RODRIGO VIEIRA LEITE

ESTIMATING STEM VOLUME OF EUCALYPT FOREST PLANTATION USING LIDAR: A COMPARISON OF THE AREA AND INDIVIDUAL TREE BASED APPROACHES

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência Florestal, para a obtenção do título de *Magister Scientiae*.

Orientadora: Cibele Hummel do Amaral

Coorientador: Hélio Garcia Leite

VIÇOSA - MINAS GERAIS
2019

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T
L533e
2019

Leite, Rodrigo Vieira, 1992-
Estimating stem volume of eucalypt forest plantation using LiDAR : a comparison of the area and individual tree based approaches / Rodrigo Vieira Leite. – Viçosa, MG, 2019.
42 f. : il. (algumas color.) ; 29 cm.

Inclui apêndice.
Orientador: Cibele Hummel do Amaral.
Dissertação (mestrado) - Universidade Federal de Viçosa.
Referências bibliográficas: f. 31-35.

1. Eucalipto. 2. Sensoriamento remoto. 3. Aprendizado do computador. I. Universidade Federal de Viçosa. Departamento de Engenharia Florestal. Programa de Pós-Graduação em Ciência Florestal. II. Título.

CDO adapt. CDD 22. ed. 634.95245

RODRIGO VIEIRA LEITE

ESTIMATING STEM VOLUME OF EUCALYPT FOREST PLANTATION USING LIDAR: A COMPARISON OF THE AREA AND INDIVIDUAL TREE BASED APPROACHES

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência Florestal, para a obtenção do título de *Magister Scientiae*.

APROVADA: 10 de julho de 2019.

Assentimento:

Rodrigo Vieira Leite

Rodrigo Vieira Leite
Autor

Cibele Hummel do Amaral
Orientadora

AGRADECIMENTOS

A Deus, por estar comigo em todos os momentos, me conduzir, proteger, me dar força para superar as dificuldades, mostrar os caminhos nas horas incertas, e me dar saúde e paz para trabalhar.

À Universidade Federal de Viçosa e ao Departamento de Engenharia Florestal, pela oportunidade de realização do meu curso e ao CNPq pela concessão da bolsa de estudos.

À professora Cibele Hummel do Amaral, pela orientação, apoio, ensinamentos e por toda paciência e ajuda durante a realização desse trabalho.

Ao professor Hélio Garcia Leite, pela coorientação, pelo grande incentivo para a realização do meu mestrado. Além dos conselhos, ensinamentos e apoio na realização desse trabalho.

Aos membros da banca de avaliação desta dissertação Carlos Pedro Boechat Soares e Carlos Alberto Silva. A todos os professores e funcionários do Departamento de Engenharia Florestal que contribuíram para a minha formação. Ao professor Elpídio Inácio Fernandes Filho pela prontidão em ajudar sempre que o procurei e suas sugestões em diversas análises.

Ao meu pai, Domingos Luiz Leite, e minha mãe, Leoneza Helena Vieira Leite, pelo apoio incondicional e por estarem sempre cuidando de mim. Obrigado, mãe e pai, por todas as orações, por todo carinho, e grandes exemplos que vocês são para mim.

À minha irmã Fernanda e cunhado Daniel pelo apoio, e presença na minha vida sempre compartilhando todos os momentos.

À minha namorada Patricia, pelo carinho, amor e cuidado. Pela paciência, pelos momentos e pelo apoio incondicional.

A toda minha família que está sempre torcendo por mim e me apoiando. Em especial ao meu tio Hélio, pelos conselhos, conversas, ensinamentos, e apoio de sempre.

Aos colegas da pós-graduação em Ciência Florestal que estiveram junto comigo durante esses anos em especial os membros do laboratório SIGMA: Lucas, Luana, Acauã, Renata, Duberli, Nero, Eduardo, Brener, Pedro, Écio, Thales e Dayane. Pela excelente

convivência e conversas no dia-a-dia. Em todos os momentos que me ajudaram e que contribuíram para o resultado deste trabalho. Aos colegas Raul e Guilherme por também terem me auxiliado em diversos momentos para análises e processamentos.

A todos que torceram por mim e que de forma direta ou indireta contribuíram para a realização deste trabalho.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001

Obrigado!

"Read critically, write consciously; speak clearly, tell your truth."

Clint Smith

"O coração do homem traça seu caminho,

mas o Senhor lhe dirige os passos."

Provérbios 16.9

RESUMO

LEITE, Rodrigo Vieira, M.Sc., Universidade Federal de Viçosa, julho de 2019. **Estimativa de volume em plantios de eucalipto utilizando LiDAR aerotransportado: Uma comparação entre as abordagens a nível de área e árvore.** Orientadora: Cibele Hummel do Amaral. Coorientador: Hélio Garcia Leite.

A estimativa de parâmetros florestais usando nuvem de pontos de LiDAR é um aprimoramento dos métodos convencionais, pois fornece informações precisas e espacializadas. Neste estudo comparamos o desempenho das abordagens a nível de área (ABA) e árvore (ITD) para esimar o volume em povoamentos de eucalipto. Foram testados cinco métodos na abordagem ABA usando *Aritificial Neural Network* (ANN), *Random Forest* (RF), *Support Vector Machine* (SVM), regressão linear e Gompertz, selecionando variáveis de entrada nos modelos usando *Recursive Feature Elimination* e correlação de Spearman para métodos não-paramétricos e paramétricos, respectivamente. Para o ITD, foram testados três métodos para detectar árvores (dois filtros máximos locais e o método da bacia hidrográfica), adicionando um novo procedimento gerando um *buffer* de exclusão de árvores para melhorar a precisão da detecção de arvores. Também foram testados quatro métodos para estimar o *dap* usando ANN e as equações de Gompertz, logística e exponencial. O volume de árvores nessa abordagem foi calculado utilizando o modelo de Schumacher e Hall. Além disso, foram avaliados os efeitos do ângulo de escaneamento, inclinação do terreno e idade do talhão na contagem de árvores, medição de altura e estimativa de volume. No geral, a abordagem ABA teve um melhor desempenho na estimativa de volume para o talhão ($r = 0,95$ e RMSE = 14,41%) não diferindo estatisticamente ($p > 0,05$) dos volumes medidos em campo para o teste F de Graybill e teste-t para o erro médio. As estimativas ITD, embora com desempenho semelhante ($r = 0,97$ e RMSE = 16,4%), tenderam a subestimar o volume em povoamentos mais jovens e em terrenos com inclinação até 25%, além disso diferiram estatisticamente ($p < 0,05$) dos volumes medidos em campo no teste F de Graybill e teste

t para o erro médio. Os mapas de volume também foram diferentes para as abordagens, em que ITD teve valores distribuídos em mais classes que ABA. É discutido neste trabalho a importância das métricas LiDAR como variáveis de entrada nos métodos de estimativa e os possíveis problemas relacionados ao desempenho do ITD.

Palavras-chave: Eucalyptus. Detecção de árvores. Aprendizado de máquina. Sensoriamento remoto

ABSTRACT

LEITE, Rodrigo Vieira, M.Sc., Universidade Federal de Viçosa, July, 2019. **Estimating stem volume of eucalypt forest plantation using lidar: a comparison of the area and individual tree based approaches.** Adviser: Cibele Hummel do Amaral. Co-adviser: Hélio Garcia Leite.

Forest plantations are globally important for the economy and are significant for carbon sequestration. Properly managing plantations requires accurate information about stand timber stocks. In this study, we used the area (ABA) and individual tree (ITD) based approaches for estimating stem volume in fast-growing Eucalyptus spp forest plantations. Herein, we propose a new method to improve individual tree detection (ITD) in dense canopy homogeneous forests and assess the effects of stand age, slope and scan angle on ITD accuracy. Field and Light Detection and Ranging (LiDAR) data were collected in Eucalyptus urophylla x Eucalyptus grandis even-aged forest stands located in the mountainous region of the Rio Doce Valley, southeastern Brazil. We tested five methods to estimate volume from LiDAR-derived metrics using ABA: Artificial Neural Network (ANN), Random Forest (RF), Support Vector Machine (SVM), and linear and Gompertz models. LiDAR-derived canopy metrics were selected using the Recursive Feature Elimination algorithm and Spearman's correlation, for nonparametric and parametric methods, respectively. For the ITD, we tested three ITD methods: two local maxima filters and the watershed method. All methods were tested adding our proposed procedure of Tree Buffer Exclusion (TBE), resulting in 35 possibilities for treetop detection. Stem volume for this approach was estimated using the Schumacher and Hall model. Estimated volumes in both ABA and ITD approaches were compared to the field observed values using the F-test. Overall, the ABA with ANN was found to be better for stand volume estimation (r = 0.95 and RMSE = 14.4%). Although the ITD results showed similar precision (r = 0.94 and RMSE = 16.4%) to the ABA, the results underestimated stem

volume in younger stands and in gently sloping terrain (< 25%). Stem volume maps also differed between the approaches; ITD represented the stand variability better. In addition, we discuss the importance of LiDAR metrics as input variables for stem volume estimation methods and the possible issues related to the ABA and ITD performance.

Keywords: Eucalyptus. Tree detection. Machine learning. Remote sensing

SUMÁRIO

Chapter 1: Estimating Stem Volume in Eucalyptus plantations Using Airborne LiDAR: A comparison of Area- and Individual Tree-based Approaches

Published in *Remote Sensing* **2020**, *12*(9), 1513; **https://doi.org/10.3390/rs12091513**

Rodrigo Vieira Leite [1,*], Cibele Hummel do Amaral [1], Raul de Paula Pires [2], Carlos Alberto Silva [3,4], Carlos Pedro Boechat Soares [1], Renata Paulo Macedo [5], Antonilmar Araújo Lopes da Silva [6], Eben North Broadbent [7], Midhun Mohan [8] and Hélio Garcia Leite [1]

[1] Department of Forest Engineering, Federal University of Viçosa, Viçosa, MG, 36570-900, Brazil; chamaral@ufv.br (C.H.A); csoares@ufv.br (C.P.B.S.); hgleite@ufv.br (H.G.L.)

[2] School of Agrifood and Forestry Science and Engineering, University of Lleida, Lleida, 25198, Spain; rdp8@alumnes.udl.cat

[3] School of Forest Resources and Conservation, University of Florida, Gainesville, FL, 32611, USA; c.silva@ufl.edu

[4] Department of Geographical Sciences, University of Maryland, College Park, Maryland, MD 20740, USA

[5] Klabin S/A, Telêmaco borba, PR, 84275-000, Brazil; renata.macedo@klabin.com.br

[6] Celulose Nipo-brasileira S.A. (CENIBRA), Belo Oriente, MG, 35196-000, Brazil; antonilmar.silva@cenibra.com.br

[7] Spatial Ecology and Conservation (SPEC) Lab, School of Forest Resources and Conservation, University of Florida, Gainesville, FL, 32611 USA; eben@ufl.edu

[8] Department of Geography, University of California – Berkeley, Berkeley, CA 94709, USA; mid_mohan@berkeley.edu

Abstract: Forest plantations are globally important for the economy and are significant for carbon sequestration. Properly managing plantations requires accurate information about stand timber stocks. In this study, we used the area (ABA) and individual tree (ITD) based approaches for estimating stem volume in fast-growing Eucalyptus spp forest plantations. Herein, we propose a new method to improve individual tree detection (ITD) in dense canopy homogeneous forests and assess the effects of stand age, slope and scan angle on ITD accuracy. Field and Light Detection and Ranging (LiDAR) data were collected in *Eucalyptus urophylla x Eucalyptus grandis* even-aged forest stands located in the mountainous region of the Rio Doce Valley, southeastern Brazil. We tested five methods to estimate volume from LiDAR-derived metrics using ABA: Artificial Neural Network (ANN), Random Forest (RF), Support Vector Machine (SVM), and linear and Gompertz models. LiDAR-derived canopy metrics were selected using the Recursive Feature Elimination algorithm and Spearman's correlation, for nonparametric and parametric methods, respectively. For the ITD, we tested three ITD methods: two local maxima filters and the watershed method. All methods were tested adding our proposed procedure of Tree Buffer Exclusion (TBE), resulting in 35 possibilities for treetop detection. Stem volume for this approach was estimated using the Schumacher and Hall model. Estimated volumes in both ABA and ITD approaches were compared to the field observed values using the F-test. Overall, the ABA with ANN was found to be better for stand volume estimation ($r_{y\hat{y}}$ = 0.95 and RMSE = 14.4%). Although the ITD results showed similar precision ($r_{y\hat{y}}$ = 0.94 and RMSE = 16.4%) to the ABA, the results underestimated stem volume in younger stands and in gently sloping terrain (< 25%). Stem volume maps also differed between the approaches; ITD represented the stand variability better. In addition, we discuss the importance of LiDAR metrics as input variables for stem volume estimation methods and the possible issues related to the ABA and ITD performance.

1. Introduction

Forest plantations are globally important for the economy and for carbon sequestration. Properly managing them requires accurate information about stand timber stocks. Foresters often need to sample trees in the field to obtain reference mean values of tree height and volume, which they later extrapolate to the entire stand. This approach, although effective, is generally laborious, costly, and may incur errors due to not measuring the entire population [1–3].

Remote sensing technologies have been evolving and can be used as tools to aid foresters in obtaining necessary parameters to improve management decision-making. In the last decades, one of the main studied technologies is the laser scanner, namely LiDAR (Light Detection and Ranging) [4,5]. LiDAR sensors work by tracking the emission and return of laser pulses (light emitted energy), determining the sensor-target distance based on the return time lag. This technique demonstrates the potential to improve estimates of forest parameters (e.g., diameter at 1.3 m above ground (*dbh*), volume, and biomass) since it extends spatial analysis beyond the x- and y-axes, generating three-dimensional information [6]. Since the '80s, LiDAR has been used to estimate forest parameters [5,7–12]. Aerial platforms are often used (named Airborne Laser Scanner —ALS) for the benefit of covering larger areas to provide spatialized estimates (or "wall-to-wall" estimates) [2].

There are primarily two main approaches to estimate forest parameters using ALS: (1) area-based approach (ABA) [13] and (2) individual tree detection (ITD) [14]. For the former, point cloud statistics, namely LiDAR metrics (e.g., total number of points, point elevation kurtosis, elevation percentiles) are often input features for allometric models. In the ITD, trees are automatically detected to obtain direct measures as stem density, tree height, and crown area. Then, indirect parameters can be derived through allometric relationships between tree data and the variable of interest, such as *dbh* and wood volume.

In general, the ABA is more widely used than ITD based approaches, due to its relative simplicity and acceptable accuracy. In planted forests of Eucalyptus spp. and its hybrids, which are over 5 million hectares in Brazil alone [15], studies have demonstrated the capability of the former approach for providing accurate-spatialized estimates of forest parameters [16–24]. Nonetheless, oftentimes there occur issues with selecting non-collinear and meaningful LiDAR-derived metrics for the models [25,26]. Additionally, this approach has limited ability in terms of offering a direct measurement of stem density, height, and *dbh* distribution. Moreover, the estimates are spatialized in tiles, often the same size as the sample plots, where accuracy could be affected by several factors such as missing variances on a smaller scale [5]. Whereas in the ITD approach, the automatic detection of the treetops seems not to be trivial and several studies have developed methods for performing this task [3,27–34]. Researchers have also applied this approach in Eucalyptus stands [35]; however, stem density and crown shape have been found to be potential issues [36–38]. Even then, the application of both approaches for the same area to account for the accuracy trade-offs has not been explored much so far.

Furthermore, both approaches are subjected to the characteristics of the point cloud. These characteristics may vary depending on the land cover (e.g., species, crown shape, and stem density) [39], site (e.g., topography) [13,39], and sensor (e.g., flight altitude, scan angle, footprint, and point density) [23,40,41]. Those conditions might make forest attributes modeling specific, that is, so the approach performance changes case-by-case [32,42]. Adding to that, plot size and geolocation might impact the building of relationships and the validation with field data [43,44].

In this study, we compare ABA and ITD approaches to estimate the stem volume of Eucalyptus forest plantations. Our objectives were to (1) find out if one approach outperforms the other, assessing the benefits of each for estimating stem volume in the context of our study site and, (2) propose a method for overcoming issues related to tree detection in homogeneous Eucalyptus forest plantations. The study site spans an area of varying slope conditions and stand age. We investigated the effects of stand age, slope, and scan angle on tree detection accuracy and tested commonly used parametric and non-parametric methods to obtain the best possible model in the ABA.

2. Materials and Methods

2.1. Study Site

The study site is located in the Rio Doce Valley, Minas Gerais, southeastern Brazil (Figure 1). This State is the first in terms of area of Eucalyptus plantations in Brazil [15]. The main characteristic of the region is the high variation in slope values and frequency of steep terrains. The total study area is 471.48 ha divided into 19 *Eucalyptus urophylla x Eucalyptus grandis* clone stands planted for pulpwood production. We used forest stands covering three age classes: young (< 4 years), intermediate (between 4 and 6 years), and mature (> 6 years) (see Table 1).

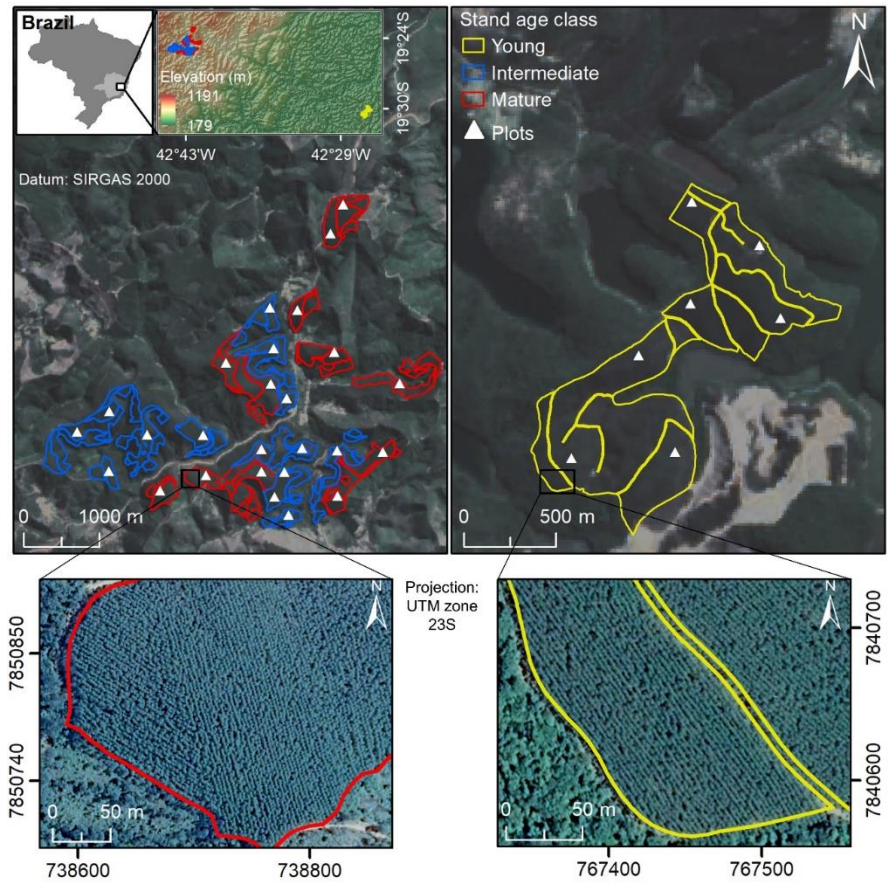

Figure 1. Map of the study area showing stand and plot locations over the mountainous region of the Rio Doce Valley, Minas Gerais, Brazil.

2.2. Field Measurements

A total of 33 rectangular field plots of ~280 m² were measured in these stands as part of previous forest inventory. Tree *dbh* was measured with tree calipers and heights with Haglof digital clinometer (Haglof Inc., Madison, MS, USA). The outside-bark volumes (m^3) were calculated using the Schumacher and Hall [45] equation fitted using historical tree scaling data in the area of study (see Supplementary Material). The field measurements were used as observed values to compare to the LiDAR volume estimates (Table 1). In order to make this comparison, one geographic position from each rectangular plot was collected with a dual-frequency GNSS receiver TOPCON HiPer II (Topcon Corporation, Tokyo, JPN) with estimated uncertainty < 2 m in stand-alone mode. We created a 9.44 m-buffer around the geographic point to represent the total area of each plot. The single point in each plot was the only positional reference the stand owners had for the plots, and previous stand harvest hampered us in obtaining more geographic positions for this study. This procedure, although not ideal, was usable because of the nature of

our target. Even-aged clone trees have considerably low structural variability, especially at this spatial scale (Table S2), and plantation spacing is standardized within stands.

Table 1. Stand characteristics and mean values by age class of field measurements in the study area. Where: *Dbh* = mean diameter at 1.3 m above ground; *H* = mean height; *V* = tree outside-bark mean volume; N = mean number of trees per hectare.

Age class	Number of plots	Total area (ha)	*Dbh* (cm)	*H* (m)	*V* (m^3ha^{-1})	N/ha
Mature	12	186.91	26.3 ± 0.91	27.7 ± 2.02	494.40 ± 38.97	1095
Intermediate	14	188.84	24.5 ± 0.92	23.0 ± 2.69	324.74 ± 61.55	1042
Young	7	95.73	13.5 ± 0.75	20.6 ± 1.22	133.11 ± 25.87	978

Note: Young (< 4 years), intermediate (between 4 and 6 years), and mature (> 6 years).

2.3. Lidar Data Acquisition and Processing

LiDAR data was acquired in an airborne mission with the aircraft Cessna model 206, which had an average height above ground of 618 m, speed of 55 m/s and pulse frequency of 300 kHz, yielding a point cloud with average 5 points per square meter. The laser beam divergence resulted in a footprint of 0.31 m on the ground. The maximum scan angle of the sensor was 30° covering an average of 713 m per swath. The survey was accomplished in February 2014 near to the dates when field data was collected (June – August 2014).

Prior to the application of the two approaches, we derived the Digital Terrain Model (DTM), normalized point cloud, and Canopy Height Model (CHM). We classified the point cloud into ground and non-ground returns with the adapted Kraus and Pfeifer [45] morphological filter, in the software FUSION [46]. This filter works by classifying points as ground (1) and non-ground (0) in an iterative process following a weight function. Detailed information on the algorithm operation is available in [47]. We tested combinations of parameters g (−4, −3, −2, −1, 0 and 1) and w (1, 2, 3 and 4) from the morphological filter equation and held the best result (smoother DTM) with the values -3 and 3, respectively. Parameters a and b were kept constant as 1 and 4 since these values have shown adequate performance in other applications [46,47]. We created the DTM by filtering the terrain points and rasterizing it to a 0.45 x 0.45 m cell size [48]. The average elevation of points within each cell is defined as the cell value. We then normalized the whole point cloud (all ground elevations equal to 0) and created the CHM. The CHM represents the canopy heights and we used it to detect trees in the ITD approach, while the normalized point cloud was used to calculate LiDAR metrics for the ABA.

2.4. Modeling Forest Stand Volume

Under ABA, we tested two parametric and three nonparametric methods: Linear regression, Gompertz model [49,50], Artificial Neural network (ANN) [51.52], Random Forest (RF) [53], and Support Vector Machine (SVM) [54,55]. Feature selection, tuning and accuracy assessment of the models described in Section 2.5.2. In the ITD, we first assessed tree detection accuracy, proposing a new method to improve it, and estimated stand volume by summing up tree volumes calculated with the Schumacher and Hall [45] model (Equation (1)).

$$\ln V = b_0 + b_1 \ln dbh + b_2 \ln h + \varepsilon \quad (1)$$

where: V = Tree volume, *dbh* = tree diameter at breast height (1.3 m above ground) and h = tree height.

2.5. Area-based Approach (ABA)

2.5.1. LiDAR Metrics and Feature Selection

From all the elevation metrics available within the software FUSION [46], we defined as candidate features those related to forest structure and that had been used in other studies of

forest applications of ALS [20,23,56] (Table 2). The lower threshold for the metrics was 1.3 m (i.e., only points above this height were considered in the metric calculation).

We selected the LiDAR metrics for the nonparametric methods considering two aspects: collinearity and importance. First, we excluded metrics that were highly correlated using a ± 0.9 correlation coefficient threshold [20,22]. Subsequently, we used a method based on Recursive Feature Elimination (RFE) [57], considering only the features not excluded in the first step to account for feature importance in the models. This method is a backward selection algorithm that computes feature importance at each iteration by ranking them from the most important to the least, removing a user-defined subset iteratively at each stage [58]. Although feature collinearity may not severely affect nonparametric methods, the exclusion of highly correlated ones was important for making RFE iterations more constant, as features could be interchangeable within the models. We ran the algorithm 100 times with each tested method and selected the most frequently held features considering their position in the importance rank defined by the algorithm. The RFE algorithm is present at the "Caret" package [59] and was run through the "recursive_feature_elimination" function of the Labgeo package [60], both in the R environment [61]. For the parametric methods, we computed the Spearman's correlation coefficient to rank the relationships between the candidate LiDAR metrics and volume. We selected the metric with the highest correlation coefficient with respect to volume as the explanatory variable in the parametric models.

Table 2. Light Detection and Ranging (LiDAR) elevation metrics used as candidate input features in volume estimate models.

Description	Metric	Acronym
Central tendency and range	Minimum	Hmin
	Maximum	Hmax
	Mean	Hmean
	Mode	Hmod
Dispersion and distribution	Standard-deviation	Hstd
	Variance	Hvar
	Coefficient of variation	Hcv
	Absolute deviation median	H.ADmed
	Absolute deviation mode	H.ADmod
	Kurtosis	Hkurt
	Skewness	Hskew
Percentiles	Height 1st percentile	H.P1
	Height 5th percentile	H.P5
	Height 10th percentile	H.P10
	Height 20th percentile	H.P20
	Height 25th percentile	H.P25
	Height 30th percentile	H.P30
	Height 40th percentile	H.P40
	Height 50th percentile	H.P50
	Height 60th percentile	H.P60
	Height 70th percentile	H.P70
	Height 75th percentile	H.P75
	Height 80th percentile	H.P80
	Height 90th percentile	H.P90
	Height 95th percentile	H.P95
	Height 99th percentile	H.P99
Canopy relief ratio	(Hmean - Hmin) / (Hmax - Hmin)	CRR
Canopy cover	% of first returns above 1.3m	Ccover

2.5.2. Estimation Methods Tuning and Fitting

We trained the nonparametric methods (i.e., ANN, RF, and SVM) varying their tuning parameters selecting the best tuning based on correlation coefficient ($r_{y\hat{y}}$; Equation (2)) and absolute and relative Root Mean Square Error (RMSE and RMSE%; Equations (3) and (4), respectively). The ANN is a multi-layer perceptron with backpropagation learning function and logistic activation function for the hidden layers. The tuning parameter is the number of units in the hidden layer (hereafter referred to as "size"). For the RF, we set the number of decision trees to 500 and varied the number of candidate features randomly sampled at each tree-node (hereafter referred to as "mtry"). The SVM was with the Radial Basis Function Kernel tuned varying the parameters "Cost" (C) and sigma. All algorithms were fitted in R [61] using the package Labgeo [60], with dependencies of the packages RSNNS [62], random Forest [63] and kernlab [64] for ANN, RF and SVM, respectively. Before model fitting, we randomly split the dataset into training and test data with 60% and 40% of the data, respectively. The training set was used to train and define the best non-parametric methods using leave-one-out cross validation and to fit the parametric models. The test set was held out from this process and used to apply and assess the methods' performance (external validation; see Section 2.7).

$$r_{y\hat{y}} = \frac{n^{-1}\sum_{i=1}^{n}(\hat{Y}_i - \hat{Y}_m)(Y_i - \bar{Y})}{\sqrt{n^{-1}\sum_{i=1}^{n}(\hat{Y}_i - \hat{Y}_m)^2 n^{-1}(Y_i - \bar{Y})^2}} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2}{n}} \quad (3)$$

$$RMSE\% = \frac{RMSE}{\bar{Y}x100} \quad (4)$$

where $\hat{Y}_i$ = estimated stem volume; Y_i = observed stem volume; $\bar{Y}$ = mean observed stem volume; $\hat{Y}_m$= estimated mean stem volume and n = number of observations.

2.6. Individual Tree Approach (ITD)

2.6.1. Tree Detection and Crown Delineation

In this study, we tested three methods to detect the treetops: two Local Maxima Filter (LMF) algorithms [27,29,31], varying window size; and the watershed method [65] for crown delineation (WTS), using the ArcGIS hydrology tools (ESRI – Redlands, CA, USA) [66]. The latter comes from watershed creation using the inverted CHM, defining its boundaries. This resembles classic watershed algorithms [28]. Treetops are the maximum value within each delineated crown in this method. To improve the tree detection performance, we proposed a new algorithm we call Tree Buffer Exclusion (TBE), to be applied after the tree detection methods.

To understand how it works, consider N as the set of all detected trees in a stand. This exclusion consists of iterations through N, selecting each individual (N_i) and creating a buffer of radius r. If the analyzed tree (N_i) is the highest among those within the buffered zone, it is considered as a real treetop, if not, it is eliminated (Figure 2). Therefore, it allows the removal of points too close to each other, which likely belong to the same crown. This is based on the a priori consideration that trees on even-aged planted stands, like those in this study, have similar canopy

structures. Such homogeneity, together with the high spatial resolution of the CHM might induce the algorithms to count more than one "treetop" per tree.

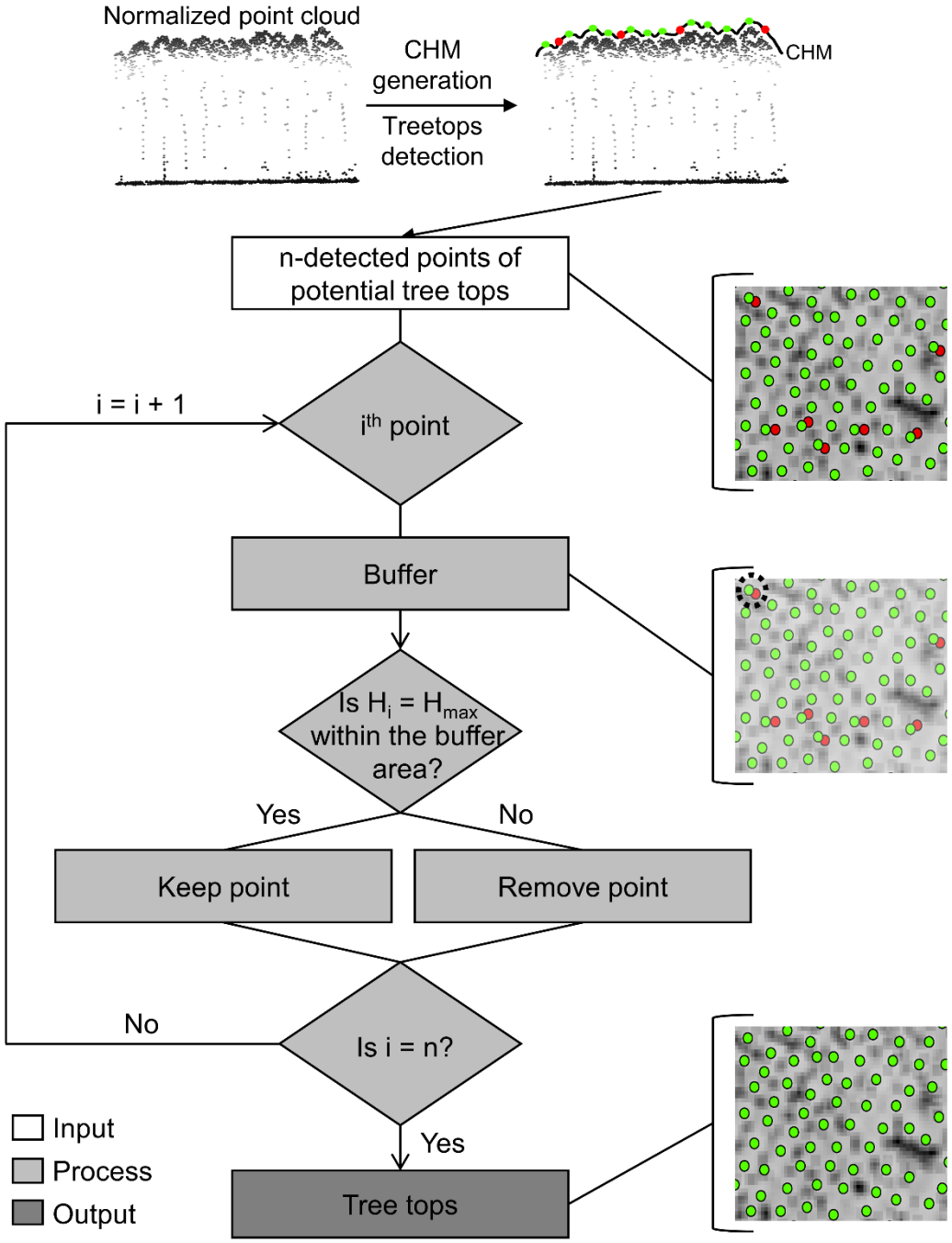

Figure 2. Flowchart representing the steps of individual tree detection and the Tree Buffer Exclusion (TBE). CHM stands for Canopy Height Model. Hi = Elevation of point i; Hmax = maximum height; n = total number of detected points.

The combination of method, window size, and TBE exclusion radius yielded 35 possibilities for treetop detection (Table 3). We defined the best method validating tree count to the estimated stem density in the stand, which was calculated using the sample plots data. We found that to be more reliable than plot validation since we did not have geolocation for trees in the plots, geolocation accuracy was low for this purpose and a slight change in plot position might severely affect the presence/absence of a tree in a plot.

Table 3. Summary of parameters used in each tree detection and delineation algorithm. NA = not applicable. TBE stands for our proposed method of Tree Buffer Exclusion.

	Method	Window Size	TBE radius
LMF1	Popescu et al. (2002)	2,3,4 meters	
LMF2	Silva et al. (2015)	3,5,7 pixels	0, 0.5, 1, 1.5, 2 meters
WTS	Watershed	NA	

We assessed tree detection accuracy by age, slope and scan angle classes. Age classes were the same as defined in Section 2.1. Slope classes were: gentle (<25%), moderate (25%–45%) and steep slope (> 45%). A relatively higher threshold (45%) is used to define mountainous slopes in Brazil [67]. The classes for scan angles were: nadir (0°–7°), off-nadir (7°–23°) and large off-nadir (> 23°) [41]. Scan angles are likely to overlap, mainly between flight tracks. Thus, it was not possible to obtain a reasonable number of plots with a single scan angle class. To assess its effects, we calculated the scan angle class mode using values retrieved for each returned point in a plot and defined it as the single scan-angle class for that plot. Trees in the border of plots were kept for the analysis only if the treetop was within the plot.

2.6.2. Individual Tree *Dbh* and Volume Estimates

We tested four methods to estimate *dbh* (see supplementary files Fig. S1) and found ANN as the best unbiased method to estimate *dbh* for trees in the stand ($r_{y\hat{y}}$ = 0.94, RMSE% = 5.52, bias% = 0.0001; Figure S1). The input variables were height, soil type, stem spacing and regeneration method. With the *dbh* and LiDAR retrieved height for all trees, we estimated volume with the Schumacher and Hall [55] equation. The parameters for this equation were provided by the stand owners and were obtained with historical tree scaling data in the area of study (Table S4). The performance assessment for this approach was conducted comparing the field observed values to the sum of the detected tree volumes in the plot and stand.

2.7. Model Accuracy, Statistic Tests and Comparison of The Approaches

Assessment of all models' performance was carried out computing the correlation coefficient ($r_{y\hat{y}}$) absolute and relative RMSE, and absolute and relative bias (Equation (2) to Equation (6)). We also tested all estimates with the statistic *F* modified by Graybill [68]. This statistic calculates the significance of a simple regression testing the null hypothesis that the parameter b_0 (intercept) is equal to zero and b_1 (slope) to one (H_0: $[\alpha_0\ \alpha_1] = [0\ 1]$) in the model $Y_i = \alpha_0 + \alpha_1\hat{Y}_i + \varepsilon$. The comparisons were for the pairwise combination of estimated ($\hat{Y}_i$) and observed (Y_i) values [69].

$$bias = \frac{\sum_{i=1}^{n}(\hat{Y}_i - Y_i)}{n} \quad (5)$$

$$bias\% = \frac{bias}{\bar{Y}x100} \quad (6)$$

where $\hat{Y}_i$ = estimated stem volume; Y_i = observed stem volume; $\bar{Y}$ = mean observed stem volume and n = number of observations.

2.8. Volume Maps

Volume maps (i.e., wall-to-wall estimates) were created by sectioning the stand into tiles and then estimating the stem volume for each one. The area of each tile was equal to those of the inventory plots (280 m^2). In the ABA, we computed the selected LiDAR metrics for each tile and then applied the estimation methods. For the ITD, we summed up all estimated tree volumes within a tile. Hence, the volume for the whole stand was calculated and can be assessed (Figure 3).

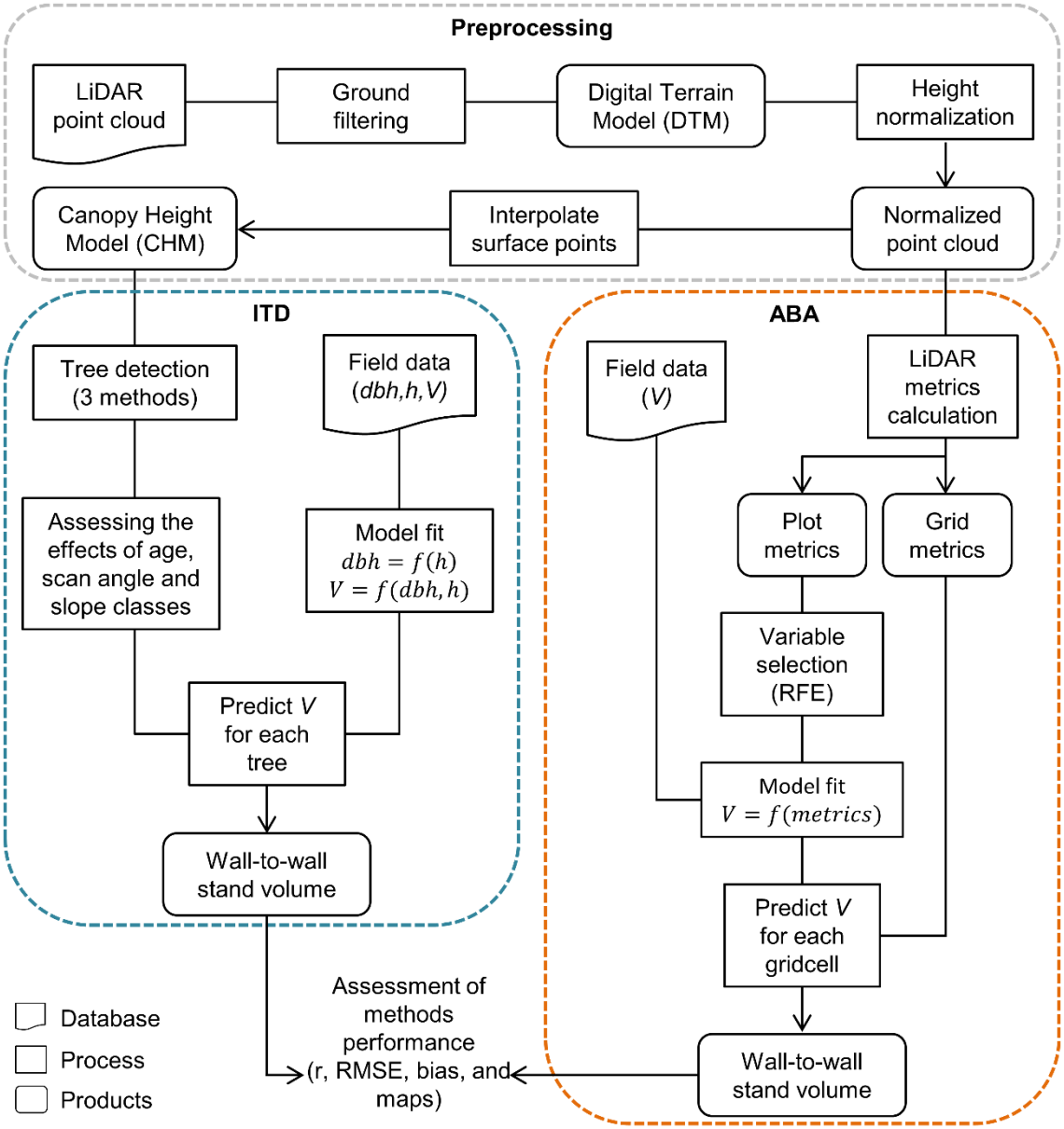

Figure 3. Flowchart showing the steps to estimate and assess stem volume for the whole stand using the area-based (ABA) and individual-tree detection (ITD) approaches.

3. Results

3.1. Area-based Approach (ABA)

In the first feature selection step for the nonparametric methods (ANN, RF, SVM), we removed 15 from the 28 candidate features, based on a correlation threshold of 0.90 (Figure 4). Feature importance ranks were then built using the RFE algorithm with the remaining 13 features.

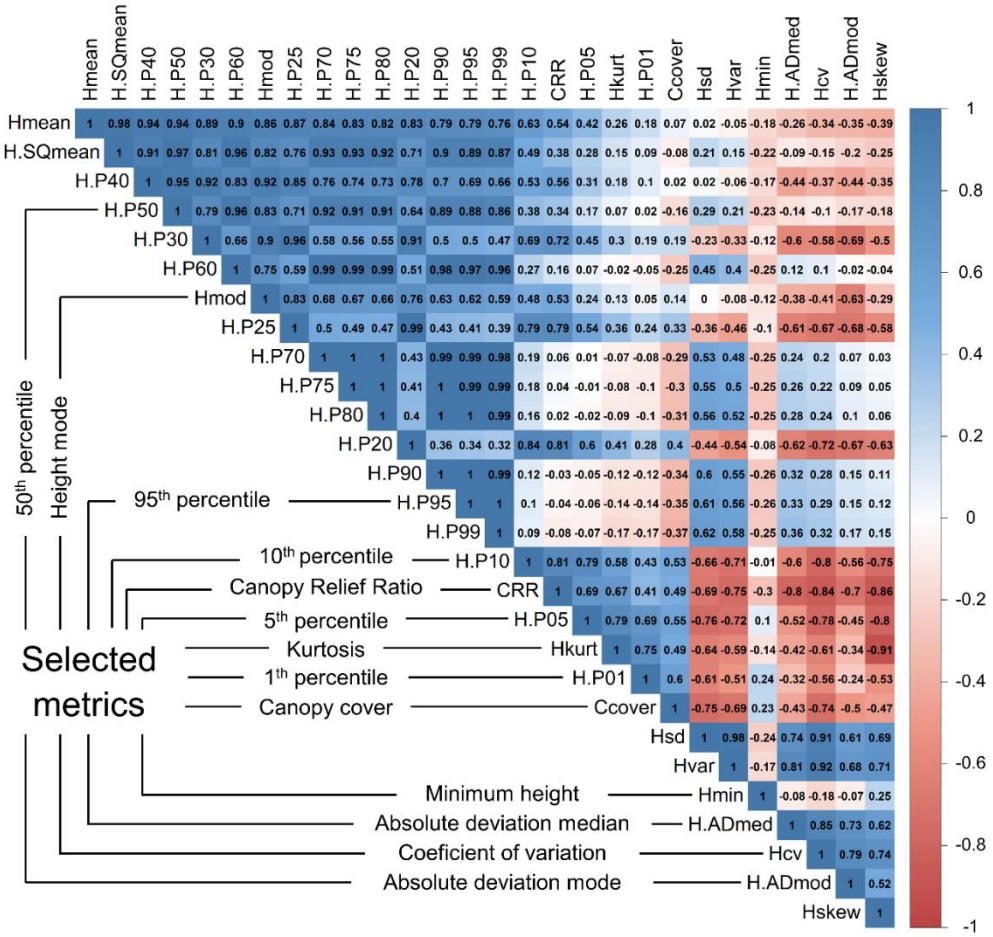

Figure 4. Correlation matrix with Pearson's correlation values amongst LiDAR metrics. The figure highlights the metrics selected for the next step of feature selection. Correlation chart built with R package corrplot [70].

The most frequently ranked features did not significantly change among estimation methods, especially until the fourth most important metric (Table S4). The four most important LiDAR metrics to estimate volume were: H.P95, Hmod, H.P50, and H.P10. Therefore, these metrics were selected as input features for the nonparametric methods. From the fifth position forward, the ranked features changed over the iterations, indicating that the importance of such features for model fitting is not as clear as the first four ranked metrics.

For the parametric methods (linear regression and Gompertz), the variable H.P95 (height at the 95th percentile) was selected based on Pearson's correlation coefficients. Its correlation with volume was the highest (ϱ = 0.8361) amongst LiDAR metrics. The estimation methods were then fitted using the selected metrics (Table 4).

Table 4. Tuning parameters and equations for nonparametric and parametric estimation methods used to estimate stem volume of the Eucalyptus stands.

Estimation Method	Tuning parameter / equation
ANN	size = 9
SVM	C = 4; sigma = 0.0926
RF	mtry = 2
Gompertz	$V_{est} = 60.8615e^{-e^{1.6089-0.03242H.P95}}$
Linear	$ln\,(V_{est}) = 0.00575 + 0.0735H.P95$

Note: Size = number of neurons in the hidden layer; mtry = number of candidate features randomly sampled at each tree-node; C = Cost.

Volume estimates in the parametric and nonparametric methods showed overall Pearson's correlation coefficient close to 0.80 in the test data (i.e., 40% of the data held out from model fitting) (Table 5). The values ranged from 0.79, with linear regression to 0.83 with ANN. The ANN also showed the lowest values for RMSE% (19.39) while the SVM showed the highest (24.71). All estimates with parametric methods were different ($p < 0.05$) from observed values, despite the lowest overall bias with the Linear model.

Table 5. Statistics for the application of estimative methods in the test partition, i.e., 40% of the data held out from fitting the estimation methods. P-values are for the Graybill (1976) F-test.

Method	$r_{y\hat{y}}$	RMSE	RMSE (%)	bias	bias (%)	p-value*
RF	0.80	46.22	22.90	13.220	6.27	0.0672*
SVM	0.80	47.81	24.71	6.026	2.86	0.0675*
ANN	0.83	40.71	19.39	6.784	3.22	0.0532*
Gompertz	0.82	45.91	23.33	7.607	3.61	0.0357
Linear	0.79	46.74	23.14	1.536	0.73	0.0291

3.2. Individual Tree Approach (ITD)

The best method to detect trees varied over the stands (Table 6). However, all tested methods had their performance improved by adding the TBE. The best overall results for the number of detected trees were observed by using LMF1 with a 2 m window and TBE radius equal to 1.5 m. This method had the lowest tree detection errors for 11/19 stands, followed by the WTS + 1.5 m buffer (7/19).

Table 6. Best method for tree detection per stand with the difference between LiDAR and field retrieved stem density. TBE stands for our proposed method of Tree Buffer Exclusion.

Stand	Method	Window Size	TBE radius	Difference (%)
S1	LMF1	2	1.5	−1.37
S2	WTS	NA	1.5	−11.85
S3	LMF1	2	1.5	−6.09
S4	WTS	NA	1.5	−1.91
S5	LMF1	2	1.5	−3.27
S6	LMF1	2	1.5	−0.83
S7	LMF1	2	1.5	−5.75
S8	LMF1	2	1.5	0.91
S9	WTS	NA	1.5	−0.36
S10	WTS	NA	1.5	0.12
S11	WTS	NA	1.5	−4.26
S12	WTS	NA	1.5	−14.02
S13	LMF1	2	1.5	12.33
S14	LMF1	2	1.5	−2.56
S15	LMF1	3	0	−9.46
S16	LMF1	NA	1.5	1.11
S17	WTS	NA	1.5	3.64
S18	LMF1	NA	1.5	1.01
S19	LMF2	3	0	−7.39

Upon comparing LiDAR- and field-measured heights by age classes in the plots, the only difference ($p < 0.05$) found was on the young stands' plots (Figure 5a), where the trend was for underestimating tree mean height (bias% = −11.17). Slope classes also affected tree mean height for the LiDAR measurements. In gentle slopes (< 25%) the differences from LiDAR to field-

measured tree heights were significant showing an underestimation pattern (Figure 5d). In scan-angle classes, the LiDAR retrieved mean height was different ($p < 0.05$) from the field measures at nadir angles (0°–7° scan angle class) (Figure 5g). Although the differences were not significant on large off-nadir angles (23°–30°), the number of observations in this class was too low to make inferences.

Using the LiDAR measured tree heights and estimated *dbh* (see Section 2.6.2), tree volumes were estimated in the plots with $r_{y\hat{y}}$ = 0.8, RMSE% = 23.96 % and bias% = -4.13 %.

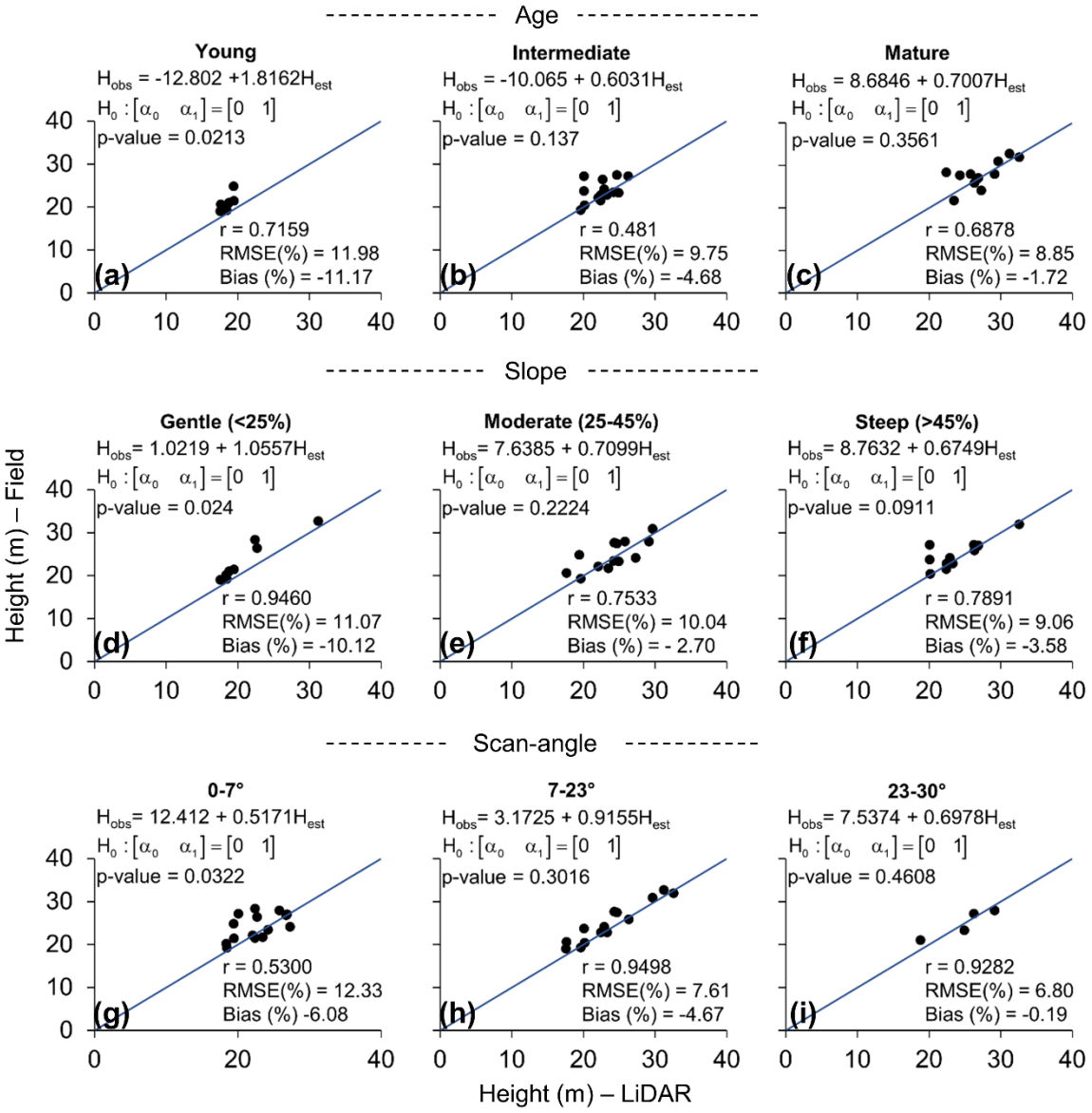

Figure 5. LiDAR and field measured mean height (m) for the sample plots. Charts (a), (b) and (c) are the results for the stand age classes Young (< 4 years), Intermediate (between 4 to 6 years) and Mature (> 6 years). Charts (d), (e) and (f) are the results for the stand slope classes varying from < 25% to > 45% and (g), (h), (i) are the results for the LiDAR data acquisition scan-angle classes 0-7°, 7-23°and 23-30°, respectively. P-values are for the Graybill [68] F-test

3.3. Stand Volume Estimates with ABA and ITD

The stand volume estimations $r_{y\hat{y}}$ and RMSE (%) were overall around 0.9% and 15%, respectively (Figure 6). The best result was obtained with the ABA with ANN where $r_{y\hat{y}}$, RMSE (%) and bias (%) were equal to 0.95, 14.41 and −1.59, respectively. The estimates with the ITD

approach, despite the relatively high $r_{y\hat{y}}$ and low RMSE (%) (0.94 and 16.4, respectively), presented a bias (%) of −9.55 which was the highest absolute value. This tendency was observed mainly in lower volumes, underestimating the observed values (Figure 6f).

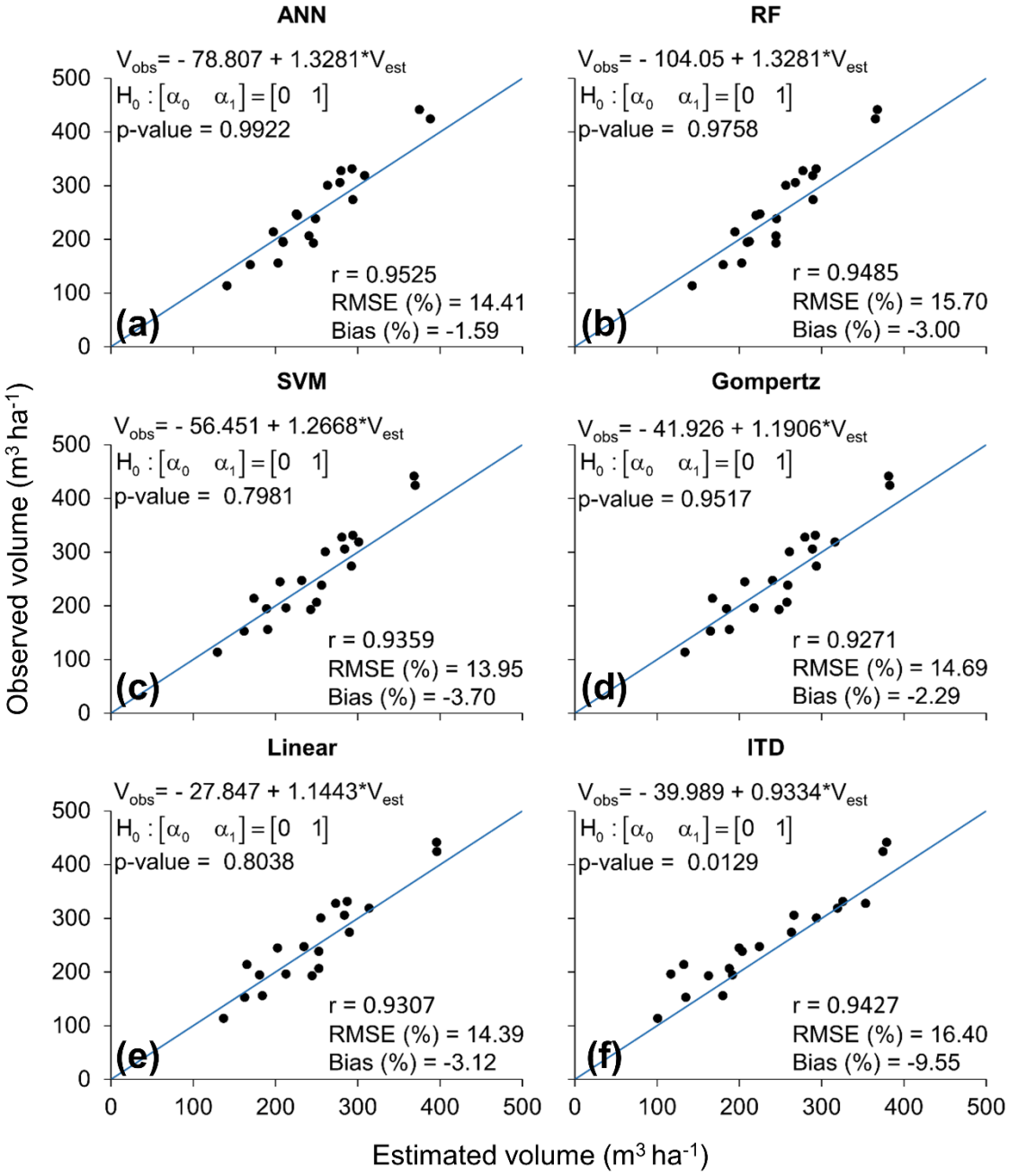

Figure 6. Estimated vs. observed stem volume for the whole stand. Charts (a), (b), (c), (d) and (e) are the estimations with the area-based approach using the methods Artificial Neural Network (ANN), Random Forest (RF), Support Vector Machine (SVM) and the Gompertz and Linear models. Chart (f) is the volume obtained with the Individual Tree Detection (ITD) approach P-values are for the Graybill [68] F-test.

3.4. Stem Volume Mapping

The volume maps showed differences in ABA and ITD estimates (Figure 7, 8 and 9). The main differences were in the volume distribution across the stand. Although both seemed to have normal distribution ITD stand volume distribution presented a wider range than ABA's.

In addition, ABA maps using ANN had the mean volume estimates closer to the ITD maps. In intermediate stands (Figure 8), the ANN had two peaks where estimates were more frequent, showing that stem volume estimates may have different behavior even within one age class.

Stem volume estimates using Gompertz differed from estimates with the ITD mainly in young stands (Figure 7). Values were more frequent in the classes around 140–160 m^3ha^{-1} using Gompertz, while they were around 110-140 m3 with ITD. This change made the distribution with Gompertz slightly skewed in relation to the estimated mean stem volume from the ITD approach. This pattern was attenuated in intermediate and mature stands (Figure 8, Figure 9).

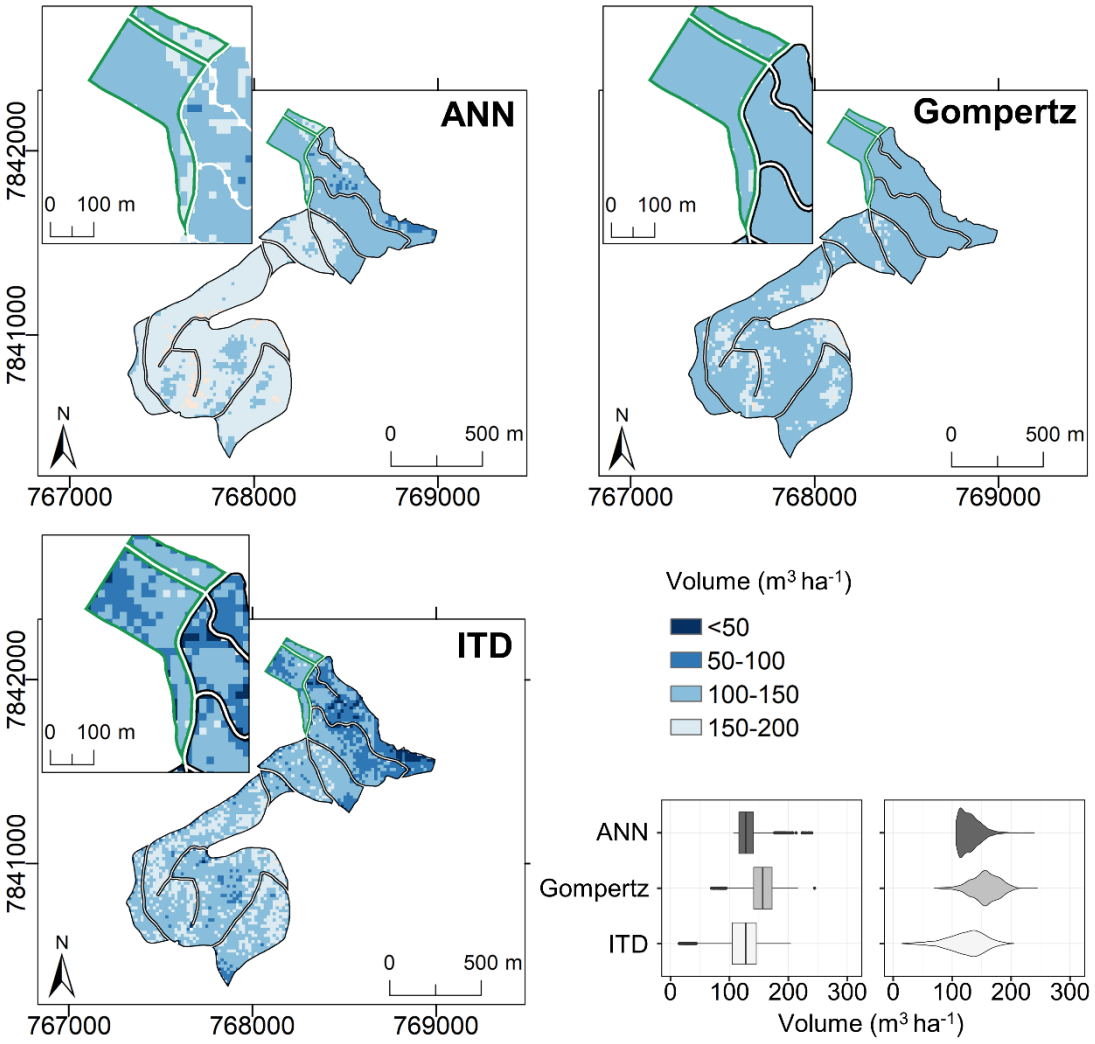

Figure 7. Stem volume map for young stands using Artificial Neural Network (ANN) and Gompertz, in the area- and individual tree (ITD) based approaches. Box and violin plots represent the volume estimates distribution across the stands and methods.

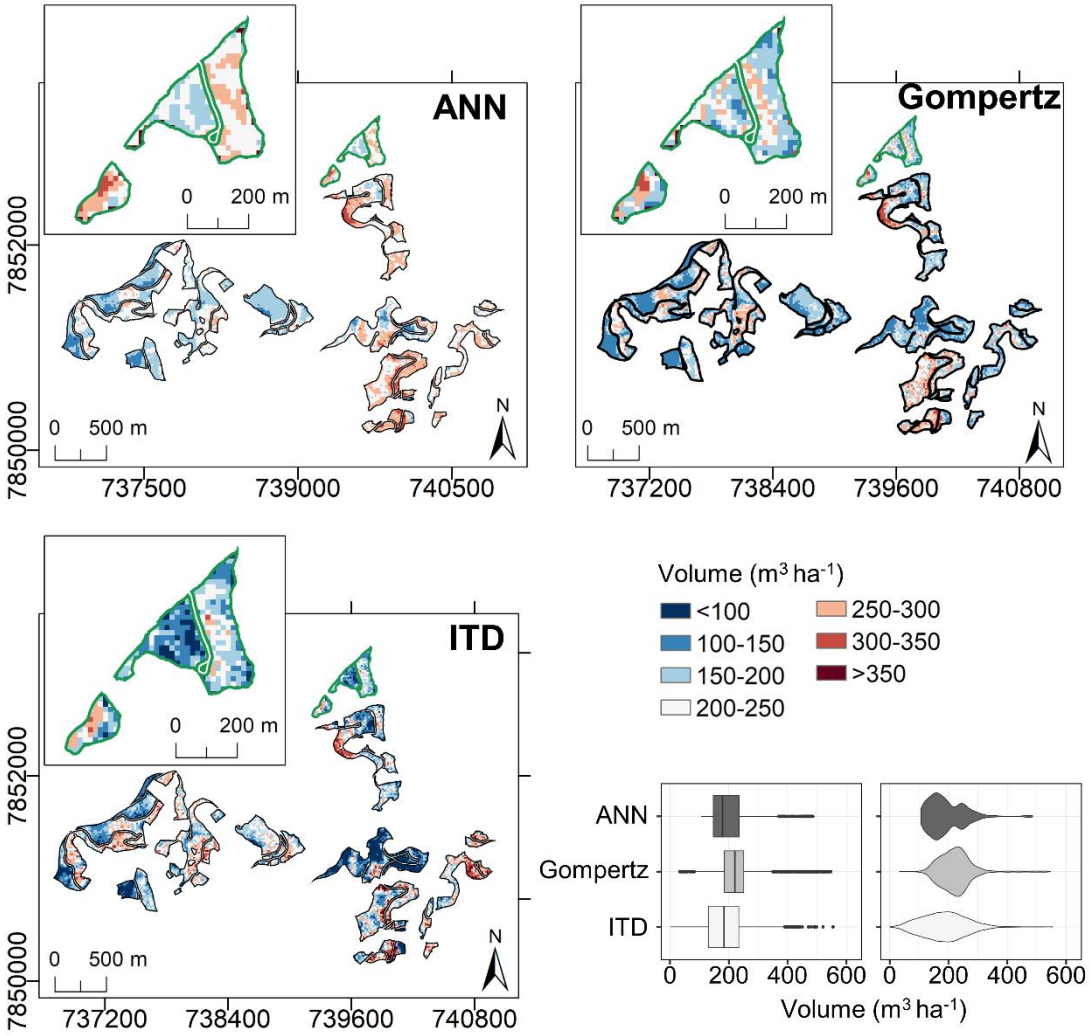

Figure 8. Stem volume map for intermediate stands using Artificial Neural Network (ANN) and Gompertz, in the area- and individual tree (ITD) based approaches. Box and violin plots represent the volume estimates distribution across the stands and methods.

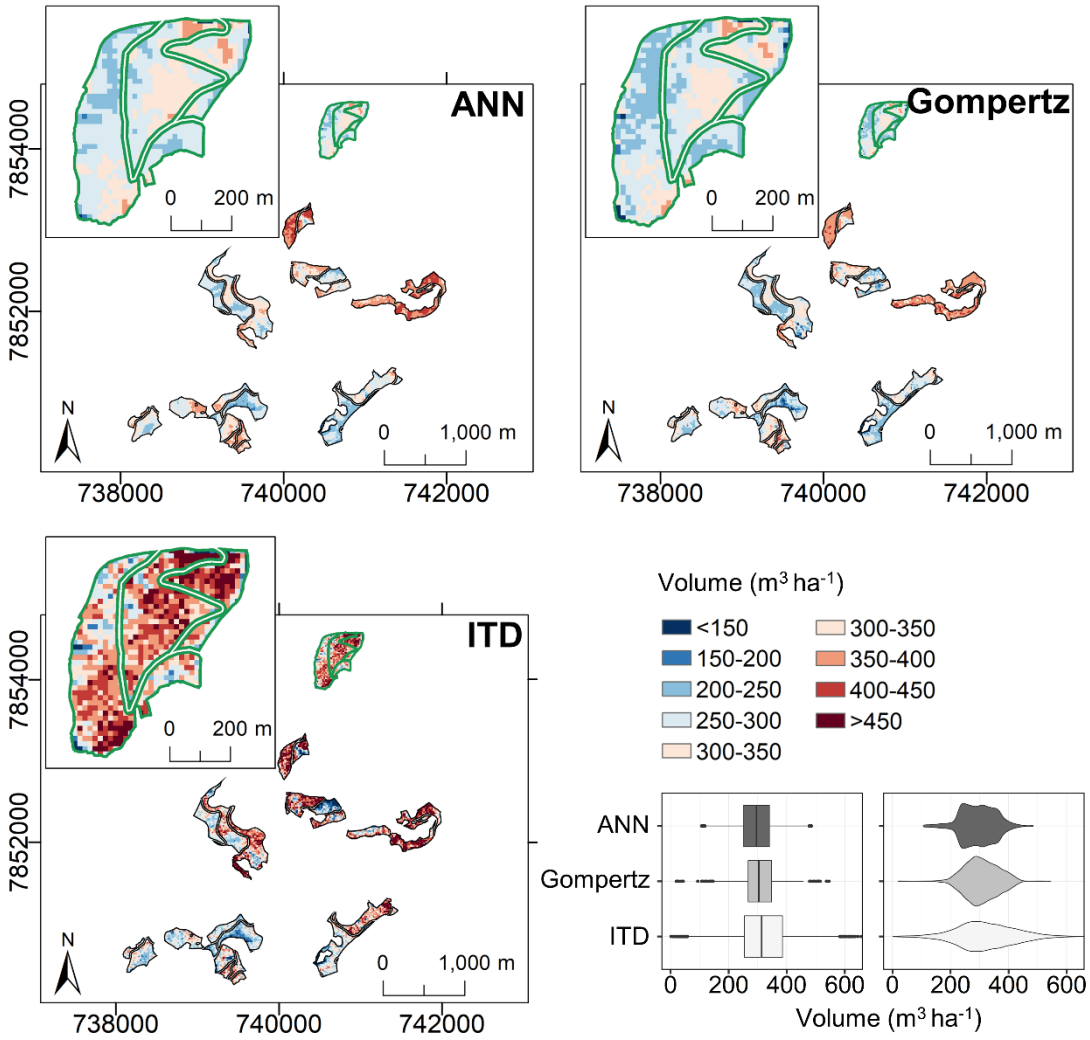

Figure 9. Stem volume map for mature stands using Artificial Neural Network (ANN) and Gompertz, in the area- and individual tree (ITD) based approaches. Box and violin plots represent the volume estimates distribution across the stands and methods.

4. Discussion

4.1. Forest Stands Strata Characterization from LiDAR Metrics

The selected metrics used as input features in the estimating methods were H.P95, Hmode, HP50 and HP10. These metrics appear frequently in other studies that estimate stand volume using LiDAR data, and typically, at least one height percentile is used [9,11,17]. The selected metrics support the concept of using metrics that are related to the canopy height and its variance for volume estimation [71]. They represent specific heights in the point cloud and the same metric can be in a different forest stratum depending on the forest structure (Figure 10). The highest height percentile used was H.P95, which corresponds roughly to the top of the canopy. However, using this metric alone might not account for the whole vertical variability in the stand, even in homogeneous forests as Eucalyptus even-aged stands. Thus, vertical changes in forest structure were detected when other height metrics were inserted into the models (Figure 10).

In this study, the most important metrics after the H.P95 were Hmode, HP50 and HP10. Figure 10 shows a representation of how these metrics account for the stand variability. If the canopy is uniform in a single layer, lower percentiles (e.g., H.P10) give us values near the highest percentiles (e.g., Figure 10a$_1$ and Figure 10d$_1$ where H.P10 = 24.23 m and H.P95 = 27.42 m, difference = 3.19 m). Therefore, in the case of a homogeneous forest, using more than one height metric will add meaning to the model if the crown height is a parameter of interest. Meanwhile,

when there are more returns (i.e., intercepted objects) between the upper stratum and the ground, differences from lower and higher percentiles will be larger (Figure 10a$_3$ and Figure 10d$_3$ where H.P10 = 9.38 m and H.P95 = 33.31 m, difference = 23.93 m). Points below the upper canopy are likely from suppressed trees or limbs not self-pruned during tree growth. Attention should be given to the fact that stands reaching the end of the rotation are frequently without recent silvicultural treatments and may contain undergrowth (e.g., herbs and shrubs) [20]. In addition, lower percentiles (below H.P50) may be unstable if the LiDAR data acquisition is taken at significantly high altitudes due to attenuation of the pulse signal [23].

Removing highly correlated metrics prior to feature selection is commonly applied when estimating forest parameters with LiDAR metrics, performed for example with PCA or the Pearson's coefficient of correlation [20,23,26,72]. Some regression models are more sensitive than others, but even for nonparametric methods, this step is important to remove redundancy and lead to more interpretability of the selected features [72]. The RFE has been applied in other fields of study to select the best subset of features to input in a model, mainly using a large number of candidate features [57,73]. This method benefits from refitting fewer models at each step due to the variable rank creation [58]. This approach directly outputs the most important features for the estimation method and is an option to make decisions when a large number of candidate features are available, which is likely to happen in studies with the integration of LiDAR data from other platforms (e.g., terrestrial and orbital) or with other sensors (e.g., optical and radar). We applied the RFE algorithm oriented by each tested nonparametric method (ANN, RF, and SVM). Although they led to similar results, computation labor was remarkably lower with RF. This algorithm has a well-known internal method for computing variable importance [74].

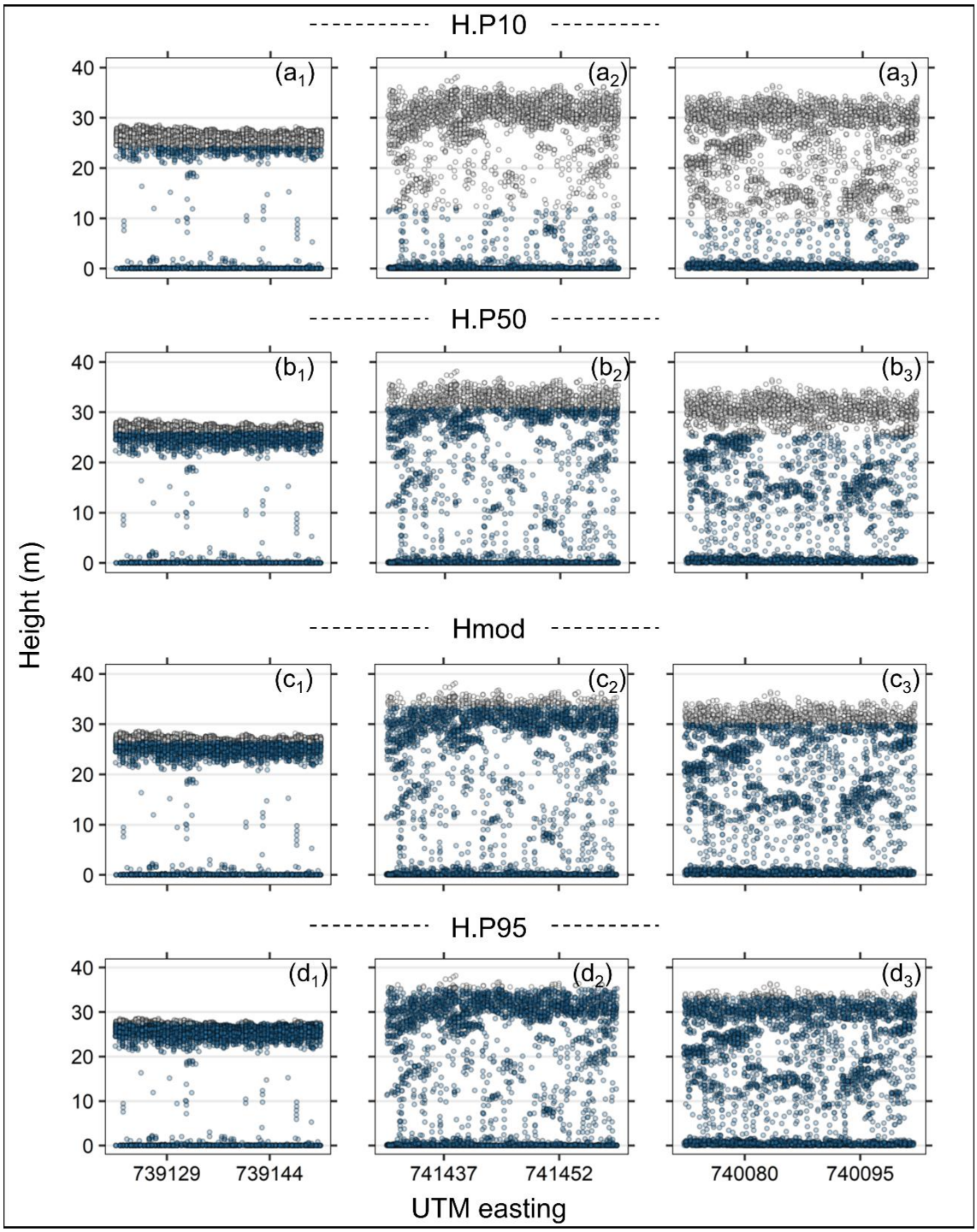

Figure 10. Sample plot transects representing metrics selected as input features for the estimation methods. Colored in blue is the metric height value in the point cloud. Charts (a), (b), (c) and (d) represent the height at the 10th (H.P10), 50th (H.P50), mode (Hmod) and 95th (H.P95) percentiles, respectively. Subscript numbers 1, 2 and 3 represent different sample plots with uniform, intermediate and stratified canopy layers, respectively.

4.2. Individual Tree Detection and TREE Buffer Exclusion

Tree detection is the most sensitive part of the ITD approach [75–77]. The most common methods to detect trees using LiDAR data are raster-based (e.g., LMF and watershed algorithms) [32] and in homogenous forests, such as even-aged Eucalyptus stands, a simple local maxima filter is often selected to detect trees [35,78]. In this study, we used three raster-based methods

and all of them benefited by adding the TBE. The best overall method was the 2-m window of [27] LMF with 1.5-m-buffer exclusion. This LMF, implemented in the software FUSION [46], has an optional enhancement to vary the window size based on an equation relating field measured height and crown diameter which we did not use due to the lack of required information. Thus, defining the best window size is done empirically by reasonably relating it to tree spacing. Wider windows than tree spacing cause omission errors (i.e., existing trees that are not detected), while narrower windows cause commission errors (i.e., detecting more than one maximum per tree). In our study, the commission errors were reduced using the TBE, based on a buffer around each treetop. Errors in tree detection here were likely caused by the dense canopy and round-shaped crown that makes treetops harder to define than in conifers, for example [38]. The watershed method also performed relatively well to detect trees and might be a good option in Eucalyptus stands. One should notice that the LMF and watershed method will not have perfect performance unless all trees are in the upper canopy (i.e., no presence of understory trees). Further, for LMF, the direction and start of the moving window is a factor not completely explored that can affect tree detection. The accurate information on stem density and tree height variation across the stand obtained by detecting the trees is important for precision forest and timber assortment planning [79,80].

Treetop detection and height retrieval are highly influenced by stand and sensor characteristics [3,39,40]. We observed significant differences between LiDAR and field measures when assessing the results in young stands (< 3 years) and gentle slopes (< 25%), underestimating tree heights. In the early stages of Eucalyptus growth, the competition still has not had its effects on the height of trees as much as in a mature forest. In addition, crown shape can affect LiDAR measurements of height [38,81]. We highlight that correctly identifying trees in young stands is important for growth and yield modeling nonetheless, for pre-harvest volume stock estimation which uses mainly mature forest data, this bias should not be an issue.

The differences were also significant in gentle slopes (< 25%). At first, we expected that steep slopes would bias tree detection and height, which we did not observe. In gentle terrains, a dense upper canopy might act as a barrier to pulse penetration and return to the sensor, likely generating errors in the DTM. Height retrieval from ALS and the forest parameter estimates are sensitive to these errors [82].

Our results show that younger stands were more affected for tree height retrieval with LiDAR. Increasing point density should be considered for example by varying acquisition parameters such as pulse frequency and flight height. However, it implies a higher cost for LiDAR data acquisition and processing. An alternative should be to acquire a first (more expensive) dense point cloud, in order to model the terrain more accurately, and successive low-density point clouds for long-term volume estimation [44].

4.3. Area- and Individual Tree-based Approach for Estimating Stem Volume

The ABA with ANN was the best overall method. ANNs are currently used in forestry in many applications, for instance, developing growth and yield models [83], predicting height [84] and stem volume [85]. As a nonparametric method, it benefits from its capacity to account for data variability and non-linear relationships. Alternatively, parametric models are simpler and of more widespread knowledge, besides being easier to share and explain. The issues arise when problem complexity increases, for instance by increasing the number of predictors, and in the presence of considerable variability and outliers.

The error associated with the stem volume estimation methods was around 15%. Although not as low as found in other studies in the same type of forest (e.g., [20]) it is reasonable given the reduced number of plots and the variability of age and regeneration methods in the stands in our study. Clearcut and coppice stands might have structural differences and can in practice be separated during modeling, for example. The recommended number of plots to fit estimation models with ALS data varies depending on the characteristics of the study area and, as in conventional inventory, the desired precision [86]. Maltamo et al. [87] demonstrated that plot

selection to train stem volume models can be improved using ALS data as a priori information. Laranja et al. [88] using a double sampling strategy with LiDAR, decreased the number of necessary plots from 35 to 10 while reducing the inventory error. Our study was carried out with few sample plots (n = 33), which can be sometimes referred to as an issue for training non-parametric methods. Nevertheless, this number is consistent with what is used to record homogeneous Eucalyptus plantation variability (one plot every 10-15 ha; [20,24]). Furthermore, the number of features (explanatory variables) was relatively low (four) and we were concerned about the tuning parameter levels to avoid overfitting.

The ITD approach had lower values for $r_{y\hat{y}}$ (0.80), and higher for RMSE (23.96%) in the plot validation. ABA benefits in this type of validation for retrieving a mean value per area. Whereas, in the ITD, each miss-detected tree can increase the stem volume estimate error per plot. In our study, we used a buffer from a single georeferenced point in the sample plots to validate estimates and a slight change in its position affects the validation. When we validate the volume for the whole stand (i.e., field measurements from sample plots interpolated to the stand), ITD yields better precision and accuracy ($r_{y\hat{y}}$ 0.94 and RMSE = 16.4%), though biased in lower values (Figure 6f). The trend to underestimate lower volumes comes from the errors in tree height measurements in younger stands discussed in the previous section. Breidenbach et al. [89] introduced a method for reducing bias from the ITD by calculating the volume by segments, integrating ABA and ITD. Some other studies have used calibration with linear regression to reduce biased predictions in ITD [76,77]. These are some of the ways to obtain unbiased estimates in the ITD approach.

In the volume maps we observe that the wall-to-wall estimates vary differently for both approaches. The main difference is the distribution of estimated volumes. The range is wider with ITD since ABA concentrate estimates only in a few volume classes. This is an indicator of how ITD can be more thorough in estimates since it is sensitive, for instance, to bigger and smaller trees, which are present in the stand. Extreme values (e.g., 1 or 640 m^3ha^{-1}) appear because one cell can concentrate more trees than others. If they appear too often it could represent estimation errors.

From an operational standpoint, ABA is related to lower costs because it is possible to create models even with low-density point clouds [16]. The drawbacks of this approach are that it is entirely dependent on field plot measurements, size and accuracy [75,76]. Because of that, reductions in fieldwork are limited. On the other hand, for the ITD, high-density point clouds are required to yield better results, implying higher costs [78]. This approach, however, gives a direct measurement of stem density and height distribution to all trees (nearly 520,000 in this study, Figure S2) in addition to *dbh* and volume estimates per tree [75,76]. Furthermore, fieldwork can be reduced to calibrate the models. Tavares Júnior et al. [85], for instance, found that tree volume could be predicted with four sample trees per diameter class using ANN.

The benefits of using LiDAR data to estimate volume is that we have precise spatially explicit estimates for the whole stand. Inventory plots have been used historically and are usually reliable but yield on a single mean value for the whole stand. They do not account for local growth differences affected by, for example, local site variations, diseases or pests spread. An enhanced forest inventory [86] for forest parameters estimation and monitoring can be reached by using the ABA and ITD approaches. Especially with the ITD, the results are promising for even-aged Eucalyptus stands. These forests are uniform, having most trees in the upper canopy favoring simple raster-based methods for tree detection (e.g., LMF and watershed algorithms), with minimized commission errors using algorithms such as the TBE presented here.

5. Conclusions

This study presented a comparison of area- (ABA) and individual tree- based (ITD) approaches for estimating stem volume in Eucalyptus forest plantations. The ABA was more accurate and precise especially with ANN for estimating stem volume as compared to the ITD approach. However, with ITD it was possible to obtain direct measures of tree density and height distribution as well as capture more variability when mapping volume, especially for

homogeneous Eucalyptus stands. The Tree Buffer Exclusion (TBE), presented and tested here as our secondary objective, improved all raster-based tree detection algorithms used in this study. It contributes by reducing treetop detection commission errors, commonly present in studies associated with even-aged stands containing trees exhibiting irregular crown shapes. We expect this method to contribute to ITD implementation in Eucalyptus forest stands in the future.

Author contributions: All the authors have made a substantial contribution towards the successful completion of this manuscript. They all have been involved in designing the study, drafting the manuscript and engaging in critical discussion. R.V.L. and C.H.A. contributed with the conceptualization, data processing, data analysis, visualization, validation, project supervision and write up. R.P.P., C.A.S., C.P.B.S., R.P.M., A.A.L.S., E.N.B., M.M. and H.G.L. contributed to the interpretation, quality control and revisions of the manuscript. All authors have read and approved the final manuscript.

Funding: The first author was granted with a scholarship from the CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico)

Acknowledgments: We acknowledge the Forestry Department at the Federal University of Viçosa for the assistance and for providing the necessary resources during the project. We also thank CENIBRA (Celulose Nipo-Brasileira S.A) for providing the data and supporting the research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hyyppä, J; Kelle, O.; Lehikoinen, M.; Inkinen, M. A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 969–975. doi:10.1109/36.921414
2. White, J.C.; Coops, N.C.; Wulder, M.A.; Vastaranta, M.; Hilker, T.; Tompalski, P. Remote Sensing Technologies for Enhancing Forest Inventories: A Review. *Can. J. Remote Sens.* **2016**, *42*, 619–641. doi:10.1080/07038992.2016.1207484
3. Wu, B.; Yu, B.; Wu, Q.; Huang, Y.; Chen, Z.; Wu, J. Individual tree crown delineation using localized contour tree method and airborne LiDAR data in coniferous forests. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *52*, 82–94. doi:10.1016/j.jag.2016.06.003
4. Wulder, M.A.; Coops, N.C.; Hudak, A.T.; Morsdorf, F.; Nelson, R.; Newnham, G.; Vastaranta, M. Status and prospects for LiDAR remote sensing of forested ecosystems. *Can. J. Remote Sens.* **2013**, *39*, 37–41. doi:10.5589/m13-051
5. Beland, M.; Parker, G.; Sparrow, B.; Harding, D.; Chasmer, L.; Phinn, S.; Antonarakis, A.; Strahler, A. On promoting the use of lidar systems in forest ecosystem research. *Forest Ecol. Manag.* **2019**, *450*. doi:10.1016/j.foreco.2019.117484
6. Lefsky, M.A.; Cohen, W.B.; Parker, G.G.; Harding, D.J. Lidar Remote Sensing for Ecosystem Studies: Lidar, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular interest to forest, landscape, and global ecologists. *BioScience* **2002**, *52*, 19–30. doi:10.1641/0006-3568(2002)052[0019:LRSFES]2.0.CO;2
7. Nelson, R.; Krabill, W.; Tonelli, J. Estimating forest biomass and volume using airborne laser data. *Remote Sens. Environ.* **1988**, *24*, 247–267.
8. Nilsson, M. Estimation of tree heights and stand volume using an airborne lidar system. *Remote Sens. Environ.* **1996**, *56*, 1–7. doi:10.1016/0034-4257(95)00224-3
9. Means, J.E.; Acker, S.A.; Fitt, B.J.; Renslow, M.; Emerson, L.; Hendrix, C.J. Predicting Forest Stand Characteristics with Airborne Scanning Lidar. *Photogramm. Eng. Remote Sens.* **2000**, *66*, 1367–1371.
10. Næsset, E.; Økland, T. Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve. *Remote Sens. Environ.* **2002**, *79*, 105–115. doi:10.1016/S0034-4257(01)00243-7
11. González-Ferreiro, E.; Diéguez-Aranda, U.; Miranda, D. Estimation of stand variables in Pinus radiata D. Don plantations using different LiDAR pulse densities. *Forestry* **2012**, *85*, 281–292. doi:10.1093/forestry/cps002
12. Nelson, R. How did we get here? An early history of forestry lidar 1. *Can. J. Remote Sens.* **2013**, *39*, S6–S17. doi:10.5589/m13-011

13. 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. doi:10.1016/S0034-4257(01)00290-5
14. Hyyppä, J., Inkinen, M. Detecting and estimating attributes for single trees using laser scanner. *Photogramm. J. Finl.* **1999**, *16*, 27–42.
15. IBÁ—Indústria Brasileira de Árvores, 2019. Relatório 2019. São Paulo. Available online: https://iba.org/publicacoes (accessed on 30 March 2019).
16. Tesfamichael, S.G.; Ahmed, F.B.; van Aardt, J.A.N. Investigating the impact of discrete-return lidar point density on estimations of mean and dominant plot-level tree height in Eucalyptus grandis plantations. *Int. J. Remote Sens.* **2010**, *31*, 2925–2940. doi:doi.org/10.1080/01431160903144086
17. Zonete, M.F.; Rodriguez, L.C.E.; Packalén, P. Estimação de parâmetros biométricos de plantios clonais de eucalipto no sul da Bahia: Uma aplicação da tecnologia laser aerotransportada. *Sci. For.* **2010**, 225–235.
18. De Oliveira, L.T.; Ferreira, M.Z.; de Carvalho, L.M.T.; Ferraz Filho, A.C.; Oliveira, T.C.D.A.; Silveira, E.M.D.O.; Junior, F.W.A. Determinação do volume de madeira em povoamento de eucalipto por escâner a laser aerotransportado. *Pesqui. Agropecuária Bras.* **2014**, *49*, 692–700. doi:10.1590/S0100-204X2014000900005
19. Silva, C.A.; Klauberg, C.; Carvalho, S.D.P.C.; Hudak, A.T.; Rodriguez, L.C.E. Mapping aboveground carbon stocks using LiDAR data in Eucalyptus spp. plantations in the state of São Paulo, Brazil. *Sci. For.* **2014**, *42*, 591–604.
20. Silva, C.A.; Klauberg, C.; Hudak, A.T.; Vierling, L.A.; Liesenberg, V.; Carvalho, S.P.C.E.; Rodriguez, L.C.E. A principal component approach for predicting the stem volume in Eucalyptus plantations in Brazil using airborne LiDAR data. *Forestry* **2016**, *89*, 422–433. doi:10.1093/forestry/cpw016
21. Shinzato, E.; Shimabukuro, Y.; Coops, N.; Tompalski, P.; Gasparoto, E. Integrating area-based and individual tree detection approaches for estimating tree volume in plantation inventory using aerial image and airborne laser scanning data. *IForest Biogeosciences For.* **2016**, *10*, 296–302. doi:10.3832/ifor1880-009
22. Silva, C.; Klauberg, C.; Hudak, A.; Vierling, L.; Jaafar, W.; Mohan, M.; Garcia, M.; Ferraz, A.; Cardil, A.; Saatchi, S. Predicting Stem Total and Assortment Volumes in an Industrial Pinus taeda L. Forest Plantation Using Airborne Laser Scanning Data and Random Forest. *Forests* **2017**, *8*, 254. doi:10.3390/f8070254
23. 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
24. Görgens, E.B.; Montaghi, A.; Rodriguez, L.C.E. A performance comparison of machine learning methods to estimate the fast-growing forest plantation yield based on laser scanning metrics. *Comput. Electron. Agric.* **2015**, *116*, 221–227. doi:10.1016/j.compag.2015.07.004
25. Li, Y.; Andersen, H.-E.; McGaughey, R. A Comparison of Statistical Methods for Estimating Data. *J. Appl. For.* **2008**, *23*, 223–231. doi:doi.org/10.1093/wjaf/23.4.223
26. Moser, P.; Vibrans, A.C.; McRoberts, R.E.; Næsset, E.; Gobakken, T.; Chirici, G.; Mura, M.; Marchetti, M. Methods for variable selection in LiDAR-assisted forest inventories. *Forestry* **2017**, *90*, 112–124. doi:10.1093/forestry/cpw041
27. Popescu, S.C.; Wynne, R.H.; Nelson, R.F. Estimating plot-level tree heights with lidar: Local filtering with a canopy-height based variable window size. *Comput. Electron. Agric.* **2002**, *37*, 71–95. doi:10.1016/S0168-1699(02)00121-7
28. Koch, B.; Heyder, U.; Weinacker, H. Detection of Individual Tree Crowns in Airborne Lidar Data. *Photogramm. Eng. Remote Sens.* **2006**, *72*, 357–363. doi:10.14358/PERS.72.4.357
29. Silva, C.A.; Crookston, N.L.; Hudak, A.T.; Vierling, L.A.; Klauberg, C. rLiDAR: LiDAR Data Processing and Visualization. R package version 0.1. 2015. Available online: https://CRAN.R-project.org/package=rLiDAR (accessed on May 8, 2020).
30. Ferraz, A.; Saatchi, S.; Mallet, C.; Meyer, V. Lidar detection of individual tree size in tropical forests. *Remote Sens. Environ.* **2016**, *183*, 318–333. doi:/10.1016/j.rse.2016.05.028
31. Silva, C.A.; Hudak, A.T.; Vierling, L.A.; Loudermilk, E.L.; O'Brien, J.J.; Hiers, J.K.; Jack, S.B.; Gonzalez-Benecke, C.; Lee, H.; Falkowski, M.J.; et al. Imputation of Individual Longleaf Pine (Pinus palustris

Mill.) Tree Attributes from Field and LiDAR Data. *Can. J. Remote Sens.* **2016**, *42*, 554–573. doi:10.1080/07038992.2016.1196582

32. Zhen, Z.; Quackenbush, L.; Zhang, L. Trends in Automatic Individual Tree Crown Detection and Delineation—Evolution of LiDAR Data. *Remote Sens.* **2016**, *8*, 333. doi:10.3390/rs8040333
33. Ayrey, E.; Fraver, S.; Kershaw, J.A.; Kenefic, L.S.; Hayes, D.; Weiskittel, A.R.; Roth, B.E. Layer Stacking: A Novel Algorithm for Individual Forest Tree Segmentation from LiDAR Point Clouds. *Can. J. Remote Sens.* **2017**, *43*, 16–27. doi:10.1080/07038992.2017.1252907
34. Mohan, M.; de Mendonça, B.A.F.; Silva, C.A.; Klauberg, C.; de Saboya Ribeiro, A.S.; de Araújo, E.J.G.; Cardil, A. Optimizing individual tree detection accuracy and measuring forest uniformity in coconut (Cocos nucifera L.) plantations using airborne laser scanning. *Ecol. Model.* **2019**, *409*, 108736.
35. Cosenza, D.N.; Soares, V.P.; Leite, H.G.; Gleriani, J.M.; do Amaral, C.H.; Júnior, J.G.; da Silva, A.A.L.; Soares, P.; Tomé, M. Airborne laser scanning applied to eucalyptus stand inventory at individual tree level. *Pesqui. Agropecu. Bras.* **2018**, *53*, 1373–1382. doi:10.1590/S0100-204X2018001200010
36. Zandoná, D.F. Varredura a Laser aerotransportado para estimativa de variáveis dendrométricas. *Sci. For.* **2008**, *36*, 295–306.
37. Vauhkonen, J.; Ene, L.; Gupta, S.; Heinzel, J.; Holmgren, J.; Pitkanen, J.; Solberg, S.; Wang, Y.; Weinacker, H.; Hauglin, K.M.; et al. Comparative testing of single-tree detection algorithms under different types of forest. *Forestry* **2012**, *85*, 27–40. doi:10.1093/forestry/cpr051
38. Guerra-Hernández, J.; Cosenza, D.N.; Rodriguez, L.C.E.; Silva, M.; Tomé; M; Díaz-Varela, R.A.; González-Ferreiro, E. Comparison of ALS- and UAV(SfM)-derived high-density point clouds for individual tree detection in Eucalyptus plantations. *Int. J. Remote Sens.* **2018**, *39*, 5211–5235. doi:10.1080/01431161.2018.1486519
39. Khosravipour, A.; Skidmore, A.K.; Wang, T.; Isenburg, M.; Khoshelham, K. Effect of slope on treetop detection using a LiDAR Canopy Height Model. *ISPRS J. Photogramm. Remote Sens.* **2015**, *104*, 44–52. doi:10.1016/j.isprsjprs.2015.02.013
40. Keränen, J.; Maltamo, M.; Packalen, P. Effect of flying altitude, scanning angle and scanning mode on the accuracy of ALS based forest inventory. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *52*, 349–360. doi:10.1016/j.jag.2016.07.005
41. Liu, J.; Skidmore, A.K.; Jones, S.; Wang, T.; Heurich, M.; Zhu, X.; Shi, Y. Large off-nadir scan angle of airborne LiDAR can severely affect the estimates of forest structure metrics. *ISPRS J. Photogramm. Remote Sens.* **2018**, *136*, 13–25. doi:10.1016/j.isprsjprs.2017.12.004
42. Roussel, J.-R.; Caspersen, J.; Béland, M.; Thomas, S.; Achim, A. Removing bias from LiDAR-based estimates of canopy height: Accounting for the effects of pulse density and footprint size. *Remote Sens. Environ.* **2017**, *198*, 1–16. doi:10.1016/j.rse.2017.05.032
43. Silva, C.A.; Hudak, A.; Klauberg, C.; Vierling, L.; Gonzalez-Benecke, C.; Carvalho, S.P.C.; Rodriguez, L.C.E.; Cardil, A. Combined effect of pulse density and grid cell size on predicting and mapping aboveground carbon in fast-growing Eucalyptus forest plantation using airborne LiDAR data. *Carbon Balance Manag.* **2017**, *12*, 13.
44. Silva, C.A.; Hudak, A.; Vierling, L.; Klauberg, C.; Garcia, M.; Ferraz, A.; Keller, M.; Eitel, J.; Saatchi, S. Impacts of airborne lidar pulse density on estimating biomass stocks and changes in a selectively logged tropical forest. *Remote Sens.* **2017**, *9*, 1068.
45. Hall, F.; Schumacher, F. Logarithmic expression of timber-tree. *J. Agric. Res.* **1933**, *47*, 719–734.
46. McGaughey, R.J. *FUSION/LDV: Software for LiDAR Data Analysis and Visualization*; US Department of Agriculture, Forest Service, Pacific Northwest Research Station: Seattle, WA, USA, 2018.
47. Kraus, K.; Pfeifer, N. Determination of terrain models in wooded areas with airborne laser scanner data. *ISPRS J. Photogramm. Remote Sens.* **1998**, *53*, 193–203. doi:10.1016/S0924-2716(98)00009-4
48. Chen, Q.; Baldocchi, D.; Gong, P.; Kelly, M. Isolating Individual Trees in a Savanna Woodland Using Small Footprint Lidar Data. *Photogramm. Eng. Remote Sens.* **2006**, *72*, 923–932. doi:10.14358/PERS.72.8.923
49. Gompertz, B. XXIV On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philos. Trans. R. Soc. Lond.* **1825**, *115*, 513–583. doi:10.1098/rstl.1825.0026
50. Nokoe, S. Demonstrating the flexibility of the Gompertz function as a yield model using mature species data. *Commonw. For. Rev.* **1978**, 35–42.
51. Lippmann, R.P. An introduction to computing with neural nets. *IEEE Assp Mag.* **1987**, *4*, 4–22.

52. Jain, A.K.; Mao, J.; Mohiuddin, K.M. Artificial neural networks: A tutorial. *Computer* **1996**, *29*, 31-44. doi:10.1109/2.485891
53. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.
54. Boser, B.; Guyon, I.; Vapnik, V. A training algorithm for optimal margin classifiers. In Proceedings of the Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, PA, USA, 27–28 July 1992; pp. 144–152.
55. Vapnik, V.N. *Statistical Learning Theory*; Wiley Interscience. New York, NY, USA. 1998.
56. Levick, S.R.; Hessenmöller, D.; Schulze, E.D. Scaling wood volume estimates from inventory plots to landscapes with airborne LiDAR in temperate deciduous forest. *Carbon Balance Manag.* **2016**, *11*. doi:10.1186/s13021-016-0048-7
57. Guyon, I.; Weston, J.; Stephen, B.; Vapnik, V. Gene Selection for Cancer Classification using Support Vector Machines. *Mach. Learn.* **2002**, 389–422. doi:10.1023/A:101248730
58. Kuhn, M.; Johnson, K. *Applied Predictive Modeling*; Springer: New York, NY, USA, 2013. doi:10.1007/978-1-4614-6849-3
59. Kuhn, M. Caret: Classification and Regression Training. R pacakge version 6.0-84. 2019. Available online: http://topepo.github.io/caret/index.html (accessed on May 8, 2020).
60. Fernandes Filho, E.I. labgeo: Collection of functions to fit models with emphasis in land use and soil mapping. R package version 0.3.9.3. 2019. Available online: https://github.com/elpidiofilho/labgeo (accessed on May 8, 2020).
61. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019.
62. Bergmeir, C.; Benitez, J.M. Neural Networks in R Using the Stuttgart Neural Network Simulator: RSNNS. *J. Stat. Softw.* **2012**, *46*, 1–26.
63. Liaw, A.; Wiener, M. Classification and Regression by randomForest. *R News* **2002**, *2*, 18–22.
64. Karatzoglou, A.; Smola, A.; Hornik, K.; Zeileis, A. kernlab—An S4 Package for Kernel Methods in R. *J. Stat. Softw.* **2004**, *11*, 1–20
65. Wang, L.; Gong, P.; Biging, G.S. Individual Tree-Crown Delineation and Treetop Detection in High-Spatial-Resolution Aerial Imagery. *Photogramm. Eng. Remote Sens.* **2004**, *70*, 351–357. doi:10.14358/PERS.70.3.351
66. ArcGIS, E.S.R.I. *10.1*; ESRI: Redlands, CA, USA, 2012
67. EMBRAPA, Empresa Brasileira de Pesquisa Agropecuária. In Proceedings of the Súmula da X Reunião Técnica de Levantamento de Solos, Rio de Janeiro, RJ, Brazil, 1979. Conference date: 26–30 March, 1979.
68. Graybill, F.A. *Theory and Application of the Linear Model*; Duxbury press. North Scituate, MA, USA, 1976.
69. Piñeiro, G.; Perelman, S.; Guerschman, J.P.; Paruelo, J.M. How to evaluate models: Observed vs. predicted or predicted vs. observed? *Ecol. Model.* **2008**, *216*, 316–322. doi:10.1016/j.ecolmodel.2008.05.006
70. Wei, T.; Simko, V. R package "corrplot": Visualization of a Correlation Matrix (Version 0.84). 2017. Available online: https://github.com/taiyun/corrplot (accessed on May 8, 2020).
71. 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
72. Dormann, C.F.; Elith, J.; Bacher, S.; Buchmann, C.; Carl, G.; Carré, G.; Marquéz, J.R.G.; Gruber, B.; Lafourcade, B.; Leitão, P.J.; et al. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **2013**, *36*, 27–46. doi:10.1111/j.1600-0587.2012.07348.x
73. Li, D.; Gu, X.; Pang, Y.; Chen, B.; Liu, L. Estimation of forest aboveground biomass and leaf area index based on digital aerial photograph data in northeast China. *Forests* **2018**, *9*, 1–23. doi:10.3390/f9050275
74. Kuhn, M.; Johnson, K. Feature Engineering and Selection: A Practical Approach for Predictive Models. Taylor & Francis Group. 2019. Available online: http://www.feat.engineering/ (accessed on May 8, 2020).
75. Vastaranta, M.; Kankare, V.; Holopainen, M.; Yu, X.; Hyyppä, J.; Hyyppä, H. Combination of individual tree detection and area-based approach in imputation of forest variables using airborne laser data. *Isprs J. Photogramm. Remote Sens.* **2012**, *7*, 73–79. doi:10.1016/j.isprsjprs.2011.10.006
76. Yu, X.; Hyyppä, J.; Holopainen, M.; Vastaranta, M. Comparison of area-based and individual tree-based methods for predicting plot-level forest attributes. *Remote Sens.* **2010**, *2*, 1481–1495. doi:10.3390/rs2061481

77. Vastaranta, M.; Holopainen, M.; Yu, X.; Haapanen, R.; Melkas, T.; Hyyppä, J.; Hyyppä, H. Individual tree detection and area-based approach in retrieval of forest inventory characteristics from low-pulse airborne laser scanning data. *Photogramm. J. Finl.* **2011**, *22*, 1–13.
78. De Oliveira, L.T.; de Carvalho, L.M.T.; Ferreira, M.Z.; Oliveira, T.C.D.A.; Acerbi Junior, F.W. Application of LIDAR to forest inventory for tree count in stands of Eucalyptus sp. *CERNE* **2012**, *18*, 175–184. doi:10.1590/S0104-77602012000200001
79. Kaartinen, H.; Hyyppä, J.; Yu, X.; Vastaranta, M.; Hyyppä, H.; Kukko, A.; Holopainen, M.; Heipke, C.; Hirschmugl, M.; Morsdorf, F.; et al. An international comparison of individual tree detection and extraction using airborne laser scanning. *Remote Sens.* **2012**, *4*, 950–974. doi:doi.org/10.3390/rs4040950
80. Vastaranta, M.; Holopainen, M.; Yu, X.; Hyyppä, J.; Mäkinen, A.; Rasinmäki, J.; Melkas, T.; Kaartinen, H.; Hyyppä, H. Effects of individual tree detection error sources on forest management planning calculations. *Remote Sens.* **2011**, *3*, 1614–1626. doi:10.3390/rs3081614
81. Nelson, R. Modeling forest canopy heights: The effects of canopy shape. *Remote Sens. Environ.* **1997**, *60*, 327–334. doi:10.1016/S0034-4257(96)00214-3
82. Leitold, V.; Keller, M.; Douglas, C.M.; Cook, B.D.; Shimabukuro, Y.E. Airborne lidar-based estimates of tropical forest structure in complex terrain: Opportunities and trade-offs for REDD+. *Carbon Balance Manag.* **2015**, *10*, 3. doi:10.1186/s13021-015-0013-x
83. Castro, R.V.O.; Boechat, C.P.S.; Leite, H.G.; de Souza, A.L.; Nogueira, G.S.; Martins, F.B. Individual Growth Model for Eucalyptus Stands in Brazil Using Artificial Neural Network. *ISRN For.* **2013**, *2013*, 1–12. doi:10.1155/2013/196832
84. Binoti, M.L.M.D.S.; Binoti, D.H.B.; Leite, H.G. Aplicação de redes neurais artificiais para estimação da altura de povoamentos equiâneos de eucalipto. *Rev. Árvore* **2013**, *37*, 639–645. doi:10.1590/S0100-67622013000400007
85. Tavares Júnior, I.D.S.; da Rocha, J.E.C.; Ebling, ângelo A.; de Souza Chaves, A.; Zanuncio, J.C.; Farias, A.A.; Leite, H.G. Artificial neural networks and linear regression reduce sample intensity to predict the commercial volume of eucalyptus clones. *Forests* **2019**, *10*, 268. doi:10.3390/f10030268
86. White, J.C.; Wulder, M.; Varhola, A.; Vastaranta, M.; Coops, N.C.; Cook, B.D.; Pitt, D.; Woods, M. A model development and application guide for generating an enhanced forest inventory using airborne laser scanning data and an area-based approach. *Nat. Resour. Can.* **2017**. doi:10.5558/tfc2013-132
87. Maltamo, M.; Bollandsås, O.M.; Næsset, E.; Gobakken, T.; Packalén, P. Different plot selection strategies for field training data in ALS-assisted forest inventory. *Forestry* **2011**, *84*, 23–31. doi:10.1093/forestry/cpq039
88. Laranja, D.C.F.; Gorgens, E.B.; Soares, C.P.B.; Silva, A.G.P.; Rodriguez, L.C.E. Redução do erro amostral na estimativa do volume de povoamentos de Eucalyptus ssp. por meio de escaneamento laser aerotransportado. *Sci. For.* **2015**, *43*, 845–852. doi:10.18671/scifor.v43n108.9
89. Breidenbach, J.; Næsset, E.; Lien, V.; Gobakken, T.; Solberg, S. Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. *Remote Sens. Environ.* **2010**, *114*, 911–924. doi:10.1016/j.rse.2009.12.004

Supplementary material

Table S1. Stand characteristics and mean values of field measurements in the study area. Where: *Dbh* = mean diameter at 1.3 m above ground; *H* = mean height; *V* = tree outside-bark mean volume; N = mean number of trees per hectare

Stand	Area (ha)	Regeneration	Spacing	Age class	*Dbh* (cm)	*H* (m)	*V* (m^3ha^{-1})	N/ha
S1	10.41	Clearcut	3x3	Mature	19.5	32.0	441.33	1028
S2	25.00	Clearcut	3x3	Mature	34.6	25.4	654.62	1179
S3	25.65	Clearcut	3x3	Mature	33.7	25.0	601.02	1178
S4	37.91	Clearcut	3x3	Mature	35.7	27.4	662.34	1075
S5	37.56	Clearcut	3x3	Mature	34.0	27.9	547.78	954
S6	18.31	Clearcut	3x3	Mature	17.1	26.3	318.41	1167
S7	22.01	Clearcut	3x3	Mature	18.9	31.2	424.15	1074
S8	10.06	Clearcut	3x3	Mature	17.0	26.6	305.56	1108
S9	37.80	Clearcut	3x3	Intermediate	28.7	19.9	311.43	1060
S10	16.21	Clearcut	3x3	Intermediate	16.0	22.4	194.45	953
S11	8.22	Clearcut	3x3	Intermediate	15.7	23.0	213.75	1056
S12	17.12	Clearcut	3x3	Intermediate	16.8	22.0	247.19	1101
S13	27.39	Clearcut	3x3	Intermediate	15.4	22.7	244.43	1259
S14	24.03	Coppice	4x3	Intermediate	32.4	24.5	412.87	956
S15	11.20	Coppice	4x3	Intermediate	16.0	23.5	192.90	953
S16	24.31	Coppice	3x3	Intermediate	47.5	24.9	714.02	1101
S17	22.56	Coppice	4x3	Intermediate	31.6	24.5	391.58	939
S18	57.40	Clearcut	3x3	Young	13.4	21.5	152.85	1056
S19	38.33	Coppice	3x3	Young	13.6	19.7	113.36	900

Table S2. Plot characteristics and mean values of field measurements in the study area. Where: Dbh = mean diameter at 1.3 m above ground; H = mean height; V = tree outside-

bark mean volume; N = mean number of trees per hectare. Partition stands for the data sets used for training (60%) and test (40%) of the volume estimation models.

Stand	Plot	Regeneration	Spacing	Age	Dbh (cm)	H (m)	V (m^3/ha)	N	Partition
S1	P1	Clearcut	3x3	Mature	19.53 ± 2.98	32.04 ± 2.2	441.33 ± 5.26	1028	Training
S2	P2	Clearcut	3x3	Mature	16.69 ± 3.68	23.7 ± 3	304.67 ± 3.84	1268	Training
S2	P3	Clearcut	3x3	Mature	17.97 ± 4.58	27.12 ± 4	349.95 ± 5.38	1090	Training
S3	P4	Clearcut	3x3	Mature	17.33 ± 1.99	27.73 ± 1.3	363.14 ± 2.66	1241	Test
S3	P5	Clearcut	3x3	Mature	16.44 ± 2.12	22.29 ± 1.2	237.88 ± 2.12	1114	Training
S4	P6	Clearcut	3x3	Mature	17.97 ± 2.34	27.91 ± 1.7	342.64 ± 2.75	1079	Training
S4	P7	Clearcut	3x3	Mature	17.75 ± 1.81	27.04 ± 1.2	319.7 ± 1.97	1070	Test
S5	P8	Clearcut	3x3	Mature	16.73 ± 1.11	25.98 ± 0.6	238.54 ± 1.48	937	Training
S5	P9	Clearcut	3x3	Mature	17.34 ± 2.88	29.85 ± 2.2	309.24 ± 4.3	971	Test
S6	P10	Clearcut	3x3	Mature	17.12 ± 2.29	26.33 ± 1.6	318.41 ± 2.71	1167	Test
S7	P11	Clearcut	3x3	Mature	18.93 ± 2.9	31.26 ± 2.2	424.15 ± 5.43	1074	Training
S8	P12	Clearcut	3x3	Mature	17.02 ± 3.1	26.63 ± 2.8	305.56 ± 3.29	1108	Training
S9	P13	Clearcut	3x3	Intermediate	14.05 ± 1.34	19.44 ± 0.8	146.41 ± 1.01	1072	Test
S9	P14	Clearcut	3x3	Intermediate	14.69 ± 2.27	20.35 ± 1.4	165.02 ± 1.96	1047	Test
S10	P15	Clearcut	3x3	Intermediate	16.03 ± 1.63	22.48 ± 0.8	194.45 ± 1.74	953	Test

S11	P16	Clearcut	3x3	Intermediate	15.79 ± 1.24	23.03 ± 0.7	213.75 ± 1.56	1056	Training
S12	P17	Clearcut	3x3	Intermediate	16.84 ± 2.89	22.06 ± 2.3	247.19 ± 2.82	1101	Training
S13	P18	Clearcut	3x3	Intermediate	15.49 ± 2.31	22.7 ± 1.4	244.43 ± 2.72	1259	Training
S14	P19	Coppice	4x3	Intermediate	16.11 ± 2.99	23.5 ± 1.8	201.43 ± 2.5	1005	Test
S14	P20	Coppice	4x3	Intermediate	16.35 ± 3.33	25.63 ± 2.6	211.44 ± 2.7	907	Training
S15	P21	Coppice	4x3	Intermediate	16.05 ± 3.54	23.58 ± 2.5	192.9 ± 3.09	953	Test
S16	P22	Coppice	3x3	Intermediate	15.34 ± 3.74	25.7 ± 2.8	242.04 ± 3.88	1149	Training
S16	P23	Coppice	3x3	Intermediate	14.89 ± 3.01	22.74 ± 2.1	181.33 ± 2.14	1084	Training
S16	P24	Coppice	3x3	Intermediate	17.32 ± 3.92	26.52 ± 2.9	290.65 ± 3.66	1068	Training
S17	P25	Coppice	4x3	Intermediate	14.97 ± 3.28	22.82 ± 2.1	166.16 ± 2.69	974	Test
S17	P26	Coppice	4x3	Intermediate	16.72 ± 3.31	26.34 ± 2.3	225.42 ± 3.2	903	Training
S18	P27	Clearcut	3x3	Young	14.02 ± 2.37	21.24 ± 2.1	140.74 ± 1.79	909	Training
S18	P28	Clearcut	3x3	Young	13.67 ± 1.67	20.3 ± 1	135.99 ± 1.22	986	Test
S18	P29	Clearcut	3x3	Young	12.55 ± 2.42	23.03 ± 2.2	181.83 ± 1.88	1271	Training
S19	P30	Coppice	3x3	Young	13.92 ± 2.02	19.31 ± 1	114.11 ± 1.4	891	Test
S19	P31	Coppice	3x3	Young	12.36 ± 2	19.96 ± 1.6	119.18 ± 1.61	1094	Training
S19	P32	Coppice	3x3	Young	13.9 ± 3.32	20.38 ± 2.2	112.81 ± 2.69	803	Test

S19	P33	Coppice	3x3	Young	14.29 ± 1.51	19.15 ± 0.8	107.32 ± 0.85	813	Training

Table S3. Description of soil types by plot and stand. Soil types classified in the Brazilian Soil Classification System [1]

Stands	**Plots**	**Soil description**
S19	P30, P31, P32, P33	Yellow Latosol: Typical dystrophic, very clayey texture, A prominent, alicic, kaolinitic, mesofferic.
S12, S15, S16, S2	P17, P2, P21, P22, P23, P24, P3	Yellow Latosol: Typical Acrylic, clayey texture, moderate A, alicic, kaolinitic or kaolinitic-oxidic, mesofferic.
S18	P27, P28, P29	Cambisols: Dystrophic latosolic, very clayey texture, A prominent or moderate, alic, kaolinitic, hypoferric or mesoferric.
S1, S17, S7	P1, P11, P25, P26	Cambisols: Dystrophic latosolic, medium / clayey texture, A prominent or moderate, kaolinitic, hypoferric or mesofferric.
S10, S13, S14, S3, S4, S9	P13, P14, P15, P18, P19, P20, P4, P5, P6, P7	Cambisols: Dystrophic latosolic, medium / clayey texture, A prominent or moderate, kaolinitic, hypoferric or mesofferric. + Cambisols: Dystrophic lithic, clayey or very clayey texture, A moderate, alicic, kaolinitic or kaolinitic-oxidic, mesofferic.
S6	P10	Cambisols: Dystrophic latosolic, very clayey gravel texture, moderate A, alic, kaolinitic, hypoferric or mesofferric.
S11, S5, S8	P12, P16, P8, P9	Red Latosol: Typical Accric, clayey texture, moderate A, alic, kaolinitic or kaolinitic-oxydic, mesofferic. + Red Yellow Latosol: Typical Accric, clayey texture, moderate A, alicic, kaolinitic or kaolinitic-oxidic, mesoferric.

Table S4. Parameters used to estimate volume in the ITD approach using Schumacher and Hall (1933) model. Stand names following Table 1. RMSE = Root Mean Square Error; RMSE(%) = relative RMSE.

Stand	b0	b1	b2	$r_{y\hat{y}}$	RMSE (m^3/tree)	RMSE(%)
S(1-13)	-10.082876	1.961099	0.978704	0.9955	0.0182	7.9
S(14-17)	-10.853507	1.811935	1.325494	0.9971	0.0230	11.8
S18	-10.300613	1.689753	1.292432	0.9920	0.0301	12.4
S19	-10.327244	1.820607	1.175653	0.9924	0.0320	13.2

Table S5. Most frequently held LiDAR metrics as input variables in an estimation method by the Recursive Feature Elimination algorithm (RFE) based on its importance. Metrics within the dashed box were used as input variables in the estimation methods

Metric	**Frequency (%)**		
	ANN	**RF**	**SVM**
H.P95	99	99	100
H.P50	91	91	70
Hmod	73	73	77
H.P10	46	46	68
H.P05	12	26	48
H.Admod	7	8	43
Ccover	6	26	18
Hkurt	5	3	27
H.Admed	1	7	16
Hcv	0	4	10
CRR	0	4	7
H.P01	0	3	4
Hmin	0	2	0

Table S6. Statistics for training, tuning and validating non-parametric methods. Method's performance was assessed with leave-one out cross validation resample method. $r_{y\hat{y}}$= Pearson's correlation coefficient between observed and estimated volume; RMSE = Root Mean Square Error; RMSE(%) = relative RMSE.

Method	$r_{y\hat{y}}$	RMSE	RMSE(%)	Bias	Bias(%)
RF	0.97	25.75	11.39	-2.576	-1.02
SVM	0.89	32.79	16.24	-2.098	-0.83
ANN	0.95	28.17	12.08	0.413	0.16

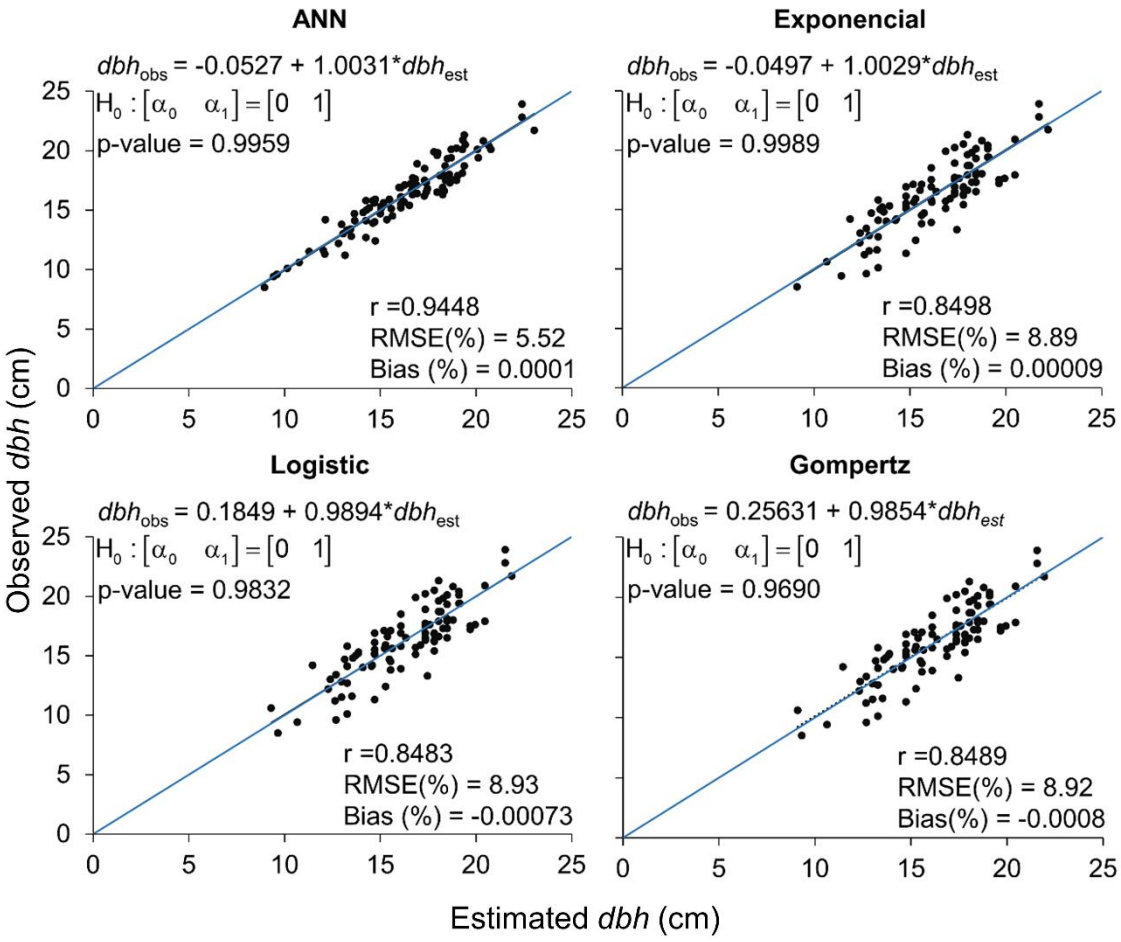

Figure S1. Estimates of diameter at breast height (*dbh*). P-values within plot are for the Graybill (1976) F-test. Parametric methods:

$$dbh = b_0 e^{-e^{b_1 - b_a H}} + \varepsilon \quad \text{(Gompertz)}$$

$$dbh = \frac{b_0}{1 + b_1 e^{-b_2 H}} + \varepsilon \quad \text{(Logistic)}$$

$$dbh = b_0 e^{\frac{b_1}{H}} + \varepsilon \quad \text{(Exponential)}$$

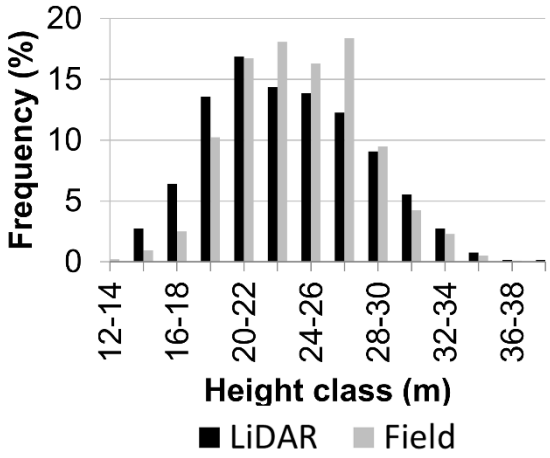

Figure S2. Height distribution of sample plot field measured trees in 33 plots and the nearly 520 thousand trees detected with LiDAR.

1. Santos, H.G., Jacomine, P.K.T., Anjos, L.H.C., Oliveira, V.A., Lumbreras, J.F., Coelho, M.R., Almeida, J.A., Araujo Filho, J.C., Oliveira, J.B., Cunha, T.J.., 2013. Brazilian Soil Classification System, 3rd ed. Embrapa, Brasília, 353 p.