Version of Record: https://www.sciencedirect.com/science/article/pii/S0168169920308838
Manuscript_1f6f4831301abd596eb2f11ea92aec36

Forest inventory with high-density UAV-Lidar: machine learning approaches for predicting individual tree attributes

Ana Paula Dalla Corte[a,*], Deivison Venicio Souza[b], Franciel Eduardo Rex[a], Carlos Roberto Sanquetta[a], Midhun Mohan[c], Carlos Alberto Silva[d,e], Angelica Maria Almeyda Zambrano[f], Gabriel Prata[f], Danilo Roberti Alves de Almeida[f,g], Jonathan William Trautenmüller[a], Carine Klauberg[h], Anibal de Moraes[a], Mateus N. Sanquetta[a], Ben Wilkinson[c] and Eben North Broadbent[e]

[a] *Federal University of Paraná - UFPR, Department of Forest Science, Prefeito Lothário Meissner Avenue - Jardim Botânico, Curitiba, PR, Brazil, 80210-170. anapaulacorte@gmail.com; francielrexx@gmail.com; carlossanquetta@gmail.com; jwtraute@gmail.com; anibalm@ufpr.br; mateus.sanquetta@gmail.com*

[b] *Faculty of Forestry Engineering, Federal University of Pará - UFPA, Coronel José Porfírio street, 2515, Altamira, Pará, Brazil, 68370-000. deivisonvs@ufpa.br*

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

[d] *School of Forest Resources and Conservation, University of Florida, Gainesville, Florida, United States of America. carlos_engflorestal@outlook.com*

[e] *Department of Geographical Sciences, University of Maryland, College Park, Maryland, MD 20740, USA.*

[f] *Spatial Ecology and Conservation Lab, School of Forest Resources and Conservation, University of Florida, Gainesville, FL 32611 USA. aalmeyda@ufl.edu; eben@ufl.edu*

[g] *Department of Forest Sciences, "Luiz de Queiroz" College of Agriculture, University of São Paulo (USP/ESALQ), Piracicaba, SP, 13418-900, Brazil. daniloraa@usp.br*

[h] *Federal University of São João Del Rei-UFSJ, Sete Lagoas, MG, Brazil, 35701-970. carine_klauberg@hotmail.com*

ARTICLE INFO

Keywords:
Forest variables
Support Vector Regression
Random Forest
Artificial Neural Networks
Extreme Gradient Boosting
Forest attributes

ABSTRACT

The high dimensionality of data generated by Unmanned Aerial Vehicle(UAV)-Lidar makes it difficult to use classical statistical techniques to design accurate predictive models from these data for conducting forest inventories. Machine learning techniques have the potential to solve this problem of modeling forest attributes from remotely sensed data. This work tests four different machine learning approaches - namely Support Vector Regression, Random Forest, Artificial Neural Networks, and Extreme Gradient Boosting - on high-density GatorEye UAV-Lidar point clouds for indirect estimation of individual tree dendrometric metrics (field-derived) such as diameter at breast height, total height, and timber volume. A total of 370 trees had their dbh and height measured for validation purposes. Using LAStools we generated normalized Light Detection and Ranging (Lidar) point clouds and created a raster canopy height model at a 0.5×0.5 m spatial resolution following the construction of a digital terrain model and a digital surface model. The R package 'lidR' was set with the functions tree_detection (local maximum filter algorithm) and lastrees. Subsequently, we applied the function tree_metrics to extract individual metrics. Machine learning techniques were applied to the derived metrics to estimate dendrometric field measures. The machine learning models (MLM) with optimal hyperparameters showed similar predictive performances for modeling the variables diameter, height, and volume. All models had a rRMSE below 15% (for diameter at breast height), 9% (for height) and 29% (for volume). The Support Vector Regression algorithm showed the best performance. Our work demonstrates that all tested machine learning models are adequate and robust to handle the high dimensionality of UAV-Lidar data for the estimation of individual attributes, with Support Vector Regression model being the best performer in terms of minimal error rates.

1. Introduction

Forest inventories are an integral component of monitoring and managing natural resources (Fankhauser et al., 2018). They are traditionally carried out through intensive field sampling, aiming to provide managers with an understanding of the composition and structure of a forest (Goodbody et al., 2017). In a traditional forest inventory, diameter at breast height (DBH) and total height are the main variables measured, as they are easier to measure and have strong relationships with other tree parameters; for example, the volume, which is of great importance for managers. Although traditional forest inventory methods can provide highly accurate vegetation parameters, collecting data in situ is time-consuming and labor intensive, especially for studies that comprise large scales (Means et al., 2000; White et al., 2016). Thus, there is a need for alternative or complementary methods that can overcome the disadvantages associated with field-survey based data acquisition (Williams et al., 1994).

In this context, remote sensing technologies provide auxiliary and valuable information that can be used to increase the accuracy and timeliness of forest parameter estimates (McRoberts et al., 2010; Kangas et al., 2018). Light Detection and Ranging (LiDAR) is an active remote sensing system, which collects ranging data utilizing the speed of light and information about the flight time of a laser pulse (Lim et al., 2003), and has emerged as a particularly useful technology for accurate characterization of forest properties

*Corresponding author
anapaulacorte@gmail.com (Ana Paula Dalla Corte)
ORCID(s): 0000-0001-8529-5554 (Ana Paula Dalla Corte)

at different resolutions. Compared with traditional optical remote sensing technologies, LiDAR has greater penetration and is not easily affected by weather conditions; therefore, this technology has unique advantages in obtaining forest structure information (Wu et al., 2019). Thus, with the possibility to receive detailed estimates of forest structure variables, the use of LiDAR in recent years has increased in forestry.

LiDAR applications have expanded rapidly in the past two decades to model forest structural attributes (Næsset, 2002; Popescu, 2007; Yu et al., 2011), aboveground biomass (AGB) (Popescu et al., 2011; Rex et al., 2019), forest fuel parameters (Kramer et al., 2014), as well as species diversity (Simonson et al., 2012; Alonzo et al., 2014). In the context of forest structural attributes, Næsset (2002) using an airborne laser scanning (ALS), found greater precision for the Lorey's mean height ($R^2 = 0.82$-0.95), followed by the mean diameter at breast height (DBH) ($R^2 = 0.39$-0.78) and volume ($R^2 = 0.80$-0.93). Popescu (2007), estimated the DBH of individual trees through regression analysis, using the LiDAR-derived height and crown diameter measurements. The findings showed a small RMSE of 4.9cm, which was approximately 18% of the dbh mean of all measured trees, with an R^2 value of 0.87. Yu et al. (2011) proposed an approach to predict forest attributes at the individual level using data from airborne LiDAR. They proposed a new detection method to find individual trees along with random forests as an estimation method. Correlation coefficients between the observed and predicted values of 0.93, 0.79 and 0.87 for individual tree height, DBH and stem volume, respectively, were achieved, based on 26 laser-derived features. Despite intense research efforts, operational applications of airborne LiDAR are wide conditioned on area-based approach (ABA), mainly due to the insufficient amount of pulses per square meter obtained by these systems, and also by specific techniques such as linear regression and random forest. In addition, the current high cost of ALS data has also limited their applications in small (up to 4 ha) and medium (4-400 ha) projects (Næsset, 2004; Reutebuch et al., 2005; Hudak et al., 2006; Belmonte et al., 2019).

However, there has been a significant increase in the use of unmanned aerial vehicles (UAVs) for forestry inventory applications due to their relative low cost, automation features, and considering the fact that they can support various types of useful sensors, such as visual, or multispectral cameras, LiDAR and radar (Morales et al., 2018). Also, when compared to traditional airborne LiDAR (a.k.a, Airborne Laser Scanning - ALS), UAV-Lidar data acquisitions provide much higher point densities (Morsdorf et al., 2017; Wieser et al., 2017), which allow the recovery of single tree-level forest inventory parameters (Wallace et al., 2014; Wieser et al., 2017; Corte et al., 2020). However, there is still a challenge in modeling the obtained high-density data. In many studies with a predictive focus using LiDAR data, the approach used involves selecting independent variables through their explanatory capacity for the dependent variable of interest (Silva et al., 2014; Mauro et al., 2017). Also, it is common to apply correlation tests to select independent variables for modeling the best parameters (Miura and Jones, 2010; Stark et al., 2012; Taylor et al., 2015; Zhang et al., 2017). In other situations, stepwise regression analysis (Means et al., 2000) or multiple regression analysis techniques (Næsset and Bjerknes, 2001; Næsset, 2002, 2004) are applied to determine the best predictive models. In such cases, it is common to face the problem of multicollinearity between predictor variables derived from the point cloud, often indicated by the statistical variance inflation factor (VIF) and multidimensionality, which imposes challenges to statistical modeling (Adam and Mutanga, 2009; Dalponte et al., 2009; Laurin et al., 2014; Junttila et al., 2015; Venier et al., 2019).

Evaluating new approaches to predictive modeling is important in the search for more accurate models and for overcoming problems common to the conventional multiple regression techniques, such as multicollinearity. This is a recurrent situation between predictor variables derived from Lidar point clouds. In addition, a large amount of data (500-1,500 or more points m^{-2}) and the high dimensionality of the resource space generated by UAV-Lidar are intrinsic characteristics that encourage the use of machine learning techniques. Recently, a new culture of statistical modeling - machine learning - has gained momentum and has been applied to solve challenging issues in several areas of science and technology. Unsurprisingly, it has shown great potential for modeling forest attributes from remotely sensed data that exhibit complex interactions (McRoberts, 2012; Valbuena et al., 2016; Jordan and Mitchell, 2015). Although there are several algorithms within the machine learning domain, Artificial Neural Networks (ANN) has been one of the most widely used in terms of forest modeling research, in addition to Support Vector Machine - SVM (Nieto et al., 2012; Montaño et al., 2017). K-nearest neighbors algorithm (k-NN) is another algorithm that has gained prominence, specifically for its high performance in predicting variables such as carbon stock, biomass, and volume (Fehrmann et al., 2008; Sanquetta et al., 2013, 2018; Souza et al., 2019; Knapp et al., 2020). Moreover, various modifications of machine learning techniques - such as deep learning - in combination with remotely sensed data have been recently used for moisture content estimation (Villacrés et al., 2019; Arevalo-Ramirez et al., 2020), tree abnormality detection (Nguyen et al., 2020), forest species classification and identification (Olschofsky and Köhl, 2020; Xi et al., 2020) and tree crown delineation (Wan Mohd Jaafar et al., 2018; Weinstein et al., 2020).

Predictive modeling using machine learning techniques can offer advantages over conventional regression. For example, the approach does not require the a priori specification of functional forms describing the predictive relationship(s) and the response variable (Fehrmann et al., 2008). They can also be superior, like artificial neural networks (ANNs), due to their ability to overcome several problems in forest data, such as nonlinear relationships, non-Gaussian distributions, multicollinearity, outliers and noise in the data

(Diamantopoulou and Milios, 2010). Additionally, they admit variables of different natures and tend to work better for multidimensional data. A potential approach to providing more accurate estimates in forest inventories would be to combine high-density point data and machine learning techniques, which could provide information at the individual tree level.

Even though the use of machine learning techniques is prevalent nowadays, similar research on deep learning models (LeCun et al., 2015; Schmidhuber, 2015) with a focus on individual tree-level modeling is scarce. In the scientific community, there are few works that address different machine learning techniques applied to LiDAR data in the context of the precise forest inventory, and most of them are focused on ABA. In addition, work from LiDAR data generally employs low or medium pulse density in its collection. Therefore, alternatives to address this problem need to be developed and tested. Thus, we intend to test the potential of four machine learning techniques to predict variables at the individual tree level through high-density pulse clouds collected from a UAV-LiDAR system.

2. Material and Methods

In this section we describe the collection mechanisms and processing of UAV-Lidar data obtained from GatorEye Unmanned Flying Laboratory and a brief review of the four machine learning techniques used to model biometric variables (dbh, ht, Vol) from individual trees. In addition, we describe the hyperparameter tuning process of the algorithms, the selection metrics and model comparison strategy, as well as mechanisms to assess the importance of predictors.

2.1. Study Site

Study site is located within Fazenda Canguiri, which belongs to Federal University of Paraná (UFPR), in the municipality of Pinhais, Paraná state, southern Brazil, in the latitude 25º24'03.48" South and longitude 49º07'08.54" West. The regional climate is classified as Cfb (humid subtropical with oceanic climate, without dry season and temperate summer), characterized by oceanic climate without a dry season, with a temperate summer, an annual average temperature of 17 °C (20.5 °C in January and 13 °C in July) and an annual rainfall of 1,550 mm, which is slightly concentrated in the warmest months, December to February. The driest months are July and August (Alvares et al., 2013). The site of study is part of a project called NITA (Center for Technological Innovation in Agriculture) (Fig. 1) is currently run, which has a forest plantation with approximately 17 hectares and corresponds to an iCLF system (integration of crop, livestock and seminal forest plantations of *Eucalyptus benthamii* Maiden et Cambage). The planting occurred in September 2013 by a contour line and had a spacing of 2×14m (357 individuals.ha^{-1}) (Porfírio-da Silva et al., 2010).

2.2. LiDAR Data Collection

UAV-Lidar data were collected using the GatorEye Unmanned Flying Laboratory (www.gatoreye.org) 'Generation 2', in October 2019. The GatorEye Gen2 system comprised a modified Phoenix Scout Ultra system, which incorporates a STIM Inertial Measurement Unit (IMU), an L1/L2 GNSS receiver, an SSD hard drive, and a Velodyne 32c Ultra Puck. The Ultra Puck has 32 individual lasers, each with a range of up to 220 m, and which are installed to provide an along-track field of view (FOV) of 40 degrees while collecting a full 360 degrees of cross-track data. The GatorEye Gen2 system also incorporated high-resolution visual and hyperspectral sensors, which were not used in this present study. The final LiDAR point cloud is based on a post-processed kinematic (PPK) flight trajectory produced using the GatorEye GNSS data fused with the IMU accelerometer and gyro data in Novatel Inertial Explorer software, and which combined with the Puck per-point laser angle and ranging information enables a final point cloud absolute spatial accuracy of approximately 5 cm RMSE (Wilkinson et al., 2019). To further maximize point cloud accuracy, we limited returns to a maximum distance of 100 meters from the sensor, and to an angular field of view (FOV) maximum of 120 degrees. The mission plan also flew slow (8 m/s) and low (45 m above-ground level), and with flight lines tightly spaced (15 meters apart), resulting in a sidelap cross-swath coverage of 93%. The final point density was 1500-2500 pts m^{-2}. Data collection and point cloud specifications are further described in Corte et al. (2020).

2.3. Field Data

Forest census data of our study area were collected shortly after the Lidar flights (October 2019). A total of 370 trees had their dbh (diameter at breast height) and total height (ht) measured by a measuring tape and a Haglöf Vertex IV hypsometer, respectively. The tree locations were obtained using a Garmin GPS receiver, model 62CSX.

The individual tree volume (Vol) was calculated for each tree by fitting a polynomial taper equation (Prodan, 1965) (5th-degree; Eq. 1) and applying the corresponding volume equation (Eq. 2, developed from 12 trees cut and cubed by the Smalian formula.

$$\frac{d_i}{dbh} = \beta_0 + \beta_1\left(\frac{h_i}{ht}\right) + \beta_2\left(\frac{h_i}{ht}\right)^2 + \beta_3\left(\frac{h_i}{ht}\right)^3 + \beta_4\left(\frac{h_i}{ht}\right)^4 + \beta_5\left(\frac{h_i}{ht}\right)^5 \quad (1)$$

$$vt = \frac{\pi}{40000} dbh^2 \int_{h_1}^{h_2} \left\{\beta_0 + \beta_1\left(\frac{h_i}{ht}\right) + \beta_2\left(\frac{h_i}{ht}\right)^2 + \beta_3\left(\frac{h_i}{ht}\right)^3 + \beta_4\left(\frac{h_i}{ht}\right)^4 + \beta_5\left(\frac{h_i}{ht}\right)^5\right\}^2 dh_i \quad (2)$$

Where: vt - the total volume in m^3; dbh - the diameter at breast height (cm); hi - the height at which the user desires a diameter prediction; di - diameter at height hi on the

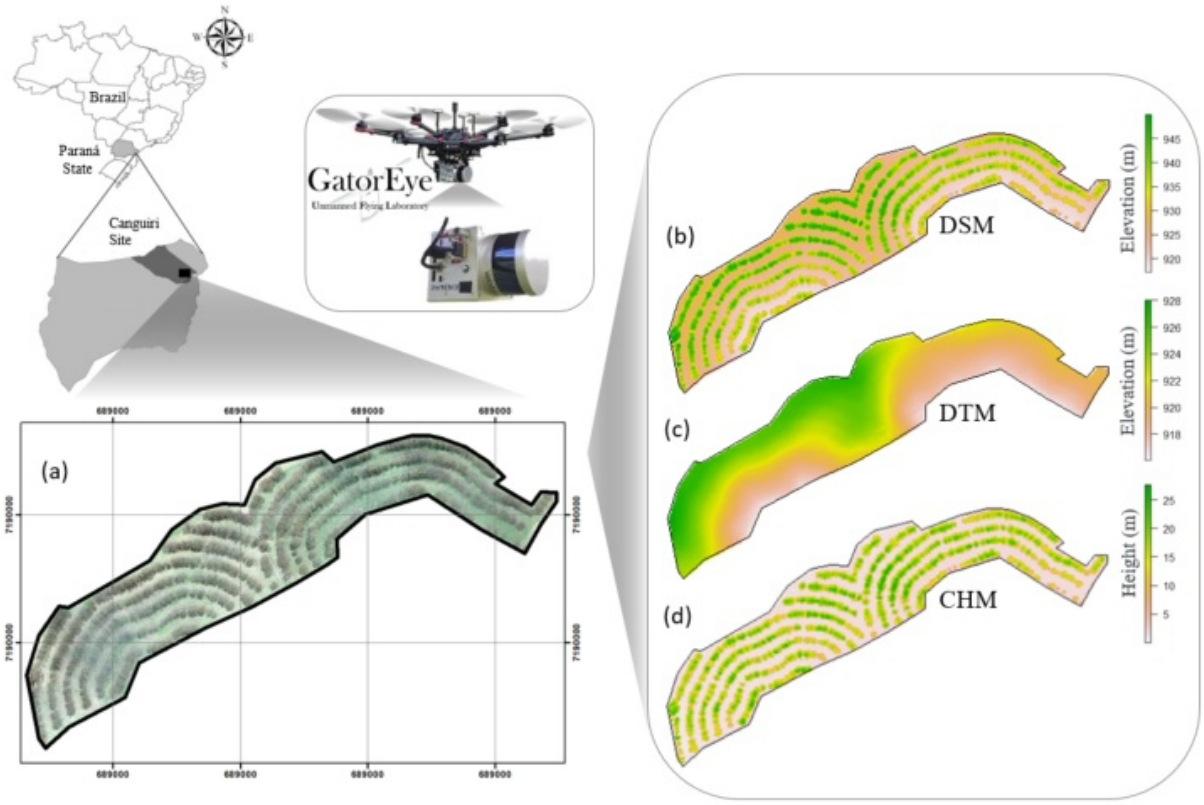

Figure 1: Study area, 17 hectares of *Eucalyptus* plantation at Canguiri Site in Pinhais City, Paraná State, Brazil. (a) Tree location; (b) DSM - Digital Surface Model; (c) DTM - Digital Terrain Model; (d) CHM - Canopy Height Model.

tree; ht - the total height or stem height; β_0=1.194915; β_1=-3.802747; β_2=15.859004; β_3=-34.711086; β_4=32.905061; β_5=-11.432145.

2.4. Lidar data and processing

The Lidar data processing and analysis were performed in an R environment (Version 3.6.1) (R Core Team, 2019) and Rstudio (Version 1.2.5001), using the functions of the LAStools software (Isenburg, 2019). First, we classified the ground returns and generated the digital terrain model (DTM). Then, we applied a spurious-return filter and generated the raster digital surface model (DSM). Based on the DTM, we generated the terrain-normalized Lidar point cloud, and the canopy height model (CHM). The CHM and the normalized point cloud were then clipped to the plantation's extent. All rasters were generated at a 0.5×0.5m spatial resolution.

To process the data, we applied: (1) the tree_detection function of the 'lidR' package using the local maximum filter (lmf) algorithm, as described by Popescu and Wynne (2004), for locating the position of trees; (2) the lastrees function, along with an algorithm for segmenting tree crowns known as dalponte2016, based on the Dalponte and Coomes (2016) algorithm; this function was used for segmenting individual tree crowns; and (3) the function tree_metrics to extract the point cloud metrics associated with each sequential tree number (called in this paper: treeID). The metrics evaluated are described in Tab. 1. Furthermore, a spatial join was performed with field data for accurate alignment of trees and comparisons of their respective derived metrics with real values of dbh, ht, and Vol.

2.5. Machine Learning Algorithms

In this study, four machine learning algorithms were tested to predict the three field-derived dentrometric metrics (dbh, ht, and Vol). The machine learning algorithms were also implemented using the R environment: Support Vector Regression ('kernlab' package) (Karatzoglou et al., 2004), Artificial Neural Networks ('nnet' package) (Venables and Ripley, 2002), Random Forest ('randomForest' package) (Liaw and Wiener, 2002), and Extreme Gradient Boosting ('xgboost' package) (Chen et al., 2019). In the following subsections, a brief description of the algorithms and their adjustment hyperparameters are presented.

2.5.1. Support Vector Regression

The method traditionally known as Support Vector Regression (SVR) (Kavaklioglu, 2011) was proposed by Vapnik 1995 as an extension to the traditional SVM algorithm, which is also referred to as the ϵ-SV algorithm (Torgo, 2017). The basic idea behind the ϵ-SV algorithm is to find an $f(x)$ function that has at most ϵ deviations from the real values y_i of the training set and, at the same time, as linear as possible (Smola and Schölkopf, 2004). In the R environ-

Table 1
Metrics extracted from the "tree_metrics" function for each tree.

Notation	Description
zmax	maximum height
zmean	mean height
zsd	standard deviation of height distribution
zskew	skewness of height distribution
zkurt	kurtosis of height distribution
zentropy	entropy of height distribution (see function entropy)
pzabovzmn	percentage of returns above zmean
pzabov2	percentage of returns above x
zq(x=5,...,95) range=5	xth percentile (quantile) of height distribution
zpcum(x=1,...,9) range=1	Cumulative percentage of return in the xth layer, according to Woods et al. (2008)
p(x=1,2)th	percentage xth returns

ment, the ϵ-SV version is available in the 'kernlab' package (Karatzoglou et al., 2004). In this study, we chose the radial-based kernel function, and two hyperparameters were tuned: *C* (cost of violation of restrictions) and *sigma* (kernel parameter of radial basis).

2.5.2. Artificial Neural Networks

"Multilayer Perceptron" (MLP) networks are among the most used networks in several fields of science. In general terms, the MLP network consists of three types of layers: an input layer, one or more hidden (or intermediate) layers, and an output layer (Fath et al., 2018). The most common architecture for an MLP is "completely connected", that is, all neurons in a layer *C* are connected to all other neurons in the layer *C+1*. Assuming that *C* is the first hidden layer, each *C* neuron is connected to all attributes of the predictor space (Gama et al., 2015). The 'nnet' package (Venables and Ripley, 2002) was used to train "feed-forward" single-layer MLP networks. Two hyperparameters were tuned: *size* (number of neurons in the hidden or intermediate layer) and *decay* (weight decay rate). The ANN model differs from traditional methods in that it uses several neurons in parallel to model a specific relationship.

2.5.3. Random Forests

The Random Forests (RF) algorithm was proposed by Leo Breiman in 2001 (Breiman, 2001). The algorithm constitutes a substantial modification of the bagging algorithm, one of the first ensemble algorithms developed and proposed by Leo Breiman in 1996 (Breiman, 1996). The main objective of Random Forests is to build a collection of "uncorrelated trees" and then average individual predictions in case of regression (Hastie et al., 2016). Thus, the main difference between bagging and Random Forests is the choice of k size (number of predictors) in the original predictor space (James et al., 2013). The 'randomForest' package (Liaw and Wiener, 2002) was used to build the models; two hyperparameters were tuned: *mtry* (number of trees to grow) and *ntree* (number of predictors used in the construction of each tree).

2.5.4. Extreme Gradient Boosting

The Extreme Gradient Boosting (XGBoost) algorithm has been widely used by data scientists to achieve excellent results in machine learning challenges, such as 'Kaggle' competitions. The algorithm was developed by Chen and Guestrin (2016) and constituted an efficient and scalable implementation of the framework Gradient Boosting Machine. In the R environment, XGBoost can be accessed through the 'xgboost' package (Chen et al., 2019) and numerous hyperparameters were available for tuning: *eta* (learning rate), *max_depth* (maximum tree depth), *min_child_weight* (minimum sum of the required instance weight on a child node), *subsample* (fraction of randomly selected training set instances), *colsample_bytree* (proportion of column subsamples to build each tree), *gamma* (minimum loss reduction required to make an additional partition in a tree node), and *nrounds* (number of trees to be grown).

2.6. Hyperparameter Tuning Process

The machine learning models were trained using the CARET package interface (Classification and Regression Training), a framework available for classification and regression tasks (Kuhn et al., 2016). Before the model learning process, the data set was divided into training (70%) and testing (30%) using stratified sampling based on tree diameter.

A common approach to estimate the expected performance of a machine learning model is to use some method for resampling the original data (Kuhn and Johnson, 2013). The tenfold cross-validation method was applied using the training set to obtain performance estimates of predictive models (Fig. 2). The final performance of each model was obtained by calculating the arithmetic mean of estimates in k cross-validation partitions. After determining the optimal hyperparameter tuning for each algorithm, the models were fit to the training set. Then, the performance of the models was evaluated on a test set with data not used in learning the predictive models. Before the process of model learning, the "center", "scale", and "BoxCox" methods were applied to transform predictors.

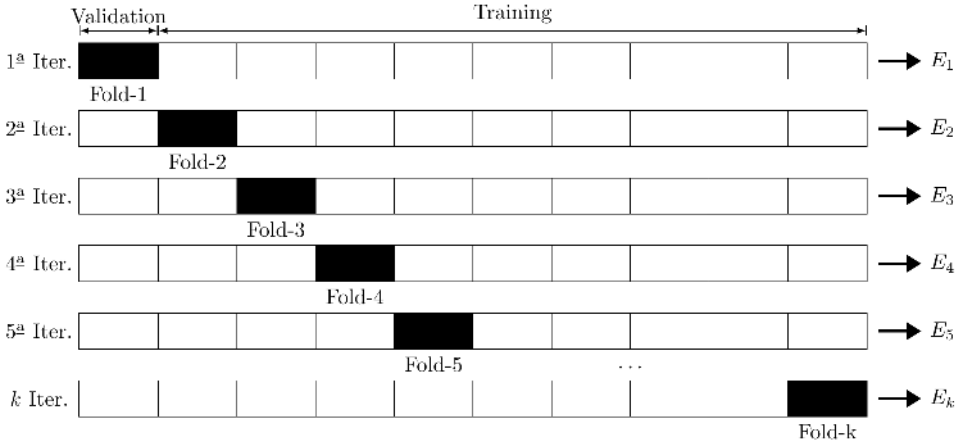

Figure 2: Representation of the k-fold cross-validation method performed.

Table 2
Candidate hyperparameters for each machine learning algorithm and libraries used for modeling field-derived dendrometric metrics (dbh, ht, and Vol).

Algorithm	Variable	Hyperparameter variants	Method/package	Author
Support Vector Regression (SVR)	dbh	C = 2^(-2,-1,0,1,2,3,4,5,6,7,8,9,10) sigma = seq(0.0001, 0.0005, 0.0001)	svmRadial/kernlab	Karatzoglou et al. (2004)
	ht	C = 2^(-2,0,2,4,5) sigma = seq(0.0001, 0.009, 0.0001)		
	vt	C = 2^c(-2,0,2,3,4,5,6,7,8,9,10) sigma = seq(0.0001,0.009,0.0001)		
Artificial Neural Network (ANN)	dbh, ht	size = seq(1,15,1) decay = seq(0.1,1,0.1)	nnet/nnet	Venables and Ripley (2002)
	vt	size = seq(1,20,1) decay = seq(0.001,0.01,0.001)		
Random Forest (RF)	dbh, ht, vt	mtry = c(1:35) ntree = c(50,100,150,200,300,400,500)	rf /randomForest	Liaw and Wiener (2002)
Extreme Gradient Boosting (XGBoost)	dbh, ht, vt	nrounds = seq(200,1000,50) eta = seq(0.01,0.07,0.01) max_depth = (1:5) gamma = c(0,1,10) colsample_bytree = seq(0.1,1,0.1) min_child_weight = seq(5,50,5) subsample = seq(0.1,1,0.1)	xgbTree/Xgboost	Chen et al. (2019)

In general, machine learning algorithms have specific hyperparameters that must be tuned to find the best predictive performance setting. Here, the grid search strategy was used. A grid of candidate hyperparameters was defined for each machine learning algorithm. The hyperparameter variants used for the design of machine learning models, as well as the respective packages (Tab. 2).

2.7. Model selection and importance of predictors

The performance of designed predictive models was evaluated using the metrics: Relative Root Mean Square Error (RMSE), Relative Root Mean Square Error (rRMSE), Coefficient of Determination (R^2), Mean Absolute Error (MAE), and Bias (Kvålseth, 1985; Pretzsch, 2009; Kuhn and Johnson, 2013; Chai and Draxler, 2014; Tanaka et al., 2015). The Pearson's correlation coefficient (r) was used to quantify the correlation between the observed and the predicted values for each predictive model (Zhang et al., 2015). An r value of 1 indicates a perfect correlation between predicted and observed values.

The Bland-Altman (B-A) method was used to assess the difference in performance between machine learning models (MLMs) with optimal hyperparameter tuning based on the estimates of the RMSE metric in the cross-validated paired partitions. In practical terms, the B-A method assesses the degree of agreement between two quantitative measures (Altman and Bland, 1983; Bland and Altman, 2010). B-A statistics were calculated using the "blandr" package (Datta,

2017). The method allows estimating bias (or average difference between models) and limits of agreement (limits and upper), which can be estimated by $\bar{d} \pm 1.96 * sd$, where: $\bar{d}$ = average difference or bias; sd = standard deviation of the differences between methods. Confidence intervals (CI) for bias and limits of agreement are also estimated. The closer the mean difference (bias) is to zero, the better the agreement between the measures will be (Hirakata and Camey, 2009; Odor et al., 2017).

In order to better understand (and/or interpret) the built machine learning models, the relative importance of predictors was calculated using the DALEX framework (Descriptive Machine Learning Explanations). This library has a collection of functions that aim to assist in providing explanations of predictive models (Biecek, 2018). Therefore, charts of the importance of predictor variables were plotted.

3. Results

In general, machine learning models, with the optimal hyperparameters tuning (machine learning models, MLMs), showed high predictive performance to modeling the three dependent variables (dbh, ht, and Vol) (Tab. 3). The implemented models showed an rRMSE below 15%, 9%, and 29% for the estimate of dbh, ht, and Vol, respectively. The MLMs built using the SVR algorithm showed the lowest rRMSE, MAE, and MAD values. Regarding the stepwise linear regression (SLR), the SVRs models showed a better performance for predicting the diameter, height and volume, with a reduction of approximately 0.85%, 0.65% and 1.5%, respectively, in the average of rRMSE estimated in the cross validation. The increase in the accuracy of machine learning models was small when compared to the SLR method. Even so, the modeling of biometric variables from the combination of high-density UAV-Lidar data and machine learning techniques can be considered very promising.

In this study, several metrics derived from UAV-Lidar clouds showed high linear correlation. We observed that the metrics that describe the percentiles of the height distribution (from zq = 35 to zq = 95) showed positive correlations between themselves and also with the metrics: zsd, zmean and zmax. Pearson's correlation coefficient between these predictors was greater than 0.75, with almost perfect correlations (Fig. 7, Appendix A). The SLR algorithm estimated final models with a large number of metrics and the vast majority of them with high collinearity. The Variance Inflation Factor (VIF) statistic indicated severe multicollinearity (VIF > 10) for the final models estimated by SLR. The variables selected by SLR modeling for each model, the coefficients significance and the VIF statistics can be found in Table 4 in the Appendix A.

Evaluating the resampling distributions of the statistical models is particularly useful to know their stability in predicting the average response in future observations. Therefore, boxplot graphs can be developed here to provide a good comparison of the distribution of cross-validation estimates (RMSE) among machine learning models with an optimal hyperparameter tuning (Fig. 3). In general, for the dbh variable, the RMSE averages among SVR, RF, and XGBoost models were similar, but the variance was higher for RF (CV = 16.89%). The other models showed a variation coefficient lower than 15%. For the variable ht, the ANN model showed less variance (CV = 12.71%), but the mean RMSE was higher. The MLMs for the variable volume had similar averages. The SVR (19.42%) and ANN (18.81%) models had greater dispersion in resampling and were influenced by influential points (outliers).

The Bland-Altman graph was used to compare MLMs designed using the SVR algorithm that had the lowest rRMSE (Fig. 4). The bias estimates were not considered significant, and the equality line was within the confidence intervals of the average difference in most cases. In addition, most differences between models were within the limits of agreement ($\bar{d} \pm 1.96 * sd$), and the normality of different distributions was admitted through the Shapiro-Wilk test (α= 0.05). Therefore, there is evidence to admit that the statistical models have a consistent average predictive performance.

The relative importance of predictors (threshold $\geq$ 80%) for MLMs with a lower rRMSE for each modeled variable is available in (Fig. 5). The "zmean" predictor (mean height) showed a greater relative importance in SVR models designed to predict the variables dbh and volume. On the other hand, "zmax" (maximum height) was the most important predictor for the SVR model learned to predict tree height. In general, the percentiles of height distribution provided good information for MLM training. However, lower percentiles (zq5 to zq20) seemed to provide less relevant information.

The MLMs with optimal hyperparameter tuning were evaluated in the test set (n = 98). In general, MLMs showed similarity in the residual distribution and in the estimation of average performance to predict the variables dbh, ht, and Vol in future samples. The performance of the models in the test set (Fig. 6) is compatible with cross-validation estimates and confirms the similar predictive capacity of the models, as evidenced by the Bland-Altman method.

4. Discussion

We examined the performance of four machine learning approaches using high-dimensional UAV-Lidar data for the estimation of structural forest attributes at the individual level in eucalyptus stands in southern Brazil. The performance of each approach was compared and rated based on RMSE, R^2 and Bias to select the most appropriate model. Our findings demonstrated that machine learning models are adequate and robust enough to handle the high dimensionality of UAV-Lidar data and are able to estimate the dendrometric metrics at the individual level. With the fast development of remote sensing technologies and given the applicability of state-of-the-art statistical analysis methods on UAV systems derived data, it is possible to examine forest structural attributes with high accuracy (Dandois and Ellis, 2013; Wallace et al., 2014). In our study, we examined three

Table 3
Optimal hyperparameter tuning and average performance estimate in cross-validation for machine learning models and stepwise linear regression built using high-density UAV-Lidar GatorEye data.

Model	Hyperparameter tuning	Statistics	Tenfold cross-validation						
			RMSE	rRMSE	MSE	r	R^2	MAE	Bias%
Diameter (cm)									
SLR	See Tab. 4 (Appendix A)	Mean	3.973	14.049	15.967	0.561	0.337	3.148	0.196
		Sd.	0.449	1.572	3.718	0.159	0.183	0.342	2.824
ANN	size = 1	Mean	4.1695	14.7416	17.6921	0.5009	0.2819	3.2728	-0.051
	decay = 0.6	Sd.	-0.5846	-2.0321	-4.9087	-0.1855	-0.1869	-0.3722	-2.0373
RF	mtry = 6	Mean	3.7559	13.2772	14.469	0.6244	0.4076	3.02	-0.0019
	ntree = 150	Sd.	-0.6345	-2.2029	-4.6873	-0.1404	-0.1678	-0.5091	-2.2963
SVR	sigma = 0.0011	Mean	**3.7343**	**13.1992**	14.2253	0.6195	0.4037	2.9475	0.0919
	C = 64	Sd.	-0.558	-1.9168	-4.1617	-0.1489	-0.1808	-0.4129	-1.8501
XGBoost	nrounds = 250 eta = 0.02 max_depth = 1 gamma = 1 colsample_bytree = 0.4 min_child_weight = 5 subsample = 0.4	Mean	3.7872	13.3942	14.6083	0.626	0.4062	3.0247	-0.5581
		Sd.	-0.5429	-1.9317	-4.1065	-0.1263	-0.1658	-0.4496	-2.1034
Total height (m)									
SLR	See Tab. 4 (Appendix A)	Mean	1.58	8.348	2.52	0.8	0.645	1.225	0.162
		Sd.	0.162	0.905	0.507	0.08	0.131	0.151	1.722
ANN	size = 1	Mean	1.5534	8.21	2.4481	0.82	0.6763	1.2381	-0.2629
	decay = 0.9	Sd.	-0.1975	-1.1076	-0.588	-0.0663	-0.1108	-0.1633	-1.9155
RF	mtry = 3	Mean	1.506	7.9597	2.3277	0.8208	0.6799	1.2069	0.0623
	ntree = 100	Sd.	-0.2574	-1.4093	-0.7523	-0.0829	-0.1374	-0.1909	-1.4136
SVR	sigma = 0.0006	Mean	**1.4581**	**7.7048**	2.1706	0.8328	0.6983	1.1491	0.1487
	C = 32	Sd.	-0.2221	-1.2121	-0.646	-0.0726	-0.1223	-0.1687	-1.7867
XGBoost	nrounds = 850 eta = 0.02 max_depth = 1 gamma = 0 colsample_bytree = 0.1 min_child_weight = 5 subsample = 0.9	Mean	1.5055	7.9564	2.3087	0.8256	0.6863	1.2011	-0.0805
		Sd.	-0.2166	-1.1966	-0.6467	-0.0725	-0.1218	-0.1405	-1.6024
Volume (m^3)									
SLR	See Tab. 4 (Appendix A)	Mean	0.15	28.24	0.023	0.659	0.455	0.116	0.588
		Sd.	0.026	4.12	0.007	0.154	0.198	0.021	6.96
ANN	size = 1	Mean	0.1539	28.9764	0.0244	0.6411	0.4351	0.1206	0.1296
	decay = 0.005	Sd.	-0.029	-4.8342	-0.0083	-0.1637	-0.2	-0.0218	-6.3006
RF	mtry = 1	Mean	0.1457	27.435	0.0219	0.6764	0.4749	0.115	0.6279
	ntree = 150	Sd.	-0.0262	-4.2859	-0.0072	-0.1387	-0.1869	-0.01978	-5.7806
SVR	sigma = 0.0029	Mean	**0.1422**	**26.7408**	0.0209	0.6908	0.4928	0.1088	-2.0909
	C = 4	Sd.	-0.0277	-4.558	-0.0071	-0.1312	-0.1764	-0.0222	-5.0925
XGBoost	nrounds = 250 eta = 0.01 max_depth = 2 gamma = 0 colsample_bytree = 0.2 min_child_weight = 15 subsample = 0.7	Mean	0.1447	27.2792	0.0214	0.6951	0.4957	0.1142	-0.121
		Sd.	-0.0229	-3.714	-0.0064	-0.1183	-0.1521	-0.016	-5.9603

Where: Sd. = Standard deviation.

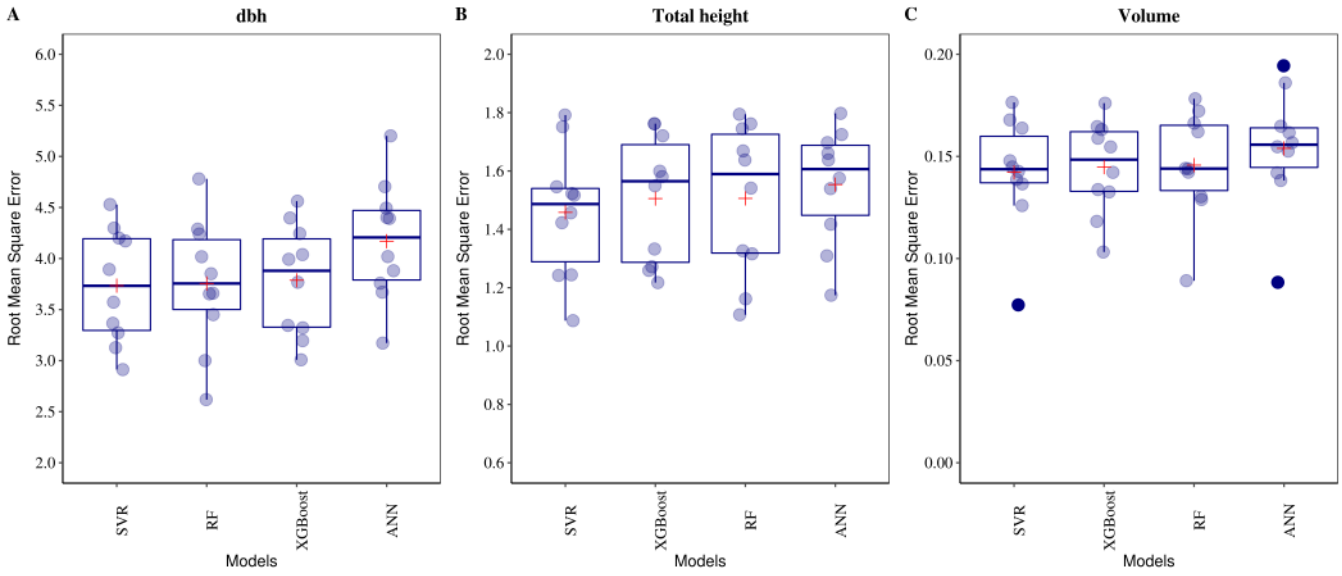

Figure 3: Distribution of resampling estimates of RMSE for machine learning models with the optimal hyperparameters tuning using high-density UAV-Lidar GatorEye data. Vertical bars (blue) represent Q1 - 1.5*IQR and Q3 + 1.5*IQR; the red cross represents the average performance of the models in resampling. Black points are outliers. ANN = Artificial Neural Networks; SVR = Support Vector Regression; XGBoost = Extreme Gradient Boosting; RF = Random Forest.

of the most commonly used parameters in forest data modeling projects - dbh, total height, and volume at an individual tree level.

In the case of SLR, after the execution of the algorithm, it is expected that highly collinear predictors are excluded from the final regression model, since the addition of a collinear metrics does not improve the model's ability to explain. However, in this study, those models that used SLR incorporated highly correlated variables and, therefore, presented severe multicollinearity, with Variance Inflation Factors (VIF) greater than 10. The most serious effect of multicollinearity is the increase in standard errors of the estimated parameters (Alin, 2010) and, consequently, the confidence intervals associated with the coefficients that tend to be broader (Gujarati and Porter, 2011). Some research suggests that ML techniques can be robust in the presence of multicollinearity, for example, artificial neural networks (RNAs) and support vector machines (SVM) (Kotsiantis et al., 2007; Kang et al., 2015). Kotsiantis et al. (2007) state that ANN and SVM perform well when multicollinearity is present and there is a non-linear relationship between the covariates and the response variable. Drake et al. (2006) modeling ecological niches with support vector machines expose that useful information can be obtained by adding more environmental variables, even if they are highly correlated. Therefore, although the ML approaches used in this study can be robust to correlated metrics, the impacts of variable selection (or elimination of highly correlated metrics) and the use of dimensionality reduction techniques are interesting aspects to be addressed in future research.

Among the factors that affect the estimates generated, selecting an ideal modeling approach is one of the most important steps in most cases (Fassnacht et al., 2014). Regarding the performance of the machine learning approaches tested here, we noticed that the models presented a very similar performance for the studied variables (Adj-R^2 = 0.28-0.69, rRMSE = 7.70-28.97%, were: Adj-R^2 = Adjusted R-Squared). The total height showed the smallest difference (ΔAdj-R^2 = 0.02, Δr RMSE = 0.51%), while volume and diameter showed greater differences. Thus, the results of the models showed that the total height estimate (r = 0.83, rRMSE = 7.70%) had the highest precision, followed by volume (r = 0.69, rRMSE = 26.74%) and dbh (r = 0.61, rRMSE = 13.19%). Despite a very close performance among the tested algorithms in terms of rRMSE and correlation coefficient (r), we found that SVR presented the best performance among all approaches. This algorithm presented the most appropriate statistics to generate estimates with fewer errors for all variables. We also noticed that ANN was not able to explain the behavior of dbh. It presented the worst performance among ML approaches in terms of r and R^2.

When comparing our approach with that of recent publications, we noticed that there is a trend and a wide discussion in the application of machine learning modeling (Dong et al., 2019; Hernando et al., 2019; Marrs and Ni-Meister, 2019; Malek et al., 2019). Despite this tendency to use more flexible approaches in relation to traditional methods, we also realize that most studies have focused on above-ground biomass (AGB). Only a few studies have aimed to further develop individual-level approaches for structural parameter estimates. Along this line, our estimated results for stem volume and dbh can be deemed very promising for individual tree modeling. Yu et al. (2011) reported that an estimate of stem volume and dbh in an area of boreal forest was achieved with a relative RMSE of 38% and 21%, respectively, in the best cases, based on 26 laser-derived characteristics. Malek

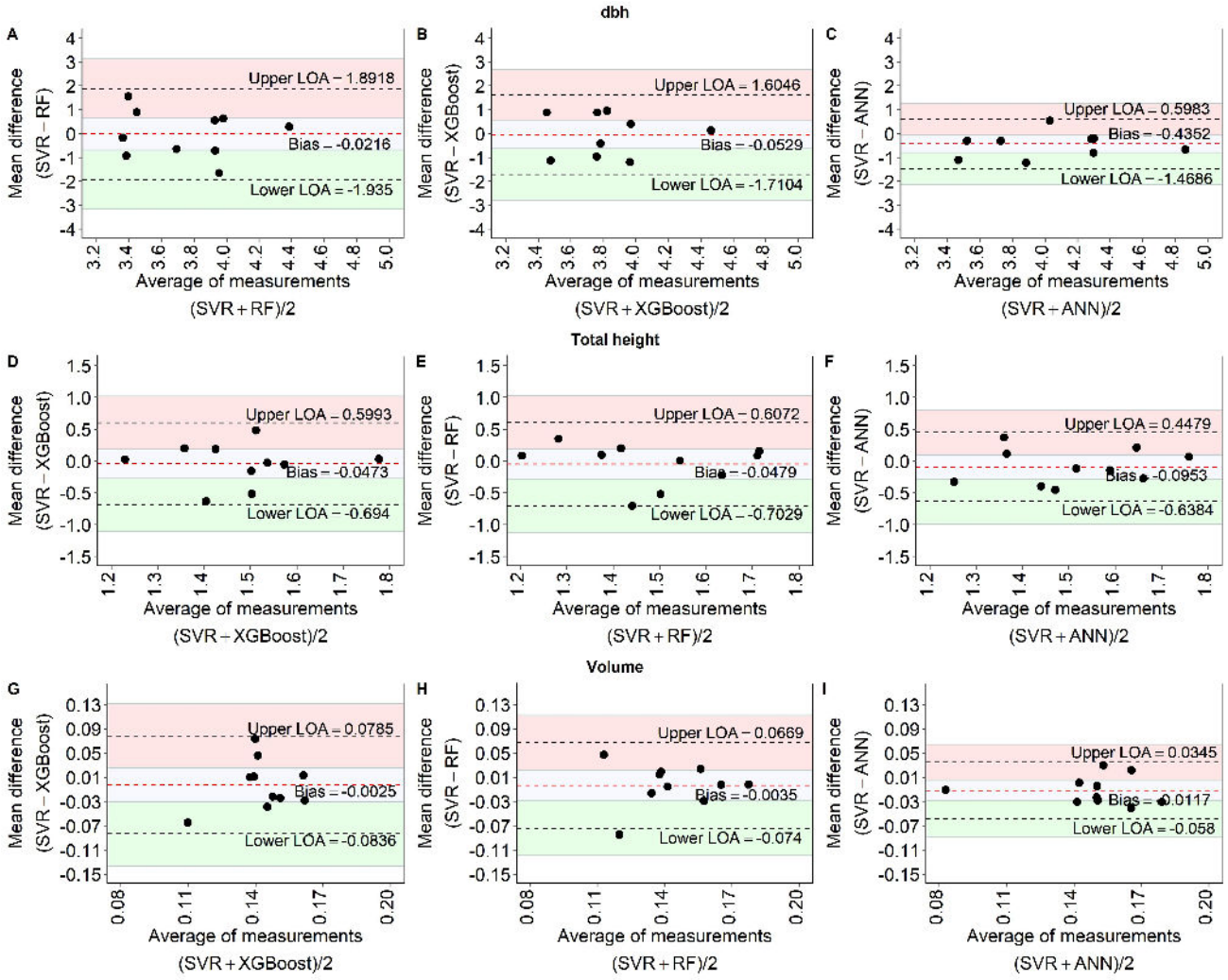

Figure 4: Bland-Altman method for analysis of agreement between machine learning models with optimal hyperparameters tuning using high-density UAV-Lidar GatorEye data. Bias = mean difference for RMSE between two machine learning models (horizontal red dotted line); Lower LOA = Lower limit of agreement; Upper LOA = Upper limit of agreement; hatched regions = 95% confidence interval for Upper LOA (red), Lower LOA (green) and bias (gray). ANN = Artificial Neural Network; SVR = Support Vector Regression; XGBoost = Extreme Gradient Boosting; RF = Random Forest.

et al. (2019) used SVR and RF to predict dbh and AGB at the Individual Tree Crown (ITC) level using metrics extracted from ALS data and achieved the best results using the SVR algorithm for both variables studied, which is in line with the results presented here. However, our findings show an improvement in RMSE compared to those authors' approaches. In the best of our cases, SVR presented an RMSE of 3.73cm for dbh (Tab. 3) while for Malek et al. (2019), the best result was 4.93cm, a difference of approximately 1.2cm in dbh estimates.

Because SVR has the best performance among machine learning models, we selected this model for the analysis of other factors that may be interesting from the point of view of designing a more robust model. Regarding the most important variables in the design of machine learning models (Fig. 5), SVR identified average height as the most important variable for modeling volume and diameter variables, while for the total height, the model identified maximum height as the most important variable, which was already expected since Hmax represents the total height of trees. These findings agree with previous studies when relating them to estimates of structural variables of forests based on LiDAR data (Popescu et al., 2003; Heurich and Thoma, 2008; Ioki et al., 2010). It is also interesting to note that the average height metric (especially high percentiles) has shown in previous studies to be a reliable proxy for forest properties, such as above-ground biomass (Fassnacht et al., 2014; Kattenborn et al., 2015; Latifi et al., 2012; Rex et al., 2020). In a similar vein, we also observed a pattern related to heights in the selected set (Li et al., 2014; García-Gutiérrez et al., 2015). Estimating the relative importance of predictor variables for MLMs learned is interesting as it is possible to identify covariables that provide minimal information to model the response variable. These variables can be excluded from pre-

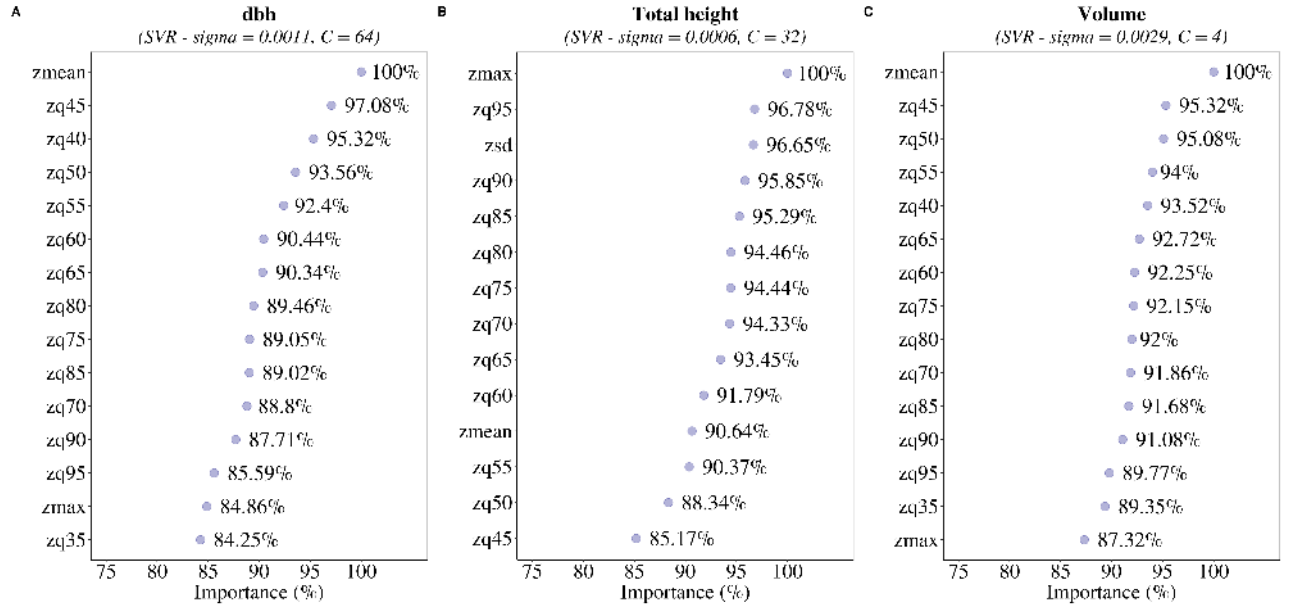

Figure 5: Demonstration of the relative importance of predictors for machine learning models with lower rRMSE. Sigma = radial based kernel function parameter; C = penalty parameter. SVR = Support Vector Regression.

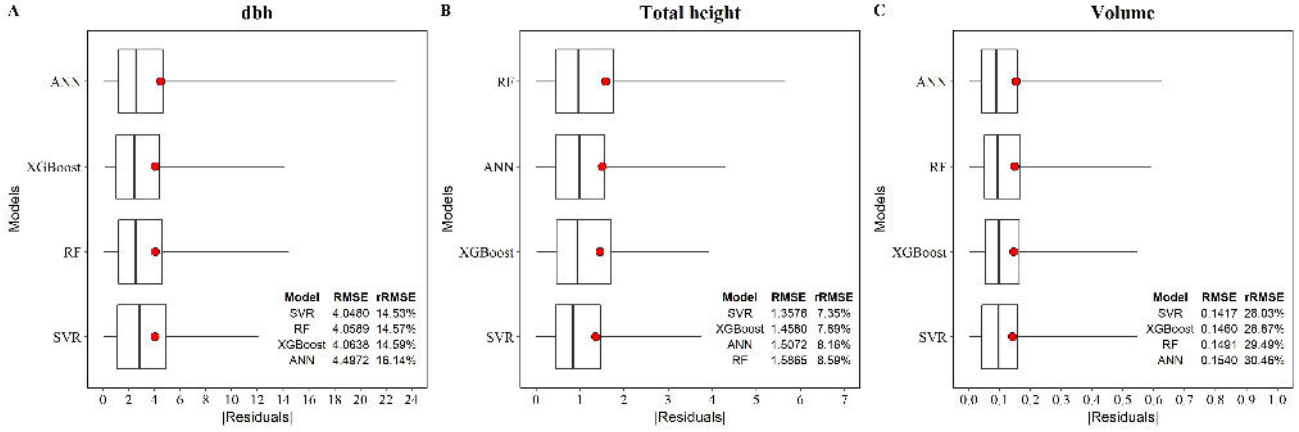

Figure 6: Comparison of residual distribution in the test set (n = 98) for machine learning models with optimal hyperparameters tuning using high-density UAV-Lidar GatorEye data. ANN = Artificial Neural Network; SVR = Support Vector Regression; XGBoost = Extreme Gradient Boosting; RF = Random Forest.

dictive models, and thereby we can decrease the computational cost of the learning process.

Furthermore, we compared our results with approaches commonly used in the literature - such as plot-based methods and found that the accuracy of our models for forest structural attributes was relatively inferior (for example, García-Gutiérrez et al. (2015); Lee et al. (2018)). García-Gutiérrez et al. (2015) conducted a plot-based study to estimate forest attributes using machine learning approaches and achieved results superior to those of this study. Lee et al. (2018) also achieved better results compared to us in terms of R^2 value; they obtained an $R^2 = 0.93$ for the average height of the plot in the best of cases. In general, forest inventory estimation using LiDAR data can be performed using two types of approaches: individual trees and plots (Hyyppä et al., 2008). The fact that plot-based approaches present results with fewer errors in relation to individual tree modeling may be related to the lower use of sample units used for the modeling, which ends up showing a higher value of R^2. In our study we used more than 300 sample units for robust modeling, this fact may have resulted in a lower R^2 value (Tab. 3) when compared to previous studies. Besides, some studies that focused on the effects of point density upon measuring forest structure suggested that increasing point density results in a greater precision at tree level (Disney et al., 2010; Lovell et al., 2005) - a fact that corroborates the findings of this study.

This study provides a fundamental reference for the selection of machine learning algorithms and their respective adjusted hyperparameters for the estimation of forest variables at the individual level based on UAV-Lidar data. Cloud metrics have been shown to be related to the main forest

measurements and can be derived with high reliability from UAV-Lidar systems, generating better estimates of forest parameters. In addition, from a technical point of view, we can highlight the possibility of creating machine learning approaches as interesting alternatives to traditional regression methods as here the metrics are chosen are not limited and consequently, this helps avoid loss of important information while predicting the variable of interest, maintaining a similar level of performance (Görgens et al., 2015). Our study demonstrated that, in iCLF system, LiDAR data and machine learning modeling could be useful and are able to assist in obtaining variables at the individual level, although care is needed to deal with the uncertainties inherent to modeling. Also, our study represents one of the first contributions in the context of forest management in ILFP-type planted forests in Brazil for the process of estimating structural variables of the forest using UAV-Lidar data.

5. Conclusion

Our work demonstrates that all tested machine learning algorithms (SVR, RF, ANN, and XGBoost) are adequate and robust to handle the high dimensionality of UAV-Lidar data for the estimation of structural attributes of the planted forest at an individual level, even in the presence of predictors with high collinearity. However, we consider that the SVR models performed slightly better to predict the response variables, even compared to the stepwise linear regression method. Also, the tested models were found to vary less among themselves for the variable height than for diameter and volume.

Predictors derived from high-density UAV-Lidar data showed high linear correlation. Unlike expected, the stepwise linear regression method using akaike information criterion (AIC) did not provide models without predictors with high collinearity. In future research it is interesting to evaluate the performance of other parametric statistical techniques not included in this study, considered robust in the presence of multicollinearity in the regressors.

Likewise, although the machine learning techniques used in this study may be robust to the high collinearity between predictors, the impacts of variable selection (or elimination of highly correlated predictors) and the use of dimensionality reduction techniques are interesting aspects to address in future research.

Author contributions

For research articles with several authors, a short paragraph specifying their individual contributions must be provided. Ana Paula Dalla Corte: conceptualization, formal analysis, investigation, methodology, project administration, resources, writing-original draft; Deivison Venicio Souza: formal analysis, methodology, resources, writing-original draft; Franciel Eduardo Rex: formal analysis, methodology, resources, writing-original draft; Carlos Roberto Sanquetta: formal analysis, methodology, resources, writing-original draft; Midhun Mohan: resources, writing-review and editing; Carlos Alberto Silva: resources, writing-review and editing; Angelica Maria Almeyda Zambrano: resources, writing-review and editing; Gabriel Prata: resources, writing-review and editing; Danilo Almeida: resources, writing-review and editing; Jonathan William Trautenmüller: writing-review; Carine Klauberg: writing-review; Anibal de Moraes: writing-review; Mateus N. Sanquetta: writing-review; Ben Wilkinson: resources, writing-review and editing; Eben North Broadbent: conceptualization, data curation, investigation, project administration, resources, supervision, writing-review and editing. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

The authors are very grateful to the Spatial Ecology and Conservation (SPEC) Lab at the University of Florida who funded and collected the GatorEye Unmanned Flying Laboratory lidar data, with support from the USDA National Institute of Food and Agriculture McIntire-Stennis program, and the Federal University of Parana (UFPR) and NITA working group coordinated by Professor Anibal de Moraes for the use of the study area.

Funding

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Finance Code 001 (A. Corte #88887.373249/2019-00), MCTIC/CNPq Nº 28/2018 (#408785/2018-7; #438875/2018-4), CNPq Nº 09/2018 (#302891/2018-8). D. Almeida was supported by the São Paulo Research Foundation (#2018/21338-3 and #2019/14697-0).

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Adam, E., Mutanga, O., 2009. Spectral discrimination of papyrus vegetation (cyperus papyrus l.) in swamp wetlands using field spectrometry. ISPRS Journal of Photogrammetry and Remote Sensing 64, 612–620. doi:10.1016/j.isprsjprs.2009.04.004.

Alin, A., 2010. Multicollinearity. Wiley Interdisciplinary Reviews: Computational Statistics 2, 370–374.

Alonzo, M., Bookhagen, B., Roberts, D.A., 2014. Urban tree species mapping using hyperspectral and lidar data fusion. Remote Sensing of Environment 148, 70–83. doi:10.1016/j.rse.2014.03.018.

Altman, D.G., Bland, J.M., 1983. Measurement in medicine: the analysis of method comparison studies. Journal of the Royal Statistical Society: Series D (The Statistician) 32, 307–317. doi:10.2307/2987937.

Alvares, C.A., Stape, J.L., Sentelhas, P.C., de Moraes Gonçalves, J.L., Sparovek, G., 2013. Köppen's climate classification map for brazil. Meteorologische Zeitschrift 22, 711–728.

Arevalo-Ramirez, T., Villacrés, J., Fuentes, A., Reszka, P., Cheein, F.A.A., 2020. Moisture content estimation of pinus radiata and eucalyptus globulus from reconstructed leaf reflectance in the SWIR region. Biosystems Engineering 193, 187–205. doi:10.1016/j.biosystemseng.2020.03.004.

Belmonte, A., Sankey, T., Biederman, J.A., Bradford, J., Goetz, S.J., Kolb, T., Woolley, T., 2019. UAV-derived estimates of forest structure to inform ponderosa pine forest restoration. Remote Sensing in Ecology and Conservation 6, 181–197. doi:10.1002/rse2.137.

Biecek, P., 2018. Dalex: explainers for complex predictive models. ArXiv e-prints arXiv:1806.08915.

Bland, J.M., Altman, D.G., 2010. Statistical methods for assessing agreement between two methods of clinical measurement. International journal of nursing studies 47, 931–936. doi:10.1016/S0140-6736(86)90837-8.

Breiman, L., 1996. Bagging predictors. Machine learning 24, 123–140. doi:10.1007/BF00058655.

Breiman, L., 2001. Random forests. Machine learning 45, 5–32.

Chai, T., Draxler, R.R., 2014. Root mean square error (rmse) or mean absolute error (mae)?–arguments against avoiding rmse in the literature. Geoscientific model development 7, 1247–1250. doi:10.5194/gmd-7-1247-2014.

Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, ACM. pp. 785–794. doi:10.1145/2939672.2939785.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., Li, Y., 2019. xgboost: Extreme Gradient Boosting. URL: https://CRAN.R-project.org/package=xgboost. r package version 0.82.1.

Corte, A.P.D., Rex, F.E., de Almeida, D.R.A., Sanquetta, C.R., Silva, C.A., Moura, M.M., Wilkinson, B., Zambrano, A.M.A., da Cunha Neto, E.M., Veras, H.F.P., de Moraes, A., Klauberg, C., Mohan, M., Cardil, A., Broadbent, E.N., 2020. Measuring individual tree diameter and height using GatorEye high-density UAV-lidar in an integrated crop-livestock-forest system. Remote Sensing 12, 863. doi:10.3390/rs12050863.

Dalponte, M., Bruzzone, L., Vescovo, L., Gianelle, D., 2009. The role of spectral resolution and classifier complexity in the analysis of hyperspectral images of forest areas. Remote Sensing of Environment 113, 2345–2355. doi:10.1016/j.rse.2009.06.013.

Dalponte, M., Coomes, D.A., 2016. Tree-centric mapping of forest carbon density from airborne laser scanning and hyperspectral data. Methods in Ecology and Evolution 7, 1236–1245. doi:10.1111/2041-210x.12575.

Dandois, J.P., Ellis, E.C., 2013. High spatial resolution three-dimensional mapping of vegetation spectral dynamics using computer vision. Remote Sensing of Environment 136, 259–276. doi:10.1016/j.rse.2013.04.005.

Datta, D., 2017. blandr: a Bland-Altman Method Comparison package for R. doi:10.5281/zenodo.824514.

Diamantopoulou, M.J., Milios, E., 2010. Modelling total volume of dominant pine trees in reforestations via multivariate analysis and artificial neural network models. Biosystems engineering 105, 306–315. doi:10.1016/j.biosystemseng.2009.11.010.

Disney, M., Kalogirou, V., Lewis, P., Prieto-Blanco, A., Hancock, S., Pfeifer, M., 2010. Simulating the impact of discrete-return lidar system and survey characteristics over young conifer and broadleaf forests. Remote Sensing of Environment 114, 1546–1560. doi:10.1016/j.rse.2010.02.009.

Dong, L., Tang, S., Min, M., Veroustraete, F., Cheng, J., 2019. Aboveground forest biomass based on OLSR and an ANN model integrating LiDAR and optical data in a mountainous region of china. International Journal of Remote Sensing 40, 6059–6083. doi:10.1080/01431161.2019.1587201.

Drake, J.M., Randin, C., Guisan, A., 2006. Modelling ecological niches with support vector machines. Journal of Applied Ecology 43, 424–432. doi:10.1111/j.1365-2664.2006.01141.x.

Fankhauser, K., Strigul, N., Gatziolis, D., 2018. Augmentation of traditional forest inventory and airborne laser scanning with unmanned aerial systems and photogrammetry for forest monitoring. Remote Sensing 10, 1562. doi:10.3390/rs10101562.

Fassnacht, F., Hartig, F., Latifi, H., Berger, C., Hernández, J., Corvalán, P., Koch, B., 2014. Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass. Remote Sensing of Environment 154, 102–114. doi:10.1016/j.rse.2014.07.028.

Fath, A.H., Madanifar, F., Abbasi, M., 2018. Implementation of multilayer perceptron (mlp) and radial basis function (rbf) neural networks to predict solution gas-oil ratio of crude oil systems. Petroleum doi:10.1016/j.petlm.2018.12.002.

Fehrmann, L., Lehtonen, A., Kleinn, C., Tomppo, E., 2008. Comparison of linear and mixed-effect regression models and ak-nearest neighbour approach for estimation of single-tree biomass. Canadian Journal of Forest Research 38, 1–9. doi:10.1139/X07-119.

Gama, J., Carvalho, A.C.P.d.L., Faceli, K., Lorena, A.C., Oliveira, M., et al., 2015. Extração de conhecimento de dados: data mining.

García-Gutiérrez, J., Martínez-Álvarez, F., Troncoso, A., Riquelme, J., 2015. A comparison of machine learning regression techniques for LiDAR-derived estimation of forest variables. Neurocomputing 167, 24–31. doi:10.1016/j.neucom.2014.09.091.

Goodbody, T.R., Coops, N.C., Marshall, P.L., Tompalski, P., Crawford, P., 2017. Unmanned aerial systems for precision forest inventory purposes: A review and case study. The Forestry Chronicle 93, 71–81. doi:10.5558/tfc2017-012.

Görgens, E.B., Montaghi, A., Rodriguez, L.C.E., 2015. A performance comparison of machine learning methods to estimate the fast-growing forest plantation yield based on laser scanning metrics. Computers and Electronics in Agriculture 116, 221–227. doi:10.1016/j.compag.2015.07.004.

Gujarati, D., Porter, D., 2011. Econometria Básica. 5 ed., AMGH Editora Ltda.

Hastie, T., Tibshirani, R., Friedman, J., 2016. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.

Hernando, A., Puerto, L., Mola-Yudego, B., Manzanera, J., García-Abril, A., Maltamo, M., Valbuena, R., 2019. Estimation of forest biomass components using airborne LiDAR and multispectral sensors. iForest - Biogeosciences and Forestry 12, 207–213. doi:10.3832/ifor2735-012.

Heurich, M., Thoma, F., 2008. Estimation of forestry stand parameters using laser scanning data in temperate, structurally rich natural european beech (fagus sylvatica) and norway spruce (picea abies) forests. Forestry 81, 645–661. doi:10.1093/forestry/cpn038.

Hirakata, V.N., Camey, S.A., 2009. Análise de concordância entre métodos de bland-altman. Clinical & Biomedical Research 29, 261–268.

Hudak, A.T., Smith, A.M., Evans, J.S., Falkowski, M.J., 2006. Estimating coniferous forest canopy cover from lidar and multispectral data, in: AGU Fall Meeting Abstracts, pp. B43D–03.

Hyyppä, J., Hyyppä, H., Leckie, D., Gougeon, F., Yu, X., Maltamo, M., 2008. Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. International Journal of Remote Sensing 29, 1339–1366. doi:10.1080/01431160701736489.

Ioki, K., Imanishi, J., Sasaki, T., Morimoto, Y., Kitada, K., 2010. Estimating stand volume in broad-leaved forest using discrete-return LiDAR: plot-based approach. Landscape and Ecological Engineering 6, 29–36. doi:10.1007/s11355-009-0077-4.

Isenburg, M., 2019. Lastools—efficient lidar processing software,(version 1.8, licensed). URL: http://rapidlasso.com/LAStools. accessed on 11 November 2019.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An introduction to statistical learning. volume 112. Springer.

Jordan, M.I., Mitchell, T.M., 2015. Machine learning: Trends, perspectives, and prospects. Science 349, 255–260. doi:10.1126/science.aaa8415.

Junttila, V., Kauranne, T., Finley, A.O., Bradford, J.B., 2015. Linear models for airborne-laser-scanning-based operational forest inventory with small field sample size and highly correlated LiDAR data. IEEE Transactions on Geoscience and Remote Sensing 53, 5600–5612. doi:10.1109/tgrs.2015.2425916.

Kang, J., Schwartz, R., Flickinger, J., Beriwal, S., 2015. Machine learning approaches for predicting radiation therapy outcomes: A clinician's perspective. International Journal of Radiation Oncology Biology Physics 93, 1127–1135. doi:10.1016/j.ijrobp.2015.07.2286.

Kangas, A., Astrup, R., Breidenbach, J., Fridman, J., Gobakken, T., Korhonen, K.T., Maltamo, M., Nilsson, M., Nord-Larsen, T., Næsset, E., Olsson, H., 2018. Remote sensing and forest inventories in nordic countries – roadmap for the future. Scandinavian Journal of Forest Research 33, 397–412. doi:10.1080/02827581.2017.1416666.

Karatzoglou, A., Smola, A., Hornik, K., Zeileis, A., 2004. kernlab – an S4 package for kernel methods in R. Journal of Statistical Software 11, 1–20. URL: http://www.jstatsoft.org/v11/i09/.

Kattenborn, T., Maack, J., Faßnacht, F., Enßle, F., Ermert, J., Koch, B., 2015. Mapping forest biomass from space – fusion of hyperspectral EO1-hyperion data and tandem-x and WorldView-2 canopy height models. International Journal of Applied Earth Observation and Geoinformation 35, 359–367. doi:10.1016/j.jag.2014.10.008.

Kavaklioglu, K., 2011. Modeling and prediction of turkey's electricity consumption using support vector regression. Applied Energy 88, 368–375. doi:10.1016/j.apenergy.2010.07.021.

Knapp, N., Fischer, R., Cazcarra-Bes, V., Huth, A., 2020. Structure metrics to generalize biomass estimation from lidar across forest types from different continents. Remote Sensing of Environment 237, 111597. doi:10.1016/j.rse.2019.111597.

Kotsiantis, S.B., Zaharakis, I., Pintelas, P., 2007. Supervised machine learning: A review of classification techniques. Emerging artificial intelligence applications in computer engineering 160, 3–24.

Kramer, H., Collins, B., Kelly, M., Stephens, S., 2014. Quantifying ladder fuels: A new approach using LiDAR. Forests 5, 1432–1453. doi:10.3390/f5061432.

Kuhn, M., Johnson, K., 2013. Applied predictive modeling. volume 810. Springer.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., the R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., Hunt., T., 2016. caret: Classification and Regression Training. URL: https://CRAN.R-project.org/package=caret. r package version 6.0-73.

Kvålseth, T.O., 1985. Cautionary note about r2. The American Statistician 39, 279–285. doi:10.2307/2683704.

Latifi, H., Fassnacht, F., Koch, B., 2012. Forest structure modeling with combined airborne hyperspectral and LiDAR data. Remote Sensing of Environment 121, 10–25. doi:10.1016/j.rse.2012.01.015.

Laurin, G.V., Chen, Q., Lindsell, J.A., Coomes, D.A., Frate, F.D., Guerriero, L., Pirotti, F., Valentini, R., 2014. Above ground biomass estimation in an african tropical forest with lidar and hyperspectral data. ISPRS Journal of Photogrammetry and Remote Sensing 89, 49–58. doi:10.1016/j.isprsjprs.2014.01.001.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. doi:10.1038/nature14539.

Lee, J., Im, J., Kim, K., Quackenbush, L., 2018. Machine learning approaches for estimating forest stand height using plot-based observations and airborne LiDAR data. Forests 9, 268. doi:10.3390/f9050268.

Li, M., Im, J., Quackenbush, L.J., Liu, T., 2014. Forest biomass and carbon stock quantification using airborne LiDAR data: A case study over huntington wildlife forest in the adirondack park. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7, 3143–3156. doi:10.1109/jstars.2014.2304642.

Liaw, A., Wiener, M., 2002. Classification and regression by randomforest. R News 2, 18–22. URL: https://www.researchgate.net/profile/Andy_Liaw/publication/228451484_Classification_and_Regression_by_RandomForest/links/53fb24cc0cf20a45497047ab/Classification-and-Regression-by-RandomForest.pdf.

Lim, K., Treitz, P., Wulder, M., St-Onge, B., Flood, M., 2003. LiDAR remote sensing of forest structure. Progress in Physical Geography: Earth and Environment 27, 88–106. URL: https://doi.org/10.1191/0309133303pp360ra, doi:10.1191/0309133303pp360ra.

Lovell, J., Jupp, D., Newnham, G., Coops, N., Culvenor, D., 2005. Simulation study for finding optimal lidar acquisition parameters for forest height retrieval. Forest Ecology and Management 214, 398–412. doi:10.1016/j.foreco.2004.07.077.

Malek, S., Miglietta, F., Gobakken, T., Næsset, E., Gianelle, D., Dalponte, M., 2019. Prediction of stem diameter and biomass at individual tree crown level with advanced machine learning techniques. iForest - Biogeosciences and Forestry 12, 323–329. doi:10.3832/ifor2980-012.

Marrs, J., Ni-Meister, W., 2019. Machine learning techniques for tree species classification using co-registered LiDAR and hyperspectral data. Remote Sensing 11, 819. doi:10.3390/rs11070819.

Mauro, F., Monleon, V., Temesgen, H., Ruiz, L., 2017. Analysis of spatial correlation in predictive models of forest variables that use LiDAR auxiliary information. Canadian Journal of Forest Research 47, 788–799. doi:10.1139/cjfr-2016-0296.

McRoberts, R.E., 2012. Estimating forest attribute parameters for small areas using nearest neighbors techniques. Forest Ecology and Management 272, 3–12. doi:10.1016/j.foreco.2011.06.039.

McRoberts, R.E., Tomppo, E.O., Næsset, E., 2010. Advances and emerging issues in national forest inventories. Scandinavian Journal of Forest Research 25, 368–381. doi:10.1080/02827581.2010.496739.

Means, J.E., Acker, S.A., Fitt, B.J., Renslow, M., Emerson, L., Hendrix, C.J., et al., 2000. Predicting forest stand characteristics with airborne scanning lidar. Photogrammetric Engineering and Remote Sensing 66, 1367–1372.

Miura, N., Jones, S.D., 2010. Characterizing forest ecological structure using pulse types and heights of airborne laser scanning. Remote Sensing of Environment 114, 1069–1076. doi:10.1016/j.rse.2009.12.017.

Montaño, R.A.N.R., Sanquetta, C.R., Wojciechowski, J., Mattar, E., Dalla Corte, A.P., Todt, E., 2017. Artificial intelligence models to estimate biomass of tropical forest trees. Polibits 56, 29–37. doi:10.17562/PB-56-4.

Morales, G., Kemper, G., Sevillano, G., Arteaga, D., Ortega, I., Telles, J., 2018. Automatic segmentation of mauritia flexuosa in unmanned aerial vehicle (UAV) imagery using deep learning. Forests 9, 736. doi:10.3390/f9120736.

Morsdorf, F., Eck, C., Zgraggen, C., Imbach, B., Schneider, F.D., Kükenbrink, D., 2017. UAV-based LiDAR acquisition for the derivation of high-resolution forest and ground information. The Leading Edge 36, 566–570. doi:10.1190/tle36070566.1.

Næsset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sensing of Environment 80, 88–99. doi:10.1016/s0034-4257(01)00290-5.

Næsset, E., 2004. Practical large-scale forest stand inventory using a small-footprint airborne scanning laser. Scandinavian Journal of Forest Research 19, 164–179. doi:10.1080/02827580310019257.

Næsset, E., Bjerknes, K.O., 2001. Estimating tree heights and number of stems in young forest stands using airborne laser scanner data. Remote Sensing of Environment 78, 328–340. doi:10.1016/s0034-4257(01)00228-0.

Nguyen, V.T., Constant, T., Kerautret, B., Debled-Rennesson, I., Colin, F., 2020. A machine-learning approach for classifying defects on tree trunks using terrestrial LiDAR. Computers and Electronics in Agriculture 171, 105332. doi:10.1016/j.compag.2020.105332.

Nieto, P.G., Torres, J.M., Fernández, M.A., Galán, C.O., 2012. Support vector machines and neural networks used to evaluate paper manufactured using eucalyptus globulus. Applied Mathematical Modelling 36, 6137–6145. doi:10.1016/j.apm.2012.02.016.

Odor, P.M., Bampoe, S., Cecconi, M., 2017. Cardiac output monitoring: Validation studies–how results should be presented. Current Anesthesiology Reports 7, 410–415. doi:10.1007/s40140-017-0239-0.

Olschofsky, K., Köhl, M., 2020. Rapid field identification of cites timber species by deep learning. Trees, Forests and People 2, 100016. doi:10.1016/j.tfp.2020.100016.

Popescu, S.C., 2007. Estimating biomass of individual pine trees using airborne lidar. Biomass and Bioenergy 31, 646–655. doi:10.1016/j.biombioe.2007.06.022.

Popescu, S.C., Wynne, R.H., 2004. Seeing the trees in the forest. Photogrammetric Engineering & Remote Sensing 70, 589–604. doi:10.14358/pers.70.5.589.

Popescu, S.C., Wynne, R.H., Nelson, R.F., 2003. Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. Canadian journal of remote sensing 29, 564–577.

Popescu, S.C., Zhao, K., Neuenschwander, A., Lin, C., 2011. Satellite lidar

vs. small footprint airborne lidar: Comparing the accuracy of aboveground biomass estimates and forest structure metrics at footprint level. Remote Sensing of Environment 115, 2786–2797. doi:10.1016/j.rse.2011.01.026.

Pretzsch, H., 2009. Forest dynamics, growth, and yield, in: Forest dynamics, growth and yield. Springer, pp. 1–39.

Prodan, M., 1965. Holzmesslehre. Technical Report. Sauerländer's Verlag: Frankfurt.

R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Reutebuch, S.E., Andersen, H.E., McGaughey, R.J., 2005. Light detection and ranging (lidar): an emerging tool for multiple resource inventory. Journal of forestry 103, 286–292.

Rex, F.E., Corte, A.P.D., do Amaral Machado, S., Silva, C.A., Sanquetta, C.R., 2019. Estimating above-ground biomass of araucaria angustifolia (bertol.) kuntze using LiDAR data. Floresta e Ambiente 26. doi:10.1590/2179-8087.110717.

Rex, F.E., Silva, C.A., Corte, A.P.D., Klauberg, C., Mohan, M., Cardil, A., da Silva, V.S., de Almeida, D.R.A., Garcia, M., Broadbent, E.N., Valbuena, R., Stoddart, J., Merrick, T., Hudak, A.T., 2020. Comparison of statistical modelling approaches for estimating tropical forest aboveground biomass stock and reporting their changes in low-intensity logging areas using multi-temporal LiDAR data. Remote Sensing 12, 1498. URL: https://doi.org/10.3390/rs12091498, doi:10.3390/rs12091498.

Sanquetta, C.R., Piva, L.R., Wojciechowski, J., Corte, A.P., Schikowski, A.B., 2018. Volume estimation of cryptomeria japonica logs in southern brazil using artificial intelligence models. Southern Forests: a Journal of Forest Science 80, 29–36. doi:10.2989/20702620.2016.1263013.

Sanquetta, C.R., Wojciechowski, J., Corte, A.P.D., Rodrigues, A.L., Maas, G.C.B., 2013. On the use of data mining for estimating carbon storage in the trees. Carbon Balance and Management 8. doi:10.1186/1750-0680-8-6.

Schmidhuber, J., 2015. Deep learning in neural networks: An overview. Neural Networks 61, 85–117. doi:10.1016/j.neunet.2014.09.003.

Silva, C.A., Klauberg, C., e Carvalho, S.d.P.C., Hudak, A.T., et al., 2014. Mapping aboveground carbon stocks using lidar data in eucalyptus spp. plantations in the state of são paulo, brazil. Scientia Forestalis. 42 (104): 591-604. 42, 591–604.

Porfírio-da Silva, V., Medrado, M.J.S., Nicodemo, M.L.F., Dereti, R.M., 2010. Arborização de pastagens com espécies florestais madeireiras: implantação e manejo. Embrapa Pecuária Sudeste-Folderes/Folhetos/Cartilhas (INFOTECA-E) .

Simonson, W.D., Allen, H.D., Coomes, D.A., 2012. Use of an airborne lidar system to model plant species composition and diversity of mediterranean oak forests. Conservation Biology 26, 840–850. doi:10.1111/j.1523-1739.2012.01869.x.

Smola, A.J., Schölkopf, B., 2004. A tutorial on support vector regression. Statistics and computing 14, 199–222. doi:10.1023/B:STCO.0000035301.49549.88.

Souza, D.V., Nievola, J.C., Santos, J.X., Wojciechowski, J., Gonçalves, A.L., Corte, A.P.D., Sanquetta, C.R., 2019. k-nearest neighbor regression in the estimation of Tectona grandis trunk volume in the state of Pará, Brazil. Journal of Sustainable Forestry 38, 755–768. doi:10.1080/10549811.2019.1607391.

Stark, S.C., Leitold, V., Wu, J.L., Hunter, M.O., de Castilho, C.V., Costa, F.R.C., McMahon, S.M., Parker, G.G., Shimabukuro, M.T., Lefsky, M.A., Keller, M., Alves, L.F., Schietti, J., Shimabukuro, Y.E., Brandão, D.O., Woodcock, T.K., Higuchi, N., de Camargo, P.B., de Oliveira, R.C., Saleska, S.R., 2012. Amazon forest carbon dynamics predicted by profiles of canopy leaf area and light environment. Ecology Letters 15, 1406–1414. doi:10.1111/j.1461-0248.2012.01864.x.

Tanaka, S., Takahashi, T., Nishizono, T., Kitahara, F., Saito, H., Iehara, T., Kodani, E., Awaya, Y., 2015. Stand volume estimation using the k-NN technique combined with forest inventory data, satellite image data and additional feature variables. Remote Sensing 7, 378–394. doi:10.3390/rs70100378.

Taylor, P., Asner, G., Dahlin, K., Anderson, C., Knapp, D., Martin, R., Mascaro, J., Chazdon, R., Cole, R., Wanek, W., Hofhansl, F., Malavassi, E., Vilchez-Alvarado, B., Townsend, A., 2015. Landscape-scale controls on aboveground forest carbon stocks on the osa peninsula, costa rica. PLOS ONE 10, e0126748. doi:10.1371/journal.pone.0126748.

Torgo, L., 2017. Data mining with R: learning with case studies. Chapman and Hall/CRC.

Valbuena, R., Maltamo, M., Packalen, P., 2016. Classification of forest development stages from national low-density lidar datasets: a comparison of machine learning methods. Revista de Teledetección , 15doi:10.4995/raet.2016.4029.

Venables, W., Ripley, B., 2002. Modern applied statistics with s. doi:10.1007/978-0-387-21706-2.

Venier, L.A., Swystun, T., Mazerolle, M.J., Kreutzweiser, D.P., Wainio-Keizer, K.L., McIlwrick, K.A., Woods, M.E., Wang, X., 2019. Modelling vegetation understory cover using LiDAR metrics. PLOS ONE 14, e0220096. doi:10.1371/journal.pone.0220096.

Villacrés, J., Arevalo-Ramirez, T., Fuentes, A., Reszka, P., Auat Cheein, F., 2019. Foliar moisture content from the spectral signature for wildfire risk assessments in valparaíso-chile. Sensors 19. doi:https://doi.org/10.3390/s19245475.

Wallace, L., Lucieer, A., Watson, C.S., 2014. Evaluating tree detection and segmentation routines on very high resolution UAV LiDAR data. IEEE Transactions on Geoscience and Remote Sensing 52, 7619–7628. doi:10.1109/tgrs.2014.2315649.

Wan Mohd Jaafar, W.S., Woodhouse, I.H., Silva, C.A., Omar, H., Abdul Maulud, K.N., Hudak, A.T., Klauberg, C., Cardil, A., Mohan, M., 2018. Improving individual tree crown delineation and attributes estimation of tropical forests using airborne lidar data. Forests 9. doi:https://doi.org/10.3390/f9120759.

Weinstein, B.G., Marconi, S., Bohlman, S.A., Zare, A., White, E.P., 2020. Cross-site learning in deep learning RGB tree crown detection. Ecological Informatics 56, 101061. doi:10.1016/j.ecoinf.2020.101061.

White, J.C., Coops, N.C., Wulder, M.A., Vastaranta, M., Hilker, T., Tompalski, P., 2016. Remote sensing technologies for enhancing forest inventories: A review. Canadian Journal of Remote Sensing 42, 619–641. doi:10.1080/07038992.2016.1207484.

Wieser, M., Mandlburger, G., Hollaus, M., Otepka, J., Glira, P., Pfeifer, N., 2017. A case study of UAS borne laser scanning for measurement of tree stem diameter. Remote Sensing 9, 1154. doi:10.3390/rs9111154.

Wilkinson, B., Lassiter, H.A., Abd-Elrahman, A., Carthy, R.R., Ifju, P., Broadbent, E., Grimes, N., 2019. Geometric targets for UAS lidar. Remote Sensing 11, 3019. doi:10.3390/rs11243019.

Williams, M.S., Bechtold, W.A., LaBau, V.J., 1994. Five instruments for measuring tree height: An evaluation. Southern Journal of Applied Forestry 18, 76–82. doi:10.1093/sjaf/18.2.76.

Woods, M., Lim, K., Treitz, P., 2008. Predicting forest stand variables from lidar data in the great lakes st. lawrence forest of ontario. The Forestry Chronicle 84, 827–839.

Wu, X., Shen, X., Cao, L., Wang, G., Cao, F., 2019. Assessment of individual tree detection and canopy cover estimation using unmanned aerial vehicle based light detection and ranging (UAV-LiDAR) data in planted forests. Remote Sensing 11, 908. doi:10.3390/rs11080908.

Xi, Z., Hopkinson, C., Rood, S.B., Peddle, D.R., 2020. See the forest and the trees: Effective machine and deep learning algorithms for wood filtering and tree species classification from terrestrial laser scanning. ISPRS Journal of Photogrammetry and Remote Sensing 168, 1–16. doi:10.1016/j.isprsjprs.2020.08.001.

Yu, X., Hyyppä, J., Vastaranta, M., Holopainen, M., Viitala, R., 2011. Predicting individual tree attributes from airborne laser point clouds based on the random forests technique. ISPRS Journal of Photogrammetry and Remote Sensing 66, 28–37. doi:10.1016/j.isprsjprs.2010.08.003.

Zhang, J., Chen, W., Sun, P., Zhao, X., Ma, Z., 2015. Prediction of protein solvent accessibility using PSO-SVR with multiple sequence-derived features and weighted sliding window scheme. BioData Mining 8. doi:10.1186/s13040-014-0031-3.

Zhang, Z., Cao, L., She, G., 2017. Estimating forest structural parameters using canopy metrics derived from airborne lidar data in subtropical forests. Remote Sensing 9, 940. doi:10.3390/rs9090940.

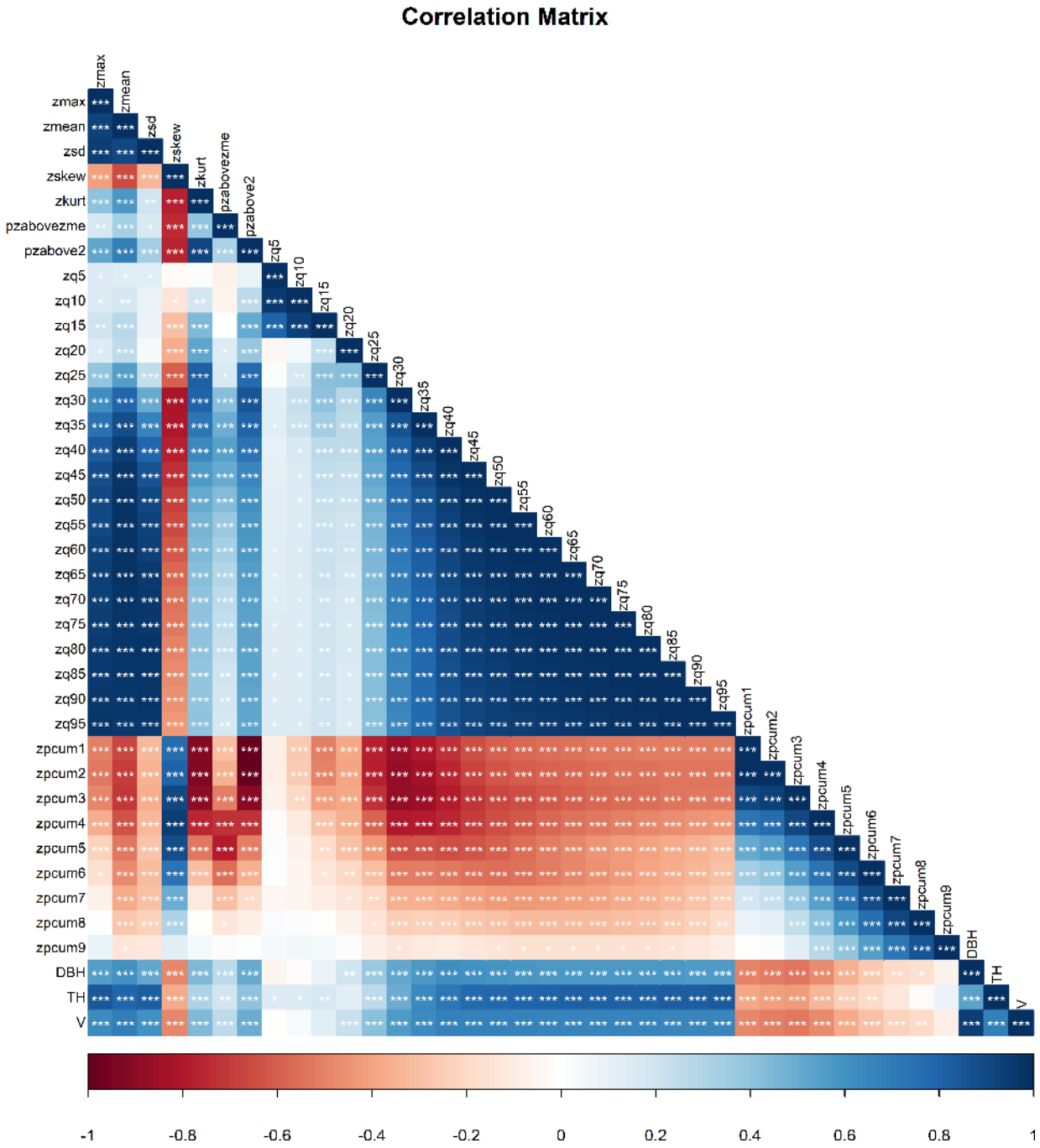

Figure 7: Pearson's correlation matrix between predictor variables derived from high-density UAV-Lidar GatorEye data and response variables. Relation with symbol = correlation coefficient is significant ($\alpha = 0.05$). Relation without symbol = correlation coefficient is regarded as insignificant.

A. Appendix

Table 4
Predictors selected using the stepwise linear regression method, estimated parameters and variance inflation factor.

Parameter	Diameter			Total height			Volume		
	EP	SP	VIF	EP	SP	VIF	EP	SP	VIF
$\hat{\beta}_0$	28.286			18.941			0.529		
$\hat{\beta}_1$	-9.600	zmax	322.486	5.795	zmean	715.725	0.883	zmean	930.323
$\hat{\beta}_2$	44.618	zmean	2828.947	8.806	zsd	441.900	0.800	zsd	540.831
$\hat{\beta}_3$	38.080	zsd	1064.911	-1.650	zskew	62.560	-0.148	zskew	35.237
$\hat{\beta}_4$	-7.621	zskew	63.290	-0.456	pzabovezme	9.232	-0.065	pzabovezme	7.324
$\hat{\beta}_5$	-2.210	zkurt	36.391	-0.615	pzabove2	15.015	0.019	zq20	1.412
$\hat{\beta}_6$	4.567	pzabove2	20.697	-0.514	zq5	6.229	-0.233	zq60	225.053
$\hat{\beta}_7$	0.757	zq20	2.121	0.815	zq15	8.511	-0.291	zq70	248.936
$\hat{\beta}_8$	-1.347	zq30	9.821	-5.356	zq60	490.583	-0.623	zq80	505.952
$\hat{\beta}_9$	-1.995	zq35	14.646	4.819	zq65	894.918	-0.406	zq95	366.376
$\hat{\beta}_{10}$	-3.726	zq40	42.556	-5.054	zq70	1155.127	0.128	zpcum6	15.299
$\hat{\beta}_{11}$	-11.060	zq60	329.965	6.444	zq75	1820.327	-0.085	zpcum8	7.089
$\hat{\beta}_{12}$	-15.611	zq70	494.137	-15.564	zq80	1958.233			
$\hat{\beta}_{13}$	-13.912	zq80	809.936	9.834	zq85	1220.944			
$\hat{\beta}_{14}$	-14.511	zq85	810.216	-7.490	zq90	754.999			
$\hat{\beta}_{15}$	-10.960	zq95	555.393	-0.990	zpcum3	27.238			
$\hat{\beta}_{16}$	3.461	zpcum5	28.124	0.947	zpcum4	18.930			
$\hat{\beta}_{17}$	4.555	zpcum6	28.023	0.581	zpcum6	5.917			
$\hat{\beta}_{18}$	1.756	zpcum7	23.674						
$\hat{\beta}_{19}$	-4.636	zpcum8	25.817						
$\hat{\beta}_{20}$	0.916	zpcum9	6.914						

Where: SP = Selected predictors; EP = Estimated parameters; VIF = Variance Inflation Factor; $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$,..., $\hat{\beta}_n$ = regression coefficients.