University of São Paulo
"Luiz de Queiroz" College of Agriculture

LiDAR technology applied to vegetation quantification and qualification

Eric Bastos Görgens

Thesis presented to obtain the degree of Doctor in Science. Area: Forest Resources. Sub-area: Silviculture and Forest Management

Piracicaba
2014

Eric Bastos Görgens
Degree in Forestry

LiDAR technology applied to vegetation quantification and qualification

Advisor:
Prof. Dr. **LUIZ CARLOS ESTRAVIZ RODRIGUEZ**

Thesis presented to obtain the degree of Doctor in Science. Area: Forest Resources. Sub-area: Silviculture and Forest Management

Piracicaba
2014

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP

Görgens, Eric Bastos
LiDAR technology applied to vegetation quantification and qualification / Eric Bastos Görgens. - - Piracicaba, 2014.
102 p. : il.

Tese (Doutorado) - - Escola Superior de Agricultura “Luiz de Queiroz”, 2014.

1. ALS 2. Laser 3. Perfil vertical 4. Métricas LiDAR 5. Estimação de volume I. Título

CDD 634.9734
G667L

DEDICATION

To **YOU**,

I dedicate this work.

ACKOWNLEDGEMENTS

I would like to thanks my parents Robert and Jacy, my sister Diana and my love Pollyanna for always support my dreams.

I thank my family (grandparents, uncles, aunts and cousins) and my second family Roberto and Tânia who helped me along this walk.

I thank Professor Luiz Carlos Estraviz Rodriguez for advices, friendship and for the stimulating discussions.

My warm thanks to my friends and scientific partners André Gracioso Peres da Silva and Matheus Henrique Nunes.

I thank my fellow labmates in GET-LiDAR: Bruno Kanieski, Julianne Oliveira, Debora Camilo, Danitiele Laranja, Tiago de Conto and Clayton Alvares.

My sincere thanks to Professors Carlos Pedro Boechat Soares (UFV), João Batista (ESALQ), Demóstenes Silva Filho (ESALQ), Sílvio Ferraz (ESALQ), Hélio Garcia Leite (UFV), Márcio Romarco de Oliveira (UFVJM), Nicholas Coops (UBC), Petteri Packalen (UEF), Timo Pukkala (UEF), Ruben Valbuena (UEF), Håkan Olsson (SLU) and Alessandro Montaghi (SLU).

I cannot forget to thank the incredible Jefferson Polizel and my labmates in CMQ (Centro de Métodos Quantitativos – ESALQ).

I thank Giovana Oliveira for all the valuable advices and talks.

I also thank Richard Kennedy and Simon Dunstin for the English assistance.

I thank the University of São Paulo and staff, on behalf of Katarina Germutz and Margarete PInese.

I thank IPEF, Virginia Tech, North Caroline State Univertiry, Fibria SA, EMBRAPA Satélites, Forest Service - US, the City of Belo Horizonte for supply the data to my research.

I thank Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for the scholarship.

SUMMARY

RESUMO....9
ABSTRACT....11
1 INTRODUCTION....13
References....20
2 ACQUIRING MORE INFORMATIVE METRICS THROUGH THE OPTIMIZATION OF THRESHOLD HEIGHTS WHEN USING AIRBORNE LASER SCANNING TO ESTIMATE THE STAND VOLUME OF A EUCALYPTUS PLANTATION....25
Abstract....25
2.1 Introduction....25
2.2 Material and Methods....27
2.2.1 Field and ALS data....27
2.2.2 Metrics extraction....28
2.3 Results....29
2.3.1 Minimum height....30
2.3.2 Height break....31
2.4 Discussion....32
2.5 Conclusions....33
Acknowledgments....34
References....34
3 DEVELOPMENT OF STAND VOLUME MODELS BASED ON STABLE METRICS AS FROM MULTIPLE ALS ACQUISITIONS IN EUCALYPTUS PLANTATIONS....37
Abstract....37
3.1 Introduction....37
3.2 Material and Methods....40
3.3 Results....44
3.4 Discussion....48
Acknowledgments....50
References....50

4 ANALYSIS OF THE PERFORMANCE OF TWO MACHINE LEARNING METHODS TO ESTIMATE THE FOREST STAND VOLUME USING AIRBORNE LASER SCANNING DATA 53
Abstract 53
4.1 Introduction 54
4.2 Material and Methods 56
4.3 Results 59
4.4 Discussion 62
4.5 Conclusion 64
Acknowledgments 64
References 65
5 CHARACTERIZATION OF FOREST TYPOLOGIES UTILIZING VERTICAL PROFILES DERIVED FROM AIRBORNE LASER SCANNING SURVEYS 69
Abstract 69
5.1 Introduction 70
5.2 Materials and Methods 71
5.2.1 Study areas 71
5.2.2 Airborne Laser Surveys 74
5.2.3 Vertical profile 75
5.2.4 Spatial dependence and sampling unit size 76
5.2.5 Sampling intensity and sampling unit size 77
5.3 Results 77
5.3.1 Vertical profiles of the topologies 77
5.3.2 Random component and sampling unit size 79
5.3.3 Sampling intensity and sampling unit size 81
5.4 Discussion 83
5.4.1 Vertical profile of the vegetation 83
5.4.2 Spatial dependence, sampling intensity and sampling unit size 86
5.5 Conclusions 89
Acknowledgments 90
References 90
6 CONCLUSIONS 97
APPENDIX 99

RESUMO

O uso de tecnologia LiDAR para quantificação e qualificação da vegetação

A metodologia para quantificar vegetação a partir de dados LiDAR (Light Detection And Ranging) está de certa forma consolidada, porém ainda existem pontos a serem esclarecidos que permanecem na lista da comunidade científica. Quatro aspectos foram estudos nesta tese. No primeiro estudo, foi investigado a influência das alturas de referência (altura mínima e altura de quebra) na qualidade do conjunto de métricas extraído visando estimação do volume de um plantio de eucalipto. Os resultados indicaram que valor mais altos de alturas de referência retornaram um conjunto de métricas melhor. O efeito das alturas de referência foi mais evidente em povoamentos jovens e para as métricas de densidade. No segundo estudo, avaliou-se a estabilidade de métricas LiDAR derivadas para uma mesma área sobrevoada com diferentes configurações de equipamentos e voo. Este estudo apresentou como a seleção de métricas estáveis pode contribuir para a geração de modelos compatíveis com diferentes bases de dados LiDAR. De acordo com os resultados, as métricas de altura foram mais estáveis que as métricas de densidade, com destaque para os percentis acima de 50% e a moda. O terceiro estudo avaliou o uso de máquinas de aprendizado para a estimação do volume em nível de povoamento de plantios de eucalipto a partir de métricas LiDAR. Ao invés de estarem limitados a um pequeno subconjunto de métricas na tentativa de explicar a maior parte possível da variabilidade total dos dados, as técnicas de inteligência artificial permitiram explorar todo o conjunto de dados e detectar padrões que estimaram o volume em nível de povoamento a partir do conjunto de métricas. O quarto e último estudo focou em sete áreas de diferentes tipologias florestais brasileiras, estudando os seus perfis verticais de dossel. O estudo mostrou que é possível diferenciar estas tipologias com base no perfil vertical derivado de levantamentos LiDAR. Foi observado também que o tamanho das parcelas possui diferentes níveis de dependência espacial. Cada tipologia possui características específicas que precisam ser levadas em considerações em projetos de monitoramento, inventário e mapeamento baseado em levantamentos LiDAR. O estudo mostrou que é possível determinar o perfil vertical de dossel a partir da cobertura de 10% da área, chegando a algumas tipologias em apenas 2% da área.

Palavras-chave: ALS; Laser; Perfil vertical; Métricas LiDAR; Estimação de volume

ABSTRACT

LiDAR technology applied to vegetation quantification and qualification

The methodology to quantify vegetation from airborne laser scanning (or LiDAR - Light Detection And Ranging) is somehow consolidated, but some concerns are still in the checklist of the scientific community. This thesis aims to bring some of those concerns and try to contribute with some results and insights. Four aspects were studied along this thesis. In the first study, the effect of threshold heights (minimum height and height break) in the quality of the set of metrics was investigated aiming the volume estimation of a eucalyptus plantation. The results indicate that higher threshold height may return a better set of metrics. The impact of threshold height was more evident in young stands and for canopy density metrics. In the second study, the stability of the LiDAR metrics between different LiDAR surveys over the same area was analyzed. This study demonstrated how the selection of stable metrics contributed to generate reliable models between different data sets. According to our results, the height metrics provided the greatest stability when used in the models, specifically the higher percentiles (>50%) and the mode. The third study was designed to evaluate the use of machine learning tools to estimate wood volume of eucalyptus plantations from LiDAR metrics. Rather than being limited to a subset of LiDAR metrics in attempting explain as much variability in a dependent variable as possible, artificial intelligence tools explored the complete metrics set when looking for patterns between LiDAR metrics and stand volume. The fourth and last study has focused upon several highly important forest typologies, and shown that it is possible to differentiate the typologies through their vertical profiles as derived from airborne laser surveys. The size of the sampling cell does have an influence on the behavior observed in analyses of spatial dependence. Each typology has its own specific characteristics, which will need to be taken into consideration in projects targeting monitoring, inventory construction, and mapping based upon airborne laser surveys. The determination of a converged vertical profile could be achieved with data representing 10 % of the area for all typologies, while for some typologies 2 % coverage was sufficient.

Keywords: ALS; Laser; Vertical profile; LiDAR metrics; Volume estimation

1 INTRODUCTION

"Try again, fail again.
Fail better."
Samuel Beckett

Soviet researchers were the first to study the use of laser range finders for measuring trees. In 1976, Soldukhin and his team performed the first laser profiling exercises designed to obtain tree attributes. Good results obtained from ground tests led in 1977 to the installation of laser equipment in a plane which took an initial survey whilst flying at 160 km h^{-1} at an altitude of 40 meters. As a result of this effort, the Soviet researchers reported that the profiles obtained could be employed towards establishing computerized or automated information of use for determining forest management (NELSON, 2013).

Over the years, several technical challenges were subsequently overcome such as increasing the energy laser pulse to allow higher flight, greater precision in the positioning of aerial shots, and improved alignment of the flight tracks. In 1980, Krabill and McDonough had already conducted studies whilst flying at heights of 300 and 900 meters respectively (NELSON, 2013).

In the 1980s, the researchers Aldred and Bonner (1985) published the first airborne laser sensor application with the ability to acquire forest information. Once confirmed that it was possible to obtain good estimates of stand height and density, the next step would naturally be to quantify the vegetation (principally biomass and volume). As demonstrated by Maclean and Krabill (1986), the profile generated by airborne laser scanning was directly and linearly related to volume ($R^2 = 0.76$).

The technology involved in laser surveys (also known as LiDAR; 'Light Detection And Ranging') continued to develop over the years, from airborne equipment through to the hardware and software necessary for storage and data processing. An airborne laser survey involves a set of tools, consisting of both for transmitting and receiving laser units, a global positioning unit (GPS), inertial measurement unity (IMU) and a central computer that controls the system and stores the data (BALTSAVIAS, 1999). The laser used with land and vegetation surveys usually operates at a near-infrared wavelength (~ 1064 nm), with low divergence (<0:20 mrad) and a high repetition rate (~ 300 kHz) (BALTSAVIAS, 1999; VASTARANTA, 2012).

Errors of up to 3 meters can occur in studies related to airborne laser scanning (ALS). These errors are consistent with the errors occurring in field surveys (COOPS et al., 2004). LiDAR errors typically underestimate the real height of trees, since the probability of the laser pulse measuring the exact apex of the treetop is very low (ST-ONGE et al., 2003). This underestimation can be reduced by increasing the pulse density per unit area via the reduction of flight altitude, via the reduction of aircraft speed, or via repeated overflights (COOPS et al., 2007).

Quantitative forest data from LiDAR is based on the high correlation between forest attributes and the statistics extracted from the point cloud formed by the laser returns (GOODWIN et al., 2006). Thus, for a project to accurately quantify vegetation parameters two elements are required: LiDAR data for the area to be quantified and a pre-calibrated model relating forest attributes to the LiDAR metrics.

The main product of a LiDAR flight is the point cloud, which is usually delivered in *las* format (maintained by the American Society for Photogrammetry and Remote Sensing - ASPRS). This provides a binary file format for storing ALS data. This format, which now exists in its version 1.4, has had to evolve to keep pace with technological developments in airborne laser sensors. To work successfully with vegetation, it is necessary to have a classified point cloud, albeit not filtered. Unlike digital terrain models (which can be derived from the last returns) or digital surface models (which may be derived from the first returns), all returns are utilized when processing LiDAR data to quantify vegetation characteristics.

The LiDAR file retains information related to flight equipment configuration, plus the recorded variables and each recorded return. The most basic information associated with each return consists of: XYZ position in space, return intensity, return number, number of returns generated by the pulse, emission angle and classification (ASPRS, 2009).

The return XY position is usually the UTM coordinate, and Z is the return elevation relative to the reference datum established for the survey. Return intensity is the proportion of original energy emitted that returned and was picked up by the sensor. Although theoretically interesting, this intensity information is not widely used to quantify and qualify biomass. The main reason for this being that intensity is not calibrated during the survey and can be influenced by several factors, even varying at times within a single flight line.

Point density is a property of cloud LiDAR returns that can be a limiting factor in its use (VAUHKONEN et al., 2008). The number of points per square meter determines the degree of detail with which objects may be imaged. Applications requiring higher detail, will thus require higher densities. Vauhkonen et al. (2008) proposed reference densities for some applications:

- 0.5 to 1 returns per m^2 = terrain modeling
- 5 to 6 returns per m^2 = forestry applications
- 11 to 12 returns per m^2 = individual tree applications
- Over 20 returns per m^2 = archaeological applications.

Cloud returns originally save elevation information. Most vegetation monitoring however uses height above ground level. In this regard, the first step after obtaining a LiDAR cloud is to generate a digital terrain model (DTM). This DTM is a raster file derived from filtered and classified returns as terrain. Different filters can be used, from simple filters based on the type of return (e.g. past returns) to filters designed based on the curvature of neighboring returns (e.g. multiscale curvature method). All of these methods aim to separate from the LiDAR cloud those returns most likely to have been backscattered by the terrain and to use them as a reference for generating the DTM.

Having filtered these returns with the highest probability of belonging to the terrain, the value of each raster cell (pixel) is then calculated by the average of the returns within the cell area (MENG; CURRIT; ZHAO, 2010). The raster produced contains the same XY dimensions as the particular cloud from which it derives, making it possible to discount from each return the value representing the ground. This process is known as the normalization of LiDAR data, the processed cloud representing the variable Z as vegetation height relative to the reference DTM.

Two methodological approaches were developed over the time for the processing of LiDAR data in order to quantify vegetation characteristics: Area Based Approach (ABA) and Individual Tree Detection (ITD). Both make correlations between data obtained from field measurements and information derived from the LiDAR cloud, albeit at different scales (VASTARANTA, 2012). Whilst ABA uses an area as the reference, ITD uses an individual tree.

The ITD approach is guided by the extraction of individual metrics such as crown diameter, tree height and canopy depth for each identified and isolated trees. These

individual tree metrics are then also linked to individual tree attributes (e.g. volume) measured in the field (HYYPPÄ; INKINEN, 1999). Through one variation known as the Tree Cluster Approach (TCA), the metrics extracted from individual trees are consolidated into an area (e.g. plot) before being linked to attributes measured in the field, (POPESCU; WYNNE; NELSON, 2002). Thus, LiDAR processing occurs at tree level (ITD) but the linkage between metrics and attributes occurs at the area level (ABA). Both approaches will result in models that can be used to estimate the attributes for areas with LiDAR data, and thus without the need for field measurements..

Detection algorithms and the intensity of clustering trees are among the most relevant factors for the success of ITD (KAARRINEN et al., 2012; VAUHKONEN et al., 2012). Applications of ITD have been successful in isolating around 70% of the trees in temperate coniferous forests, which in turn represent approximately 91% of the total volume (PERSSON; SODERMAN; HOLMGREN, 2002). In mixed temperate forests the detection success rate came close to 40% of the trees (PITKANEN et al., 2004).

In practice however, ITD is not yet widely used due to the low availability of field data at an individual level, and also to the high accuracy required by individual geo-referencing. ABA is thus the most frequently used technique for vegetation quantification, since it takes much of the effort out of traditional forest inventory. Just as with ITD, field data give context to the statistics derived from LiDAR data, although these linkages are made at the plot level.

After clipping of the normalized LiDAR cloud, which is based on georeference and field plots size, the statistics are then extracted from this cloud and linked to the forest attributes measured in field plots. The models obtained are then used to estimate the vegetation stock from any other areas sensed via laser scanning (WHITE et al., 2013).

These extracted statistics are referred to as metrics; grouped in terms of height and density. The height metrics consist primarily of measures of location (e.g. mean, median, percentiles) and measures of dispersion (e.g. standard deviation, interquartile range, variance). The density metrics compute ratios of returns above a height break.

A common (though not constant) practice is to remove returns below a minimum height whilst computing height metrics. Elimination of these returns is defended by the argument that this action eliminates any returns that have not been backscattered by vegetation, but by the ground, rocks or understory. Different minimum height settings can be applied depending on the approach of the analyst (NILSSON, 1996; NÆSSET;

BJERKNES, 2001; ZIMBLE et al., 2003; HUDAK et al., 2006; JENSEN et al., 2006). In case of density metrics, height break can be fixed when determined by the analyst, or dynamic when determined based on other metrics from the cloud (e.g. mean, median, mode) (NÆSSET, 2002; ANDERSEN; MCGAUGHEY; REUTEBUCH, 2005; SULLIVAN, 2008). In Chapter 2 we investigate the effect of these threshold heights (minimum height and height break) on the quality of a set of metrics generated to estimate the volume of a eucalyptus plantation.

Filters can even be defined for the type of returns to be used during metrics calculation. Besides performing a calculation that considers all returns, it is common in vegetation studies to discard the first returns, since they are usually backscattered due to interactions between the laser pulses and the vegetation (NÆSSET, 2002).

If combining all of the available options that concern threshold heights and filters, multiple metrics can be generated from a LIDAR cloud. As a result, most of these metrics are collinear. The question that persists is then which metrics should be selected?

Different methods can be employed to select or discard metrics when composing a final estimation model. Amongst parametric methods employed for model-building are the stepwise method (NÆSSET, 2002), the inter-correlation and correlation method (STEPHENS et al., 2012.) and metric selection via multivariate analysis (LI; ANDERSEN; MCGAUGHEY, 2008). Amongst non-parametric methods are Nearest Neighbors (NN), Neural Network (ANN) and Random Forest® (RF). Some nonparametric methods allow the researcher to work with any set of metric size without the need to either select or discard variables. Another approach is to follow pre-established principles in the literature; for instance Lefsky et al. (2005) recommended that volume estimation models should contain metrics related to height variability (site index) and the amount of vegetation present (stand density).

The choice of metrics has varied widely within the published literature. For example Næsset (2002) established three different models for volume estimation of differently-aged stands of a forest composed predominantly of Scots Pine (*Pinus sylvestris* L.) and Norway Spruce (*Picea abies* (L.) Karst.). Volume of a younger stand was best modeled by the metrics: average height of last returns and proportion of the first returns above the 50th percentile. Volume of a mature stand growing a lower quality site was best modeled via the metrics: 30th percentile, 50th percentile and 60th percentile and proportion of returns above the 90th percentile, all derived from first

returns. Volume of a mature stand in a good quality site was best modeled via metrics: 80th percentile of last returns, maximum height of first returns and proportion of the first returns above the 50th percentile. Volume of a eucalyptus forest plantation in Brazil was best modeled via the 10th percentile for the first returns (ZONETE, 2009). A study by Stephens et al. (2012) to establish national model to estimate carbon in New Zealand, revealed the 30th percentile and the canopy cover metric to be the most accurate predictors.

Numerous studies have been conducted to determine how the properties of LiDAR equipment and the aircraft can influence the accuracy of metrics, and to study the effects on vegetation quantification efforts. Goodwin et al. (2006) established that the altitude, the scanning angle and the footprint size do not influence the vertical distribution of LiDAR returns at stand level. However, Næsset (2002) found that the use of different sensors, the pulse emission frequency and flight altitude influenced estimation of vertical height profile and also effected the penetration level of the pulse. He emphasized that last returns are more sensitive to these flight parameters than first returns. Crash et al. (2011) found that data acquired with constant flight parameters and laser equipment settings remains practically unchanged over time. Complexity increased when different flight parameters, or different sensor configurations were taken into account. For large-scale monitoring efforts (i.e. national level) equipment settings are almost impossible to keep constant, and flight parameters are even further less likely to be consistent. In Chapter 3, we discuss the resilience of metrics derived from flights made with different equipment configurations and flight parameters.

Usually, field information used to adjust the models arises from traditional forest inventories. Tree diameter and height is measured and consolidated into a plot level attribute (e.g. volume, biomass and others). It is essential that all field plots to be precisely georeferenced in order to allow correct geolocation into the LiDAR cloud. It is also important to collect the field measurements at a similar time to LiDAR survey. In the case of physiological seasonality, this feature should be considered when planning both the field work and the LiDAR survey.

The main method used when modeling the relationship between LiDAR and field data is parametric regression. However, recent studies have shown that non-parametric techniques such as Random Forest® and nearest neighbor can be successfully used for the estimation of forest attributes (HUDAK et al., 2008; LATIFI; NOTHDURFT;;; KOCH, 2010; SIMARD et al., 2011). Amongst the mathematical

criteria used for model comparison are the root mean square error (RMSE), coefficient of determination (R^2) and the bias. Model choice should not rely solely on mathematical criteria, but should also take into account the context within which the model will be applied; for instance the hardware, software and the amount of data available.

In Chapter 4, a model developed through selecting metrics based on inter-correlation and correlation is compared with two models that do not require any reduction in the number of variables (neural networks and Random Forest®). The aim was to model wood volume in eucalyptus plantations based on LiDAR metrics.

Once adjusted and validated, the model estimates forest attributes for the entire area, allowing the production of prediction maps (GREGOIRE et al., 2011). In some cases, is possible to use LiDAR metrics in the first phase of a double sampling procedure (two-phase), especially if highly correlated to the attribute of interest (e.g. volume). This results in a higher sampling intensity (STEPHENS et al., 2012). This double sampling allows for estimation of a forest attribute through relating its value to the sampling error.

The quantification of vegetation characteristics is undoubtedly of major interest for forest managers. In addition to establishing the availability of wood supply, accumulation of this information over time allows for calculations of growth, replacement rate and mortality rate. Quantification and qualification of forest vertical structure are also essential for providing an understanding of the functioning of forest ecosystems. Information on the forest vertical structure is difficult to quantify through direct field sampling however, since it can be highly labor intensive (COOPS et al., 2007).

Modeling the vertical structure of the forest can add an important element to the understanding of functional behavior and the growth of the trees. It can also provide important information regarding assessment of the nature of forest response to different levels of disturbance (tree population, community and ecosystem), whether natural or anthropogenic (PARKER et al., 2004).

Passive remote sensing techniques have limited ability to assess structural changes since they cannot penetrate the vegetation canopy (COOPS et al., 2007). That said, Coops et al. (2007) and Drake et al. (2002) demonstrated that LIDAR sensors could be successfully used to estimate important parameters of different structural attributes, which in turn could be presented as effective predictors for forest attributes.

Each vegetation type typically has a particular vertical structure that influences the distribution of laser returns. Thus, through fitting a theoretical distribution to the vertical profile of the laser returns, the parameters obtained can reflect in some way the vertical structure of forest surveyed. These parameters can be used directly within models that estimate forest attributes or to classify vertical forest structures (LOVELL et al., 2003; COOPS et al., 2007; SILVA, 2014). Chapter 5 presents a study on the usage of vertical profiles for classification and differentiation of Brazilian forest types, including a discussion about the sample design required for vertical profile estimation and about its geospatial dependence.

This thesis focus on four specific topics regarded to vegetation quantification and qualification through airborne laser surveys:

- Investigate the effect of threshold heights (minimum height and height break) on the explanation capacity of a set of metrics aimed at estimating the volume of a eucalyptus plantation.
- Investigate the stability of LiDAR metrics through comparison of different LiDAR surveys over the same area.
- Evaluate the application of machine learning tools in estimating the wood volume of eucalyptus plantations.
- Differentiate Brazilian forest types based on vertical profiles derived from airborne laser surveys and determine a related sampling design and spatial dependence.

References

ALDRED, A.; BONNOR, G. **Application of airborne lasers to forest surveys.** Ontario: Petawawa National Forestry Institute, Canadian Forestry Service, 1985. 62 p. (Information Report, PI-X-51).

ANDERSEN, H.E.; MCGAUGHEY, R.J.; REUTEBUCH, S.E. Estimating forest canopy fuel parameters using LIDAR data. **Remote Sensing of Environment**, Amsterdam, v. 94, n. 4, p. 441-449, 2005.

AMERICAN SOCIETY FOR PHOTOGRAMMETRY & REMOTE SENSING. **LAS specification version 1.3**. Bethesda, 2009. 17 p. (Technical Report).

BALTSAVIAS, E.P. Airborne laser scanning: existing systems and firms and other resources. **ISPRS Journal of Photogrammetry and Remote Sensing**, Amsterdam, v. 54, n. 2/3, p. 164-198, 1999.

BATER, C.W.; WULDER, M.A.; COOPS, N.C.; NELSON, R.F.; HILKER, T.; NÆSSET, E. Stability of sample-based scanning-LiDAR-derived vegetation metrics for forest monitoring. **IEEE Transactions on Geoscience and Remote Sensing**, New York, v. 49, n. 6, p. 2385-2392, 2011.

COOPS, N.C.; HILKER, T.; WULDER, M.A.; ST-ONGE, B.; NEWNHAM, G.; SIGGINS, A.; TROFYMOW, J.T. Estimating canopy structure of Douglas-fir forest stands from discrete-return LiDAR. **Trees-Structure and Function**, Berlin, v. 21, n. 3, p. 295-310, 2007.

COOPS, N.C.; WULDER, M.A.; CULVENOR, D.S.; ST-ONGE, B. Comparison of forest attributes extracted from fine spatial resolution multispectral and lidar data. **Canadian Journal of Remote Sensing**, Ottawa, v. 30, n. 6, p. 855-866, 2004.

DRAKE, J.B.; BUBAYAH, R.O.; CLARK, D.B. KNOX, R.G.; BLAIR, J.B.; HOFTON, M.A.; CHAZDON, R.L.; WEISHAMPEL, J.F.; PRINCE, S. Estimation of tropical forest structural characteristics using large-footprint lidar. **Remote Sensing of Environment**, Amsterdam, v. 79, n. 2, p. 305-319, 2002.

GOODWIN, N.R.; COOPS, N.C.; CULVENOR, D.S. Assessment of forest structure with airborne LiDAR and the effects of platform altitude. **Remote Sensing of Environment**, Amsterdam, v. 103, n. 2, p. 140-152, 2006.

GREGOIRE, T.G.; STÅHL, G.; NÆSSET, E.; GOBAKKEN, T.; NELSON, R.; HOLM, S. Model-assisted estimation of biomass in a LiDAR sample survey in Hedmark County, Norway. **Canadian Journal of Forest Research**, Ottawa, v. 41, n. 1, p. 83-95, 2011.

HUDAK, A.T.; CROOKSTON, N.L.; EVANS, J.S.; FALKOWSKI, M.J.; SMITH, A.M.; GESSLER, P.E.; MORGAN, P. Regression modeling and mapping of coniferous forest basal area and tree density from discrete-return lidar and multispectral satellite data. **Canadian Journal of Remote Sensing**, Ottawa, v. 32, n. 2, p. 126-138, 2006.

HUDAK, A.T.; CROOKSTON, N.L.; EVANS, J.S.; FALKOWSKI, M.J. Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. **Remote Sensing of Environment**, Amsterdam, v. 112, n. 5, p. 2232-2245, 2008.

HYYPPÄ, J.; INKINEN, M. Detecting and estimating attributes for single trees using laser scanner. **The Photogrammetric Journal of Finland**, Aalto, v. 16, n. 2, p. 27-42, 1999.

JENSEN, J.L.R.; HUMES, K.S.; CONNER, T.; WILLIAMS, C.J.; DEGROOT, J. Estimation of biophysical characteristics for highly variable mixed-conifer stands using small-footprint lidar. **Canadian Journal of Forest Research**, Ottawa, v. 36, n. 5, p. 1129-1138, 2006.

KAARTINEN, H.; HYYPPA, J.; YU, X.; VASTARANTA, M.; HYYPPA, H.; KUKKO, A.; HOLOPAINEN, M.; HEIPKE, M.H.; MORSDORF, F.; NÆSSET, E.; PITKÄNEN, J.; POPESCU, S.; SOLBERG, S.; WOLF, B.M.; WU, J.C. An international comparison of individual tree detection and extraction using airborne laser scanning. **Remote Sensing**, Basel, v. 4, n. 4, p. 950-974, 2012.

LATIFI, H.; NOTHDURFT, A.; KOCH, B. Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: application of multiple optical/LiDAR-derived predictors. **Forestry**, London, v. 83, n. 4, p. 395-407, 2010.

LEFSKY, M.A.; HUDAK, A.T.; COHEN, W.B.; ACKER, S.A. Patterns of covariance between forest stand and canopy structure in the Pacific Northwest. **Remote Sensing of Environment**, Amsterdam, v. 95, n. 4, p. 517-531, 2005.

LI, Y.; ANDERSEN, H.E.; MCGAUGHEY, R. A comparison of statistical methods for estimating forest biomass from light detection and ranging data. **Western Journal of Applied Forestry**, Bethesda, v. 23, n. 4, p. 223-231, 2008.

LOVELL, J.L.; JUPP, D.L.; CULVENOR, D.S.; COOPS, N.C. Using airborne and ground-based ranging lidar to measure canopy structure in Australian forests. **Canadian Journal of Remote Sensing**, Ottawa, v. 29, n. 5, p. 607-622, 2003.

MACLEAN, G.A.; KRABILL, W.B. Gross-merchantable timber volume estimation using an airborne LIDAR system. **Canadian Journal of Remote Sensing**, Ottawa, v. 12, p. 7-18, 1986.

MENG, X.; CURRIT, N.; ZHAO, K. Ground filtering algorithms for airborne LiDAR data: a review of critical issues. **Remote Sensing**, Basel, v. 2, n. 3, p. 833-860, 2010.

NÆSSET, E. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. **Remote Sensing of Environment**, Amsterdam, v. 80, n. 1, p. 88-99, 2002.

_____. Effects of different sensors, flying altitudes, and pulse repetition frequencies on forest canopy metrics and biophysical stand properties derived from small-footprint airborne laser data. **Remote Sensing of Environment**, Amsterdam, v. 113, n. 1, p. 148-159, 2009.

NÆSSET, E.; BJERKNES, K. O. Estimating tree heights and number of stems in young forest stands using airborne laser scanner data. **Remote Sensing of Environment**, Amsterdam, v. 78, n. 3, p. 328-340, 2001. ISSN 0034-4257. Disponível em: < <Go to ISI>://WOS:000172633700011 >.

NELSON, R. How did we get here? An early history of forestry lidar1. **Canadian Journal of Remote Sensing**, Ottawa, v. 39, suppl. 1, p. S6-S17, 2013.

NILSSON, M. Estimation of tree heights and stand volume using an airborne lidar system. **Remote Sensing of Environment**, Amsterdam, v. 56, n. 1, p. 1-7, 1996.

PARKER, G.G.; HARMON, M.E.; LEFSKY, M.A.; CHEN, J.; VAN PELT, R.; WEIS, S.B.; THOMAS, S.C.; WINNER, W.E.; SHAW, D.C.; FRANKLING, J.F. Three-dimensional structure of an old-growth Pseudotsuga-Tsuga canopy and its

implications for radiation balance, microclimate, and gas exchange. **Ecosystems**, New York, v. 7, n. 5, p. 440-453, 2004.

PERSSON, A.; HOLMGREN, J.; SODERMAN, U. Detecting and measuring individual trees using an airborne laser scanner. **Photogrammetric Engineering and Remote Sensing**, Bethesda, v. 68, n. 9, p. 925-932, 2002.

PITKÄNEN, J.; MALTAMO, M.; HYYPPA, J.; YU, X. Adaptive methods for individual tree detection on airborne laser based canopy height model. **International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences**, Amsterdam, v. 36, n. 8, p. 187-191, 2004.

POPESCU, S.C.; WYNNE, R.H.; NELSON, R.F. Estimating plot-level tree heights with lidar: local filtering with a canopy-height based variable window size. **Computers and Electronics in Agriculture**, Amsterdam, v. 37, n. 1/3, p. 71-95, 2002.

SILVA, A.G.P. **Estimativa da biomassa de lenho em povoamentos de *Eucalyptus grandis* baseada em estatísticas do perfil de dossel geradas por escaneamento a laser aerotransportado.** 2014. 146 p. Dissertação (Mestrado em Recursos Florestais) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba.

SIMARD, M.; PINTO, N.; FISHER, J.B.; BACCINI, A. Mapping forest canopy height globally with spaceborne lidar. **Journal of Geophysical Research: Biogeosciences**, New York, v. 116, p. 1-12, 2011.

STEPHENS, P.R.; KIMBERLEY, M.O.; BEETS, P.N.; PAUL, T.S.H.; SEARLES, N.; BELL, A.; BRACK, C.; BROADLEY, J. Airborne scanning LiDAR in a double sampling forest carbon inventory. **Remote Sensing of Environment**, Amsterdam, v. 117, p. 348-357, 2012.

ST-ONGE, B.; TREITZ, P.; WULDER, M.A. Tree and canopy height estimation with scanning lidar. In: WULDER, M and FRANKLIN, S. R. (Ed.). **Remote sensing of forest environments.** New York: Springer, 2003. p. 489-509. 2003.

SULLIVAN, A.A. **LIDAR based delineation in forest stands.** 2008. 90 p. Dissertation (Master of Science in Forest Resources - College of Forest Resources, University of Washington, Washington, 2008.

VASTARANTA, M. **Forest mapping and monitoring using active 3D remote sensing.** 2012. 45 p. Thesis (Doctorate in Forest Science) – Faculty of Agriculture and Forestry, University of Helsinki, Helsinki, 2012.

VAUHKONEN, J.; ENE, L.; GUPTA, S.; HEINZEL, J.; HOLMGREN, J.; PITKÄNEN, J.; SOLBERG, S.; WANG, Y.; WEINACKER, H.; HAUGLIN, K.M.; LIEN, V.; PACKALÉN, P.; GOBAKKEN, T.; KOCH, B.; NÆSSET, E.; TOKOLO, T.; MALTAMO, M. Comparative testing of single-tree detection algorithms under different types of forest. **Forestry**, London, v. 85, n. 1, p. 27-40, 2012.

WHITE, J.C. ; WULDER, M.A.; VARHOLA, A.; VASTARANTA, M.; COOPS, N.C.; COOK, B.D.; PITT, D.; WOODS, M. A best practices guide for generating forest

inventory attributes from airborne laser scanning data using an area-based approach. **Forestry Chronicle**, Ottawa, v. 89, n. 6, p. 722-723, 2013.

ZIMBLE, D.A.; EVANS, D.L.; CARLSON, G.C.; PARKER, R.C.; GRADO, S.C.; GERARD, P.D. Characterizing vertical forest structure using small-footprint airborne LiDAR. **Remote Sensing of Environment**, Amsterdam, v. 87, n. 2/3, p. 171-182, 2003. ISSN 0034-4257. Disponível em: < <Go to ISI>://WOS:000186447400005 >.

ZONETE, M.F. **Análise do uso da tecnologia laser aerotransportado para inventários florestais em plantios clonais de *Eucalyptus sp.* no sul da Bahia**. 2009. 95 p. Dissertação (Mestrado em Recursos Florestais) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2009.

2 ACQUIRING MORE INFORMATIVE METRICS THROUGH THE OPTIMIZATION OF THRESHOLD HEIGHTS WHEN USING AIRBORNE LASER SCANNING TO ESTIMATE THE STAND VOLUME OF A EUCALYPTUS PLANTATION.

"We can easily forgive a child who is afraid of the dark.
The real tragedy of life is when men are afraid of the light"
Plato

Abstract

Applications of airborne laser scanning (ALS) are quickly moving from research to the operational status. ALS-derived canopy height and density metrics provide means of assessing forest attributes by modelling techniques. This study investigates the influence of threshold height during ALS metrics extraction. The forest used is formed by *Eucalyptus* plantations which age varied between 2 and 8 years old. The threshold heights could refer to minimum height and establishes a lower bound to height metrics computation, or refer to height break which establishes a reference to canopy density metrics calculation. Predictive modelling was repeated for several stand ages changing these parameters when processing ALS metrics, observing the differences in terms of metrics set quality. The results indicate that higher threshold height may return a better set of metrics. The impact of threshold height was more evident in young stands and for canopy density metrics. It is essential to separate the returns from canopy to below canopy aiming tree volume and biomass modelling, and the biological aspect should be preferred during threshold height definition.

Keywords: Minimum height; Height break; Modelling; AIC

2.1 Introduction

The main approach towards estimating forest attributes from airborne laser scanning (ALS) datasets is to try and find relationships between metrics derived from laser returns and stand attributes measured in the field (NÆSSET, 1997). ALS metrics are statistics that distill the information available in cloud points assessed from a specific area (e.g. a forest stand, or a sample plot). A variety of modelling techniques relate forest attributes to specific ALS-based metrics (VALBUENA et al., 2013; HUDAK et al., 2008; HYYPPA et al., 2012; NÆSSET, 2002), in a so-called area-based approach (NÆSSET, 2002). It has become the most widespread approach in ALS-assisted forest assessments, in the main due to its flexibility, since it can yield

satisfactory results whilst at the same time dealing with a wide range of vegetation types.

After discounting the elevation of each laser return by the underlying terrain, the ALS returns that correspond to the georeferenced field plot are usually clipped out and the metrics computed. The most used ALS-metrics are: location and dispersion measures defined as canopy height metrics (mean, standard deviation, percentiles, etc) (MAGNUSSEN; BOUDEWYN, 1998), and proportions among vertical strata defined as density metrics (NÆSSET, 1997; NÆSSET, 2002; KORHONEN et al., 2011).

During the extraction of ALS metrics two threshold heights (*ThresH*) are usually used. Minimum height (*MinH*) thresholds are commonly specified as lower bounds to calculate height metrics. In other cases heights are defined as breaks (*HBreak*) or limits to separate the cloud into two sets, as when the canopy returns are separated from the under canopy group. The *MinH* and *HBreak* may have the same value, or they can be set independently. A wide range of values can be applied, especially for mature forests where the majority of the returns are from biological material located some meters above ground (NÆSSET, 2011).

A *MinH* usually defines points that do not match a certain criterion. Three types of approaches for setting that criterion can be found in the literature. The first approach selects a threshold value which splits the cloud into below-canopy and canopy returns, i.e. *MinH* eliminates points not related to canopy. The two meters value, for instance, was introduced by Nilsson (1996) as a minimum height threshold and is now one of the most common choices used to split the canopy (NÆSSET, 2011; MONTAGHI et al., 2013). Another common threshold is the 1.3 m value separating below-canopy and canopy returns. A second scenario is when the *MinH* is used to exclude elements not considered as necessary for particular field measurements, as established in the work of Zimble (2003) who adopted a *MinH* of 5.7m, corresponding to the lower limit of commercial viability as expressed by the average height of the smallest measured trees. Jensen and Humes (2006) used a *MinH* of 1.3 m arguing that field measurements do not consider trees below that height. The third scenario deals with noise from ground classification errors and aims to eliminate returns from boulders, stones, dead wood and other non-relevant elements. The choices here are based on visual inspection or empirical decisions and usually converge to heights as low as 0.5

meters (NÆSSET and BJERKNES, 2001), 0.3 meters (GARCIA et al., 2010), or even 0.17 meters (HUDAK et al., 2006).

HBreaks splits the cloud into two sets in order to calculate proportions (canopy density metrics). The threshold may be fixed, or otherwise defined by a function of other metrics, e.g. mean, mode or percentiles (NÆSSET 2002) defined specifically for the whole study area (MCGAUGHEY 2013). Sullivan (2008) applied a fixed *HBreak* of 2 meters, emulating the eye of a human observer inside the forest with the intention of comparing the resulting data to traditional canopy density measurements. Flaherty considered a fixed *HBreak* of 1.3 m. Andersen (2005) defined the *HBreak* with regard to the height of the canopy base (FLAHERTY; LURZ; PATENAUDE, 2012). A mixed approach can be followed via the definition of fixed intervals relative to the maximum return elevation, a metric which can change among forest areas (NÆSSET, 2011).

The objective of this research was to investigate the influence of *ThresH* , when *ThresH* is defined as both (or either) *MinH* and *HBreak* during ALS metrics extraction from a set of *Eucalyptus* plantations located in Brazil. Predictive modelling was repeated for several stand ages, altering the processing *ThresH*, and observing the qualitative differences of the resulting statistical models.

2.2 Material and Methods

2.2.1 Field and ALS data

The study encompassed five different sites within an industrial eucalyptus plantation managed by a large pulp mill company located in the Eastern region of the State of Sao Paulo, in Brazil. The climate is subtropical, with an annual mean temperature of 21ºC. July is the coldest month with temperature averaging 17ºC and February the hottest with temperatures varying around 24ºC. The annual rainfall is 1300 mm with wet and hot summers and dry and cold winters. The plantations consist of a mixture of hybrid eucalyptus clones obtained via the crossing of E. urophyla and E. gradis. The planting density results in a land occupancy of 1666 trees per hectare (3 x 2 meters). Age per stand ranges from 2 to 8 years old (Table 3.1).

Table 3.1- Summary of stands characteristics

Age (yr)	Basal Area ($m^2.ha^{-1}$)	Height (m)	Dominant Height (m)	Volume ($m^3.ha^{-1}$)
2	7.9 (2.5 – 12.0)	12.1 (5.1 – 14.8)	13.4	46.3 (6.9 – 87.8)
3	20.8 (16.6 – 23.6)	18.6 (16.1 – 20.7)	20.1	173.8 (117.3 – 217.5)
5	27.2 (20.1 – 31.6)	21.5 (18.8 – 23.0)	24.2	266.5 (200 – 337.8)
6	27.3 (24.2 – 30.6)	23.7 (22.0 – 25.3)	26.5	292.1 (241.7 - 344)
8	30.2 (26 – 36.1)	25.2 (19.8 – 31.8)	30.2	362.3 (277.7 – 452.8)

The circular field plots were georeferenced and established with an area of 400m². All tree diameters were measured at breast height using a calliper. Tree heights were determined by an electronic clinometer for 15% of trees within a plot. Heights of unmeasured trees were estimated using locally adjusted allometric models (same age and genetic material). Next, tree volumes were determined by applying the respective diameter and height into a volumetric model. Finally, plot stand volume was calculated and transformed into volume per hectare. All of the field measurements, and the allometric and the volumetric models were supplied by the local foresters in charge of forest inventory operations.

The ALS survey was conducted in April 2012, and produced clouds with an average density of 5 points·m-². The laser returns were filtered and classified as ground and non-ground using the interpolation algorithm proposed by Kraus and Pfeifer (1998), considering a moving window equal to 6 meters and nine interactions. A digital terrain model (DTM) with a 1-m cell size was created which assigned to each cell the average elevation of the corresponding ground returns. The ALS heights were normalized by the respective DTM to obtain a ground-based cloud.

2.2.2 Metrics extraction

After clipping out the georeferenced plots, both the height metrics and the canopy density metrics were extracted using a particular threshold height t (*MinH* for height metrics and *HBreak* for canopy density metrics) which systematically increased from zero in intervals of 0.5 meters. The height metrics included in this study were minimum, maximum, mean, mode, standard deviation, variance, skewness, kurtosis, absolute average deviation, median absolute deviation from median, median absolute deviation

from mode, L-moments, percentiles (HP05 up to HP99), canopy relief ratio, quadratic mean, cubic mean (HCM). The canopy density metrics consisted of the percentage of first returns above *HBreak*, the percentage of all returns above *HBreak* and the ratio all returns above *HBreak* divided by the total of first returns. The ALS metrics were processed in FUSION (MCGAUGHEY 2013).

For a specific t, each extracted metric m was regressed to the stand volume and the Akaike Information Criteria ($AIC(m_t)$) was calculated to measure the relative quality of the statistical model (AKAIKE, 1974; BOZDOGAN, 2000). The importance of a specific *t* was determined by the cumulative Akaike Information Criteria ($cAIC_t$) and calculated as the sum of the respective AICs obtained individually for each height or canopy density metric (m_t):

$$cAIC_t = \sum AIC(m_t) \quad (2.1)$$

The sets were evaluated based on their loss of information with respect to the best t subset. The relative increase in cumulative Akaike Information Criteria (Δ_t) for each set of metrics derived for t was performed with respect to the most parsimonious model ($\min(cAIC_t)$) and was evaluated by:

$$\Delta_t = cAIC_t - \min(cAIC_t) \quad (2.2)$$

2.3 Results

Five canopy profiles were analyzed for each stand age (Figure 2.1). The typical bimodal curve commonly present in most ALS data is observed clearly in the five age classes. The first peak represents the ground returns and the second peak represents the crown returns. The proportion of ground returns was around 10% (8-12%) and the canopy returns ranged from 9 to 15%.

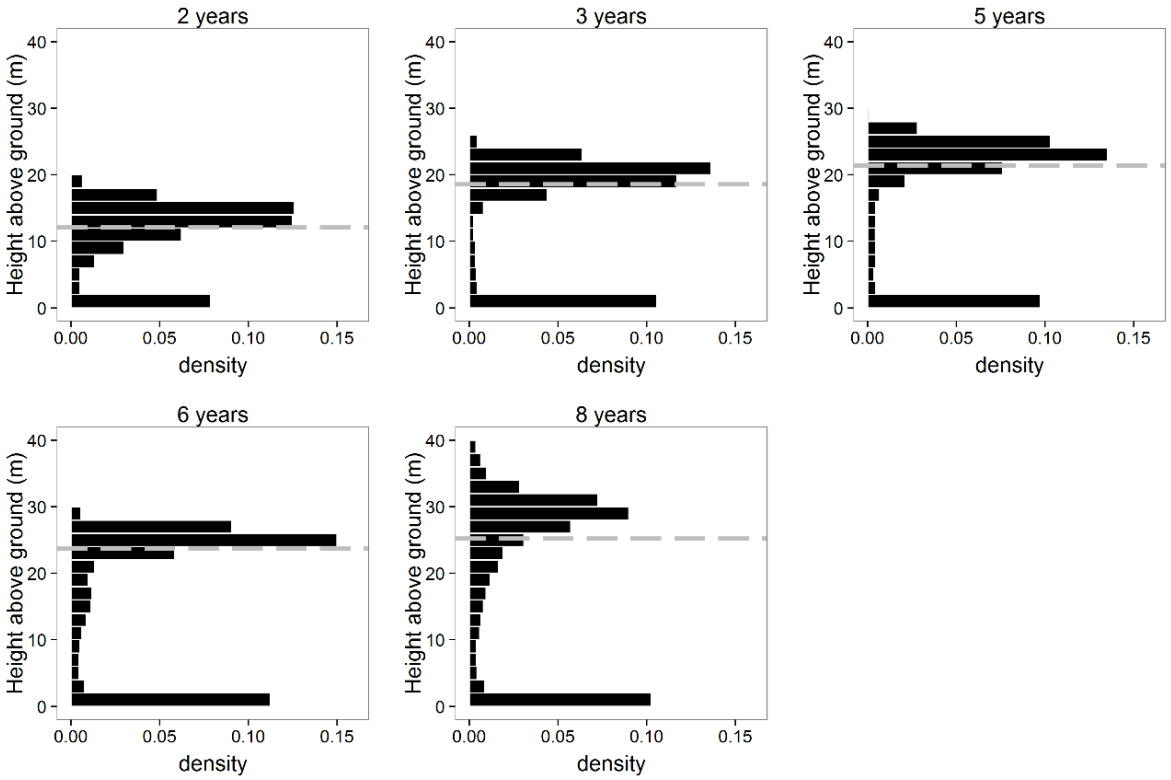

Figure 2.1 - Average vertical profile for 2, 3, 5, 6 and 8 year old stands. Heights above ground in meters (range: 0 – 40) and density in absolute values (range: 0 - 0.15). Dashed gray line indicates observed stand mean height (in meters)

Threshold height analysis was carried out by plotting Δ_t (equation 2.2) for the increasing *ThresH* values. In order to allow direct comparison across all study sites in Table 2.1, the *ThresH* values were expressed relative to the dominant stand height of each site. The change in $\Delta_t$was thus observed when compared against the ratio between each *ThresH* value and the dominant stand height. The optimal *ThresH*, which generates the best set of height metrics was selected as that one with the lowest Δ_t value (zero).

2.3.1 Minimum height

The optimal value to *MinH* varied among stand ages from between 60% to 80% of the dominant height, where 80% of the dominant height was used as the upper bound (i.e. the cutoff level) for the *MinH* variation (Figures 2.2). For all ages, Δ_t was higher for lower values of *MinH* and decreased when *MinH* rose. For age classes 2 and 5, the optimal *MinH* value was approximately equal, but slightly lower than the

upper bound of the interval (9 to 19 meters). For age class 6, the minimum heights varied significantly and presented a very distinctive curve shape.

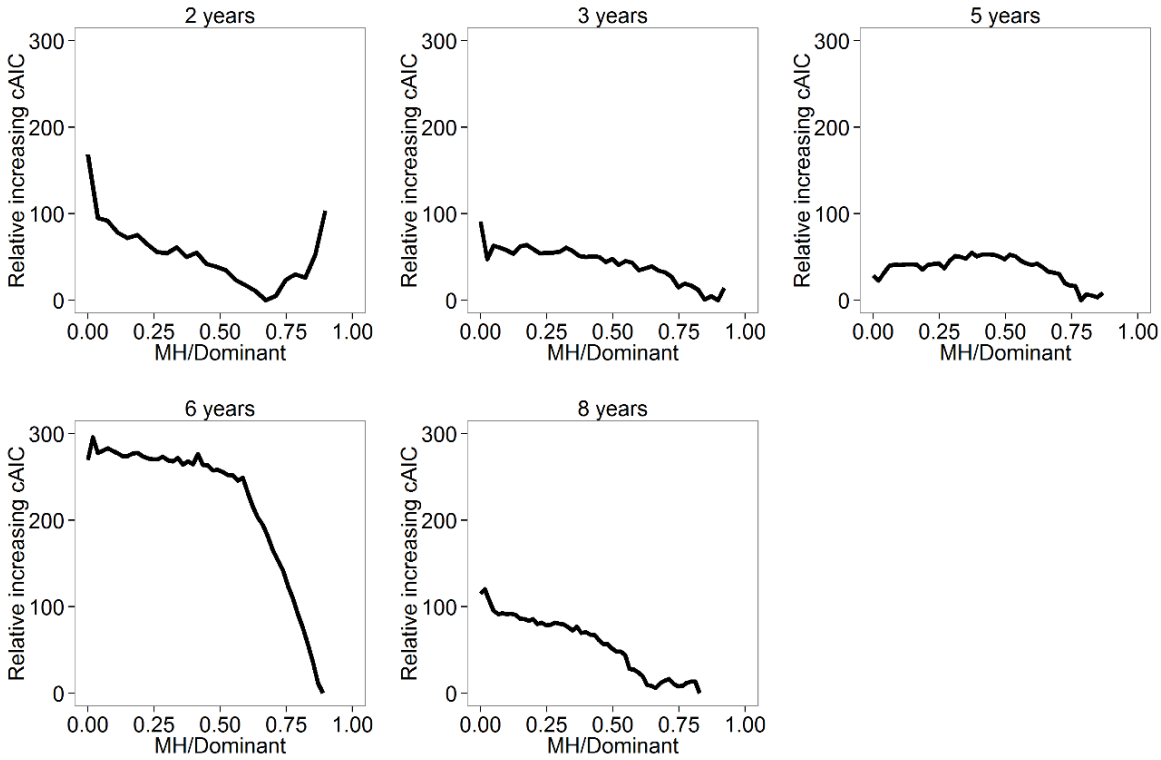

Figure 2.2 - The relative increasing of cumulated Akaike Information Criteria (Δ_t) plotted against the ratio between the MinH value and the dominant stand height. The zero value for Δ_t indicates the best MinH in terms of dominant height

2.3.2 Height break

Younger stands showed larger variation in Δ_t among *HBreak* values (Figure 2.3). The optimal *HBreak* value (considering the studied age classes) was located close to its upper bound, i.e. the mean stand height. The improvements related to *HBreak* were much higher in young stands than in mature stands (Figure 2.3). Δ_t decreased by age class. The difference between the best and worst canopy density metrics was close to 40 units for the $cAIC_t$, in comparison to the 5 units observed in the 8 year old age class.

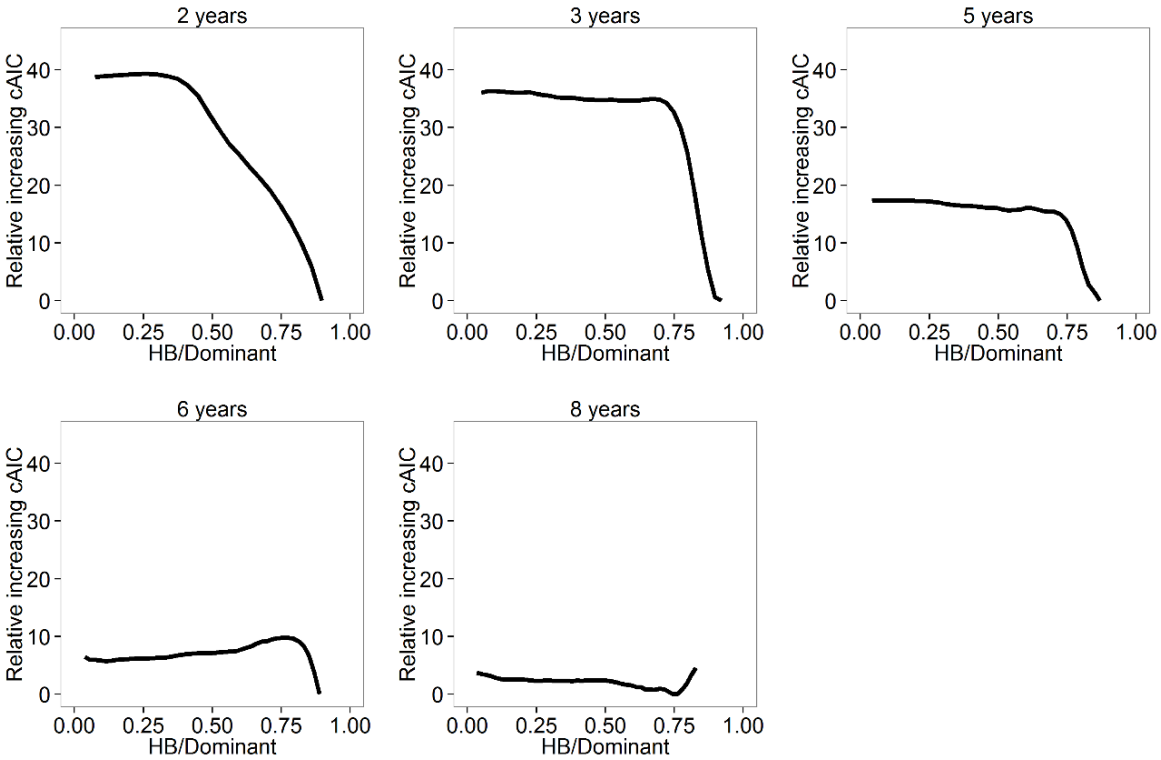

Figure 2.3 - The relative increase of cumulated Akaike Information Criteria (Δ_t) plotted against the ratio between the HBreak value and the dominant height. The null value to Δ_t indicates the best HBreak in the dominant height perspective

2.4 Discussion

Few studies exist that explore the impact of the use of different height thresholds on calculation of ALS metrics. Previous studies on *ThresH* were not able to assess clearly the nature of its influence on LiDAR metrics extraction and the ability of those metrics to explain the variability of forest attributes. Whilst investigating *ThresH* values (0.5, 1.3 and 2 meters) for biomass quantification of three young temperate forests, Næsset found in one of them a slight tendency towards a lesser proportion of the variance being explained, and a higher mean square error as *ThresH* increases (NÆSSET, 2011). In his approach, the best model for each studied *ThresH* was determined using stepwise regression analysis. He concluded that all of the tested threshold values generated models which performed equally well in estimating biomass and suggested that high values for the threshold height may discard useful information, whilst very low values may introduce noise into the canopy data set. The optimal solution would be to consider a compromise between these two criteria: eliminating noise without losing explanatory power.

Indeed, the curves constructed for many of the studied age class showed a stable $cAIC$ for lower *ThresH* values. However, when considering the entire studied range the best set of height metrics showed AIC values 100 units higher than the least helpful set of metrics. For the canopy density metrics, the best results scored a maximum of 40 AIC units. This suggests that for ALS metrics the higher *ThresH* returned a better set of metrics.

Canopy density metrics are indicative of structural features, they are thus of more use when estimating the basal area, diameter and number of trees, in contrast to volume and biomass where height metrics are usually more effective (MALTAMO et al., 2006; NÆSSET, 2002; ZONETE; RODRIGUEZ; PACKALEN, 2010; GONCALVES-SECO et al., 2011). This may explain why the greater influence of the *ThresH* value in relation to volume when calculating height metrics (*MinH*), compared to a lesser influence when calculating canopy density metrics (*HBreak*). For mature stands, a wide range of suitable *HBreak* are possible, in contrast to young stands where unique values may have a larger impact on the metrics quality (NÆSSET, 2011).

A useful and relevant technique concerning *MinH* is to exclude the low part of the cloud, changing the usual bimodal distribution (one peak of ground and one peak of canopy – Figure 2.1) to a uni-modal distribution. This modification is especially useful in the modeling of canopy profiles since normal distributions (e.g. Weibull) typically indicate uni-modal curves (COOPS et al., 2007).

Caution should be attached to any extrapolation of our results towards the assessments of forests with more densely vegetated vertical structures, since the understory does not form a significant proportion of overall volume in a plantation management regime. ALS points backscattered from below the canopy have little explanatory power for stand volume. Another perspective is that in the absence of an understory, the number of returns backscattered from tree trunks can differ greatly according to the scanning angle. The elimination of those returns could be one of the reasons for the improved effectiveness of metrics found in this study.

2.5 Conclusions

This study helps to expand understanding on the influence of *ThresH* choice in metrics extraction. Better metrics can be obtained if echoes from the canopy are

separated from those emanating from below the canopy. This separation is highly recommendable, above all for young stands, but still and becoming more flexible in mature plantations. This conclusion is of special importance to particular applications where the understory has not yet been assessed, such as the estimation of tree volumes.

Acknowledgments

This research was funded by CAPES by a student grant; ForEAdapt (FP7-PEOPLE-2009-IRSES) and supported with a visit to UEF/Finland for the processing of data and for a discussion of the results. Fibria SA with ALS data sets, and field data. We also wish to thank Simon Dunster for his help and advice regarding the presentation of this document in English.

References

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, New York, v. 19, n. 6, p. 716-723, 1974.

ANDERSEN, H.E.; MCGAUGHEY, R.J.; REUTEBUCH, S.E. Estimating forest canopy fuel parameters using LIDAR data. **Remote Sensing of Environment**, Amsterdam, v. 94, n. 4, p. 441-449, 2005.

BOZDOGAN, H. Akaike's Information Criterion and Recent Developments in Information Complexity. **Journal of Mathematical Psychology**, Amsterdam, v. 44, n. 1, p. 62-91, 2000. ISSN 0022-2496. Disponível em: < http://www.sciencedirect.com/science/article/pii/S0022249699912774 >.

COOPS, N.C.; HILKER, T.; WULDER, M.A.; ST-ONGE, B.; NEWNHAM, G.; SIGGINS, A.; TROFYMOW, J.T. Estimating canopy structure of Douglas-fir forest stands from discrete-return LiDAR. **Trees-Structure and Function**, Berlin, v. 21, n. 3, p. 295-310, 2007.

FLAHERTY, S.; LURZ, P.W.W.; PATENAUDE, G. Implementation of a general linear model using LiDAR derived explanatory variables: a case study in Scotland. In: SPIE 8531, Remote Sensing for Agriculture, Ecosystems, and Hydrology XIV, 2012, Edinburgh. **Proceedings…** Edinburgh: SPIE, 2012. p. 1-28.

GARCÍA, M.; RIAÑO, D.; CHUVIECO, E.; DANSON, F.M. Estimating biomass carbon stocks for a Mediterranean forest in central Spain using LiDAR height and intensity data. **Remote Sensing of Environment**, Amsterdam, v. 114, n. 4, p. 816-830, 2010.

GONÇALVES-SECO, L.; GONZÁLEZ-FERREIRO, E.; DIÉGUEZ-ARANDA, U.; FRAGA-BUGALLO, B.; CRECENTE, R.; MIRANDA, D. Assessing the attributes of high-density *Eucalyptus globulus* stands using airborne laser scanner data. **International Journal of Remote Sensing**, London, v. 32, n. 24, p. 9821-9841, 2011.

HUDAK, A.T.; CROOKSTON, N.L.; EVANS, J.S.; FALKOWSKI, M.J.; SMITH, A.M.; GESSLER, P.E.; MORGAN, P. Regression modeling and mapping of coniferous forest basal area and tree density from discrete-return lidar and multispectral satellite data. **Canadian Journal of Remote Sensing**, Ottawa, v. 32, n. 2, p. 126-138, 2006.

HUDAK, A.T.; CROOKSTON, N.L.; EVANS, J.S.; HALL, D.E.; FALKOWSKI, M.J. Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. **Remote Sensing of Environment**, Amsterdam, v. 112, n. 5, p. 2232-2245, 2008.

HYYPPÄ, J.; YU, X.; HYYPPÄ, H.; VASTARANTA, M.; HOLOPAINEN, M.; KUKKO, A.; KAARTINEN, H.; JAAKKOLA, A.; VAAJA, M.; KOSKINEN, J.; ALHO, P. Advances in forest inventory using airborne laser scanning. **Remote Sensing**, Basel, v. 4, n. 5, p. 1190-1207, 2012.

JENSEN, J.L.R.; HUMES, K.S.; CONNER, T.; WILLIAMS, C.J.; DEGROOT, J. Estimation of biophysical characteristics for highly variable mixed-conifer stands using small-footprint lidar. **Canadian Journal of Forest Research**, Ottawa, v. 36, n. 5, p. 1129-1138, 2006.

KORHONEN, L.; KORPELA, I.; HEISKANEN, J.; MALTAMO, M. Airborne discrete-return LIDAR data in the estimation of vertical canopy cover, angular canopy closure and leaf area index. **Remote Sensing of Environment**, Amsterdam, v. 115, n. 4, p. 1065-1080, 2011.

KRAUS, K.; PFEIFER, N. Determination of terrain models in wooded areas with airborne laser scanner data. **ISPRS Journal of Photogrammetry and Remote Sensing**, Amsterdam, v. 53, n. 4, p. 193-203, 1998.

MAGNUSSEN, S.; BOUDEWYN, P. Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators. **Canadian Journal of Forest Research**, Ottawa, v. 28, n. 7, p. 1016-1031, 1998.

MALTAMO, M.; EERIKÄINEN, K.; PACKALÉN, P.; HYYPPÄ, J. Estimation of stem volume using laser scanning-based canopy height metrics. **Forestry**, London, v. 79, n. 2, p. 217-229, 2006.

MCGAUGHEY, R. J. **FUSION/LDV:** software for LiDAR Data Analysis and Visualization. Seattle: USDA, Forest Service, 2013. v. 3, 171 p.

MONTAGHI, A.; CORONA, P.; DALPONTE, M.; GIANELLE, D.; CHIRICI, G.; OLSSON, H. Airborne laser scanning of forest resources: An overview of research in Italy as a commentary case study. **International Journal of Applied Earth Observation and Geoinformation**, Amsterdam, v. 23, p. 288-300, 2013.

NÆSSET, E. Estimating timber volume of forest stands using airborne laser scanner data. **Remote Sensing of Environment**, Amsterdam, v. 61, n. 2, p. 246-253, 1997.

_____. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. **Remote Sensing of Environment**, Amsterdam, v. 80, n. 1, p. 88-99, 2002.

_____. Estimating above-ground biomass in young forests with airborne laser scanning. **International Journal of Remote Sensing**, London, v. 32, n. 2, p. 473-501, 2011.

NÆSSET, E.; BJERKNES, K.O. Estimating tree heights and number of stems in young forest stands using airborne laser scanner data. **Remote Sensing of Environment**, Amsterdam, v. 78, n. 3, p. 328-340, 2001.

NILSSON, M. Estimation of tree heights and stand volume using an airborne lidar system. **Remote Sensing of Environment**, Amsterdam, v. 56, n. 1, p. 1-7, 1996.

SULLIVAN, A.A. **LIDAR based delineation in forest stands.** 2008. 90 p. Dissertation (Master of Science in Forest Resources) – College of Forest Resources, University of Washington, Washington, 2008.

VALBUENA, R.; MALTAMO, M.; MARTÍN-FERNÁNDEZ, S.; PACKALEN, P.; PASCUAL, C.; NABUURS, G. J. Patterns of covariance between airborne laser scanning metrics and Lorenz curve descriptors of tree size inequality. **Canadian Journal of Remote Sensing**, Ottawa, v. 39, n. S1, p. S1-S14, 2013.

ZIMBLE, D.A.; EVANS, D.L.; CARLSON, G.C.; PARKER, R.C.; GRADO, S.C.; GERARD, P.D. Characterizing vertical forest structure using small-footprint airborne LiDAR. **Remote Sensing of Environment**, Amsterdam, v. 87, n. 2/3, p. 171-182, 2003.

ZONETE, M.F.; RODRIGUEZ, L.C.; PACKALEN, P. An estimate of biometric parameters in eucalyptus clone plantations in Southern Bahia: an application of the airborne laser scanning (ALS) technology. **Scientia Forestalis**, Piracicaba, v. 38, n. 86, p. 225-235, 2010.

3 DEVELOPMENT OF STAND VOLUME MODELS BASED ON STABLE METRICS FROM MULTIPLE ALS ACQUISITIONS IN EUCALYPTUS PLANTATIONS

"What we do for ourselves dies with us.
What we do for other and the world,
remains and is immortal"
Mason Albert Pike

Abstract

Some problems limit the ability to transfer predictive relations obtained from ALS point cloud descriptive metrics obtained from non-standardized flight missions, greatly limiting the benefits of forest ALS monitoring assessments. To mitigate this problem and reduce multi-temporal ALS assessments discrepancies, studies are needed to identify the most resilient set of metrics that produce stable and reliable estimates. The aim of this paper is to identify stable metrics derived from different ALS data sets, allowing them to be used as an independent variable when modeling stand volume, and thus assisting the management of Eucalyptus forest plantations. Three consecutive ALS missions were carried out at the same site on a particular day, but using different equipment and flight settings. Canopy height and density metrics were calculated for the ALS data sets. A sequence of statistical analyses were implemented in order to assess the variability and relevance of metrics extracted from the three datasets. The stability of ALS metrics differed at plot and grid levels. The selection of stable metrics can generate reliable models between different data sets. According to our results, the height metrics provide the greatest stability when used in the models, specifically the higher percentiles (>50%) and the mode. According to our results, some plausible stability can be achieved when the model relies more strongly on canopy height metrics than on density metrics.

Keywords: Light detection and ranging (LiDAR); Vegetation mapping; Monitoring; Forest measurement; Returns density

3.1 Introduction

A goal of remote sensing in forestry is the reliable prediction of key vegetation parameters from data gathered using aircraft and satellite-borne instruments. In recent years airborne laser scanning (ALS), otherwise referred to as light detection and ranging (LiDAR), has emerged as the leading technology for the construction of high resolution three-dimensional maps (LIM; HOPKINSON; TREITZ, 2008; MONTAGHI et al., 2013).

The technology works by sending out short pulses of laser light downwards from an airborne platform (airplane). Part of the light is reflected back to the platform from features on the ground, with the return time giving the distance from the platform to the feature. At the moment each laser pulse is sent out, the position of the platform, obtained from an on-board GPS unit, is logged together with the direction of propagation of the light pulse from a highly sensitive inertial measurement unit attached to the beam steering optics (BALTSAVIAS, 1999a, 1999b). In practice a single outgoing pulse may generate multiple returns. This arises when there are partially transmitting strata of objects, such as the leaves and branches of trees, above the ground. Reflected pulses will come from these strata as well as from the light, which reaches and is reflected from the ground. The raw vertical heights of the points in the 3D cloud are relative to a geo-fixed level. These heights can be normalized by subtracting the altitude of the ground (available from the pulses reflected at the ground). This post-processing of the ALS data yields a ground-referenced cloud showing the positions, heights and shapes of partially transmitting objects, such as trees. It is these unique capabilities, which have made ALS a useful remote sensing technology for forestry monitoring, and a potential source of vegetation structure information.

Since the number of points in an ALS cloud is very high, for the evaluation of forest vegetation, it is usual to reduce the point cloud over each plot to a set of metrics. These metrics are statistics that describe the distribution of the ALS returns. Each return is basically characterized by a ground position, a height above the ground, and an intensity. The point height records for a plot generate a discrete distribution of number of returns for different heights above the ground. The characterization of this distribution is achieved through height metrics of location (for example: mean and percentile boundaries), height metrics of dispersion and form (for example: standard deviation and skewness), and density metrics (for example: the fraction of points in the distribution with heights greater than the mean) (DALPONTE et al., 2009; NÆSSET, 2009; GONCALVES-SECO et al., 2011; VAUHKONEN; MEHTATALO; PACKALEN, 2011).

The distribution of return intensities, relative to the intensity of the incident pulse, for a plot can also be extracted from an ALS data set, and summarized by metrics analogous to those used to describe the height distribution. However, many factors influence the intensity of a return, making it hard to obtain reproducible distributions of

return intensities. Therefore, intensity metrics were not computed or analyzed in this study.

The content of an ALS data set is as much influenced by the laser source, the sensor parameters, and the flight settings as by the features present on the ground. Considering a given location on the ground, the number of pulses incident on unit area perpendicular to the vertical (the pulse density) depends on the height and speed of the airborne platform over the ground, the pulse repetition frequency, the angles defining the path of the incident pulses and the speeds at which the laser scanning optics change these angles. All of these factors are mathematically defined, so that the incident pulse density can be calculated (BALTSAVIAS, 1999a).

Forests are considered semipermeable for the laser beam. Trees are spatially complex objects capable of generating many returns from each pulse, thus predicting which of these many returns will be logged by an ALS sensor is extremely challenging and dependent on many factors. One return point will be recorded when the energy reflected back to the sensor has power enough to trigger the detection algorithm (CHASMER et al., 2006). In order to receive a useful signal from higher-flying heights, both the transmitted power and the dimensions of the receiver optics need to be increased, and the beam divergence decreased (BALTSAVIAS, 1999). It is important to mention that the capacity to register a return point is not reliant solely upon the ALS equipment, but also upon the object. Object reflectivity thus has a fundamental role concerning the quality of the signal.

The width of the full area to be surveyed by ALS is usually greater than the swath of a single pass, which is determined by the platform height and the limits of the angular sweep provided by a single cycle of the beam steering optics. The ALS data set from a single survey is thus normally a combination of several, laterally displaced, flights over the survey area, with overlap between the strips scanned on successive passes (REUTEBUCH; ANDERSEN; MCGAUGHEY, 2005; MALTAMO et al., 2012).

Differences in cloud properties (e.g. point density, pulse density, overlap areas) and between raw data sets collected at different times are extremely common, if not inevitable. These differences in survey results over time have implications that must be considered carefully. The most frequent causes of these differences include the use of different instruments and modifications to the flight settings. A straightforward approach in reducing these impacts is to adjust the model parameters using new field data after a particular ALS survey. Keeping these flight settings constant, or collecting

a new set of field data will substantially impact upon the cost of surveys. Ideally, metrics describing the height distribution of returns should not vary in relation to the flight settings.

In this study, we examined data sets corresponding to consecutive scans of the same Eucalyptus plantation carried out on the same day using different settings for the equipment and flight. Our aim was to identify stable metrics that could serve as independent variables in stand volume modelling, the final objective being to assist Eucalyptus plantation managers.

3.2 Material and Methods

The study area is located at the geographical coordinates (22º 58' 04'' S, 48º 43' 40'' W), which lies within the municipality of Itatinga in the state of São Paulo, Brazil. It comprises a plantation of 7-year-old *Eucalyptus grandis*, covering an area of 198.9 ha. The tree spacing was 3.75 m x 1.60 m, resulting in an area of 6 m² per tree. The terrain displays gentle undulations (height difference between the highest and lowest points = 57 m), with an average elevation of 750 m above sea level (CAMPOE et al., 2012). The mean annual rainfall was 1391 mm, with 75% concentrated from October to March. Mean annual temperature was 19.2 °C, ranging from 13.3 °C in the coldest months (June to August) to 27.2 °C in the warmest months (December to February) (ALVARES et al., 2013).

The study site was formed of evenly planted trees from superior genetic material; the trees were of the same age and the mortality rate was low. These forest plantation conditions make it reasonable to assume a linear relationship between stand volume and metrics expressing the height of the trees. We found that fitting linear regression models between ALS metrics and stand volume resulted in statistically significant coefficients and models.

The field data were acquired in July 2009. We established 23 rectangular plots with areas ranging from 500 to 900 m²; the limits of each plot were marked by painting the trees which corresponded to the boundaries. The height of each tree within a given plot was measured with a hypsometer Haglof HEC-R for height and the diameter at 1.30 m above the ground with a diameter tape. The timber volume for the tree was then estimated from its height and diameter using a formula developed by DURATEX S.A. (forest manager) for Eucalyptus grandis established in this region of Brazil. The

variable of interest, stand volume (total timber volume per unit area), was calculated as the total estimated timber volume for the trees of the plot divided by the respective area. Uniform clear-cutting of the entire plantation occurred a few days after the tree height and diameter measurements were completed. The geographical positions of the vertices of each plot could then be determined by differential GPS, with no risk of crown interference. At each vertex, GPS signals were logged using a roving receiver with an external antenna (Trimble Pro XR), these records were post-processed with correction data retrieved from the fixed base station in Presidente Prudente (station number 93900, 22º 07' 09.9679'' S, 51° 24' 28.9700'' W and altitude of 435.40 meters) operated by the Brazilian Institute of Geography and Statistics (IBGE) to give final plot corner positions. Location statistics of the stand volume over the 23 rectangular plots are presented in Table 3.1. The mean volume of the study plots is equal to 410.6 m3 ha-1, the minimum volume is 278 and the maximum is 536.6 m³ ha-1, indicating that the stand volume is an important parameter for be estimated against the ALS metrics.

Table 3.1 – Location statistics of the stand volume over 23 plots ($m^3 ha^{-1}$)

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
278.0	354.2	402.4	410.6	471.9	536.6

Three ALS data sets (DS) were acquired on April 2009, with a Leica ALS50-II carried on an EMB 810 C SENECA II aircraft with slightly different settings (Table 3.2). Each data set contains the results from a series of parallel sweeps over the study area, the following parameters were the same for every flight: flight speed (140 km h-1), half angle (±7.5°), scan rate (74 Hz). The lateral displacement of successive sweeps over the study area was set to give 30% overlap.

Table 3.2 – Settings between the three data sets (DS)

	DS1	DS2	DS3
Flight height (m)	2485	1658	900
PRF (Hz)	100.7	137.3	109.7
Laser Power (%)	77	54	28
Number of strips	4	10	22
Point density average	2.8	5.6	8.7

The altitude of the terrain beneath each point was subtracted from the altitude of the point. The digital elevation model required for this normalization was obtained to a horizontal resolution of 1 m from the ground returns of each data set. Each georeferenced field plot was subsequently extracted and its respective metrics calculated. The corresponding metrics extracted from plot level were also calculated for a grid consisting of 20-meter square cells, each encompassing a single data set.

We calculated the following height metrics (HM) from the point height data: mean, quadratic mean, cubic mean, and mode (measures of central tendency); percentiles: 10, 25, 50, 75, 90, 95, and 99; and standard deviation, coefficient of variation, skewness, and kurtosis (measures of dispersion). These 15 metrics were calculated from the height values of all the points above 1 meter in each cloud, and from the heights of the subset of points corresponding to first returns, giving a total of 30 height metrics for each plot and data set. The density metrics (DM) are ratios (expressed as percentages) of a relation between the total number of points, and the numbers of points satisfying the following height-value criteria: greater than 2 m, greater than the mean for the plot, and greater than the mode for the plot. We generated a total of 12 density metrics for each plot and data set. The complete list of metrics can be found in Appendix.

The best model for estimating stand volume based on the metrics calculated from all 23 plots was determined for each particular data set. The correlation matrix between metrics was calculated, and all metrics with a correlation greater than 90% were grouped together. The metric that was most highly correlated with the dependent variable stand volume was chosen to represent the group. All of these metrics were then used to perform a regression analysis on the stand volume. Every possible combination was investigated via limiting the regression size up to three independent variables. The best overall regression, and the best regression that included one high metric and one density metric, were then selected through Akaike Information Criteria (AIC) and the adjusted coefficient of determination (adjR2).

After determination of the best models considering all metrics for each data set, we started to study the stability of the metrics amongst the data sets. Initially, we compared the similarity of height metrics at plot level amongst the three data sets for all returns and for the first returns. The same comparison was then performed on density metrics, comparing again amongst the three data sets.

We are interested in how the variability in the four groups of metrics (height metrics and density metrics for all returns and first returns) can be explained by each ALS data set (DS1, DS2 and DS3 as described in Table 3.2). Because there are multiple dependent variables that cannot be combined, a multivariate analysis of variance (MANOVA) was chosen. The null hypothesis in this analysis is that different acquisition settings have no effect on any of the four types of metrics.

A MANOVA was used to search for evidence of significant differences in the sub-vectors of plot metrics describing the data sets. We divided the vector of plot metrics into four sub-vectors: all return height metrics (HM | AR), first return height metrics (HM | FR), all return density metrics (DM | AR), and first return density metrics (DM | FR), and applied MANOVA to each sub-vector separately.

For each sub-vector, and assuming that (i) there are q dependent variables in the MANOVA; (ii) all hypothesis tests may be performed on matrix A = HE-1, with H denoting the hypothetical sums of squares and cross products matrix and E denoting the error sums of squares and cross products; and (iii) letting λ_i denote the ith eigenvalue of matrix A, a Pillai's trace can be computed as:

$$Pillai's\ trace = trace[\mathrm{H}(\mathrm{H}+\mathrm{E})^{-1}] = \sum_{i=1}^{q} \frac{\lambda_i}{1+\lambda_i} \quad (3.1)$$

The *Pillai's trace* is one of the four multivariate criteria test statistics used in MANOVA and is considered the most powerful and robust for testing the null hypothesis. In combination with the appropriate numbers of degrees of freedom, each *Pillai´s trace* yielded a p-value from which the significance of the data set effect on the components of the sub-vector was assessed.

The similarity between a set of metrics at plot level does not guarantee stability for the entire study area, even when the plots are an asymptotically unbiased sample of the population. The ALS data sets, though, covers the entire area returning an dense cloud point sample from the population, capturing much more details and resulting in differences not captured at the plot level. Therefore, the interest on investigating the stability amongst data sets via a comparison of the metrics extracted from 20-meter cell grids.

A paired *Kolmogorov-Smirnov test* (KS) determined which metrics had non-significant differences in the distribution between data sets. Non-significant differences

between the possible paired comparisons identified the stable metrics. The best model for estimating stand volume was fitted again as described before, however this time only the stable metrics were used.

The best models obtained for each data set based on non-discriminated metrics and the best models using the most stable metrics were then compared and evaluated according to the adjusted coefficient of determination (adjR2), Akaike Information Criteria (AIC), and the relative root mean square error (rRMSE).

3.3 Results

The groups of correlated metrics varied widely between the data sets. Metrics derived from DS1 and DS2 were grouped into 7 groups, and DS3 into 9 groups. The group members showed some interesting differences however. Whereas HSD, HCV, HSDf and HCVf metrics formed a separate group in DS1, they were grouped with the other height metrics for DS2. The metric HP10f was also included in an isolated group for DS2. When analyzing the groups formed for DS3, the metric HP25 went to an isolated group, HP10f remained isolated as in DS2 and HSD, HCV, HSDf and HCVf metrics formed a separate group as in DS1. For all of them, the density metrics were similarly grouped: PARAMO and PFRAMO together, PARA2 and PFRA2 together and another group consisting of PARAM and PFRAM (Table 3.3).

Table 3.3 - Groups formed by data set considering all metrics extracted. In bold, the highest correlated metrics to volume representing the group. The metrics are identified in APPENDIX, and the data sets in Table 3.2

	DS1	**DS2**	**DS3**
Group 1	HM, HMO, HP25, HP50, HP75, HP90, HP95, HP99, HSQ, HC, HMf, HMOf, HP10f, HP25f, **HP50f**, HP75f, HP90f, HP95f, HP99f, HSQf, HCf	HM, HMO, HP25, HP50, HP75, HP90, HP95, HP99, HSQ, HC, HMf, HMOf, **HP25f**, HP50f, HP75f, HP90f, HP95f, HP99f, HSQf, HCf, HSD, HCV, HSDf, HCVf	HM, HMO, HP25, HP50, HP75, HP90, HP95, HP99, HSQ, HC, HMf, HMOf, HP10f, **HP25f**, HP50f, HP75f, HP90f, HP95f, HP99f, HSQf, HCf
Group 2	**HSD**, HCV, HSDf, HCVf	**HS**, HK, HSf, HKf	**HSD**, HCV, HSDf, HCVf
Group 3	HS, **HK**, HSf, HKf	**PARA2**, PFRA2	**HS**, HK, HSf, HKf
Group 4	**PARA2**, PFRA2	**PARAM**, PFRAM	**PARA2**, PFRA2
Group 5	**PARAM**, PFRAM	**PARAMO**, PFRAMO	**PARAM**, PFRAM
Group 6	**PARAMO**, PFRAMO	**HP10**	**PARAMO**, PFRAMO
Group 7	**HP10**	**HP10f**	**HP10**
Group 8			**HP25**
Group 9			**HP10f**

Similar to group formation, the metrics which represented each group also differed between the data sets. The metrics representing the groups formed for DS1 were HP10f, HSD, HK, PARA2, PARAM, PARAMO and HP10. For DS2 the metrics were: HP25f, HS, PARA2, PARAM, PARAMO, HP10 and HP10f. Finally, for DS3, the representative metrics were HP25f, HSD, HS, PARA2, PARAM, PARAMO, HP10, HP10f, HP25.

The *Pillai's trace* testing of the metrics for sensitivity to the data set was carried out comparing three ALS data sets. Examining first the height metrics (HM), when these were computed from all the returns (HM | AR), the metrics showed no significant differences between data sets, with p-values of 0.161. The height metrics for the first returns (HM | FR) displayed weakly significant differences (p-value = 0.038). The density metrics (DM) computed both from all the returns (DM | AR) and from the first returns (DM | FR) were significantly different between the ALS data sets (Table 3.4).

Table 3.4 – Results of Pillai's Trace (Λ) used to test for significant differences in the metrics for plot level between data sets, followed by computation of an approximate F-ratio (Fapprox) and a p-value to establish significance. The metrics are identified in APPENDIX, and the data sets in Table 3.2

	Group of Metrics			
	Height Metrics		Density Metrics	
	HM | AR (15 metrics)	HM | FR (15 metrics)	DM | AR (6 metrics)	DM | FR (6 metrics)
Λ	0.5427	0.6288	0.7871	0.8248
F_{approx}	1.3077	1.6202	4.2542	4.6014
***p*-value**	0.161^{ns}	0.0385	7e-07**	1e-07**

^{ns} No significant differences (*p*-value > 0.1) and ** Highly significant differences (*p*-value < 0.001)

The *Kolmogorov-Smirnov test* provided a statistical test of equivalence of one metric at a time between pairs of data sets. From 30 HM and 12 DM including all and first returns, just 10 HM and 1 DM displayed non-significant differences of cumulative probability distribution between data sets at a confidence level of 95% (Table 3.5).

Table 3.5 – P-value for the non-significant metrics by the Kolmogorov Smirnov test comparing two-by-two distribution metrics extracted for a grid by 20x20m superimposed to the study area. The metrics are identified in APPENDIX, and the data sets (DS) in Table 3.2

	HMO	**HP50**	**HP75**	**HP90**	**HP95**	**HP99**	**HMOf**	**HP90f**	**HP95f**	**HP99f**	**PARAMO**
DS1 x DS2	0.25	0.08	0.02	0.03	0.06	0.19	0.25	0.03	0.03	0.1	0.07
DS2 x DS3	0.74	0.81	0.56	0.32	0.2	0.3	0.55	0.31	0.24	0.24	0.33
DS1 x DS3	0.45	0.05	0.48	0.72	0.98	0.99	0.62	0.59	0.93	0.99	0.01

If considering only the non-significant metrics established by a KS test between DS, the groups formed and the representative metrics are the same. The 11 stable metrics between data sets were split into two groups. One containing the height metrics (HMO, HP50, HP75, HP90, HP95, HP99, HMOf, HP90f, HP95f and HP99f) and the second group containing the density metric PARAMO (Table 3.6).

Table 3.6 – Groups formed by data set considering just non-significant metrics under KS test. In bold the highest correlated metrics to volume representing the group. The metrics are identified in APPENDIX, and the data sets in Table 3.2

	DS1	**DS2**	**DS3**
Group 1	HMO, **HP50**, HP75, HP90, HP95, HP99, HMOf, HP90f, HP95f, HP99f		
Group 2	**PARAMO**		

While considering all 42 metrics (first and all returns) for both the grouping of metrics and for the evaluation the best potential model, we did not find any consistency between the best model per data set. The best model for DS1 was built with HP50f and HP10 metrics. For DS2, the best model had one unique metric, the HP25f, and for the DS3 the stand volume was regressed against HP25f and HSD. Between the 11 stable metrics, the metrics that make up the best model were consistently the same the HP50 (Table 3.7).

Table 3.7 – The best model to estimate stand volume based on ALS metrics. The metrics are identified in APPENDIX, and the data sets in Table 3.2

DS | Metrics	Best model	R^2adj	AIC	rRMSE
DS1 | all metrics	V = HP50f + HP10	0.78	224.7	7.51
DS2 | all metrics	V = HP25f	0.81	220.4	7.14
DS3 | all metrics	V = HP25f + HSD	0.79	223.4	7.31
DS1 | stable metrics	V = HP50	0.77	224.4	7.8
DS2 | stable metrics	V = HP50	0.78	223.8	7.69
DS3 | stable metrics	V = HP50	0.76	225.41	7.97

When investigating the scenario where the model must be based in one height metric and one density metric, and where all of the 42 metrics are eligible, the best model for DS1 consisted of the metrics HP50f and PARA2. The best model for DS2 consisted of the metrics HP25f and PARAM. The best model for DS3 consisted of the metrics HP25f and PARA2. If narrowing the eligible metrics to solely the stable group, the best model was the same across all data sets, and was formed by metrics: HP50 and PARAMO (Table 3.8).

Table 3.8 – The best model with one height metrics and one density metric to estimate stand volume based on stable ALS metrics. The metrics are identified in APPENDIX, and the data sets in Table 3.2

DS | Metrics	Best model	R^2adj	AIC	rRMSE
DS1 | stable metrics	V = HP50 + PARAMO	0.76	226.1	7.74
DS2 | stable metrics	V = HP50 + PARAMO	0.77	225.3	7.61
DS3 | stable metrics	V = HP50 + PARAMO	0.75	227.2	7.93

3.4 Discussion

Three data sets obtained for the same area in the same day, but employing different equipment configurations and different flight settings, generated slightly different metrics. These differences appeared both in the correlation between metrics, as well as in the correlation to the dependent variable (stand volume). The implication here is that the best model in estimating stand volume also differs between the data sets.

Thomas et al. (2006) found different models for each ALS data set comparing the coefficient of determination and the root mean square error. The best model to estimate stem volume by their low-density data set (0.035 pts m^{-2}) used the first returns percentile 50 and the best model to their high-density (4 pts m^{-2}) data set used the all returns percentile 50. Gonzalez-Ferreiro et al. (2012) have also explored the selection of height metrics for the prediction of stand volumes from different LiDAR data sets. The best model for their 0.5 pts m^2 data set used as predictors the height metrics: standard deviation, skewness and the 5% percentile. In contrast, the stand volume model for the 8 pts m^{-2} data set did not include the standard deviation.

The modelling of stand volume was based on multiple linear regression, based upon an earlier investigation in which both the linear and nonlinear models were individually adjusted in relation to the stand volume and to the metrics. The test for linearity was found to be significant, and the nonlinear models (quadratic and cubic) did not increase the explained variance significantly. The forest type studied can explain this linear behaviour: homogeneous and fast growing stands, with same age.

A *Pillai's Trace test* revealed that the density metrics (for both cases: all returns and first returns) differed between the data sets, which did not occur when we performed the test with the height metrics. This high sensitivity is reflected in the absence of density metrics among the best models selected for each data set (Table 3.7).

The density metrics were highly sensitive to the point density, due probably to the fact that as altitude increases footprints also increase, returns densities decrease and irradiance per unit area decreases. Therefore, as the flight altitude varies, full waves may be trigged and decomposed into returns differently from flight to flight. Nonetheless, Bater et al. (2011), when testing the reproducibility of ALS metrics with multiple scans over the same location taken on the same day, found that the majority

of the height metrics were not significantly different. However, the differences for the density metrics were highly significant, and arose from large differences in point density between the four flight lines. They argued that the observed differences in point density found during their campaign might have been caused by unintentional variations in the orientation and altitude of the platform.

Application of the *Kolmogorov-Smirnov test* to compare cell metric values between pairs of data sets showed statistically significant differences between the data sets for most metrics. The exceptions were the high height percentiles, the mode of the heights and the proportion of all returns above mode, which displayed statistical similarity between all the data sets.

Once the candidate metrics as explanatory variable for stand volume had been limited to metrics that were considered stable, similar models were determined for the data sets, both in respect to performance and to selected metrics. Even equipment configurations and flight parameters that resulted in a higher return density did not show any significant differences. Thomas et al. (2006) also found that an increase in point density does not necessarily improve the prediction accuracy of mean dominant height, basal area, crown closure, and average aboveground biomass.

Even when restricting the candidate metrics to just the stable group, the model performances are still very close (Table 3.7 and Table 3.8). One particular height metric and one particular density metric were imposed onto the model in order to simulate the biological condition, concerning the site quality and the stand density respectively. The models even in this scenario maintained a similar level of performance to models based upon all of the metrics.

Foody et al. (2003) explored the problems associated with attempts to transfer predictive models of vegetation towards combined uses with remote sensing data, and demonstrated that there exist many complicating factors. The inability to transfer these data relationships satisfactorily seriously limits the contribution that remote sensing can make in environmental applications. Our study has shown that some metrics do nevertheless have the desirable property of returning similar values across different data sets, and that some post-processing of data sets can improve model transferability. Once variance in airborne laser scanner sensors and flight settings are considered, the scope for metric values to change in response to properties unrelated to the vegetation being surveyed increases (MAGNUSSON; FRANSSON; HOLMGREN, 2007), and so also do the difficulties in the search for transferable

models. A straightforward approach towards consistently producing good models is to always use field data when resetting model parameters after a LiDAR data survey. However, a more low-cost method over the long term would be the discovery and development of transferable models. This would require stable metrics, such as those, which show the greatest similarity between different data sets and different methods of remote data gathering.

Even the ALS sensors have evolved over the last decades, the biomass quantification remains holding in the metrics correlation to biophysical parameters. The use of stable metrics can increase the compatibility between surveys, making the results comparable and allowing a historical series studies.

This study investigated candidate metrics to form transferable models based on ALS-derived metrics derived from multiple surveys of data collected at the same site on a particular date. We have investigated the stability of metrics at both the plot and grid levels, and have demonstrated how the selection of stable metrics can contribute to generate reliable models between different data sets. According to our results, the height metrics provide the greatest stability when used in the models, specifically the higher percentiles (>50%) and the mode.

Acknowledgments

This paper is part of the research program developed by GET-LiDAR. The EUCFLUX project, Duratex S.A., the Brazilian Forestry Science and Research Institute (IPEF), North Carolina State University (NCSU), and Virginia Polytechnic Institute and State University (VT) are acknowledged for their support of this research through the provision of the laser and field data sets. We thank the anonymous reviewers for their insightful suggestions. The Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES, supported this study via a student grant and ForEAdapt (FP7-PEOPLE-2009-IRSES) supported this work through a visit to UEF/Finland during data processing.

References

ALVARES, C. A.; STAPE, J. L.; SENTELHAS, P. C.; GONÇALVES, J. L. M.; SPAVOREK, G. Koppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, Sttutgart, v. 22, n. 6, p. 711-728, 2013.

BALTSAVIAS, E.P. Airborne laser scanning: basic relations and formulas. **ISPRS Journal of Photogrammetry and Remote Sensing**, Amsterdam, v. 54, n. 2/3, p. 199-214, 1999a.

_____. Airborne laser scanning: existing systems and firms and other resources. **ISPRS Journal of Photogrammetry and Remote Sensing**, Amsterdam, v. 54, n. 2/3, p. 164-198, 1999b.

BATER, C.W.; WULDER, M.A.; COOPS, N.C.; NELSON, R.F.; HILKER, T.; NÆSSET, E. Stability of sample-based scanning-LiDAR-derived vegetation metrics for forest monitoring. **IEEE Transactions on Geoscience and Remote Sensing**, New York, v. 49, n. 6, p. 2385-2392, 2011.

CAMPOE, O.C.; STAPE, J.L.; LACLAU, J.P.; MARSDEN, C.; NOUVELLON, Y. Stand-level patterns of carbon fluxes and partitioning in a *Eucalyptus grandis* plantation across a gradient of productivity, in São Paulo State, Brazil. **Tree Physiology**, Oxford, v. 32, n. 6, p. 696-706, 2012.

DALPONTE, M.; COOPS, N.C.; BRUZZONE, L.; GIANELLE, D. Analysis on the use of multiple returns LiDAR data for the estimation of tree stems volume. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, New York, v. 2, n. 4, p. 310-318, 2009.

FOODY, G.M.; BOYD, D.S.; CUTLER, M.E.J. Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. **Remote Sensing of Environment**, Amsterdam, v. 85, n. 4, p. 463-474, 2003.

GONÇALVES-SECO, L.; GONZÁLEZ-FERREIRO, E.; DIÉGUEZ-ARANDA, U.; FRAGA-BUGALLO, B.; CRECENTE, R.; MIRANDA, D. Assessing the attributes of high-density *Eucalyptus globulus* stands using airborne laser scanner data. **International Journal of Remote Sensing**, London, v. 32, n. 24, p. 9821-9841, 2011.

GONZALEZ-FERREIRO, E.; DIEGUEZ-ARANDA, U.; MIRANDA, D. Estimation of stand variables in *Pinus radiata* D. Don plantations using different LiDAR pulse densities. **Forestry**, London, v. 85, n. 2, p. 281-292, 2012.

LIM, K.; HOPKINSON, C.; TREITZ, P. Examining the effects of sampling point densities on laser canopy height and density metrics. **Forestry Chronicle**, Ottawa, v. 84, n. 6, p. 876-885, 2008.

MAGNUSSON, M.; FRANSSON, J.E.S.; HOLMGREN, J. Effects on estimation accuracy of forest variables using different pulse density of laser data. **Forest Science**, Bethesda, v. 53, n. 6, p. 619-626, 2007.

MALTAMO, M.; MEHTÄTALO, L.; VAUHKONEN, J.; PACKALEN, P. Predicting and calibrating tree attributes by means of airborne laser scanning and field measurements. **Canadian Journal of Forest Research**, Ottawa, v. 42, n. 11, p. 1896-1907, 2012.

MONTAGHI, A.; CORONA, P.; DALPONTE, M.; GIANELLE, D.; CHIRICI, G.; OLSSON, H. Airborne laser scanning of forest resources: An overview of research in Italy as a commentary case study. **International Journal of Applied Earth Observation and Geoinformation**, Amsterdam, v. 23, p. 288-300, 2013.

NÆSSET, E. Effects of different sensors, flying altitudes, and pulse repetition frequencies on forest canopy metrics and biophysical stand properties derived from small-footprint airborne laser data. **Remote Sensing of Environment**, Amsterdam, v. 113, n. 1, p. 148-159, 2009.

REUTEBUCH, S.E.; ANDERSEN, H.E.; MCGAUGHEY, R.J. Light detection and ranging (LIDAR): sn emerging tool for multiple resource inventory. **Journal of Forestry**, Bethesda, v. 103, n. 6, p. 286-292, 2005.

THOMAS, V.; TREITZ, P.; MCCAUGHEY, J.H.; MORRISON, I. Mapping stand-level forest biophysical variables for a mixedwood boreal forest using lidar: an examination of scanning density. **Canadian Journal of Forest Research**, Ottawa, v. 36, n. 1, p. 34-47, 2006.

VAUHKONEN, J.; MEHTÄTALO, L.; PACKALEN, P. Combining tree height samples produced by airborne laser scanning and stand management records to estimate plot volume in Eucalyptus plantations. **Canadian Journal of Forest Research**, Ottawa, v. 41, n. 8, p. 1649-1658, 2011.

4 ANALYSIS OF THE PERFORMANCE OF TWO MACHINE LEARNING METHODS TO ESTIMATE FOREST STAND VOLUME USING AIRBORNE LASER SCANNING DATA

"Fairy tales do not tell children that dragons exist.
Children already know that dragons exist.
Fairy tales tell children that dragons can be killed."
G. K. Chesterton

Abstract

Machine learning models appear to be an attractive route towards tackling high-dimensional problems, particularly within areas where a lack of knowledge exists regarding the development of effective algorithms, and where programs must dynamically adapt to changing conditions. The core objective of this study was to evaluate the performance of two machine learning tools in predicting stand volume, based on statistical vegetation metrics extracted by an Airborne Laser Scanning (ALS) survey. The forests used in this study were composed of 1,138 hectares of commercial plantations that consisted of hybrids of *Eucalyptus grandis* and *Eucalyptus urophylla*, managed for pulp production. Two machine learning tools were implemented, neural network (NN) and random forest (RF), and their performance was compared to a regression model (RM). The RF and the RM presented a RMSE of 31.85 m^3 ha^{-1} (13.1%) and 31.09 m^3 ha^{-1} (12.8%) respectively. The NN presented a higher RMSE than the others, equal to 50.23 m^3 ha^{-1} (20.7%). The ranking of ALS metrics based on their relative importance for the estimation of stand volume differed significantly between the methods used. Rather than being limited to a subset of predictor variables when explaining as much of the variability of the dependent variable as possible, machine learning techniques explored the complete metrics set, looking for patterns between them and the dependent variable.

Keywords: Stand volume; LiDAR; Artificial intelligence; Neural network; Random forest

4.1 Introduction

Remote Sensing (RS) has been used as an efficient assessment tool to monitor large forest areas. RS techniques allow the retrieval of spatial data from the environment like trees, roads, stream flow, and other objects located over the ground surface (ZHOU et al., 2013). The available expertise in multi-spectral image acquisition, processing, interpretation, and its relatively lower cost have resulted in the high use of this method within forest monitoring activities (PRASAD; BRUCE; CHANUSSOT, 2011). However, multi-spectral RS encounters problems when assessing vertical information directly (i.e. incorporating a third dimension) since it performs less impressively in sensing structure under medium to high leaf area conditions. The radiometric interference from the surface slope, the weather conditions, the atmospheric turbidities, and the angles of solar incidence also present problems to multi-spectral RS (PROY; TANRE; DESCHAMPS, 1989; STOJANOVA et al., 2010; PFLUGMACHER; COHEN; KENNEDY, 2012).

Airborne Laser Scanning (ALS) has been employed to generate Digital Elevation Models (DEM) throughout the last 20 years (MONTAGHI et al., 2013). Due to its ability to penetrate the forest canopy, ALS technology has become the primary data source for characterizing vertical forest structure (WHITE et al., 2013), and its use has expanded towards new applications such as monitoring vegetation. Based on Light Detection and Ranging (LiDAR) technology, this sensor provides horizontal and vertical information at high spatial resolution and high vertical accuracies that significantly increase our understanding about the real 3D structure of the forest (NÆSSET, 2004; PATENAUDE et al., 2004). The large amount of data constrains the direct use of ALS as an input to the modelling of forest parameters. After collection, the raw ALS data must subsequently be reduced and represented numerically within the calculation of several spatial ALS-metrics that can then be used to create predictive equations for forestry inventory attributes. The number of metrics can easily reach hundreds of variables, and the selection of these metrics remains an empirical process highly dependent of human intervention. After eliminating all but the most descriptive metrics, forest attributes are then estimated through statistical regression analyses that explore the correlation between field measurements and ALS metrics (NELSON; KRABILL; TONELLI, 1988; NÆSSET, 1997; LEFSKY et al., 1999; NÆSSET; BJERKNES, 2001; REUTEBUCH; ANDERSEN; MCGAUGHEY, 2005; ZHAO;

POPESCU; NELSON, 2009; GLEASON; IM, 2012). Linking ALS-metrics to field data is an effective method for estimating several forest attributes (e.g., stem volume, basal area, biomass, etc.) at the stand or regional level, but there remains a large set of assumptions and site-specific considerations that must be made (ZHAO et al., 2011). In fact, a large number of variables could theoretically improve the precision of the models, but models with fewer variables are much easier to interpret (MURPHY; EVANS; STORFER, 2010). It is thus important to develop parsimonious models, mainly because prediction models should be valid for general conditions, and degrees of freedom should not be unreasonably discarded. Finally, large sets of predictor variables often bear strong inter-correlations, which can lead to unstable predictions (LATIFI; NOTHDURFT; KOCH, 2010).

Three approaches have been used to select metrics and develop regression models with ALS data. One is to adjust models based on empirical pre-established relationships between field data and ALS metrics established by other studies (ZHAO et al., 2011). Another is to determine the best relationship between ALS metrics and field data through optimizing a certain statistical measure (stepwise or exhaustive search) (NÆSSET, 1997; PATENAUDE et al., 2004). A third approach involves the use of multivariate statistical analysis, based on the assumption that multi correlated ALS metrics should be used to estimate forest parameters (VALBUENA et al., 2013).

Within the context of remote sensing forest structures via ALS data, studies have shown that the majority of estimation models are not only site- or species-specific, but also scale-dependent, which indicates that the models should be applied at a scale or pixel (cell of the grid) size commensurate to the plot size used in the model fitting (NÆSSET, 2002). Zhao (2009) demonstrated that is possible to reduce the effects of plot scale by using machine learning tools such as neural network (NN) or random forest (RF).

These machine learning tools are effective when producing software, since they have the ability to tackle high-dimensional problems, poorly understood domains where there is lack of knowledge needed to develop effective algorithms, or domains where programs must dynamically adapt to changing conditions (MITCHELL, 1997). An additional advantage is that machine learning allows the user to implement a continuous learning process. Previous remote sensing studies have shown a superior or promising level of performance by artificial intelligence techniques over more

classical methods (FANG; LIANG; KUUSK, 2003; ATZBERGER, 2004; DURBHA; KING; YOUNAN, 2007; ZHAO; POPESCU; ZHANG, 2008; ZHAO et al., 2011).

The primary objective of this study is to evaluate the performance of two machine learning tools (NN and RF) to stand volume estimation. The secondary objectives are: (1) to compare the performance of AI tools to the usual regression model and (2) to assess the relative importance of ALS metrics through AI.

4.2 Material and Methods

The study area is located in the State of São Paulo, characterized by a mountainous topography, ranging from 589 to 1294 meters above the sea level. The area covers approximately 1,138 hectares. The forest consists of a commercial plantation, the stock being hybrids of *Eucalyptus grandis* W.Hill ex Maiden and *Eucalyptus urophylla* S. T. Blake managed by the Fibria SA to supply a pulp production company. The trees spacing is 3 x 2 meters, resulting in 1666 plants per hectare, with an age range of 2 to 8 years old.

The ground data were collected using 112 georeferenced circular plots of 400 square-meters. The plot volume was determined from the sum of individual trees, based on their diameter, height and specific volume equation linked to the total area.

The LiDAR data were acquired over the 2012 summer by an RIEGL LMS Q860I sensor combined with an Applanix 510 IMU/GPS installed on a Piper Seneca II aircraft. The mean posting density was 5 points per square meters. The main ALS survey parameters are reported in Table 4.1.

Table 4.1 - Parameters of ALS campaign used in this work

Parameter	Unit	Values
Average points density	$pts.m^{-2}$	5
Flight height	M	792
Flight speed	$km.h^{-1}$	148
Pulse rate	kHz	200
Datum	UTM	SIRGAS2000
Year of flight	Year	2012
Overlapping	%	30

In order to retrieve canopy height in reference to the ground, the 1-m resolution digital elevation model derived from the ground returns was used to normalize the point cloud. The understory influence was then eliminated by excluding all echoes below 1-m from the calculation of ALS-metrics (NILSSON, 1996).

The ALS metrics for the corresponding field plots were initially extracted from just the first returns (single and first of many) and then later from all returns. A total of 104 ALS-metrics, divided into either height metrics or canopy density metrics, were extracted for each plot. The ALS-metrics considered in this study were widely used in the existing literature, and have been established as effective for stem volume estimation (LIM; TREITZ, 2004; NÆSSET, 2004; PARKER; RUSS, 2004; HUDAK et al., 2006; NÆSSET; GOBAKKEN, 2008).

The metrics used were: mean of heights (HMEAN), quadratic mean (HSQ), cubic mean (HCUB), mode (HMODE), median (HMEDIAN), median of the absolute deviations from the overall median (HMADIAN), median of the absolute deviations from the overall mode (HMADODE), percentiles of height (HP##), maximum (HMAX), interquartile distance (HIQ), standard deviation (HSD), variance (HVAR), average absolute deviation (HAAD), kurtosis (HKUR), skewness (HSKW), L-moments (HL1, HL2, HL3, HL4, HLSKW, HLKUR), canopy relief ratio (CRR), percentage of all returns above a particular height (ARA2FR, PARA2, ARAMOTFR, PARAM, ARAMO, ARAMTFR, PARAMO, ARAM, TARF), percentage of first returns above a particular height (PFRAMO, PFRA2, PFRAM, FRAM, FRA2, TFRF).

In our experiment, we used to machine learning both based on supervised learning: neural network (NN) and random forest (RF). The underlying goal of artificial intelligence is to perform acceptably (or even optimally) well at a specific task in a particular environment (WHITE, 1989). The learning algorithm used as a target the stand volume from the field plots to adjust the AI tool whilst receiving the ALS metrics as inputs. Usually AI tools are not sensible to collinearity, normality or linearity. RF is not sensitive to the unit dimension as well.

The NN is a mathematical model of the biological neuron. The input layer receives all available ALS metrics and the output layer returns the stand volume. All data involved was standardized (4.1) by:

$$Standard\ Score = \frac{X_i - \bar{X}}{s} \quad (4.1)$$

where $\bar{X}$ = metric mean, x_i = i-metric value, and *s* = metric standard deviation.

The neural network was set up with 30 neurons on the hidden layer. Input and hidden nodes computed the logistic function, and the output node computes linear function. The initial weights were set randomly, and no decay parameter was used. The supervised training of the neural network adjusted the weights by minimizing the squared error (ε) (4.2):

$$\varepsilon = |y - \hat{y}|^2/2 \quad (4.2)$$

where *y* the observed values and $\hat{y}$ is the NN output.

During the NN learning process, the weights were adjusted to return a result as similar as possible to the training set and to dictate the relative influence of the variables. The variable importance in NN was determined by the deconstruction of the model weights as proposed by Goh (1995). A measure of importance was thus obtained for each of the ALS metrics with regard to stand volume.

RF is a combination of a decision tree (i.e., a forest) with a value of a random independently sampled vector, and with same distribution for all trees in the forest (BREIMAN, 2001). RF used a series of binary rule-based decisions to indicate which particular tree is used as a specific data input. During random forest adjustment, if the error associated with splitting a single rule into multiple rules is lower than the error associated with using just a single rule, the regression tree will “branch out” and the tree will grow (GLEASON; IM, 2012).

A script compiled using the statistical software R (R CORE TEAM, 2013) was used to set the RF parameters of random Forest library (LIAW; WIENER, 2002). RF was adjusted using the data from 600 trees, with the minimum size of terminal nodes corresponding to 5. The predictor variables were all available ALS metrics, but in contrary to the neural network, it was not necessary to scale them. During the supervised training, the algorithm adjusted multiple regression trees for different subsamples of the data set. The result is a combination average of all the trees.

We used the notion of node impurity from RF to evaluate the variable importance (CUTLER et al., 2007). Every time a split of a node is made the impurity for the two descendent nodes is less than the parent node. With the addition of a new node, the impurity thus decreases for each individual variable in relation to all of the trees in the

forest and the relative variable importance is thus determined. For regression, the node impurity is measured by the residual sum of squares.

The methods performance was assessed using 'leave-one-out' cross validation during training and the indicators root mean square error (RMSE) and relative RMSE (rRMSE) (4.3) and the bias (4.4) calculated by:

$$rRMSE = \frac{\sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n}}}{\bar{y}} * 100 \quad (4.3)$$

$$bias = \frac{\sum(y_i - \hat{y}_i)}{n} \quad (4.4)$$

where y_i is the observed value, $\hat{y}_i$ is the estimated value and n is the number of observations. We also used two graphic analyses to assess the performance: observed and predicted scatter-plots and bias histograms.

The machine learning methods were compared to a regression model (RM) for volume estimation, adapted from Stephens (2012). Initially, the 104 ALS metrics were grouped if correlation was above 80% in absolute value. For each group, the metric which most strongly correlated to volume was assigned to represent it in further analyses. Finally, an exhaustive search was performed in order to find the best model derived from the reduced set of metrics. The search was limited to models composed by one or two explanatory variables.

4.3 Results

The estimation based on artificial intelligence tools used all 104 metrics extracted from ALS data sets to estimate stand volume, exempting both metrics selection and reduction. The RF and the RM presented an RMSE of 31.85 $m^3\ ha^{-1}$ (13.1%) and 31.09 $m^3\ ha^{-1}$ (12.8%) respectively. The NN showed a higher RMSE then RM, equal to 50.23 $m^3\ ha^{-1}$ (20.7%). The bias calculated from all techniques resulted in a normal distribution, centered on zero (Figure 4.1). However, the RF bias distribution was lightly narrow then RM, and then NN.

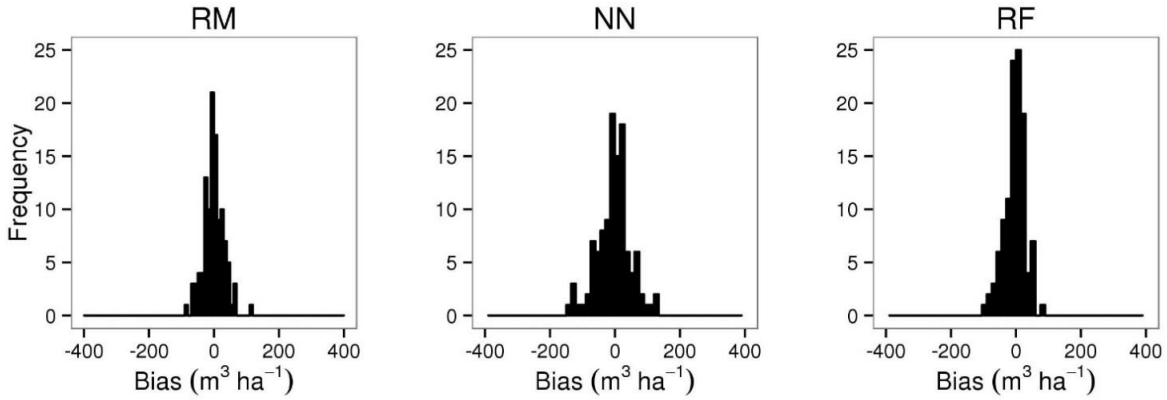

Figure 4.1 - Bias histogram stand volumes as predicted by the ALS-based regression model (RM), the neural network (NN) and the random forest tool (RF) via 'leave one out' cross-validations

Scatter plots of the observed stand volumes against the model estimations display and reinforce the larger deviations observed for NN in comparison to RF and RM (Figure 4.2).

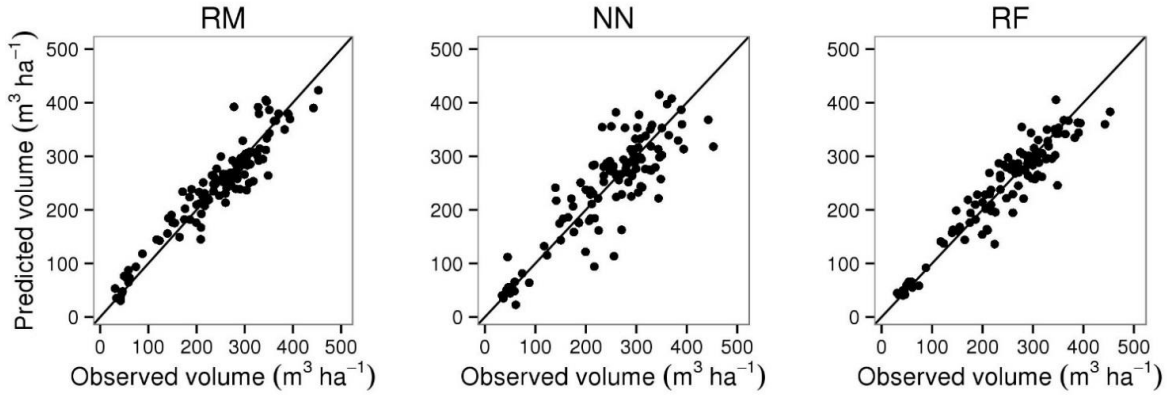

Figure 4.2 - Scatterplot of volumes as predicted by the ALS-based regression model (RM), the neural network (NN) and the random forest tool (RF) via 'leave one out' cross-validations

The 104 ALS metrics were subsequently divided into 15 groups based on the correlation matrix, and the metric most strongly correlated to volume was assigned to represent each one of them (Table 4.2). An exhaustive search combined 15 metrics into every possible combination and then looked for the best model with no more than two explanatory variables. The best model to estimate the stand volume, based on adjusted coefficient of determination, was formed through using the HLSKW and HMODEF metrics together.

Table 4.2 - Co-linearity groupings of ALS metrics with the representative of each group and the correlation to stand volume in parenthesis

Groups	Metrics	Representative
1	HMAX, HKUR, HKURF, HMAXF, HL4, HL4F	HL4F (0.72)
2	HMEAN, HMODE, HL1, HP20, HP25, HP30, HP40, HP50, HP60, HP70, HP75, HP80, HP90, HP95, HP99, HSQ, HCUB, HMEANF, HMODEF, HL1F, HP20F, HP25F, HP30F, HP40F, HP50F, HP60F, HP70F, HP75F, HP80F, HP90F, HP95F, HP99F, HSQF, HCUBF,	HMODEF (0.95)
3	HSD, HVAR, HCV, HAAD, HL2, HSDF, HVARF, HCVF, HAADF, HL2F	HSD (0.75)
4	HIQ, HMADIAN, HLCV, HIQF, HLCVF	HIQ (0.31)
5	HSKW, HSKWF	HSKWF (-0.11)
6	HMADODE, HMADIANF, HMADODEF, HL3F, HL3	HL3 (-0.79)
7	HLSKW	HLSKW (-0.65)
8	HP10F, HLKUR	HP10F (0.44)
9	HP01, HP01F	HP01 (-0.38)
10	HP05, HP05F, HP10	HP10 (0.22)
11	CRR, CRRF	CRRF (0.14)
12	PFRA2, PARA2, ARA2FR, PFRA2F, ARA2FRF, PARA2F	ARA2FR (-0.46)
13	FRA2, ARA2, FRAM, FRAMO, ARAM, ARAMO, FRA2F, ARA2F, FRAMF, FRAMOF, ARAMF, ARAMOF, TFRF, TARF	ARAMO (-0.39)
14	PFRAM, PARAM, ARAMTFR, PFRAMF, PARAMF, ARAMTFRF	PARAMF (0.52)
15	PFRAMO, PARAMO, ARAMOTFR, PFRAMOF, PARAMOF, ARAMOTFRF	ARAMOTFR (-0.36)

The choice of linear model was based on the relation observed between HLSKW, HMODEF and stand volume (Figure 4.3). The bias histogram and the scatterplot did not show either homoscedasticity nor normality deviation.

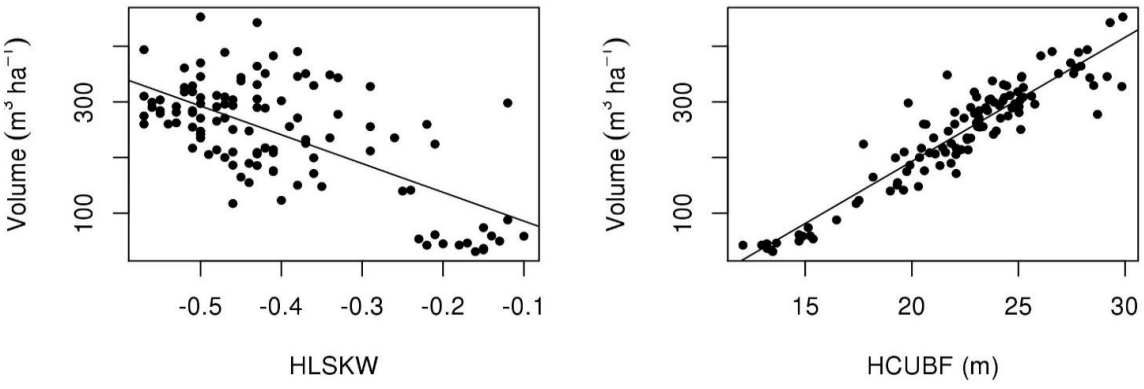

Figure 4.3 - Scatterplot between stand volume and the explanatory metrics HLSKW and HMODEF

The ranking of each ALS metric for relative importance in estimating stand volume differed significantly between artificial intelligence tools (Figure 4.4). The positions of HMODEF and HLSKW, the two best metrics under RM methodology, were

first and 20th respectively in RF rank. For NN, the HMODEF was ranked as low as 18th place and the HLSKW rank was third to last.

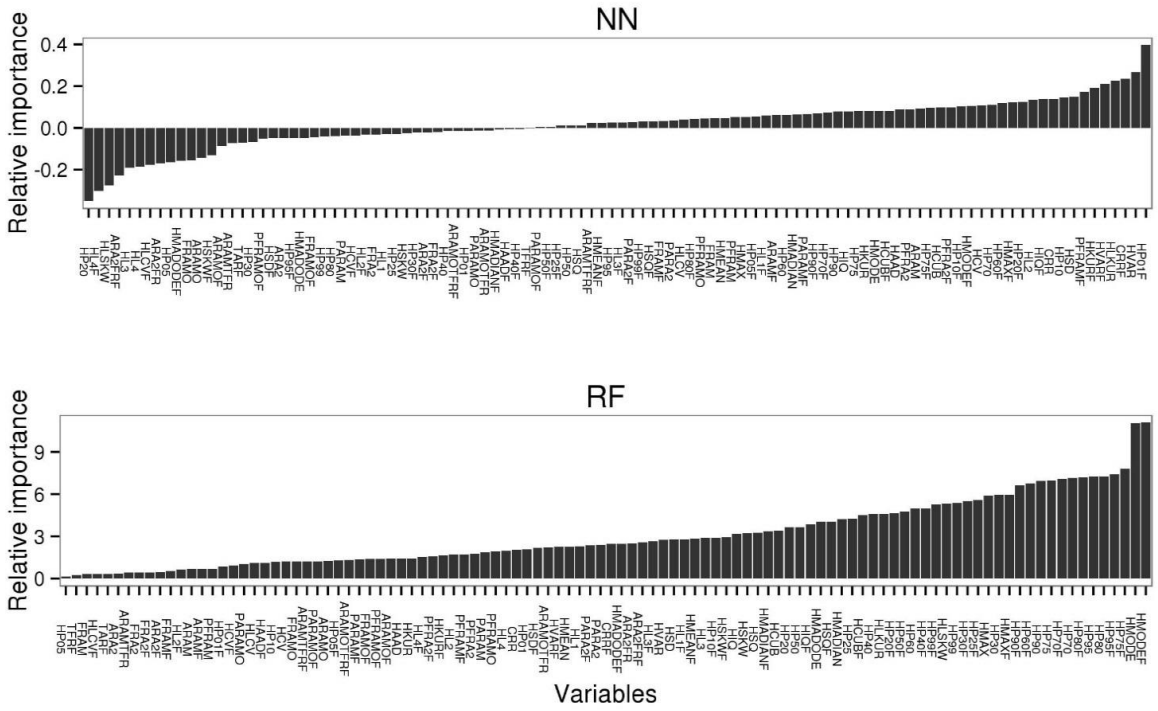

Figure 4 - Relative importance of the metrics within each artificial intelligence technique: neural network (NN) and random forest (RF)

4.4 Discussion

This study examined the performance of two artificial intelligence tools regarding the area-based estimation of stand volume in eucalyptus plantation forests. Their performance was compared to traditional methodology that uses regression models.

The regression models require direct interaction and calibration from the modeler, since these models are built upon relationships and assumptions that are not always repeated in other data sets. During regression parameter estimation, the ordinary least squares (the most used method) minimizes the residual sum of squares, and it will always fit at least as well as, and generally better than, the true mean function. Any two data sets for which these are identical will therefore give the same fitted regression, even if a straight-line model is appropriate for one but not the other. In this study, the relationship of the strongest metrics (HLSKW and HMODEF) was linear with the stand volume and the residue did not show any deviation from normality or from apparent

homoscedasticity. However, several studies showed the need for transformation of the variables in order to meet some of the deviations mentioned above (NÆSSET, 2004).

When the process of selecting and reducing variables is automated, the metrics chosen might not rely on biological assumptions and often differ completely between studies. The stand volume reflects the site quality and the stand density (CLUTTER et al., 1983). Traditional stand volume models are constructed upon the adoption of basal area as a measure of stand density and on dominant height as a measure of site quality. For example, a previous study developed in another eucalyptus plantation established HP90F and HP30F as the best predictors (ZONETE, 2009), which was not found to be the case in this study.

The similarity between RF and the traditional RM method is very interesting from an implementation point of view. The RF does not contain assumptions (i.e. unity dimension, residual normality and homoscedasticity) and this feature allows it to work with all of the metrics without losing any information regarding variable selection and reduction. The NN performance was not as impressive as RF, but nevertheless still displayed interesting properties such as bias centered in zero.

Regression techniques are the most frequently used method to estimate plantation forest attributes from ALS metrics. In a previous study in Brazil, the rRMSE was equal to 11.86% of a model built with HP10, HP50 and PARAH90 (including ground) (PACKALEN; MEHTATALO; MALTAMO, 2011). One studies in Spain found an RMSE of 53.6 m^3 ha-1 to estimate stand volume for Eucalyptus globulus stand. In this study, the model had designated the height corresponding to the third tallest LiDAR-detected trees in the plots as an explanatory variable (GONCALVES-SECO et al., 2011).

The AI tools have been also applied within temperate forests. The ability of artificial intelligence for predicting stand volume was tested for a natural forest in Germany, achieving an rRMSE of 23.3% in RF (LATIFI; NOTHDURFT; KOCH, 2010). When studying another forest dominated by Norway spruce (*Picea abies*) and Scotch pine (*Pinus sylvestris*), which together comprised 91% of total volume, the RF estimated the stand volume with an rRMSE of 17.1% to volume prediction in the best case (BREIDENBACH et al., 2010). The AI tools presented a lower RMSE when modeling vegetation height and canopy cover in the Kras region of Slovenia (STOJANOVA et al., 2010). RF and Support Vector Regression successfully predicted biomass in forest containing both coniferous and deciduous trees in Upstate New York

(United States). The best biomass prediction was performed by SVR with a rRMSE of 18.1% (GLEASON; IM, 2012).

It is important to overcome certain obstacles if aiming to adopt these AI techniques widely, especially related to the transparency of the modeling results. One good alternative is to assess the relative importance of explanatory variables. Allowing this interpretation is extremely important to ensure that the statistical technique in question is not driving the survey towards biologically incorrect and non-sensical information (KOZAK; AZEVEDO, 2011). Each of these tools has its own strategy to investigate the influence of variables during prediction. RF is an interesting alternative to the traditional regression methods, since it does not choose just a few metrics, and consequently avoids losing important information when predicting the variable of interest, while maintaining a similar level of performance.

4.5 Conclusion

Rather than being limited to a subset of ALS metrics in attempting explain as much variability in a dependent variable as possible, artificial intelligence tools explore the complete metrics set when looking for patterns between ALS metrics and stand volume. AI tools can easily be employed towards dealing with the problems inherent in estimating stand volume from an ALS data set via the use of a machine learning approach. This approach can be implemented through software, continuously learning patterns from processed data sets. This characteristic is especially interesting to an ALS approach where the cloud point can be distilled into a huge number of metrics, and where the relationships between forest parameters and ALS metrics vary widely.

Acknowledgments

This paper is part of the research program developed by GET-LiDAR. FIBRIA SA is acknowledged for their funding of this study regarding the laser and field data set acquisition. The Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES, supported this study via a student grant and ForEAdapt (FP7-PEOPLE-2009-IRSES) supported this work through a visit to SLU/Sweden during data processing. We would also like to thank Simon Dunster for his review of our translation of this paper into English.

References

ATZBERGER, C. Object-based retrieval of biophysical canopy variables using artificial neural nets and radiative transfer models. **Remote Sensing of Environment**, Amsterdam, v. 93, n. 1/2, p. 53-67, 2004.

BREIDENBACH, J.; NÆSSET, E.; LIEN, V.; GOBAKKEN, T.; SOLBERG, S. Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. **Remote Sensing of Environment**, Amsterdam, v. 114, n. 4, p. 911-924, 2010.

BREIMAN, L. Random forests. **Machine Learning**, New York, v. 45, n. 1, p. 5-32, 2001.

CLUTTER, J.L.; FORTSON, J.C.; PIENAAR, L.V.; BRISTER, G.H.; BAILEY, R.L. **Timber management**: a quantitative approach. Chichester: John Wiley, 1983. 333 p.

CUTLER, D.R.; EDWARDS JR., T.C.; BEARD, K.H.; CUTLER, A.; HESS, K.T.; GIBSON, J.; LAWLER, J.J. Random forests for classification in ecology. **Ecology**, Ithaca, v. 88, n. 11, p. 2783-2792, 2007.

DURBHA, S.S.; KING, R.L.; YOUNAN, N.H. Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer. **Remote Sensing of Environment**, Amsterdam, v. 107, n. 1/2, p. 348-361, 2007.

FANG, H.L.; LIANG, S.L.; KUUSK, A. Retrieving leaf area index using a genetic algorithm with a canopy radiative transfer model. **Remote Sensing of Environment**, Amsterdam, v. 85, n. 3, p. 257-270, 2003.

GLEASON, C.J.; IM, J. Forest biomass estimation from airborne LiDAR data using machine learning approaches. **Remote Sensing of Environment**, Amsterdam, v. 125, p. 80-91, 2012.

GOH, A. Back-propagation neural networks for modeling complex systems. **Artificial Intelligence in Engineering**, New York, v. 9, n. 3, p. 143-151, 1995.

GONÇALVES-SECO, L.; GONZÁLEZ-FERREIRO, E.; DIÉGUEZ-ARANDA, U.; FRAGA-BUGALLO, B.; CRECENTE, R.; MIRANDA, D. Assessing the attributes of high-density *Eucalyptus globulus* stands using airborne laser scanner data. **International Journal of Remote Sensing**, London, v. 32, n. 24, p. 9821-9841, 2011.

HUDAK, A.T.; CROOKSTON, N.L.; EVANS, J.S.; FALKOWSKI, M.J.; SMITH, A.M.; GESSLER, P.E.; MORGAN, P. Regression modeling and mapping of coniferous forest basal area and tree density from discrete-return lidar and multispectral satellite data. **Canadian Journal of Remote Sensing**, Ottawa, v. 32, n. 2, p. 126-138, 2006.

KOZAK, M.; AZEVEDO, R.A. Does using stepwise variable selection to build sequential path analysis models make sense? **Physiologia Plantarum**, New York, v. 141, n. 3, p. 197-200, 2011.

LATIFI, H.; NOTHDURFT, A.; KOCH, B. Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: application of multiple optical/LiDAR-derived predictors. **Forestry**, London, v. 83, n. 4, p. 395-407, 2010.

LEFSKY, M.A.; COHEN, W.B.; ACKER, S.A.; PARKER, G.G.; SPIES, T.A.; HARDING, D. Lidar remote sensing of the canopy structure and biophysical properties of Douglas-fir western hemlock forests. **Remote Sensing of Environment**, Amsterdam, v. 70, n. 3, p. 339-361, 1999.

LIAW, A.; WIENER, M. Classification and regression by random forest. **R News**, Vienna, v. 2, n. 3, p. 18-22, 2002.

LIM, K.S.; TREITZ, P.M. Estimation of above ground forest biomass from airborne discrete return laser scanner data using canopy-based quantile estimators. **Scandinavian Journal of Forest Research**, Oslo, v. 19, n. 6, p. 558-570, 2004.

MITCHELL, T.M. **Machine learning.** Boston: McGraw-Hill, 1997. 414 p.

MONTAGHI, A.; CORONA, P.; DALPONTE, M.; GIANELLE, D.; CHIRICI, G.; OLSSON, H. Airborne laser scanning of forest resources: An overview of research in Italy as a commentary case study. **International Journal of Applied Earth Observation and Geoinformation**, Amsterdam, v. 23, p. 288-300, 2013.

MURPHY, M.A.; EVANS, J.S.; STORFER, A. Quantifying Bufo boreas connectivity in Yellowstone National Park with landscape genetics. **Ecology**, New York, v. 91, n. 1, p. 252-261, 2010.

NÆSSET, E. Estimating timber volume of forest stands using airborne laser scanner data. **Remote Sensing of Environment**, Amsterdam, v. 61, n. 2, p. 246-253, 1997.

_____. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. **Remote Sensing of Environment**, Amsterdam, v. 80, n. 1, p. 88-99, 2002.

_____. Practical large-scale forest stand inventory using a small-footprint airborne scanning laser. **Scandinavian Journal of Forest Research**, Oslo, v. 19, n. 2, p. 164-179, 2004.

NÆSSET, E.; BJERKNES, K.O. Estimating tree heights and number of stems in young forest stands using airborne laser scanner data. **Remote Sensing of Environment**, Amsterdam, v. 78, n. 3, p. 328-340, 2001.

NÆSSET, E.; GOBAKKEN, T. Estimation of above- and below-ground biomass across regions of the boreal forest zone using airborne laser. **Remote Sensing of Environment**, Amsterdam, v. 112, n. 6, p. 3079-3090, 2008.

NELSON, R.; KRABILL, W.; TONELLI, J. Estimating forest biomass and volume using airborne laser data. **Remote Sensing of Environment**, Amsterdam, v. 24, n. 2, p. 247-267, 1988.

NILSSON, M. Estimation of tree heights and stand volume using an airborne lidar system. **Remote Sensing of Environment**, Amsterdam, v. 56, n. 1, p. 1-7, 1996.

PACKALEN, P.; MEHTATALO, L.; MALTAMO, M. 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**, Paris, v. 68, n. 6, p. 1085-1092, 2011.

PARKER, G.G.; RUSS, M.E. The canopy surface and stand development: assessing forest canopy structure and complexity with near-surface altimetry. **Forest Ecology and Management**, Amsterdam, v. 189, n. 1/3, p. 307-315, 2004.

PATENAUDE, G.; HILL, R.A.; MILNE, R.; GAVEAU, D.L.A.; BRIGGS, B.B.J.; DAWSON, T.P. Quantifying forest above ground carbon content using LiDAR remote sensing. **Remote Sensing of Environment**, Amsterdam, v. 93, n. 3, p. 368-380, 2004.

PFLUGMACHER, D.; COHEN, W.B.; KENNEDY, R. Using Landsat-derived disturbance history (1972–2010) to predict current forest structure. **Remote Sensing of Environment**, Amsterdam, v. 122, p. 146-165, 2012.

PRASAD, S.; BRUCE, L.M.; CHANUSSOT, J. **Optical remote sensing:** advances in signal processing and exploitation techniques. Berlin: Springer, 2011. 352 p.

PROY, C.; TANRE, D.; DESCHAMPS, P. Evaluation of topographic effects in remotely sensed data. **Remote Sensing of Environment**, Amsterdam, v. 30, n. 1, p. 21-32, 1989.

R CORE TEAM. **R**: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing 2013.

REUTEBUCH, S.E.; ANDERSEN, H.E.; MCGAUGHEY, R.J. Light detection and ranging (LIDAR): an emerging tool for multiple resource inventory. **Journal of Forestry**, Bethesda, v. 103, n. 6, p. 286-292, 2005.

STEPHENS, P.R.; KIMBERLEY, M.O.; BEETS, P.N.; PAUL, T.S.H.; SEARLES, N.; BELL, A.; BRACK, C.; BROADLEY, J. Airborne scanning LiDAR in a double sampling forest carbon inventory**. Remote Sensing of Environment**, Amsterdam, v. 117, p. 348-357, 2012.

STOJANOVA, D.; PANOV, P.; GJORGJIOSKI, V.; KOBLER, A.; DŽEROSKI, S. Estimating vegetation height and canopy cover from remotely sensed data with machine learning. **Ecological Informatics**, Amsterdam, v. 5, n. 4, p. 256-266, 2010.

VALBUENA, R.; MALTAMO, M.; MARTÍN-FERNÁNDEZ, S.; PACKALEN, P.; PASCUAL, C.; NABUURS, G.J. Patterns of covariance between airborne laser scanning metrics and Lorenz curve descriptors of tree size inequality. **Canadian Journal of Remote Sensing**, Ottawa, v. 39, n. S1, p. S1-S14, 2013.

WHITE, H. Learning in artificial neural networks: a statistical perspective. **Neural computation**, Cambridge, v. 1, n. 4, p. 425-464, 1989.

WHITE, J.C.; WULDER, M.A.; VASTARANTA, M.; COOPS, N.C.; PITT, D.; WOODS, M. The utility of image-based point clouds for forest inventory: a comparison with airborne laser scanning. **Forests**, Basel, v. 4, n. 3, p. 518-536, 2013.

ZHAO, K.; POPESCU, S.; NELSON, R. Lidar remote sensing of forest biomass: A scale-invariant estimation approach using airborne lasers. **Remote Sensing of Environment**, Amsterdam, v. 113, n. 1, p. 182-196, 2009.

ZHAO, K.; POPESCU, S.; ZHANG, X. Bayesian learning with Gaussian processes for supervised classification of hyperspectral data. **Photogrammetric Engineering and Remote Sensing**, Bethesda, v. 74, n. 10, p. 1223-1234, 2008.

ZHAO, K.; POPESCU, S.; MENG, X.; PANG, Y.; AGCA, M. Characterizing forest canopy structure with lidar composite metrics and machine learning. **Remote Sensing of Environment**, Amsterdam, v. 115, n. 8, p. 1978-1996, 2011.

ZHOU, J.; PROISY, C.; DESCOMBES, X.; LE MAIRE, G.; NOUVELLON, Y.; STAPE, J. L. VIENNOIS, G.; ZERUBIA, J.; COUTERON, P. Mapping local density of young Eucalyptus plantations by individual tree detection in high spatial resolution satellite images. **Forest Ecology and Management**, Amsterdam, v. 301, p. 129-141, 2013.

ZONETE, M.F. **Análise do uso da tecnologia laser aerotransportado para inventários florestais em plantios clonais de *Eucalyptus sp.* no sul da Bahia**. 2009. 95 p. Dissertação (Mestrado em Recursos Florestais) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2009.

5 CHARACTERIZATION OF FOREST TYPOLOGIES UTILIZING VERTICAL PROFILES DERIVED FROM AIRBORNE LASER SCANNING SURVEYS

"In the end, it's not the year in your life that count.
It's the life in your years."
Abraham Lincoln

Abstract

This article explores the use of airborne laser scanning (ALS) surveys to describe four different groups of forest typologies in Brazil, and the spatial dependencies related to the computation of canopy profiles. The Weibull distribution with two parameters was used to describe the profile of vertical heights for seven different forests, four native types: rupestrian fields, semi-deciduous, open ombrophilous and dense ombrophilous, plus three eucalyptus plantations (ages 3 and 8 years for one clone, and age 8 years for a second clone). Grid cell sizes were varied from 10 to 80 meters, and semivariograms were built to determine the cell size that resulted in adjusted distribution models with the lowest random errors. Simulation techniques were subsequently used to determine the minimum number of grid cells needed to define a reliable sampling intensity for each cell size. All combinations of forest typology (7 typologies), grid cell size (8 sizes) were submitted to 1000 iterations of 6 sampling intensities. The Weibull parameters for the rupestrian fields and the eucalyptus plantations clearly assigned an exponentially shaped curve (shape parameter ≤ 1) for the cumulative density function (cdf) of the vertical profile. In contrast, the cdf curves for the semi-deciduous and the two ombrophilous forests have a sigmoidal shape (Weibull shape parameter greater than one). Differences between the savannah, native and planted forest profiles were evident. More subtle were the differences detected inside each group. Dense and open ombrophilous showed a subtle difference in the Weibull scale parameter. The main source of difference in the vertical profiles of the two eucalyptus plantations of the same age, but different clones, proved to be in the Weibull scale parameter. The inventory assessments for large areas of forest can benefit from the results observed in this study, as they should help in the development of recommendations for grid cell size and sampling intensity. The results also provide an abbreviated characterization (two parameters: shape and scale) of the vertical profile for the different typologies: open ombrophilous forest (shape = 1.78, scale = 22.74 m), dense ombrophilous forest (1.84, 17.50 m), semi-deciduous forest (1.35, 9.92 m), young eucalyptus plantation (0.86, 12.21 m), mature eucalyptus plantations (0.60, 12.45 m, and 0.79, 16.63 m) and rupestrian fields (0.45, 0.37 m).

Keywords: Light Detection And Ranging; LiDAR; ALS; Canopy profile

5.1 Introduction

The vertical canopy structure is one of the most important characteristics for the discrimination of different types of forest. This structure expresses functional attributes and processes related to tree growth, as well as reflecting disturbances impacting on different levels of the landscape: tree, population, community, and ecosystem (PARKER et al., 2004).

An increase in the height of the trees causes modifications in the structure and microclimate within a forest, which together contribute to the development of more organized vertical structures. However, quantification of the vertical structure of a forest through in-field surveys has proven very difficult, requiring not only high levels of manpower, but also the production of data sets on a scale and with sufficient intensity to represent the area of interest (BAKER; WILSON, 2000).

The use of remote sensing (RS) for forestry monitoring is not new, but continues to concentrate on the interpretation of optical images, principally obtained from satellites and aircraft. Although these images are able to provide data covering large areas for relatively low effort, the optical sensors have not proven effective in the monitoring of vertical structure, for they do not offer vertical penetration (VASTARANTA, 2012). Three-dimensional sensors, such as those based around the use of lasers and radar, do not suffer this limitation, and are gaining increasing attention. Montaghi et al. (2013) have found that airborne laser scanning (ALS) surveys are still focused on the determination of forest attributes quantification. Applications directed at ecological issues and the detection of ecological change are still rare, principally due to the low availability of laser overflights as a function of time covering large areas.

Although airborne laser sensors would seem to be the ideal tool for the monitoring and mapping of forests, very little is currently known about the vertical signatures of different forest typologies (ASNER et al., 2012). A forest typology displays spatial structure, and robust sampling methods are required in order to detect different levels of spatial autocorrelation (GOSLEE, 2006). Many of the published studies that address forest ecology do not consider spatially dependent structure (DORMANN, 2007).

Studies of forest typology must necessarily pass through four issues: (1) determination of the spatially dependent structure of the typology; (2) sampling optimization (scale of the sampling scheme); (3) interpolation and map construction;

and (4) estimation and inference (AUBRY; DEBOUZIE, 2000, 2001). In ecological surveys, the concept of scale refers principally to the components defining the sampling scheme (HE; LAFRANKIE; SONG, 2002), these are the size and shape of the sampling unit, and the distance between the sampling units (BELLEHUMEUR; LEGENDRE, 1998; DUNGAN et al., 2002). Geostatistics and simulation are the essential tools for the investigation of these issues. The term geostatistics is used to identify those statistical methods that provide descriptions of observations which present a continuous variation across space. As such, geostatistics models the spatial distributions, assigning a degree of dependence to the measurements based on distance and direction between the sampled points (VIOLA, 2007).

It may be expected that laser surveys will become increasingly utilized for the monitoring of forest coverage. The ability of a laser sensor to penetrate into forest vegetation and to generate a high density of records assures the collection of information from the different strata composing a forest area, permitting the determination of the vertical profile on different scales.

One fundamental task is to establish the relationships between the information furnished by an airborne laser sensor with the forest parameters obtained from traditional field inventories. Such relationships lie at the heart of the interpretation of the information extracted from a sensor, and addressing the impacts of spatial dependence and sample size on surveys obtained from laser sensors will permit the creation of sampling designs that are compatible between field and airborne surveys.

The objective of this study was the characterization of the vertical profile of seven forest typologies through surveys with airborne laser instruments, the analysis of the spatial dependence of vertical profile parameters and the investigation of the effects of the sampling intensity.

5.2 Materials and Methods

The present study was performed for seven forest typologies, four corresponding to different biomes found in Brazil, and three to fast growing eucalyptus plantations.

5.2.1 Study areas

Area 1 is located in the state of Pará (Figure 5.1 – location 1), and is currently managed by the Instituto de Floresta Tropical (IFT), a non-governmental organization. The area comprises an open ombrophilous forest, and constitutes a part of the Brazilian biome known as the Amazonian Rainforest. The terrain is gently undulating, with a predominance of latosols. There is a dry period, which lasts for more than 60 days. Prevalent genera are *Pouteria, Licania, Inga, Protium* and *Miconia* (OLIVEIRA-FILHO, 2000). It is common to find communities of vines covering these shrubs, especially in depressions in the landscape.

Areas 2 and 3 are located within a conservation area in the state of Minas Gerais (Figure 5.1 – location 2), where the altitude is approximately 800 m (VELOSO et al., 1992). For area 2, the vegetation is composed of secondary, seasonal semi-deciduous forest, typical of regions with a two-season climate pattern, which together with dense and mixed ombrophilous formations are among the forest typologies which make up the Atlantic Forest (VIANA, 1996). The climate has two distinct seasons: a tropical season with heavy summer rains followed by an accentuated drought, and a subtropical season provoked by the cool winters, when temperatures can be below 15 °C. The areas of semi-deciduous forest are frequently associated with a transition between ombrophilous forest and open savannah, the Brazilian Cerrado (LEITÃO FILHO, 1987). The distribution of rains is the principal factor that distinguishes the ombrophilous and semi-deciduous forests (OLIVEIRA FILHO; FONTES, 2000). The proportion of deciduous trees within the semi-deciduous forests of the Atlantic Forest falls within the range 20 to 50 %. Families which are common and considered most representative of the seasonal semi-deciduous forests of the region are *Myrtaceae, Fabaceae, Rubiaceae, Lauraceae*, and *Melastomataceae* (OLIVEIRA FILHO; FONTES, 2000).

Area 3 presents a distinctively Brazilian typology, known locally as 'campo rupestre', for which the literal translation is rupestrian field, the latter designation is followed in this work. This formation occurs at higher altitudes in the eastern region of Brazil, and rupestrian field areas are recognized as important centers of endemism for neotropical flora (GIULIETTI et al., 2000; CONCEIÇÃO; PIRANI, 2007). Veloso et al. (1991) have described the rupestrian fields as "plant refuges", for the plant species found in these areas are isolated, in the sense that they are completely distinct from the dominant flora found in the regions where these rupestrian fields are located. Souza et al. (2010) have associated rupestrian fields with the occurrence of surface

soils or quartzite outcrops which affect plant growth. The rupestrian fields found in the region which contains location 2 are situated in areas of transition between tropical savannah (Cerrado), semi-arid scrub forest (Caatinga) and the Atlantic Forest (VASCONCELOS, 2000), though they are commonly considered a component part of the Cerrado (SILVA; BATES, 2002; ALVES; KOLBEK, 2010).

Area 4 is located in the state of São Paulo, within the Serra do Mar State Park, which is part of the Brazilian Atlantic Forest biome (Figure 5.1 – location 3). The vegetation corresponds to dense ombrophilous forest, characterized by the presence of phanerophytes, woody lianas, and an abundance of epiphytes. The principal features of the characteristic ecology of this part of the Atlantic Forest are directly related to the high temperatures (average of 25 °C) and the absence of a significant dry period, since rains are evenly distributed throughout the year. Due to the high availability of water, far greater structural complexity is found in comparison to seasonal semi-deciduous forests (OLIVEIRA FILHO; FONTES, 2000). The soil is predominantly formed of latosols with some cases of cambisols, both with low natural fertility. Families which are frequently found and so considered characteristic of such ombrophilous forest regions of the Brazilian Atlantic Forest are *Myrtaceae, Melastomataceae, Rubiaceae, Fabaceae,* and *Lauraceae* (OLIVEIRA FILHO; FONTES, 2000).

Areas 5, 6, and 7 are forest plantations of fast growing trees in the state of São Paulo (Figure 5.1 – location 4). By definition, forest plantations are stands of trees artificially introduced into a given region, whose composition is represented by a unique species. Plantations are synonymous with high levels of human intervention, and are created for a specific market destination (FOX, 2000). In Brazil, forest plantations have the additional characteristics of rapid growth and short rotations. The prime genus is *Eucalyptus*, covering 5.1 million hectares, representing 70 % of the area given over to forest plantations. The eucalyptus clone P4295H was being raised in areas 5 and 6, at the time the data were collected the ages of the plantations were three and eight years, respectively. Area 7 also corresponded to a forest plantation with an age of 8 years, in this case made up of TC31H clones. For all three forest plantations, the planting grid was 3 m x 2 m, and the expected clear cut of eucalyptus plantation generally occur when the forest reaches 7 to 8 years old.

Figure 5.1 - Location of the study areas: 1 - open ombrophilous forest (area 1), 2 – semi-deciduous seasonal forest (area 2) and rupestrian fields (area 3), 3 – dense ombrophilous forest (area 4), and 4 – forest plantations of rapidly growing eucalyptus (areas 5, 6, and 7)

5.2.2 Airborne Laser Surveys

The laser surveys of the seven study areas, representing the different typologies studied, were performed during three data gathering campaigns, with different equipment and operational parameters (Table 5.1). The three eucalyptus plantations (location 4 – areas 5, 6, and 7) were flown over in April 2012, the flight over the dense ombrophilous forest (location 3 – area 4) was conducted in September 2010, while the other overflights (location 1 – area 1, and location 2 – areas 2 and 3) took place in September 2009.

Table 5.1 - Areas covered by the surveys over the seven forest typologies, together with identification of the equipment used, the operating parameters for the sensor, and the flight plan

Typology	Area / ha	Equipment	Parameters				
			Sensor			Flight	
			PRF[1/] / kHz	Scan frequency / Hz	Scan angle	Overlap	Height / m
Open ombrophilous	949.15	ALTM 3100	50	25	10°	30 %	1500
Rupestrian fields	22.79	ALTM 3100	70	55.9	7.8°	30 %	1371.60
Seasonal semi-deciduous	114.00	ALTM 3100	70	55.9	7.8°	30 %	1371.60
Dense ombrophilous	1170.31	ALTM 3100	50	25	10°	30 %	1500
Eucalyptus plantation: clone: P4295H, age: 3 years	72.19	RIEGL LMS Q680I	200	30	30°	30 %	629.24
Eucalyptus plantation: clone: P4295H, age: 8 years	42.89	RIEGL LMS Q680I	200	30	30°	30 %	629.24
Eucalyptus plantation. clone: TC31H, age: 8 years	18.42	RIEGL LMS Q680I	200	30	30°	30 %	629.24

[1/] PRF = pulse repetition frequency

5.2.3 Vertical profile

The LiDAR (Light Detection And Ranging) point clouds for each of the seven typologies were normalized to digital models of the respective underlying terrain. The vertical profile for the whole area of each studied typologies was modelled by means of the two-parameter Weibull cumulative distribution function:

$$F(h) = 1 - \left[e^{-\left(\frac{h}{scale}\right)^{shape}} \right] \quad (5.1)$$

in which $h \geq 0$, and $F(h) = 0$ for $h < 0$ (BAILEY; DELL, 1973).

The Weibull cumulative distribution (equation 5.1) function was selected for its ability to assume different forms. The shape parameter (*shape*) is dimensionless and by changing the value of *shape*, the Weibull cumulative distribution function can be used to model the cumulative distribution function of a wide variety of distributions, including normal and exponential. The scale parameter (*scale*) has the same units as

the variable (*x*) whose cumulative distribution is to be modelled; increasing or diminishing the value of *scale* for a given value of *shape* has the effect of widening or narrowing, respectively, the range of *x* values over which *F*(*x*) increases from ~0 to ~1.

5.2.4 Spatial dependence and sampling unit size

Eight *raster* files, comprised by a grid of regular cells ranging from 10 to 80 m, in equal intervals of 10 m, were created for each of the seven typologies, resulting in a total of 56 files (8 rasters × 7 typologies). For a given typology, the two parameters (*scale*, *shape*) of the Weibull cumulative distribution function were estimated for each cell of each of the raster files. Starting from these cell values, semivariograms were computed for *scale* and *shape* for each *raster* file. The semivariogram of a random variable z known at a set of locations x_i comprises values for the semivariance of z as a function of the distance u between pairs of locations, where the semivariance of z for distance u is defined as one half of the variance of the differences between the z values for pairs of locations separated by the distance u:

$$\hat{\gamma}(z;u) = \frac{1}{2N(u)} \sum_{a=1}^{N(u)} \big(z(x_a) - z(x_a + u)\big)^2 \tag{5.2}$$

where $z(x_a)$ is the value of z at location x_a, and $z(x_a + u)$ the value at $x_a + u$, $N(u)$ is the number of pairs of locations separated by the distance u, so that the summation extends over all pairs of locations separated by the distance u (OLIVER, 2010). For the present case, the locations corresponded to the centers of the cells making up the *raster* files. Low values for $\hat{\gamma}(z;u)$indicate low spatial variability of the variable z, that is similarity between locations.

In the presence of spatial stationarity, there is a relationship between the structure of the variances and covariances and that of the semivariance, which is given by the following expression:

$$\hat{\gamma}(z;u) = \tau_z^2 + \sigma_z^2 \times \big(1 - \rho_z(u)\big) \tag{5.3}$$

in which τ_z^2 is the *nugget effect*, σ_z^2 is the partial *sill*, and the function $\rho_z(u)$ expresses the manner in which the semivariance approaches the semivariogram upper bound or sill, $(\tau_z^2 + \sigma_z^2)$. The distance from the origin at which the semivariance reaches the sill

will be called the *range* of the semivariogram. The nugget effect (τ_z^2) represents the random component of the total variance (OLIVER, 2010). An empirical semivariogram (equation 5.2) can assume various forms, among the principal models, corresponding to different functions $\rho_z(u)$ in equation 5.3, are Gaussian, linear, exponential, and spherical.

5.2.5 Sampling intensity and sampling unit size

For a given forest typology and cell size (ranging from 10, to 80 m, in steps of 10 m), *n* cells were selected at random 1000 times. The returns contained within the selected cells were concatenated, a height profile was generated, and values for the *scale* and *shape* parameters of the Weibull cumulative distribution function obtained. The *n* increased from 3 up to 10 % of the total number of cells (*N*) in the *raster* file, where *N* depended on the area surveyed and the cell size.

At the end of each *n*-selection, the estimated values for the *scale* and *shape* parameters were compared against the values obtained when the returns of the entire area (*N*), and not only the *n*-cells, were used in the Weibull adjust. The Weibull cumulative distribution function generated from (*scale*(*n*), *shape*(*n*)), estimated from the returns for the *n* selected cells, was compared against that from (*scale*(*N*), *shape*(*N*)) using the Kolmogorov - Smirnov test(KS). The number of times that the KS-test was not significant among 1000 iterations was expressed as percentage of adhesion of the sample (n) to the population (N). The adhesion of each *n* generated an adherence curve for each typology and each cell size.

5.3 Results

5.3.1 Vertical profiles of the topologies

The cumulative distribution obtained from the heights of the returns is an expression of the vertical profile of a typology (Table 5.2). In this study of seven typologies, each typology presented distinct characteristics (Figure 5.2). The cumulative distribution curves can be described in terms of level of saturation and gradient. Saturation corresponds to the curve reaching its upper bound, which in the

case of a cumulative distribution is the value one. The gradient indicates the extent to which the value of the function changes in response to a change in the height of the canopy profile.

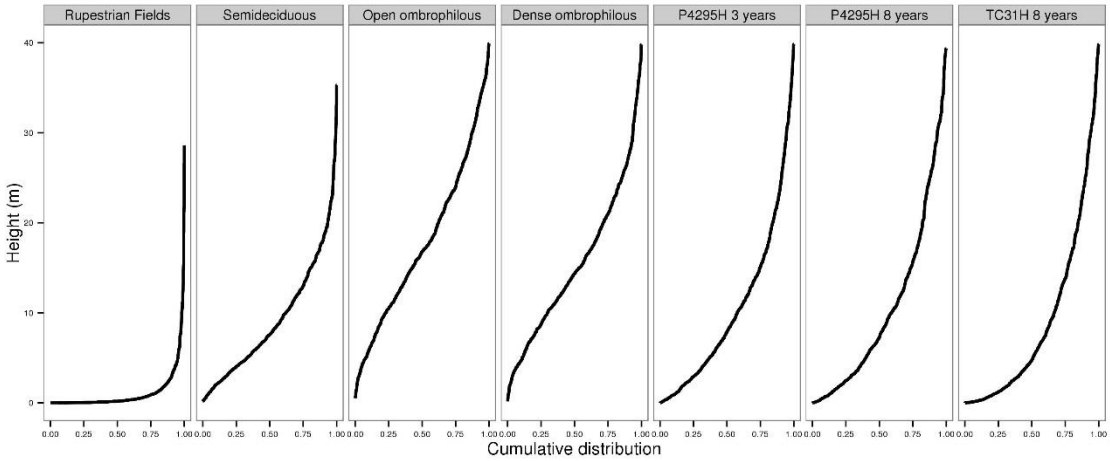

Figure 5.2 - Observed cumulative distributions for the seven studied typologies

The cumulative distribution for the rupestrian fields presented rapid saturation, showing that 96% of the returns were from heights below 5 m. For both of the ombrophilous typologies, the curves were sigmoidal; however, the dense typology presented a slightly more rapid approach to saturation than the open typology. The empirical cumulative distribution curve for the semi-deciduous forest is also sigmoidal, with a point of inflection located at a height of about 3 m, saturation has been approached by 20 m, where the distribution curve has reached 92.3 % of its upper bound.

A cursory inspection of the curves for the three eucalyptus forest plantations shows that they do not present points of inflection, this immediately differentiates them from the curves for the ombrophilous and semi-deciduous forest typologies. There is also a clear difference from the curve for the rupestrian fields, as the approach to saturation occurs over a much greater range of heights. However, it is more difficult to discern differences between the three plantations. When the two plantations of the clone P4295H are compared (the difference is significant with p-value ~ 0.013) with the curve for the 8-year old plantation reaching the saturation quite slowly than the 3-year old plantation. Comparing the curve for the 8-year old plantation of the clone TC31H with that for the plantation of the same age of P4295H, the most obvious differences are in the gradients of the curves at smaller heights (< 7.5 m), where the gradient for the clone TC31H is greater.

Table 5.2 - Cumulative Weibull distribution parameters describing the vertical profiles for the seven studied forest typologies, obtained from cumulative histograms generated using all the returns in the ALS surveys

Typology	Shape	Scale (m)
Rupestrian field	0.45	0.37
Semi-deciduous	1.35	9.92
Open ombrophilous	1.78	22.74
Dense ombrophilous	1.84	17.50
P4295H (3 years)	0.86	12.21
P4295H (8 years)	0.79	16.63
TC31H (8 years)	0.60	12.45

5.3.2 Random component and sampling unit size

The only typologies which presented semivariograms without a sill were the dense ombrophilous typology, for both the shape and scale parameters, and the 3-year old plantation of the P4295H clone, for the scale parameter. All the other typologies presented typical semivariograms with a sill, for both the shape and scale parameters (Figure 5.3).

The open and dense ombrophilous forests and the semi-deciduous forest presented steady decreases in the random variance (nugget effect) towards an asymptotic value as the size of the sampling unit was increased. This behavior was observed for both the scale and shape parameters. For the rupestrian field typology, the random variance at first fell with increasing cell size, passed through a minimum for a cell size of 60 m, before increasing again for the largest cell sizes. For the eucalyptus plantations, the form of the dependence of the random variance on cell size was less clear, though it did appear that there was a tendency for the random variance to decrease with increasing cell size, possibly towards a constant value.

There was also evidence that the values for the sill of each of the semivariograms decreased towards an asymptotic value with increasing cell size. As for the nugget effect, this trend was less clear for the eucalyptus plantations than for the other typologies. The best cell dimensions to all the typologies fell between 30 and 60 m for the random variance and the sill.

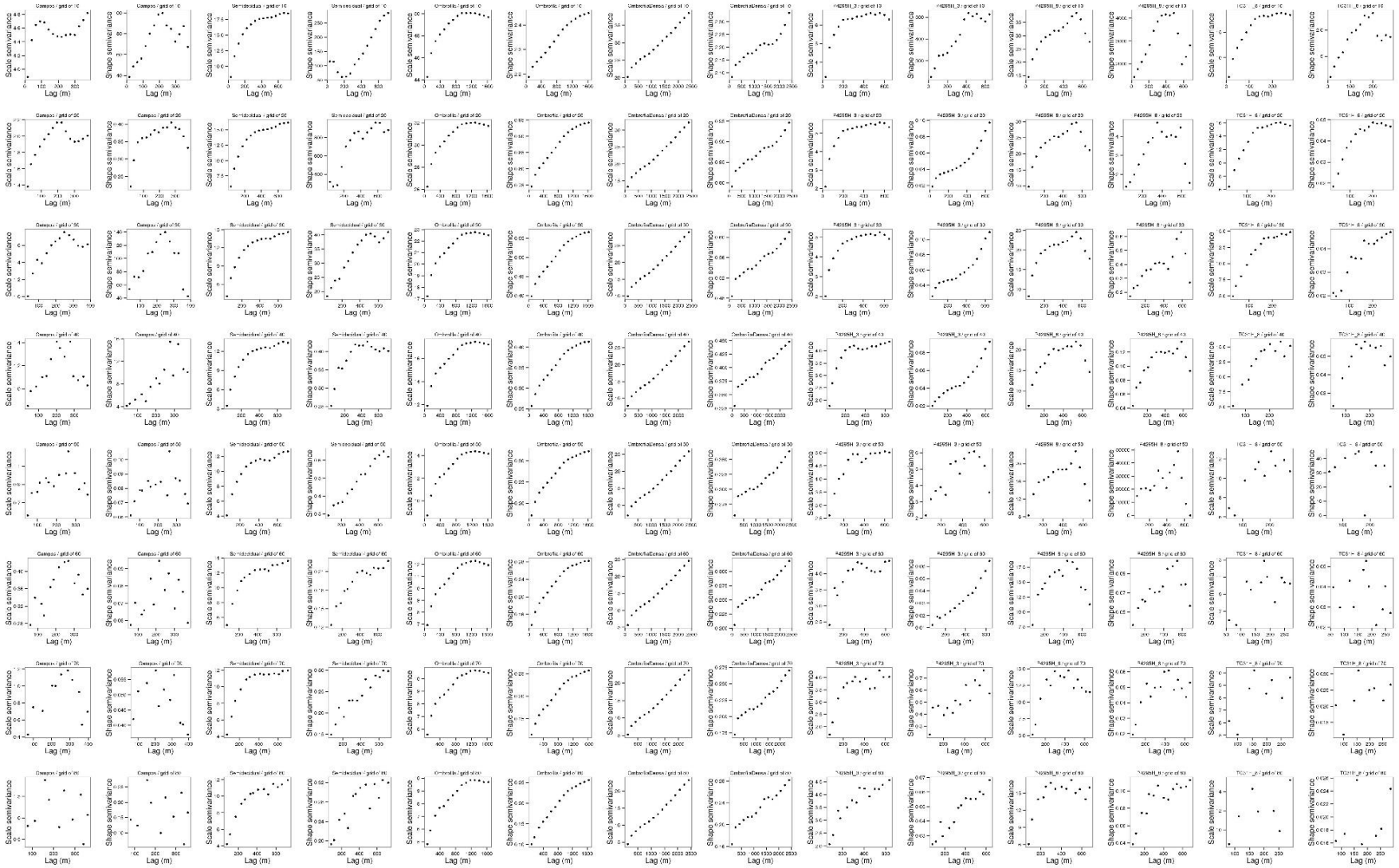

Figure 5.3 - Scale and shape semivariograms for the seven studied forest typologies, per grid cell size

The analysis of the semivariograms for the open ombrophilous typology returned values for the range parameter which were invariant to the grid size: equivalent to 1000 m for the Weibull scale parameter and 1600 m for the shape parameter. Cell size invariant estimates for the range parameter were also obtained from the scale semivariograms for the semi-deciduous typology and the 8-year old plantation of the eucalyptus clone TC31H: 400 m and 150 m, respectively.

It has previously been established that some typologies could be grouped together on the basis of similarities between their vertical profiles, specifically the dense and open ombrophilous typologies formed one group, and the three plantations of eucalyptus clones another. Within group similarities were also apparent in the semivariograms and their properties.

5.3.3 Sampling intensity and sampling unit size

For all of the studied typologies, the adherence to the vertical profile obtained from the heights of all the returns improved as the sampling intensity increased (Figure 5.4). With the exception of the rupestrian fields typology, the adherences converged to a value of approximately 90 % for all tested sampling unit sizes (10 m to 80 m), within the range of tested sampling intensities, which extended to 10 % coverage. For the rupestrian fields typology, convergence was not achieved through the unit sizes, and the best adherence (~ 80 %) was obtained for the combination of a sampling unit size of 10 m and a sampling intensity of 225 units (~10 % of the area covered by the ALS survey). With larger sampling unit sizes, although the adherence improved as the sampling intensity increased, none delivered an adherence of greater than 80 %.

For the other typologies, the combinations of sampling unit size and sampling intensity that were sufficient to achieve at least 90% of adherence were identified, and the proportion of the total area of the ALS survey covered by the randomly selected cells evaluated. In the case of the semi-deciduous forest, convergence was achieved for sampling cell sizes of 10, 20, and 30 m, at sampling intensities equivalent to 8 % of the area covered by the ALS survey. For the dense and open ombrophilous typologies, combinations of sampling intensity and sampling unit size equivalent to 2 % of the surveyed area were sufficient to give convergence (adherence ~90 %), independent of the sampling unit size. The plantation of young eucalyptus (P4295H, 3 years)

presented convergence of the adherence to a level above 90 % for cell dimensions of 10, 20, and 30 m combined with sampling intensities equivalent to ≥ 2 % of the population, for larger cell dimensions (40 to 70 m) greater sampling intensities (equivalent to ≥ 8 % of the area) were required to achieve convergence. For the mature plantations (P4295H, 8 years and TC31H, 8 years) convergence was not achieved for larger cell sizes: ≥ 50 m for P4295H and ≥ 40 m for TC31H. With smaller cell sizes, the sampling intensity necessary to achieve convergence depended on the clone. Thus, the clone P4295H presented convergence of the percent adherence for sampling intensities equivalent to ≥ 2 % of the area for cell dimensions of 10 and 20 m. Clone P4295H with cell dimensions of 30 and 40 m presented convergence for sampling intensities equivalent to ≥ 8 % of the area. The sampling intensities necessary to achieve convergence for the clone TC31H for cell dimensions of 20 and 30 m were equivalent to ≥ 5 % and ≥ 10 % of the area of the population, respectively.

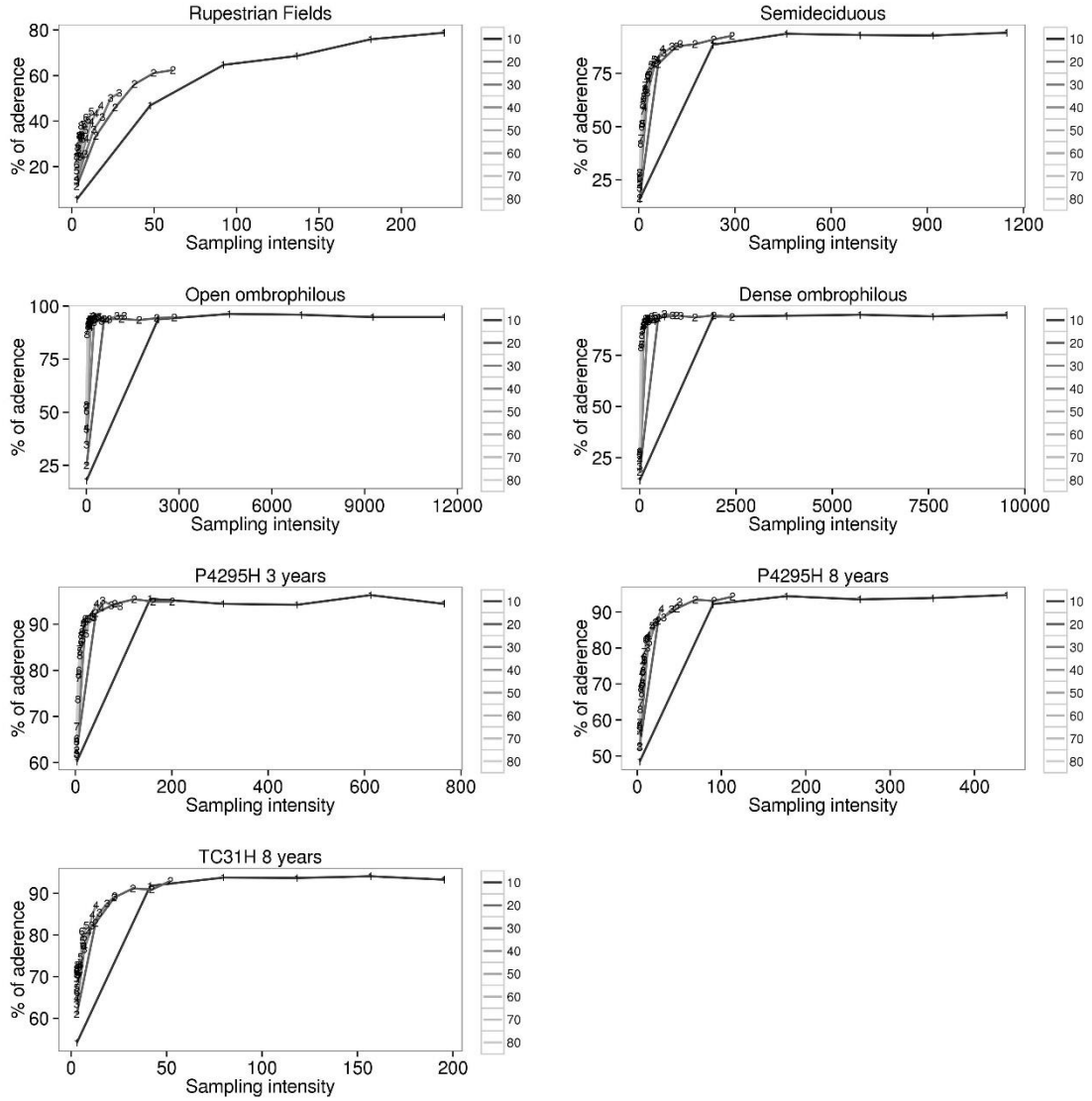

Figure 5.4 - Convergence of the vertical profiles, expressed as percent adherence to the all returns profile, as a function of cell size and sampling intensity

5.4 Discussion

5.4.1 Vertical profile of the vegetation

The cloud of points from an ALS survey represents the distribution in three dimensions of the returns coming from objects above the surface of the underlying terrain. In the case of the clouds analyzed in this study, these objects are elements of the vegetation, including leaves, stems, branches, and trunks. Our results show that there exist substantial variations in the vertical structure of the vegetation, permitting the identification of similarities and differences between the studied forest typologies.

The Weibull distribution function was utilized for the modelling of the vertical profile due to its flexibility, which makes possible the representation of a variety of asymmetric distributions. The Weibull function contains a scale parameter, which corresponds to the 63.2th percentile of the probability density distribution (MCCOOL, 2012). In assessing vertical profiles, such as those examined in this work, the scale parameter can be used as a point of reference for describing the position (or height) of the profile. For comparison, in a normal distribution, the 50th percentile would indicate both the mean and the median of the distribution.

Mean heights obtained in phytosociological surveys of the typologies studied in the present work are related to the scale parameter values we have estimated from ALS surveys. Valle et al. (2006) found, for the area of open ombrophilous forest at location 1, a canopy height varying between 25 and 35 m (scale parameter = 22.74 m). Medeiros (2009) observed, in field measurements of plots of dense ombrophilous vegetation, that 70 % of the sampled individuals fell within height classes between 4 and 11.9 m and reported the maximum sampled height as 29.9 m (scale parameter = 17.50 m). In the survey conducted by Jacobi et al. (2008) for an area of rupestrian field close to the area covered by the ALS survey (location 2, area 3), it was identified that 80 % of the individuals had heights below 25 cm, with a few individual trees rising to heights of 10 m (scale parameter = 0.37 m). In the same municipality, two areas of secondary forest of semi-deciduous typology, similar to that overflown at location 2 (area 2), presented canopy heights varying between 12 and 16 m for one area, and between 9 and 11.5 m for the other (scale parameter for location 2, area 2 = 9.92 m). Macedo (2010) found a mean height of 22.19 m for 4-year old clonal plantations of eucalyptus on properties neighboring the plantations surveyed in this work (scale parameters: 12.21, 12.45, and 16.63 m).

Within the long list of factors that are responsible for variations in the vertical structure, profile architecture, and floristic composition of plant communities, the following are recognized as important: the distribution and quantity of rain and the annual cycle of temperature changes (OLIVEIRA FILHO; FONTES, 2000), soil conditions and characteristics (RICKLEFS, 2001; ASSIS et al., 2011; HOFMOCKEL et al., 2011), natural and anthropogenic disturbances (CHAMBERS; SILVER, 2004; HIGUCHI et al., 2008; SHIELS et al., 2010), competition between individuals (COOMES; LINES; ALLEN, 2011), and the individuals composing the local area (URIARTE et al., 2004).

In natural forests, there is a common pattern as to how the trees fill the vertical spaces, which is not applicable to forest plantations, where the space and vertical structure are necessarily very homogeneous (ASNER et al., 2012). Once the vertical profile of a forest typology has been determined, procedures for the monitoring, mapping, and classification of the vegetation can be created, since the vertical profile is directly linked to the inferential processes applied to the distribution and quantification of the principal components of the canopy (LEFSKY et al., 2002; ASNER et al., 2012). The vertical structure contains information at a series of levels that reveal the state of development of the vegetation (LEFSKY et al., 1999), and is related to the height of the trees (LEFSKY et al., 2002).

The height of a plant varies according to the species and its genetic characteristics, in association with various ecological factors (BOHLMAN; O'BRIEN, 2006; VAN GELDER; POORTER; STERCK, 2006). The geomorphological and tectonic processes that occurred during the quaternary period played a fundamental role in soil formation, and so continue to be visible today in allometric relationships and through their influence on the vertical profile of the vegetation (FELDPAUSCH et al., 2011).

The theory of hydraulic limitation hypotheses that tree heights are limited by the availability of water, so that maximum tree height patterns should coincide with the spatial distribution of rains (RYAN; PHILLIPS; BOND, 2006). Thus, the distribution of rains is an influential factor in stratification and, therefore, on the vertical profile of a forest typology; for instance, the number of strata is larger when the number of dry months is smaller (LOPES; FERRAZ; ARAUJO, 2008). The species growing in more humid areas invest more resources in vertical growth in their search for light than species found in dry areas. The proportion of deciduous trees within different forests, which can be related to rain distribution, is a factor in determining the amount of light that penetrates through the canopy: with more deciduous trees, there is an increase in the amount of light. There is a consequent impact on the height profile: with more light (more deciduous trees), fewer individuals reach the highest strata than for ombrophilous forests. Examining the results obtained in the present study, these effects are reflected in the differences in the scale and shape parameters between the ombrophilous forests (shape = 1.78 and 1.84, scale = 22.74 and 17.50) the semi-deciduous forest (shape = 1.35, scale = 9.92).

In this work, we have observed sigmoidal cumulative distributions, and consequently Weibull shape parameters > 1, for the forest typologies which naturally present greater vertical stratification. While, for less stratified forests, such as homogeneous plantations, there was no point of inflection in the cumulative distribution (Weibull shape parameter ≤ 1), so that the gradient of the cumulative distribution curve falls continuously with increasing height. For the rupestrian fields, the vegetation is dominated by plants of low stature, leading to low values for both the shape and scale parameters of the Weibull distribution. The extreme shallowness and low nutrient levels of the soil are significant factors distinguishing rupestrian fields from other plant habitats and accordingly determine the structure of the vegetation (BENITES et al., 2007).

The use of the Weibull function to characterize the vertical profile of each typology has permitted their differentiation based upon the cumulative probability distribution. This function has been shown to be the appropriate choice for the description of the vertical profiles of specific species and forest stands (COOPS et al., 2007; DEAN et al., 2009). The correlations between the parameters of the Weibull distribution and the structural attributes of a forest indicate that the laser returns have a predictive power (COOPS et al., 2009).

Drake et al. (2002), in a study that compared a primary tropical forest with a young secondary tropical forest, have also perceived that LiDAR clouds are sensitive to differences in canopy structure. These authors have demonstrated the importance of establishing the relationships between the vertical profile and forest parameters in different ecosystems. Airborne laser scanning can be used not only to derive volumetric and gravimetric information about a forest, but also for evaluations of the canopy structure (COOPS et al., 2007). Although the profiles were equally effective for the estimation of biomass, Asner et al. (2012) found differences in the profiles between tropical forests in four regions of the world. As a consequence, a single generic model is unable to return results of the same quality as specific models adjusted to account for regional differences.

5.4.2 Spatial dependence, sampling intensity and sampling unit size

The determination of the size of the sampling cells and the sampling intensity has two fundamental implications. The first is related to the design and planning of the data

collecting flight, and the second to the collection of field data for the calibration of models linking LiDAR and field. For very large scale surveys, such as a national survey, it is not economically viable to conduct flights over the entire study area, instead some sampling scheme is demanded (WULDER et al., 2007). The second implication arises because airborne laser surveys have generally taken advantage of previously existing forest inventories, which were not planned with consideration for the spatial characteristics of laser data.

The results obtained in this article reveal a spatial dependence for the parameters, which describe the vertical profile of the studied forest typologies, for it, has been found that their properties (nugget, sill and range) are directly influenced by the size of the sampling cell. Saatchi et al. (2011) observed spatial dependences of structural variables for mature and secondary forests in Panama; this high spatial variability was considerably reduced when the sampling cell size was increased. This disappearance of spatial variability occasioned by an increase in the size of the sampling unit has also been found by Clark and Clark (2000).

The systematic division of the LiDAR cloud by the overlaying of grids permitted the investigation of the spatial behavior of the parameters that described the vertical profile for different cell dimensions. For two typologies (dense ombrophilous forest and a 3-year old plantation of the clone P4295H), the shape's semivariograms did not converge to a sill, for these cases the maximum distance between the cells was not sufficient to stabilize the spatial dependence. For the typologies which presented semivariograms with a sill, convergence of the nugget effect and of the sill was obtained for cells sizes of 50 m $\times$ 50 m and above. Working with cells that lead to lower random variance is recommended, since these larger cells determine the part of the total variance which is not explained by the spatial phenomenon.

Environmental gradients regulate spatial variability and contribute to the aggregation of adult trees (MURRELL, 2009). Topography (BAGCHI et al., 2011), light (WRIGHT, 2002), water (ENGELBRECHT et al., 2007), and nutrients from the soil (JOHN et al., 2007) are regulators of spatial variability and species distribution, therefore, the spatial dependence of the arboreal component is dependent upon the environmental heterogeneity. The range was greatest for the open ombrophilous typology, 1000 m for the scale parameter of the vertical profile and 1600 m for the shape parameter. The semi-deciduous typology presented a range of 400 m for both parameters. For the rupestrian field, when smaller cell dimensions were utilized (< 30

m) the range was around 200 m, but with larger cells the range fell drastically. The forest stands in plantations presented ranges for the scale parameter which increased with increasing size of the sampling unit, for example the initial range (cell 10 m × 10 m) for the 3-year old plantation of the clone P4295H was 200 m, rising to 350 m for the largest cell size. The reverse trend was observed for the range of the shape parameter, in that the range decreased as the sampling cell dimension increased. Although it was possible to detect some similarities with the patterns displayed by the other typologies, the forest plantations did present some inconsistent points and a behavior that was sometimes chaotic. These anomalies were probably caused by the smaller areas of the plantation surveys (< 80 ha) compared to the natural forest surveys (> 100 ha), and the spatial discontinuities associated with the corridors of native vegetation that permeate the landscape.

It is important to consider which factors that alter the dynamics of a forest, influencing the rates of mortality, recruitment, gain and loss of biomass, and the velocity of change, affect the spatial homogeneity of the structural components (KELLNER et al., 2011). Therefore, the spatial dependence of the vertical profile can be influenced by the degree of disturbance (natural or not) of the forest, which are able to modify the spatial organization of the structural components of the vegetation (CLARK et al., 1996; NICOTRA; CHAZDON; IRIARTE, 1999).

The simulation of the sampling reinforces the view that it is not necessary to survey an entire area to determine the vertical profile of the area based upon LiDAR data. This corroborates the suggestion that large area surveys can be conducted through the planning of partial flights, for example by following the method discussed by Wulder et al. (2012). The adherence between the estimated parameters for a set of randomly selected cells and the parameters of the entire population converged, regardless of the size of the cells, as the sampling intensity increased. In some cases, convergence was achieved when the sampling intensity represented 2 % of the total area. In general, lower percentage sampling intensities for convergence were obtained for the typologies with greater forest coverage, in particular the ombrophilous typologies. While, for the rupestrian fields, which are composed of a large number of habitats in a complex mosaic (CARVALHO et al., 2012), higher coverage with the LiDAR sensor was necessary to obtain a better convergence index.

The compatibility between ALS and field surveys permits the use of LiDAR metrics as high correlation secondary variables to improve the estimation of a primary

variable (TSUI et al., 2013). The use of airborne laser surveys for inventory construction, monitoring, and mapping of large areas is a consequence of the high correlation that has been found between LiDAR data and forest parameters such as biomass, volume, and basal area (REUTEBUCH; ANDERSEN; MCGAUGHEY, 2005). Gains may be possible when consideration of the spatial dependence observed in the semivariograms of the vertical profile obtained from LiDAR data is exploited in the design of field sampling.

Although this work has demonstrated that it is possible to differentiate forest typologies by comparing their vertical profiles as derived from a LiDAR overflight, it is essential that the study is replicated using other locations. This further work will explore the range of variations in the vertical profile that may occur between locations for a single typology.

5.5 Conclusions

Knowledge of the characteristics that each typology will display when surveyed using a specific class of sensor is an essential step in the implementation of effective monitoring. This study has focused upon several highly important forest typologies, and shown that it is possible to differentiate the typologies through their vertical profiles as derived from airborne laser surveys. The size of the sampling cell does have an influence on the behavior observed in analyses of spatial dependence. Each typology has its own specific characteristics, which will need to be taken into consideration in projects targeting monitoring, inventory construction, and mapping based upon airborne laser surveys. The determination of a converged vertical profile could be achieved with data representing 10 % of the area for all typologies, while for some typologies 2 % coverage was sufficient. When the vertical profile results are combined with field information, it will be possible to determine the degree of spatial correlation between the attributes of a forest and ALS data.

Acknowledgments

This paper is part of the research program developed by GET-LiDAR. CAPES is thanked for support of this research through a PhD scholarship grant. ALS data provided by Fibria SA, the Sustainable Landscapes Project (EMBRAPA and United States Forest Service), and the City of Belo Horizonte aided in making this research possible.

References

ALVES, R.J.V.; KOLBEK, J. Can campo rupestre vegetation be floristically delimited based on vascular plant genera? **Plant Ecology**, Boston, v. 207, n. 1, p. 67-79, 2010.

ASNER, G.P.; MASCARO, J.; MULLER-LANDAU, H.C.; VIEILLEDENT, G.; VAUDRY, R.; RASAMOELINA, M.; HALL, J.S.; VAN BREUGEL, M. A universal airborne LiDAR approach for tropical forest carbon mapping. **Oecologia**, Berlin, v. 168, n. 4, p. 1147-1160, 2012.

ASSIS, A.C.C.; COELHO, R.M.; PINHEIRO, E.S.; DURIGAN, G. Water availability determines physiognomic gradient in an area of low-fertility soils under Cerrado vegetation. **Plant Ecology**, Boston, v. 212, n. 7, p. 1135-1147, 2011.

AUBRY, P.; DEBOUZIE, D. Geostatistical estimation variance for the spatial mean in two-dimensional systematic sampling. **Ecology**, New York, v. 81, n. 2, p. 543-553, 2000.

AUBRY, P.; DEBOUZIE, D. Estimation of the mean from a two-dimensional sample: the geostatistical model-based approach. **Ecology**, New York, v. 82, n. 5, p. 1484-1494, 2001.

BAGCHI, R.; HENRYS, P.A.; BROWN, P.E.; BURSLEM, D.F.R.P.; DIGGLE, P.J.; GUNATILLEKE, C.V.S.; GUNATILLEKE, I.A.U.N.; KASSIM, A.R.; LAW, R.; NOOR, W.; VALENCIA, R.L. Spatial patterns reveal negative density dependence and habitat associations in tropical trees. **Ecology**, New York, v. 92, n. 9, p. 1723-1729, 2011.

BAILEY, R.L.; DELL, T. Quantifying diameter distributions with the Weibull function. **Forest Science**, Bethesda, v. 19, n. 2, p. 97-104, 1973.

BAKER, P.J.; WILSON, J.S. A quantitative technique for the identification of canopy stratification in tropical and temperate forests. **Forest Ecology and Management**, Amsterdam, v. 127, n. 1/3, p. 77-86, 2000.

BELLEHUMEUR, C.; LEGENDRE, P. Multiscale sources of variation in ecological variables: modeling spatial dispersion, elaborating sampling designs. **Landscape Ecology**, Hague, v. 13, n. 1, p. 15-25, 1998.

BENITES, V.M.; SCHAEFER, C.E.G.; SIMAS, F.N.; SANTOS, H.G. Soils associated with rock outcrops in the Brazilian mountain ranges Mantiqueira and Espinhaço. **Brazilian Journal of Botany**, São Paulo, v. 30, n. 4, p. 569-577, 2007.

BOHLMAN, S.; O'BRIEN, S. Allometry, adult stature and regeneration requirement of 65 tree species on Barro Colorado Island, Panama. **Journal of Tropical Ecology**, Cambridge, v. 22, p. 123-136, 2006.

CARVALHO, F.; SOUZA, F.A.; CARRENHO, R.; MOREIRA, F.M.D.S.; JESUS, E.D.C.; FERNANDES, G.W. The mosaic of habitats in the high-altitude Brazilian rupestrian fields is a hotspot for arbuscular mycorrhizal fungi. **Applied Soil Ecology**, Amsterdam, v. 52, p. 9-19, 2012.

CHAMBERS, J.Q.; SILVER, W.L. Some aspects of ecophysiological and biogeochemical responses of tropical forests to atmospheric change. **Philosophical Transactions B**, London, v. 359, n. 1443, p. 463-476, 2004.

CLARK, D.B.; CLARK, D.A. Landscape-scale variation in forest structure and biomass in a tropical rain forest. **Forest Ecology and Management**, Amsterdam, v. 137, n. 1/3, p. 185-198, 2000.

CLARK, D.B.; CLARK, D.A.; RICH, P.M.; WEISS, S.; OBERBAUER, S.F. Landscape scale evaluation of understory light and canopy structure: Methods and application in a Neotropical lowland rain forest. **Canadian Journal of Forest Research**, Ottawa, v. 26, n. 5, p. 747-757, 1996.

CONCEIÇÃO, A.A.; PIRANI, J.R. Diversidade em quatro áreas de campos rupestres na Chapada Diamantina, Bahia, Brasil: espécies distintas, mas riquezas similares. **Rodriguésia**, Rio de Janeiro, v. 58, n. 1, p. 193-206, 2007.

COOMES, D.A.; LINES, E.R.; ALLEN, R.B. Moving on from metabolic scaling theory: hierarchical models of tree growth and asymmetric competition for light. **Journal of Ecology**, Oxford, v. 99, n. 3, p. 748-756, 2011.

COOPS, N.C.; HILKER, T.; WULDER, M.A.; ST-ONGE, B.; NEWNHAM, G.; SIGGINS, A.; TROFYMOW, J. T. Estimating canopy structure of Douglas-fir forest stands from discrete-return LiDAR. **Trees-Structure and Function**, Berlin, v. 21, n. 3, p. 295-310, 2007.

COOPS, N.C.; VARHOLA, A.; BATER, C.W.; TETI, P.; BOON, S.; GOODWIN, N.; WEILER, M. Assessing differences in tree and stand structure following beetle infestation using lidar data. **Canadian Journal of Remote Sensing**, Ottawa, v. 35, n. 6, p. 497-508, 2009.

DEAN, T.J.; CAO, Q.V.; ROBERTS, S.D.; EVANS, D.L. Measuring heights to crown base and crown median with LiDAR in a mature, even-aged loblolly pine stand. **Forest Ecology and Management**, Amsterdam, v. 257, n. 1, p. 126-133, 2009.

DORMANN, C.F. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. **Global Ecology and Biogeography**, Oxford, v. 16, n. 2, p. 129-138, 2007.

DRAKE, J.B.; DUBAYAH, R.O.; KNOX, R.G.; CLARK, D.B.; BLAIR, J.B. Sensitivity of large-footprint lidar to canopy structure and biomass in a Neotropical rainforest. **Remote Sensing of Environment**, Amsterdam, v. 81, n. 2/3, p. 378-392, 2002.

DUNGAN, J.L.; PERRY, J.N.; DALE, M.R.T.; LEGENDRE, P.; CITRON-POUSTY, S.; FORTIN, M.-J.; JAKOMULSKA, A.; MIRITI, M.; ROSENBERG, M.S. A balanced view of scale in spatial statistical analysis. **Ecography**, Copenhagen, v. 25, n. 5, p. 626-640, 2002.

ENGELBRECHT, B.M.J.; COMITA, L.S.; CONDIT, R.; KURSAR, T.A.; TYREE, M.T.; TURNER, B.L.; HUBBELL, S.P. Drought sensitivity shapes species distribution patterns in tropical forests. **Nature**, London, v. 447, n. 7140, p. 80-82, 2007.

FELDPAUSCH, T.R.; BANIN, L.; PHILLIPS, O.L.; BAKER, T.R ; LEWIS, S.L.; QUESADA, C.A.; AFFUM-BAFFOE, K.; ARETS, E.J.M.M.; BERRY, N.J.; BIRD, M.; BRONDIZIO, E.S.; CAMARGO, P.; CHAVE, J.; DJAGBLETEY, G.; DOMINGUES, T.F.; DRESCHER, M.; FEARNSIDE, P.M.; FRANÇA, M.B.; FYLLAS, N.M.; LOPEZ-GONZALEZ, G.; HLADIK, A.; HIGUCHI, N.; HUNTER, M.O.; IIDA, Y.; SALIM, K.A.; KASSIM, A.R.; KELLER, M.; KEMP, J.; KING, D.A.; LOVETT, J.C.; MARIMON, B.S.; MARIMON-JUNIOR, B.H.; LENZA, E.; MARSHALL, A.R.; METCALFE, D.J.; MITCHARD, E.T.A.; MORAN, E.F.; NELSON, B.W.; NILUS, R.; NOGUEIRA, E.M.; PALACE, M.; PATIÑO, S.; PEH, K.S.-H.; RAVENTOS, M.T.; REITSMA, J.M.; SAIZ, G.; SCHRODT, F.; SONKÉ, B.; TAEDOUMG, H.E.; TAN, S.; WHITE, L.; WÖLL, H.; LLOYD, J. Height-diameter allometry of tropical forest trees. **Biogeosciences**, Katlenberg-Lindau, v. 8, n. 5, p. 1081-1106, 2011.

FOX, T.R. Sustained productivity in intensively managed forest plantations. **Forest Ecology and Management**, Amsterdam, v. 138, n. 1, p. 187-202, 2000.

GIULIETTI, A.; HARLEY, R.M.; QUEIROZ, L.D.; WANDERLEY, M.G.L.; PIRANI, J.R. Caracterização e endemismos nos campos rupestres da Cadeia do Espinhaço. In: CAVALCANTI, T. B. **Tópicos atuais em botânica.** Brasília: SBB; EMBRAPA, 2000. p. 311-318.

GOSLEE, S.C. Behavior of vegetation sampling methods in the presence of spatial autocorrelation. **Plant Ecology**, Boston, v. 187, n. 2, p. 203-212, 2006.

HE, F.L.; LAFRANKIE, J.V.; SONG, B. Scale dependence of tree abundance and richness in a tropical rain forest, Malaysia. **Landscape Ecology**, Hague, v. 17, n. 6, p. 559-568, 2002.

HIGUCHI, P.; OLIVEIRA-FILHO, A.T.; BEBBER, D.P.; BROWN, N.D.; SILVA, A.C.; MACHADO, E.L. Spatio-temporal patterns of tree community dynamics in a tropical forest fragment in South-east Brazil. **Plant Ecology**, Boston, v. 199, n. 1, p. 125-135, 2008.

HOFMOCKEL, K.S.; ZAK, D.R.; MORAN, K.K.; JASTROW, J.D. Changes in forest soil organic matter pools after a decade of elevated CO2 and O-3. **Soil Biology & Biochemistry**, Oxford, v. 43, n. 7, p. 1518-1527, 2011.

JACOBI, C.M.; CARMO, F.F.D.; VINCENT, R.D.C. Estudo fitossociológico de uma comunidade vegetal sobre canga como subsídio para a reabilitação de áreas mineradas no Quadrilátero Ferrífero, MG. **Revista Árvore**, Viçosa, v. 32, n. 2, p. 345-353, 2008.

JOHN, R.; DALLING, J.W.; HARMS, K.E.; YAVITT, J.B.; STALLARD, R.F.; MIRABELLO, M.; HUBBELL. S.P.; VALENCIA, R.; NAVARRETE, H.; VALLEJO, M.; FOSTER, R.B. Soil nutrients influence spatial distributions of tropical tree species. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 104, n. 3, p. 864-869, 2007.

KELLNER, J.R.; ASNER, G.P.; VITOUSEK, P.M.; TWEITEN, M.A.; HOTCHKISS, S.; CHADWICK, O.A. Dependence of forest structure and dynamics on substrate age and ecosystem development. **Ecosystems**, New York, v. 14, n. 7, p. 1156-1167, Nov 2011.

LEFSKY, M.A.; COHEN, W.B.; ACKER, S.A.; PARKER, G.G.; SPIES, T.A.; HARDING, D. Lidar remote sensing of the canopy structure and biophysical properties of Douglas-fir western hemlock forests. **Remote Sensing of Environment**, Amsterdam, v. 70, n. 3, p. 339-361, 1999.

LEFSKY, M.A.; COHEN, W.B.; PARKER, G.G.; HARDING, D.J. Lidar remote sensing for ecosystem studies. **BioScience**, Oxford, v. 52, n. 1, p. 19-30, 2002.

LEITÃO FILHO, H.D.F. Considerações sobre a florística de florestas tropicais e subtropicais do Brasil. **IPEF,** Piracicaba, v. 35, p. 41-46, 1987.

LOPES, C.G.R.; FERRAZ, E.M.N.; ARAUJO, E.D.L. Physiognomic-structural characterization of dry- and humid-forest fragments (Atlantic Coastal Forest) in Pernambuco State, NE Brazil. **Plant Ecology**, Boston, v. 198, n. 1, p. 1-18, 2008.

MACEDO, R. **Estimativa volumétrica de povoamento clonal de *Eucalyptus sp.* através de laserscanner aerotransportado**. 2009. 143 p. Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2009. Disponível em:<http://urlib. net/sid. inpe. br/mtc-m18@ 80/2009/02.13.04.40>. Acesso em: 10 mar. 2010.

MCCOOL, J.I. **Using the Weibull distribution**: reliability, modeling and inference. New Jersey: John Wiley, 2012. 368 p.

MEDEIROS, M. **Caracterização fitofisionômica e estrutural de áreas de Floresta Ombrófila Densa Montana no Parque Estadual da Serra do Mar, SP, Brasil.** 2009. 85 p. Dissertação (Mestrado em Biodiversidade Vegetal e Meio Ambiente) - Instituto de Botânica, Secretaria do Estado do Meio Ambiente, São Paulo, 2009.

MONTAGHI, A.; CORONA, P.; DALPONTE, M.; GIANELLE, D.; CHIRICI, G.; OLSSON, H. Airborne laser scanning of forest resources: An overview of research in Italy as a commentary case study. **International Journal of Applied Earth Observation and Geoinformation**, Amsterdam, v. 23, p. 288-300, 2013.

MURRELL, D.J. On the emergent spatial structure of size-structured populations: when does self-thinning lead to a reduction in clustering? **Journal of Ecology**, Oxford, v. 97, n. 2, p. 256-266, 2009.

NICOTRA, A.B.; CHAZDON, R.L.; IRIARTE, S.V.B. Spatial heterogeneity of light and woody seedling regeneration in tropical wet forests. **Ecology**, New York, v. 80, n. 6, p. 1908-1926, 1999.

OLIVEIRA FILHO, A.T.; FONTES, M.A.L. Patterns of floristic differentiation among Atlantic forests in southeastern Brazil and the influence of climate. **Biotropica**, Washington, v. 32, n. 4B, p. 793-810, 2000.

OLIVER, M.A. **Geostatistical applications for precision agriculture**. Amsterdam: Springer Netherlands, 2010. 300 p.

PARKER, G.G.; HARMON, M.E.; LEFSKY, M.A.; CHEN, J.; VAN PELT, R.; WEIS, S.B.; THOMAS, S.C.; WINNER, W.E.; SHAW, D.C.; FRANKLING, J.F. Three-dimensional structure of an old-growth Pseudotsuga-Tsuga canopy and its implications for radiation balance, microclimate, and gas exchange. **Ecosystems**, New York, v. 7, n. 5, p. 440-453, 2004.

REUTEBUCH, S.E.; ANDERSEN, H.E.; MCGAUGHEY, R.J. Light detection and ranging (LIDAR): An emerging tool for multiple resource inventory. **Journal of Forestry**, Bethesda, v. 103, n. 6, p. 286-292, 2005.

RICKLEFS, R.E. **The economy of nature.** New York: WH Freeman, 2001. 620 p.

RYAN, M.G.; PHILLIPS, N.; BOND, B.J. The hydraulic limitation hypothesis revisited. **Plant Cell and Environment**, New York, v. 29, n. 3, p. 367-381, 2006.

SAATCHI, S.; MARLIER, M.; CHAZDON, R.L.; CLARK, D.B.; RUSSELL, A.E. Impact of spatial variability of tropical forest structure on radar estimation of aboveground biomass. **Remote Sensing of Environment**, Amsterdam v. 115, n. 11, p. 2836-2849, 2011.

SHIELS, A.B.; ZIMMERMAN, J.K.; GARCÍA-MONTIEL, D.C.; JONCKHEERE, I.; HOLM, J.; HORTON, D.; BROKAW, N. Plant responses to simulated hurricane impacts in a subtropical wet forest, Puerto Rico. **Journal of Ecology**, Oxford, v. 98, n. 3, p. 659-673, 2010.

SILVA, J.M.C.; BATES, J.M. Biogeographic patterns and conservation in the South American Cerrado: a tropical Savanna hotspot. **Bioscience**, Oxford, v. 52, n. 3, p. 225-233, 2002.

SOUZA, A.A.; GALVAO, L.S.; SANTOS, J.R. Relationships between Hyperion-derived vegetation indices, biophysical parameters, and elevation data in a Brazilian savannah environment. **Remote Sensing Letters**, Paris, v. 1, n. 1, p. 55-64, 2010.

TSUI, O.W.; COOPS, N.C.; WULDER, M.A.; MARSHALL, P.L. Integrating airborne LiDAR and space-borne radar via multivariate kriging to estimate above-ground biomass. **Remote Sensing of Environment**, Amsterdam, v. 139, p. 340-352, 2013.

URIARTE, M.; CONDIT, R.; CANHAM, C.D.; HUBBELL, S.P. A spatially explicit model of sapling growth in a tropical forest: does the identity of neighbours matter? **Journal of Ecology**, Oxford, v. 92, n. 2, p. 348-360, 2004.

VALLE, D.; SCHULZE, M.; VIDAL, E.; GROGAN, J.; SALES, M. Identifying bias in stand-level growth and yield estimations: a case study in eastern Brazilian Amazonia. **Forest Ecology and Management**, Amsterdam, v. 236, n. 2, p. 127-135, 2006.

VAN GELDER, H.A.; POORTER, L.; STERCK, F.J. Wood mechanics, allometry, and life-history variation in a tropical rain forest tree community. **New Phytologist**, Oxford, v. 171, n. 2, p. 367-378, 2006.

VASCONCELOS, M. Reserva do Caraça: história, vegetação e fauna. **Aves**, São Paulo, v. 1, n. 1, p. 3-7, 2000.

VASTARANTA, M. **Forest mapping and monitoring using active 3D remote sensing.** 2012. 45 p. Thesis (Doctorate in Forest Science) – Faculty of Agriculture and Forestry, University of Helsinki, Helsinki, 2012.

VELOSO, H.P.; RANGEL FILHO, A.L.R.; LIMA, J.C.A. **Classificação da vegetação brasileira, adaptada a um sistema universal.** Brasília: Ministério da Economia, Fazenda e Planejamento; Rio de Janeiro: Fundação Instituto Brasileiro de Geografia e Estatística, Diretoria de Geociências, Departamento de Recursos Naturais e Estudos Ambientais, 1991. 123 p.

VELOSO, H.P.; OLIVEIRA-FILHO, L.D.; VAZ, A.M.S.F.; LIMA, M.P.M.; MARQUETE, R.; BRAZÃO, J.E.M. **Manual técnico da vegetação brasileira**. Rio de Janeiro: IBGE, 1992. 92 p.

VIANA, V.M.; TABANEZ, A.A.J. Brazilian Atlantic moist forest. In: BRUSH, S.B.; SCHELHAS, J.; GREENBERG, R.S.; STABINSKY, D. (Ed.). **Forest patches in tropical landscapes**. Washington: Island Press, 1996. p. 151.

VIOLA, D.N. **Detecção e modelagem de padrão espacial em dados binários e de contagem.** 2007. 118 p. Tese (Doutorado em Agronomia) – Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, São Paulo, 2007.

WRIGHT, S.J. Plant diversity in tropical forests: a review of mechanisms of species coexistence. **Oecologia**, Berlin, v. 130, n. 1, p. 1-14, 2002.

WULDER, M.A.; HAN, T.; WHITE, J.C.; SWEDA, T.; TSUZUKI, H. Integrating profiling LIDAR with Landsat data for regional boreal forest canopy attribute estimation and change characterization. **Remote Sensing of Environment**, Amsterdam, v. 110, n. 1, p. 123-137, 2007.

WULDER, M.A.; WHITE, J.C.; NELSON, R.F.; NÆSSET, E.; ØRKA, H.O.; COOPS, N.C.; HILKER, T.; BATER, C.W.; GOBAKKEN, T. Lidar sampling for large-area forest characterization: a review. **Remote Sensing of Environment**, Amsterdam, v. 121, p. 196-209, 2012.

6 CONCLUSIONS

"It's not so important who starts the game,
but who finishes it."
John Wooden

The capacity of a sensor to acquire three-dimensional information instead of traditional two-dimensional data has brought a completely new perspective to forestry mapping and monitoring. LiDAR is not the only source of three-dimensional data, although it is by far the most recommended system when it comes to technologies that are easier to operate and that produce results simple to interpret and useful for forest managers and researchers.

The overall methodology applied to quantification of vegetation characteristics is somewhat consolidated, but certain concerns do persist within areas of the scientific community. This thesis aimed to deal with some of these concerns and to try to contribute some new results and insights. The key concerns addressed by this study are: the effect of threshold heights on the quality of the set of metrics; the mechanism by which selection of stable metrics can contribute towards the generation of reliable models between different data sets; the use of machine learning tools in estimating the wood volume of eucalyptus plantations from LiDAR metrics; and the differentiation of varying vegetation communities through vertical profiles derived from airborne laser surveys.

The results revealed some interesting phenomena. The results presented in Chapter 2 demonstrated that the choice of threshold heights concerning the extraction of metrics could influence the set of metrics extracted from LiDAR data. Better metrics can be obtained if echoes from the canopy are separated from those that emanate from below the canopy. This separation is especially recommended for young stands, but has less influence in mature plantations.

Chapter 3 presented metrics that could potentially be used to create transferable models that are based on ALS-derived metrics taken from multiple surveys of the same site on a particular date. The stability of these metrics at both plot and grid levels might contribute towards the generation of reliable models between different data sets. According to our results, height metrics provide the greatest stability when used in these models, specifically the higher percentiles (>50%) and the mode.

Instead of being limited to just a subset of ALS metrics when attempting explain as much variability in a dependent variable as possible, Chapter 4 showed that artificial intelligence tools provide a reliable strategy to explore the existence of any patterns between ALS metrics and stand volume. These tools can be easily directed towards dealing with the problems inherent in estimating stand volume from an ALS data set when they are implemented through the use of a machine learning approach. Such an approach can be implemented through software, via continuous learning of patterns from processed data sets. This characteristic is especially interesting regarding an ALS approach where the point cloud can be processed into a huge number of metrics, and where the relationships between forest parameters and ALS metrics vary widely.

Chapter 5 demonstrated that it is possible to identify various different vegetation communities through their vertical profiles as derived from airborne laser surveys. Determination of the vertical profile was able to be achieved with data representing at least 10% of the area, although for some typologies as little as 2% coverage was sufficient.

The investigations of this thesis reveal that the ALS technology can be successfully applied to the quantification of vegetation characteristics, especially concerning forest formations. Several approaches can be identified, from classical parametric models up to nonparametric artificial intelligence tools. Whilst the former models function with little but the best possible metrics, the latter models look for patterns within a large amount of metrics. Not only should measures of accuracy and precision be considered as guides to the choice of ideal technique, but also the amount of data available, the model transferability and the available computational processing power. This study managed to differentiate between varying vegetation communities based on ALS data sets, paving the way for its usage in large classification and mapping projects.

APPENDIX

List of computed height and density metrics

Metric	Description/Statistic	Symbol for all returns (HM | AR)	Symbol for first returns (HM | FR)
	Height Metrics (HM)		
Location "Center"			
Mean	$\bar{x} = \frac{1}{n}\sum x_i$	HME	HMEf
Quadratic mean	$\sqrt{\frac{1}{n}\sum x_i^2}$	HSQ	HSQf
Cubic mean	$\sqrt[3]{\frac{1}{n}\sum x_i^3}$	HC	HCf
Mode	$Mo =$ Most frequent value in data set	HMO	HMOf
Location "Percentile"			
Percentile	Height separating lowest h% of data from remainder	HP*h*	HP*h*f
Dispersion			
Standard Deviation	$\sqrt{\frac{1}{n-1}\sum(x_i - \bar{x})^2}$	HSD	HSDf
Coefficient of Variation	$\frac{\sqrt{\frac{1}{n-1}\sum(x_i - \bar{x})^2}}{\bar{x}}$	HCV	HCVf
Skewness	$\frac{\frac{1}{n}\sum(x_i - \bar{x})^3}{\left(\frac{1}{n}\sum(x_i - \bar{x})^2\right)^{3/2}}$	HS	HSf
Kurtosis	$\frac{\frac{1}{n}\sum(x_i - \bar{x})^4}{\left(\frac{1}{n}\sum(x_i - \bar{x})^2\right)^2} - 3$	HK	HKf

Cover Metrics (CM)

All returns differentiated by height (CM | AR)

Percentage of all returns above a specified height	$\frac{N(> h_j)}{N} \times 100$	PARA2
Percentage of all returns above the mean height	$\frac{N(> \bar{h})}{N} \times 100$	PARAM
Percentage of all returns above the mode height	$\frac{N(> Mo(h))}{N} \times 100$	PARAMO
Percentage of all returns above the specified height related to first returns	$\frac{N(> h_j)}{N_1} \times 100$	AR2FR
Percentage of all returns above the mean related to first returns	$\frac{N(> \bar{h})}{N_1} \times 100$	ARMFR
Percentage of all returns above the mode related to first returns	$\frac{N(> Mo(h))}{N_1} \times 100$	ARMOFR

First returns differentiated by height (CM | FR)

Percentage of first returns above a specified height	$\frac{N_1(> h_j)}{N} \times 100$	PFRA2
Percentage of first returns above the mean height	$\frac{N_1(> \bar{h})}{N} \times 100$	PFRAM
Percentage of first returns above the mode height	$\frac{N_1(> Mo(h))}{N} \times 100$	PFRAMO
Percentage of first returns above the specified height related to first returns	$\frac{N_1(> h_j)}{N_1} \times 100$	FR2FR
Percentage of first returns above the mean related to first returns	$\frac{N_1(> \bar{h})}{N_1} \times 100$	FRMFR
Percentage of first returns above the mode related to first returns	$\frac{N_1(> Mo(h))}{N_1} \times 100$	FRMOFR