Article

Effect of Field Sample Size on Large-Scale Subtropic Forest Inventory Attribute Estimation Based on Airborne Laser Scanner Data

Zhu Yu [1], Xiangbei Zhou [1], Xin Lin [2], Zhongchao Chen [1], Mei Zhou [3] and Chungan Li [1,*]

[1] Forestry College , Guangxi University, 100 Daxue Road, Nanning, Guangxi 530004, China
[2] Guangxi Forest Inventory and Planning Institute, 14 Zhonghua Road, Nanning, Guangxi 530011, China
[3] School of Computer,Electronics and Information, Guangxi University, 100 Daxue Road, Nanning, Guangxi 530004, China
[*] Correspondence: gxali@126.com

Abstract: Exploring the effect of the sample size on the estimation accuracy of airborne LiDAR forest attributes in a large-scale area can help in optimizing the technical application scheme of operational ALS-based large-scale forest stand inventories. In our study, sample datasets composed of different sample plots were constructed by repeated sampling from 1003 sample plots in a subtropical study area covering 2376×10^3 km². Sixteen multiplicative power models were built in each forest type consisting of four forest attributes. Through these models, the variations of standard deviation (SD) and coefficient of variation (CV) of R² and rRMSE of forest attribute estimation models for different quantity levels of sample plots were also analyzed. The results showed that, first, when the sample size increased from 30 to the top limit, the SD of the forest attributes and LiDAR variables showed a decreasing trend. Second, as the sample size increased, the rRMSE of the 16 forest attribute estimation models gradually decreased, while the R² gradually increased. Third, when the sample size was small, both the SD of R² and rRMSE of the models were large, and the SD of R² and rRMSE gradually decreased as the sample size increased. In 50 models conducted for each attribute at the same sample size, for the mean standard deviations of forest attributes, the ten best performing models were lower than those of the total 50 models, and the worst ten models were the opposite. When the sample size increased, the accuracy of each forest attribute estimation model for each forest type gradually improved. The variation of forest attributes and the LiDAR variable of the construction model are critical factors that affect the model's accuracy. To efficiently apply airborne LiDAR in order to survey large-scale subtropical forest resources, the sample size of the Chinese fir forest, pine forest, eucalyptus forest, and broad-leaved forest should be 110, 80, 85, and 70, respectively.

Keywords: airborne LiDAR; forest attributes; multivariate power model; sample size

1. Introduction

Airborne laser scanning (ALS) was applied to carry out forest inventories in Norway, Finland, and Sweden from 2002[1, 2]. Since then, the reliability of the ALS estimation of large-scale forest attributes (diameter at breast height, DBH; mean tree height, H; stand basal area, BA; stand volume, VOL; above-ground biomass, AGB) has been confirmed in some countries, such as Australia, Canada and Spain[3-5]. ALS forest attribute estimation and mapping usually employ an ABA (an area-based approach)[6-9], and this is based on the excellent statistic relationship between LiDAR variables and stand attributes of sample plots [10-12]. This is generally carried out in two stages: the first stage is to establish an empirical relationship between LiDAR variables and forest attributes using field plot data; the second is to estimate the attributes in all forest areas using practical experience[13]. Therefore, field measurements of each plot are critical in the estimation of ABA forest

attributes. However, these methods are expensive, time-consuming, and labor-intensive [14-16]. The cost of field measurements accounts for a large proportion of the cost of airborne LiDAR forest attribute estimation. The optimization of sample plot surveys is a problem in the application of airborne LiDAR for forest resource investigation and monitoring in a large region[17, 18].

The sample size, the field plot size, and the requirement of location precision are the main factors that affect the cost of the sample plot survey. Gobakken and Næsset (2008) used regression models to estimate the DBH, BA, and VOL of forests and found that in most cases, larger plots (400 m²) can be compared with smaller plots (200 m²), the rRMSE of the model decreases, and R² increases. When the sample plot area increased from 200 m² to 300–400 m², the model accuracy improved[19]. Many studies also support this conclusion[20-24]. Ruiz et al. (2014) pointed out that for the estimation of VOL, AGB, and BA, the minimum size of the sample plot should be between 500 and 600 m². Lombardi et al. (2015) believe that the sample plot area should be at least 500 m² to evaluate forest investigated indicators. Adnan et al. (2017) suggest that when estimating the Gini coefficient of forest diameter distribution, the optimal sample plot area is 250–400 m²[25]. It was found that regardless of the size of the sample plot and the condition of the forest, the position error of up to 1 m has little effect on prediction accuracy. Relaxing the requirement of positioning accuracy and using cheaper GPS receivers help in reducing the survey cost. The sample size is the most critical factor affecting the price of the sample plots survey. Through Monte Carlo simulation, it was found that the estimation accuracy of forest attributes was moderate when the sample size decreased from the initial 50, 34, and 48 by 75 or even 50%[19]. The results of Silva et al. (2020) on the combined impact of sample size and modeling approaches for predicting stem volume in Eucalyptus spp. show that by using 63 sample plots, the model can achieve the same accuracy as that of the traditional forest inventory method[26]. With the increase in the number of sample plots, the model accuracy gradually improves. The sample size depends on the size of the study area, the complexity of forest structure, modeling methods, sampling methods, etc. Generally speaking, when the study area is large, the sample size required is large; otherwise, it is small [27, 28]. The coverage of the sample plot in the geospatial and variable feature space is also essential [17]. Some studies believe that forest attribute estimation technology impacts sample efficiency more significantly than the sample site selection method does. For example, the random forest (RF) accountability model is most effective when the sample size is minimal (less than 50) [29]. Some studies also believe that the least square method requires fewer sample plots than RF does [26]. Of course, different modeling methods do have other demands for sample plots. However, the majority of research on this topic has focused on sampling methods of sample plots. Maltamo et al. (2011) analyzed the influences of random sampling, forest type pre-stratified random sampling, selection based on geographical location, and sampling-based on ALS data on the estimation accuracy of the nonparametric model, and the results showed that the accuracy is the highest when using ALS data to assist sampling[30]. Many studies have shown that, in the stage of technical scheme design, ALS data are used as prior information to stratify the forests in the study area, and then, the sample plots are laid out, which helps to reduce the sample size under the premise of ensuring accuracy [30-32]. According to this method, in the Norwegian forest resource survey, there are about 50 sample plots in each layer [33]. However, tropical and subtropical regions, due to heavy rainfall, rain and fog, and weather influences, require a long time for airborne lidar data acquisition and sample surveys in large areas. Coupled with the rapid growth of forest trees (e.g., eucalyptus plantations), growth reaches 5–8 m/year. If LiDAR data are used to assist sampling, a long interval between LiDAR data and sample survey data can easily occur, so it is not easily adopted. Relevant research on the impact of the number of sample plots on the performance of airborne LiDAR subtropical forest attribute estimation models in a large area is insufficient.

In this study, 1003 sample plots (typical sampling) were sampled from 4 forest types in a large regional subtropical study area to explore the effect of sample size on airborne

LiDAR forest attribute estimation. The specific objectives were as follows: (1) to investigate the effects of different sample sizes on airborne LiDAR variables and forest attributes in sample plots; (2) to investigate the effect of sample size on the estimation accuracy for different forest attributes in different forest types through the multivariate power model; (3) to explore the mechanism of the influence of different sample sizes on the estimation accuracy for forest attributes; (4) to determine the minimum and maximum sample size in the large-area forest resource survey application of airborne LiDAR.

2. Materials and Methods

2.1. Study Area

The study site covers the whole of Guangxi Zhuang Autonomous Region in South China (104°28′–112°04′ E, 20°54′–26°24′ N), spanning an area of 23.76 km² (Figure 1). In this study, airborne LiDAR data acquisition and sample plot measurements were carried out over three years; each year was divided into one region, namely, the Nanning region (an area of 22,100 km²), the eastern region (128,400 km²), and the western region (87,100 km²).

The Tropic of Cancer traverses the middle of the study area, belonging to the subtropical monsoon climate zone. The average annual temperature is 16.5–23.1 °C, the accumulated temperature ≥10 °C is 5000–8300 °C, the average annual precipitation is 1080–2760 mm, 70–80% of the rain occurs in the rainy season (from April to September), and the annual sunshine duration is 1213–2135 h.
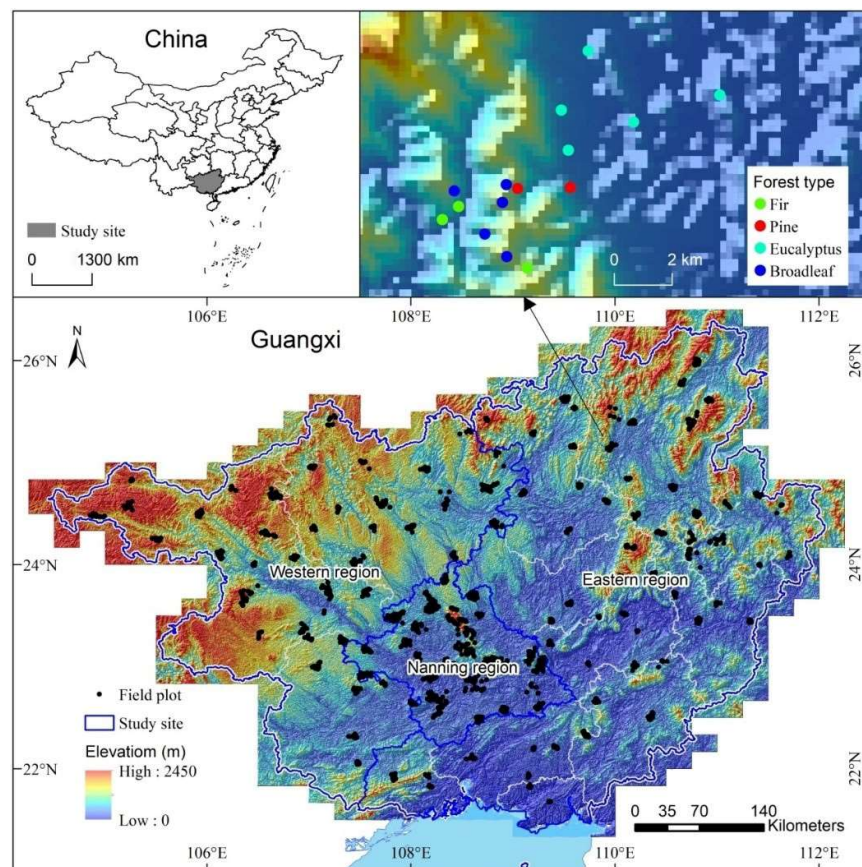

Figure 1. Study site and distribution of sample plot.

The landform and forest composition of the three regions differ to some extent. The central and northern parts of the eastern region and the north parts of the western region belong to the subtropical zone, with continuous mountains. The zonal vegetation is an

evergreen broad-leaved forest. *Schima superba* forest, *S. argentea* forest, *Castanopsis faberi* forest, *C. lamontii* forest, *C. eyrie* forest, and *C. fargesii* forest are the representative forests. *Pinus massoniana* (about 80% natural and 20% plantations) and Chinese fir (Cunninghamia lanceolata) plantations are relatively large in number, with a few Eucalyptus plantations (mainly Eucalyptus urophylla and Eucalyptus grandis × E. urophylla). The central part of the western region and the northern and eastern part of the Nanning region belong to the south subtropical region, alternating mountains, hills, and plains. The zonal vegetation is a monsoon evergreen broad-leaved forest. There are Castanopsis hystrix forests, C. tonkinensis forests, C. kawakamii Hay., Cryptocarya concinna forests, and Machilus forests. The chinensis forest is the most suitable area for Chinese fir. The natural Pinus massoniana forest and the Chinese fir plantation have a large area, and the eucalyptus plantation occupies a large proportion. The eastern, western, and southern parts of Nanning are mainly mountainous and hilly, belonging to the northern tropical zone, and the zonal vegetation is located in the seasonal rain forest. The Hopea chinensis forest, Parashorea chinensis forest, olive (Canarium album) forest, olive (C. pimela) forest, Erythrophleum forest, Fordii forest, and Madhuca pasquieri forest are the representative forests. Eucalyptus plantation occupies the most significant area, natural horsetail forest also occupies a particular proportion, and there are a few Chinese fir plantations. Still, this area is a suitable general area for Chinese fir. Chinese fir plantation accounts for 16.5% in the study area, pine plantation accounts for 17.5%, eucalyptus plantation accounts for 24.8%, and broad-leaved forest accounts for 41.2%. A few fir Masson pine mixed forests (natural) and fir hardwood mixed forests (natural) exist, as most Chinese fir plantations are virgin forests. The Pine Linda area is a pure natural forest and is the same age as that of some pristine artificial forests, and there are also several Masson pine-hardwood mixed forests, broad-leaved–most natural mixed forests, and a small plantation.

2.2. Plot Data

The arbor forest in the study area was divided into four types, namely, Chinese fir forest, pine forest, eucalyptus forest, and broad-leaved forest, and sample plots were arranged on forest distribution mapped in 2015 according to the principle of typicality and cluster distribution. About 100 sample plots were laid for each forest type in each region. The minimum spacing of sample plots was 500 m (Figure 1). The hand-held dual-frequency differential GPS was used to navigate to the center of the sample plot site. The forest compass and laser rangefinder (Leica Disto™ X30) set the plot boundaries and separated the sub-plots. The size of the plot was 30 × 20 m, and the plot was divided into four 15 × 10 m sub-plots in a north-south direction. RTK was used to measure the coordinates of the northwest and southeast corner of the sample plot, and the error of the coordinate of the sample plot was less than 1 m. The trees with a diameter at breast height of ≥5.0 cm were measured in each sub-plot. Three average trees, one tree with dominant height, and one tree with dominant diameter were selected, and the height of the trees was measured using an ultrasonic altimeter (HagLof Vertex IV). The investigated indicators of the sub-plot include mean square diameter, mean tree height, stand basal area, tree density, and stand volume, in which stand volume is calculated according to the cross-basal area and average height using the local general height model. The following methods were used to calculate the forest attributes of the sample plot: stand basal area and volume were obtained by totaling four sub-plots; the average diameter at the breast height of the sample plots was obtained through the measurement of every individual tree, and the average height of the sample plots was calculated using the weighted-average method. Moreover, the forest origin, average age, and canopy coverage (CC) of the trees were investigated and recorded.

Sample plots were surveyed from October 2016 to January 2017 in Nanning, from November 2018 to May 2019 in the eastern region, and from August 2019 to January 2020 in the western area. Due to the large interval between the survey time of some sample plots in the east of China and the acquisition time of LiDAR data (the proportion of sample

plots with a gap of ≥90 d ≥70%), to objectively reflect the forest structure state during Li-DAR scanning, a stand growth model was established based on the sample plot reexamination data of Guangxi forest resources' continuous inventory system from 2005 to 2015. The diameter at the breast height and the mean tree height of the sample plots in the eastern region were adjusted via the forest growth rhythm table. The basic information of sample plots is shown in Table 1.

Table 1. Summary of the field plots.

Forest Type	Sample Size	DBH		H		BA		VOL	
		Mean (cm)	CV (%)	Mean (m)	CV (%)	Mean (m²/ha)	CV (%)	Mean (m³/ha)	CV (%)
Fir	222	12.28	27.51	10.88	27.19	31.91	29.34	193.60	46.58
Pine	260	19.05	28.52	13.79	26.99	27.95	31.71	192.03	46.94
Eucalyptus	269	11.27	21.50	16.12	20.03	17.24	34.12	141.91	44.16
Broadleaf	252	13.70	36.86	10.50	27.15	19.64	41.07	111.80	58.85

2.3. Airborne Lidar Data

The acquisition time of airborne LiDAR in the Nanning region, eastern region, and western region is from October 2016 to April 2017, October 2018 to October 2019, and August 2019 to January 2020, respectively. RIEGL LMS-1560 and RIEGL LMS-1560I laser radar scanners, with a scattering angle of 0.5 mrad, were used in this study. The mean footprint was 30 cm on average, the pulse repetition frequency was 1000 kHz, the scan frequency was 146 Hz, and the maximum scanning angle was ±30°. The LiDAR data of the three regions were obtained according to the same standards: the relative altitude of the flight was about 2500 m, the flight speed was 200–240 km/h, the overlap was 21–25%, and the airstrip interval was 1.3–2.2 km—average point density ≥2.0 points/m². The error in the elevation of the laser point is better than 0.15 m.

Since the first LiDAR echo represents the first important part of the reflected signal, the variables extracted from the first echo are all able to meet the requirements of biomass estimation compared, while those from the subsequent echoes are not [34-36]. In this study, the first echo was used to extract 13 LiDAR variables, including the mean height of point (Hmean) and its standard deviation (Hstdev); the coefficient of variation (HCV); the 95% quantile height (HP95, canopy coverage (CC); 50 and 75% quantile densities (DP50 and DP75), the mean value of the leaf area density profile (LADmean) and its standard deviation (LADstdev) and coefficient of variation (LADcv); and the mean value of the vertical foliage profile (VFPmean), its standard deviation (VFPstdev), and coefficient of variation (VFPcv).

2.4. Statistics of Airborne LiDAR Variables and Forest Attributes

Among the total number of sample plots of each forest type, the datasets with different sample sizes were constructed using the repeated sampling method in steps of 5, starting from 30 sample plots, such as 30, 35,...200, 205..., etc. The statistical values (mean and standard deviation) of 13 LiDAR variables and 4 forest attributes (DBH, H, BA, and VOL) in samples with different sizes were calculated. Their range of variation was analyzed via ((max − min) × 200/(max + min)). The mean value of LiDAR variables and forest attributes in samples with different sizes and the corresponding variables and attributes in the whole number of plots was analyzed by adopting the t-test method. In the above calculation, for different sizes of the sample, random sampling was repeated 50 times to reduce the unexpected error.

2.5. Model Fitting and Validating

A large number of airborne lidar forest attribute estimation models have been published, including linear models, nonlinear models, parametric models, and nonparametric

models [24, 37-40]. These models have achieved favorable estimation effects for specific study areas, specific forest types, and specific forest attributes—maximum coefficient of determination (R^2), minimum estimation error (rRMSE), and minimum systematic error [24].

In another unpublished study, we divided 13 airborne LiDAR variables into 3 groups, namely, a highly variable group, a density variable group, and vertical structure variables. Each group contains one or two variables, and these variables have a combination of rules and a total of 84 model formulae. Using the same LiDAR data, the generalized, predictive models of the forest attributes for 4 forest types were established (Table 2; the establishment of the model is described in another article).

Table 2. Structural formula of forest attribute estimation model.

Forest Type	Attribute	Model
Fir	DBH	$DBH = a_0 Hp95^{a_2} CC^{a_3} Hcv^{a_4} dp50^{a_5} LADcv^{a6}$
	H	$H = a_0 Hp95^{a_2} CC^{a_3} Hstdev^{a_4} dp50^{a5} LADcv^{a6}$
	BA	$BA = a_0 Hp95^{a_2} CC^{a_3} Hcv^{a_4} LADmean^{a_5}$
	VOL	$VOL = a_0 Hmean^{a_1} CC^{a_2} Hcv^{a_3} LADmean^{a4}$
Pine	DBH	$DBH = a_0 Hmean^{a_2} CC^{a_3} Hcv^{a_4} LADmean^{a_5}$
	H	$H = a_0 Hmean^{a_2} CC^{a_3} Hcv^{a_4} VFPmean^{a_5}$
	BA	$BA = a_0 Hmean^{a_2} CC^{a_3} Hcv^{a_4} dp75^{a_5} VFPmean^{a_6}$
	VOL	$VOL = a_0 Hmean^{a_2} CC^{a_3} Hcv^{a_4} dp50^{a_5} LADmean^{a_6}$
Eucalyptus	DBH	$DBH = a_0 Hp95^{a_2} CC^{a_3} Hstdev^{a_4} dp75^{a_5} LADstdev^{a_6}$
	H	$H = a_0 Hmean^{a_2} CC^{a_3} Hcv^{a_4} dp50^{a_5} VFPmean^{a_6}$
	BA	$BA = a_0 Hp95^{a_2} CC^{a_3} Hstdev^{a_4} dp75^{a5} VFPstdev^{a6}$
	VOL	$VOL = a_0 Hp95^{a_2} CC^{a_3} Hcv^{a_4} dp75^{a5} VFPcv^{a6}$
Broadleaf	DBH	$DBH = a_0 Hmean^{a_2} CC^{a_3} LADmean^{a_4}$
	H	$H = a_0 Hmean^{a_2} CC^{a_3} VFPmean^{a_4}$
	BA	$BA = a_0 Hmean^{a_2} CC^{a_3} VFPstdev^{a_4}$
	VOL	$VOL = a_0 Hmean^{a_2} CC^{a_3} VFPstdev^{a_4}$

Samples with different sizes were used to fit and validate the above models. The Newton–Gaussian iterative method was used for model fitting, and the leave-one-out cross-validation method was used to validate the model. To reduce random error, 50 tests were carried out, and the R^2 and rRMSE of each fitting and validating sample were calculated.

2.6. Determination of Minimum and Maximum Sample Size for Forest Attribute Estimation

After 50 times of model fitting for each forest attribute with different sample sizes, the coefficient of variation (CV_rRMSE) of the rRMSE was calculated. The regression method was used to fit the sequence of CV_rRMSE that changed with the sample size, and the theoretical model of the relationship between CV_rRMSE and sample size was obtained. When CV_rRMSE was less than 5%, the accuracy of the estimation model for this forest attribute tended to be stable, and increasing the number of sample plots had little effect on improving the estimation accuracy of the model. In this case, the corresponding number of sample plots is the maximum sample size. After a considerable number of tests, the relationship between CV_rRMSE and the sample size was best fitted by the LOGISTIC regression, and its model is

$$CV\_rRMSE(\%) = 1/(a_0 + a_1 \times a_2^n) \tag{1}$$

In the formula, n is the sample size, and a_0, a_1, and a_2 are model parameters.

Increasing the sample size helps to improve the accuracy of the forest attribute estimation model [26], but increasing the sample size means increasing the field measurements cost. Considering model accuracy and survey costs together, the number of sample plots is determined as the minimum sample size when the forest attribute estimation model reaches an acceptable accuracy. Many studies have shown that the estimation accuracy of forest attributes varies significantly in different study areas, different forest types, and different forest attributes [11, 41, 42]. Therefore, it is difficult to determine an acceptable minimum estimation accuracy for each forest type and each forest attribute, such as rRMSE lower than 20% or R^2 higher than 0.7. We believe that, when CV_rRMSE is less than 10%, the model accuracy is acceptable. Based on this, the minimum sample size can be determined.

3. Results

3.1. The Influence of the Variation of the Sample Size on the LiDAR Variables and Forest Attributes

In each forest type, the mean value of the LiDAR variables was very close in the different sample sizes. The range of mean value was generally lower than 1.0% and rarely higher than 2.0%, and the maximum was not higher than 5.0% (Table 3). The T-test results show that there was no significant difference between the mean value of each LiDAR variable in samples with different sizes and the corresponding variable in the total sample in each forest type (α = 0.05). It was observed that the variation of the sample size did not affect the stability of the LiDAR variable.

Table 3. Range of mean and SD of 13 LiDAR variables in the different sample sizes in each forest type (%).

Stratum	Static	hp95	Hmean	Hstdev	Hcv	CC	dp50	dp75	LADm	LADsd	LADcv	VFPm	VFPsd	VFPcv
Fir	Mean	0.91	0.97	1.05	0.77	0.36	1.00	1.94	4.81	2.72	0.55	0.91	0.68	0.59
	Std.	1.65	3.92	2.15	3.12	6.85	5.34	3.30	16.72	10.38	0.84	2.78	0.73	2.20
Pine	Mean	0.82	0.83	1.29	0.53	0.25	0.67	1.34	3.73	4.41	0.41	0.89	0.44	2.86
	Std.	2.04	1.52	1.31	2.32	2.79	2.11	0.72	11.04	10.44	2.68	4.70	2.41	1.82
Eucalyptus	Mean	0.71	0.67	1.16	0.46	0.23	0.33	0.93	2.14	3.66	1.13	0.81	0.37	0.56
	Std.	5.41	1.87	2.58	3.12	1.21	3.10	2.64	6.55	6.48	1.85	1.62	2.02	1.35
Broadleaf	Mean	0.95	0.88	0.99	0.35	0.30	0.52	1.62	1.26	1.41	0.47	1.34	0.44	0.75
	Std.	2.25	2.08	1.93	2.43	4.38	10.15	2.03	3.08	8.64	1.13	1.42	1.06	1.20

* Note: LADm is LADmean, LADsd is LADstdev, VFPm is VFPmean, and VFPsd is VFPstdev.

In general, the SD of LiDAR variables was also close in different sample sizes in each forest type, with most of their range being lower than 5.0%. Their range of variation was higher than the range of the mean value, with the maximum reaching 16.72% (Table 3). With the increase in sample size, the SD of each variable showed a decreasing trend. After

calculating the relative difference between the SD of each LiDAR variable in different sample sizes and the corresponding variable in the total sample, it could be observed that with the increase in sample size, their relative differences gradually decreased (Figure 2).
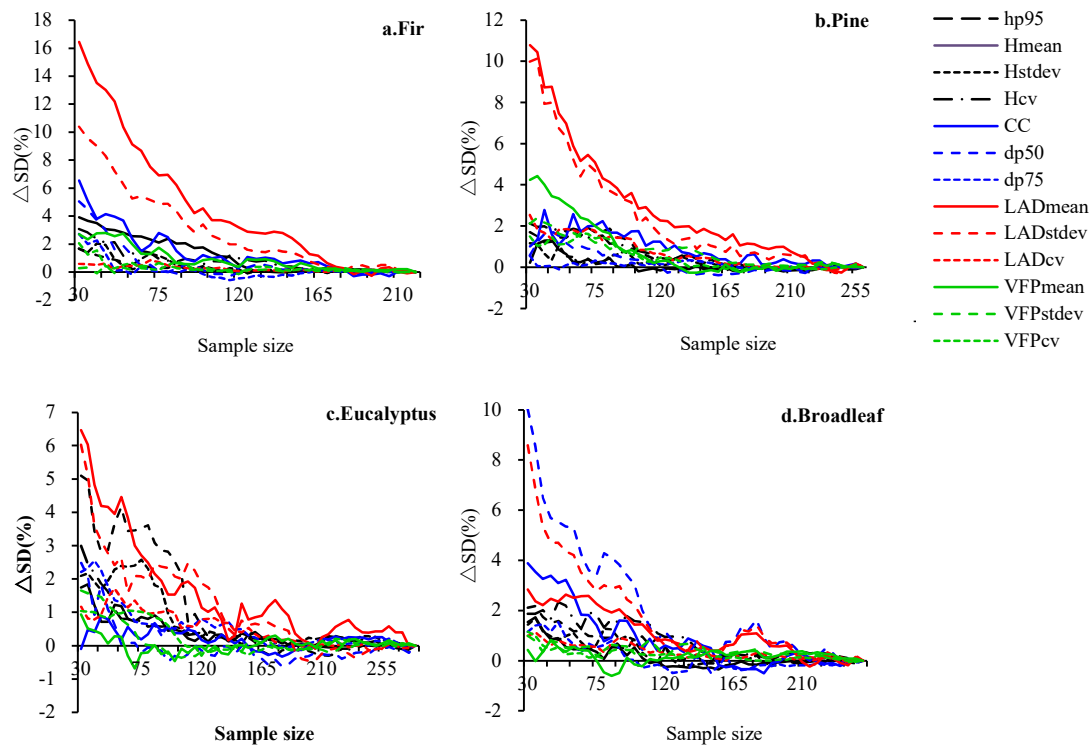

Figure 2. In each forest type, the relative difference in the SD between LiDAR variables of different sample sizes and the corresponding variables in the total sample showed a decreasing trend. Fir (a), Pine(b),Eucalyptus(c) ,Broadleaf(d)

In general, The SD of each LiDAR variable in eucalyptus forests had the smallest range of variation, which decreased rapidly with the increase in sample plots. When the sample size reached approximately 110, the relative difference of SD for most variables was lower than 1%. The relative difference of SD of each LiDAR variable in the broad-leaved forest was slightly higher than that in the eucalyptus forest. When the sample size reached approximately 115, the relative difference of most variables was lower than 1%. When the sample size reached approximately 120 and 125, the relative difference of most variables was lower than 1% in the pine forest and Chinese fir forest, respectively.

The variation of forest attributes was similar to that of LiDAR variables, but the variation range was smaller than that of LiDAR variables (no more than 1.5%). There was no significant difference between the mean value of each forest attribute in the samples of different sizes and the corresponding variables in the total sample in each forest type (α = 0.05). With the increase in sample size, the relative difference of SD of forest attributes in each forest type also decreased (Figure 3).
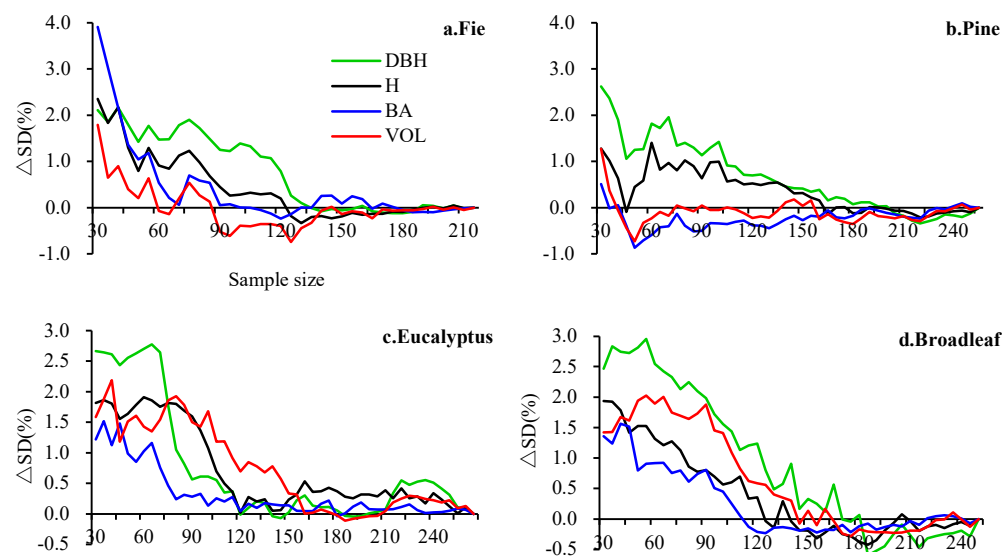
Figure 3. In each forest type, the relative difference of the SD between forest attributes in different sample sizes and the corresponding variables in the total sample showed a decreasing trend. Fir (a), Pine(b),Eucalyptus(c) ,Broadleaf(d)

In general, the variation of the standard deviation of the diameter at breast height was the largest between different sample sizes, followed by the volume, and the range of variation in the standard deviation of the basal area and mean tree height was relatively small.

The mean values of the LiDAR variables and forest attributes were very close among the different sample sizes. Their ranges of variation were small and not significantly different (α = 0.05). Their standard deviations were also relatively close, but their ranges of variation were significantly larger than their means. With the increase in sample size, the range of variation in the standard deviation of LiDAR variables and forest attributes decreased, but at different rates in different forest types.

3.2. The Influence of the Variation of Sample Plots on the Estimation Accuracy of Forest Attributes

The results (R^2 and rRMSE) of the 50 rounds of training and testing were analyzed in-depth in the datasets with different sample sizes. When the sample size was 30 or the total number, the validation results of each forest attribute estimation model of each forest type were as shown in Table 4.

Table 4. The mean values of R^2 and rRMSE of the test samples of each attribute estimation model of each forest type were obtained when the number of plots was 30 and the total number of plots was whole.

Forest Type	Sample Size	VOL		BA		H		DBH	
		R^2	rRMSE (%)	R^2	rRMSE (%)	R^2	rRMSE (%)	R^2	rRMSE (%)
Fir	215	0.761	22.18	0.582	11.78	0.804	11.78	0.617	16.79
	30	0.651	24.08	0.422	19.88	0.717	12.78	0.444	18.62
Pine	255	0.824	19.69	0.655	18.62	0.854	10.30	0.506	20.01
	30	0.756	22.04	0.542	20.41	0.779	11.72	0.295	22.63
Eucalyptus	275	0.811	19.09	0.716	18.17	0.790	8.95	0.714	11.32
	30	0.715	22.16	0.608	20.61	0.701	9.82	0.582	12.51
Broadleaf	250	0.617	36.16	0.484	29.37	0.595	17.09	0.428	31.16

30	0.478	39.57	0.361	31.01	0.415	19.12	0.379	32.06

When the sample size increased from 30 to the whole number, the mean R^2 gradually increased in the test samples, and the mean rRMSE gradually decreased (Figure 4), indicating that the accuracy of all models improved.
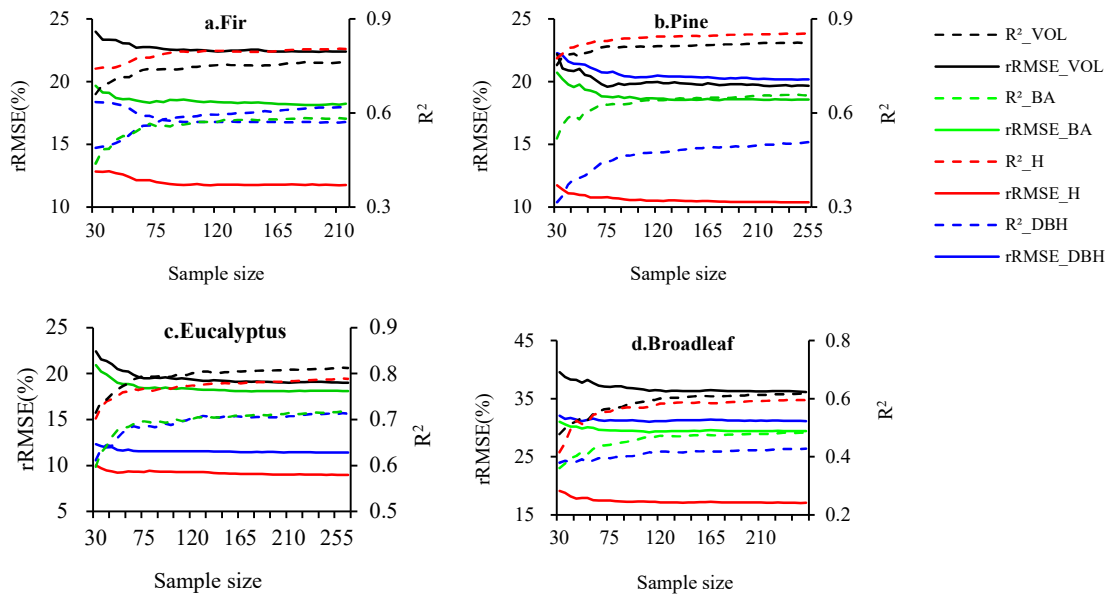

Figure 4. The mean values of R^2 and rRMSE change with the increase in test sample sizes. Fir (a), Pine(b),Eucalyptus(c) ,Broadleaf(d)

As shown in Figure 4, the mean value of R^2 and rRMSE varied significantly with the increase in sample size. When the sample size increased from 30 to the total number, the R^2 of the fir forest volume model increased by 16.89%, and the rRMSE decreased by 7.87%. In contrast, the R^2 of the eucalyptus forest volume model increased by 13.40%, and the rRMSE decreased by 13.86% (Table 4). The results show that the sample size had different effects on the performance of each forest attribute estimation model in each forest type.

When further analyzing the results of 50 rounds of model fitting and validating for different sample sizes of each forest parameter for different forest types, noteworthy distribution of R^2 and rRMSE was observed when the sample size was small, the average R^2 was low, and rRMSE was high. With the increase in sample size, their range of variation gradually decreased. Moreover, the mean value of R^2 gradually increased, and the mean value of rRMSE gradually decreased (Figure 5). Taking the pine forest volume estimation model as an example, when the sample size was 30, the variation (mean value) of R^2 and rRMSE in the test samples were 0.419–0.886 (0.756) and 14.27–32.63% (22.04%), respectively. When the sample size was 255, the R^2 and rRMSE ranged from 0.816 to 0.839 (0.824) and 18.92 to 19.93% (19.69%), respectively.
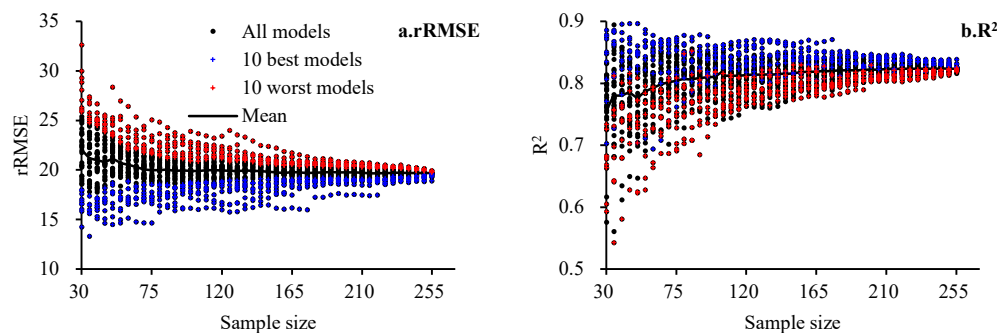
Figure 5. Distribution of rRMSE (a) and R^2 (b) of pine forest stand volume estimation model in the test samples when 50 repeated sampling models were fitted and validated for different sample sizes.

Of course, with the increase in sample plots, the coefficient of variation of R^2 and rRMSE of each forest attribute estimation model in each forest type also showed a decreasing trend (Figure 6).

Possible reasons for the above phenomenon are as follows (1) Datasets with different sizes were constructed by the random sampling method from the whole sample plot. When the sample size is small (e.g., 30), the sample composition significantly varies among different datasets. Some datasets achieve higher model accuracy, while others have lower accuracy. Thus, R^2 and rRMSE vary over a wide range, and their standard deviations are large. When the sample size is large, there is a minimal difference among different datasets. The model accuracies obtained by each sample dataset are close to each other, and the range of variation is small. (2) When the sample size is small, the mean value of the model accuracy is low due to the wide range of variation of the model accuracy. By contrast, when the sample size is large, the model accuracy range is small, so the mean value is high.

During the 50 rounds of model fitting and validating, rRMSE in the test samples was taken as the evaluation standard. The range of variation of R^2 and rRMSE in the ten models with the best performance also decreased with the increase in sample size. Taking pine forest as an example, when the sample size was 35, the ranges of variation (mean value) of R^2 and rRMSE of the ten best performing models were 0.727–0.907 (0.851) and 13.31–17.39% (16.01%), respectively, and when the sample size was 255, their ranges of variation (mean value) were 0.825–0.839 (0.828) and 18.92–19.61% (19.46%), respectively. With the increase in sample size, the mean value of R^2 and rRMSE of the ten models with the best performance showed a gradually decreasing trend. That is, the model accuracy showed this trend. The mean value of R^2 and rRMSE of the ten models with the worst performance was the opposite (Figure 5). The above figure shows that provided the sample plots are properly selected, even if the sample size is small, high model accuracy can be achieved. It also shows that the model accuracy is closely related to the selection of sample plots. An increase in sample plots is not a necessary condition for improving model accuracy.

The ten best and worst-performing models were selected among the various forest attribute estimation models for different forest types. Their changes in target attributes and LiDAR variables were analyzed. The results showed that (1) when the sample sizes were the same, for the mean standard deviations of forest attributes, the ten best performing models were lower than those of the total 50 models, and the worst ten models were the opposite; (2) when the sample sizes were the same, there was always one or more LiDAR variables, and for the mean standard deviations, the ten best performing models were lower than those of the total 50 models and the worst ten models were the opposite; (3) with the increase in sample size, the standard deviations of the target attributes and LiDAR variables in the samples of the ten best and worst performing models were close to those of the corresponding variables in the total 50 samples. Figure 7 shows the variation of the standard deviation of the attribute, 5 LiDAR variables in the ten best and worst

performing models, and the overall 50 models for the pine forest stand volume estimation in different samples. When the sample size was not large, for the samples of the ten best performing models, the standard deviations of their VOL, Hcv, and DP50 were significantly smaller than those in the overall 50 samples, while the change in the ten models with the worst performance was the opposite.
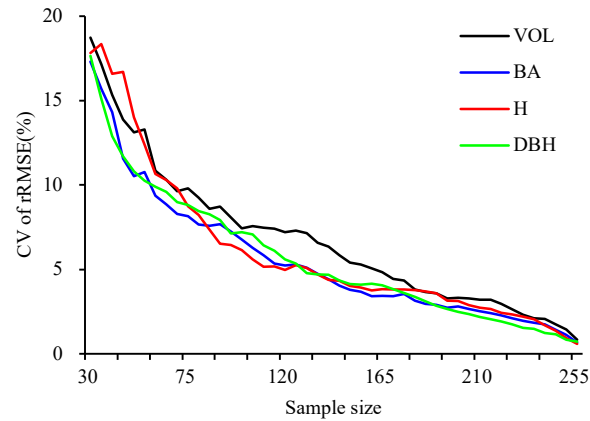

Figure 6. The variation coefficient of rRMSE of the four forest attribute estimation models of pine forest all showed a decreasing trend with an increase in sample size.
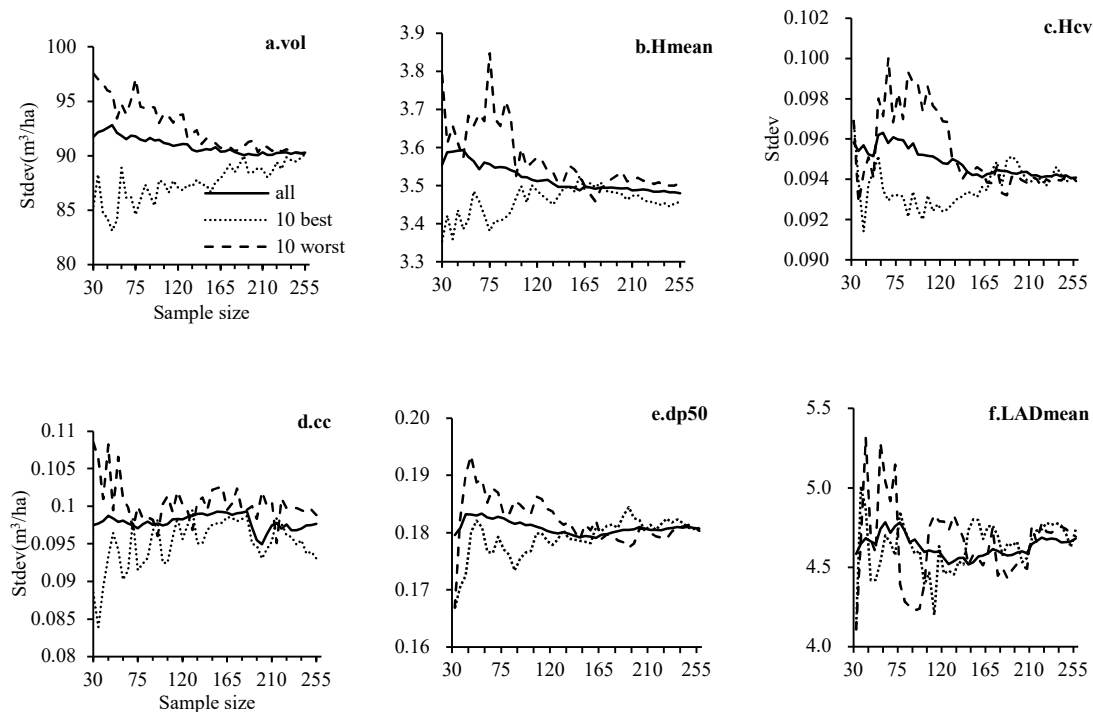

Figure 7. Variation trend of standard deviations of target variables (**a**) and LiDAR variables, Hmean (**b**), HCV (**c**), CC (**d**), DP50 (**e**), and LADmean (**f**) of the ten models with the best and worst performance in the estimation of pine forest stand volume with the increase in the number of sample plots.

According to Figures 5 and 7, it can be inferred that the standard deviation of the target variable and the LiDAR variable in the sample is a critical factor in determining the

accuracy of the model. When the standard deviation of the target variable and the LiDAR variable is small, the model accuracy is high; otherwise, the model accuracy is low.

3.3. Determination of Minimum Plot Quantity and Maximum Plot Quantity

For the relation between the coefficient of variation of the rRMSE (CV_rRMSE) and sample size, logistic regression was most suitable. The range of variation of R^2 was 0.943–0.990, and rRMSE was less than 15%, as shown in Figure 8, indicating that with the increase in sample size, the CV_rRMSE in each forest attribute estimation model had good regularity.
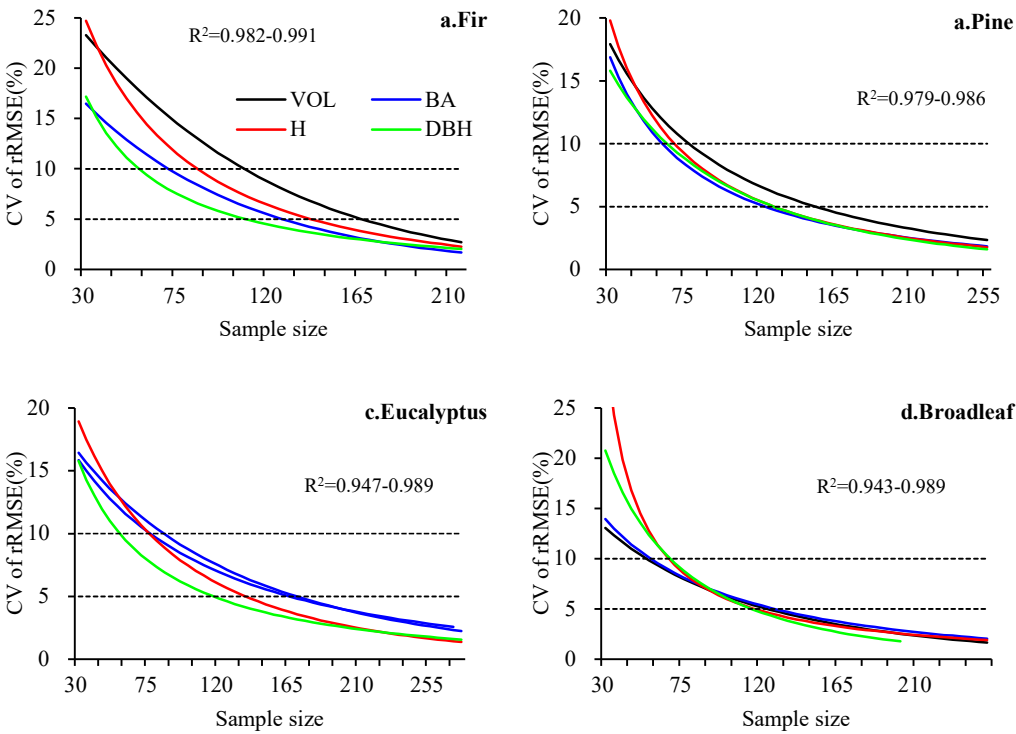

Figure 8. Logistic curve of variation coefficient of rRMSE of four forest attribute estimation models of four forest types with an increase in sample size. Fir (a), Pine(b),Eucalyptus(c) ,Broadleaf(d)

According to the variation curve of CV_rRMSE of each forest type and each forest attribute, we can determine the sample size when CV_rRMSE is less than 10 and 5%. The two sample sizes are the minimum sample size required for each forest type and the sample size required when the model accuracy is stable (the maximum number) (Table 5).

Table 5. Minimum number of sample plots required for each forest attribute estimation of each forest type and the maximum number of sample plots needed when the model accuracy is stable.

Forest Type	VOL		BA		H		DBH		All Attributes	
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
Fir	110	170	75	130	85	145	60	105	110	170
Pine	80	155	65	125	70	130	65	130	80	155
Eucalyptus	80	165	85	170	80	140	60	120	85	170
Broadleaf	55	125	60	130	70	120	70	115	70	130

It can be seen from Table 5 that the minimum sample size required for estimating different forest attributes varies among forest types, and the sample size required to

achieve stable model accuracy (maximum sample size) also differs. In general, when considering the minimum sample size, broad-leaved forests require the least number of plots, and fir forests need the most. When considering the maximum sample size, fir forests and eucalyptus forests require the largest number of plots, and broad-leaved forests, the smallest. For the same forest type, the minimum and maximum sample sizes required by different forest attributes greatly vary. The minimum and the maximum sample plots needed by estimating the volume are the largest, and those required for the diameter at breast height are the smallest. The maximum sample size is similar to the minimum sample size regarding estimating different forest types and different forest attributes, and it is the smallest for the broad-leaved forest, while its highest values can be found for the Chinese fir forest and eucalyptus forest. The sample size needed for the estimation of volume is the highest, while that required for the DBH is the lowest. When all four forest attributes needed to be estimated, the minimum numbers of plots needed for fir, pine, eucalyptus, and broadleaf forests were 110, 80, 85, and 70, respectively. For each forest attribute estimation of the same forest type, the minimum sample size required differed by 55–120% compared with the sample size when the model accuracy was stable, where the difference between the minimum and maximum sample plots of each forest attribute in Chinese fir forest was the smallest at 55–75%, while the difference between broad-leaved forest and eucalyptus forest was the largest at 64–127% and 75–106%, respectively, indicating that when the number of sample plots reaches a specific number, with the increase in this number, the accuracy of each forest attribute estimation model of each forest type improves at different speeds.

For the minimum sample size, the differences in the rRMSE between each forest attribute model and the corresponding model for the whole number of plots are mostly less than 5%, and the maximum is 6.12%. When the model accuracy reaches a stable state, the differences between the rRMSE of each forest attribute model of each forest type and the corresponding model for the whole number of plots are slight, most of which are less than 1%, and the maximum difference is no more than 3%. This indicates that even if the number of sample plots increases further, the improvement to model accuracy is minimal.

4. Discussion

Sample plot survey is an essential element of airborne LiDAR forest attribute estimation in large areas. It is also a work with a heavy workload, high labor intensity, and high cost. Therefore, the sample size is one of the critical issues to be considered in the formulation of operational ALS-based large-scale forest stand inventories. Although some scholars have studied the influence of sample size on the estimation accuracy of forest attributes, according to our limited knowledge, these studies are mainly focused on temperate forests, and their study areas are small. Studies on subtropical forests in large regions have not been reported. In this study, we conducted in-depth research on this issue through a large number of samples in a large regional subtropical forest research area with diverse forest types and complex landscapes. We preliminarily proved the influence of sample size on the performance of the estimation model of LiDAR variables and forest attributes. The findings of this study are of the universal reference value.

Based on 1003 sample plots in the subtropical study area (237.6×103 km²), datasets composed of different sample plots were constructed by repeated sampling. The variations of LIDAR variables and forest attributes in these datasets were analyzed. For the same order of magnitude, the repeated sampling method was used to obtain 50 datasets. These datasets were used for modeling and testing, and the ten best and worst performing models from these 50 models were analyzed. On this basis, the differences in target variables and LiDAR variables of the ten models with the best and worst performers were analyzed. Finally, the accuracy of models with different sample sizes was analyzed, and logistic regression models between CV_rRMSE and sample size were established. The minimum and the maximum sample size needed for each forest type and each forest attribute estimation in large-scale forest resource survey application of airborne LiDAR was

further discussed. Such a large study area, such a complex forest background, such a large sample size, and such a large number of forest attributes have not previously been reported.

This study shows that the mean values of LiDAR variables and forest attributes in samples with different sizes were very close to each other. However, with the increase in sample size, their standard deviations rapidly decreased at the beginning and then decreased slowly after reaching a certain sample size. The variation trend of the relative root means square error (rRMSE) of each forest attribute estimation model was the same as the variation trend of the standard deviation of the LiDAR variable. By contrast, the R^2 of the forest attribute estimation model was the opposite, indicating that the performance of the forest attribute estimation model gradually improved with the increase in sample size. Due to the lack of similar research reports, we are unable to conduct an in-depth comparative analysis. White et al. (2017) pointed out that the number of sample plots required by the ABA method depends on the complexity of the forest environment, modeling method, estimated target attributes, number of layers, and expected accuracy. Our research also partially supports this conclusion[12]. For the R^2 and rRMSE of the same forest attribute estimation model, the variation trends of Chinese fir forest, pine forest, eucalyptus forest, and broad-leaved forest were significantly different. For the same forest type, the performance of the mean diameter at breast height, mean height, stand basal area, and stand volume models were also significantly different. Norway spruce (Picea abies (L.) Karst), located in the southeastern part of Norway, was observed in this study. When the number of sample plots decreased from 132 to 66, the estimation accuracy of average height, stand basal area, and stand volume minimally decreased. This conclusion is not entirely different from the conclusion of our research. The complexity of the forest structure in the study area may be one of the main factors affecting the estimation accuracy of the model. In general, the larger the study area, the greater the heterogeneity of forest structure. In many small area studies, even with less than 50 sample plots per layer, airborne LiDAR forest parameter estimation can achieve good accuracy [11, 43-47]. In some studies, the number of sample plots was less than 25 [21, 28, 48], and even only 10 [49]; the accuracy of the model is good. Many studies have shown that in the stage of technical scheme design, the ALS data (the mean height of point et al.) is used as prior information to stratify the forests in the study area, and the sample plots are then laid out, which helps in reducing the number of sample plots while maintaining accuracy [30-32]. According to this method, in the Norwegian forest inventory, there are approximately 50 sample plots of each layer. However, it is difficult to use this method in large subtropical regions with high rainfall and rapid forest growth (the high growth rate of eucalyptus plantations is 5–8 m/year). For such a large study area, acquisition of airborne LiDAR data takes at least six months, and the sample plot survey takes as long as four months due to weather conditions. When the interval between the acquisition time of LiDAR data and the measurement time of the sample plot is long, the laser point has a large deviation to the forest structure, especially for the forest type with rapid growth, which leads to an increase in the uncertainty of the estimation results. Therefore, in practical application, we adopt the scheme in which airborne LiDAR data acquisition is carried out simultaneously with sample site investigation. The most effective and safe way to ensure the accuracy of forest attribute estimation is to increase the number of sample plots appropriately faced with such a large area. However, the conclusion that the method of post-stratification of LIDAR data can also improve the estimation accuracy is worthy of further experimental study[50]. Another finding of this study is that the number of sample plots is not necessary to affect the accuracy of the forest attribute estimation model. Provided the sample plots are properly selected, high model accuracy can be achieved even if the number of sample plots is small (e.g., 30), which shows that the model accuracy is closely related to the selection of sample plots. This also explains why better estimation accuracy can be achieved even with a small sample size in some studies of small areas. However, in large-area forest resource investigation and monitoring applications, facing huge overall ranges, sample layout can only rely

on the reliability of limited data through typical sampling methods to reduce the uncertainty of forest attribute estimation results. Increasing the number of samples as much as possible is the recommended approach due to its reliability. Therefore, for different forest types in subtropical regions, the minimum number of sample plots required for operational ALS-based large-scale forest stand inventories is proposed, and it is expected to provide a reference for the development of this technical solution.

Our study showed that, with the increase in sample size, the improvement trend for the accuracy of each forest attribute estimation model for each forest type was essentially the same as the decreasing trend for the standard deviation of the target attributes and the LiDAR variable. Changes in the forest attributes and the LiDAR variables of the construction model were key factors affecting the estimation accuracy of the model. This also illustrates the mechanism of the influence of sample size on the accuracy of forest attribute estimation—the variation of target attributes and LiDAR variables determines the model accuracy. This finding effectively explains why LiDAR-stratified selection of plots can reduce the sample size and has instructive significance for other kinds of plot layout methods, such as target sampling and forest attribute estimation after LiDAR data stratification.

In this study, we used the exhaustive method to establish the multivariate power model of each forest attribute estimation of each forest type based on 13 LiDAR variables that describe the three-dimensional structure of the forest canopy. On this basis, we explored the effect of different sample sizes on the estimation of forest attributes. However, there are many types of forest attribute estimation models based on LiDAR. Although it is impossible to exhaust all of the models, it is still necessary to conduct more model fitting and validating tests to further understand the impact of sample size on the accuracy of forest attribute estimation.

5. Conclusions

Through the study on the estimation of forest attributes in a large-scale subtropical forest with a complex structure and significant heterogeneity, we preliminarily clarified the influence law of the sample size on the estimation accuracy of airborne LiDAR forest attributes. With the increase in sample size, the accuracy of each attribute estimation model of each forest type gradually improved. The main reason for this is that the variation of target attributes and LiDAR variables decreased with an increase in sample size. By applying airborne LiDAR to large-scale subtropical forest attribute estimation, the sample size of the Chinese fir forest, pine forest, eucalyptus forest, and broad-leaved forest should be no less than 110, 80, 85, and 70, respectively. The results of this paper have significant reference value for similar research and applications in large-scale areas, and they suggest that it is beneficial to optimize the cost and overall efficiency of forest inventory.

References

1. Næsset, E. Practical large-scale forest stand inventory using a small airborne scanning laser. *Scand. J. For. Res.* **2004**, *19*, 164–179.
2. Maltamo, M.; Packalen, P. *Species-Specific Management Inventory in Finland. Chapter 11*; Maltamo, M., Næset, E., Vauhkonen, J., Eds.; Forestry Applications of Airborne Laser Scanning, Concepts and Case Studies: Dordrecht, The Netherlands, 2014; 464p.
3. Turner, R.; Goodwin, N.; Friend, J.; Mannes, D.; Rombouts, J.; Haywood, A. A national overview of airborne lidar applications in Australian forest agencies. In Proceedings of the SilviLaser 2011, Hobart, Tasmania, Australia, 16–19 October 2011; 13p.
4. White, J.C.; Wulder, M.A.; Varhola, A.; Vastaranta, M.; Coops, N.C.; Cook, B.D.; Pitt, D.; Woods, M. A best practices guide for generating forest inventory attributes from airborne laser scanning data using the area-based approach. In *Information Report FI-X-10*; Canadian Forest Service, Canadian Wood Fibre Centre, Pacific Forestry Centre: Victoria, BC, Canada, 2013; 50p. Available online: http://www.cfs.nrcan.gc.ca/pubwarehouse/pdfs/34887.pdf (accessed on 20 October 2020).
5. Novo-Fernández, A.; Barrio-Anta, M.; Recondo, C.; Cámara-Obregón, A.; López-Sánchez, C.A. Integration of national forest inventory and nationwide airborne laser scanning data to improve forest yield predictions in North-Western Spain. *Remote Sens.* **2019**, *11*, 1693, doi:10.3390/rs11141693.
6. Næsset, E. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote. Sens. Environ.* **2002**, *80*, 88–99.
7. Jensen, J.L.R.; Humes, K.S.; Conner, T.; Williams, C.J.; DeGroot, J. Estimation of biophysical characteristics for highly variable mixed-conifer stands using small-footprint Lidar. *Can. J. For. Res.* **2006**, *36*, 1129–1138
8. Maltamo, M.; Eerikainen, K.; Packalen, P.; Hyyppa, J. Estimation of stem volume using laser scanning-based canopy height metrics. *Forestry* **2006**, *79*, 217–229, doi: 10.1093/forestry/cpl007.
9. Thomas, V.; Treitz, P.; McCaughey, J.; Morrison, I. Mapping stand-level forest biophysical variables for a mixedwood boreal forest using Lidar: an examination of scanning density. *Can. J. For. Res.* **2006**, *36*, 34–47.
10. Montaghi, A.; Corona, P.; Dalponte, M.; Gianelle, D.; Chirici, G.; Olsson, H. Airborne laser scanning of forest resources: An overview of research in Italy as a commentary case study. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *23*, 288–300.
11. Bouvier, M.; Durrieu, S.; Fournier, R.A.; Renaud, J.-P. Generalizing predictive models of forest inventory attributes using an area-based approach with airborne LiDAR data. *Remote Sens. Environ.* **2015**, *156*, 322–334, doi:10.1016/j.rse.2014.10.004.
12. White, J.C.; Tompalski, P.; Vastaranta, M.; Wulder, M.A.; Saarinen, N.; Stepper, C.; Coops, A.C. A model development and application guide for generating an enhanced forest inventory using airborne laser scanning data and an area-based approach. In *Natural Resources Canada, Canadian Forest Service, Canadian Wood Fibre Centre, Information Report FI-X-018*; 2017.
13. Nasset, E.; Gobakken, T.; Holmgren, J.; Hyyppa, H.; Hyyppa, J.; Maltamo, M.; Nilsson, M.; Olsson, H.; Persson, A.; Soderman, U. Laser scanning of forest resources: The nordic experience. *Scand. J. For. Res.* **2004**, *19*, 482–499.
14. Luo, S.Z.; Wang, C.; Zhang, G.B.; Xi, X.H.; Li, G.C. Forest leaf area index (LAI) estimation using airborne discrete-return LiDAR data. *Chin. J. Geophys.* **2013**, *56*, 233–243.
15. Dube, T.; Sibanda, M.; Shoko, C.; Mutanga, O. Stand-volume estimation from multi-source data for coppiced and high forest Eucalyptus spp. silvicultural systems in KwaZulu-Natal, South Africa. *ISPRS J. Photogramm. Remote Sens.* **2017**, *132*, 162–169.
16. Jarron, L.R.; Coops, N.G.; MacKenzie, W.H.; Tompalski, P. Detection of sub-canopy forest structure using airborne LiDAR. *Remote Sens. Environ.* **2020**, *244*, 111770, doi:10.1016/j.rse.2020.111770.
17. Junttila, V.; Finley, A.O.; Bradford, J.B.; Tuomo Kauranne, T. Strategies for minimizing sample size for use in airborne LiDAR-based forest inventory. *For. Ecol. Manag.* **2013**, *292*, 75–85.
18. Fassnacht, F.E.; Hartig, F.; Latifi, H.; Berger, C.; Hernández, J.; Corvalán, P.; Koch, B. Importance of sample size, data type and prediction method for remote sensing-based estimations of above-ground forest biomass. *Remote Sens. Environ.* **2014**, *154*, 102–114, doi:10.1016/j.rse.2014.07.028.
19. Gobakken, T.; Næsset, E. Assessing effects of laser point density, ground sampling intensity, and field sample plot size on biophysical stand properties derived from airborne laser scanner data. *Can. J. For. Res.* **2008**, *38*, 1095–1109.
20. Watt, P.; Watt, M.S. Development of a national model of *Pinus radiate* stand volume from lidar metrics for New Zealand. *Int. J. Remote Sens.* **2013**, *34*, 5892–5904.
21. Ruiz, L.A.; Hermosilla, T.; Mauro, F.; Godino, M. Analysis of the influence of plot size and LiDAR density on forest Structure attribute estimates. *Forests* **2014**, *5*, 936–951, doi:10.3390/f5050936
22. Hernández-Stefanoni, J.L.; Reyes-Palomeque, G.; Castillo-Santiago, M.Á.; George-Chacón, S.P.; Huechacona-Ruiz, A.H.; Tun-Dzul, F.; Rondon-Rivera, D.; Dupuy, J.M. Efects of sample plot size and GPS location errors on aboveground biomass estimates from LiDAR in tropical dry forests. *Remote Sens.* **2018**, *10*, 1586.
23. Lombardi, F.; Marchetti, M.; Corona, P.; Merlini, P.; Chirici, G.; Tognetti, R.; Burrascano, S.; Alivernini, A.; Puletti, N. Quantifying the effect of sampling plot size on the estimation of structural indicators in old-growth forest stands. *For. Ecol. Manag.* **2015**, *346*, 89–97.
24. Zolkos, S.G.; Goetz, S.J.; Dubayah, R. A meta-analysis of terrestrial above-ground biomass estimation using lidarremote sensing. *Remote Sens. Environ.* **2013**, *128*, 289–298.
25. Adnan, S.; Maltamo, M.; Coomes, D.A.; Valbuena, R. Effects of plot size, stand density, and scan density on the relationship between airborne laser scanning metrics and the Gini coefficient of tree size inequality. *Can. J. For. Res.* **2017**, *47*, 1590–1602, doi:10.1139/cjfr-2017-0084
26. Da Silva, V.S.; Silva, C.A.; Mohan, M.; Cardil, A.; Rex, F.E.; Loureiro, G.H.; De Almeida, D.R.A.; Broadbent, E.N.; Gorgens, E.B.; Corte, A.P.D.; et al. Combined Impact of Sample Size and Modeling Approaches for Predicting Stem Volume in Eucalyptus spp.

Forest Plantations Using Field and LiDAR Data. *Remote Sens.* **2020**, *12*, 1438.

27. Xu, G.; Manley, B.; Morgenroth, J. Evaluation of modelling approaches in predicting forest volume and stand age for small-scale plantation forests in New Zealand with RapidEye and LiDAR. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *73*, 386–396

28. Ioki, K.; Imanishi, J.; Sasaki, T.; Morimoto, Y.; Kitada, K. Estimating stand volume in broad-leaved forest using discrete-return LiDAR: plot-based approach. *Landsc. Ecol. Eng.* **2010**, *6*, 29–36, doi:10.1007/sl1355-009-0077-4.

29. Yang, T.-R.; Kershaw, J.A., Jr.; Weiskittel, A.R.; Lam, T.Y.; McGarrigle, E. Influence of sample selection method and estimation technique on sample size requirements for wall-to-wall estimation of volume using airborne LiDAR. *Forestry* **2019**, *92*, 311–323, doi:10.1093/forestry/cpz014.

30. Maltamo, M.; Bollandså, O.M.; Næset, E.; Gobakken, T.; Packalén, P. Different plot selection strategies for field training data in ALS-assisted forest inventory. *Forestry* **2010**, *84*, 23–31.

31. Hawbaker, T.; Keuler, N.; Lesak, A.; Gobakken, T.; Contrucci, K.; Radeloff, V. Improved estimates of forest vegetation structure and biomass with a LiDAR-optimized sampling design. *J. Geophys. Res.* **2009**, *114*, 1–11.

32. Grafströn, A.; Ringvall, A.H. Improving forest field inventories by using remote sensing in novel sampling designs. *Can. J. For. Res.* **2013**, *43*, 1015–1022.

33. Næsset, E. *Area-Based Inventory in Norway-From Innovation to an Operational Reality. Chapter 11*; Maltamo, M., Næset, E., Vauhko-nen, J., Eds.; Forestry Applications of Airborne Laser Scanning, Concepts and Case Studies; Springer: Dordrecht, The Nether-lands, 2015; 464p.

34. Chen, Q.; Laurin, G.V.; Battles, J.J.; Saah, D. Integration of airborne lidar and vegetation types derived from aerial photography for mapping above-ground live biomass. *Remote Sens. Environ.* **2012**, *121*, 108–117

35. Kim, E.; Lee, W.K.; Yoon, M.; Lee, J.Y.; Son, Y.; Abu Salim, K. Estimation of Voxel-based above-ground biomass using airborne LiDAR data in an intact tropical rain forest, Brunei. *Forests* **2016**, *7*, 259.

36. Singh, K.K.; Chen, G.; Vogler, J.B.; Meentemeyer, R.K. When big data are too much: Effects of LiDAR returns and point density on estimation of forest biomass. *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* **2016**, *9*, 3210–3218.

37. Asner, G.P.; Mascaro, J. Mapping tropical forest carbon: calibrating plot estimates to a simple LiDAR metric. *Remote Sens Environ.* **2014**, *140*, 614–624.

38. Goetz, S.; Dubayah, R. Advances in remote sensing technology and implications for measuring and monitoring forest carbon stocks and change. *Carbon Manag.* **2011**, *2*, 231–244.

39. Latifi, H.; Fassnacht, F.E.; Hartig, F.; Berger, C.; Hernández, J.; Corvalán, P.; Koch, B. Stratified aboveground forest biomass estimation by remote sensing data. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *38*, 229–241.

40. Maltamo, M.; Bollandsas, O.M.; Gobakken, T.; Nasset, E. Large-scale prediction of above-ground biomass in heterogeneous mountain forests by means of airborne laser scanning. *Can. J. For. Res.* **2016**, *46*, 1138–1144, doi:10.1139/cjfr-2016-0086.

41. Wang, C.; Glenn, N.F. A linear regression methods for tree canopy height estimation using airborne lidar data. *Can. J. Remote Sensing* **2008**, *34*, S217–S227.

42. Kauranne, T.; Pyankov, S.; Junttila, V.; Kedrov, A.; Tarasov, A.; Kuzmin, A.; Peuhkurinen, J.; Villikka, M.; Vartio, V.-M.; Sanna Sirparanta, S. Airborne laser scanning based forest inventory: Comparison of experimental results for the Perm Region, Russia and prior results from Finland. *Forest* **2017**, *8*, 72, doi:10.3390/f8030072.

43. Gobakken, T.; Korhonen, L.; Næsset, E. Laser-assisted selection of field plots for an area-based forest inventory. *Silva Fenn.* **2013**, *47*, 943, doi:10.14214/sf.943.

44. Zhang, Z.; Cao, L.; She, G. Estimating Forest Structural Parameters Using Canopy Metrics Derived from Airborne LiDAR Data in Subtropical Forests. *Remote. Sens.* **2017**, *9*, 940, doi:10.3390/rs9090940.

45. Patenaude, G.; Hill, R.A.; Milne, R.; Gaveau, D.L.A.; Brihhs, B.B.J.; Daeson, T.P. Quantify forest above ground carbon content using LiDAR remote sensing. *Remote. Sens. Environ.* **2004**, *93*, 368–380.

46. Silva, C.A.; Hudak, A.T.; Klauberg, C.; Vierling, L.A.; Gonzalez-Benecke, C.; de Padua Chaves Carvalho, S.; Rodriguez, L.C.E.; Cardil, A. Combined effect of pulse density and grid cell size on predicting and mapping above-ground carbon in fast-growing Eucalyptus forest plantation using airborne LiDAR data. *Carbon Balance Manag.* **2017**, *12*, 13.

47. Montealegre, A.L.; Lamelas, M.T.; de la Riva, J.; García-Martín, A.; Escribano, F. Use of low point density ALS data to estimate stand-level structural variables in Mediterranean Aleppo pine forest. *Forestry* **2016**, *89*, 373–382, doi:10.1093/forestry/cpw008.

48. Görgens, E.B.; Packalen, P.; da Silva, A.G.P.; Alvares, C.A.; Campoe, O.C.; Stape, J.L.; Rodriguez, L.C.E. Stand volume models based on stable metrics as from multiple ALS acquisitions in Eucalyptus plantations. *Ann. For. Sci.* **2015**, *72*, 489–498, doi:10.1007/s13595-015-0457-x.

49. Giannico, V.; Lafortezza, R.; John, R.; Sanesi, G.; Pesola, L.; Chen, J. Estimating Stand Volume and Above-Ground Biomass of Urban Forests Using LiDAR. *Remote Sens.* **2016**, *8*, 339, doi:10.3390/rs8040339.

50. McRoberts, R.E.; Gobakken, T.; Næsset, E. Post-stratified estimation of forest area and growing stock volume using lidar-based stratifications. *Remote Sens. Environ.* **2012**, *125*, 157–166.