http://sbsr.com.br/en/node/1443?q=/br/node/1443

APPLYING MIXED-EFFECTS MODEL FOR ESTIMATING INDIVIDUAL TREE ATTRIBUTES IN *Eucalyptus spp*. FOREST PLANTATIONS FROM FIELD AND AIRBORNE LIDAR DATA

Midhun Mohan [1,2],Carlos Alberto Silva [3,4], Carine Klauberg [5], Adrián Cardil[6], Danilo Almeida[7], Samuel de Padua Chaves e Carvalho [8], Wan Shafrina Wan Mohd Jaafar[9], Andrew T. Hudak[10]

[1]Department of Forestry and Environmental Resources, North Carolina State University, 2800 Faucette Drive, Raleigh, NC 27695, USA; mmohan2@ncsu.edu; [2]Drone Academy Thailand, Ratchpruek Village Sukkhaprachasun Rd., Bangpood, Prakred, Nonthaburi 11120, Thailand; [3]Biosciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD 20707, USA, carlos_engflorestal@outlook.com ; [4]Department of Geographical Sciences, University of Maryland, College Park, Maryland, MD 20740, USA. [5]Federal University of São João Del Rei – UFSJ, Sete Lagoas, MG, Brazil - 35701-970; carine_klauberg@hotmail.com; [6]Tecnosylva. Parque Tecnológico de León. 24009, León, Spain; adriancardil@gmail.com; [6]Department of Forest Sciences, University of São Paulo, "Luiz de Queiroz" College of Agriculture (USP/ESALQ), Av. Pádua Dias, 11, 13418-900 Piracicaba, SP, Brazil, daniloflorestas@gmail.com ; [8]College of Forestry, Federal University of Mato Grosso, Av. Fernando Correa da Costa, 2367, Boa Esperança, Cuiabá, MT 78060-900, Brazil - sam.padua@gmail.com; [9]School of Geosciences, University of Edinburgh, Edinburgh EH8 9XL, UK; wanshafrina@gmail.com ; [10]USDA Forest Service, Rocky Mountain Research Station, 1221 South Main Street, Moscow, ID 83843 USA, ahudak@fs.fed.us

ABSTRACT

Forest plantations cover a large part of tropical countries such as Brazil and eucalyptus plantations in particular account for 57% of Brazil's reforested area. As plantations offer multiple benefits – such as an option to offset natural forests, simplify otherwise complex forest ecosystems, meet energy, pulp and paper demands, restore ecological services and combat climate change by sequestering carbon – to the society, monitoring and tracking the growth and productivity of forest plantations should be given high priority. In this regard, remote sensing techniques have been found highly efficient. The core objective of this study is to estimate individual tree attributes, such as tree height, diameter at breast height (dbh) and above ground carbon (AGC) stocks of eucalyptus plantations from lidar data using linear mixed effects (LME) models; Ordinary Least Square (OLS) regression models are also built for comparison purposes. From our results, it can be inferred that hierarchy existing within the plantation datasets can be well handled by LME models and predictive models, for tracking tree level AGC and forest productivity, with satisfactory accuracies possible by combining lidar and LME modeling techniques.

Key words — plantation forestry, forest productivity, above ground carbon estimation, tree level attributes

1. INTRODUCTION

Forest plantations cover approximately 50 million ha in the tropics (1% of total global forest area) and are expected to continue expanding at rapid rates in the coming years due to their multiple benefits – such as the ability to offset natural forest exploitation, simplify otherwise complex forest ecosystems, meet energy, pulp and paper, and several wood products demand, restore ecological services initially offered by natural forests, as well as to offset greenhouse gas increases and combat climate change by sequestering carbon or by avoiding deforestation [1-6]. In this context, monitoring and tracking the growth and productivity of forest plantations has become a high priority.

In tropical countries like Brazil, *Eucalyptus* spp. (e.g. *Eucalyptus urograndis*) are the preferred species due to their high wood quality and fast growth rate; it accounts for approximately 57% (around 3.1 million ha) of the country's total reforested area [2]. Therefore, quantifying and modeling forest attributes – especially tree height (Ht), diameter at breast height (dbh) and above ground carbon (AGC) content - in eucalyptus plantations is crucial for optimizing management efficiency in a sustainable manner. In this regard, remote sensing applications can be deemed an invaluable resource, as they are able to provide comparable results to traditional field survey based methods, while being more economical, less laborious, repeatable, and highly time efficient [5-7]. However, existence of hierarchical structure (such as sample plots within forest stands) and crossed grouping structures (such as tree increments for different calendar years) and associated site and genetic variations [8-10] need to be considered while modeling eucalyptus plantations' forest attributes. This nesting happens due to multiple reasons: sometimes from the copious sample of felled trees from multiple plantation sources and at times, because of the different management regimes and multiple products generated from these plantations [12]. In general, linear mixed effects (LME) models [12,13] are found to be more flexible, precise, and accurate compared to Ordinary Least Square (OLS) regression models while working with plantation datasets, as these models account for the "random effects" and thereby enable us to generate multiple models for different sites and clones [14,15].

The primary objective of this study is to estimate individual tree attributes, such as Ht, dbh and AGC, from airborne lidar (Light Detection and Ranging) data using LME modeling. We hypothesize that incorporating LME models would allow us to better account for the hierarchical structure existing within the dataset and as such, provide us with estimates of tree level AGC and forest productivity, compared to OLS regression models.

2. MATERIAL AND METHODS

2.1. Study Area, Data Collection and Lidar Processing

The study area is located within the Telêmaco Borba municipality in the state of Paraná, Brazil. In total, there were 23 Eucalyptus spp stands, which included 8 different species and 5 different variety of clones; the plantations are managed by a pulp and paper company named Klabin S.A. As a part of field data collection, individual trees were measured for diameter dbh at 1.30 m, and a random subsample (15%) of trees were measured for Ht; heights of remaining trees were predicted using hypsometric equations employing dbh as the independent variable, and Ht as the dependent variable [16]. For estimating AGC stocks at tree level in the field, allometry equations were employed [19].

Lidar and high-resolution image data acquisitions were done using a Harrier 68i sensor mounted on a CESSNA 206 aircraft. The main lidar products - digital terrain model (DTM), canopy height model (CHM) and the lidar-derived canopy structure metrics - were derived from lidar point clouds by utilizing the tools available in US Forest Service FUSION/LDV 3.42 software [20].

2.2. Tree Detection and Crown Metrics Computation

For individual tree detection (ITD) and crown metrics, the rLiDAR package [21] available in R (R Development Core Team 2015) was used. First, smoothening of CHM by a 3 × 3 mean filter was done to remove spurious local maxima caused by tree branches. Second, ITD from the smoothed CHM was performed using the local maxima algorithm based FindTreeCHM function [21]. Third, detected tree crowns were delineated using the Voronoi tessellation algorithm based ForestCAS function. Afterwards, all returns of the normalized 3D heights within the tree crowns were extracted. Finally, crown metrics - maximum height (HMAX, m), crown projected area (CPA, m2) and crown volume (CV, m3) – were computed for individual trees using the CrownMetrics function. For accuracy assessment, comparisons were done between the number of trees detected (NTD) per plot from lidar and the numbers of trees observed in the field plots; exact tree locations and crown boundaries were later verified by overlaying it on high-resolution imagery as well.

2.3. Individual Tree Attribute Modeling

For developing and organizing the dataset relevant for the study, field and lidar-derived individual tree lists for each plot were sorted by crown height for tree matching and combined into a single table initially. This was further split into training (75%) and testing (25%) datasets for fitting the models and for validation purposes. Since data hierarchy was observed in the dataset through exploratory data analysis, linear mixed effects (LME) modeling was presumed appropriate for this study. Using lmer function within the lme4 package [22] in [23], 3 LME models were developed; one for each response variable: Ht, dbh and AGC. Here, the independent (fixed) variables included the lidar metrics (HMAX, CPA and CV) and field parameters - Species (SP), Clone (MG) and Age (AGE; in years) - were considered as random variables. Two sets of OLS regression models were also developed, for predicting each response variable, for comparison purposes. The first set of models considered HMAX, CPA and CV as independent variables and the second set of models were built with HMAX, CPA, CV, and AGE as independent variables. The Akaike information criterion (AIC) was calculated [17,18] to compare and rank the proposed models, and accuracy of each model was assessed with respect to the coefficient of determination (R^2), absolute and relative root mean square error (RMSE), and Bias (See [5] for more details).

3. RESULTS AND DISCUSSION

On preliminary visual analysis – employing scatter plots – we identified strong patterns between HMAX and field parameters with respect to the response variables (see figure 1), which implies possible data clustering within the data; no specific patterns were observed in the cases of CPA or CV. Therefore, while creating LME models, random effects were nested only for HMAX. We observed the LME models outperforming alternative OLS models in all the three cases. LME models explained 87%, 83% and 77% of the variation while predicting AGC, Ht and dbh respectively. In the case of the first set of OLS regression models, with HMAX, CV and CPA as parameters, satisfactory results were obtained, although accompanied by comparatively higher RMSE values while predicting Ht and dbh, and very poor results (RMSE of 23.95%) in the case of AGC prediction. The second set of OLS regression models, which had the field parameter age included in the model, showed a slight improvement in the results compared to the former model; however, the results were not as good as the ones obtained using LME models (see table 1), which shows the superiority of LME models while working with nested data.

After performing cross-validation of modeling data, RMSE values revealed that the mixed effects model predictions for dbh had the lowest score (9.75%) followed by Ht (11.6 %), both with low bias (0.49% and 0.39%

http://sbsr.com.br/en/node/1443?q=/br/node/1443

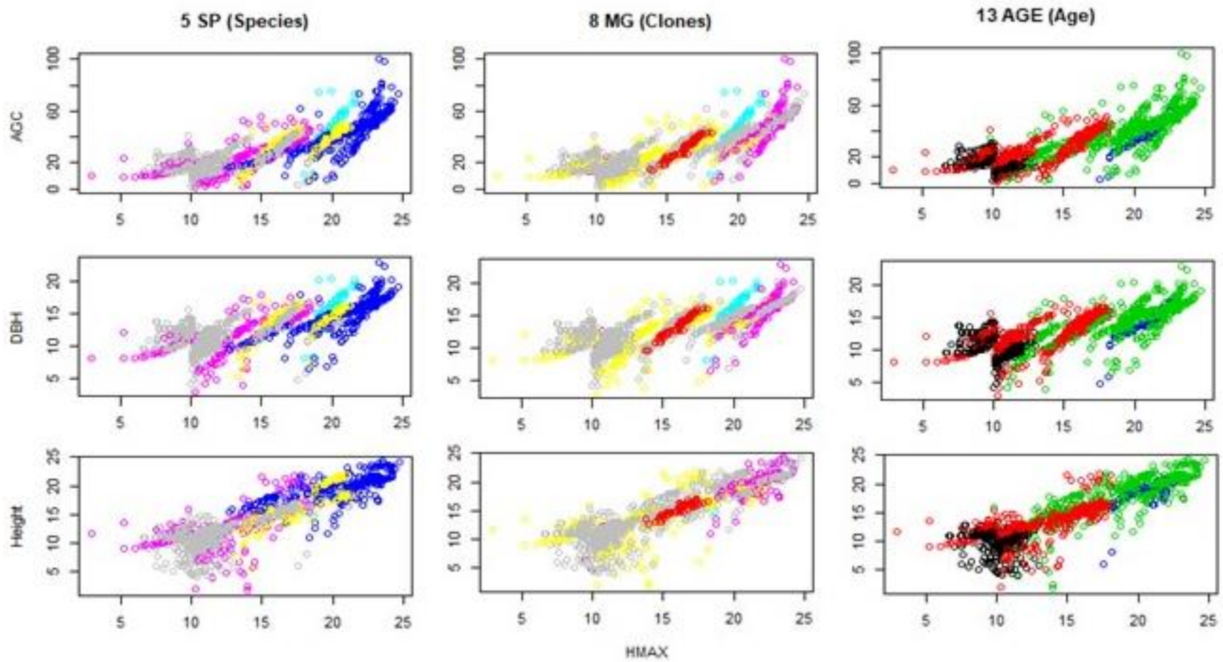

Figure 1. Scatter plots showing the hierarchy within data (HMAX|SP/MG/AGE)

Table 1: Model results and performance comparisons

Predictive Models	Model Equations	RMSE (Kg; m; cm)	RMSE %	Bias (Kg; m; cm)	Bias %	R^2
LM – lidar metrics	AGCt = HMAX + CPA + CV	7.51	23.95	-0.2	-0.74	0.73
	Ht = HMAX + CPA + CV	1.96	12.77	0.03	0.17	0.79
	dbh = HMAX + CPA + CV	1.66	12.65	-0	-0.02	0.61
LM – lidar metrics + age	AGCt = HMAX + CPA + AGE	6.8	21.78	-0.1	-0.34	0.78
	Ht = HMAX + CPA + AGE	1.82	11.93	0.07	0.44	0.82
	Ht = HMAX + CPA + AGE	1.82	11.93	0.07	0.44	0.82
MEM - lidar metrics + field parameters	AGCt = HMAX + CPA + CV + (1 + HMAX | SP/MG/AGE)	5.14	16.61	0.2	0.65	0.87
	Ht = HMAX + CPA + CV + (1 + HMAX | SP/MG/AGE)	1.78	11.6	0.06	0.39	0.83
	dbh = HMAX + CPA + CV + (1 + HMAX | SP/MG/AGE)	1.28	9.75	0.06	0.49	0.77

respectively). From the R^2 values for dbh and Ht - .77 and .83 – it became clear that the addition of random effects reduced the RMSE values and helped in defining the model more effectively. This happens as LME models work better in complicated forest datasets with groupings (e.g., trees within plots, sample plots within stands, etc.) or that have multiple nested levels (e.g., repeated observations of trees in successive years or in different aerial imageries or even within same sample plots) [8-10,24]. In these cases, the OLS regression model performances are usually limited, since, they are not able to account for the non-independence in the data arising from the inherent hierarchical structure. In the case of dbh predictions, R^2 values for the second set of OLS regression models and LME models were found to be identical, which implies that the influence random effects have in this case is nominal. RMSE of AGC was above the accepted limit of 15% for all the models; this higher RMSE might have occurred as a result of errors incurred through the indirect calculations, using allometric equations. The maximum R^2 value was observed for the LME model (.87).

5. CONCLUSIONS

Estimation and tracking of forest attributes are pivotal for effective decision making and management of eucalyptus plantations in Brazil and other tropical countries. The findings from our study highlight the efficiency of LME models in predicting forest attributes – such as tree height, dbh and above ground carbon – in plantations, with the support of lidar data. LME models are able to accommodate the hierarchical differences within data (which usually happens in the case of plantations) as random effects in the model; therefore, they are recommended as a better alternative to OLS regression models while working with predictive models for plantations.

6. REFERENCES

[1] Stape, J. L., Binkley, D., & Ryan, M. G. (2008). Production and carbon allocation in a clonal Eucalyptus plantation with water and nutrient manipulations. *Forest Ecology and Management*, 255(3-4), 920-930.

[2] Ibá (Indústria Brasileira de Árvores). Brazilian Tree Industry. 2015. Available online: http://www.iba.org/images/shared/iba_2015.pdf (accessed on 10 November 2016)

[3] Paquette, A., & Messier, C. (2010). The role of plantations in managing the world's forests in the Anthropocene. *Frontiers in Ecology and the Environment*, 8(1), 27-34

[4] Roise, J. P., Harnish, K., Mohan, M., Scolforo, H., Chung, J., Kanieski, B., ... & Shen, T. (2016). Valuation and production possibilities on a working forest using multi-objective programming, Woodstock, timber NPV, and carbon storage and sequestration. *Scandinavian Journal of Forest Research*, 31(7), 674-680.

[5] Silva, C. A., Klauberg, C., Hudak, A. T., Vierling, L. A., Liesenberg, V., Carvalho, S. P. E., & Rodriguez, L. C. (2016). A principal component approach for predicting the stem volume in Eucalyptus plantations in Brazil using airborne LiDAR data. *Forestry: An International Journal of Forest Research*, 89(4), 422-433.

[6] Silva, C. A., Klauberg, C., Hudak, A. T., Vierling, L. A., Jaafar, W. M., Shafrina, W., Mohan, M., ... & Saatchi, S. (2017). Predicting Stem Total and Assortment Volumes in an Industrial Pinus taeda L. Forest Plantation Using Airborne Laser Scanning Data and Random Forest. *Forests* (19994907), 8(7).

[7] Mohan, M., Silva, C. A., Klauberg, C., Jat, P., Catts, G., Cardil, A., ... & Dia, M. (2017). Individual Tree Detection from Unmanned Aerial Vehicle (UAV) Derived Canopy Height Model in an Open Canopy Mixed Conifer Forest. *Forests*, 8(9), 340.

[8] Packalén, P., Mehtätalo, L., & Maltamo, M. (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. *Annals of Forest Science*, 68(6), 1085.

[9] Mehtätalo, L. (2004). A longitudinal height–diameter model for Norway spruce in Finland. *Canadian Journal of Forest Research*, 34(1), 131-140.

[10] de Souza Vismara, E., Mehtätalo, L., & Batista, J. L. F. (2015). Linear mixed-effects models and calibration applied to volume models in two rotations of Eucalyptus grandis plantations. *Canadian Journal of Forest Research*, 46(1), 132-141.

[11] Calegario, N., Daniels, R. F., Maestri, R., & Neiva, R. (2005). Modeling dominant height growth based on nonlinear mixed-effects model: a clonal Eucalyptus plantation case study. *Forest Ecology and Management*, 204(1), 11-21.

[12] Lindstrom, M. J., & Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics*, 673-687

[13] Pinheiro, J. C., & Bates, D. M. (2000). Linear mixed-effects models: basic concepts and examples. *Mixed-effects models in S and S-Plus*, 3-56.

[14] Hedges, L. V., & Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological methods*, 3(4), 486.

[15] Gardiner, J. C., Luo, Z., & Roman, L. A. (2009). Fixed effects, random effects and GEE: what are the differences?. *Statistics in medicine*, 28(2), 221-239.

[16] Curtis, R. O. (1967). Height-diameter and height-diameter-age equations for second-growth Douglas-fir. *Forest science*, 13(4), 365-375.

ISBN: 978-85-17-00097-3

http://sbsr.com.br/en/node/1443?q=/br/node/1443

[17] Akaikei, H. (1973). Information theory and an extension of maximum likelihood principle. In Proc. *2nd Int. Symp. on Information Theory* (pp. 267-281).

[18] Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6), 716-723.

[19] Silva, C. A., Klauberg, C., Pádua, S. C. D., Piccolo, M., & Rodriguez, L. C. E. (2015). Estoque de carbono na biomassa aérea florestal em plantações comerciais de Eucalyptus spp. *Sci For*, 43, 135-46.

[20] McGaughey, R. J. (2009). FUSION/LDV: Software for LIDAR data analysis and visualization. *US Department of Agriculture, Forest Service, Pacific Northwest Research Station: Seattle, WA, USA*, 123(2).

[21] Silva, C.A., Crookston, N.L., Hudak, A.T., and Vierling, L.A. 2015. rLiDAR: An R package for reading, processing and visualizing lidar (Light Detection and Ranging) data, version 0.1, accessed Oct. 15 2015, < http://cran.rproject.org/web/packages/rLiDAR/index.html

[22] Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint* arXiv:1406.5823.

[23] . R Core Team. 2015. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing, Vienna, Austria*, accessed Oct. 15 2015, <http://www.R-project.org≥.

[24] Mehtätalo, L. (2016). Mixed-Effect Models for Prediction of Tree Attributes