018. The images were resampled to 25 m spatial resolution and converted to the EPSG:3301 coordinate system. The methods for image processing and data analysis resemble those recently described by Lang *et al.* (2018).

Software and implementation

The data processing system is implemented with existing software, but with the addition of a calling script written in Python. Each task (e.g. the prediction of forest height, of wood volume, or of tree species composition; or the detection of changes; or the appraisal of predictions) is structured as a sequence of specific steps (e.g. a preparation phase, the extraction of features from remote sensing data, the fitting of a model, the application of the selected model, and validation). The user edits the script template of a specific step, inserts proper parameter values, and executes the script. Each processing execution is assigned a unique ID, which is then used in a file-storage folder structure. The scripts produce an execution history log as a directed acyclic graph (DAG). Version control, the generated data-folder structure, and the generated DAG jointly ensure traceability and repeatability of the data-processing steps. In considering alternatives to this scheme, we evaluated some formal workflow tools (Airflow, Luigi, Jenkins), but found that they failed to add sufficient value. The principal software tools used in our scheme are QGIS, GRASS GIS, Python, GDAL, ESA SNAP, the Orfeo Toolbox (OTB), LAsTools, FUSION, R, RStudio, and Git.

During execution, all intermediate results are stored in the local computer. In addition, the intermediate and final results are copied to a backup file server. Intermediate results are stored for 2 years (this being the longest interval between successive measurement updates). Final results are

stored for some longer period, as decided by the system administrator. The hardware configuration comprises an 8-core workstation CPU, 24 GB of RAM, and 6 TB of workstation disk storage, with additionally 2.3 TB of backup storage. All hardware is virtualized.

Model statistics and assessment of predicted values

For characterization of forest height and wood volume models, the mean residual squared error

$$RSE = \sqrt{\frac{\sum(\hat{Y} - Y)^2}{N}} \quad (3)$$

and mean residual error

$$MRE = \frac{\sum(\hat{Y} - Y)}{N} \quad (4)$$

were calculated, with N the number of observations, \hat{Y} the predicted value for each sample plot in the sum, and Y the corresponding measured value.

For the assessment of forest height and wood volume predictions, we used a set of forest stands from the overlap area of the NE and NW blocks, and additionally a set of forest stands from the overlap area of the SE and SW blocks (Figure 4). We applied the following selection criteria: the stand polygon area was required to lie in the range of 1.2–8.0 ha; the count of 10 m pixels within the stand polygon was required to be > 100 ; the number of pixels within the stand polygon without value (i.e. with NoData label) was required to be < 10 in the corresponding comparison maps; and the proportion of evergreen tree species in the stand was required to be either $\geq 75\%$ or $\leq 25\%$. The last of these filters (the disjunctive criterion regarding the proportion of evergreens) was applied to select contrasting sets of stands according to tree species. On the overlap area of NW

and NE blocks were 1,157 evergreen and 605 deciduous stands. On the overlap area of SW and SE blocks were 644 evergreen and 1,275 deciduous stands. For each stand, the mean value of pixels was calculated from forest-height and wood-volume prediction maps.

The tree species composition was validated on NFI sample plots, because we used FMI data for the prediction model, and the NFI and FMI datasets are independent. Two selection criteria were used: it was required that the sample plot be described as forest in the NFI, and the probability of the plot being non-forest was required to be < 25% according to predicted pixel values in the tree species map. This pair of filters helped to remove disturbed areas and validation points in which the spectral signature is mainly influenced by objects other than forest trees. We analyzed the prediction for dominant tree species and the prediction for the proportion of evergreen coniferous trees in the species composition.

Results

The forest-height prediction models described 89.5–94.8% of the variability in the empirical data (Table 1). The residual standard error of the models remained below 2.4 m. The comparison of predicted forest height on ALS dataset overlap areas revealed a small systematic difference, dependent on the dominant species (Figure 5). With models fitted for each ALS dataset individually, the predicted values for forests dominated by broadleaf deciduous trees tended to be greater when a midsummer ALS dataset was used. For evergreen coniferous trees, with the midsummer ALS dataset, the species-independent prediction model yielded an indication of underestimation (Table 4).

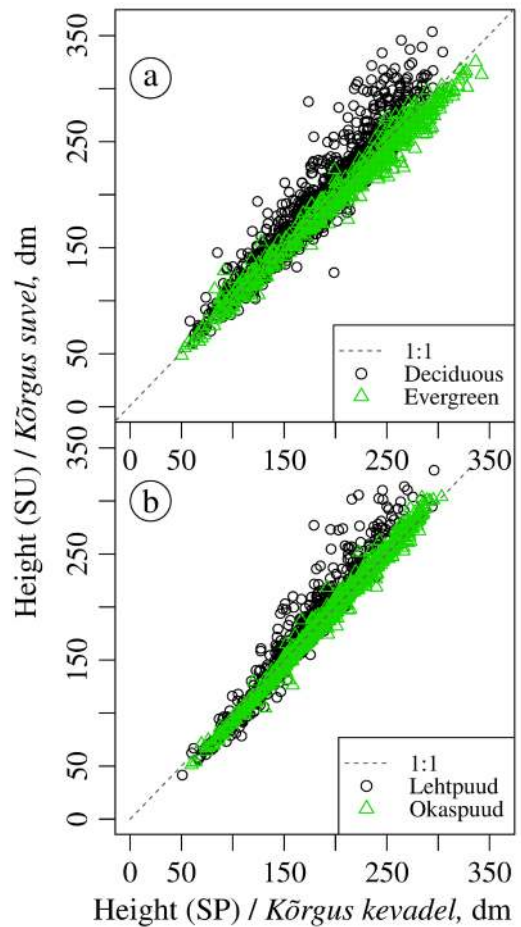


Figure 5. Predicted forest height on the overlap area of ALS data: (a) blocks SW and SE, (b) blocks NW and NE. The blocks correspond to lidar data from summer (SU) and springtime (SP).

Joonis 5. Lidarmõõtmise alade ülekatedel asuvatele puistutele prognoositud kõrgus: (a) plokid SW ja SE, (b) plokid NW ja NE. Plokid vastavad suvistele (SU) ja kevadistele (SP) lidarandmetele.

Table 1. Parameters for the forest-height H (dm) prediction model (1). ALS data are from spring (SP) and summer (SU), with RSE the model residual standard error, R^2 the coefficient of determination, and DF the number of degrees of freedom. Insignificant values ($p > 0.05$) are in italics.

Tabel 1. Metsa kõrguse H (dm) prognoosmudeli (1) parameetrid. Eristatud on suvised (SU) ja kevadised (SP) laserskaneerimise lennud. RSE on mudeli jääkhälve, R^2 on determinatsioonikordaja ja DF on vabadusastmete arv. Mitteilulised väärtused ($p > 0,05$) on kursiivis.

ALS flight <i>ALS lend</i>	Block <i>Ala</i>	Model parameters* / <i>Mudeli parameetrid*</i>				
		<i>a</i>	<i>b</i>	RSE (dm)	R^2 %	DF
SP 2017	SW	9.05	11.86	24	89.5	292
SU 2017	SE	6.09	11.58	21	94.7	281
SP 2018	NE	13.80	11.47	17	94.8	313
SU 2018	NW	<i>3.94</i>	12.07	21	91.8	312

*Height percentiles are in metres. / *Punktipilve kõrgusprotsentiilid on meetrites.*

Table 2. Parameters for the standing-wood volume M ($m^3 ha^{-1}$) prediction model (2) for the upper layer. ALS data are from springtime (SP) and summer (SU). RSE is the model residual standard error, R^2 the coefficient of determination, and DF the number of degrees of freedom. Insignificant values ($p > 0.05$) are in italics.

Tabel 2. Esimese rinde tüvemahu M ($m^3 ha^{-1}$) prognoosmudeli (2) parameetrid. Eristatud on suvised (SU) ja kevadised (SP) laserskaneerimise lennud. RSE on mudeli jääkhälve, R^2 on determinatsioonikordaja ja DF on vabadusastmete arv. Mitteilulised väärtused ($p > 0,05$) on kursiivis.

ALS flight <i>ALS lend</i>	Model parameters* / <i>Mudeli parameetrid*</i>						
	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	RSE ($m^3 ha^{-1}$)	R^2	DF
SP 2017	1.438	1.264	<i>0.200</i>	0.416	78.6	84.2	287
SU 2017	<i>0.083</i>	1.452	<i>0.121</i>	0.861	97.0	85.2	276
SP 2018	1.277	1.233	<i>0.308</i>	0.454	65.7	91.7	308
SU 2018	<i>0.298</i>	1.492	<i>0.003</i>	0.598	76.3	85.1	304

*Height percentiles are in metres, and canopy cover values within the range of 0–100. / *Punktipilve kõrgusprotsentiilid on meetrites ja katvuse väärtus on vahemikus 0–100.*

The fitted models for standing-wood volume prediction described 84.2–91.7% of the variability in the empirical data (Table 2, Figure 6). The residual standard error of the models remained in the range of 66–97 $m^3 ha^{-1}$. Model (2) is constructed with regard to the prevalence of multi-layer canopies in Estonia. In the current exploratory study, on the other hand, the variable predicted is wood volume for the dominant tree layer. We consider this to

be the reason why the parameter c for the lower quartile of the point-cloud height distribution was not significant (Table 2). The model analysis indicated that predicted values may be dependent on dominant tree species (Table 3, Figure 5), as the mean residual error (MRE) had values in the range of (–51)–54 $m^3 ha^{-1}$ when calculated for sample plots with dominance of a particular tree species.

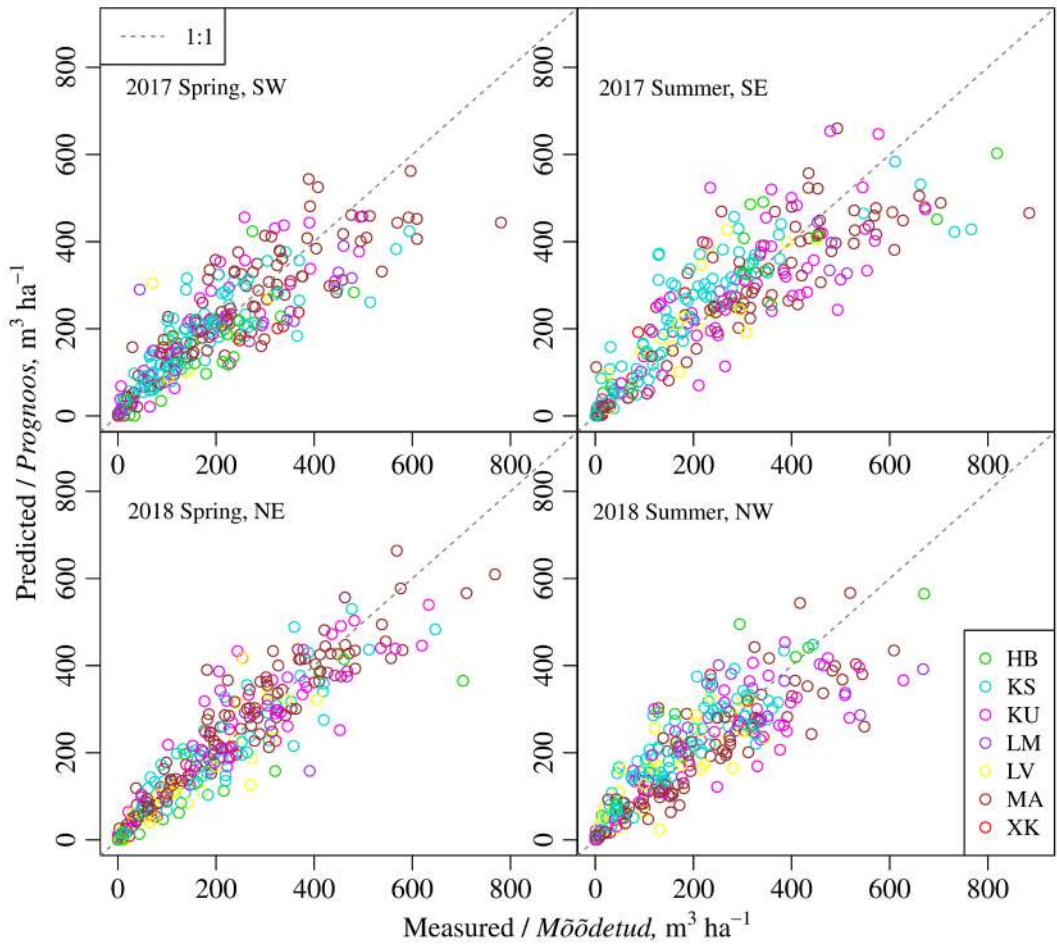


Figure 6. Measured and predicted wood volume for NFI sample plots. Dominant species codes: HB = European aspen; KS = silver birch; KU = Norway spruce; LM = black alder; LV = grey alder; MA = Scots pine; XK = other.

Joonis 6. Mõõdetud ja prognoositud tüvemaht SMI proovitükkidel. Enamuspuuliikide koodid on metsakorralduse juhendi järgi.

Table 3. Standing-wood volume M ($\text{m}^3 \text{ha}^{-1}$) prediction errors under model (2), by dominant tree species, with ALS data distinguished according to spring (SP) and summer (SU).

Tabel 3. Esimese rinde tüvemahu M ($\text{m}^3 \text{ha}^{-1}$) prognoosumudeli (2) vead peapuuliikide järgi. Eristatud on suvised (SU) ja kevadised (SP) laserskaneerimise lennud.

ALS data <i>ALS andmed</i>	Target area <i>Piirkond</i>	Variable* <i>Tunnus*</i>	Dominant species** / <i>Enamuspuuliik**</i>						
			HB	KS	KU	LM	LV	MA	XK
SP 2017	SW	\bar{M}	190	161	179	263	152	250	233
SP 2017	SW	RSE	77	67	70	116	96	83	81
SP 2017	SW	MRE	-37	11	18	-45	25	-1	-59
SP 2017	SW	N	22	93	43	13	7	106	7
SU 2017	SE	\bar{M}	265	190	287	331	199	302	87
SU 2017	SE	RSE	102	92	106	90	76	98	73
SU 2017	SE	MRE	-3	37	-13	-51	22	-26	57
SU 2017	SE	N	18	83	57	8	18	92	4
SP 2018	NE	\bar{M}	157	197	235	188	145	237	166
SP 2018	NE	RSE	95	60	73	78	53	56	30
SP 2018	NE	MRE	-37	-3	6	-26	-9	10	-26
SP 2018	NE	N	19	59	66	10	31	120	3
SU 2018	NW	\bar{M}	239	162	248	314	158	211	235
SU 2018	NW	RSE	104	59	91	125	59	72	144
SU 2018	NW	MRE	54	31	-29	-17	12	-24	144
SU 2018	NW	N	9	99	54	16	36	93	1

* \bar{M} = mean of measured values ($\text{m}^3 \text{ha}^{-1}$); N = number of observations; RSE = residual standard error ($\text{m}^3 \text{ha}^{-1}$); MRE = mean residual error ($\text{m}^3 \text{ha}^{-1}$). ** Species codes: HB = European aspen; KS = silver birch; KU = Norway spruce; LM = black alder; LV = grey alder; MA = Scots pine; XK = other. / * \bar{M} : vaatluste aritmeetiline keskmine ($\text{m}^3 \text{ha}^{-1}$), N: vaatluste arv, RSE: keskmine ruutviga ($\text{m}^3 \text{ha}^{-1}$), MRE: keskmine viga ($\text{m}^3 \text{ha}^{-1}$). ** Puuliikide koodid on metsakorralduse juhendi järgi.

The wood volume prediction for evergreen stands was greater when estimated from springtime ALS maps (Table 4, Figure 7). The opposite was observed for deciduous stands. The apparent change in predicted values for evergreen forests is the result of using a prediction model that does not account for a change in the leaf area index. A second factor may also be relevant: canopy cover estimates based on Riegl VQ-1650i laser-scanner data may be expected to be sensitive to the contribution from evergreen coniferous and deciduous broadleaf trees due to their morphological differences in the shoot and crown structure.

Table 4. The average difference between predicted forest height \hat{H} (m) and standing-wood volume \hat{M} ($\text{m}^3 \text{ha}^{-1}$) in the case of both evergreen (EGR) and deciduous (DEC) stands, upon comparing ALS springtime (SP) against summer (SU) data.

Tabel 4. Prognoositud kõrguse \hat{H} (m) ja esimese rinde tüvemahu \hat{M} ($\text{m}^3 \text{ha}^{-1}$) keskmise erinevus okaspuuistutes (EGR) ja lehtpuuistutes (DEC) kasutades suvised (SU) ja kevadisi (SP) laserskaneerimise andmeid.

Block SP <i>Ala SP</i>	Block SU <i>Ala SU</i>	$(\hat{H}_{SP} - \hat{H}_{SU})$, m		$(\hat{M}_{SP} - \hat{M}_{SU})$, $\text{m}^3 \text{ha}^{-1}$	
		EGR	DEC	EGR	DEC
SW	SE	0.53	-0.90	34.5	-46.5
NE	NW	0.13	-0.71	40.4	-38.0

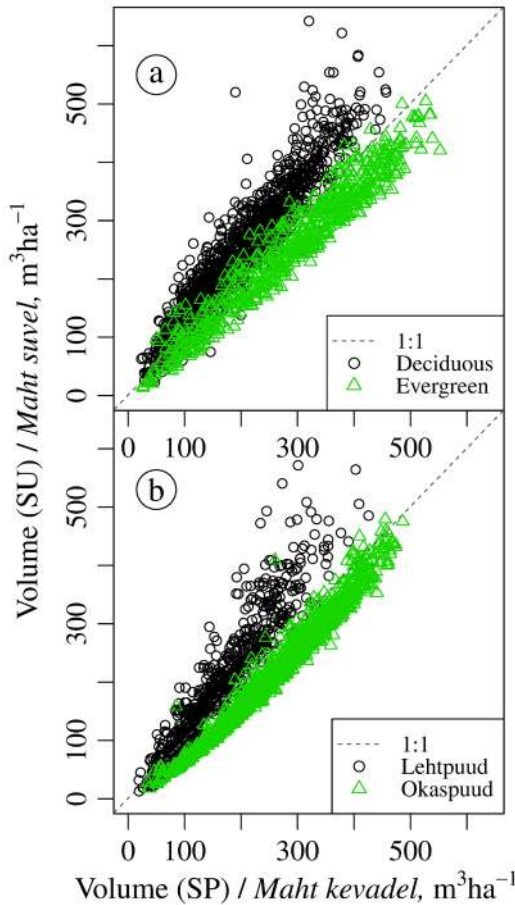


Figure 7. Predicted standing-wood volume on the overlap area of ALS data: (a) blocks SW and SE, (b) blocks NW and NE. The blocks correspond to lidar data from summer (SU) and spring-time (SP).

Joonis 7. Prognoositud puistu tüvemaht lidar-mõõtmise alade ülekatedel: (a) plokid SW ja SE, (b) plokid NW ja NE.

The tree species composition prediction was first compared to a dataset published by Lang *et al.* (2018), using 6,239 pixels drawn according to NFI sample-plot location coordinates. The Cohen's kappa value 95% confidence interval for dominant species was 0.69–0.72. When the prediction for dominant tree species was compared to the NFI sample-plot data (Table 5), the Cohen's kappa value 95% confidence interval was 0.51–0.54. The somewhat small kappa

value can be well explained by the fact that the validation dataset contained both pure and mixed stands. The NFI sample plot coordinates typically have a positioning error of 5–10 m, and, in addition, the sample plots are small (7–10 m in radius). Consequently, the validation dataset itself has some uncertainties: a small sample plot describes a particular point in the given forest which may or may not be representative of the surrounding forest stand, even though the surrounding stand contributes to the spectral signature in the corresponding satellite-image pixel. The second validation was done using the predicted proportion of evergreen coniferous tree species (taken as pine and spruce summed), and was based on the 5,011 NFI sample plots with forests aged over 24 years. Although there was substantial scatter at the single-pixel level, the predictions were consistent (Figure 8), and, on average, the increased proportion of evergreen coniferous species on NFI sample plots was found to be present on corresponding pixels of the tree species composition map.

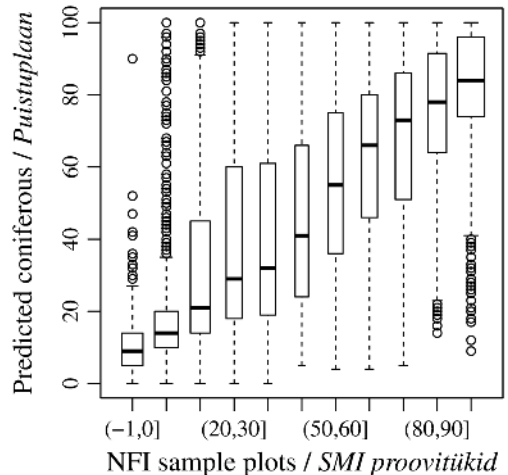


Figure 8. Predicted proportion (%) of evergreen conifers for 5,011 NFI sample plots with stands older than 24 years.

Joonis 8. Prognoositud okaspuude osakaal 5011 SMI proovitükil, kus kasvab 24-aastane või vanem puistu.

Table 5. Predicted dominant species and field observations on 6,239 NFI sample plots.

Tabel 5. Prognoositud enamuspuuliik võrrelduna 6239 SMI proovitüki andmetega.

NFI / SMI	Predicted dominant species* / Prognoositud enamuspuuliik*						
	KS	KU	MA	LV	LM	HB	XK
KS	1,110	186	114	153	137	106	16
KU	201	694	162	47	12	37	6
MA	173	211	1,532	4	8	15	2
LV	93	25	3	334	14	23	7
LM	67	10	7	25	123	20	4
HB	81	34	5	65	23	108	13
XK	47	13	15	72	30	35	17

*Species codes: HB = European aspen; KS = silver birch; KU = Norway spruce; LM = black alder; LV = grey alder; MA = Scots pine; XK = other. / Puuliikide koodid on metsakorralduse juhendi järgi.

Discussion

Forest height, standing-wood volume, and tree species composition

The NFI is based on contact measurements of trees on a set of small sample plots. For spatial units smaller than the individual county, feature variables from remote-sensing data can be used to construct maps of forest-inventory variables. As in the case of many other countries, Estonia has a national programme of springtime aerial photography and ALS for topographic mapping, with special flights additionally performed at higher altitudes in the summertime for forest-inventory purposes.

To construct maps of forest height based on the 3-dimensional ALS point clouds and NFI sample-plot measurements, it suffices to apply a linear model that takes as argument an upper percentile of the point-cloud height distribution. Upon taking a model independent of tree species and comparing forest height predictions based on springtime against predictions based on summer ALS data, we found a slight bias in the predicted forest height, dependent on the proportion of evergreen coniferous species in the subject forests. The difference proved more pronounced in deciduous forests, where greater heights were predicted when summer ALS data were used.

Lidar pulse returns are rarely recorded from tree stems during airborne measurements. The prediction of standing-wood volume therefore requires a model that accounts for tree height and forest density (defined as the number of trees per unit area). With detection of single trees not reliable from sparse point clouds with about 1 point per m², canopy cover can be taken as a proxy for forest density. However, the interaction of laser pulse with forest canopy, including the degree of pulse penetration to the ground, is substantially dependent on the amount of foliage, in addition to suffering dependence on flight parameters and scanner settings. There are two options for incorporating canopy cover information into the standing-wood volume model. The first option is to use a variable that predicts canopy cover. The second option is to calculate forest height metrics from point clouds which include near-to-ground points. The former option has been adopted for Estonian forests, while jurisdictions adopting the second option include Finland (Kotivuori *et al.*, 2016). Arumäe and Lang (2018) analyzed ALS-based canopy cover (CC_{ALS}) estimation errors using digital hemispherical images and found the random variability of CC_{ALS} to be about 10–12% for sample plots 30 m in radius. In view of the random variability in CC_{ALS} , the combined effect of lo-

cation errors and small sample-plot sizes, and the failure of tree stems to exert direct influence on the sparse ALS point clouds, the fitted for wood volume prediction are reliable for practical use for Estonian forests. The relative residual error of the models remained in the range of 30–38% of the mean measured value. A further decrease of random errors in predicted values, although possible, is difficult to achieve at pixel level (Arumäe & Lang, 2016).

A bias similar to the result encountered with evergreen coniferous-dominated forest height prediction was found in the standing-wood volume prediction, with the common model yielding an smaller by about $40 \text{ m}^3 \text{ ha}^{-1}$ for summer than for springtime ALS data. On the other hand, a systematically greater, with a discrepancy of the same order of magnitude, was found for forests dominated by deciduous trees for summer as compared against springtime ALS data. The problem could be addressed by dividing the empirical data into subsets according to the dominant tree species and fitting model parameters for each subset. However, in this approach the number of observations decreases and it becomes necessary to know the tree species distribution with high accuracy before the model can be applied. The number of observations for model fitting can be increased by including sample-plot measurements from several years. If information about forest disturbances is reliable, then forest height or wood-volume data for the sample plots measured a few years before the collection of ALS data can be updated with a forest growth model.

Tree species composition prediction was tested in this study for 25 m target pixels. Although the uncertainty in species was substantial at the pixel level, the proportion of aggregated evergreen coniferous species in the composition was predicted reliably, and therefore could in future versions of the system be used for correcting phenology effects.

Options for annual updating of forest inventory data

The NFI is intended to provide timely forest data for the whole country and to furnish published maps of the main variables, namely forest height, standing-wood volume, tree species composition, forest age, basal area, relative density, and location of disturbances. However, even for a small country, such as Estonia it is not economically feasible to carry out ALS measurements every year for the whole area, making it necessary in the case of many Estonian locations to update the data by other means. With a geographically comprehensive prediction of tree species composition, it becomes possible to use forest growth modelling to update forest height and wood-volume predictions for those areas lacking current-year ALS measurements. The current Estonian Land Board flight schedule makes it necessary to apply the modelling over a period of up to two years. For the determination of forest height, an algebraic difference model is used to simulate the growth of Estonian forests (Kiviste, 1997):

$$H_{ALS,updated} = f(A_1, H_1, A, OHOR, PE, TEKE). \quad (5)$$

Here $H_{ALS,updated}$ the updated height for a forest of some specified age A , is calculated from the known value of height H_1 at age A_1 , the known soil organogenic layer thickness $OHOR$ (in cm), and the known stand origin $TEKE$ (given as natural or cultivated) for a known dominant species PE . A model similar to (5) is available for wood volume. For areas without current FMI data, forest age can be calculated from a site fertility-index model used in Estonia, based, in turn, on forest height and stand age (Forestinv, 2017).

The remaining two variables to be predicted for forest management planning are stand basal area G and relative density T . For forest inventory, the relative density is calculated as the ratio of stand basal area

to a reference (a normal stand in the given place). It is, in principle, possible to predict G and T from ALS data. However, we propose to use the relationship $M = G \cdot H \cdot F$ where wood volume is the value predicted from ALS data, and where the stand form factor height $H \cdot F$ is predicted with a forest-height-based model taken from forest inventory regulations (Forestinvent, 2017). For each target pixel, the forest height is then predicted from ALS data with model (1).

As an essential part of our remote-sensing-based prediction system for forest variables, we designed a module for the estimation of prediction errors. This error-budget module is based on the k -NN method and on the assumption that NFI sample-plot field measurement data provide 'true' values for the forest variables. According to the proposed method, the prediction error of a forest variable for a given pixel is calculated from known errors of similar pixels (found by the k -NN algorithm) in the NFI data. The validation of prediction models is still based on simple validation datasets. Nevertheless, for practical use, the described method makes it possible to construct uncertainty maps of predicted values for the entire target area.

Outlook

In this paper, we have presented a fully functional remote-sensing support module for the Estonian NFI, supporting the construction of maps of forest height, standing-wood volume, and tree species composition. Our study has, however, revealed the need for correction of phenology effects in the ALS data. This can be achieved by increasing the accuracy of tree species composition estimates, using longer time series of multispectral satellite data and dense time series of Sentinel-1 SAR data as proposed by Dostálová *et al.* (2018).

When the system becomes fully operational, it will be possible to determine also forest age at high accuracy, by using time series of multi-temporal satellite images for the detection of forest regeneration fell-

ings (Peterson *et al.*, 2004; Liira *et al.*, 2006). Weak disturbances, such as normal thinnings, can also be detected from bi-temporal multispectral satellite image pairs (Uiga *et al.*, 2003), as well as from sparse point clouds obtained through repeated ALS measurements (Arumäe, *et al.*, 2020). With the combination of these various refinements in technique, Estonian forestry is entering an era in which data for forests at every point of the country are updated yearly and can be used for sustainable forest-management planning, on the basis of 10–30 m spatial units.

Conclusions

The remote-sensing support module offered in this paper for the Estonian NFI accepts open-source data. Our models for sparse ALS point clouds yield coefficients of determination in the ranges of 89.5–94.8% for height and 84.2–91.7% for wood volume. However, validation of the common model prediction results has revealed systematic bias, upon comparing predictions from summer against predictions from springtime ALS data. The bias problem could be addressed by including the share of evergreen tree species in the model for springtime ALS data. At the pixel level, the prediction of dominant tree species, from a set of six possible options, has been found to give a Cohen's kappa value of 95% for the confidence interval 0.51–0.54 when validated on small NFI sample plots, with the determination of the share of aggregated evergreen tree species in forests found to be reliable. Further studies are required to better account for phenology effects influencing ALS data and to increase the precision of tree species composition predictions. Already, however, our proposed solution indicates a path forward for Estonian forestry, offering the prospect of annually updated forest-inventory data for all Estonian forest at 10–30 m spatial resolution, to support sustainable

forest-management planning.

Acknowledgements. The study was financed by the European Regional Development Fund within the National Activity “Smart development (including analysis) of existing and new information systems”. The Estonian Land Board releases its ALS data, its digital base map, and its digital soil map for public use. Landsat-8 OLI and Sentinel-2 MSI images are released for public use in the USGS and Copernicus framework. We thank Enn Pärt from the Estonian Environment Agency for helping us find unpublished reports about Estonian forest resources and technical reports about early NFI test-case work in Estonia. We thank Dr. Toomas Karmo for the language proofreading of the manuscript. The map of country borders is provided by ESRI for public use. We thank reviewers for their comments.

References

- Adermann, V. 2010. Development of Estonian National Forest Inventory. – Tomppo, E., Gschwantner, T., Lawrence, M., McRoberts, R.E. (eds.). National Forest Inventories: Pathways for Common Reporting. Heidelberg, Springer, 171–184.
- Aru, A., Okas, J. 1959. Forestry development plan for Estonian SSR 1959–1975. (Eesti NSV metsamajanduse arendamise generaalplaan 1959–1975.a. Seletuskiri). Ü/K “Metsaprojekt” Eesti Aerofoto-metsakorralduse Kontor. Tallinn. 171 pp. (In Estonian).
- Aru, A., Tappo, E., Lall, E. 1975. Forestry development plan for Estonian SSR 1976–1990. (Eesti NSV metsamajanduse arendamise perspektiivplaan 1976.–1990.a.). Ü/K “Metsaprojekt” Eesti Metsakorralduskeskus. Tallinn, Eesti NSV Metsamajanduse ja Looduskaitse Ministerium. 350 pp. (In Estonian).
- Arumäe, T., Lang, M. 2016. ALS-based wood volume models of forest stands and comparison with forest inventory data. – *Forestry Studies / Metsanduslikud Uurimused*, 64, 5–16. <https://doi.org/10.1515/fsmu-2016-0001>. (In Estonian with English summary).
- Arumäe, T., Lang, M. 2018. Estimation of canopy cover in dense mixed-species forests using airborne lidar data. – *European Journal of Remote Sensing*, 51(1), 132–141. <https://doi.org/10.1080/22797254.2017.1411169>.
- Arumäe, T., Lang, M., Laarmann, D. 2020. Thinning- and tree-growth-caused changes in canopy cover and stand height and their estimation using low-density bitemporal airborne lidar measurements – a case study in hemi-boreal forests. – *European Journal of Remote Sensing*, 53(1), 113–123. <https://doi.org/10.1080/22797254.2020.1734969>.
- Barrett, F., McRoberts, R.E., Tomppo, E., Cienciala, E., Waser, L.T. 2016. A questionnaire-based review of the operational use of remotely sensed data by national forest inventories. – *Remote Sensing of Environment*, 174, 279–289. <https://doi.org/10.1016/j.rse.2015.08.029>.
- Dostálová, A., Wagner, W., Milenković, M., Hollaus, M. 2018. Annual seasonality in Sentinel-1 signal for forest mapping and forest type classification. – *International Journal of Remote Sensing*, 39(21), 7738–7760. <https://doi.org/10.1080/01431161.2018.1479788>.
- ESA. 2015. Sentinel-2 user handbook. – ESA Standard Document. Date 24 July 2015, Issue 1, Rev 2. 64 pp.
- ESA. 2020. Sentinel-1 product specification. – Ref: S1-RS-MDA-52-7441. Date 27 February 2020, Issue 3/7. 197 pp.
- ESTHub. 2016. Smart infrastructure development plan for year 2016. (2016. a nutika teenuste taristu arendamise toetamise investeeringute kava nr 5). Majandus- ja Kommunikatsiooniministerium. – *Käskkiri* 22.11.16 nr 16-0334. (In Estonian).
- Etverk, I. 2003. 20th Century in Estonian Forests. Painküla, Eesti Metsaselts. 193 pp.
- Fazakas, Z., Nilsson, M., Olsson, H. 1999. Regional forest biomass and wood volume estimation using satellite data and ancillary data. – *Agricultural and Forest Meteorology*, 98/99, 417–425. [https://doi.org/10.1016/S0168-1923\(99\)00112-4](https://doi.org/10.1016/S0168-1923(99)00112-4).
- Franco-Lopez, H., Ek, A.R., Bauer, M.E. 2001. Estimation and mapping of forest sand density, volume, and cover type using the *k*-nearest neighbors method. – *Remote Sensing of Environment*, 77, 251–274. [https://doi.org/10.1016/S0034-4257\(01\)00209-7](https://doi.org/10.1016/S0034-4257(01)00209-7).
- Forest. 1995. Yearbook of Forest 1994. (Aastaraamat Mets '94). – Hepner, H. (ed.). Tallinn, Metsamajanduse Ökoonoomika- ja Informaatikakeskus. 107 pp.
- Forest. 2020. Yearbook Forest 2018. (Aastaraamat Mets 2018). – Raudsaar, M., Siimon, K.L., Valgepea, M. (eds.). Tallinn, Keskkonnaagentuur. 279 pp.
- Forestact. 2016. Forest act. (Metsaseadus). – RT I 2006, 30, 232. (In Estonian).
- Forestinv. 2017. Forest inventory act. (Metsa korraldamise juhend). – RT I, 22.02.2017, 11. (In Estonian).
- Forestman. 2007. Forest management regulation. (Metsa majandamise eeskiri). – RTL 2007, 2, 16. (In Estonian).

- Gustafsson, L., Baker, S.C., Bauhus, J., Beese, W.J., Brodie, A., Kouki, J., Lindenmayer, D.B., Löhmus, A., Martínez Pastur, G., Messier, C., Neyland, M., Palik, B., Sverdrup-Thygeson, A., Volney, J., Wayne, A., Franklin, J.F. 2012. Retention forestry to maintain multifunctional forests: a world perspective. - *BioScience*, 62(7), 633–645. <https://doi.org/10.1525/bio.2012.62.7.6>.
- Howard, J.A. 1991. *Remote Sensing of Forest Resources. Theory and Application*. London, Chapman & Hall. 420 pp.
- Jõgiste, K., Frelich, L.E., Laarmann, D., Vodde, F., Baders, E., Donis, J., Jansons, A., Kangur, A., Korjus, H., Köster, K., Kusmin, J., Kuuluvainen, T., Marozas, V., Metslaid, M., Metslaid, S., Polyachenko, O., Poska, A., Rebane, S., Stanturf, J.A. 2018. Imprints of management history on hemiboreal forest ecosystems in the Baltic States. - *Ecosphere*, 9(11), e02503. <https://doi.org/10.1002/ecs2.2503>.
- Kiviste, A. 1997. An algebraic difference model for the forest growth simulation in Estonia. (Eesti riigimetsa puistute kõrguse, diameetri ja tagavara vanuseridade diferentsmudel 1984.–1993. a. metsakorralduse takseerkirjelduste andmeil). - *EPMÜ teadustööde kogumik*, 189, 63–75. Tartu, Eesti Põllumajandusülikool. (In Estonian).
- Kohava, P. 1998. A report about statistical forest inventory on Hiiumaa. (Hiiumaa statistilise metsainventeerimise katsetööde aruanne). - Technical Report. Tallinn, OÜ Eesti Metsakorralduskeskus. Unpublished. 51 pp. (In Estonian).
- Kohava, P. 2000. *Forests in Estonia 1999*. (Eesti metsad 1999). Tallinn, OÜ Eesti Metsakorralduskeskus. 44 pp. (In Estonian).
- Kotivuori, E., Korhonen, L., Packalén, P. 2016. Nationwide airborne laser scanning based models for volume, biomass and dominant height in Finland. - *Silva Fennica*, 50(4), 1–28. <http://dx.doi.org/10.14214/sf.1567>.
- Kremser, W. 1998. *History of Estonian Forestry*. (Epochen der Forstgeschichte Estlands). Tallinn, Tallinna Raamatutrükikoda. 280 pp. (In German).
- Krigul, T. 1972. *Forest Mensuration*. (Metsatakseerimine). Tallinn, Valgus. 359 pp. (In Estonian).
- Lang, M., Arumäe, T., Anniste, J. 2012. Estimation of main forest inventory variables from spectral and airborne lidar data in Aegviidu test site, Estonia. - *Forestry Studies / Metsanduslikud Uurimused*, 56, 27–41. <https://doi.org/10.2478/v10132-012-0003-7>. (In Estonian with English summary).
- Lang, M., Arumäe, T., Lükk, T., 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. <https://doi.org/10.2478/fsmu-2014-0010>.
- Lang, M., Gulbe, L., Traškovs, A., 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.
- Lang, M., Kaha, M., Laarmann, D., Sims, A. 2018. Construction of tree species composition map of Estonia using multispectral satellite images, soil map and a random forest algorithm. - *Forestry Studies / Metsanduslikud Uurimused*, 68, 5–24. <https://doi.org/10.2478/fsmu-2018-0001>.
- Liira, J., Püssa, K., Peterson, U. 2006. The radiance contrast of forest-to-clearcut edges on a medium resolution Landsat Enhanced Thematic Mapper satellite winter image. - *International Journal of Remote Sensing*, 27(13), 2753–2766. <https://doi.org/10.1080/01431160600593023>.
- Maa-amet. 2019. Map sheet indexes and coordinate systems. [WWW document]. - URL <https://geoportaal.maaamet.ee/eng/Spatial-Data/Map-Sheet-Indexes-and-Coordinate-Systems-p359.html>. [Accessed 20 May 2020].
- Maa-amet. 2020. Orthophoto metadata by year. [WWW document]. - URL <https://geoportaal.maaamet.ee/eng/Spatial-Data/Orthophoto>. [Accessed 04 June 2020].
- Mandel, M., Maasikamäe, S., Lang, M. 2019. Land abandonment in Tartu County. - *Forestry Studies / Metsanduslikud Uurimused*, 71, 30–47. <https://doi.org/10.2478/fsmu-2019-0011>. (In Estonian with English summary).
- McGaughey, R.J. 2018. FUSION/LDV: Software for LIDAR Data Analysis and Visualization. August 2018 – FUSION Version 3.80. United States Department of Agriculture Forest Service, Pacific Northwest Research Station. 211 pp.
- McRoberts, R.E., Tomppo, E.O. 2007. Remote sensing support for national forest inventories. - *Remote Sensing of Environment*, 110, 412–419.
- Meikar, T. 1998. *Forest Inventory in Estonia*. (Metsakorraldus Eestis). Tartu, Eesti Põllumajandusülikooli Metsandusteaduskond. 96 pp. (In Estonian).
- Metsainventuur. 2000a. Forest statistics by using statistical sampling. (Metsade inventeerimine statistilise valikmeetodiga. Riigihankelepingu raport). - Technical Report. Tartu, Metsaekspert OÜ. Unpublished. 65 pp. (In Estonian).
- Metsainventuur. 2000b. A sampling design for National Forest Inventory in Estonia. (Eesti metsade statistilise inventeerimise meetodika). Ver 1.0. Seisuga 15.02.2000. Tartu, Metsaekspert OÜ. Unpublished. 55 pp. (In Estonian).
- Næsset, E. 1997. Determination of mean tree height of forest stands using airborne laser scanner data. - *ISPRS Journal of Photogrammetry & Remote Sensing*, 52, 49–56. [https://doi.org/10.1016/S0924-2716\(97\)83000-6](https://doi.org/10.1016/S0924-2716(97)83000-6).

- Næsset, E. 2005. Assessing sensor effects and effects of leaf-off and leaf-on canopy conditions on biophysical stand properties derived from small-footprint airborne laser data. - *Remote Sensing of Environment*, 95, 356-370. <https://doi.org/10.1016/j.rse.2005.07.012>.
- Næsset, E., Gobakken, T., Holmgren, J., Hyyppä, H., Hyyppä, J., Maltamo, M., Nilsson, M., Olsson, H., Persson, Å., Söderman, U. 2004. Laser scanning of forest resources: the Nordic experience. - *Scandinavian Journal of Forest Research*, 19(6), 482-499. <https://doi.org/10.1080/02827580410019553>.
- NFIEUROPE. 1997. Study on European Forestry Information and Communication System - Reports on Forest Inventory and Survey Systems. - Päivinen, R., Köhl M. (eds.). Luxembourg, Office for Official Publications of the European Communities. 1328 pp.
- Nilson, T., Peterson, U. 1994. Age dependence of forest reflectance: Analysis of main driving factors. - *Remote Sensing of Environment*, 48, 319-331. [https://doi.org/10.1016/0034-4257\(94\)90006-X](https://doi.org/10.1016/0034-4257(94)90006-X).
- Packalén, P., Maltamo, M. 2007. The k-MSN method for the prediction of species-specific stand attributes using airborne laser scanning and aerial photographs. - *Remote Sensing of Environment*, 109, 328-341. <https://doi.org/10.1016/j.rse.2007.01.005>.
- Peterson, U., Aunap, R. 1998. Changes in agricultural land use in Estonia in the 1990s detected with multitemporal Landsat MSS imagery. - *Landscape and Urban Planning*, 41, 193-201. [https://doi.org/10.1016/S0169-2046\(98\)00058-9](https://doi.org/10.1016/S0169-2046(98)00058-9).
- Peterson, U., Püssa, K., Liira, J. 2004. Issues related to delineation of forest boundaries on Landsat Thematic Mapper winter images. - *International Journal of Remote Sensing*, 25(24), 5617-5628. <https://doi.org/10.1080/01431160412331291314>.
- Polli, M., Viilup, Ü. 1989. Forestry Development Plan "Estonian Forests 2010". (Eesti NSV metsade majandamise ja puidukasutuse arenduskava "Eesti mets 2010"). Tallinn, Eesti Metsakorralduskeskus, Eesti NSV Riiklik Looduskaitse ja Metsamajanduse Komitee. 187 pp.
- Poso, S., Karlsson, M., Pekkonen, T., Härmä, P. 1990. A system for combining data from remote sensing, maps and field measurement for forest planning purposes. - *Research Notes*. Helsinki, University of Helsinki, Department of Forest Mensuration and Management. 40 pp.
- Tamm, T., 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. <https://doi.org/10.1016/j.jag.2009.03.006>.
- Tappo, E. 1982. Mean Characteristics of Forest Stands in Estonia by Dominant Species, Site Fertility, and Age. (Eesti NSV puistute keskmised takseertunnused puistu enamuspuuliigi, boniteedi ja vanuse järgi). Tallinn, Eesti NSV Põllumajandusministeeriumi Informatsiooni ja Juurutamise Valitsus. 72 pp. (In Estonian).
- Tomppo, E., Schadauer, K., McRoberts, R.E., Gschwantner, T., Gabler, K., Ståhl, G. 2010. History of NFIs. - Tomppo, E., Gschwantner, T., Lawrence, M., McRoberts, R.E. (eds.). *National Forest Inventories*. Heidelberg, Springer, 1-2.
- Uiga, R., Lang, M., Lük, T., Anniste, J. 2003. Mapping the location of forest cuttings using Landsat 5 TM and Landsat 7 ETM+ imagery. - Kiviste, K., Lember, S., Kivastik, S. (eds.). *Postgraduate Research Papers at the Faculty of Forestry*, 36, 142-161.
- USGS. 2019. Landsat 8 (L8) Data Users Handbook. *LSDS-1574, Version 5.0*. Department of the Interior, U.S. Geological Survey. 114 pp.
- Vaus, M. 2005. *Forest Mensuration*. (Metsatakseerimine). Tartu, OÜ Halo kirjastus. 178 pp. (In Estonian).
- Viilup, Ü. 1995. *Estonian Forest Resources 1994*. (Eesti metsavarud seisuga 01. jaanuar 1994. a.). Tallinn, Eesti Metsakorralduskeskus. 42 pp. (In Estonian).

Kaugseirel põhinev lahendus Eesti statistilise metsainventuuri jaoks puistute kõrguse, tüvemahu ja liigilise koosseisu kaartide koostamiseks

Mait Lang, Allan Sims, Kalev Pärna, Raul Kangro, Märt Möls, Marta Möistus, Andres Kiviste, Mati Tee, Toivo Vajakas ja Mattias Rennel

Kokkuvõte

Eestis on metsade sihipärase takseerimisega (Krigul, 1972; Vaus, 2005) tegeletud juba peaaegu 150 aastat (Meikar, 1998). Metsataksaatori kohapeal koostatud puistute kirjeldused on küll piisavalt hea andmestik konkreetse metsaosajandusotsuste kavandamiseks ning on varasemalt olnud ainuke allikas üleriigilise metsafondi iseloomustamiseks (Aru & Okas, 1959; Aru *et al.*, 1975; Tappo, 1982; Polli & Viilup 1989; Viilup, 1995), aga tavaliselt on kümneaastane või pikem kordusmõõtmise periood, tööde kõrge hind, mittemajandatavate metsade alakaetus ja omandist sõltuv andmete uuendamise strateegia põhjused, miks ka Eestis on võetud kasutusele instrumentaalselt mõõdetavate proovitükkide valimil põhinev statistiline metsainventuur (SMI) (NFIEUROPE, 1997; Kohava, 1998; Adermann, 2010; Tomppo *et al.*, 2010). SMI vaatluste arv aastas võimaldab saada metsaressursi hinnanguid kuni maakonna tasemeni. Siinse töö eesmärgiks oli luua Eesti SMI jaoks meetodika ja tarkvaraline lahendus kaugseireandmete (multispektraalsed satelliidipildid, satelliitradari pildid ja aerolaserskaneerimise andmed) abil puistute kõrguse, tüvemahu ja liigilise koosseisu prognooskaartide koostamiseks. Sarnaseid lisasid statistilisele metsainventuurile leidub paljudes riikides (Poso *et al.*, 1990; McRoberts & Tomppo, 2007; Barrett *et al.*, 2016).

Maapealse mõõtmisandmestiku moodustavad SMI proovitükid, mis paiknevad kobaratena üle Eesti (joonis 1). Igal aastal mõõdetakse umbes 5000 proovitükki (joonis 2), mille raadius on 7 või 10

meetrit. Proovitükkide klastrid on maastikul 800×800 meetri suurused ruudud (joonis 3), mille külgedel paiknevad mõõtmiskohad.

Eesti Maa-ameti tehtav lidarmõõdistus (joonis 4) võimaldab saada metsa struktuuri kajastava kolmemõõtmelise andmestiku, mille meetrikutele (punkt pilvi kirjeldavad statistikud) tuginedes saab luua mudelid puistute kõrguse (1) ja tüvemahu (2) prognoosimiseks SMI proovitükkide suurusega võrreldavatele pinnaühikutele (10–30 m pikslid). Kõrguse ja tüvemahu mudelid ei ole praegu puistu enamuspüüliigist sõltuvad. Puistute koosseisu prognoosimiseks kasutakse multispektraalseid satelliidipilte ja metsaregistri andmeid (vt Lang *et al.*, 2018). Kõrguse H ja tüvemahu M mudelite iseloomustamiseks kasutati töös keskmist ruutviga (RSE) (3) ja keskmist viga (MRE) (4) ning prognoose analüüsiti ka lidarmõõdistuse objektide ülekattealadel (joonis 4), mis võimaldab saada tagasisidet kevadise ja suvise lidarandmestiku kasutamise tõttu tekkivatest nihetest prognoosides. Puistute liigilise koosseisu prognoosi täpsust näitab kapp indeks ning okaspuude osakaalu kokkulangevus valideerimisandmetega, mis saadi SMI proovitükkidelt.

Puistute kõrguse prognoosimudel (1) kirjeldas rohkem kui 89% empiirilise andmestiku variatsioonist (tabel 1). Kevadiste ja suviste lidarmõõdistusalade ülekatetel asuvatele puistutele prognoositud kõrguste võrdlemisel ilmnes mõningane süstemaatiline nihe sõltuvalt puistu enamuspüüliigist (tabel 4, joonis 5). Puistute tüvemahu prognoosimudel kirjeldas üle

82% empiirilise andmestiku variatsioonist (tabel 2). Kuigi mõõdetud ja prognoositud tüvemahus enamupuuliigist sõltuvust visuaalselt ei paista (joonis 6), siis puuliikide järgi rühmitatud empiirilise andmestiku korral sõltub mudeli keskmine viga siiski enamupuuliigist (tabel 3). Puuliigi mõju ilmneb selgelt kevadiste ja suviste lidarmõõdistusalade ülekatetel kasvavatele puistutele prognoositud tüvemahtude võrdlemisel (tabel 4, joonis 7). Prognooside täpsustamiseks tuleb edaspidi uurida võimalusi puuliigist sõltuvate mudelite kasutamiseks, kuigi oluliseks takistuseks on siin väike SMI proovitükkide arv. Puistu liigilise koosseisu prognoosi võrreldi enamupuuliigi järgi Lang *et al.* (2018) saadud tulemustega kasutades 6239 SMI proovitüki asukohalt võetud piksleid ning saadi Coheni kapp 95%-lised usalduspiirid 0,69–0,72. SMI proovitükkide takseerkirjel-

dustes oleva enamupuuliigiga oli kokkulangevus (tabel 5) veidi väiksem ja Coheni kapp 95%-lised usalduspiirid olid 0,51–0,54. Selline tulemus on igati aktsepteeritav arvestades prognoositud puuliikide arvu ja võimalikke segusid Eestis ning SMI proovitükkide väiksust ja asukohatäpsust ning satelliidipildi piksliste spektraalsete signatuuride kujunemist mõjutavate tegurite suurt arvu. Seda näitas ka prognoositud okaspuude osakaalu keskmine kokkulangevus 5011 SMI proovitükil, kus kasvas 25-aastane või vanem puistu (joonis 8). Kokkuvõtteks tuleb nentida, et käesoleva rakendusuuringu tulemusena on Eestis jõutud tasemele, kus säästliku metsamajanduse kavandamiseks on metsakorralduses nüüd võimalik kasutada üle riigi 10–30 meetrise ruumilahutusega andmeid, mille vanus pole rohkem kui üks kuni kaks aastat.

Received July 19, 2020, revised November 21, 2020, accepted November 23, 2020