

Article

Planning of Commercial Thinnings Using Machine Learning and Airborne Lidar Data

Tauri Arumäe ^{1,2,*}, Mait Lang ^{1,3}, Allan Sims ¹  and Diana Laarmann ¹

¹ Institute of Forestry and Engineering, Estonian University of Life Sciences, 51006 Tartu, Estonia; mait.lang@emu.ee (M.L.); allan.sims@emu.ee (A.S.); Diana.Laarmann@emu.ee (D.L.)

² Estonian State Forest Management Centre, Mõisa, Sagadi Village, 45403 Haljala Municipality, Estonia

³ Tartu Observatory, University of Tartu, Observatooriumi 1, 61602 Tõravere, Estonia

* Correspondence: tauri.arumae@rmk.ee

Abstract: The goal of this study was to predict the need for commercial thinning using airborne lidar data (ALS) with random forest (RF) machine learning algorithm. Two test sites (with areas of 14,750 km² and 12,630 km²) were used with a total of 1053 forest stands from southwestern Estonia and 951 forest stands from southeastern Estonia. The thinnings were predicted based on the ALS measurements in 2019 and 2017. The two most important ALS metrics for predicting the need for thinning were the 95th height percentile and the canopy cover. The prediction accuracy based on validation stands was 93.5% for southwestern Estonia and 85.7% for southeastern Estonia. For comparison, the general linear model prediction accuracy was less for both test sites—92.1% for southwest and 81.8% for southeast. The selected important predictive ALS metrics differed from those used in the RF algorithm. The cross-validation of the thinning necessity models of southeastern and southwestern Estonia showed a dependence on geographic regions.

Keywords: forest management planning; machine learning; airborne lidar; commercial thinning; sparse lidar point clouds



Citation: Arumäe, T.; Lang, M.; Sims, A.; Laarmann, D. Planning of Commercial Thinnings Using Machine Learning and Airborne Lidar Data. *Forests* **2022**, *13*, 206. <https://doi.org/10.3390/f13020206>

Academic Editor: Henning Buddenbaum

Received: 22 November 2021

Accepted: 25 January 2022

Published: 29 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Forest management planning and decision making are mainly based on forest inventory (FI) data. It is still common for FI data to be collected by specialized personnel via field survey, but today these data are increasingly obtained via remote sensing [1–8]. In addition to stereo- and orthophotos, airborne laser scanning (ALS; [9]) has gained a leading role as an FI data source, as it is the main basis for describing forest structure in remote-sensing-based inventories. However, with the large number of metrics and combinations of data sources, the need for machine learning algorithms to extract significant information from large datasets has increased [10,11]. There are many machine learning algorithms, among which the most used are k-NN [12], k-MSN [13], various types of neural networks [14], and random forests [15].

Forest management in Estonia is carried out at the stand level. A forest stand is the elementary management unit of a forest area. A stand is defined as a small area of homogeneous forest delineated on a map as a polygon, and is described by forest mensuration parameters—age, site type, height, growing stock, basal area, etc. [16]. During the fieldwork of an FI, expert suggestions for management (e.g., thinning, sanitary cutting) can also be made; however, since large areas need to be covered, and the average revision cycle of stands is 5–10 years, management decisions are made based on somewhat outdated FI data. The management plan includes commercial thinnings, which are planned based on stands' relative density and age; clear-cuts are planned according to the combined rule of forest age and mean tree diameter at breast height (DBH), and pre-commercial thinnings are planned based on DBH and field inspections [17].

Commercial and pre-commercial thinnings are carried out with the goal of providing the best growing conditions for the remaining trees and maximizing their growth potential and wood quality [18,19]. It has also been stated that the correct timing of the commercial thinning has a positive effect on the growth potential and development of the stand [20]. Thinnings are said to improve the mechanical stability of the stands [21], and are foremost a silvicultural practice carried out from an economic standpoint of improving the quality of the timber at final felling [19].

Therefore, an up-to-date description of the forest stands is required for optimal and timely management—especially in the young stands, when growth is rapid and changes in structure are fast. Such timely information is best obtained using remote sensing data, but decision making and interpretation of large amounts of data from different technologies is time-consuming and complex. Machine learning has gained users along with traditional statistical methods, as indicated by success in combining and studying large amounts of data through supervised learning without the need to fulfil assumptions about distributions of the data values. Machine learning in forestry is mainly used for monitoring forest ecology, species distribution, carbon stock, natural hazard prediction, and estimation of forest structure variables [22].

The goal of this study was to test and build a supervised machine learning model to predict the need for commercial thinning of forest stands for practical forest management planning. Descriptive stand metrics were calculated based on nationwide sparse ALS point clouds and used with the random forest (RF) machine learning algorithm to predict the need for commercial thinning. Results from RF were also compared with a general linear model (GLM) predicting the same thinning need.

2. Materials and Methods

2.1. Forest Inventory and Management Data

Forest inventory (FI) data from the Estonian State Forest Management Centre (RMK) database were used—specifically of 1053 forest stands from southwestern Estonia and 951 forest stands from southeastern Estonia (Figure 1). The terrain in southwestern Estonia is mostly flat, with slightly more variation in southeastern Estonia. The average site index (H_{100}) in the southeast test site was 29.0 m, and the main dominant tree species were Scots pine (*Pinus sylvestris* L., 346 stands), Norway spruce (*Picea abies* (L.) H. Karst, 326 stands), and birch (*Betula pendula* Roth and *Betula pubescens* Ehrh., 212 stands). The main forest site types according to the Estonian classification system [23] were *Oxalis* (235 stands) and *Oxalis-Myrtillus* (164 stands). The average H_{100} for the southwest area was 27.9 m, and the most common dominant tree species were similar to those in the southeast area—Scots pine (356 stands), Norway spruce (332 stands), and birch (322 stands). The prevailing forest site types were *Filipendula* (189 stands), *Oxalis-Myrtillus* (187 stands), *Myrtillus* (96 stands), and *Aegopodium* (89 stands). The mixed-species forests usually have a second layer of Norway spruce, which is more common in the fertile site types.

Commercial thinnings by the RMK are usually planned for stands with a relative density above 75%, considering the dominant tree species, stand age, and expected time remaining to the final felling. Decisions for the thinnings are then made based on the FI data and confirmed via field visits. For our study, we took the thinning data for one year before and one year after the ALS measurements, allowing us to construct a list of the following training stands: (1) stands thinned after the flight of ALS, and assumed to have undisturbed structure between the ALS measurements; and (2) stands that were thinned before the flight of ALS. The first set of stands, which were thinned after the ALS measurements, were labelled as “1”, representing stands where a thinning should be carried out according to forest inventory experts. The second set of stands, which were thinned up to a year before the ALS measurements, were labelled as “0”—training data of stands that do not need to be thinned. Additional stands for the class “0” had to be randomly selected from the RMK’s FI database because initial tests with RF predicted that thinning was necessary in bog areas, young stands, stands growing on poor soil, and for the stands already reaching the age of

clear-cutting. The additional stands were selected to be greater than 1 hectare and with a relative density below 75%, or stands reaching the clear-cut stage and young stands (age less than 30 years). The query resulted in 207 additional stands for the southwest area and 199 stands for the southeast area.

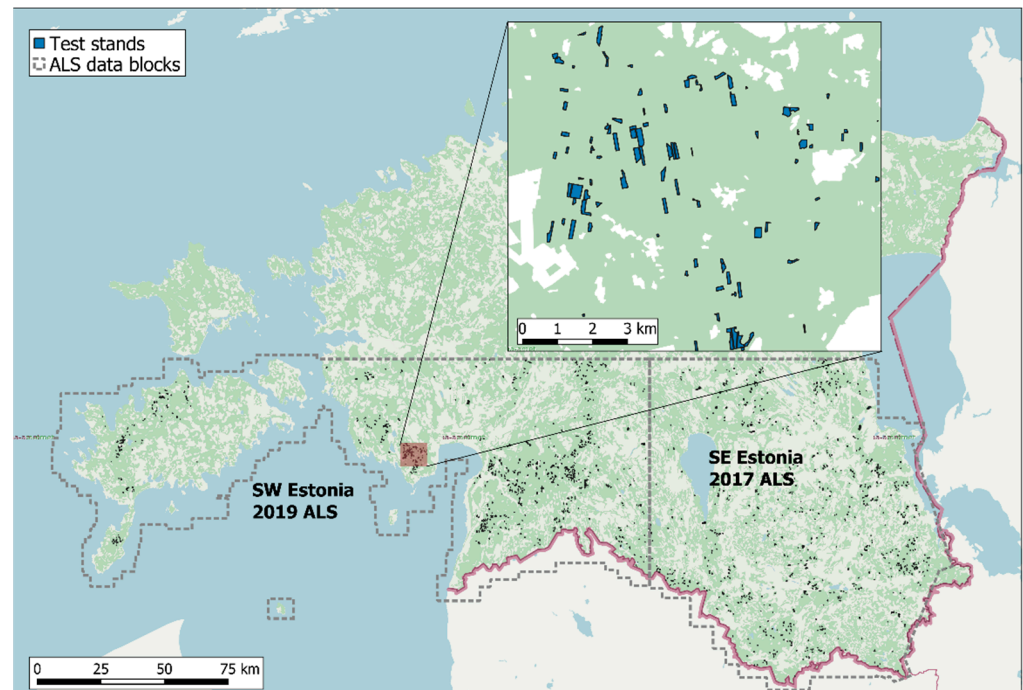


Figure 1. The airborne laser scanning (ALS) areas in 2017 in southeastern Estonia and 2019 in southwestern Estonia. Indication of the stands used for machine learning experiments.

2.2. Airborne Lidar Data

The Estonian Land Board carried out ALS measurements in southeastern Estonia in 2017 and in southwestern Estonia in 2019 after the final leaf unfolding [24], when the foliage of deciduous trees was fully developed. A RIEGL VQ-1560i scanner [25] was used from a flight altitude of 3100 m, which provided an average ALS data point density of 0.8 m^{-2} .

ALS data were processed using FUSION/LDV [26]. Stand borders were flanked with a buffer of 10 m to reduce border errors, and then the ALS data point clouds were extracted for each stand. The ALS point clouds were normalized using the digital terrain model constructed by the Estonian Land Board. A stand-based canopy cover proxy was calculated using a threshold of 1.3 meters ($CC_{ALS,1.3}$; Equation (1)), similar to that used by Arumäe and Lang [7]:

$$CC_{ALS,z} = \frac{100 \cdot P(h_p > \zeta)}{P}, \quad (1)$$

where h_p is the pulse return height above ground, P is the number of echoes, and ζ is the set threshold. Other metrics and height percentiles (H_{px}) were calculated, with points excluded below $h_p \leq 1.3 \text{ m}$, and canopy cover was also calculated, with thresholds set at mean height. The $CC_{ALS,1.3}$ was substantially higher for stands that needed to be thinned (Table 1) for both test sites, and the difference was visually discernible on point clouds (Figure 2). The forest height indication metric H_{p95} at both test sites was lower for stands that needed to be thinned.

Table 1. The average of the 95th height percentile (H_{P95}) and canopy cover ($CC_{ALS_1.3}$) based on thinning necessity and test site. Standard deviation is given in brackets.

Test Site	Thinning Necessity	Number of Stands	ALS Metrics	
			Canopy Cover (%)	H_{P95} (m)
Southwestern Estonia	Yes	637	85.1 (9.2)	18.2 (3.6)
	No	416	71.9 (12.9)	20.1 (5.7)
Southeastern Estonia	Yes	360	89.1 (7.8)	20.5 (3.5)
	No	591	71.1 (27.1)	25.4 (6.3)

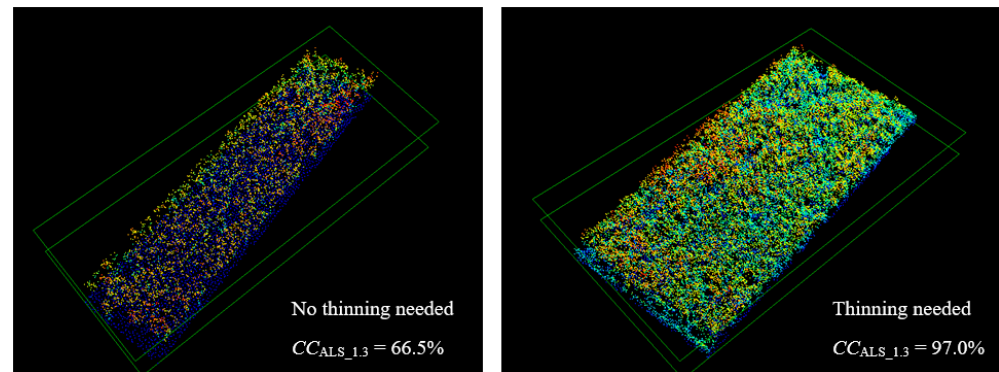


Figure 2. Example of two point clouds from southeastern Estonia.

2.3. Random-Forest-Based Model Construction

The RF machine learning package in R [27] was used for constructing the model for thinning necessity. The main parameters for model optimization in the R package randomForest are the number of feature variables at each split (RF_{mtry}) and the number of trees to grow (RF_{ntree}). With fixed values of $RF_{mtry} = 5$ and $RF_{ntree} = 500$, the first stage of machine learning was to select a suitably small set of the ALS metrics with significant predictive value; this was done in order to reduce the time of further processing. We excluded the ALS intensity metrics and the height metrics that were highly correlated with one another before the metric selection, leaving 41 ALS metrics for the search of the most significant metric. The final selection of ALS metrics was made based on the mean decrease in accuracy (MDA; [28]) and mean decrease in Gini coefficient (MDG; [29]).

The southeast and southwest datasets were then randomly split into two subsets, leaving the share of 70% and 30% for the model training and validation samples, respectively. The training was run 10 times on 70% randomly selected stands and validated on the rest of the stands. The final model parameter tuning was carried out using the R Caret package [27] to find the optimal values for RF_{mtry} , RF_{ntree} , and the other two parameters—the minimum number of observations to construct a separate terminal node ($RF_{nodesize}$), and the maximum number of terminal nodes ($RF_{maxnodes}$). The parameter values were varied, and optimal values were chosen based on the accuracy (k-fold cross-validation; Caret package method = “repeatedcv”). This procedure was repeated with the 5 pre-selected ALS metrics for both the southwest and southeast areas 10 times, as the randomness of training stands gives somewhat different accuracy each time.

For cross-validation of the area-based model, we selected all of the stands exclusively from the southeast or southwest Estonia test sites, and fitted the models using the previously estimated optimal values of the parameters (RF_{ntree} , RF_{mtry} , $RF_{maxnodes}$, $RF_{nodesize}$). These models were then applied to the neighboring area stands—the southeast model to the southwest stands and the southwest model to the southeast stands; this was done in order to assess whether the models were area-specific, or whether a universal model could be used. Additional tests were carried out by introducing errors to the training data, by swapping the decision from one (thinning needed) to zero (no thinning required)—or vice

versa—firstly on 30 stands, and then by increasing the falsely assigned stands to 30% of the training stands.

2.4. General-Linear-Model-Based Prediction

For comparison with the RF model, we applied the general linear model (GLM) method to the thinning data, as follows:

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \dots + \beta_n \cdot X_n + \varepsilon \quad (2)$$

where β_0 is the model intercept, $\beta_1 \dots n$ represents the slope coefficients, $X_1 \dots n$ represents the variables (ALS metrics), ε is the model residual error, and Y is the dependent variable (thinning necessity, with the similar coding values of 0 and 1). As with the RF model, the most significant ALS metrics were selected according to p -value, excluding intercorrelated metrics or the metrics with no logical meaning to thinning necessity (i.e., total number of echoes). The GLM model (2) coefficients for both southeastern and southwestern Estonia were fitted on 70% randomly selected stands, and then validated on the remaining 30% of observations. For the validation, the predicted log-odds of the model were calculated as a probability of thinning necessity. The probability was then classified as a Boolean decision of need for thinning, with probability greater than or equal to 0.5 targeted for thinning, and a probability less than 0.5 for the forest stands where no thinning was needed.

3. Results

3.1. ALS Metrics

Based on the random forest algorithm's variable importance indicators MDG and MDA, the five most significant ALS metrics were independently selected for both test areas. The two most significant predictive feature variables in both test areas were $CC_{ALS_1.3}$ using only the first echoes and the 95th percentile of point cloud height distribution (H_{P95}), with both showing substantially greater importance than the following three metrics. The fifth ALS metric accounted for five times less of the importance compared to the top two metrics. For both test sites—southwestern and southeastern Estonia—a lower percentile (H_{P20} and H_{P25}) and the median of the absolute deviations from the pulse return height mode value (H_{MAD_mode} ; [26]) were among the five most significant metrics (Table 2).

Table 2. The ALS metrics with the greatest predictive power for assessing the need for thinning using the random forests algorithm in southwestern and southeastern Estonia. MAD is the median of the absolute value.

Test Site	ALS Metrics	Mean Decrease in Accuracy	Mean Decrease in Gini
Southwestern Estonia	Canopy cover	138.4	115.8
	95th height percentile	56.1	65.4
	Height MAD mode	42.6	29.1
	Coefficient of height variation	23.3	17.6
	25th height percentile	19.6	15.6
Southeastern Estonia	Canopy cover	99.5	109.5
	95th height percentile	79.8	111.4
	Height MAD mode	36.8	34.3
	Height skewness	26.4	30.3
	20th height percentile	20.9	29.1

3.2. Random Forest Model Optimization

The crucial step in constructing the model was to select the best combination of RF model parameters to potentially increase the accuracy of the prediction. We first used the default values of RF (the RF algorithm in R uses $RF_{ntree} = 500$ and RF_{mtry} value as

the square root of the number of predictor variables rounded down), and then varied the RF_{ntree} from 500 to 2500 with a 500 step and RF_{mtry} from one to five. The highest accuracy was found using RF_{mtry} set at 2 and RF_{ntree} at 1000 (Figure 3) for southwestern Estonia, but based on the validation stands, the accuracy values did not significantly differ for other values of RF_{ntree} or RF_{mtry} (91.4%–92.0%; Figure 3). The southeastern Estonia model behaved similarly, although the overall accuracy was smaller (84.4%–85.0%) but, similarly to the southwest test site, the influence of the model parameter RF_{ntree} and RF_{mtry} values was not proven to be significant with regards to model accuracy. The optimal values for the southeast area were RF_{mtry} of 2 and RF_{ntree} of 1500.

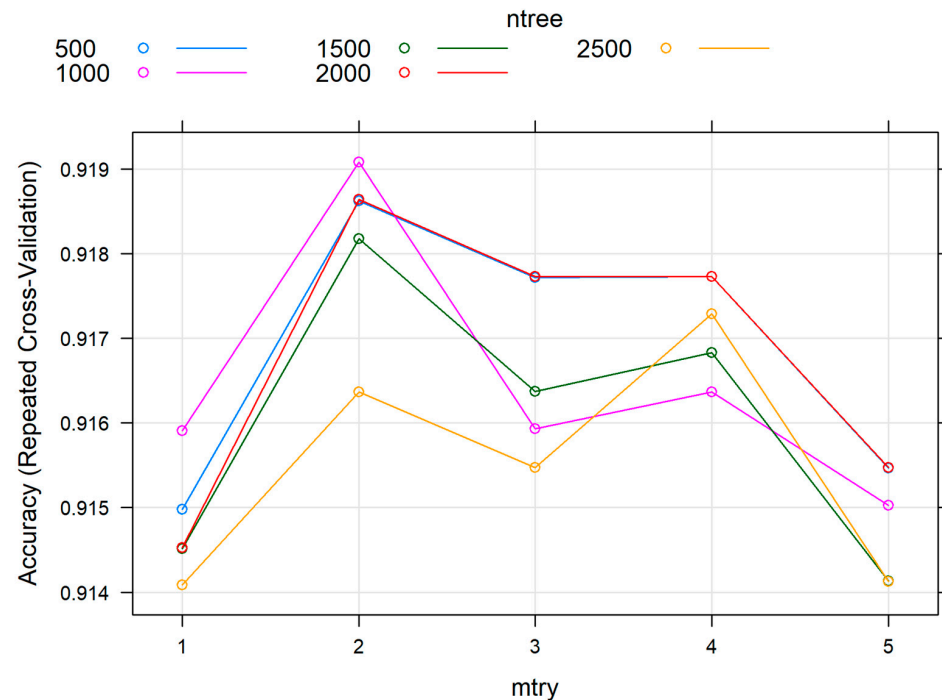


Figure 3. Random forest model training accuracy for the southwestern Estonia dataset, influenced by varying the number of variables at each split ($mtry$) and the number of trees to grow ($ntree$).

With the RF_{mtry} and RF_{ntree} optimal values fixed, the next parameters we varied were $RF_{maxnodes}$ (search from 2 to 50, by 1) and $RF_{nodesize}$ (search from 1 to 20, by 1). The accuracy did not increase significantly compared to the given RF default values; the highest accuracy of 92.2% for southwestern Estonia was shown using $RF_{maxnodes}$ set at 46 and $RF_{nodesize}$ at 9. Similar to the RF_{ntree} and RF_{mtry} variation, no significant impact on model accuracy was found (91.3%–92.2%). In southeastern Estonia, the accuracy was less than in southwestern Estonia, but varying the $RF_{maxnodes}$ and $RF_{nodesize}$ did not significantly affect the accuracy (84.3–85.5%). The best accuracy was obtained with $RF_{maxnodes}$ set at 43 and $RF_{nodesize}$ at 8. When using the RF model with default values, the accuracy of the models was similar, but slightly higher compared to our selected parameters.

3.3. Validation and Decision Error Sensitivity Test

The validation process was repeated 10 times at both test sites, randomly selecting 70% of the stands for training data, and validating the model on the other 30% of the stands accordingly. The southwest model showed an average classification accuracy of 91.9% (87.1%–94.0%) using our selected parameter values ($RF_{mtry} = 2$, $RF_{ntree} = 1000$, $RF_{maxnodes} = 46$, $RF_{nodesize} = 9$), and 93.5% (91.2%–95.3%) when using the default parameter values chosen by the RF algorithm. The southeast model's average classification accuracy over 10 runs was 85.2% (81.8%–87.4%) using the parameter values we selected ($RF_{mtry} = 2$,

$RF_{ntree} = 1500$, $RF_{maxnodes} = 43$, $RF_{nodesize} = 8$, and 85.7% (83.2%–89.2%) for the default values of RF.

Cross-validating the southwest model and applying it to the southeast stands showed an accuracy of 80.3%. Vice versa, the southeast model applied to southwest stands showed a slightly higher accuracy of 83.1%. Considering the possible confidence intervals, the prediction accuracy was no different; however, as might be expected, both test-area-specific models performed worse on the other area than on the test site with the model empirical data.

In the forced estimation error test, in which field decisions of 30 randomly selected stands from the training set in the southwest test site were altered, the final prediction accuracy dropped by 10% on the validation set. Increasing the number of incorrect decisions on 30% of the training set stands produced a model with prediction accuracy of only 21%. The results in the southeast test site were similar when thinning necessity errors were introduced into the training data.

3.4. General-Linear-Model-Based Predictions

The three most significant (p -value < 0.05) ALS metrics for the southwest area were the coefficient of height variation (H_{Var}), H_{P50} , and the first echo-based canopy cover, calculated above the mean height threshold (CC_{ALS_mean}). The prediction accuracy on the validation set when run 10 times with the random set of training stands was 92.1% (89.9%–95.3%) for the southwest area (Table 3). The most significant ALS metrics for the southeast area were the ALS point cloud mean height (H_{Pmean}), the kurtosis of the point cloud (H_{Kurt}) and, again, CC_{ALS_mean} . The prediction accuracy based on the validation stands in the southeast area was on average 81.75% (77.3%–85.3%; Table 3). The excluded ALS metrics were either correlated with the three already-selected metrics, or had p -values greater than 0.05. When the significant ALS metrics from the southeast area were chosen for the southwest—and vice versa—for model construction, the accuracy did not significantly differ from the area-specific prediction accuracy (p -value > 0.05). Using the ALS metrics selected with the RF algorithm (Table 2) and with the GLM, the prediction had a slight decrease in prediction accuracy—3.5% in southeastern Estonia, but not significantly in the southwestern Estonia test site.

Table 3. General linear models for thinning necessity prediction based on ALS metrics and their estimated values of model parameters (p -values given in brackets) for the southwest and southeast test sites.

Test Area	Model Accuracy	ALS Metric	Parameter Value
Southwestern Estonia	92.1%	50th height percentile	−0.04817 (0.004)
		Coefficient of height variation	−0.00209 (0.001)
		Canopy cover above mean height	0.02321 (0.001)
		Intercept	−0.28000 (0.045)
Southeastern Estonia	81.8%	Mean height	−0.04887 (0.004)
		Kurtosis of height	0.06119 (0.011)
		Canopy cover above mean height	0.01193 (0.001)
		Intercept	0.31203 (0.088)

4. Discussion

Forest management planning and decision making can be guided by several objectives—maximizing ecological or economic values, concentrating on social needs and, in most cases, the need to balance all of them. Decisions are usually made based on measurable parameters, i.e., forest age, mean tree diameter at breast height, additional constraints arising from neighboring stands, or standing wood volume; all of this requires analysis of large quantities of data. In the Estonian State Forest Management Centre (RMK), for example, the total thinning area in 2018 was around 10,000 ha, with an average stand being two hectares in size [30]. The total area of commercial thinnings for each year is limited by the annual allowable cut

given to RMK, and is then divided for each management region separately based on the regional forest inventory data. The final decision of which FI data-based pre-selected stands are thinned is made during final field visits by the regional foresters. This study offers a novel approach for planning commercial thinnings using a machine learning algorithm alongside ALS-based structure metrics in order to most effectively utilize the information with the highest acquisition cost—the decisions made in situ by forest inventory experts.

The two most descriptive ALS metrics using the RF model for thinning necessity were the 95th height percentile and canopy cover, which can also be defined as stand height and stand relative density; these are also the main criteria used by the RMK for indicating the need for thinning. Additionally, forest height is a good indicator of stand age [31], which is one of the criteria for thinning assignment. The skewness, variation in height, and other metrics describing the point distribution can be related to the density of tree crowns and the length of crowns [32], showing the competition for light and, thus, being another indication of thinning necessity. The accuracy for both test sites was surprisingly high, and the effort to further increase the accuracy by fine-tuning the random forest algorithm's parameters yielded no significant improvement in prediction accuracy. With small differences between the linear model prediction accuracy and RF accuracy, the RF model outperformed the linear model approach.

An indication that the prediction accuracy was approaching its maximum with a risk of overfitting was addressed by inserting a small error into the model training data, after which the accuracy dropped significantly by 10%. Overfitting could be addressed by setting the $RF_{nodesize}$ to be at least three, which would result in a small decrease in model training accuracy but would yield more reliable results for later application to the target area. Another issue is the assumption of homogeneity in tree cover within the forest stands. Although, by definition, forest stands should be homogeneous, in Estonian semi-natural forests we find large variation in canopy cover, tree height, soil properties, terrain, etc., within a single stand. However, the stands still are different from neighboring stands, and a somewhat subjective delineation is made in the forest. An option to better target the areas where thinning is needed would be to use a smaller unit than a stand and to apply the RF model to the new subunits. The solution would be a pixel-based approach with clustering and segmenting the pixels first according to their ALS metrics, and the segments could then be used as new possible silvicultural treatment polygons.

The later implementation of the stand-level RF model for nationwide application showed a need to diversify the training data sample, as the model was still assigning thinnings to stands belonging to the class ready for final felling, or to stands that had not reached the age of commercial thinning. Indication of thinning necessity in younger stands than those at commercial thinning age could be useful from the forest management point of view; however, our input training dataset did not include enough young stands and, therefore, the model predictions may have large uncertainties. In addition to training data diversification, an even more practical solution would be to apply the stand-age-dependent post-process filter in the prediction of thinning. Similarly to age restrictions, the model predicts thinnings for stands with management restrictions, remote areas out of management, stands already thinned after the ALS flight, etc., all of which could be filtered using the forest management inventory data. According to our experience, it is not feasible to predict such limitations using remote sensing data or add corresponding variables to the model. During its application, the model should also be developed as a continuous project, where field-expert-corrected predictions are added from time to time to the training data, thus improving the new version of the predictions.

The average relative density was 82 in the stands with classification errors during later model applications. Considering that the relative stand density for assigning thinning in the RMK is 75, we can conclude that the falsely assigned thinnings in many cases are borderline decisions. Some errors occurred in stands thinned, along with other thinnings carried out in that region in stands that do not necessarily need to be thinned at the time, but would need to be in the upcoming years. Such an approach is common for optimizing

the cost of silvicultural treatments for the purposes of not having to return to the same area in upcoming years, saving expenses on logistic operations.

With fieldwork becoming more and more expensive, such pre-targeted areas are of great assistance for time management and workflow planning. Another aspect is that the forest inventories are carried out at a 5–10-year interval, compared to the Estonian cycle of ALS data of four years. With the ALS data freely available, prediction of thinning necessity using machine learning has great potential for becoming a new method of forest management planning, and could be used for other silvicultural treatments such as tending of seedling stands.

Our study showed that due to regional differences in management practices, subjective decision making, forest growth, forest types, and species composition, the prediction accuracy of random forest models was also region-specific. It must be noted that such regional differences exist already in model input data, and may be an indicator of the subjective decision making of the regional foresters. Therefore, the continuously updated machine learning model could also homogenize the silvicultural practices between different regions and foresters.

5. Conclusions

The random-forest-algorithm-based model showed sufficient accuracy for simulating a forest manager, planning the commercial thinnings with an average of 85% accuracy in the southeastern Estonian test site and an average of 91% accuracy in the southwestern Estonian test site. The random-forest-based model outperformed the general linear model's thinning prediction by 4% in southeastern Estonia, but showed similar accuracy in southwestern Estonia. Both thinning necessity predictions by the random forest and general linear models showed some dependence on geographic region when cross-validated between the two test sites in southwestern and southeastern Estonia.

Author Contributions: Conceptualization, T.A. and M.L.; methodology, T.A. and M.L.; software, T.A.; validation, T.A. and M.L.; formal analysis, T.A.; investigation, T.A.; resources, D.L.; data curation, T.A. and M.L.; writing—original draft preparation, T.A., M.L., A.S. and D.L.; writing—review and editing, T.A., M.L., A.S. and D.L.; visualization, T.A.; supervision, T.A., M.L., A.S. and D.L.; project administration, D.L.; funding acquisition, D.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: We thank the State Forest Management Centre for providing us with the forest management data, and the Estonian Land Board for providing the ALS data. The authors acknowledge the helpful comments received from the anonymous reviewers, and thank John A. Stanturf for the language editing.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Nelson, R.; Krabill, W.; MacLean, G. Determining forest canopy characteristics using airborne laser data. *Remote Sens. Environ.* **1984**, *15*, 201–212. [[CrossRef](#)]
2. Tiner, R.W. Use of high-altitude aerial photography for inventorying forested wetlands in the United States. *For. Ecol. Manag.* **1990**, *33–34*, 593–604. [[CrossRef](#)]
3. Lang, M.; Arumäe, T.; Anniste, J. Estimation of main forest inventory variables from spectral and airborne lidar data in Aegviidu test site, Estonia. *For. Stud.* **2012**, *56*, 27–41. [[CrossRef](#)]
4. Arumäe, T.; Lang, M. ALS-based wood volume models of forest stands and comparison with forest inventory data. *For. Stud.* **2016**, *64*, 5–16. [[CrossRef](#)]
5. Olesk, A.; Praks, J.; Antropov, O.; Zalite, K.; Arumäe, T.; Voormansik, K. Interferometric SAR Coherence Models for Characterization of Hemiboreal Forests Using TanDEM-X Data. *Remote Sens.* **2016**, *8*, 700. [[CrossRef](#)]
6. Lang, M.; Kaha, M.; Laarmann, D.; Sims, A. Construction of tree species composition map of Estonia using multispectral satellite images, soil map and a random forest algorithm. *For. Stud.* **2018**, *68*, 5–24. [[CrossRef](#)]

7. Arumäe, T.; Lang, M. Estimation of canopy cover in dense mixed-species forests using airborne lidar data. *Eur. J. Remote Sens.* **2018**, *51*, 132–141. [\[CrossRef\]](#)
8. Guerra-Hernández, J.; Arellano-Pérez, S.; González-Ferreiro, E.; Pascual, A.; Altelaarrea, V.S.; Ruiz-González, A.D.; Álvarez-González, J.G. Developing a site index model for P. Pinaster stands in NW Spain by combining bi-temporal ALS data and environmental data. *For. Ecol. Manag.* **2021**, *481*, 118690. [\[CrossRef\]](#)
9. Large, A.R.G.; Heritage, G.L. Laser Scanning for the Environmental Sciences. In *Laser Scanning—Evolution of the Discipline*; Heritage, G.L., Large, A.R.G., Eds.; John Wiley & Sons, Ltd.: West Sussex, UK, 2009; pp. 1–20.
10. Zhao, Q.; Yu, S.; Zhao, F.; Tian, L.; Zhao, Z. Comparison of machine learning algorithms for forest parameter estimations and application for forest quality assessments. *For. Ecol. Manag.* **2019**, *434*, 224–234. [\[CrossRef\]](#)
11. Hawryło, P.; Francini, S.; Chirici, G.; Giannetti, F.; Parkitna, K.; Krok, G.; Mitelsztedt, K.; Lisańczuk, M.; Stereńczak, K.; Ciesielski, M.; et al. The Use of Remotely Sensed Data and Polish NFI Plots for Prediction of Growing Stock Volume Using Different Predictive Methods. *Remote Sens.* **2020**, *12*, 3331. [\[CrossRef\]](#)
12. McRoberts, R.E.; Tomppo, E.O. Remote sensing support for national forest inventories. *Remote Sens. Environ.* **2007**, *110*, 412–419. [\[CrossRef\]](#)
13. Packalén, P.; Maltamo, M. The k-MSN method for the prediction of species-specific stand attributes using airborne laser scanning and aerial photographs. *Remote Sens. Environ.* **2007**, *109*, 328–341. [\[CrossRef\]](#)
14. Ayrey, E.; Hayes, D.J. The Use of Three-Dimensional Convolutional Neural Networks to Interpret LiDAR for Forest Inventory. *Remote Sens.* **2018**, *10*, 649. [\[CrossRef\]](#)
15. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
16. Krigul, T. *Metsatakseerimine*; Valgus: Tallinn, Estonia, 1972.
17. Rules of Forest Management. Available online: <https://www.riigiteataja.ee/en/eli/ee/KKM/reg/52112017002/consolide> (accessed on 22 November 2021).
18. Zeide, B. Thinning and growth: A full turnaround. *J. For.* **2001**, *99*, 20–25. [\[CrossRef\]](#)
19. Cameron, A.D. Importance of early selective thinning in the development of long-term stand stability and improved log quality: A review. *Forestry* **2002**, *75*, 25–35. [\[CrossRef\]](#)
20. Bose, A.K.; Weiskittel, A.; Kuehne, C.; Wagner, R.G.; Turnblom, E.; Burkhart, H.E. Does commercial thinning improve stand-level growth of the three most commercially important softwood forest types in North America? *For. Ecol. Manag.* **2018**, *409*, 683–693. [\[CrossRef\]](#)
21. Slodicak, M.; Novak, J. Silvicultural measures to increase the mechanical stability of pure secondary Norway spruce stands before conversion. *For. Ecol. Manag.* **2006**, *224*, 252–257. [\[CrossRef\]](#)
22. Liu, Z.; Peng, C.; Work, T.; Candau, J.-N.; DesRochers, A.; Kneeshaw, D. Application of machine-learning methods in forest ecology: Recent progress and future challenges. *Environ. Rev.* **2018**, *26*, 339–350. [\[CrossRef\]](#)
23. Lõhmus, E. *Eesti Metsakasvukohatüübid*; Eesti Loodusfoto: Tartu, Estonia, 2004.
24. Orthophoto Metadata by Year. Available online: <https://geoportaal.maaamet.ee/eng/Spatial-Data/Orthophotos/Orthophoto-metadata-by-year-p350.html> (accessed on 22 November 2021).
25. Dual Channel Waveform Processing Airborne Lidar Scanning System for High-Point Density and Ultra-Wide Area Mapping: Riegl VQ-1560i Datasheet. Available online: <http://www.riegl.com/nc/products/airborne-scanning/produktdetail/product/scanner/55/> (accessed on 22 November 2021).
26. McGaughey, R.J. *FUSION/LDV: Software for LIDAR Data Analysis and Visualization*. March 2014—FUSION; Version 4.00; United States Department of Agriculture Forest Service Pacific Northwest Research Station: Portland, OR, USA, 2020.
27. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria. Available online: <https://www.R-project.org/> (accessed on 22 November 2021).
28. Strobl, C.; Boulesteix, A.-L.; Kneib, T.; Augustin, T.; Zeileis, A. Conditional variable importance for random forests. *BMC Bioinform.* **2008**, *9*, 307. [\[CrossRef\]](#)
29. Louppe, G.; Wehenkel, L.; Suter, A.; Geurts, P. Understanding variable importances in forests of randomized trees. In *Advances in Neural Information Processing Systems*; Burges, C.J.C., Bottou, L., Welling, M., Ghahramani, Z., Weinberger, K.Q., Eds.; Curran Associates Inc.: New York, NY, USA, 2013; pp. 431–439.
30. Aastaraamat “Mets 2019”. Available online: <https://keskkonnaagentuur.ee/media/882/download/> (accessed on 22 November 2021).
31. Schumacher, J.; Hauglin, M.; Astrup, R.; Breidenbach, J. Mapping forest age using National Forest Inventory, airborne laser scanning, and Sentinel-2 data. *For. Ecosyst.* **2020**, *7*, 60. [\[CrossRef\]](#)
32. Arumäe, T.; Lang, M. A simple model to estimate forest canopy base height from airborne lidar data. *For. Stud.* **2013**, *58*, 46–56. [\[CrossRef\]](#)