BALTIC FORESTRY

http://www.balticforestry.mi.lt
ISSN 1392-1355
eISSN 2029-9230

Baltic Forestry 2023 29(1): 2–11
Category: Research article
https://doi.org/10.46490/BF663

The influence of forest tree species composition on the forest height predicted from airborne laser scanning data – A case study in Latvia

JANIS IVANOVS [1]*, ANDIS LAZDIŅŠ [1] AND MAIT LANG [2, 3]

[1] *Latvian State Forest Research Institute "Silava", Riga Str. 111, Salaspils LV-2169, Latvia*
[2] *Tartu Observatory, University of Tartu, Tõravere, Observatooriumi 1, Tartumaa 61602, Estonia; e-mail: lang@to.ee*
[3] *Institute of Forestry and Engineering, Estonian University of Life Sciences, Kreutzwaldi 5, Tartu 51006, Estonia*

* *Corresponding author: janis.ivanovs@silava.lv; phone: +371 25254403*

Ivanovs, J., Lazdiņš, A. and Lang, M. 2023. The influence of forest tree species composition on the forest height predicted from airborne laser scanning data – A case study in Latvia. *Baltic Forestry* 29(1): 2–11; https://doi.org/10.46490/BF663.

Received 22 August 2022 **Revised** 9 May 2023 **Accepted** 14 May 2023

Abstract

Airborne laser scanning (ALS) is used to predict different forest inventory parameters; however, the ALS point cloud properties depend on various parameters such as the type of ALS scanner employed, flight altitude and scanning angle, forest stand structure, forest tree species composition, vegetation season, etc. This study used national coverage high-resolution ALS data with minimum point density of 4 points per square meter in combination with field data from the National Forest Inventory (NFI) to build forest stand height models for forest stands dominated by 6 most common tree species in Latvian mixed forest stands, viz. *Pinus sylvestris* L., *Betula pendula* Roth, *Picea abies* (L.) Karst., *Populus tremula* L., *Alnus incana* (L.) Moench and *Alnus glutinosa* (L.) Gaertn. for the various ALS scanners employed and at different growing seasons. The selected NFI plots are divided into modelling and validation datasets in a ratio of 3 : 1. The results show that for a universal forest stand height model, the RMSE value is 1.91 m and the MAE is 1.41 m. For the forest stand height models, which are stratified by scanner, individual tree species and seasons, the RMSE value is within the limits of 1.4 m for forest stands dominated by Scots pine in leaf-on canopy condition to 3.8 m for birch in leaf-off canopy condition.

Keywords: forest inventory, airborne laser scanning, phenology, large scale forest mapping

Introduction

National and international environmental policy goals and agreements require environmental mapping and monitoring, but field surveys and the collection of information on a variety of important parameters of forest ecosystems over large areas are prohibitively expensive. In recent decades, the collection of such information has been revolutionized by remote sensing technologies such as LiDAR (**Li**ght **D**etecting **a**nd **R**anging), which provide accurate distance measurements based on the return time of emitted electromagnetic pulse. Nowadays, at a relatively lower cost, it is possible to accurately scan the surface of the earth and the vegetation growing on it, thus obtaining statistical data about the vegetation cover (Vauhkonen et al. 2014). Using ALS (**A**erial **L**aser **S**canning) technology, it is possible to obtain information about the three-dimensional structure of forest canopy, and such information is not easily obtained from other remote sensing (RS) data sources (Maltamo et al. 2011).

There is an obvious correlation between the height of the tree crowns and the height of the trees themselves in the stand, but these indicators are not identical (Magnussen and Boudewyn 1998). The measured ALS pulses in the crown of the tree are reflected at different heights. When analysing the ALS data, there is no guarantee that the higher points represent the tops of the trees themselves, because the density of the reflected ALS pulses is not sufficient to cover the entire area of the tree crown, therefore, statistical models must be used to predict the actual height of trees (Hollaus et al. 2009).

Previous studies show that using ALS data together with field measurements, the average forest height and other forest stand inventory parameters can be predicted with relatively high accuracy, a root mean square error (RMSR) is in the range of 3.7–6.4% (Næsset et al. 2004). For large areas, national forest inventory (NFI) data can be used in combination with ALS data to forecast forest heights (Hollaus et al. 2009). The uncertainty of the developed forest stand height models may be affected by different forest stand parameters and other conditions. The main parameters influencing the precision are the composition of tree species in the forest stand, vegetation season, tree canopy density, ALS point cloud density and their vertical distribution, number of echoes, etc. (Næsset 2005). Using ALS data collected over a long period of time and combining it with NFI data at the level of individual tree species on a large scale, Hauglin et al. (2021) achieved a RMSE of 12% for Lorey's height. Nilsson et al. (2017) also tested the effect of different grid cell sizes on forest stand height predictions over large areas and achieved a RMSE of 7.3% at a grid cell size of 20 × 20 m, which was recognized as the most suitable for large area mapping.

Latvia is located in the hemiboreal forest zone and both coniferous and deciduous species are widely represented in forests, so it is necessary to develop forest stand height models that are suitable for local conditions. The aim of the study is to develop separate forest stand height models for the territory of Latvia using ALS and NFI plot data for forest stands dominated by 6 main tree species. Currently, there are no such models available in the country, which would be based on such accurate data as NFI. The ALS data used in the study were obtained over a period of seven years using different ALS scanners under different vegetation season conditions. There is also problem of time difference between ALS and NFI measurements and limited amount of NFI sample plots for each stratified data set.

In this study, we used a statistical approach to develop forest stand height models based on the vertical distribution of ALS point cloud points and considering the effects of different tree species, different scanners and vegetation season. Alternatives for building forest stand height models using ALS data are local maxima (Melniks et al. 2019), inverted watershed (Miraki et al. 2021), or triangular irregular networks (Asner and Mascaro 2014) methods. These methods mostly rely on the determination of tree crown maxima and do not generally use ALS reflections from the lower layers of the forest stand.

Materials and methods

National Forest Inventory in Latvia

This study covers the entire territory of Latvia within the NFI sample plots. In Latvia, the NFI has been conducted since 2004 and includes more than 16,000 plots of land in forest land, agricultural land, inland waters and other land types. The plots are surveyed on a 5-year cycle, with 20% surveyed each year. The plots are arranged in a 4 × 4 km grid and 4 plots are located in each cell. Tree inventory parameters such as height, age, increment, quality and damage are determined for selected trees, representing all tree classes, based on their diameter at breast height. For more detailed information on NFIs in Latvia, see (Jansons and Licite 2010).

Table 1. The NFI sample plots by the dominant tree species

Tree species	NFI plot count
Black alder (*Alnus glutinosa* (L.) Gaertn.)	269
Grey alder (*Alnus incana* (L.) Moench)	342
Birch (*Betula pendula* Roth)	1268
Norway spruce (*Picea abies* Karst.)	893
Scots pine (*Pinus sylvestris* L.)	1433
European aspen (*Populus tremula* L.)	365

One NFI sample plot covers an area of 500 m^2 and the type of land use is determined by field measurements. If more than one type of land use is recorded within the boundaries of the sample plot, or significant differences are observed within the boundaries of one type of land use, the sample plot is divided into smaller sections. All undivided NFI sample plots in forest and agricultural lands were selected for the study, where the dominant tree species are Scots pine (*Pinus sylvestris* L.), Norway spruce (*Picea abies* Karst.), birch (*Betula pendula* Roth), black alder (*Alnus glutinosa* (L.) Gaertn.), European aspen (*Populus tremula* L.) or grey alder (*Alnus incana* (L.) Moench). The total study area consists of 4,570 NFI plots (total number of permanent NFI sample plots in Latvia is 16,173). The number of the NFI plots by individual tree species is shown in Table 1.

In the 4th cycle of the NFI (2019–2023), GPS coordinates were measured for the plots using a high-resolution GPS device. The Topcon GRS-1 receiver with a Trimble R1 external antenna were used to measure the coordinates and the data were post-processed in the Trimble GPS Pathfinder Office software package (KOREC 2023) using the LatPos network of base stations, thus obtaining an accuracy that is within 1 metre. At the time of data processing for current study, exact coordinate data were available for 2 of the 5 years of the NFI cycle. Accurate coordinate surveying indicated that until then the coordinates of the NFI plot centres were determined with an average accuracy of 2 metres.

ALS data

Aerial laser scanning in Latvia was performed from 2013 to 2019 and was carried out using 3 different scanners, viz. Leica ALS70-HP, Riegl LMS Q680i and Riegl LMS Q780i. Aerial laser scanning is performed independently of the NFI measurements and the area covered by the different scanners is shown in Figure 1. The points represent undivided NFI plots with forest cover. A total of 1,460 undivided NFI plots were scanned with a Leica ALS70-HP scanner, 3,075 undivided NFI plots were scanned with a Riegl LMS Q680i scanner, and 35 undivided NFI plots were scanned with a Riegl LMS Q780i scanner.

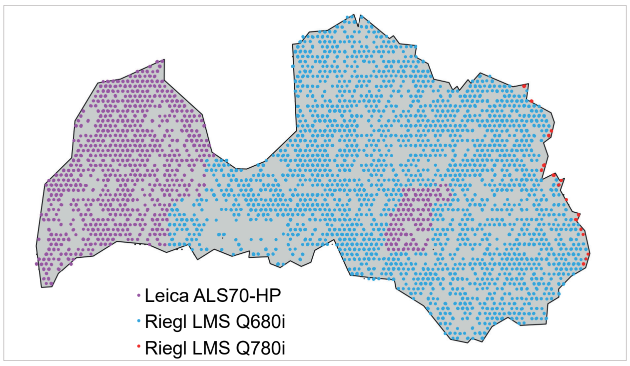

Figure 1. Density map of undivided NFI sample plots measured with different ALS

Data processing

The ALS data required for the study and their metadata layer were obtained from the Latvian Geospatial Information Agency (LGIA). Using FUSION/LDV software (McGaughey 2007), ALS point clouds were cut out along the boundaries of the NFI plots. For the NFI plots with precise coordinates, the LiDAR point cloud was extracted by a circle with a radius of 12.62 m (1,147 plots). For all the NFI plots, regardless of plot centre coordinate precision, the ALS point cloud was extracted with a radius of 14.62 m (4,570 plots). Two different data sets with different sample plot sizes have been created, because only part of the sample plots have been measured with high precision GPS antenna and the exact coordinates of their centres are known. For the rest of the sample plots, they are determined on average with an accuracy of 2 metres. During clipping of the ALS point cloud, the point cloud was normalized with the digital terrain model, which was created with the GridSurfaceCreate tool. Using the CloudMetrics tool, we obtained statistical information about the vertical distribution of ALS point cloud points. Different height percentiles such as p70, p75, p80, p90, p95 and p99 were calculated and used to develop stand height models. In addition, statistics on the vertical distribution of the ALS point cloud were obtained by selecting only points that are at least 1.5 m above the ground.

ALS measurements were taken during both the leaf-off and leaf-on period, as well as when the leaves are in bloom or are being shed during leaf fall by trees (Table 2). The earliest time of year for ALS measurements was week 15 (April), and the latest time was week 49 (December). Most of the data were obtained during the transition period between full leaf-on and leafless canopy conditions. ALS point density for the Leica ALS70-HP and Riegl LMS Q680i scanners are similar, while the data obtained with the Riegl LMS Q780i scanner have significantly higher point densities. All used scanners are capable of recording at least 7 individual returns per signal.

Table 2. ALS measurement technical details

Scanner	Leica ALS70-HP	Riegl LMS Q680i	Riegl LMS Q780i
Total NFI sample plots	1460	3075	35
Leaf-off canopy conditions (count)	155	603	0
Leaf-of canopy conditions (count)	493	626	0
Transition period (count)	812	1846	35
Scanning angle (degrees)	45	45	45
Flight altitude (m)	680–1500	650–1500	980
Min. point density (m^{-2})	2.95	2.29	11.49
Max. point density (m^{-2})	86.47	27.3	23.42
Average point density (m^{-2})	8.04	6.68	17.09
Average 1st return density (m^{-2})	5.69	4.64	8.6
Average 2nd return density (m^{-2})	1.83	1.46	5.42
Average 3rd return density (m^{-2})	0.35	0.15	2.06

The NFI plots from the 2nd, 3rd and 4th NFI cycles within ± 2 years with ALS measurements were selected for the data set to be used. Data on the height of the vertical distribution of the ALS point cloud by the 75th, 80th, 90th, 95th and 99th percentiles and the time of ALS data survey, based on the ALS metadata layer maintained by the LGIA, were added to the NFI plot information. This provided a mutually comparable database with information on the dominant tree species, tree height, ALS point height percentiles in all the NFI plots and ALS point height percentiles in the specified NFI plots.

The NFI plots, where the measurements in the field differed substantially from the ALS height percentile heights, were checked and those plots, where the field situation changed between the NFI measurements and the ALS measurements were discarded from the data set. Possible reasons for the changes in the plots are logging, damage caused by wind or snow, the establishment of new stands

Table 3. Measurement years of NFI and ALS

ALS	NFI									
	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
2013	32	36	28	28	34	0	0	0	0	0
2014	0	72	70	67	72	77	0	0	0	0
2015	0	0	99	110	138	101	132	0	0	0
2016	0	0	0	271	295	299	274	265	0	0
2017	0	0	0	0	120	131	117	106	93	0
2018	0	0	0	0	0	142	114	125	122	107
2019	0	0	0	0	0	0	233	243	205	212

forest height increment, too large error in the NFI plot centre coordinates, etc. The measurement time of the NFI plots with field measurements and with the ALS scanner is shown in Table 3.

Seasonal differences are defined according to the principle – leaf-on canopy period, leaf-off canopy period, or the transition state. The seasonal affiliation of each NFI plot was defined by the week number in a year. The leaf-off period includes weeks from 0–15 and from 43–52, the transition state includes weeks from 16–21 and 41–43, but the leaf-on period includes weeks from 22–40, respectively.

For statistical analysis of the data, four data sets have been created: (1) all the NFI plots with a radius of 14.62 m (further as the text goes, it is shown with label all_xx, where xx is substituted with height percentile identifier), (2) the NFI plots with precise coordinates of plot centres (radius is 12.62 m) (pre_xx), (3) all the NFI plots with a radius of 14.62 m with ALS points of at least 1.5 m above ground (allm_xx), and (4) the NFI plots with precise centre coordinates (radius is 12.62 m) and ALS points of at least 1.5 m above the ground (prem_xx). Statistical analysis of the data was performed in the R software (R Core Team 2019), NFI plots were selected according to various parameters (used type of ALS scanner, conifers or deciduous trees, vegetation season, dominant tree species, etc.). Selected datasets then were divided into model fitting and validation datasets in a ratio of 3 : 1. Linear regression models in each category were developed for all height percentiles of the ALS point cloud from the model fitting dataset. Only the height percentile from all four data sets that showed the highest coefficient of determination was used.

Model fitting and validation

Linear model parameters (model data volume, regression coefficient, slope and intercept values) and model validation coefficients (*RMSE* and *MAE,* mean average error) as well as a normal quantile graph indicating the level of normalized distribution of the data were analysed. The model parameters were validated and calculated for *RMSE* and *MAE* are as follows:

$$y = ax + b + \varepsilon \ , (1)$$

where:
y – the predicted value,
a – the slope of the line,
x – the given value,
b – the intercept of the line, and
ε – the residual error.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left(\hat{y}_i - y_i\right)^2}{n}} \ , \qquad (2)$$

where:
$\hat{y}_i$ – the predicted value of i-th observation,
y_i – the observed value for the i-th observation,
n – the total number of observations; and

$$MAE = \frac{1}{n} \sum_{i=1}^{n} y_i - \hat{y}_i \ , \qquad (3)$$

where:
n – the number of errors,
y_i – the predicted value, and
$\hat{y}_i$ – the actual value.

All the forest stand height models developed at a more detailed level were compared with the more general models to assess differences between model parameters and influencing factors, such as the effect of the used ALS scanner, vegetation season, or tree species. The normality, heteroscedasticity and independence test of residuals were analysed according to the Kolmogorov-Smirnov, Breusch-Pagan and Durbin-Watson tests, respectively, at 0.05 significance level. Finally, the maps of the sample areas were created using GridMetrics and the CSV2Grid tools in FUSION/LDV software for the forest height models of varying detail. The raster layer zonal statistics tool in the QGIS application (QGIS Development Team 2023) was used to compare the different obtained forest height maps with each other. The forest stands were stratified according to the dominant tree species, and the average predicted height value was calculated within the sample area. This value was calculated for the stand height models at the different levels of detail and compared with each other.

Results

Universal model

The best determination coefficient for the universal model is at the 90th percentile of the precisely measured NFI plots, and it is based on 1,147 plots. This indicates that accurate measurements of plot centres play an important role in the development of more accurate models. The graphic representation of the universal model is shown in Figure 2.

The coefficient of determination is 0.942. The validation results indicate that the height of the forest stand can

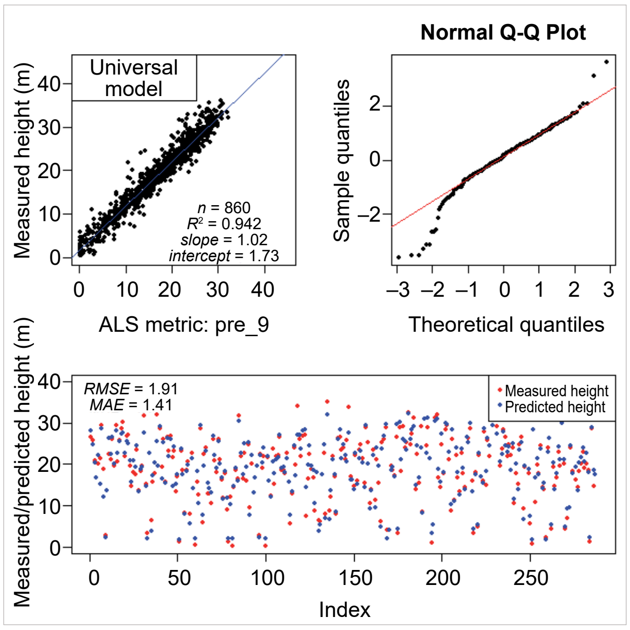

Figure 2. Parameters of the universal model

be determined with an average error of 1.41 m. The q-q plot shows that the absolute majority of the points correspond to the normal distribution. The model parameters for all other models and validation results (*RMSE* and *MAE*) are shown in Table 4. Abbreviations used in this table are as follows: Q780i stands for Riegl LMS Q780i, Q680i stands for Riegl LMS Q680i, ALS70 stands for Leica ALS70-HP, DEC stands for deciduous tree species, CON stands for coniferous tree species, LO stands for leaf-on canopy conditions, LF stands for leaf-off canopy conditions, TR stands for transition period, SP stands for Scots pine, NS stands for Norway spruce, B stands for birch, BA stands for black alder, EA stands for European aspen and GA stands for grey alder. Models with insufficient amount of data for model fitting were emitted from further analysis.

Comparing the differences of different ALS scanners within the universal forest stand height model, no statistically significant differences were observed ($p > 0.05$). In contrast, the time differences between NFI and ALS measurements and for conifers and deciduous trees are statistically significant ($p = 1.82e\text{-}05$).

Forest height models by individual ALS scanners

When splitting the data set by individual ALS scanners, small differences can be observed between the coefficients of determination of forest stand height models. The Riegl LMS Q780i scanner has the least data and the size data set is too small to draw conclusions. The performance of the Leica ALS70 scanner is also better than the universal model, R^2 values are 0.95 and 0.942, respectively, although the *RMSE* and *MAE* values are slightly higher. With a large amount of data, the most accurate models are still those prepared using NFI plots with the precise coordinates. Comparing the significance of the influencing factors of the model parameter, tree species such as birch, black alder and European aspen, tree types and seasonality have statistically significant differences within the model.

Comparison of the effects of different species and time differences between measurements on Riegl LMS Q780i data were not done, because this subsample does not have enough data and conclusions cannot be drawn. For the Riegl LMS Q680i and Leica ALS70, the data sets are larger and both show a statistically significant difference between deciduous and coniferous trees ($p = 0.0186$ and 0.0273, respectively). Comparing the significance of the influencing factors of the model parameter, tree species such as grey alder and European aspen, tree types and seasonality have statistically significant differences within the model.

Comparison of forest height models by conifers and deciduous dominated forest stands

Comparing the height model of the forest stands created on the Riegl LMS Q680i aerial laser scanner with the models created by dividing this data set into conifers and deciduous trees, an increase in the R^2 value in determining the height of coniferous forest stands can be noted. The value of the coefficient of determination increases from 0.939 to 0.955, while in the deciduous segment it decreases slightly and is 0.925. In the case of the coniferous model,

Table 4. Overview of different forest height model parameters

Model	Percentile	R^2	Slope	Inter-cept	*RMSE*	*MAE*
Universal model	Pre_90	0.942	1.02	1.73	1.91	1.41
Q780i	Allm_90	0.952	1.11	–0.694	1.68	1.29
Q680i	Pre_90	0.939	1.01	1.91	2.13	1.5
ALS70	Pre_90	0.95	1.04	1.34	1.95	1.47
Q680i-CON	Prem_90	0.955	1.05	0.071	1.49	1.12
Q680i-DEC	Pre_90	0.925	1.01	2.09	2.35	1.65
ALS70-CON	Pre_95	0.945	1.02	0.309	1.78	1.25
ALS70-DEC	Pre_95	0.946	1.03	0.628	1.82	1.46
Q680i-SP	Pre_95	0.972	1.04	0.12	1.27	0.93
Q680i-NS	Prem_90	0.943	1.01	0.56	1.83	1.56
Q680i-B	Pre_95	0.941	0.971	1.43	1.7	1.35
Q680i-EA	Pre_90	0.937	1.05	3.51	2.96	2.44
Q680i-BA	All_90	0.909	0.953	2.72	2.6	1.88
Q680i-GA	Pre_90	0.925	0.861	3.35	2.52	1.97
ALS70-SP	Pre_95	0.947	1.02	0.404	1.43	1.02
ALS70-NS	Pre_90	0.745	1.02	1.54	2.33	1.55
ALS70-B	Pre_95	0.961	1.03	0.768	2.16	1.68
ALS70-EA	Allm_90	0.907	1.11	0.394	2.5	1.92
ALS70-BA	Allm_80	0.907	1.04	1.46	1.73	1.43
ALS70-GA	All_90	0.845	1.01	0.788	2.55	1.97
Q680i-CON-LF	Prem_90	0.947	1.06	0.211	1.33	1.13
Q680i-CON-LO	Pre_90	0.968	0.998	1.46	2.05	1.5
Q680i-CON-TR	All_90	0.957	1.03	1.04	1.71	1.29
Q680i-DEC-LF	Pre_90	0.905	0.999	3.03	2.68	2.19
Q680i-DEC-LO	Pre_99	0.894	0.961	–0.637	2	1.44
Q680i-DEC-TR	Pre_95	0.922	1	1.18	2.01	1.51
ALS70-CON-LF	All_95	0.961	1.05	–0.501	2.17	1.56
ALS70-CON-LO	Pre_95	0.908	1.03	0.026	1.68	1.29
ALS70-CON-TR	Pre_95	0.956	0.994	0.871	2	1.16
ALS70-DEC-LF	Allm_90	0.86	1.12	–1.31	2.33	1.78
ALS70-DEC-LO	Pre_95	0.933	1.11	–0.763	1.82	1.6
ALS70-DEC-TR	Pre_95	0.969	0.984	1.41	1.47	1.07
Q680i-SP-LF	All_90	0.971	1.06	0.838	2.13	1.56
Q680i-SP-LO	All_95	0.973	1.05	–0.408	1.45	1.17
Q680i-SP-TR	All_95	0.972	1.03	–0.006	1.51	1.16
Q680i-NS-LF	All_90	0.863	1.06	0.968	2.36	1.73
Q680i-NS-LO	All_80	0.943	1.12	0.943	2.74	1.91
Q680i-NS-TR	Prem_90	0.933	0.994	1.03	1.62	1.28
Q680i-B-LF	All_95	0.867	1.03	0.7	3.85	2.41
Q680i-B-LO	All_95	0.921	1.03	0.228	2.09	1.62
Q680i-B-TR	Pre_95	0.951	0.995	1.11	1.62	1.37
Q680i-BA-TR	All_90	0.907	0.908	3.54	1.99	1.44
Q680i-EA-LF	All_90	0.92	1.07	3.63	3.33	2.52
Q680i-EA-LO	Allm_90	0.962	1.19	–1.32	2.86	2.12
Q680i-EA-TR	All_95	0.926	1.05	1.38	2.82	1.97
Q680i-GA-LF	All_90	0.867	0.996	1.64	3.82	2.95
Q680i-GA-LO	All_80	0.962	1.12	0.639	2.55	2.07
Q680i-GA-TR	Pre_90	0.893	0.896	2.8	1.41	1.13
ALS70-SP-LF	All_99	0.941	1.02	–1.94	1.57	1.1
ALS70-SP-LO	Pre_95	0.954	1.04	0.107	1.4	1.07
ALS70-SP-TR	Pre_95	0.944	1.01	0.467	1.49	1.14
ALS70-NS-LO	All_80	0.944	1.04	2.61	2.15	1.63
ALS70-NS-TR	Pre_95	0.935	0.955	1.41	1.18	1.02
ALS70-B-LF	All_95	0.894	1.1	0.399	2.25	2
ALS70-B-LO	All_90	0.937	1.09	0.927	2.14	1.49
ALS70-B-TR	Pre_95	0.967	0.99	1.18	1.77	1.54
ALS70-BA-TR	Allm_80	0.918	0.951	3.15	1.53	1.45
ALS70-EA-TR	Allm_90	0.965	1.17	0.037	3.05	2.37
ALS70-GA-TR	All_90	0.899	0.97	2.28	2.43	1.62

the *RMSE* and *MAE* values also increase, while in the case of deciduous trees, they decrease.

Comparing the performance of the Leica ALS70 aerial laser scanner, the values of the coefficient of determination, when splitting the model into coniferous and deciduous tree segments, decrease from 0.95 to 0.945 for conifers and 0.946 for deciduous trees, respectively.

Comparing the differences between different species and vegetation seasons within the models, it can be seen that in the coniferous models, for both the Leica ALS70-HP and the Riegl LMS Q600i, no significant differences are observed at either the species or seasonal level. In turn, in the case of deciduous tree models, there are significant differences between the performance of the models in different vegetation seasons, and, in the case of the Riegl LMS Q680i scanner, also at the level of different species.

Comparison of forest height models by tree species

Dividing the Riegl LMS Q680 model data set into groups by species shows an improvement in the performance of the model in the stands, where the dominant tree species are Scots pine, Norway spruce and birch. The coefficient of determination increases from 0.939 to 0.972, 0.943 and 0.941, respectively. The RMSE value also decreases in the validation results. For the numerically less represented species – black alder, European aspen and grey alder – the coefficients of determination of the forest stand height model in the case of the Riegl LMS Q680i ALS scanner are slightly lower than at the level of all species, 0.909, 0.937 and 0.925, respectively.

In the case of the Leica ALS70-HP ALS scanner, the height models of the forest stands of individual tree species show a higher coefficient of determination only in the case of birch (0.961 vs. 0.95); in the case of Scots pine and Norway spruce it is slightly lower, 0.947 and 0.945, respectively, while for black alder, European aspen and grey alder it is 0.907, 0.907 and 0.845, respectively. Validation results improve for Scots pine and black alder while decreasing for Norway spruce, birch, European aspen and grey alder.

Statistically significant seasonal influence affects the parameters of individual species forest stand height models for black alder and birch (in the case of Riegl LMS Q680i), while for Norway spruce and European aspen in the case of Leica ALS70-HP. At this stage, the forest stand height models for black alder, European aspen and grey alder are designed with a relatively small number of NFI sample plots, so a plausible seasonal effect on the performance of the models may not be observed.

Comparison of forest stand height models by conifers and deciduous trees depending on the season

It can be seen that the models generated from the data collected with the Riegl LMS Q680i ALS scanner from spring to autumn have a higher coefficient of determination than the previously discussed model without seasonal effects, but RMSEs and MAEs for the validation values are lower. For models based on Leica ALS70-HP data, R^2 values are better in the leaf-off canopy period and in the transition period, while in the leaf-on canopy period, this factor is significantly lower (0.908 vs. 0.945). The best results on the validation sample are shown by the data collected during the leaf-on canopy period and the transition period.

The results shown for deciduous tree models vary significantly from season to season. The R^2 values of the Riegl LMS Q6801 aerial laser scanner range from 0.894 to 0.922, while those of the Leica LMS70-HP ALS scanner range from 0.86 to 0.939. In all the developed deciduous seasonal models, the values of the coefficient of determination are lower than those of the deciduous models discussed above.

The seasonal conifer models based on the Riegl LMS Q680i ALS scanner data do not show significant differences between the Scots pine and Norway spruce datasets in any of the seasons, although a sufficiently large data set is only available for inter-seasonal data to reliably estimate differences. Also, for deciduous tree models, the data set is large enough only in the leaf-off season, and in this case, there is a significant difference between the effects of different species. The effect of conifer species on the Leica ALS70-HP models is also insignificant, while in the context of deciduous trees, there are statistically significant differences between some species during the leaf-off canopy period and in the transition period.

Comparison by species seasonally

For the Scots pine stands scanned with the Riegl LMS Q680i aerial laser scanner, the R^2 value is very high in all seasons (0.971–0.973), which is not significantly different from the model that does not consider seasonal effects (0.972), while the results of validation decrease slightly. In the case of the Leica ALS70-HP ALS scanner, the R^2 values range from 0.941 to 0.954, which is also similar to the non-seasonal model (0.947) and the validation results are similar.

In Norway spruce stands, results of the Riegl LMS Q680i scanner are seasonally different. While in the leaf-on canopy period and in the transition state R^2 values are very high, 0.943 and 0.933, respectively, in the leaf-off canopy period it is lower, – 0.863. This is probably due to the admixture of deciduous tree species in the forest stands, where Norway spruce is the predominant species. Validation coefficients compared to the Norway spruce stand model, which was not seasonally distributed, show better results only in the transition state between leaf-on and leaf-off canopy periods, while in other cases they are slightly worse. In the case of the Leica ALS70-HP ALS scanner, sufficient data for model development are available only during the leaf-on canopy period and in the transition period, and both models have relatively high values ($R^2 > 0.935$) but lower than those in the non-seasonal model. In turn, the results of the validation are better for transition period data.

In the forest stands, where the dominant tree species is birch, the seasonal models grounded on the Riegl LMS Q680i ALS scanner show better results in the transition and in the leaf-on canopy period ($R^2 > 0.92$), however, only in the leaf-on canopy period the result is better than in the undivided birch model, 0.951 and 0.941, respectively. The models based on the Leica ALS70-HP scanner data also show the worst results during the leaf-off canopy period, which can be explained by the difficulty for ALS rays to reflect against bare tree branches (R^2 – 0.894). In the transition period and during the leaf-on canopy period, the performance of the models improves significantly and R^2 values exceed 0.93.

In plots, where the dominant tree species is black alder, there are sufficient data to develop forest height models only in the transition period. The R^2 value for the Riegl LMS Q680i ALS scanner data-based model is 0.907, while for the Leica ALS70-HP data it is 0.918. These values are close to the black alder models, which were not divided into seasons, however, in both cases the validation result shows higher *RMSE* coefficients, so an improvement is observed.

In the plots, where the dominant tree species is European aspen, the models based on the Riegl LMS Q680i data show high results in all seasons, but in the leaf-on canopy period the results are very good ($R^2 = 0.962$), which is significantly higher than was shown by the non-seasonal model. In the validation data, the *RMSE* value exceeds 2.8 m in all cases and the *MAE* value is up to 2.5 m, indicating a relatively large variation in tree height predictions. For the models based on the Leica ALS70-HP data, the data set size is sufficient only for the transition period model, and the R^2 value reaches 0.965, which is significantly higher than for the non-seasonal model (0.907).

In the plots, where the dominant tree species is grey alder, models based on the Riegl LMS Q680i ALS scanner dataset show the best results during the leaf-on canopy period ($R^2 = 0.962$). While at other seasons the values of the coefficient of determination range from 0.867 to 0.893. Thus, a significant improvement of the coefficient of determination is observed in the leaf-on canopy period compared to the non-seasonally distributed model (0.925). In the Leica ALS70-HP ALS scanner dataset, the number of plots is sufficient for model development only in the transition period, and the R^2 value reaches 0.899, which is higher than in the non-seasonal model (0.845).

In order to visualize the obtained results, two maps have been created that show the height of the forest stands with different tree species. The area depicted in the Figure 3 was scanned with the Leica ALS70-HP ALS system in leaf-on canopy conditions and the tree species found in it are Scots pine, Norway spruce and birch. The left image shows forest stand spatial distribution with dominant tree species, but the right image shows predicted forest height according to the developed models. The

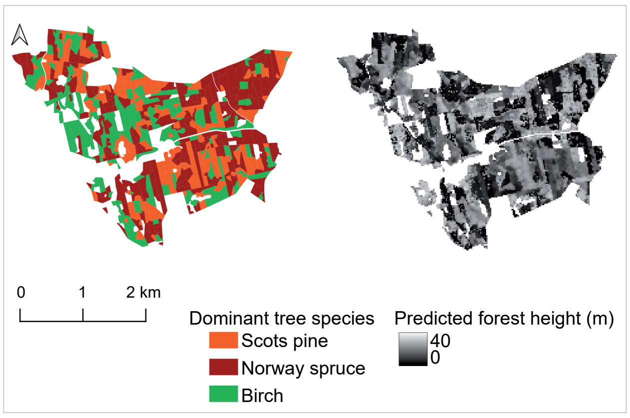

Figure 3. Forest height map for Scots pine, Norway spruce and birch dominated forests in leaf-on canopy conditions

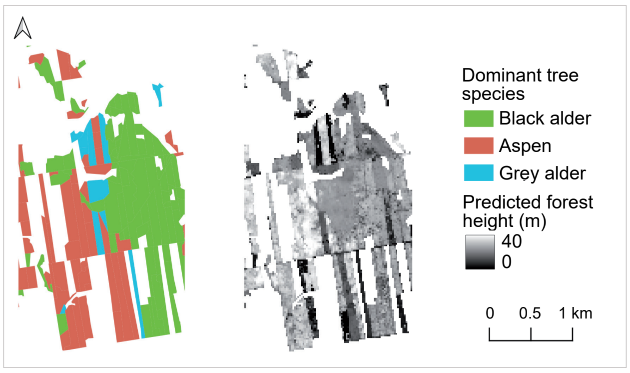

Figure 4. Forest height map for black alder, aspen and grey alder dominated forests in transition canopy conditions

area shown in the Figure 4 was scanned with the Riegl LMS Q680i ALS system and the tree species found in it are black alder, European aspen and gray alder. The horizontal resolution of the forest stand height maps are 20 × 20 m.

Table 5 shows the differences in the predicted heights of the forest stands in the sample locations, comparing them to the models derived in different levels of detail. The forest stands, in which the dominant tree species is Scots pine, the predicted heights are within 0.5 m when looking between the developed models of all details. In the forest stands, where the dominant tree species is Norway spruce, the differences in the predictions range from 1.4 to 2.2 m. Since in this case the situation is that the ALS datasets are acquired in transition period, i.e. during leaf unfolding, it can be concluded that the differences in the results could be influenced by the vegetation season. The biggest differences between the height differences of stands are observed in the stands, where the dominant tree species is European aspen. The smallest difference can be observed in comparison with the model developed based on the data divided at the scanner and species level (0.6 m), but in comparison with the universal model, the difference in the height of the forest stands exceeds the margin of 3 m.

Table 5. Forest height prediction differences between the most detailed forest height models predictions against the less detailed forest height prediction models

	N (pixel count)	Models by different scanners, tree types and seasons	Models by different ALS scanners and tree species	Models by different scanners and tree types	Models by ALS scanners	Universal model
Scots pine	4602	0.25	0.04	0.14	0.46	0.38
Norway spruce	7455	–2.12	–1.45	–2.24	–1.52	–1.64
Birch	4965	–0.92	–1.08	–0.94	0.34	0.25
Black alder	4710	–2.36	–0.12	–0.67	–0.5	–0.53
European aspen	3718	1.46	0.62	2.94	3.12	3.07
Grey alder	756	–0.14	1.78	1.27	1.45	1.51

Discussion

The national ALS campaign in Latvia used three different scanners and the average point density per square meter for different scanners ranges from 6.68 to 17.09. The vast majority of the area is scanned with the Leica ALS70-HP and Riegl LMS Q680i scanners, and the areas covered have similar point densities (8.04 and 6.68 points per square metre, respectively). The 1st, 2nd and 3rd reflections also have a slightly higher point density in the area covered by the Leica ALS scanner. The average point density in the areas scanned with the Riegl LMS Q780i is 17.09 points per square metre. Roussel et al. (2017) point out that the tree heights obtained from the ALS data are systematically lower unless high-density ALS data are available. Also, in this study, it can be concluded that the height models of forest stands developed in the areas, where the Riegl LMS Q780i scanner is employed, are more accurate, despite the fact that the amount of data is much smaller than in the areas covered by other scanners.

In various studies, the stand height is determined based on the 99th percentile (García et al. 2010, Singh et al. 2017), thus trying to determine the forest stand height based on its highest points, however, this approach is also subject to underestimation. In this study, by performing statistical analysis and creating forest stand height models with different phenological, tree species composition and different scanner data, we have shown that the 99th height percentile shows the best results in only two cases: Scots pine dominated forest stands in the leaf-off canopy period with the Leica ALS70-HP scanner and deciduous tree dominated forest stands in the leaf-on canopy period with the Riegl LMS Q680i scanner. In other cases, using the lower height percentile in the models, it is possible to determine the height of the forest more accurately.

The reflection height of ALS points changes due to phenology, so in this study we distinguished 3 different groups: leaf-off, leaf-on canopy and transition periods. Simonson et al. (2018) also points to this effect. If the height of ALS reflections in coniferous forests is similar in all vegetation seasons, then differences in the vertical distribution of ALS points are observed in deciduous forests. During the leaf-off canopy period, the pulses emitted by the ALS scanner are able to pass more easily through the tree crowns and are reflected in the lower layers of the forest stand, while as the leaf volume increases, the height of the reflections gradually increases. In this study, it can be observed that in the leaf-on canopy period in the deciduous forest stands the forest stand height can be most accurately determined using relatively higher height percentiles than in the leaf-off canopy period.

Determining the height of forest stands using ALS and NFI datasets is an accepted practice and various published results are found in the literature. Nord-Larsen and Riis-Nielsen (2010) used more than 2,000 NFI plots and ALS data with a minimum point density of 0.5 pulses per square meter and determined the dominant tree height with an *RMSE* value of 2.34 m. When the results were stratified by forest stands dominated by conifers, deciduous trees or mixed stands, the *RMSE* values were 2.05 m, 2.49 m, and 2.47 m, respectively. In this study, the *RMSE* value for the universal stand height model reached 1.91 m, and this improvement is probably due to the higher density of ALS points in the available laser scanning data set. However, it can also be observed in this study that the best *RMSE* values are observed in the forest stands dominated by conifers.

By using machine learning methods and combining LiDAR and hyperspectral data, Arjasakusuma et al. (2020) was able to achieve an *RMSE* value of 1.7 m, while using only ALS data, the *RMSE* value reached 1.82 m which is also comparable to our study, although on different forest types.

In order to apply the developed forest stand height models in practice, information on the spatial distribution of tree species within the study area is required. At the moment, such information is not available on a national scale in Latvia. The closest data source that can be utilised is the layer of forest stands maintained by the State Forestry Service, which stores information about state and private forests in the form of polygons, but it does not contain information about all areas covered with trees in the country. Therefore, it is necessary to develop a wall-to-wall map that contains classified tree species information for all areas in the country that are covered with trees.

Conclusions

In this study, the forest stand height models were developed using ALS data and NFI field data information. No statistically significant differences were observed between the two most representative data sets of ALS scanners, viz. Leica ALS70-HP and Riegl LMS Q680i, within the universal forest stand height model, however, when analyzing the data in a finer division, significant differences can be observed in the characteristics of the stand height models.

With a large amount of data in the development of forest stand height models, the best results are usually always achieved using only those NFI plots for which the coordinates of their centres have been measured with high precision. *RMSEs* for the universal model reached 1.91 m, while for the stratified models *RMSE* values were from 1.18 m to 3.85 m. As the accuracy of determining the coordinates of the centre of the sample plots increases, the accuracy of the developed stand height models also improves, as the models are built on spatially accurate datasets. This also reduces the impact of heterogeneous forest structure. As the amount of the precise coordinates of the NFI plot centres continues to increase during ongoing NFI cycle, it will be possible to develop even more accurate forest stand height models in perspective.

In the forest stands dominated by coniferous tree species, seasonal effects are observed in Norway spruce-dominated stands (*RMSEs* range from 1.18 to 2.74 m) possibly due to the admixture of deciduous trees. In contrast, lesser effect is observed in the Scots pine stands (*RMSEs* range from 1.4 to 2.13 m).

In the forest stands, where the dominant tree species are deciduous trees, a seasonal effect on the parameters of forest stand height models was revealed, and they are usually more precise during the leaf-on canopy period. *RMSE* value for the areas scanned with the Riegl LMS Q980i scanner range from 2 to 2.68 m and that for the areas scanned with the Leica ALS70-HP scanner range from 1.82 to 2.33 m, respectively. The accuracy of the best forest stand height models in the leaf-on canopy period is mainly obtained from the higher percentiles of the vertical distribution of ALS points than in the leaf-off canopy period.

The obtained results in combination with the growth models of different tree species can be used to refine the forest inventory data. Carrying out repeated ALS campaigns and regular updating of ALS data on a national scale could reduce the cost of field work in the inventory of forest stands in large areas.

Acknowledgements

This study was supported through funding within the scope of European Regional Development Fund project "Remote Sensing-based System for Forest Risk Factor Monitoring" (No. 1.1.1.1/21/A/040).

References

Arjasakusuma, S., Kusuma, S.S. and Phinn, S. 2020. Evaluating variable selection and machine learning algorithms for estimating forest heights by combining LiDAR and hyperspectral data. *ISPRS International Journal of Geo-Information* 9(9): 507; https://doi.org/10.3390/IJGI9090507.

Asner, G.P. and Mascaro, J. 2014. Mapping tropical forest carbon: Calibrating plot estimates to a simple LiDAR metric. *Remote Sensing of Environment* 140: 614–624; https://doi.org/10.1016/j.rse.2013.09.023.

García, M., Riaño, D., Chuvieco, E. and Danson, F.M. 2010. Estimating biomass carbon stocks for a Mediterranean forest in central Spain using LiDAR height and intensity data. *Remote Sensing of Environment* 114(4): 816–830; https://doi.org/10.1016/j.rse.2009.11.021.

Hauglin, M., Rahlf, J., Schumacher, J., Astrup, R. and Breidenbach, J. 2021. Large scale mapping of forest attributes using heterogeneous sets of airborne laser scanning and National Forest Inventory data. *Forest Ecosystems* 8(1): 1–15; https://doi.org/10.1186/S40663-021-00338-4/FIGURES/5.

Hollaus, M., Dorigo, W., Wagner, W., Schadauer, K., Höfle, B. and Maier, B. 2009. Operational wide-area stem volume estimation based on airborne laser scanning and national forest inventory data. *International Journal of Remote Sensing* 30(19): 5159–5175; https://doi.org/10.1080/01431160903022894.

Jansons, J. and Licite, I. 2010. Latvia. In: Tomppo, E., Gschwantner, T., Lawrence, M. and McRoberts, R.E. (Eds.) National Forest Inventories: Pathways for Common Reporting. Dordrecht (The Netherlands): Springer, p. 341–350.

KOREC. 2023. Trimble® GPS Pathfinder® Office software package [Computer software]. Survey Supplies Ltd., 34 44 Mersey View, Liverpool L22 6QB, United Kingdom. URL: www.korecgroup.com.

Magnussen, S. and Boudewyn, P. 1998. Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators. *Canadian Journal of Forest Research* 28(7): 1016–1031.

Maltamo, M., Malinen, J., Packalén, P., Suvanto, A. and Kangas, J. 2011. Nonparametric estimation of stem volume using airborne laser scanning, aerial photography, and stand-register data. *Canadian Journal of Forest Research* 36(2): 426–436; https://doi.org/10.1139/X05-246.

McGaughey, R.J. 2007. FUSION/LDV: Software for LiDAR Data Analysis and Visualization. U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Seattle, WA (USA), 119 pp. URL: https://w3.ual.es/GruposInv/ProyectoCostas/FUSION_manual.pdf.

Melniks, R., Ivanovs, J. and Lazdins, A. 2019. Estimation of dominant tree height in forest stands using LiDAR data. In: Proceedings of 18th International Scientific Conference "Engineering for Rural Development", 22–24 May 2019, Jelgava (Latvia). Jelgaba: LUA, p. 1667–1672; https://doi.org/10.22616/ERDev2019.18.N462.

Miraki, M., Sohrabi, H., Fatehi, P. and Kneubuehler, M. 2021. Individual tree crown delineation from high-resolution UAV images in broadleaf forest. *Ecological Informatics* 61: 101207; https://doi.org/10.1016/j.ecoinf.2020.101207.

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* 98(2–3): 356–370; https://doi.org/10.1016/J.RSE.2005.07.012.

Næsset, E., Gobakken, T., Holmgren, J., Hyyppä, H., Hyyppä, J., Maltamo, M., 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.

Nilsson, M., Nordkvist, K., Jonzén, J., Lindgren, N., Axensten, P., Wallerman, J. and Olsson, H. 2017. A nationwide forest attribute map of Sweden predicted using airborne laser scanning data and field data from the National Forest Inventory. *Remote Sensing of Environment* 194: 447–454; https://doi.org/10.1016/J.RSE.2016.10.022.

Nord-Larsen, T. and Riis-Nielsen, T. 2010. Developing an airborne laser scanning dominant height model from a countrywide scanning survey and national forest inventory data. *Scandinavian Journal of Forest Research* 25(3): 262–272; https://doi.org/10.1080/02827581.2010.486000.

QGIS Development Team. 2023. QGIS application, version 3.30.2 [Computer software]. URL: https://qgis.org/en/site/.

R Core Team 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

Roussel, J.R., Caspersen, J., Béland, M., Thomas, S. and Achim, A. 2017. Removing bias from LiDAR-based estimates of canopy height: Accounting for the effects of pulse density and footprint size. *Remote Sensing of Environment* 198: 1–16; https://doi.org/10.1016/j.rse.2017.05.032.

Simonson, W., Allen, H. and Coomes, D. 2018. Effect of Tree Phenology on LiDAR Measurement of Mediterranean Forest Structure. *Remote Sensing* 10(5): 659; https://doi.org/10.3390/rs10050659.

Singh, M.P., Bhojvaid, P.P., Ashraf, J. and Reddy, S.R. 2017. Forest transition and socio-economic development in India and their implications for forest transition theory. *Forest Policy and Economics* 76: 65–71; https://doi.org/10.1016/j.forpol.2015.10.013.

Vauhkonen, J., Maltamo, M., McRoberts, R.E. and Næsset, E. 2014. Introduction to Forestry Applications of Airborne Laser Scanning. In: Maltamo, M., Næsset, E. and Vauhkonen, J. (Eds.) Forestry Applications of Airborne Laser Scanning. Managing Forest Ecosystems, Vol. 27. Dordrecht (The Netherlands): Springer, p. 1–16; https://doi.org/10.1007/978-94-017-8663-8_1.