



Estimating Stand Height and Tree Density in *Pinus taeda* plantations using *in-situ* data, airborne LiDAR and *k*-Nearest Neighbor Imputation

CARLOS ALBERTO SILVA^{1,2}, CARINE KLAUBERG¹, ANDREW T. HUDAK¹, LEE A. VIERLING², VERALDO LIESENBERG³, LUIZ G. BERNETT⁴, CLEWERTON F. SCHERAIBER⁴ and EMERSON R. SCHOENINGER⁴

¹USDA Forest Service, Rocky Mountain Research Station/RMRS, 1221, South Main Street, 83843 Moscow, Idaho, USA

²Department of Natural Resources and Society, College of Natural Resources, University of Idaho/UI, 875 Perimeter Drive, 83843 Moscow, Idaho, USA

³Departamento de Engenharia Florestal, Universidade do Estado de Santa Catarina/ UDESC, Avenida Luiz de Camões, 2090, 88520-000 Lages, SC, Brazil

⁴Klabin SA, Av. Araucária, s/n, 84260-001 Telêmaco Borba, PR, Brazil

Manuscript received on February 12, 2016; accepted for publication on April 29, 2016

ABSTRACT

Accurate forest inventory is of great economic importance to optimize the entire supply chain management in pulp and paper companies. The aim of this study was to estimate stand dominant and mean heights (HD and HM) and tree density (TD) of *Pinus taeda* plantations located in South Brazil using *in-situ* measurements, airborne Light Detection and Ranging (LiDAR) data and the non-*k*-nearest neighbor (*k*-NN) imputation. Forest inventory attributes and LiDAR derived metrics were calculated at 53 regular sample plots and we used imputation models to retrieve the forest attributes at plot and landscape-levels. The best LiDAR-derived metrics to predict HD, HM and TD were H99TH, HSD, SKE and HMIN. The Imputation model using the selected metrics was more effective for retrieving height than tree density. The model coefficients of determination ($\text{adj.}R^2$) and a root mean squared difference (RMSD) for HD, HM and TD were 0.90, 0.94, 0.38m and 6.99, 5.70, 12.92%, respectively. Our results show that LiDAR and *k*-NN imputation can be used to predict stand heights with high accuracy in *Pinus taeda*. However, further studies need to be realized to improve the accuracy prediction of TD and to evaluate and compare the cost of acquisition and processing of LiDAR data against the conventional inventory procedures.

Key words: forest inventory, LiDAR metrics, *k*-NN Imputation, Remote Sensing.

INTRODUCTION

Light Detection and Ranging (LiDAR) is an optical remote sensing technology which measures properties of scattered light to find range and/or other information of a distant target. LiDAR

measurements is usually acquired at airborne level and is usually also referred to Airborne laser scanning (ALS). It has been widely used in mapping the Earth's surface and especially in forest applications due its capacity to retrieve three-dimensional information of the vegetation (Yao et al. 2014).

Correspondence to: Carlos Alberto Silva
E-mail: csilva@uidaho.edu

The resulted cloud points acquired by a LiDAR system allows the reconstruction of the vertical structure of forests at different strata that cannot be obtained from any other remote sensing system. Indeed, several LiDAR-derived metrics can be derived out from these point clouds (McGaughey 2015). These metrics can be used indirectly to predict several other parameters of the forests using either regression or classification approaches to spatially represent these selected attribute over large areas (Dubayah and Drake 2000). Applications and advantages of LiDAR technology can be seeing over both natural and planted forest worldwide (Nilsson 1996, Næsset 1997, Næsset and Bjercknes 2001, Popescu et al. 2003, Popescu 2007).

Pine plantations are the most important long fiber source for pulp and paper production in South Brazil. It covers nowadays an area of 1.59 million hectares, accounting for approximately 20.54% of the country's total reforested area (Ibá 2015). Most of the pine plantations are concentrated in South Brazil, with 34.1 and 42.4% of the total reforested area located in Paraná and Santa Catarina States (Ibá 2015). *Pinus taeda* L., also known as loblolly pine, is the most planted forest specie in these regions. It has high economic importance due to its high volumetric increment in the colder regions of the southern plateau (Kohler et al. 2014). It has fast growing rates presenting increments up to $50 \text{ m}^3 \cdot \text{ha}^{-1} \cdot \text{year}^{-1}$ (Ibá 2015).

Usually, forest companies conduct their own permanent forest inventory in planted forest at annual basis in order to monitor the forest growth, to identify problematic conditions during initial growth stages, and to predict the optimal harvesting time. Important attributes describing the state of the current stand that affect thinning decisions includes both stand height and tree density. On the other hand, stand height is a useful information wefor both dominant heights and site index determination. Whereas on the other hand, tree density is a key information for thinning management. Initially,

the seedlings are planted using a higher density configuration and afterwards thinning is conducted in order to proportionate primary and secondary growths, respectively.

Therefore, accurate forest inventory is of great economic importance to optimize the entire supply chain management in pulp and paper companies (Falkowski et al. 2008, Silva et al. 2016a). Although the use of LiDAR technology is relatively new in Brazil, different investigations have been made to evaluate the ability of airborne LiDAR for forest inventories studies. They were mostly restricted to natural forest environments and/or *Eucalyptus* spp. plantations (e.g., Packalén et al. 2011, Carvalho et al. 2015, Silva et al. 2015, 2017a). As far as we know, Zandoná et al. (2008) were the first and only authors to use airborne LiDAR data for forest inventory purposes in a *Pinus taeda* plantation located in South Brazil. In their study, they derived individual tree height and crown area measurements using an individual tree approach to model the diameter at the breast height (DBH). However, early studies using both individual and plot based approaches also have demonstrated the potential of airborne LiDAR data to quantify forest attributes in conifer and deciduous forests in Northern Hemisphere (e.g., Nilsson 1996, Næsset 1997, Næsset and Bjercknes 2001, Popescu et al. 2003, Roberts et al. 2005, Popescu 2007). Therefore, a better understanding of how airborne LiDAR data can be used to extract forest attributes from planted forest of *Pinus taeda* in South Brazil is still necessary.

Currently, the most common approach to retrieve forest attributes from LiDAR data has been the area-based approach in which forest properties for an area of interest are inferred based on the relationship between field measurements and the LiDAR-derived metrics (Nilsson 1996, Næsset 1997, Hudak et al. 2006, Silva et al. 2017b). These relationships are usually linear multiple regression models, which afterwards, can be applied to the

entire LiDAR dataset for predicting forest attributes at stand level.

One emerging method for forest attribute modeling from LiDAR data is the k -nearest neighbor (k -NN) imputation. It is a nonparametric machine learning method which has widely been used to predict forest inventory attributes in un-inventoried areas at either plot- or stand-levels (Falkowski et al. 2010, Hudak et al. 2014, Racine et al. 2014, Silva et al. 2016b). The non-parametric k -NN imputation method uses a set of predictor feature variables (X) to match each target pixel to a number (k) of most similar (nearest neighbors or NN) reference pixels for which values of response variables (Y) are known (McRoberts 2012). It allows estimating a variable of interest through a weighted average of the known variables of the k -nearest neighbouring plots. The weighted average could be done using either an inverse distance weighting or the square of the inverse distance (McRoberts 2009). Examples of forest applications using k -NN imputation, including mathematical formulation may be found in Hudak et al. (2008), Racine et al. (2014), and Fekety et al. (2014).

Therefore, the aim of this study was to evaluate the use of airborne LiDAR data and k -NN imputation for predicting and mapping stand-level forest variables, which included mean height (henceforth HM), dominant height (HDOM) and stem density (TD). This work is based on the hypothesis that LiDAR data and k -NN analysis can produce precise and accurate inferences of HM, HDOM and TD in *Pinus taeda* plantations in South Brazil.

MATERIALS AND METHODS

STUDY AREA DESCRIPTION

The study area consisted of four *Pinus taeda* stands located within the Telêmaco Borba municipality in the State of Paraná, Southern Region of Brazil (Fig. 1). The trees were planted using a 3.0 x 2.0

m or 2.5 x 2.5 m grid configuration, resulting in an average tree density of 1,667 and 2,000 trees per ha, respectively.

The climate of the region is characterized as warm and temperate (Cfa) (Köppen and Geiger 1928), with annual average precipitation approximately of 1378 mm and the annual average temperature of 18.4 °C. The topography in the study ranges from mildly to very hilly with elevation ranging of 618 m to 905 m. The *Pinus taeda* stands are on the plateau where the topography is relatively flat. The plantations are managed by Klabin S.A., a pulp and paper company.

FIELD DATA COLLECTION

A total of 53 rectangular plots of approximately 500 - 620 m² each were established across the four plantations for stand measurement ranging in ages from three to nine years old. All plots were georeferenced with a geodetic GPS with differential correction capability (Trimble Pro-XR), and in each sample plot, individual trees were measured for DBH (diameter at breast height) and a subsample (15%) of trees for tree height (Ht). For trees in the plot that were not directly measured for Ht, the inventory team of Klabin S.A. predict heights from hypsometric equations. The hypsometric models were adjusted according to Curtis (1967), employing as independent variables DBH, and as dependent variables the Ht, following the model below (Eq. 1).

$$\ln(Ht) = \beta_0 + \beta * \left(\frac{1}{DBH}\right)^{+c} \quad (\text{Eq.1})$$

where: $\ln(Ht)$ is the logarithm of tree height (m); and β_0 and β are the intercept and the slope of the model; DBH is the diameter at breast height (1.30 m) and ϵ is the random error of the model. The coefficients of the hypsometric models were not available, however, according to the company, the adj.R² and Syx% of the models ranged from 0.96 to 0.98

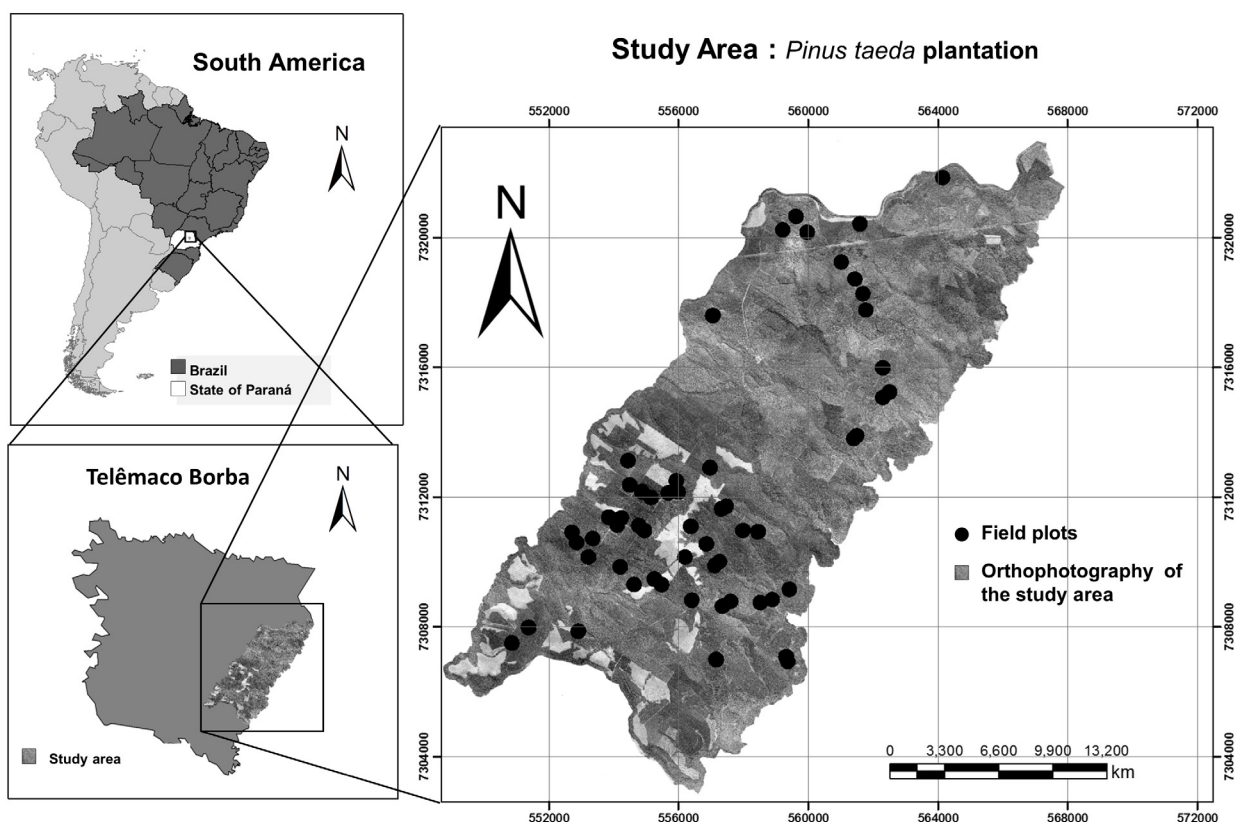


Figure 1 - Location of study area. The circles indicate the location of the *Pinus taeda* plantations.

and 5.1 to 6.5, respectively, according to the farm location.

The mean height (HM; m), dominant height (HDOM; m) and tree density (TD; N/ha) were summarized for each sample plot, and the summary statistics of HM, HDOM, and TD per stand ages are presented in Table I.

LIDAR DATA ACQUISITION AND DATA PROCESSING

LiDAR data were obtained by a o Harrier 68i sensor mounted on a CESSNA 206 aircraft. The characteristics and precision of the LiDAR data are listed in Table II. LiDAR data processing consisted of several steps that ingested the LiDAR point cloud data and provided two major outputs: the digital terrain model (DTM), and the LiDAR-derived canopy structure metrics. All of the

LiDAR processing phase was performed using the US Forest Service FUSION/LDV 3.42 software (McGaughey 2015).

Initially the *catalog* function in FUSION/LDV was used to generated a descriptive report of the point cloud, which was used to identify the LiDAR coverage extent and outliers. When detected

TABLE I
 Characteristics of the *Pinus taeda* plantations. The values are based on *in-situ* measured sample plots (n=53). Standard deviation values are given in brackets.

Ages (I)	<i>Pinus taeda</i>			Plot (N)
	HM (m)	HDOM (m)	TD (N/ha)	
3 ≤ I < 5	6.72[0.29]	7.59[0.59]	1298[86]	19
5 ≤ I < 7	9.91[0.73]	11.05[1.10]	1361[159]	18
7 ≤ I < 9	11.61[0.79]	12.87[1.05]	1266[243]	16
Average	9.28[2.14]	10.36[2.45]	1413[215]	---

TABLE II
Airborne LiDAR system characteristics.

Parameter	Value
Scan angle (°)	60°
Footprint (m)	0,33m
Flight speed (km/h)	234,00 km h ⁻¹
Horizontal accuracy	10cm
Elevation accuracy	15cm
Operating altitude	666,17 m
Scan frequency	300 kHz
Pulse density	4 pulse m ⁻²

IMU=Inertial measurement unit; GPS=Global Positioning System.

outliers, we removed them using the ClipData tool. A filtering algorithm based on Kraus and Pfeifer (1998) and available in the Groundfilter function was applied to differentiate between ground and vegetation points. DTMs were generated using the classified ground points with a spatial resolution of one meter using GridSurfaceCreate. Afterwards, the ClipData function was applied to normalize heights and to assure that the z coordinate for each point corresponded to the height above ground and not the orthometric elevation of the single point. The PolyClipdata function was then used to subset of the LiDAR points within each of the 53 measured sample plots, and the Cloudmetrics tool was applied to compute the LiDAR-derived metrics using all the available returns of the point cloud. Finally, the GridMetrics function were used to generate the same LiDAR metrics computed in CloudMetrics, but now within grid cells of 5 m spatial resolution, across the landscape. The list of the LiDAR metrics used in this are presented in the Table III.

LIDAR METRICS SELECTION AND IMPUTATION MODELING DEVELOPMENT

The LiDAR metrics selection was performed through two major steps. First, Pearson's correlation (r) was used to identify highly correlated predictor variables ($r > 0.9$) as presented in Hudak et al.

TABLE III
LiDAR-derived canopy height metrics considered as potential candidate variables for predictive imputation models (McGaughey 2015).

Variable	Description
HMIN	Height Minimum
HMAX	Height Maximum
HMEAN	Height Mean
HMAD	Height median absolute deviation
HSD	Height standard deviation
HSKE	Height skewness
HKURT	Height kurtosis
HCV	Height coefficient of variation
HMODE	Height mode
H01TH	Height 1 st percentile
H05TH	Height 5 th percentile
H10TH	Height 10 th percentile
H15TH	Height 15 th percentile
H20TH	Height 20 th percentile
H25TH	Height 25 th percentile
H30TH	Height 30 th percentile
H35TH	Height 35 th percentile
H40TH	Height 40 th percentile
H45TH	Height 45 th percentile
H50TH	Height 50 th percentile
H55TH	Height 55 th percentile
H60TH	Height 60 th percentile
H65TH	Height 65 th percentile
H70TH	Height 70 th percentile
H75TH	Height 75 th percentile
H80TH	Height 80 th percentile
H90TH	Height 90 th percentile
H95TH	Height 95 th percentile
H99TH	Height 99 th percentile
COV	Canopy Cover (Percentage of first return above 1.30m)

(2012), Silva et al. (2014, 2017c). If a given group (2 or more) of LiDAR metrics were highly correlated, we retained only one metric by excluding the others that were most highly correlated with the remaining metrics. Second, the *varSelection* function from the *yaImpute* (Crookston and Finley 2008) package in the R statistical software (R Development Core Team 2015) was used to select the most important not highly correlated metrics for the imputation modeling. The *varSelection* function computes the generalized root mean square distance (grmsd) as variables are added to an *k*-NN imputation model. By adding variables, the *varSelection* function keeps variables that strengthen imputation and exclude variables that weaken the imputation the least (Crookston and Finley 2008).

In this study, *k*-NN imputation was conducted using the *yai* and *impute.yai*, both also from in the *yaImpute* package in the R. Many imputation methods can be used for associating target and reference observations, however, recent studies have shown that the Random Forest (Breiman 2001) approach generally produces better results compared to other imputation methods (Hudak et al. 2008, Nelson et al. 2011, Waske et al. 2012). Therefore, for this study we used Random Forest based *k*-nearest neighbours (RF-*k*NN) to characterize the relationships between predictor (LiDAR-derived metrics) and response (HM, HDOM and STP) variables used for NN (*k*=1) imputation.

MODEL ASSESSMENT

The precision of the model predictions was evaluated in terms of adjusted coefficient of determination (adj.R²), Root Mean Square Difference (RMSD) and BIAS (both absolute and relative):

$$RMSD = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (\text{Eq.2})$$

$$Bias = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i) \quad (\text{Eq.3})$$

where *n* is the number of plots, *y_i* is the observed value for plot *i*, and *ŷ_i* is the predicted value for plot *i*. The RMSD is analogous to the RMSE used to assess regression model accuracy (Stage and Crookston 2007). The relative RMSD (%) and Bias (%) were calculated by dividing the absolute values (Eqs. 2 and 3) by the mean of the response variable and multiplied by 100. Ninety-five percent confidence intervals (CI) for the line of best fit were also plotted to determine whether the relationship between observed and imputed values was significantly different. The acceptable model precision was set up a relative RMSD and Bias of ≤ 15% to have a model precision equal or higher to the conventional forest inventory in the *P. taeda* plantations in Brazil. We also performed a cross validation on the RF *k*-NN imputation model with bootstrapping simulations using 1000 iterations. For each iteration we built an imputation model from randomly selected of 53 observations with replacement. The remaining observations (out of bag) were used to validate the model and to compute the RMSD and Bias. After all iterations, the average and stand deviation of the RMSD and Bias were computed.

Finally, we used the *AsciiGridPredict* function also from the *yaImpute* package in R (Crookston and Finley 2008) to apply the model across to the landscape to map the spatial distribution of HM, HDOM and SD of pine at the stand level for the benefit of forest managers. An overview of the methodology is outlined in Fig. 2.

RESULTS

LIDAR-DERIVED CANOPY HEIGHT STRUCTURE OF THE *Pinus taeda* PLANTATIONS

Young forest stands of *Pinus taeda* are normally defined by a lower canopy coverage, high tree

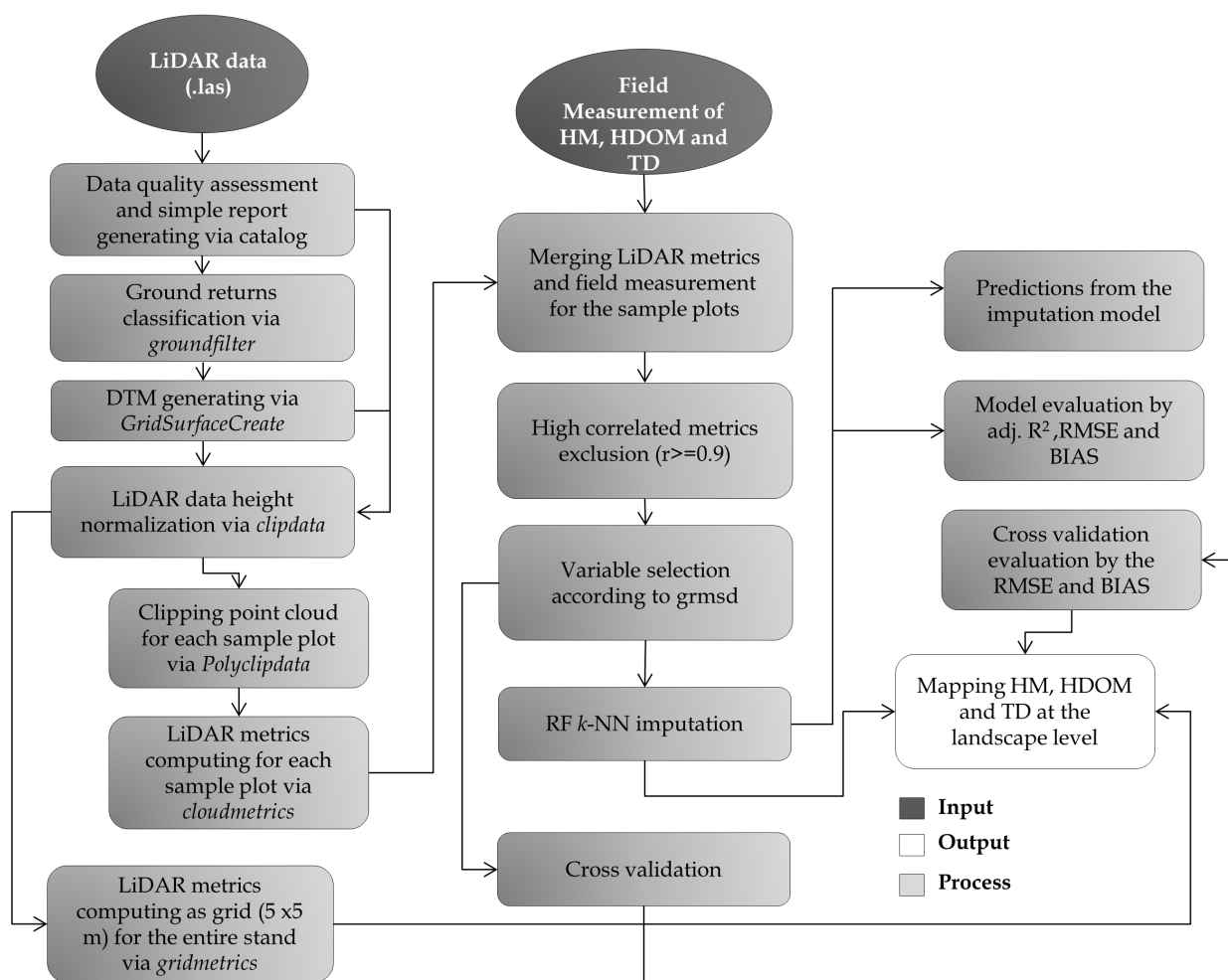


Figure 2 - Flowchart of the LiDAR data processing and forest attributes modeling.

density and low height values. By reaching maturity, intermediate and advanced stands, present higher canopy coverage values, higher height values and lower tree density when compared with young stands due to forest thinning and mortality. LiDAR derived metrics such as H99TH was high correlated ($r > 0.9$) with field heights (HM, HDOM). An example of variations in height ranging from early (i.e., 4.0 years) to intermediate (i.e., 6.0 and 8.0 years) ages of development are shown in Fig. 3. Although located in different farms and therefore under distinct site index conditions, the LiDAR derived height increased from early (Fig. 3a) to intermediate ages (Fig. 3b, c).

LIDAR METRICS SELECTION

Pearson's correlation test (r) applied to the 30 candidate LiDAR metrics determined that 24 were highly correlated each other ($r > 0.9$). We kept one of the highly correlates metrics (i.e., H99TH), and the six remaining metrics not highly correlated to each other were also used as input in the *VarSelection* function. Some LiDAR-derived metrics were positively correlated, such as H99TH and HSD ($r = 0.81$), whereas both HSKE and COV ($r = -0.67$) were negatively correlated as shown in Table IV.

The most important metrics selected from the *VarSelection* to be included in the imputation

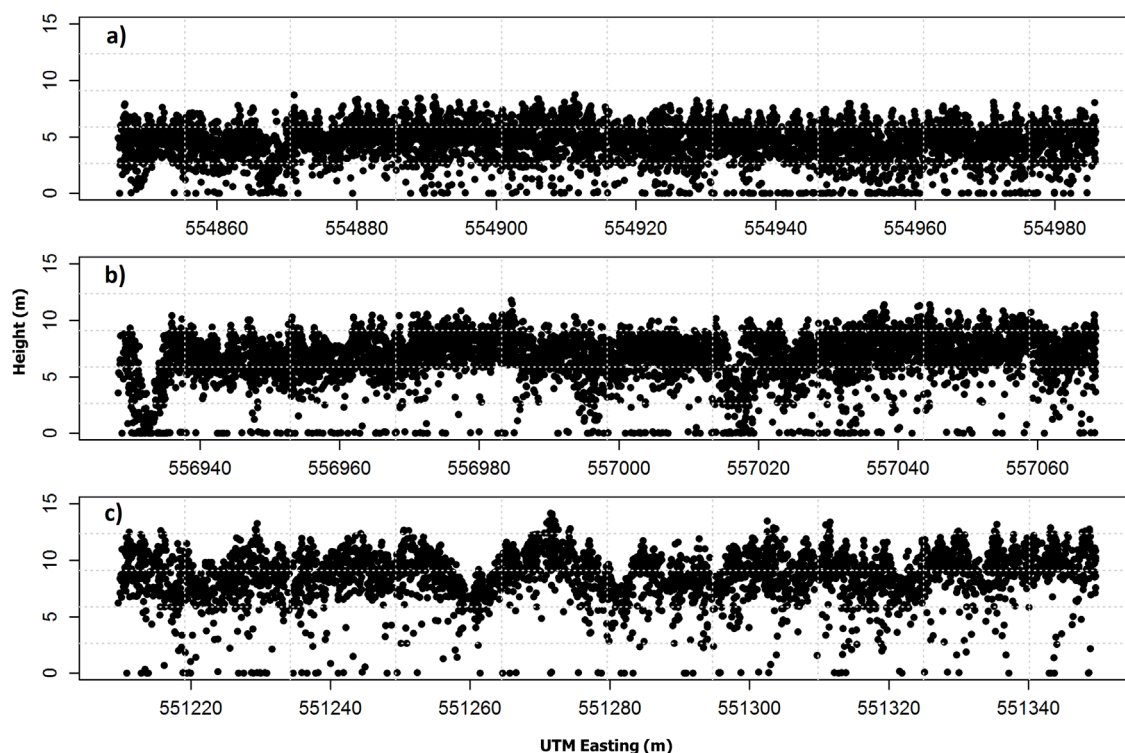


Figure 3 - LiDAR vertical profiles of three randomized transects (1.5 x 140 m) of *Pinus taeda* plantations. The transects represent stands with 4.0 years (a); 6.0 years (b) and 8.0 years (c).

TABLE IV
Pearson's correlation coefficients among the selected LiDAR-derived metrics.

r	HMIN	HSD	HSKEW	HKUR	H99TH	COV
HMIN	1.00					
HSD	0.21	1.00				
HSKEW	-0.31	-0.61	1.00			
HKUR	0.23	0.31	-0.88	1.00		
H99TH	0.37	0.81	-0.87	0.75	1.00	
COV	0.23	0.31	-0.67	0.54	0.57	1.00

model as predictor variables for estimating HM, HDOM and TD according to the gmsd statistics were H99TH, HSD, SKE and HMIN (Fig. 4). When added the HKUR and COV metrics to the model, they did not improve the fit, and therefore, they were automatically removed by the algorithm.

IMPUTATION MODEL PRECISION AND BIAS

The LiDAR-derived metrics were better predictors of HM and HDOM as compared to TD. The imputed

model produced a relative RMSD of 6.99, 5.70 and 12.92%; relative Bias of -0.36, -1.02 and -1.00, and adj.R² of 0.90, 0.94 and 0.61 for the HM, HDOM, and TD attributes respectively (Table V). With the exception of TD, the RMSD and Bias were less than 10% in the cross validation procedure. The adj.R² of the cross validation for the imputation model of TD was markedly lower at 0.40 (± 0.17).

The Figs. 5, 6, and 7 show the relationship between the observed and imputed values of the

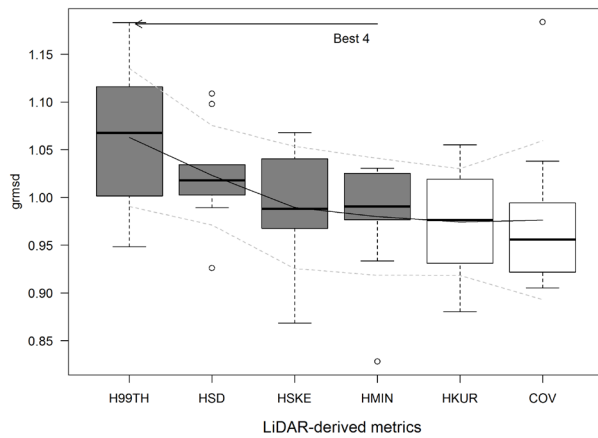


Figure 4 - LiDAR-derived metrics selection according to grmsd (generalized root mean square distance). The gray boxplots represent the four best LiDAR-derived metrics for the imputation modeling. The solid black and the dotted gray lines represent the mean and standard deviation of the grmsd.

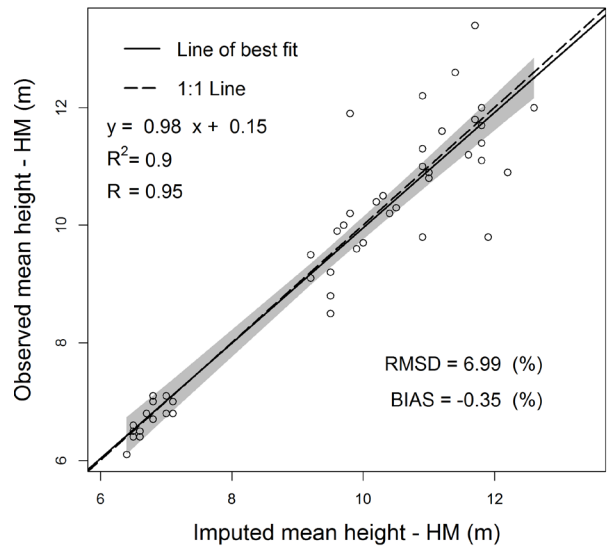


Figure 5 - Imputed versus Observed HM (m); (n=53). Solid line depicts regression fit; dashed line is the 1:1 line. Shaded areas represent 95% regression confidence intervals.

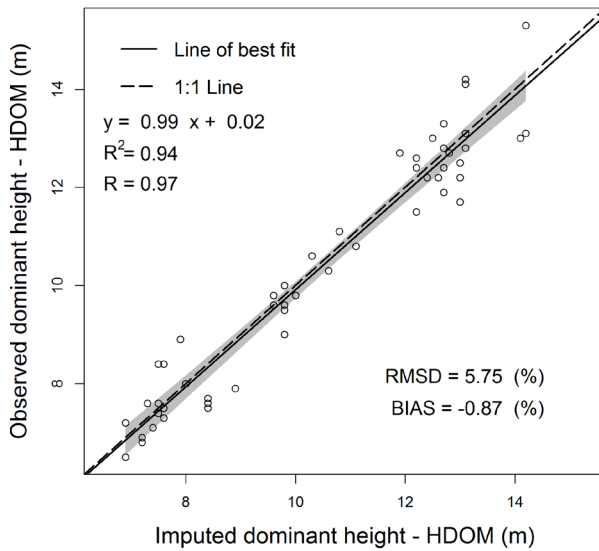


Figure 6 - Imputed versus Observed HDOM (m); (n=53). Solid line depicts regression fit; dashed line is the 1:1 line. Shaded areas represent 95% regression confidence intervals.

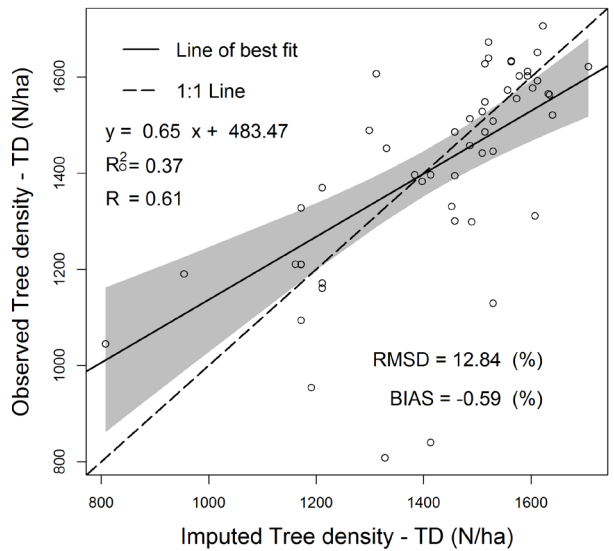


Figure 7 - Imputed versus Observed TD (N/ha); (n=53). Solid line depicts regression fit; dashed line is the 1:1 line. Shaded areas represent 95% regression confidence intervals.

forest attributes studied herein. The regression intercepts differed significantly from zero ($p \geq 0.7$), and the 1:1 line fell within the 95% regression confidence envelope for all regressions, with the exception of the TD regression.

According to the cross validation performance, the LiDAR metrics were also better predictors for

HM and HDOM rather than TD (Table VI). The HM and HDOM imputation models produced estimates that were strongly correlated ($r > 0.93$) with the validation dataset, whereas the TD imputation model were weakly correlated ($r = 0.62$). This resulted consequently in high values of RMSD and Bias, respectively. On the other hand, for both

HM and HDOM, the $\text{adj.}R^2$ values were high and RMSD and Bias values were low (Table VI).

IMPUTED FOREST ATTRIBUTES

Imputed HM, HDOM and TD of *Pinus taeda* for the 53 sample plots ranged from 6.1 to 13.4 m; 8.9 to 15.5 m and 1266 to 1585 N/ha, respectively. In general, imputed values are overestimated compared to the reference forest attribute values. Forest stands containing younger trees (3 to 5 years) showed the highest TD values, while the plantations with intermediated stands ages (6-9 years) contained the lowest TD values (Table VII).

The k -NN imputed model created was applied across the extent of the study area to map the forest attributes at the landscape level. The Fig. 8 is showing the spatial distribution of the imputed HM, HDOM and TD values at three stands with ages ranging from 4 to 8 years old. The stand "A" with the age of four years and area of 14.4 ha (Fig. 8a1-a3) has imputed average values of 7.18 (± 0.35) m, 8.02 (± 0.32) m, and 1574 (± 27) N/ha for HM,

HDOM and TD, respectively; and the stand "B" with the age of six years and area of 8.5 ha (Fig. 8b1-b3) imputed average values of 10.80 (± 0.56) m; 11.78 (± 0.75) m, and 1260 (± 93) N/ha for HM, HDOM and TD, respectively; while the stand "C" with the age of eight years and area of 15.4 ha (Fig. 8c1-c3), presented imputed average values of 11.44 (± 0.66) m; 12.66 (± 0.84) m, and 1253 (± 88) N/ha for HM, HDOM and TD; respectively. The predicted total number of tree in the stands a), b) and c) were approximately 22,651; 10,710; 19,296 trees, respectively. Differences in HM, HDOM and TD are mainly related to the ages, site index and management practices of each stand. Moreover, differences also depend on other factors such as the soil type and intensity of land use before the establishment of the stands.

DISCUSSION

LiDAR is increasingly being used as a means for obtaining forest structural measurements over large areas, and LiDAR-derived metrics can be used to

TABLE V

Adjusted coefficients of determination ($\text{adj.}R^2$), pearson correlation (r), root mean square difference (RMSD) and BIAS of both observed and predicted forest attributes values (HM, HDOM and TD) from the imputation model.

Model	LiDAR derived metrics	Adj. R^2	r	RMSD		BIAS	
				m; N·ha ⁻¹	%	m; N·ha ⁻¹	%
HM		0.90	0.95	0.65	6.99	-0.03	-0.36
HDOM	H99TH + HSD + HSK + HMIN	0.94	0.97	0.59	5.70	-0.11	-1.02
TD		0.38	0.61	183	12.92	-14.11	-1.00

TABLE VI

Cross validation statistics. Number of bootstrapping = 1000. The values represent the average, and the values inside of the brackets represent the standard deviation.

Model	Adj. R^2	R	RMSD		BIAS	
			m; N·ha ⁻¹	%	m; N·ha ⁻¹	%
HM	0.87[0.53]	0.93[0.03]	0.79[0.18]	8.55[1.80]	0.05[0.22]	0.54[2.42]
HDOM	0.83[0.06]	0.93[0.03]	0.93[0.19]	9.05[1.82]	-0.01[0.27]	-0.18[2.60]
TD	0.40[0.17]	0.62[0.14]	186.33[34.04]	13.19[2.59]	-10.99[47.86]	-0.82[3.37]

TABLE VII
Summary of the statistics for the observed and imputed forest attributes at different age intervals.

Age	Attributes		Statistics (m; and N·ha ⁻¹)			
			MIN	MAX	MEAN	SD
3-5	HM	Observed	6.4	7.1	6.76	0.25
		Imputed	6.1	7.1	6.72	0.29
	HDOM	Observed	6.9	8.9	7.69	0.54
		Imputed	6.5	8.9	7.59	0.59
	TD	Observed	1299	1707	1574	83
		Imputed	1299	1707	1585	86
5-7	HM	Observed	9.2	12.2	10.3	0.93
		Imputed	8.5	11.1	9.91	0.73
	HDOM	Observed	9.6	13.1	11.16	1.37
		Imputed	9	13.1	11.05	1.4
	TD	Observed	840	1607	1377	210
		Imputed	954	1607	1361	160
7-9	HM	Observed	9.7	12.6	11.24	0.74
		Imputed	10	13.4	11.61	0.79
	HDOM	Observed	11.1	14.2	12.94	0.81
		Imputed	10.8	15.3	12.88	1.06
	TD	Observed	808	1514	1289	188
		Imputed	808	1628	1266	243

predict forest attributes using imputation modeling (Fig. 8). A proper data mining is required in order to select the variables that have a physical mean or relationship with the desired forest attributes. In this investigation, LiDAR-derived metrics such as H99TH, HSD, HSKE and HMIN were found to be the most relevant variables to predict HM, HDOM and TD. These variables were the most efficiently to describe the canopy structure of the forest stands by capturing the majority of variation contained in the reference dataset. Moreover, these LiDAR metrics have been frequently reported as predictor variable for forest attributes modeling from LiDAR data (Hudak et al. 2012, Silva et al. 2014).

LiDAR is an important tool for forest inventory when combined with appropriately designed sample plots in the field and with appropriate modeling tools (Hudak et al. 2008, Silva et al. 2017d). Since this is most probably the first time that *k*-NN has been applied to predict forest attributes from airborne

LiDAR data in Brazil, the obtained results strongly corroborates with those presented by Roberts et al. (2005), Hudak et al. (2008, 2014), Falkowski et al. (2009, 2010), Fekety et al. (2014), and Gagliasso et al. (2014) in loblolly pine environments.

The main advantage of imputation modeling is the simultaneous coherent prediction of multiple response variables simultaneously (Crookston and Finley 2008). This is a significant asset in forest inventory where typically multiple forest attributes, in particular the HM, HDOM and TD, are of interest.

In this study, the prediction of TD based on LiDAR-derived predictors has proven to be more difficult than predicting HM and HDOM. The lower performance to retrieve this parameter over planted forest is due to the lower variability among forest stands even under different ages. The difficulties in predicting TD has been also highlighted in other studies (Maltamo et al. 2004, Woods et al.

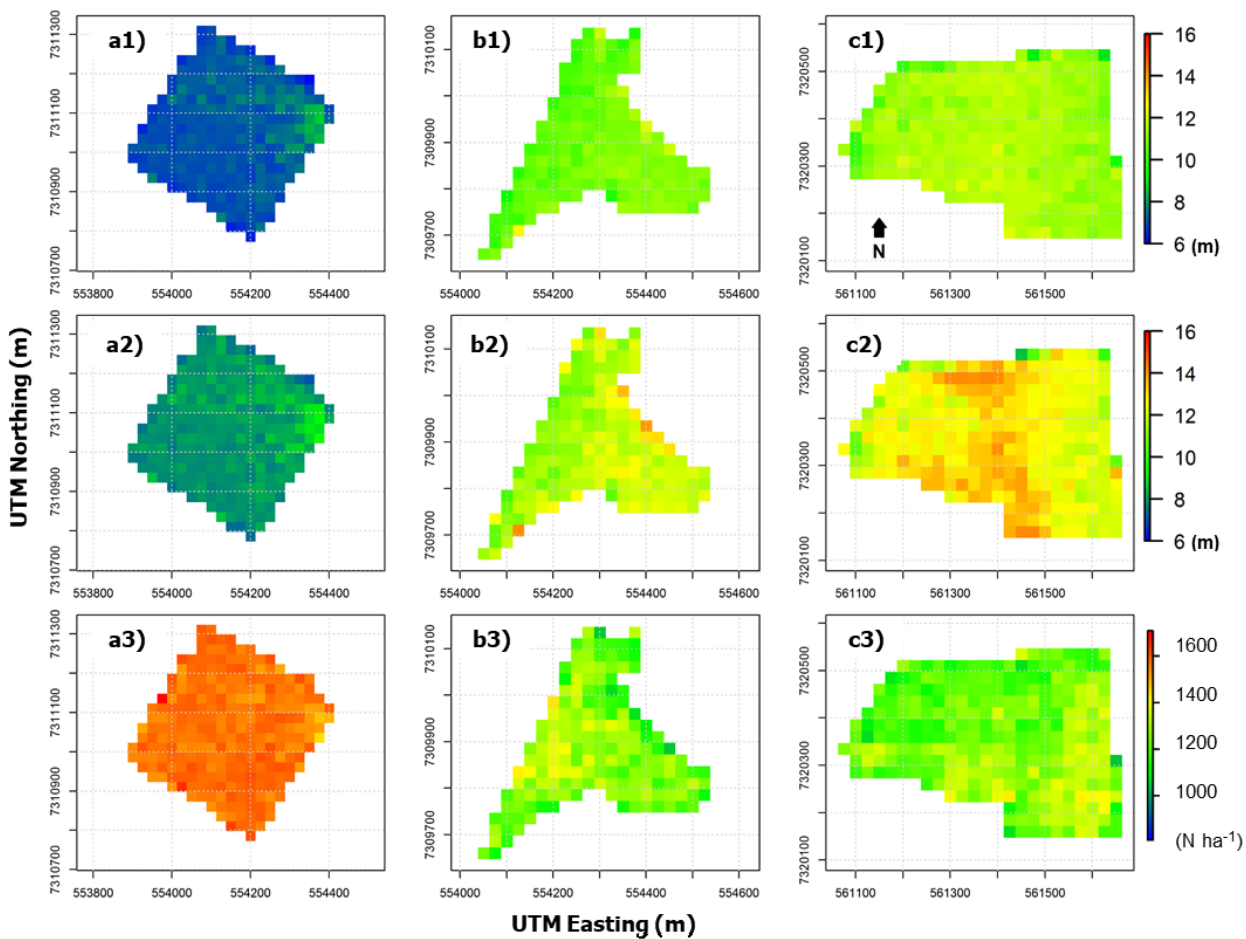


Figure 8 - Imputed forest attributes at landscape level. Maps are with 25 meters of spatial resolution. Examples of *Pinus taeda* stands with 4.0 years (a); 6.0 years (b); and 8.0 years (c). The numbers 1, 2 and 3 represent in order HM, HDOM and TD, respectively.

2008, Hirata et al. 2009). The main reason for this discrepancy are most probably because the LiDAR data distribution accounts more for vertical profile what confirms the need of other methodologies for predicting TD rather than plot-area based approach in this type of environment. The methodology already demonstrated by Zandoná et al. (2008) or more recently by Strimbu and Strimbu (2015) based directly on point cloud segmentation have shown that the individual tree based approach provides accurate estimates of tree density in pine plantations. Therefore, future research is required to evaluate new approaches using individual tree detection.

Although showing great potential for forest inventory procedures, the use of airborne LiDAR in Brazil is still considered experimental due to its expensive cost (Hummel et al. 2011). This is because the number of companies is low, when compared with other countries where this technology is already being used intensively. LiDAR has the advantage of capturing variability across forest stands and mapping forest attributes at the landscape level with high accuracy (Silva et al. 2017a).

In *in-situ* conventional forest inventory procedures, the variability of the forest attributes at stand level demands time and is costly, and is

not always perceptive and therefore less studied. LiDAR demonstrated to be an established technology that is increasingly and being used to characterize spatial variation (Watt et al. 2014). In addition, LiDAR permits an accurate production of digital elevation models, such as DTM and digital surface model (DSM) that are used for hydrological, geomorphological, and other applications (Hudak et al. 2009). In particular, LiDAR-derived DTM has attracted great interest to support forestry operations industrial plantation, like in Brazil, since it provides thorough and detailed information about terrain topography, which in turn is used to road planning and to choose the best skidding system in complex forest areas (Sterenczak and Moskalik 2014).

CONCLUSIONS

This paper describes a framework for predicting forest inventory data in unsampled areas via an RF-*k*-NN imputation modeling procedure. This procedure incorporates airborne LiDAR-derived predictor metrics. Airborne LiDAR has proven to be an accurate method to spatially estimate stand heights at regular basis. The tree density was not totally explained by the LiDAR-derived metrics into an imputation model in the area-based approach presented herein, and therefore, future research will be carried out to test new approaches using individual tree framework. In many countries, LiDAR has already moved from the research arena to operational reality in the area of forest resource inventory. We hope that the promising results for forest inventory modeling presented herein will stimulate to operational this technology in Brazil.

ACKNOWLEDGMENTS

This research was funded through a PhD scholarship from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) via the Science Without Borders Program (Process 249802/2013-

9). The authors are very grateful for the LiDAR and field inventory data collections funded by Klabin S.A., a pulp and paper company. We thank Vincent Jansen and the anonymous reviewers for their helpful suggestions on an earlier draft of the manuscript.

REFERENCES

- BREIMAN L. 2001. Random Forest. *Mach Learn* 45(1): 5-32.
- CARVALHO SPC, RODRIGUEZ LCE, SILVA LD, CARVALHO LMT, CALEGARIO N, LIMA MP, SILVA CA, MENDONÇA AR AND NICOLETTI MF. 2015. Predição do volume de árvores integrando LiDAR e Geoestatística. *Scien For* 43(107): 627-637.
- CROOKSTON NL AND FINLEY AO. 2008. yaImpute: An R Package for kNN Imputation. *J Stat Soft* 23(10): 1-16.
- CURTIS RO. 1967. Height-diameter and height-diameter-age equations for second-growth Douglas-fir. *For Sci* 13: 365-375.
- DUBAYAH RO AND DRAKE JB. 2000. LiDAR Remote Sensing for Forestry. *J For* 98(6): 44-46.
- FALKOWSKI MJ, EVANS JS, MARTINUZZI S, GESSLER PE AND HUDAK AT. 2009. Characterizing forest succession with LiDAR data: An evaluation for the Inland Northwest, USA. *Remote Sens Env* 113(5): 946-956.
- FALKOWSKI MJ, HUDAK AT, CROOKSTON N, GESSLER PE AND SMITH AMS. 2010. Landscape-scale parameterization of a tree-level forest growth model: a *k*-NN imputation approach incorporating LiDAR data. *Can J For Res* 40: 184-199.
- FALKOWSKI MJ, SMITH AMS, GESSLER PE, HUDAK AT, VIERLING LA AND EVANS JS. 2008. The influence of conifer forest canopy cover on the accuracy of two individual tree measurement algorithms using LiDAR data. *Can J Remote Sens* 34(2): 338-350.
- FEKETY PA, FALKOWSKI MJ AND HUDAK AT. 2014. Temporal transferability of LiDAR-based imputation of forest inventory attributes. *Can J For Res* 45: 422-435.
- FERRAZ A, SAATCHI S, MALLET C, JACQUEMOUD S, GONÇALVES G, SILVA C, SOARES P, TOMÉ M AND PEREIRA L. Airborne LiDAR estimation of aboveground forest biomass in the absence of field inventory. *Remote Sens* 8: 653-670.
- GAGLIASSO D, HUMMEL S AND TEMESGEN H. 2014. A comparison of selected parametric and non-parametric imputation methods for estimating forest biomass and basal area. *Open J For* 4(1): 42-48.
- HIRATA T, FURUYA N, SUZUKI M AND YAMAMOTO H. 2009. Airborne laser scanning in forest management: individual tree identification and laser pulse penetration in

- a stand with different levels of thinning. For *Ecol Manage* 258(5): 752-760.
- HUDAK AT, BRIGHT BC, POKSWINSKI SM, LOUDERMILK EL, O'BRIEN JJ, HORNSBY BS, KLAUBERG C AND SILVA CA. 2016. Mapping forest structure and composition from low-density LiDAR for informed forest, fuel, and fire management at Eglin Air Force Base, Florida, USA. *Can J Remote Sens* 42: 411-427.
- HUDAK AT, CROOKSTON NL, EVANS JS, FALKOWSKI MJ, SMITH AMS AND GESSLER PE. 2006. Regression modeling and mapping of coniferous forest basal area and tree density from discrete-return LiDAR and multispectral satellite data. *Can J Remote Sens* 32: 126-138.
- HUDAK AT, CROOKSTON NL, EVANS JS, HALL DE AND FALKOWSKI MJ. 2008. Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. *Remote Sens Environ* 112: 2232-2245.
- HUDAK AT, EVANS JS AND SMITH AMS. 2009. Review: LiDAR utility for natural resource managers. *Remote Sens* 1: 934-951.
- HUDAK AT, HAREN AT, CROOKSTON NL, LIEBERMANN RJ AND OHMANN JL. 2014. Imputing Forest Structure Attributes from Stand Inventory and Remotely Sensed Data in Western Oregon, USA. *For Sci* 60(2): 253-269.
- HUDAK AT, STRAND EK, VIERLING LA, BYRNE JC, EITEL JUH, MARTINUZZI S AND FALKOWSKI MJ. 2012. Quantifying aboveground forest carbon pools and fluxes from repeat LiDAR surveys. *Remote Sens Environ* 123: 25-40.
- HUMMEL S, HUDAK AT, UEHLER EH, FALKOWSKI MJ AND MEGOWN KA. 2011. A comparison of accuracy and cost of LiDAR versus stand exam data for landscape management on the Malheur National Forest. *J For* 109(5): 267-273.
- IBÁ. 2015. Brazilian Tree Industry. Brasília: IBÁ, 80 p. http://www.iba.org/images/shar ed/iba_2015.pdf/.
- KÖPPEN W AND GEIGER R. 1928. *Klimate der Erde*. Gotha: Verlag Justus Perthes. Wall-map 150cm×200cm.
- KOHLER SV, WOLFF N I, FILHO AF AND ARCE JE. 2014. Dinâmica do sortimento de plantios de *Pinus taeda* L. em diferentes classes de sítio no Sul do Brasil. *Scien For* 42(103): 403-410.
- KRAUS K AND PFEIFER N. 1998. Determination of terrain models in wooded areas with airborne laser scanner data. *ISPRS J Photogram Remote Sens* 53(4): 3193-3203.
- MALTAMO M, EERIKÄINEN K, PITKÄNEN J, HYYPPÄ J AND VEHMAS M. 2004. Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions. *Remote Sens Environ* 90: 319-330.
- MCGAUGHEY RJ. 2015. FUSION/LDV: software for LiDAR Data Analysis and Visualization [Computer program]. Washington: USDA, Forest Service Pacific Northwest Research Station, 2015. http://http://forsys.cfr.washington.edu/fusion/FUSION_manual.pdf/.
- MCREOBERTS RE. 2009. A two-step nearest neighbors algorithm using satellite imagery for predicting forest structure within species composition classes. *Remote Sens Environ* 113: 532-545.
- MCREOBERTS RE. 2012. Estimating forest attribute parameters for small areas using nearest neighbors techniques. *For Ecol Manag* 272: 3-12.
- NÆSSET E. 1997. Determination of mean tree height of forest stands using airborne laser scanner data. *ISPRS J Photogram Remote Sens* 52(2): 49-56.
- NÆSSET E AND BJERKNES KO. 2001. Estimating tree heights and number of stems in young forest stands using airborne laser scanner data. *Remote Sens Environ* 78: 328-340.
- NELSON MD, HEALEY SP, MOSER WK, MASEK JG AND COHEN WB. 2011. Consistency of forest presence and biomass predictions modeled across overlapping spatial and temporal extents. *Math Comput For Nat-Resour Sci* 3(2): 102-113.
- NILSSON M. 1996. Estimation of tree heights and stand volume using an airborne LiDAR system. *Remote Sens Environ* 56(1): 1-7.
- PACKALÉN P, MALTAMO M AND MEHTATALO L. 2011. ALS-based estimation of plot volume and site index in a Eucalyptus plantation with a nonlinear mixed-effect model that accounts for the clone effect. *Ann For Sci* 68: 1085-1092.
- POPESCU SC. 2007. Estimating biomass of individual pine trees using airborne LiDAR. *Biomass and Bioenergy* 31(9): 646-655.
- POPESCU SC, WYNNE RH AND NELSON RF. 2003. Measuring individual tree crown diameter with LiDAR and assessing its influence on estimating forest volume and biomass. *Can J Remote Sens* 29: 564-577.
- R DEVELOPMENT CORE TEAM. 2015. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. <http://www.rproject.org/>.
- RACINE EB, COOPS NC, ST-ONGE B AND BEGIN J. 2014. Estimating Forest Stand Age from LiDAR Derived Predictors and Nearest Neighbor Imputation. *For Sci* 60(1): 128-136.
- ROBERTS SD, DEAN TJ, EVANS DL, MCCOMBS JW, HARRINGTON RL AND GLASS PA. 2005. Estimation individual tree leaf area in loblolly pine plantations using LiDAR derived measurements of height and crown dimensions. *For Ecol Manag* 2013: 54-70.

- SILVA CA, HUDAK AT, KLAUBERG C, VIERLING LA, GONZALEZ-BENECKE C, DE PADUA CHAVES CARVALHO S, RODRIGUEZ LCE AND CARDIL A. 2017a. 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* 12: 13-29.
- SILVA CA, HUDAK AT, VIERLING LA, KLAUBERG C, GARCIA M, FERRAZ A, KELLER M, EITEL J AND SAATCHI S. 2017c. Impacts of Airborne LiDAR Pulse Density on Estimating Biomass Stocks and Changes in a Selectively Logged Tropical Forest. *Remote Sens* 9: 1068-1069.
- SILVA CA, HUDAK AT, VIERLING LA, LOUDERMILK EL, O'BRIEN JJ, HIERS JK, JACK SB, GONZALEZ-BENECKE C, LEE H AND FALKOWSKI MJ. 2016b. Imputation of Individual Longleaf Pine (*Pinus palustris* Mill.) Tree Attributes from Field and LiDAR Data. *Can J Remote Sens* 42: 554-573.
- SILVACA, KLAUBERG C, CARVALHO SPC, HUDAK AT AND RODRIGUEZ LCE. 2014. Mapping aboveground carbon stocks using LiDAR data in *Eucalyptus* spp. plantations in the State of São Paulo, Brazil. *Scien For* 42(104): 591-604.
- SILVA CA, KLAUBERG C, CARVALHO SPC, PICCOLO MC AND RODRIGUEZ LCE. 2015. Estoque de carbono na biomassa área florestal em plantações comerciais de *Eucalyptus* spp. *Scien For* 43(105): 301-309.
- SILVA CA, KLAUBERG C, HUDAK A, VIERLING L, FENNEMA S AND CORTE A. 2017c. Modeling and mapping basal area of *Pinus taeda* L. plantation using airborne LiDAR data. *An Acad Bras Cienc* 89: 1895-1905.
- SILVA CA, KLAUBERG C, HUDAK AT, VIERLING LA, JAAFAR WSWM, MOHAN M, GARCIA M, FERRAZ A, CARDIL A AND SAATCHI S. 2017b. Predicting Stem Total and Assortment Volumes in an Industrial *Pinus taeda* L. Forest Plantation Using Airborne Laser Scanning Data and Random Forest. *Forests* 8: 254-267.
- SILVA CA, KLAUBERG C, HUDAK AT, VIERLING LA, LIESENBERG V, CARVALHO SP AND RODRIGUEZ LC. 2016a. A principal component approach for predicting the stem volume in Eucalyptus plantations in Brazil using airborne LiDAR data. *Forestry* 89: 422-433.
- STAGE AR AND CROOKSTON NL. 2007. Partitioning error components for accuracy-assessment of near neighbor methods of imputation. *For Scie* 53: 62-72.
- STERENCZAK K AND MOSKALIK T. 2014. Use of LiDAR-based digital terrain model and single tree segmentation data for optimal forest skid trail network. *iForest* 8: 661-667.
- STRÎMBU VF AND STRÎMBU BM. 2015. A graph-based segmentation algorithm for tree crown extraction using airborne LiDAR data. *ISPRS J Photogramm Remote Sens* 104: 30-43.
- WASKE B, VAN DER LINDEN S, OLDENBURG C, JAKIMOW B, RABE A AND HOSTERT P. 2012. imageRF - A user-oriented implementation for remote sensing image analysis with Random Forests. *Environ Model Software* 35: 192-193.
- WATT MS, MEREDITH A, WATT P AND GUNN A. 2014. The influence of LiDAR pulse density on the precision of inventory metrics in young unthinned Douglas-fir stands during initial and subsequent LiDAR acquisitions. *NZ J For Sci* 44(1): 1-18.
- WOODS M, LIM K AND TREITZ P. 2008. Predicting forest stand variables from LiDAR data in the Great Lakes-St. Lawrence forest of Ontario. *For Chr* 88(6): 827-838.
- YAO W, KRULL J, KRZYSZEK P AND HEURICH M. 2014. Sensitivity Analysis of 3D Individual Tree Detection from LiDAR Point Clouds of Temperate Forests. *Forests* 5: 1122-1142.
- ZANDONÁ DF, LINGNAU C AND NAKAJIMA NY. 2008. Varredura a laser aerotransportado para estimativa de variáveis dendrométricas. *Scien For* 36: 295-306.