Improvements in forest structural type assessment using airborne laser scanning

Syed Adnan
School of Forest Sciences
Faculty of Science and Forestry
University of Eastern Finland

Academic dissertation

To be presented, with the permission of faculty of Science and Forestry of the University of Eastern Finland, for public criticism via an online Lifesize conference, on 11th December 2020, at 12:00 o'clock noon.

2

ABSTRACT

Accurate forest structural type (FST) assessment provides a valuable support tool to distinguish the different structures in forest stands, achieve sustainable forest management and formulate effective decisions. Data from four research sites within three biogeographical regions – Boreal, Mediterranean and Atlantic – were used in this study, and reliable methodologies were developed for FST assessment. First, the Gini coefficient (GC) of tree size inequality was used for the structural characterisation, and the effects of plot size, stand density and point density of airborne laser scanning (ALS) on the ALS-assisted GC estimations were evaluated for the Boreal region. Second, four forest structural attributes – quadratic mean diameter (QMD), GC, basal area larger than the mean ($BALM$) and stand density (N) – from the three biogeographical regions were used to develop region-independent methods for FST assessment. Lastly, a threshold value to represent maximum entropy was determined and was used to classify the various FST directly from ALS data using L-coefficient of variation and L-skewness of ALS echo heights. Aboveground biomass (AGB) was predicted for each FST and was compared with the AGB predictions without pre-stratification. The results showed that (a) plot size had a greater effect on the ALS-assisted GC estimation compared to stand size and point density, and that 250–450 m^2 plot size (radius 9–12 m for circular plots) is the optimal plot size for reliable ALS-assisted GC estimations, (b) GC and $BALM$ are the most reliable bivariate descriptors for FST assessment, and single storey, multi-storey and reversed-J type forest structures can be separated by lower, medium and upper GC and $BALM$ values, respectively, while QMD and N are relevant for the separation of young/mature and sparse/dense subtypes, and (c) based on the mathematical proofs, the threshold values calculated from ALS echo heights and tree basal areas to represent maximum entropy should be 0.33 and 0.50, respectively. Moderate improvements were observed in the AGB predictions from FST classified directly from ALS data compared to the full dataset but critical differences were identified in the selection of ALS metrics by the prediction models. For example, higher percentiles were more relevant in uneven-sized structures and open canopy areas, while cover metrics and average percentiles were important in the even-sized structures and closed canopy areas. Thus, these results are very useful in improving our understanding of the relationships that underpin the choice of ALS predictors in structurally complex forests.

4

ACKNOWLEDGEMENTS

I wholeheartedly thank my family, parents, sisters, grandfather and uncle, Mr. Qasim Bacha, for their support throughout my life. I dedicate this achievement to my brother Mr. Nasar Khisro Bacha, and my last and most special thanks go to him for always believing in me and encouraging and supporting me to follow my dreams.

Joensuu, December 11, 2020

Syed Adnan

LIST OF ORIGINAL ARTICLES

This thesis is supported by the following three articles that are referred to throughout as Roman numerals in bold. They are reproduced with kind of permission from the publishers. The thesis body summarises the overall objectives, methodologies and results presented in all three articles. Therefore, this thesis should be read in combination with the articles.

I. Adnan, S., Maltamo, M., Coomes, D. A., and Valbuena, R. (2017). Effects of plot size, stand density, and scan density on the relationship between airborne laser scanning metrics and the Gini coefficient of tree size inequality. Canadian Journal of Forest Research, 47(12), 1590–1602. https://doi.org/10.1139/cjfr-2017-0084

II. Adnan, S., Maltamo, M., Coomes, D.A., García-Abril, A., Malhi, Y., Manzanera, J.A., Butt, N., Morecroft, M. and Valbuena, R. (2019). A simple approach to forest structure classification using airborne laser scanning that can be adopted across bioregions. Forest Ecology and Management, 433, pp.111–121. https://doi.org/10.1016/j.foreco.2018.10.057

III. Adnan, S., Maltamo, M., Packalen, P., Mehtätalo, L., Ammaturo, N., and Valbuena. R. (2020). Determining maximum entropy in 3D remote sensing height distributions and using it to improve aboveground biomass modelling via stratification. (Submitted to Remote Sensing of Environment)

GLOSSARY OF ABBREVIATIONS

ABA	Area-based approach
AGB	Aboveground biomass
ALS	Airborne laser scanning
BA	Basal area
BALM	Basal area larger than mean
CART	Classification and regression tree
dbh	Diameter at breast height
DTM	Digital terrain model
FST	Forest structural types
GC	Gini coefficient
GEDI	Global Ecosystem Dynamics Investigation
HCA	Hierarchal clustering analysis
InSAR	Interferometric synthetic aperture radar
ITD	Individual tree detection
kNN	k-nearest neighbour
LiDAR	Light detection and ranging
MD	Mean difference
QMD	Quadratic mean diameter
RMSD	Root mean square difference
SSR	Sum of square ratio
UAV	Unmanned aerial vehicles

TABLE OF CONTENTS

1 INTRODUCTION

1.1 Background

Forest ecosystems are generally described by three main characteristics: composition, function and structure (Franklin 1986). Woody species and various biodiversity variables determine the composition, the rate of ecological processes, such as carbon sequestration, nutrient cycling and species interactions are depicted by function, and the physical characteristics and forest components represent the structure of the forest. All three characteristics are important for forest management and mapping (Latifi 2012). The structural heterogeneity of a forest is a multi-dimensional term and further consists of three main components (Maltamo et al. 2005).

- Vertical component is "a bottom-to-top configuration of the aboveground vegetation within a forest stand" (Brokaw and Lent 1999), for example, understorey vegetation and the number of tree layers/storeys (single storey, two-storeys and multi-storeys). Different vertical structures can be produced by different soil types, climate and tree species, and varies among stands.
- Horizontal component is the spatial distribution of vegetation (Moss 2012).
- Species richness is the total individual species per unit area (Magurran 2005; Pascual et al. 2008). However, when evaluating the structural variability, it can be interpreted as the total number of diameter at breast height (dbh) or height classes (Lexerød and Eid 2006).

Thus, a forest structure is the arrangement and distribution of different tree layers/storeys and variation in species, age and diameter classes (Smith 1997). It is important to evaluate forest structural variations as they create spatial variation in light availability and affect the growth and mortality of seedlings and saplings (Montgomery and Chazdon 2001; Donato et al. 2012). Forest structures also affect the wildlife habitat (food availability, nesting, resting, basking and perching) and the distribution of animal prey (Bell et al. 1991; Hyde et al. 2006), plant habitats (old and damaged trees provide habitats for epiphytic bryophytes and lichens) (Fritz and Brunet 2010), biodiversity (Lelli et al. 2019), long-term biomass predictions (Clark and Clark 2000) and carbon storage (Gove et al. 1995; Marvin et al. 2014). Within stands, the structural components vary in terms of height, canopy, branches and species type, and it is essential to develop objective quantitative approaches using concise indicators that accurately describe the structural heterogeneity. This would provide valuable support tools to (a) distinguish between the different structures in forest stands, (b) encourage sustainable forest management, and (c) promote effective decision making (Bergeron et al. 2002; Coomes and Allen 2007a).

1.2 Approaches and indicators for the evaluation of forest structural diversity

Disparate approaches are available in the literature to describe the complex forest structures and the possible changes that result from natural (growth and mortality) or anthropogenic activities (harvesting) (Pommerening 2002). The definition of forest structure is not explicit as with other forest attributes (e.g. diameter, basal area, dominant height, biomass) and it depends on the observer and the application (Maltamo et al. 2005). These approaches include

tree diameter distributions (Aguirre et al. 2003), age of the forest stand (Spies and Franklin 1991), stand density (Shupe and Marsh 2004) and developmental stage (Valbuena et al. 2013). Similar differences also exist in the quantitative assessment of forest structures (Valbuena et al. 2014). These scientific approaches make it possible to establish, manage and maintain complex forest structures and to achieve sustainability in forest management and planning. Although these approaches are based on small-scale datasets and can only provide variability within a given data range, in practice, they are particularly important when applied *in situ* in forests.

According to McElhinny et al. (2005), various distance-dependent (spatial) and distance-independent (non-spatial) attributes that could be used to evaluate the structural heterogeneity of a forest include:

1. *Abundance.* All common attributes that can be calculated from a given forest stand are included in this category, such as stand density (N; stems ha^{-1}), quadratic mean diameter (QMD; cm), biomass, volume, basal area and dominant height. In operational airborne laser scanning (ALS) forest inventories, these attributes have been well studied (e.g. Maltamo et al. 2014).
2. *Horizontal structure:* This category includes all distance-dependent functions that describe the positional dispersion of components in a population, for example, nearest neighbour analysis (Valbuena 2015), and pair correlation functions (Pommerening 2002). These functions are used to determine variability in the spatial positions of the trees. The indicators included in this category are valuable and could be estimated from ALS data, but they are beyond the scope of this Ph.D. dissertation.
3. *Differentiation:* All distance-independent attributes that compare the relative amount and proportion of variables in a population are included in this category. Differentiation could either be *horizontal* or *vertical* when it is based on tree *dbh* or height, respectively.

Similarly, various biodiversity indicators have been developed to describe species richness and their relative abundance, dominance, diversity and homogeneity (Magurran 2004), but they have also been applied to evaluate forest structural diversity. For the latter, richness describes the number of height or diameter classes, and abundance refers to the relative proportion of stems, basal area, biomass or volume (Pommerening 2002). Popular indicators that are used to evaluate species richness, dominance, diversity and homogeneity are shown in Table 1. Pommerening (2002) and Valbuena (2015) have provided a detailed overview of the various indicators and, based on their reviews, the most suitable indicators that have been used in this research are presented in more detail in the following sections.

1.2.1 Gini coefficient of tree size inequality

The Gini coefficient (GC) was originally developed by Gini (1921) to evaluate inequality in income distribution. Due to its robust statistical properties, researchers highlighted its usefulness in other fields, such as variability in wastewater discharge (Sun et al. 2010), variation in land uses (Zheng et al. 2013), microbial diversity (Harch et al. 1997; Cai et al. 2019) and inequality in the quality of health (Asada 2005).

Table 1. Summary of the popular indicators used for the species richness, dominance, diversity and homogeneity/inequality assessment.

Indicator	Assessment	References
Margalef (D_{Mg})	Species richness	Clifford and Stephenson (1975); Lexerød and Eid (2006)
Menhinick (D_{Mn})		Whittaker (1977)
Berger-Parker index (D_{BP})	Dominance	Berger and Parker (1970); Lexerød and Eid (2006)
Simpson index (D_{Si}) / Simpson evenness ($E_{1/D}$)	Diversity	Simpson 1949; Smith and Wilson (1996); Lexerød and Eid (2006)
McIntosh (D_{MI})/ McIntosh evenness (E_{MI})		McIntosh (1967); Lexerød and Eid (2006)
Shannon Index (H')/ Shannon evenness (J')		Shannon (1948); MacArthur and MacArthur (1961); Neumann and Starlinger (2001); Gove et al. (1995); Rouvinen and Kuuluaainen (2005); O'Hara et al. (2007); Motz et al. (2010); von Gadow et al. (2012)
de Camino homogeneity (CH)	Homogeneity/Inequality	de Camino (1976); Bachofen and Zingg (2001)
Structural index based on variance ($STVI$)		Staudhammer and LeMay (2001)

In plant sciences, GC has been applied, for example, when evaluating inequality in plant size (Weiner and Solbrig 1984; Knox et al. 1989), successional stages (Valbuena et al. 2013) or competition (Cordonnier and Kunstler 2015). In forest sciences, GC is used to appraise inequality among trees sizes growing in a forest area (Weiner and Thomas 1986) and is calculated as follows (Glasser 1962):

$$GC = \frac{n}{(n-1)} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |g_i - g_j|}{2n^2 \bar{g}} \tag{1}$$

where, n represents the total number of trees, $\bar{g}$ is the mean basal area, and g_i and g_j are the basal areas of the ith and j$^{\text{th}}$ trees.

Thus, GC describes the shape of tree diameter distribution, which is influenced by tree interaction and competition (Valbuena et al. 2016a), discriminates between stands with different diameter distributions (Cordonnier and Kunstler 2015) and provides logical ranking for different forest structural types (FST) (Lexerød and Eid 2006; Lei et al. 2009; Adhikari et al. 2020). The GC values range from 0 to 1 (perfect equality to maximum inequality) (Gini 1921), while Valbuena et al. (2012) argue that 0.50 represents maximum entropy and the boundary line between even-sized and uneven-sized forest structures. In practice, GC values < 0.50, close to 0.50 or much > 0.50 demonstrate normal distribution found in even-sized stands (Coomes and Allen 2007b), irregular size distribution (Duduman 2009) and reversed-J shaped distributions, respectively (Valbuena et al. 2013).

1.2.2 Basal area larger than the mean

Basal area larger than the mean ($BALM$) is an indicator of the structural heterogeneity of a forest and had been largely ignored by the scientific community until Gove (2004) demonstrated its usefulness as a structural guide for the decision-making process in the prescription of silvicultural activities (Ginrich 1967). It is calculated as the sum of basal area (BA: m^2 ha^{-1}) of all trees whose diameter is > the quadratic mean diameter (QMD; cm), as shown in Figure 1 (Gove 2004). $BALM$ describes the skewness of the tree diameter distribution and high $BALM$ values indicate competitive conditions that exist in the closed canopies dominated by mature trees. In contrast, lower $BALM$ values denote open canopies with dense understorey ingrowths because the proportion of trees with basal areas > QMD increases, for example, in reversed-J type forest structures. It can also be used to assess the relative dominance of tree layers, whether the biomass is stored in one or many vegetation layers/storeys, and the ecology of species with a preference for forests with single storey or multi-storeys structures (Mononen et al. 2018). Valbuena (2015) has postulated that $BALM$, together with the GC of tree size inequality, could be used as an independent bivariate descriptor to fully describe forest structures, and indicate whether tree interactions are dominated by symmetric (resource depletion) or asymmetric competition (resource pre-emption).

1.2.3 Quadratic mean diameter and stand density

Two other common forest descriptors that describe the location and density of diameter distributions are QMD and N (Gove 2004). These descriptors are crucial in forest structure characterisation. The QMD can be defined as the dbh of a tree that has an average basal area, while N is the stem number per unit area (Curtis 1982; Curtis and Marshall 2000). These descriptors are useful in determining the occurrence of mortality and the need for thinning or planting in forest stands, determination of aboveground biomass (Vincent et al. 2014), influence of fragmentation on species and forest structure (Echeverría et al. 2007), the maximum limits of density and the development of stand density management diagrams, which are used to illustrate the relationships between density, mortality and yield throughout the stand development period. These descriptors help to minimise the trees competition for resources and optimise the wildlife habitat by regulating the density of stems and their spatial arrangement (Newton 1997).
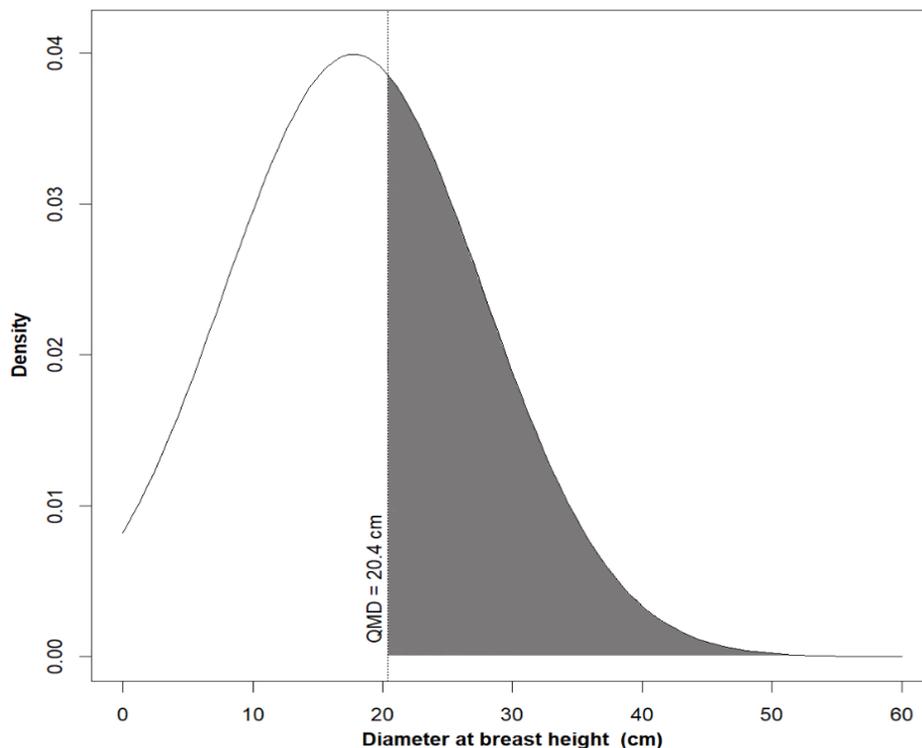
Figure 1. Graphical representation (shaded region) of basal area larger than the mean (*BALM*).

1.3 Assessment of forest structural attributes from ALS

Airborne laser scanning (ALS) produces three-dimensional (3D) canopy information and is considered as a highly effective tool because it provides numerous opportunities to monitor forest stands and obtain reliable results of forest structural properties (Gobakken and Nasset 2008; Latifi 2012). In forest monitoring, detailed canopy information is more useful than other remote sensing approaches (Maltamo et al. 2006). The ALS-derived metrics describe the key characteristics of a forest and are valuable for the prediction and monitoring of various attributes, such as tree species (Van Aardt et al. 2008), height (Maltamo et al. 2004), diameter distribution (Räty et al. 2018), volume (Næsset, 1997), spatial patterns of the trees (Packalen et al. 2013), structural complexity of the forests (Valbuena et al. 2013), biomass and carbon stocks (Næsset and Gobakken 2008; Valbuena et al. 2017a), and wildlife habitats (Hagar et al. 2020). Moreover, ALS data is also reliable for the evaluation of canopy changes and to compare different forest areas (McInerney et al. 2010). ALS-based retrieval and inventory of these structural attributes can be accomplished by two main approaches.

1.3.1 Area-based approach

In the area-based approach (ABA), the ALS metrics that describe the vegetation components are derived from a given field plot or grid cell and are then linked with the forest attributes derived from the same field plot (Maltamo et al. 2014, Chapter 1). ALS metrics, such as dominant tree species or mean height and height percentiles are used as predictors and forest attributes are used as response variables (Yu et al. 2010). Several studies have used ABA to describe the relationship between forest variables and ALS metrics. These include the prediction of dbh, basal area, volume, biomass or height using linear regression (Means et al. 2000; Næsset 2002), non-linear regression (Packalén et al. 2011) or non-parametric approaches (Packalén and Maltamo 2006; Yu et al. 2010; Andersen et al. 2011; Räty et al. 2020). Some studies have also identified the factors that affect the performance of ABA, such as plot size (Gobakken and Næsset 2008), sample size (Junttila et al. 2013), errors in plot positions (Gobakken and Næsset 2009; Rana et al. 2014), and the resolution of the cell (Packalen et al. 2019). However, this method is most often applied in operational forest inventories that employ ALS data (Maltamo et al. 2014), and is more flexible and robust for diameter predictions, for example, in boreal managed forests dominated by coniferous species (Räty et al. 2020).

1.3.2 Individual tree detection approach

In the individual tree detection (ITD) approach, the individual treetops are detected and a set of allometric models is then used for features extraction and tree attribute measurements (Maltamo et al. 2014, Chapter 1), which can later be aggregated to the plot or stand level. In these models, tree height and crown dimensions are used as inputs (Yu et al. 2010). The ITD approach depends on the canopy-height model, which is obtained by interpolating ALS heights. However, not all trees can be detected with the ITD approach as the performance of this method depends on the detection algorithm and its parameterisation (Kaartinen et al. 2012), and on forest conditions, such as stand density, canopy closure and the spatial arrangement of trees (Vauhkonen et al. 2012). Nevertheless, ITD is a suitable alternative to extract and monitor forest attributes at a much finer spatial scale (Kukkonen et al. 2019).

1.4 Existing gaps in forest structural heterogeneity assessments

1.4.1 Factors that influence estimation of the Gini coefficient

The Gini coefficient of tree size inequality is one of the best indicators for the evaluation of the structural heterogeneity of a forest (Lei et al. 2009; Valbuena et al. 2013), although the ALS-assisted GC estimates are affected by plot size and stand density (Matos 2014). In forest science, circular or rectangular sample plots are typically used to measure forest attributes (Whittaker 1972; Kent and Coker 1992) and they range from finer to coarser scales (Chytrý and Otýpková 2003), but forest attribute monitoring at larger scales in field inventories is economically and operationally limited (Almeida et al. 2019). As the size of the sample plot increases, its effect decreases (Barbeito et al. 2009), therefore, an optimal plot size is needed that should be sufficiently large to obtain reliable measurements but not larger than the required size due to the costs involved (Chytrý and Otýpková 2003). The structural diversity

obtained by an indicator, for example GC, also relies on the ALS spatial resolution (Mascaro et al. 2011) and the information retrieved may change if the scale of the observation is changed due to the aggregation of various stand conditions (Coomes and Allen 2007b). Spatial resolution stands for the plot size or the pixel size at which ALS metrics are computed (Ruiz et al. 2014; Packalen et al. 2019). Similarly, scan density is an important aspect of ALS that affects both the processing and the cost of the ALS data (Thomas et al. 2006; Kandare et al. 2016). Various studies have evaluated the effect of ALS scan density on the accuracy of digital terrain models (DTM) (Liu et al. 2007) and the measurement of ALS heights and biophysical stand properties (Gobakken and Næsset 2008). However, there is a gap in the existing scientific literature as to how plot size, stand density and ALS scan density affect the GC estimates.

1.4.2 Cross-bioregional assessment of forest structure

Forest structure is one of the essential properties of a forest ecosystem and influences the microclimate, carbon storage, wildlife habitats and biodiversity (Hyde et al. 2006; Hansen et al. 2014). Forest researchers have developed various approaches in the past to measure the structural properties of a forest, but these approaches were often laborious and restricted to small sampling areas (Weltz et al. 1994; Chytrý and Otýpková 2003). In Finland, various development classes, such as seedling, sapling, young thinning, advanced thinning, mature stands, seed-trees and multi-storeys have been used to separate the different stands, which assists in the management, planning and decision making for large forest areas (Valbuena et al. 2016b). With the advent of remote sensing, the ability to quantify forest structural changes has improved considerably (Hyde et al. 2006). For example, Næsset and Gobakken (2008) used photo interpretation of stereo images and classified various inventory plots according to the site index, age class and tree species composition before biomass estimation, while Nelson et al. (2008) estimated the aboveground biomass in predefined aerial-photo-based forest classes. Similarly, ALS has been used to quantify structural properties, such as tree height, canopy cover and layering in specific forest stands (Hansen et al. 2014). Forests have also been classified into various FST in the literature: regeneration/understorey growth (Gougeon et al. 2001), sparse and dense forest stands (Fassnacht et al. 2017), young and mature forest stands (Spies and Franklin 1991; Næsset 2002), single layer/storey to multi-layers/storey forest structures (O'Hara and Gersonde, 2004; Zhang et al. 2011), and reversed-J type forest structures, which are characterised by a peak on the right side of the distribution curve where mature trees account for the maximum proportion of the basal area (Valbuena et al. 2013). However, the forest attributes or indicators, and the approaches used for such forest structural assessments, are disparate and the definition of FST varies from one application to another (Latifi 2012; Valbuena et al. 2013). Therefore, a region-independent objective quantitative approach is needed for the structural assessment of forests, which could be applicable across different forest types and biogeographical regions.

1.4.3 Aboveground biomass predictions from FST detected directly from ALS data

Aboveground biomass (AGB) estimation from the local to the global scale is important because it quantifies carbon sequestration in forests and assists in better forest management and planning (Boudreau et al. 2008). Remote sensing technologies in general and ALS in particular play key roles in the monitoring of forest resources at the regional scale (Næsset et al. 2011) and contribute to better global policies and decision-making, e.g., in REDD (Reduce

Emissions from Deforestation and forest Degradation) activities (Angelsen et al. 2009). Various studies have used remotely sensed data and have estimated forest biomass with varying degrees of success (Foody et al. 2001; Kankare et al. 2013; Su et al. 2016). Researchers have also employed ALS data and predicted forest attributes, including AGB (Kankare et al. 2013; Maltamo et al. 2016; Bouvier et al. 2015; Nguyen et al. 2019; Knapp et al. 2020), although the prediction precision depends on the relationship between the foliage observed by ALS and the various AGB components, because the ALS pulses are mainly blocked by foliage (Næsset and Gobakken 2008; Rocha de Souza Pereira et al. 2018). Similarly, the structural complexity of a forest can cause difficulties in modeling. For example, a general equation cannot be applied to all regions, to both sparse and dense forests or to even- and uneven-sized forest structures (Chave et al. 2005; Häbel et al. 2019). This problem can be solved by stratifying the forest into different FST using a threshold value to represent maximum entropy, and a separate biomass prediction model can be developed for each stratum (Valbuena 2017b), or the forest structural information may be included in the AGB modeling (Bouvier et al. 2015; Knapp et al. 2020). Valbuena et al. (2017b) identified various FST directly from ALS data using the L-coefficient of variation (L_{cv}), which is equivalent to the Gini coefficient calculated from ALS echo heights and L-skewness (L_{skew}) of ALS echo heights. They used a threshold value of $L_{cv} = 0.50$ to represent maximum entropy and to separate even- and uneven-sized FST, although determining maximum entropy from a distribution of ALS echo heights should be different than tree basal areas. Therefore, it is important to use appropriate methods for the structural classification of forests and to understand how forest structural information are related to AGB estimation. This would provide useful information for the enhancement of forest structural characterisation and improve large scale biomass mapping and their integration in better forest management and planning (Wulder et al. 2008; Knapp et al. 2020).

1.5 Objectives of the research

The basic aim of this doctoral dissertation is to improve FST assessment, by the development of consistent, replicable and region-independent methodologies. To ensure consistency, simple indicators and forest attributes that can be easily obtained from forest inventory data have been used, while replicability and region-independency has been achieved by using ALS data in all studies. Methodologies developed in this doctoral dissertation have the potential to assist in the large-scale mapping and regional comparison of forest structures. The specific objectives of the research are:

1) To study plot size, stand density and ALS density effects on the relationship between GC of tree size inequality and ALS metrics, and to develop a simple method to select the optimal plot size for GC estimation from field data and its prediction from ALS data (**I**).
2) To develop region-independent methodologies by using four forest attributes – GC, $BALM$, QMD and N– obtained from Boreal, Mediterranean and Atlantic biogeographical regions, achieve a full description of FST, which contains all possible forest structural components, and evaluate the capacity and reliability of ALS data in acquiring those FST (**II**).
3) To detect the various FST directly from ALS data using L-coefficient of variation and L-skewness of ALS echo heights, develop an AGB prediction model for each FST and compare that model with a general AGB prediction model that contains the full dataset without prior stratification (**III**).

2 MATERIALS AND METHODS

2.1 Research sites and data collection

As the main goal of this dissertation was to develop region-independent methodologies for the structural characterisation of forests, I used field and ALS data from four research sites within three biogeographical regions (Boreal, Mediterranean and Atlantic) (Figure 2).

2.1.1 Kiihtelysvaara inventory area, Finland (Boreal)

Kiihtelysvaara is a boreal inventory area located in the eastern region of Finland (62°31' N, 30°10' E) and is managed for ecological sustainability and timber production. Scots pine (*Pinus Sylvestris* L.) is the main tree species and constitutes 73 % of the total wood volume, while Norway spruce (*Picea abies* (L.) Karst.) accounts for 16 %. The remaining 11 % is derived from deciduous species; downy birch (*Betula pubescens* Ehrh.) and silver birch (*B. pendula* Roth.) (Packalen et al 2013). A field inventory was carried out from May to June 2010 and data were collected from 79 squared field plots of various dimensions (20×20 m, 25×25 m, 30×30 m) (Maltamo et al. 2012). First, stratified random sampling was employed and the forest stands were selected and plots were then deliberately established at representative locations to avoid the placing of plots at the border of the stands due to the high costs and efforts required to measure all the trees. Before field data collection, the position (latitude and longitude) of all trees was recorded from high resolution ALS data using the ITD method (Packalen et al. 2013). Those tree positions were validated in the field, and the *dbh* of all trees with a height > 4 m or *dbh* > 5 cm were then measured. ALS data were collected on June 29, 2009 using an ATM Gemini sensor (Optech, Canada) from 600–700 m above ground surface with 26° field of view and 125 kHz pulse rate. The scan width and overlap between the strips were 320 m and 55 %, respectively. The average density of the ALS data was 11.9 points m^{-2}. Field and ALS data from the Kiihtelysvaara inventory area were used in studies **I** and **II,** but the larger field plots were reduced to 20×20 m to ensure consistency with the other two regional sites (Valsaín forest, Spain, and Wytham Woods, United Kingdom) used in study **II**.

2.1.2 Joensuu inventory area, Finland (Boreal)

This inventory area is located in the North Karelia region of eastern Finland (62°15' N, 30°13' E). The total area is approximately 252,000 ha and Scots pine, Norway spruce and birch species are the dominant species. Other deciduous species, such as *Alnus* and *Populus* are present but at a minor scale. The whole inventory area was divided into eight different strata based on development classes, such as seedling, sapling, young thinning, advanced thinning, mature, seed trees, shelterwood and multi-storey, and 244 field plots were randomly collected by University of Eastern Finland and Finnish Forest Centre (Suomen Metsäkeskus; SMK) in a joint collaboration in 2013. An approximately equal number of sample plots were collected from each stratum and the field data included species, *dbh* and height information. The detailed field data acquisition strategy is described in Valbuena et al. (2016b). For ALS data collection, a Leica ALS60 system was used at 2300 m above ground surface in May 2012

under leaf-off conditions, and the average point density of the ALS data was 0.91 points m^{-2}. Data from the North Karelia inventory area were used in study **III**.

2.1.3 Valsaín forest, Spain (Mediterranean)

Valsaín is located in the Segovia province, Spain (40°48′N, 4°01′W) at 300–1500 m above sea level. It is a drought-adapted Scots pine shelterwood managed forest (Valbuena et al. 2013). In summer 2006, field data were collected in 37 circular field plots (20 m radius). All seedlings and saplings were recorded in the inner 10 m radius of the sample plot, while trees with $dbh > 10$ cm were measured in the outer 20 m radius. In the same year, ALS data were obtained in September using ALS50-II Leica Geosystems (Switzerland) from 1500 m above ground surface. The field of view was 25° and the scan was performed in a bidirectional manner with 665 m width and 40 % side overlap. The average point density was 1.15 points m^{-2}. Data from Valsaín forest was used in study **II**.

2.1.4 Wytham Woods, UK (Atlantic)

Wytham woods is a deciduous forest located in the Oxfordshire, UK (51°46′N, 1°20′W). Ash (*Fraxinus excelsior*), sycamore (*Acer pseudoplatanus*), maple (*Acer campestre*), oak (*Quercus robur*) and hazel (*Corylus avellana*) and are the dominant species in this forest (Savill et al., 2011). The data, which included dbh of stems > 1 cm, were collected in 2010 from an 18-ha permanent plot. This permanent plot was divided into 450 subplots (20×20 m each). Low-resolution ALS data, with 0.198 points m^{-2} average point density, were collected in June 2014 using a Leica ALS50-II LiDAR system from 2500 m above sea level. The field of view and pulse rate were 35° and 69.8 kHz, respectively.

2.2 Processing of ALS data

In all studies (**I–III**), FUSION software of the USDA Forest Service (McGaughey 2015) was used and area-based metrics were calculated from ALS echo heights > 0.1 m. The 0.1 m limit was used to avoid the lower echo heights, which could be reflected from the ground surface. Prior to ALS metrics calculation, the last echoes of ALS data were extracted and interpolated into a DTM, which was then subtracted from the ALS echo heights to avoid terrain effects on ALS metrics calculations. The ALS metrics are the statistics of ALS height distribution that could be related to various forest attributes (Table 2). For example, minimum, mean and maximum ALS echo heights are related to minimum, mean and dominant tree heights, cover (percentage of all returns above a specified height) is used to represent stand density, standard deviation of ALS echo heights is related to variation in tree heights, and L_{cv} and L_{skew} is used to assess tree size inequality and dominance, respectively. These metrics are used as auxiliary information in ALS-assisted estimation of forest variables (Næsset 2002).

2.3 Optimal plot size selection for ALS-assisted Gini coefficient estimation (I)

The first task in optimal plot size selection was to simulate concentric circular plots (hereafter referred to as simulated circular plots) that ranged from 1–15 m radius within each original field plot (79 plots in total). The number of simulations (n=700) were selected based on a
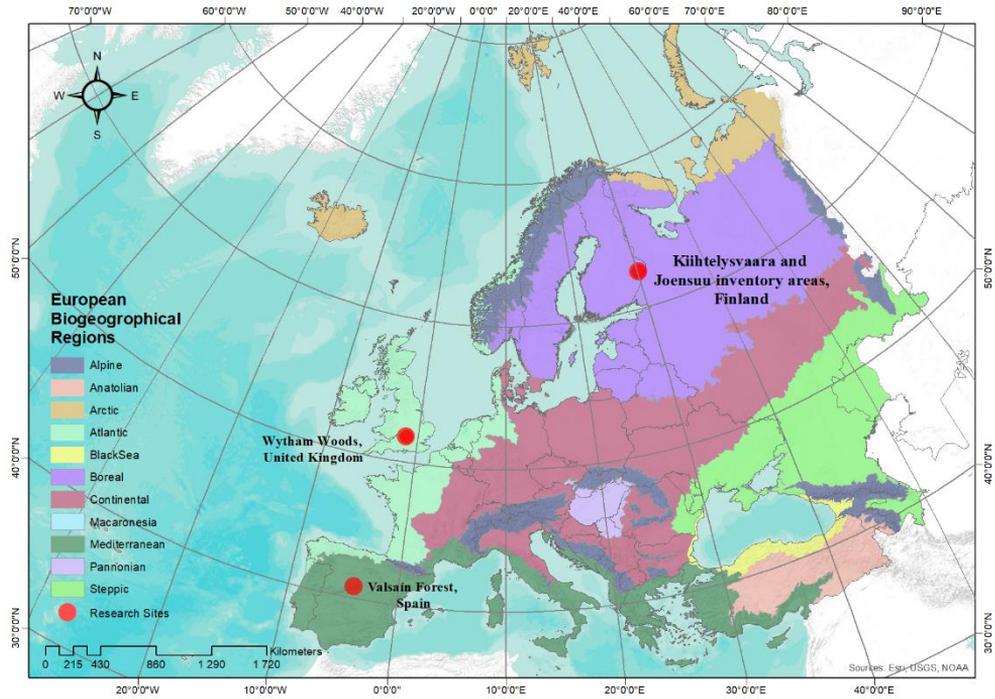
Figure 2. Map showing the location of the research sites used in this doctoral dissertation within three biogeographical regions (European Environmental Agency 2020).

Table 2. Airborne laser scanning (ALS) metrics and their corresponding forest characteristics.

Notation	Explanation	Relevant forest characteristics
$Max/P99$	Maximum ALS height over an area/99th percentile	Dominant height of tree
$P50$	50th percentile of ALS echo heights	Mean height of tree
$P25$	25th percentile (1st quartile)	Understorey growth
$Cover$	Percentage of all returns above 0.1 m	Canopy cover/stand density
$StdDev$	Standard deviation in ALS echo heights	Variation in tree heights
CRR	Canopy relief ratio	Vertical structure
L_{cv}	L-coefficient of variation of ALS echo heights	Tree size inequality
$L_{skew}/Skew$	L-skewness/skewness of ALS echo heights	Tree dominance

sensitivity analysis. The spatial distribution of the trees was replicated around the original field plot to overcome the edge effects that produce bias in statistical calculations (Diggle 2003; Pommerening and Stoyan 2006). Then, a random position was selected within each original field plot and the GC calculation was repeated within the simulated circular plots (1–15 m radius) using equation 1. The absolute position (latitude and longitude) of all simulations were recorded and was used to extract the corresponding ALS metrics at a later stage. The average $\overline{GC}$ value was computed for each simulated circular plot within each original field plot. Thereafter, all GC values were directly compared using the absolute GC differences ($\overline{GC}_{diff}$). The $\overline{GC}_{diff}$ was calculated by subtracting the $\overline{GC}$ value of all simulated circular plots from a reference GC_{ref} value (calculated from a reference field plot). The $\overline{GC}_{diff}$ value was useful for the evaluation of all the simulations, and provided the first stabilisation criterion for stable GC estimation.

$$\overline{GC}_{diff} = |GC_{ref} - \overline{GC}| \qquad (2)$$

The GC calculation and the accuracy of ALS-assisted estimation of any forest attribute depends on a basic relationship that exists between the plot size and the sample size. Therefore, the stand density (n) in a simulated circular plot size s (radius) is related to the stand density (N) of the original field plot by:

$$n = N\pi s^2 \qquad (3)$$

A similar relationship between the point density (p) within the same simulated circular plot s (radius; m) is also tied to the point density (d; points m^{-2}) of the original field plot.

$$p = d\pi s^2 \qquad (4)$$

An interesting question emerges here as to whether the optimisation should be based on a plot size (spatial resolution in the case of ALS-assisted estimations) or a sample size (N or d) because they are both directly related to each other. Therefore, the same procedure was replicated to select the optimal plot size and sample size for reliable ALS-assisted GC estimation.

2.3.1 Criteria for plot size and sample size optimisation

Two criteria were set for the plot size and sample size optimisation. First, stabilisation of the GC values at a given plot size (s) or sample size (n) was achieved by observing the $\overline{GC}_{diff}$ value for increasing s or n, where the estimation of the GC value (at $\overline{GC}_{diff} = 0.05$) was considered to be stable. Second, maximisation of the absolute correlation $|r|$ between the GC values and ALS metrics was calculated. Any plot size s or sample size n that fulfilled the above two criteria were considered optimal plot size s^* or sample size n^*.

$$s^* = \overline{GC}_{diff} < 0.05 \,\big|\, max|r| \qquad (5.1)$$

$$n^* = \overline{GC}_{diff} < 0.05 \,\big|\, max|r| \qquad (5.2)$$

After optimal plot size s^* was selected, the varying ALS point density (d) effects were investigated. The original ALS point density (11.9 point m^{-2}) was decreased to 0.50, 0.75, 1, 3, 5, 7.5 and 10 points m^{-2} using the appropriate thinning factor (Ruiz et al. 2014) and the methods included in the LAStools software (Jakubowski et al. 2013; RapidLasso GmbH Inc.: Isenburg 2016). For each reduced point density, new ALS metrics and their correlation with the GC values were calculated. In addition, the effect of changing ALS point densities on the absolute correlation $|r|$ between the GC values and the new ALS metrics was examined.

2.4 Cross-bioregional FST assessment (II)

Four forest attributes – GC, BALM, QMD and N– were calculated from the three biogeographical regions –Boreal, Mediterranean and Atlantic, and they were grouped into two broad categories: coniferous forest, which included data from Boreal (Finland) and Mediterranean (Spain) regions, and deciduous forests, which included data from the Atlantic bioregion (UK). In the first task, hierarchal clustering analysis (HCA), which merges (agglomerative procedure) or splits (divisive procedure) all observations on the basis of proximity measures, such as Euclidean distance, was applied and potential clusters (FST) were obtained for both coniferous and deciduous forests using the aforementioned four forest attributes. However, since the data were in different units, treating them in their original scale would place an unreasonable weighting on some forest attributes over others. To overcome this bias, standardisation of the original attributes using a range-equalisation method was performed prior to Euclidean distance calculation, and each attribute was normalised to a 0– 1 scale. Then, the optimum number of clusters c was decided based on a distortion curve (Sugar and James 2003; Everitt et al. 2011), and the *hclust* function included in the R package *fastcluster* (Müllner 2013) was applied to separate both coniferous and deciduous forests into the optimum number of clusters.

Since my interest was to determine the empirical threshold values of the forest attributes and use them to separate the various FST, the CART analysis (classification and regression tree) included in the R package *rpart* (Breiman et al. 1984) was applied. In this analysis, the four forest attributes (GC, $BALM$, QMD and N) were used as explanatory variables, and the potential clusters (FST) obtained from HCA were used as response variables. The data were split into the optimum number of clusters that were identified at the HCA stage, and the results resembled a tree where the classification decision (threshold values of the forest attributes) was given at each node between the two branches.

The FST obtained in the previous stage were finally predicted from the ALS data by applying the widely used k-nearest neighbour (kNN) method included in the R package *class* (Venables and Ripley 2001). This method is a supervised machine learning method and is widely used for the prediction of various forest attributes, such as volume, biomass, stand density, and basal area (Maltamo and Kangas 1998; Franco-Lopez et al. 2001; Breidenbach et al. 2012). In the kNN method, four area-based ALS metrics, such as maximum ALS return, percentage of all returns > 0.1 m, L-coefficient of variation and L-skewness of ALS echo heights were used because these metrics could be related to the QMD, N, tree size inequality and tree dominance, respectively (Zimble et al. 2003; Valbuena et al. 2017b). Leave-one-out cross validation was used for accuracy assessment, and the bias was determined as the difference between producer and user accuracies. For the former, accuracy is the proportion of the observed field plots for a FST that are classified as correct, whereas for the latter, accuracy is the proportion of field plots that are correctly classified as FST (Story and

Congalton, 1986). The kappa coefficient (k) and overall accuracy (OA), which are included in the R package *vcd* (Meyer et al. 2014), were used to evaluate potential misclassification.

2.5 ALS-based forest structural type assessment and aboveground biomass prediction (III)

Two L-moment ratios (L-coefficient of variation and L-skewness) were used to classify different FST directly from ALS data. L-moments are the same as the conventional moments but are more reliable and robust to measure the properties of a probability density distribution (Frazer et al. 2011). L-moments are based on the expected value $E(X_{n:s})$ in a sample order statistic $X_{n:s}$, where n is the smallest observation in sample size s, and are restricted by fixed intervals (Hosking 1990). L_{cv} is the ratio between the second ($L2$) and the first ($L1$) L-moments (equation 6), while L_{skew} is the ratio between the third ($L3$) and the second ($L2$) L-moments (equation 7).

$$L_{cv} = \frac{L2}{L1} = \frac{E(X_{2:2}) - E(X_{1:2})}{2E(X)} \tag{6}$$

$$L_{skew} = \frac{L3}{L2} = \frac{E(X_{3:3}) - 2E(X_{2:3}) + E(X_{1:3})}{E(X_{3:3}) - E(X_{1:3})} \tag{7}$$

where $E(X)$ is the expected values of X which represents the ALS echo heights.

L_{cv} is mathematically equivalent to the GC of tree size inequality calculated from ALS echo heights, bounded between [0,1] intervals and is useful to discriminate between even- and uneven-sized FST (Valbuena et al. 2017b: Appendix A3). Valbuena et al. (2017b) used a threshold value of $L_{cv} = 0.50$ to represent maximum entropy. However, my mathematical findings (see Appendix A in **III**) showed that the threshold value to represent the maximum entropy calculated from ALS echo heights should be 0.33, as compared to the 0.50 threshold value calculated from tree basal areas. Therefore, the $L_{cv} = 0.33$ threshold value was used in this study to represent maximum entropy and to separate the even- ($L_{cv} < 0.33$) and uneven-sized ($L_{cv} > 0.33$) structures. On the other hand, L_{skew} is bounded between [-1,1] intervals (Hosking, 1989) and can be useful to evaluate canopy closure (open canopies vs closed canopies) (Lefsky et al. 2002). $L_{skew} = 0$, which represents the symmetric distribution, was used to separate the open canopies ($L_{skew} > 0$) from closed canopies ($L_{skew} < 0$).

2.5.1 Aboveground biomass prediction and accuracy assessment

In this step, tree level aboveground biomass (kg) was calculated using species-specific biomass equations, such as for birch (Repola 2008) and Scots pine and Norway spruce (Repola 2009). These equations require the dbh and height of each species as inputs. Missing tree heights were predicted using Näslund's height curve model (1936) as presented by Siipilehto (1999). Prior to the tree height predictions, the species-specific height (H_{gM}) and diameter (D_{gM}) with median basal area were calculated and were used to determine the parameters of Näslund's height curve model. These parameters were used in the model to predict the missing tree height from tree dbh. Finally, aboveground biomass estimates were

aggregated to the plot level (Mg ha^{-1}) and used as a response variable in the subsequent prediction models.

To predict AGB from ALS data, a best subset of ALS metrics (predictors) was first selected for the general model including the full dataset (without pre-stratification) and for each FST (even- and uneven-sized, and open and closed canopy FST) using function "*regsubset*" of the R package "*leaps*". Then, the kNN method was applied and the AGB was predicted from the best subset of ALS predictors in each model. The results of the observed and predicted AGB were evaluated using root mean square difference (RMSD) and mean difference (MD):

$$RMSD = \sqrt{\frac{\sum_{i=1}^{n}((y_i^{cv} - \hat{y}_i)^2}{n}} \qquad (8)$$

$$MD = \frac{\sum_{i=1}^{n}((y_i^{cv} - \hat{y}_i)}{n} \qquad (9)$$

where n is the total number of observations (field plots), y_i^{cv} and $\hat{y}_i$ are the predicted values using cross validation and the observed value of AGB for observation i.

An additional restriction (sum of square ratio (SSR)) was used to avoid overfitting of the models. SSR is the ratio between the squared root sum of square obtained from cross validation (SS_{cv}) and without cross validation (SS_{fit}).

$$SSR = \sqrt{SS_{cv}}/\sqrt{SS_{fit}} \qquad (10)$$

$$SS_{cv} = \sum_{i=1}^{n}(y_i^{cv} - \hat{y}_i)^2 \qquad (11)$$

$$SS_{fit} = \sum_{i=1}^{n}(y_i^{fit} - \hat{y}_i)^2 \qquad (12)$$

where $\hat{y}_i$ is the observed value of AGB, and y_i^{cv} and y_i^{fit} are the predicted AGB values with cross validation and without cross validation for observation i, respectively.

3 RESULTS

3.1 Optimising the ALS-assisted Gini coefficient estimation (I)

3.1.1 Plot and sample size optimisation for the Gini coefficient of tree size inequality

The results of the first criterion used to devise the minimum plot size or sample size that could produce a stable GC estimation of the population are shown in Figure 3. The GC estimation at the smaller plot sizes and sample sizes were very unstable and only a few smaller simulated circular plots produced a stable GC estimation, most likely in the very even-sized stands. The larger simulated circular plots produced stable GC estimations (see Figure 3a in **I**). The GC stabilisation started at the 6 m radius plot size where 100 % of the original field plots were below the $\overline{GC}_{diff} < 0.05$ limit (Figure 3). Thus, the minimum plot

size should be at least 6 m in radius (approximately 113 m^2) to achieve a stable GC estimation. A similar trend was found for the number of trees (sample size) because both the plot size and sample size are related to each other, according to equation 3 (see Figure 3b in **I**). It was observed that the minimum plot size ($s = 6$ m radius) requires an average 15 trees to obtain a stable GC estimation (Figure 3). However, the average number of trees (sample size) could also be dependent on the heterogeneity of the forest, and stands with a greater inequality would require a greater number of trees, as compared to more homogeneous stands.

In regard to the second criterion, which shows the evolution of absolute correlation $|r|$ of the GC estimates with the selected ALS metrics (P25, P50, P99, Skew, StdDew, Cover, CRR in Table 2), irregular fluctuations were observed in the smaller plot sizes ($s < 6$ m radius) (see Figure 4a of **I**), which could possibly be due to the unstable GC estimations in the smaller plots sizes. Once the GC estimation stabilised under the first criterion, the correlation of GC values with the selected ALS metrics produced a convex curve with increasing plot sizes. Thus, it was possible to decide the optimal plot size for the GC estimation based on the greatest absolute correlation $|r|$. The maximum correlation was observed for the plot size with 9–12 m radius, which were considered as the optimal plot size s^* for reliable GC estimation (Table 3).

In the sample size optimisation, the absolute correlation $|r|$ of GC values with the same ALS metrics (P25, P50, P99, Skew, StdDew, Cover, CRR) (second criterion) but with an increasing number of trees (sample size) showed that the absolute correlation between GC and ALS metrics with a smaller number of trees ($n < 15$) was also irregular and should be avoided according to the first criterion, as some of the plots were above the $\overline{GC}_{diff} < 0.05$ limit (Figure 3). However, beyond $n = 15$, the correlation stabilised (see Figure 4b in **I**). The optimal sample size n^* for reliable GC estimation should range from 30–60 trees because both the plot size and sample size are related to each other, according to equation 3 (Table 3).
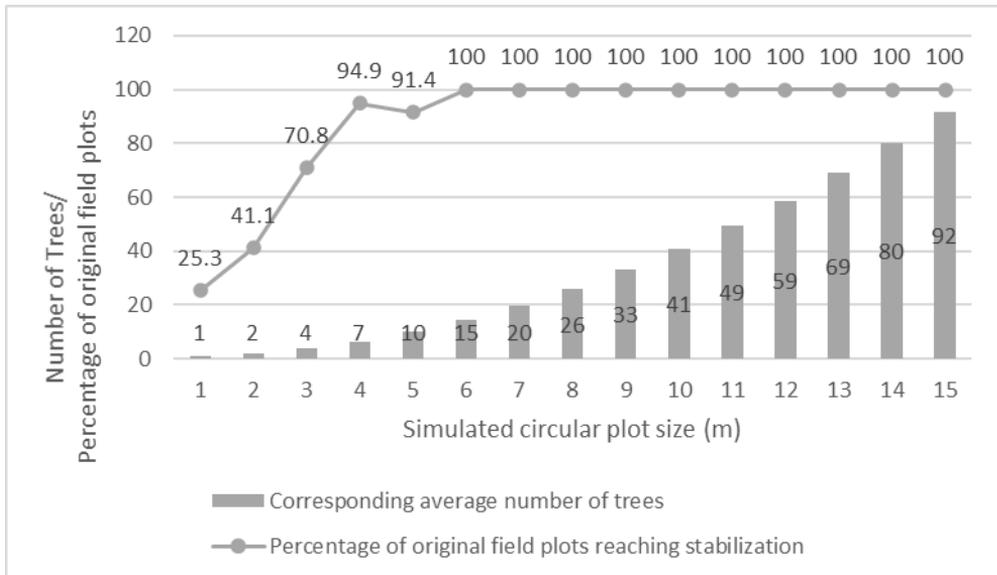
Figure 3. Average number of trees in each simulated circular plot and the proportion of original field plots that fell within the $\overline{GC}_{diff} < 0.05$ limit and reached stabilisation (first criterion).

Table 3. Results of the second criterion showing the maximum absolute correlation of the field GC with the airborne laser scanning (ALS) metrics in the optimal plot sizes and their corresponding number of trees (second criterion).

ALS metric	max$|r|$	s^*	Plot area (m^2)	n^*
Skew	0.58	10	314.16	41
Cover	0.45	12	452.39	59
CRR	0.42	9	254.47	33

$|r|$: absolute correlation; s^*: optimal plot radius (m); n^*:optimal number of trees

3.1.2 Effects of ALS point density on the relationship between GC values and ALS metrics

Once the optimal plot size was determined (in the previous stage), the s*= 9 m radius was selected as the optimal plot size to analyse the effects of the changing ALS point densities. To help in the direct comparison, the same ALS metrics (i.e. P25, P50, P99, Skew, StdDew, Cover, CRR) were also selected in this case. The relationship ($|r|$) between the GC values and the selected ALS metrics with increasing point densities was assessed (see Figure 6 in **I**). No substantial changes in the relationship were found, which suggests that the relationship between the GC values and the ALS metrics is not affected by point density d. However, point density $d < 3$ points m^2 showed a decreasing trend in the relationship, which should be avoided.

3.2 Cross-bioregional FST assessment (II)

3.2.1 Determination of FST from field data

In the cross-bioregional FST assessment, five optimum clusters were initially selected for the hierarchal clustering analysis (HCA) because HCA completely merges or splits all individual observations. Then, both the coniferous and deciduous forests were divided into those five optimum clusters (FST), and the threshold values of the four forest attributes – GC, BALM, QMD and N– (explanatory variables) were identified using CART analysis. The explanatory variable at each node maximises the inter-cluster variability, therefore, the order of these explanatory variables shows their importance in determining the different FST, both in coniferous and deciduous forests. The first cluster, which had the lowest intra-group variability in the coniferous forest, was produced by $GC \geq 0.51$, while in the deciduous forest, $BALM \leq 0.87$ produced the first cluster (Table 4). This was an iterative procedure that eventually resulted in five homogeneous clusters (FST) with the lowest intra-group variability in both forests.

The threshold values of all explanatory variables determined at each node were used to identify the different FST (Table 4; see Figure 2 in **II** for a graphical representation of the classification tree and the diameter distributions of each FST). In the coniferous forest, greater GC values (≥ 0.51) at the first node separated the peaked reversed J-type FST (#1.2) from the single storey and multi-layered FST. The next node was based on stand density ($N \geq 1339$ stems ha^{-1}), which separated out the young, dense single storey (#2.1).

Table 4. Exact threshold values that separated forest structural types (FST) in the coniferous and deciduous forests. See Figure 2 in **II** for a graphical representation of the classification tree and the diameter distribution of the FST.

Split/Node	Coniferous Forest		Deciduous Forest	
	Condition	FST	Condition	FST
1	$GC \geq 0.51$	peaked reversed J (#1.2)	$BALM > 0.87$	peaked reversed J FST (#1.2)
2	$N \geq 1339$	young dense single storey (#2.1)	$N > 1998$	young dense single storey (#2.1)
3	$QMD > 36.60$	very mature single storey (#2.3)	$GC < 0.55$	mature sparse multi-layered (#3.2)
4	$BALM \geq 0.67$	mature sparse multi-layered (#3.2)	$QMD > 24.50$	young dense multi-layered (#3.1)
	$BALM < 0.67$	mature single storey (#2.2)	$QMD < 24.50$	young dense reversed J (#1.1)

GC: Gini coefficient: N: stand density (stems ha^{-1}); QMD: quadratic mean diameter (cm); $BALM$: basal area larger than mean.

Thereafter, a high QMD ($> 36.60\ cm$) separated out the very mature single storey (#2.3). The last node was based on $BALM$, which separated the mature sparse multi-layered (#3.2) from the mature single storey (#2.2) (by $BALM > 0.67$). In the deciduous forest, the first node was based on $BALM$, which separated out the peaked reversed J-type FST (#1.2) by $BALM > 0.87$. The next two nodes were based on N and GC and they separated the young, dense single storey (#2.1) and the mature, sparse multi-layered (#3.2) by $N > 1998$ stems ha^{-1} and $GC < 0.55$, respectively. The final node was based on QMD and the young, dense reversed J-type forest structure (#1.1) was separated from the young, dense multi-layered (#3.1) by $QMD < 24.50$ cm. The characteristics that were useful to denominate the various FST in this study could be valuable in other relevant studies, and are summarised in Table 5.

3.2.2 Forest structural types prediction from ALS data

The observed and predicted FST in the coniferous forests (Finland: Boreal, and Spain: Mediterranean) are shown in Table 6 wherein the peaked reversed J-type FST (#1.2) was accurately predicted. A slight underprediction was observed in the young, dense single storey (#2.1) and mature single storey (#2.2) FST, while the very mature single storey (#2.3) and the mature, sparse multi-layered (#3.2) were slightly overpredicted. The overall accuracy in the coniferous forest was $OC = 0.73$ and $k = 0.64$ (Table 6a). In the deciduous forest (Table 6b), reversed J-type FST, such as the young, dense reversed J-type (#1.1) and the peaked reversed J-type (#1.2), were accurately predicted, while the remaining three FST (#2.1: young, dense, single storey; #3.1: young, dense, multi-layered; #3.2: mature sparse multi-

layered) were slightly underpredicted. However, the overall accuracy in the deciduous forests was better than for the coniferous forest ($OC = 0.87$ and $k = 0.81$).

3.3 Aboveground biomass estimation from FST detected directly from ALS data (III)

3.3.1 Forest structural types detection from ALS data

Figure 4 shows the forest development classes that represent the various FST on either side of $L_{cv} = 0.33$ and $L_{skew} = 0$, which are the boundary lines that separate even- and uneven-sized FST, and open and closed canopy FST, respectively. Seedling, young thinning, advanced thinning and mature development classes usually consist of even-sized diameter distributions and the majority were correctly assigned below $L_{cv} = 0.33$, whereas the seed tree and multi-storied FST, which consist of uneven-sized diameter distributions, were above $L_{cv} = 0.33$. However, in the sapling development class, where the diameter distribution is usually even, the majority was classified as uneven-sized ($L_{cv} > 0.33$), and the shelterwood development class, which comprises uneven-sized diameter distribution, was assigned as even-sized ($L_{cv} < 0.33$). In contrast, seedling, sapling and seed tree development classes, which have smaller aboveground biomass, open canopies and low ALS returns, were correctly assigned above $L_{skew} = 0$. Moreover, the young thinning, advanced thinning, mature and shelterwood development classes, which consist of large aboveground biomass, closed canopies and high ALS returns were assigned below $L_{skew} = 0$.

Table 5. Denomination of forest structural types (FST) based on their diameter distribution in each classification tree provided in Figure 2 in **II**.

FST#	Denomination	Characteristics
#1.1	Young dense reversed J	High GC, medium/high $BALM$, high N, and low QMD
#1.2	Mature sparse reversed J (Peaked reversed J)	High GC, high $BALM$, medium/low N and high QMD
#2.1	Young dense single storey	Medium GC, medium $BALM$, high N and low QMD
#2.2	Mature single storey	Low GC, low $BALM$, medium N and medium QMD
#2.3	Very mature single storey	Low GC, medium/low $BALM$, low N and high QMD
#3.1	Young dense multi-layered	Medium GC, medium $BALM$, low N and high QMD
#3.2	Mature sparse multi-layered	Medium GC, medium $BALM$, medium N and medium QMD

GC: Gini coefficient; $BALM$: Basal area larger than the mean; QMD: quadratic mean diameter (cm); N: stand density (stems ha^{-1})

Table 6. Contingency matrix showing the observed and predicted forest structural types (FST) in (a) coniferous, and (b) deciduous forests using the nearest neighbour imputation method.

(a)

	Observed FST					
Predicted FST	#1.2	#2.1	#2.2	#2.3	#3.2	User Accuracy
#1.2	26	7	0	0	1	0.76
#2.1	4	11	0	0	1	0.69
#2.2	0	0	3	0	3	0.50
#2.3	4	0	0	19	0	0.83
#3.2	0	4	8	0	25	0.68
Producer Accuracy	0.76	0.50	0.27	1.00	0.83	

(b)

	Observed FST					
Predicted FST	#1.1	#1.2	#2.1	#3.1	#3.2	User Accuracy
#1.1	40	0	2	1	0	0.93
#1.2	0	41	0	2	2	0.91
#2.1	0	0	5	1	0	0.83
#3.1	1	1	0	10	2	0.71
#3.2	0	1	0	2	5	0.62
Producer Accuracy	0.98	0.95	0.71	0.62	0.56	

#1.1: young, dense reversed J; #1.2: mature, sparse reversed J (peaked reversed J); #2.1: young, dense single storey; #2.2: mature, single storey; #2.3: very mature, single storey; #3.1: young, dense multi-layered; #3.2: mature, sparse multi-layered.

3.3.2 Aboveground biomass prediction in the detected FST

The calculated AGB and other forest attributes in the full dataset (without pre-stratification), and each FST separated by the $L_{cv} = 0.33$ and $L_{skew} = 0$ boundary lines, are shown in Table 7. The average AGB in the full dataset was 89.5 Mg ha^{-1}. When the full dataset was classified, AGB in the even- and uneven-sized, and open and closed canopy FST were 105.2 Mg ha^{-1} and 74.2 Mg ha^{-1}, and 51.5 Mg ha^{-1} and 129.4 Mg ha^{-1}, respectively. The average GC calculated from the basal areas in the even-sized (0.41) and uneven-sized FST (0.67) also affirmed the correct classification by the $L_{cv} = 0.33$ boundary line. The average QMD in the even-sized FST was greater (11.7 cm) than the uneven-sized FST (7.5 cm), which may be due to the inclusion of the mature development class in the even-sized FST. The average QMD in the closed canopy was similarly greater (13.5 cm) than the open canopy FST (5.7 cm) due to the inclusion of the mature development class in the closed canopy FST. The

average number of trees (stem density) in all FST in this study was substantial due to the inclusion of the seedling class.

Once the AGB was calculated from the field data, the next step was to predict AGB from the ALS data in the full dataset (without pre-stratification) using a general model, and in each of the pre-stratified ALS-detected FST using specific models. Before modelling, various ALS metrics were selected in each model using the "best subset" method. Overall, the metrics consisted of a measure of the central tendencies and dispersions, height percentiles and cover metrics of ALS echo heights, but there were major differences in the selected ALS metrics in even-sized vs uneven-sized FST, and open vs closed canopy FST. For example, in even-sized and closed canopy FST, the average percentile (50th percentile) and cover metrics were important compared to the uneven- and open canopies where high percentiles (70th and 99th percentiles) and variance were important (see Table 3 in **III**).
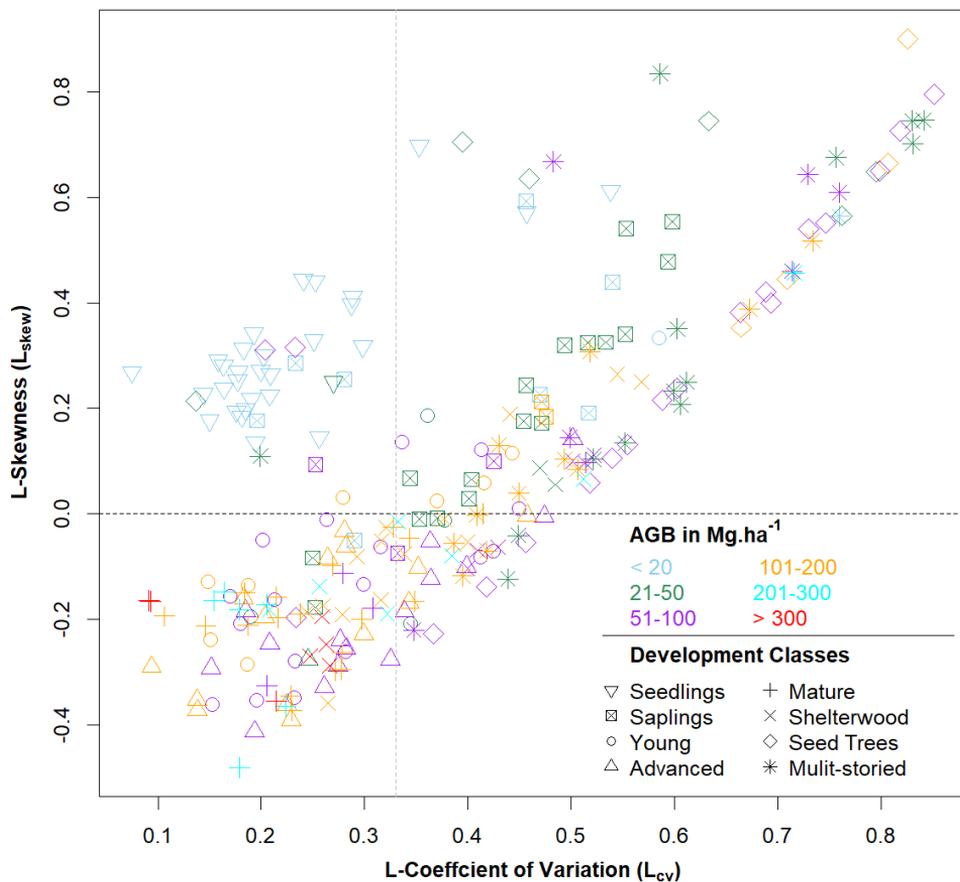
Figure 4. Separation of the development classes that represent the various forest structural types (FST) in a boreal forest by the boundary lines $L_{cv} = 0.33$ and $L_{skew} = 0$.

Table 7. Aboveground biomass and other forest attribute estimations in the full dataset (without classification), and forest structural types (FST) detected from airborne scanning (ALS) data.

		Field plots	AGB (Mg ha^{-1})	QMD (cm)	GC	Trees ha^{-1}
Full dataset	Min		2.5	0.1	0	117
	Mean		89.5	9.6	0.54	20091
	Max	244	410.5	38.1	0.99	182522
	SD		74.1	8.7	0.35	24587
Even ($L_{cv} < 0.33$)	Min		2.5	0.1	0	117
	Mean		105.2	11.7	0.41	17564
	Max	120	410.5	38.1	0.99	182522
	SD		89.6	9.7	0.32	27215
Uneven ($L_{cv} > 0.33$)	Min		6.9	0.1	0	157
	Mean		74.2	7.5	0.67	22538
	Max	124	271.4	38.0	0.99	110774
	SD		50.8	7.3	0.33	21575
Open canopy ($L_{skew} > 0$)	Min		2.5	0.1	0	117
	Mean		51.5	5.7	0.51	28624
	Max	125	271.4	38.1	0.99	182522
	SD		46.6	8.2	0.40	26765
Closed canopy ($L_{skew} < 0$)	Min		12.4	2.0	0.10	314
	Mean		129.4	13.5	0.59	11129
	Max	119	410.5	32.4	0.99	108805
	SD		76.7	7.8	0.27	18274

AGB: aboveground biomass; QMD: quadratic mean diameter; GC: Gini coefficient of basal area; L_{cv}: L-coefficient of variation of LiDAR heights; L_{skew}: L-skewness of LiDAR heights SD: standard deviation,

Accuracy assessments of the general model (Figure 5) shows that the RMSD value obtained from the observed vs predicted AGB was 37.4 Mg ha^{-1}, and similar RMSD values were obtained in the even- and uneven-sized FST. However, the RMSD values in the open and closed canopy FST were slightly reduced to 35.6 and 35.3 Mg ha^{-1}, respectively. The MD values in all FST were greater than in the full dataset (-3.55 Mg ha^{-1}), except in the even-sized FST (-2.09 Mg ha^{-1}). The models were not overfitted, as SSR was < 1.10, which is deemed an acceptable limit.

In the specific models developed for the even- and uneven-sized FST (Figure 6a), the RMSD values for the full dataset, even-sized and uneven-sized FST were reduced to 34.9, 34.6 and 35.3 Mg ha^{-1}, respectively, compared to 37.4, 37.1 and 37.6 Mg ha^{-1}, respectively, in the general model (Figure 5). The MD values also improved and the SSR value was < 1.10, which showed that the FST specific models developed for even- and uneven-sized FST were not overfitted.
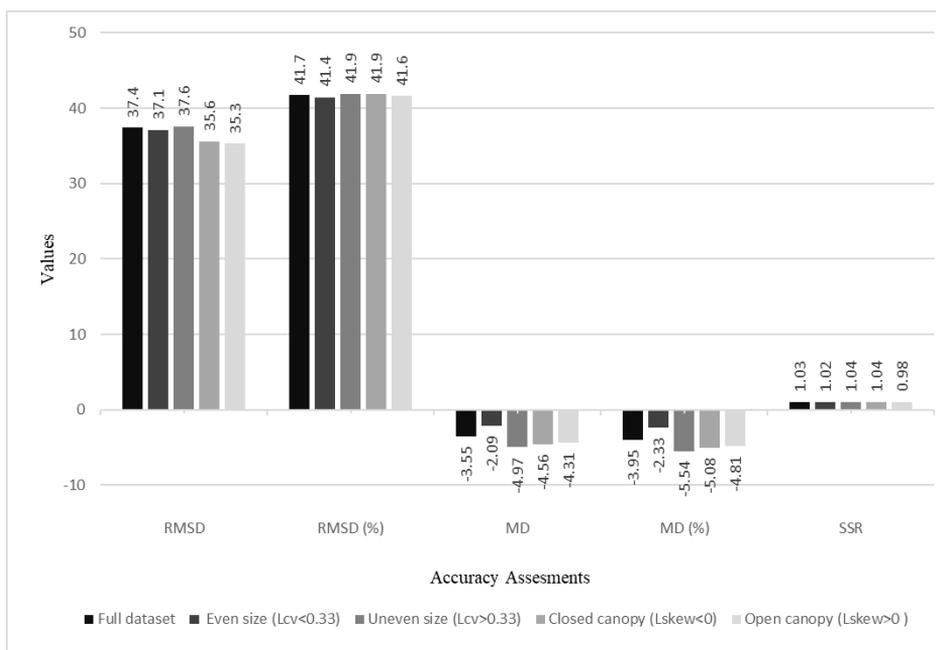
Figure 5. Accuracy assessment of the observed and predicted aboveground biomass of the general model. L_{cv}: L-coefficient of variation of LiDAR heights; L_{skew}: L-skewness of LiDAR heights; RMSD: relative mean square difference; MD: mean difference; SSR: sum of square ratio.

Further improvements were observed when the dataset was divided into closed and open canopy FST (Figure 6b). In this case, the RMSD values for the full dataset, closed canopy and open canopy were much better (32.2, 33.5 and 32.9 Mg ha^{-1}, respectively) than both general models (Figure 5) and the FST specific model developed for the even- and uneven-sized FST (Figure 6a). The MD value also improved and the SSR value was < 1.10, which was deemed an acceptable level of divergence.

4 DISCUSSION

4.1 Improving the estimation of the Gini coefficient of tree size inequality (I)

The Gini coefficient of tree size inequality has been used for structural heterogeneity assessment of forests and is considered as one of the best indicators (Lexerød and Eid 2006; Lei et al. 2009; Cordonnier and Kunstler 2015). While characterising the various FST using GC, Valbuena et al. (2013) found that the size of the plots affected the GC estimates. Matos (2014) later employed various stabilisation criteria for the GC of tree size inequality but the study was based on field information only and did not provide satisfactory results. Therefore, this study has endeavoured to tackle the question of plot size effects on GC estimates from the perspective of its practical estimation using ALS data. For reliable information, an optimal plot size is always needed for any forest attribute, including GC, because

(a)
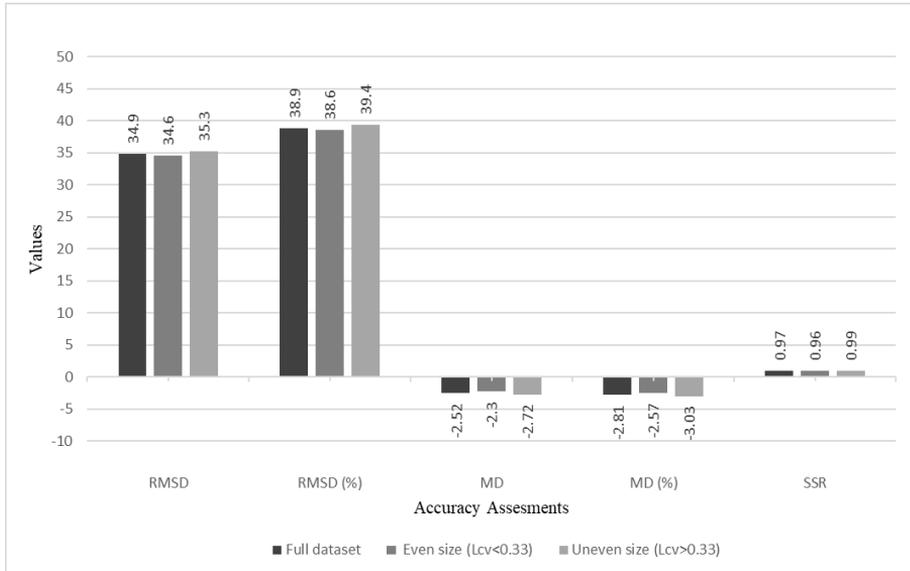
(b)
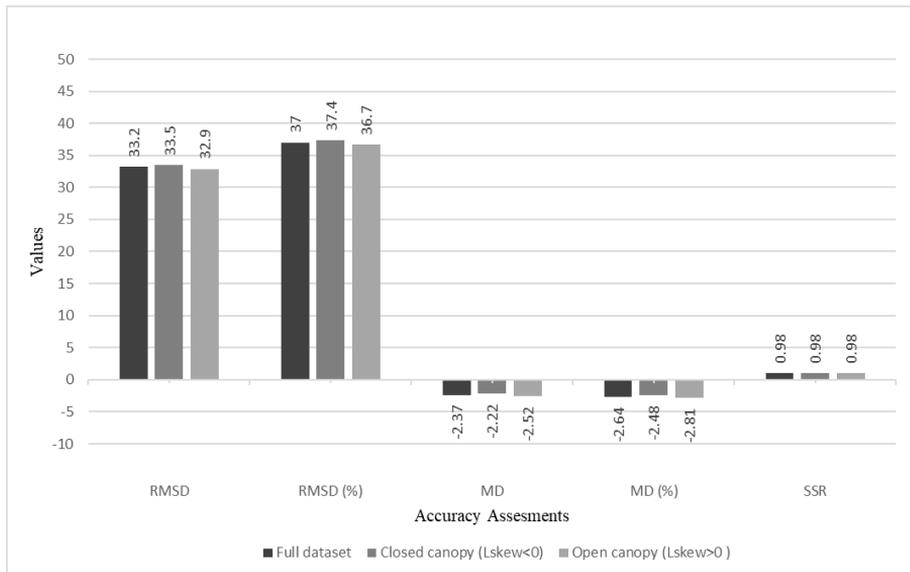
Figure 6. Accuracy assessment of the observed and predicted aboveground biomass of specific models developed for (a) even- and uneven-sized forest structural types (FST), and (b) closed and open canopy FST. L_{skew}: L-skewness of LiDAR echo heights; L_{cv}: L-coefficient of variation of LiDAR echo heights; RMSD: relative mean square difference; MD: mean difference; SSR: sum of square ratio.

inappropriate plot sizes provide unreliable results that may lead to inaccurate forest management decisions (Eid 2000). In this study, two criteria were imposed to optimise the plot size and sample size for reliable GC estimation; (1) stabilisation of the GC estimation from field information, and (2) maximising the absolution correlation of GC values with the ALS metrics. The GC value stabilises when the size of the plot or the stem density increases (Figure 3). In the smaller plot sizes ($s < 6$ m radius) or sample sizes ($n < 15$ tree), GC estimation is unstable (see Figure 3 in **I**). This is because the smaller number of trees within the smaller radius are not representative of the total population. The stabilisation of the GC estimates started at plot size $s = 6$ m or at sample size $n = 15$ trees, which should be the minimum plot size or sample size for stable GC estimation when all the original field plots decreased below the $\overline{GC}_{diff} < 0.05$ limit and fulfilled the first criterion (Figure 3). However, the minimum plot size s also depends on the stand density N of the original field plot, which can be adjusted according to equation 3. This conclusion brings generality in the method used in this study and can be extended to other forest areas.

Maximising the relationship between the GC calculated from field information and ALS metrics was assumed to be a suitable criterion to optimise the plot size and to obtain a reliable ALS-assisted GC estimation. This assumption was correct because the absolute correlation $|r|$ of GC values with the ALS metrics, in particular the most correlated metrics (i.e. Skew and CRR), followed a convex curve (see Figure 4a in **I**) where it was possible to select the optimal plot size by searching for the maximum correlation (Table 3). Similarly, the less correlated metrics (i.e. StdDev, cover, lower height percentiles) showed that once GC is stabilised under the first criterion, the correlation remains unchanged. Therefore, it was important to impose the first criterion and select the minimum plot size or sample size that could represent the total population and produce a stable GC estimation (Motz et al. 2010). Thus, plot sizes smaller than $s < 5\ m$ were dismissed under the first criterion (Figure 3 and Figure 4 in **I**). When integrating the first and second criteria, the optimal plot size and sample size in this study was found to be 9–12 m radius (250–450 m^2 area) and 30–60 trees, respectively (Table 3). However, the optimal plot size also depends on the spatial pattern of the trees, species diversity and the density of the stands (Häbel et al. 2019). The optimal plot size obtained in this study is similar to the plot sizes adopted in current forest inventory practices (Tomppo et al. 2017; Maltamo et al. 2019). Further research should be carried out to evaluate how different combinations of ALS metrics could be used for plot size optimisation and accurate GC predictions.

In national forest inventories (NFI), the sampling design is optimised to obtain highly accurate results within a given fixed budget or to obtain results with a desired precision at the lowest cost (Päivinen 1987). The optimal sampling design is an important goal of NFI (Mandallaz 2007), however, the usability of NFI plots is limited in management inventories due to heterogenous plot types and sizes, and to the small sampling intensity. If the plot size is small, it may have negative effects, for example, on the accuracy of stem diameter distributions estimated from ALS (Maltamo et al. 2019). On the other hand, if the plot size is larger than the required size, it will increase the cost, time and efforts of forest inventories (Chytrý and Otýpková 2003), and different stand conditions may also aggregate (Coomes and Allen 2007b). The optimal plot size also depends strongly on the purpose of the forest inventory, the attribute of interest and other various factors, such as time used to construct and arrange the plots, tree measurements and the disparity of each of the variables between plots (Henttonen and Kangas 2015). For example, variables such as land use change can be determined at a very small plot scale but larger plots are required for biomass and volume estimations. Therefore, prioritisation of the variable of interest is needed before plot size

optimisation. Häbel et al. (2019) provided similar arguments and included additional factors, such as stand density, diameter distribution, basal areas and spatial structures in their analyses. Various studies have been carried out to determine the optimal plot size for different forest attributes. For example, Lombardi et al. (2015) conducted a study to select an optimal plot size s* for the indicators that describe heterogeneity in old growth stands, such as number of large and dead trees, and the total volume of living and dead trees. They found that 13–15 m radius was the optimal plot size, although this could be due to the lower stand density N in their study (according to equation 3). Similarly, Tomppo et al. (2017) studied ALS-assisted forest resource estimation (volume and basal area) and recommended concentric circular plots with a maximum 9 m outer radius and 5.64 m inner radius to reduce the cost of forest inventories, but they also appraised that a fixed radius plot with a 9 m maximum radius is still feasible for ALS-assisted inventories. Maltamo et al. (2019) also estimated and compared stand level stem diameter distributions using various plot sizes (200, 400, 900 and 1600 m^2) and concluded that 200–400 m^2 is the optimal plot size for stand level estimation. Increasing the plot size also shows an averaging effect on some of the forest attributes, such as biomass and volume (Gobakken and Næsset 2008; Ruiz et al. 2014), however, in a forest structural assessment, the averaging effect is not applicable (Coomes and Allen 2007b). In fact, variables such as species richness increase as the size of the plot increases (Humphrey et al. 2000; Otypková and Chytry 2006).

The relationship of GC values with ALS metrics is not affected by ALS point density, unless the point density d is < 3 points m^2 (see Figure 6 in **I**). Similar results have been observed when the effects of point density on biomass and volume estimations have been evaluated (Maltamo et al. 2006; Ruiz et al. 2014), which demonstrates that varying ALS point density has no real effect in practical applications. Based on these results, all national programs that render ALS data with a point density < 3 points m^2 are unsuitable for the structural heterogeneity assessment of forests. Indeed, Valbuena et al. (2017b) found that the understorey development class was critically omitted due to the lower point density. Therefore, to ensure that ALS data is suitable for forest structural assessments in the future, the point density must be increased to at least 3 points m^2 in national ALS survey programmes.

4.2 Simplifying the cross-bioregional assessment of FST (II)

Forest structural types assessment is important for wildlife habitat management (Hagar et al. 2020), biodiversity (Lelli et al. 2019), biomass and carbon storage (Clark and Clark 2000; Marvin et al. 2014), natural dynamics in forests, such as thinning and disturbances (Coomes and Allen 2007a) or if these dynamics are artificially modified (Valbuena et al. 2016a). In this study, various FST were identified in a simple two-tier approach by utilising four forest attributes – GC, $BALM$, QMD and N– obtained from the Boreal, Mediterranean and Atlantic biogeographical regions, which made it feasible for a regional assessment of the FST. In the upper tier, GC and $BALM$ were used to identify the reversed J-type, single storey and multi-layer FST. These two attributes have been suggested by Valbuena (2015) as bivariate descriptors of FST. In the lower tier, traditional attributes, such as QMD and N, were used to identify forest structures according to the development stage and stand density. The QMD separates young and mature FSTs while N is used to separate sparse and dense FST, respectively. For this purpose, HCA (an unsupervised method) was used and all individual observations were assigned to five clusters/groups (Bien and Tibshirani 2011). Then, CART

analysis was used to identify the exact threshold values of the four forest attributes and their importance in determining the FST (Breiman et al. 1984). Lastly, the kNN method (Venables and Ripley 2001) was used to predict the field information-based FST from the ALS data.

The threshold values for the four forest attributes were obtained from HCA and CART analyses, and was used for the forest structural classification (Table 4; see Figure 2 in **II**). The first nodes in the coniferous and deciduous forests were based on the GC and $BALM$, respectively, which shows that these two attributes were the most important (Gove, 2004; Lexerød and Eid, 2006) in the forest structural classification, as proposed by Valbuena (2015). The GC values obtained in both coniferous ($GC = 0.51$) and deciduous ($GC = 0.57$) forests were very close to the theoretical GC values ($GC = 0.50$) used by Valbuena et al. (2012) as a boundary line to separate even- and uneven-sized FST. Thus, multi-layered FST are observed around these values, and GC values > 0.50 denote a reversed J-type FST, while $GC < 0.50$ identify an even-sized distribution, such as single storey FST (Duduman 2011; Valbuena et al. 2013). Simpson et al. (2017) obtained a similar forest structural classification in deciduous forests, however, their classification was based on vertical gap probability. The empirical values of $BALM$, which have not been identified in the literature (Valbuena 2015) were obtained and used in the forest structural classification in this study. High $BALM$ values separate the mature, sparse reversed J-type/peaked reversed J-type FST (#1.2), while lower $BALM$ values identify closed canopies such as those present in mature and single storey FST (#2.2). Traditional forest attributes, such as QMD and N, were useful to identify the mature/young and dense/sparse FST in the lower tier (Dodson et al. 2012), e.g. the separation of the very mature single storey (#2.3: Figure 7) in the Valsaín forest (Spain: Mediterranean bioregion), which contains mature trees of approximately 100 years old, by $QMD > 36.6\ cm$ (Valbuena et al. 2013), or the separation of the young, dense single storey (#2.1) by $N > 1339$ and 1998 trees ha^{-1} in the coniferous and deciduous forests, respectively.

For regional assessment and mapping, ALS is a useful remote sensing technique for FST assessment due to its large-scale coverage and ability to produce 3D canopy information (Zimble et al. 2003; Latifi 2012; Asner and Mascaro 2014). Therefore, it is important to predict field information-based forest structures from ALS data. The observed and predicted FST in the coniferous and deciduous forests, respectively, are shown in Table 6. In general, the estimations were unbiased, and the errors mostly occurred between structurally similar FST, for example, the mature single storey (#2.2) was misclassified as mature sparse multi-layered (#3.2). As these FST are typically difficult to discriminate by forest attributes, the poor prediction from ALS was not surprising. However, such misclassification of structurally similar FST has less impacts in practical forest management. The overall accuracy OC and kappa coefficient k in the deciduous forest ($OC = 0.87$ and $k = 0.81$) was greater than the coniferous forest ($OC = 0.73$ and $k = 0.64$), which may be due to differences in the ALS data. These results are particularly important for organisations involved in forest monitoring and conservation, biodiversity and landscape planning, and can assist in the development of ALS-based essential biodiversity variables (Pereira et al. 2013), regional forest structural maps (Adhikari et al. 2020), effective forest management policies and decisions towards sustainable development goals (Vihervaara et al. 2017).

Remote sensing data other than ALS have also been used in the local and global mapping of various attributes that include forest structures. For example, Pippuri et al. (2016) employed a combination of ALS and Landsat images, and various forest land attributes were classified based on national (forest land, agricultural land and built-up areas) and FAO (Food and Agriculture Organization) land use land cover classifications (forest and non-forest areas), peatland types (open, pine and spruce) and site types (poor, medium and rich). They

concluded that the prediction of these land attributes could be incorporated into forest management inventories in Finland. Adhikari et al. (2020) also used a combination of ALS and Landsat time series images, and found that the spectro-temporal metrics of Landsat images that are sensitive to tree species and forest density improved the model accuracies. Various researchers have also used Interferometric Synthetic aperture radar (InSAR) data to study vegetation properties (Ranson et al. 1995; Harrell et al. 1997; Castel et al. 2001) as the InSAR is an active remote sensing platform and has the advantage of strong penetration capability through clouds, high temporal and spatial resolution, and vast geographical coverage (Næsset et al. 2011). Indeed, Castel et al. (2002) found a significant relationship between InSAR data and stand biomass and density. Thus, InSAR application is a useful alternative to ALS for regional biomass mapping, particularly in countries where ALS data are unavailable due to the high cost (Le Toan et al. 2011). The use of unmanned aerial vehicles (UAV) equipped with digital cameras (multi- or hyperspectral) or with ALS sensors has recently increased in the forest sector (Colomina and Molina 2014; Sankey et al. 2017), which could also be useful in forest structural heterogeneity assessments. Ni et al. (2015) reported exceptional levels of accuracy when calculating tree heights, and Alonzo et al. (2018) achieved similar accuracies when modelling tree density, basal area, aboveground biomass, and species composition in boreal conditions. While UAV-based forest inventories are limited to small areas due to the high cost, they could be particularly effective, for example, in predicting the diameter distribution or in the determination of dbh (Puliti et al. 2020).

4.3 Aboveground biomass predictions in ALS-based direct FST (III)

Pre-stratification of data has been used to examine whether it could improve estimates of forest characteristics. In Norway, models for forest attributes are constructed for pre-classified strata. For the pre-classification of strata, visual interpretation of stereo-images and age classes, such as young stands, mature stands or mature stands dominated by spruce or pine are used. These strata are further divided on the basis of site quality. Errors always occur in the pre-classification of images, although various stand characteristics can be determined by this two-stage procedure with better accuracy than with conventional methods (Næsset 2002). Indeed, Maltamo et al. (2015) pre-classified their data based on tree species and stand development classes. Pre-stratification based on aerial images did not improve the model accuracies, as the optical imagery was not able to detect the small variations at the plot level, although the ALS-based pre-stratification (into mature and young stands) resulted in a slight improvement. Similarly, ALS-based pre-stratification can also reduce the sampling efforts by up to 41 % (de Almeida Papa et al. 2020).

In this study, a two-stage procedure was adopted. The ALS data was used for pre-stratification instead of the optical images, as ALS is one of the key technologies that can provide better results (than optical images), produce detailed canopy information, describe the biophysical stand properties, monitor forest changes over large geographical areas and, most importantly, predict various forest attributes with excellent levels of accuracy (Maltamo et al. 2004; Næsset and Gobakken 2008; Valbuena et al., 2013). In the pre-stratification stage, various FST are initially detected directly from ALS data (Valbuena et al. 2017b) because structural heterogeneity is an important morphological trait of ecosystems (Fahey et al. 2019) and it affects various ecological processes, such as carbon sequestration, nutrient cycling and species interactions (Brokaw and Lent 1999; McElhinny et al. 2005). Next, the aboveground

biomass predicted in each FST was compared with the AGB prediction in the full dataset without pre-stratification.

Two L-moment ratios of the ALS echo heights (L_{cv} and L_{skew}) were used for the pre-stratification as they characterise tree size inequality (Valbuena et al. 2013) and dominance (Lefsky et al. 2002), and are useful to separate even- and uneven-sized, and open and closed canopy FST, respectively. In the previous study by Valbuena et al. (2017b), $L_{cv} = 0.50$ was used to represent maximum entropy, while $L_{cv} = 0.33$ was used for the first time in my study. Mathematical proofs were provided to demonstrate that if a two-dimensional attribute, such as basal area, is used to calculate an indicator in a vicinity (e.g. GC), then maximum entropy is reached at 0.50. However, if one-dimensional attributes, such as tree diameter or ALS echo heights, are used, maximum entropy is reached at 0.33 in the same area (see Appendix A in **III**). The diameter and basal distribution in the even- and uneven-sized FST, separated by $L_{cv}(GC$ of LiDAR$) = 0.33$ and GC of basal area $= 0.50$, also illustrates the empirical equivalence of the ALS and field approaches (see Figure 2 in **III**). Figure 4 shows how the different forest structures can be separated by $L_{cv} = 0.33$ and $L_{skew} = 0$. The majority of the seedlings, young and mature development classes are below $L_{cv} = 0.33$ because they backscatter high ALS returns due to even-sized diameter distribution. As a consequence, they exhibit a small variance in ALS echo heights. On the other hand, seedlings or multi-layered development classes, which have uneven diameter distributions, are distributed above $L_{cv} = 0.33$, they have a wide variance in ALS echo heights due to the low ALS returns. The sapling development class, which is usually even-sized and where the GC of basal is below 0.50 (Valbuena et al. 2013), was separated out as uneven-sized ($L_{cv} > 0.33$). Therefore, L_{skew} is an important additional metric because the L_{cv} and GC values of basal are only similar if the greater L_{cv} values are due to canopy gaps (Stark et al. 2012). Thus, by looking at the L_{skew} values, the saplings are separated as open canopies ($L_{skew} > 0$). In addition, the shelterwood development class was not properly separated, which could be due to the omission of the understorey vegetation and the lower point density of the ALS data, which could be improved by increasing the ALS point density (Ruiz et al. 2014). The seedlings, saplings, seed tree, shelterwood and multi-storey development classes all show values above $L_{skew} > 0$ (open canopy FST) as only a small portion of the ALS returns are due to sparse vegetation. The young thinning, advanced thinning and mature development classes exhibit less gaps and a greater proportion of ALS returns as L_{skew} is below 0 (closed canopy FST).

In order to compare the AGB predicted from the full dataset without pre-stratification with the AGB predicted in the pre-classified FST (even- and uneven-sized, and open and closed canopy FST), the first step was the *a priori* selection of ALS predictors, as suggested by Bouvier et al. (2015). From various alternatives, such as best subset, stepwise, and most similar neighbour (MSN) selection methods (Valbuena et al. 2017a), the best subset was selected, as it is the most frequent method for variable selection (Hudak et al. 2006). Six ALS metrics were selected from more than 100 metrics from the full dataset, as well as from each FST (see Table 3 in **III**) because it is important to reduce the number of meaningful ALS metrics in remote sensing to avoid extensive model selection procedures (Hudak et al. 2006; Bouvier et al. 2015; Knapp et al. 2020). In the even-sized ($L_{cv} < 0.33$) and closed canopy ($L_{skew} < 0$) FST, the average percentile (P50), which represents the average tree height and cover metrics that are related to stand density, were important compared to the uneven-sized ($L_{cv} > 0.33$) and open canopy ($L_{skew} > 0$) FST where the high percentiles (P70, P99) that represent the dominant trees and variance were selected. These differences in ALS metrics were very relevant in AGB modeling and in the changes in subsequent AGB predictions.

40

The AGB predictions in the general model and in the FST specific models were compared using the widely used RMSD and MD (Van Aardt et al. 2008; Straub et al. 2013; Räty et al. 2018) with the inclusion of an additional restriction, such as SSR, to avoid overfitting of the models (Tedeschi 2006; Mauro et al. 2016). From the perspective of RMSD and MD, the results might look reliable but the models could be overfitted (Valbuena et al. 2017b). Therefore, the additional restriction (i.e. SSR) employed here was important for the most reliable results. As assumed, the AGB predictions in the FST specific models (Figure 6) were improved in comparison to the general model (Figure 5), which could be due to differences in the ALS metrics, as they contribute differently to the improvement in AGB predictions (Knapp et al. 2020). The RMSD and MD values for the even- and uneven-sized FST improved to 34.6 and 35.3 Mg ha^{-1} in the FST specific models (Figure 6a) as compared to 37.1 and 37.6 9 Mg ha^{-1} in the general model (Figure 5). In the open and closed canopy FST, the RMSD and MD values further improved to 33.5 and 32.9 Mg ha^{-1} (Figure 6b) from 35.6 and 35.3 9 Mg ha^{-1} in the general model (Figure 5). Similar improvements were also observed for the full dataset; the RMSD and MD values improved to 34.9 and -2.52 Mg ha^{-1} (Figure 6a) and 33.2 and -2.37 Mg ha^{-1} (Figure 6b) in the FST specific models, and from 37.4 and -3.55 Mg ha^{-1} in the general model (Figure 5). SSR is used to adjust the increase in unexplained variance or decrease the explained variance to a preferable limit (e.g. 10 %) and should be < 1.10 to avoid overfitting (Lipovetsky 2013). The SSR value in this study was < 1.10 in all models. The improvement in the AGB predictions were still minor, as observed by Maltamo et al. (2015) for specie-specific attributes, such as volume, but the differences in the selection of ALS metrics were critical in this case. Further research could focus on how ground-based metrics of forest structures are related to ALS metrics and whether AGB predictions could be improved if forest structure-related ALS metrics are used in modelling instead of pre-stratification. This study also suggests that the most relevant ALS metrics should be selected as predictor variables for AGB predictions.

4.4 Future Research work

1. In this doctoral research, the relationship between GC of tree size inequality and ALS data was evaluated under boreal conditions, which contained only three tree species (spruce, pine and birch). For the other biogeographical regions, a similar conclusion can be drawn that plot size has a greater influence than stand density and scan density of the ALS data, as the GC compares dbh or basal areas of individual trees growing in the vicinity and is not affected by species diversity. However, this still needs to be investigated in other biogeographical regions with more diverse forests and a wider range of stand development classes.
2. For FST prediction, only ALS data was used in these studies. However, to predict forest structural heterogeneity, the performance of other remote sensing approaches that provide tree height profiles (e.g. InSAR, NASA's Global Ecosystem Dynamics Investigation (GEDI), and a combination of InSAR, GEDI and UAV) should not be overlooked, and could be used as alternatives in developing countries where ALS data is still not available.
3. Various ALS metrics have been used in this doctoral dissertation for the structural heterogeneity assessment. Further evaluation is needed to show how these ALS metrics can be useful in the development of essential biodiversity variables related to ecosystem structure, from the local to the global scale.

4. The four forest attributes – GC, BALM, QMD and N that were used in **II** for the bioregional assessment of forest structure, the estimation of variables, such as QMD and $N,$ and how changes in plot size affect their estimation, have been well studied, and the GC of tree size inequality was evaluated in **I**. However, the estimation of $BALM$ and how various factors, such as the plot size, stand density and scan density affect the $BALM$ estimation, is still unknown and needs to be studied.

5. The potential of the FST assessed in these studies could be investigated to determine whether they could improve the diameter distribution assessment, which is a routine operation in forest inventories that provides timber biomass and volume estimations across different size classes.

5 CONCLUSIONS

The following conclusions can be outlined from each objective of this study.

1. The Gini coefficient (GC) of tree size inequality is one of the best indicators of forest structural heterogeneity. This study examines how the GC values and their relationship with ALS metrics are affected by plot size, stand density and point density of the ALS data. Plot sizes have a greater effect on the relationship between GC and the ALS metrics, as compared to the number of trees (stand density) and ALS point density. The GC estimation is very unstable in the smaller plot sizes because they are unrepresentative of the total area, while the number of trees (sample size) within the smaller plot sizes also under-represents the total population. As the size of plots increases, its effects decrease because larger plot sizes and the greater number of trees (sample size) better represent the total population. For the optimal plot size and sample size, two criteria were implemented; stabilisation of the GC value, and maximising the relationship between GC values and the ALS metrics. In boreal conditions, a minimum 6 m radius plot size (113 m^2) and 15 trees are needed to achieve a stable GC estimation. The correlation between the GC values and the ALS metrics was described by a convex curve and the maximum correlation was found between plot sizes that ranged from 9 to 12 m radius (250–450 m^2 area), which is the optimal plot size for a reliable ALS-assisted GC estimation. However, the plot size can also be adjusted in forests with different stand densities using a basic relationship between stand density and plot size. In regard to the point density effects, it was found that point density had no effect on the relationship between GC values and ALS metrics, unless the point density is < 3 points m^2. Thus, to make ALS data suitable for the structural heterogeneity assessment of forests, nationwide ALS point densities must be increased to at least 3 points m^2. As this study was based on data from boreal ecosystems, the results can only be extended to the boreal region. Similar studies must be conducted in other biogeographical regions with more diverse forests and a wider range of development classes.

2. The study based on four forest structural attributes – GC, BALM, QMD and N– obtained from the –Boreal, Mediterranean and Atlantic biogeographical regions concludes that these four forest structural attributes can be used in a simple two-tier approach for the bioregional FST assessment that covers both coniferous and deciduous forests. In the upper tier, GC and $BALM$ (which identified reversed J-type, single storey and multi-layered FST) are useful, while in the lower tier, the most traditional attributes, QMD and

N, separated the young and mature, and sparse and dense FST. These FST can also be reliably predicted from the ALS data. The methodology developed for the FST assessment in this study can be adopted across other biogeographical regions and it can also be useful to assess the effects of management practices. For countries where ALS data is not available, other remote sensing approaches, such as InSAR, NASA's GEDI and UAV could be used as alternatives to determine the different FST.

3. L-coefficient of variation (L_{cv}) and L-skewness (L_{skew}) are two prominent ALS metrics that can be used as analogous to the GC of tree size inequality to detect various FST directly from ALS data. A threshold value of $L_{cv} = 0.33$ should be used to represent maximum entropy, rather than the 0.50 value used in previous literature, provided the inequality is calculated from tree heights or dbh. Lower L_{cv} (<0.33) and L_{skew} (<0) values separate the even-sized and closed canopy FST, while higher values separate the uneven-sized ($L_{cv} > 0.33$) and open canopy ($L_{skew} > 0$) FST. The aboveground biomass predicted in these FST using their specific models were evaluated and compared with the AGB prediction in the full dataset without pre-stratification using a general model. The aboveground biomass predictions in the FST specific models were minor as compared to the general model but the ALS metrics selected in each model using the best subset procedure were critical. The selection of the relevant ALS metric in any model could play a vital role in AGB predictions in large geographical areas. Therefore, this study suggests that forest areas and the selection of the most relevant ALS metrics should be pre-stratified before AGB predictions. This would further improve our understanding of the structural and AGB dynamics within a large geographical area.

REFERENCES

Adhikari, H., Valbuena, R., Pellikka, P. K., & Heiskanen, J. (2020). Mapping forest structural heterogeneity of tropical montane forest remnants from airborne laser scanning and Landsat time series. Ecological Indicators, 108, 105739. https://doi.org/10.1016/j.ecolind.2019.105739

Almeida, D.R.A., Broadbent, E.N., Zambrano, A.M.A., Wilkinson, B.E., Ferreira, M.E., Chazdon, R., Meli, P., Gorgens, E.B., Silva, C.A., Stark, S.C. and Valbuena, R. (2019). Monitoring the structure of forest restoration plantations with a drone-lidar system. International Journal of Applied Earth Observation and Geoinformation, 79, pp.192-198. https://doi.org/10.1016/j.jag.2019.03.014

Alonzo, M., Andersen, H. E., Morton, D. C., & Cook, B. D. (2018). Quantifying boreal forest structure and composition using UAV structure from motion. Forests, 9(3), 119. https://doi.org/10.3390/f9030119

Angelsen, A., Brown, S., & Loisel, C. (2009). Reducing emissions from deforestation and forest degradation (REDD): an options assessment report. https://agris.fao.org/agris-search/search.do?recordID=GB2013201976 [accessed March 15, 2020]

Andersen, H-E, Strunk J, Temesgen H, Atwood D,Winterberger K (2011). Using multi-level remote sensing and ground data to estimate forest biomass resources in remote regions: a case study in the boreal forests of interior Alaska. Canadian Journal of Remote Sensing 37(6):596–611. https://doi.org/10.5589/m12-003

Asada, Y. (2005). Assessment of the health of Americans: the average health related quality of life and its inequality across individuals and groups. Population Health Metrics, 3: 7. https://doi.org/10.1186/1478-7954-3-7

Asner, G. P., & Mascaro, J. (2014). Mapping tropical forest carbon: Calibrating plot estimates to a simple LiDAR metric. Remote Sensing of Environment, 140, 614-624. https://doi.org/10.1016/j.rse.2013.09.023

Aguirre, O., Hui, G., von Gadow, K., & Jiménez, J. (2003). An analysis of spatial forest structure using neighbourhood-based variables. Forest Ecology and Management, 183(1-3), 137-145. https://doi.org/10.1016/S0378-1127(03)00102-6

Bachofen, H. &, Zingg A. (2001). Effectiveness of structure improvement thinning on stand structure in subalpine Norway spruce (Picea abies (L.) Karst.) stands. Forest Ecology and Management 145, 137–149. http://dx.doi.org/10.1016/s0378-1127(00)00581-8

Barbeito, I., Cañellas, I., and Montes, F. (2009). Evaluating the 43tilizat of vertical structure indices in Scots pine forests. Annals of Forest Science. 66: 710–710. https://doi.org/10.1051/forest/2009056

Bell, S. S., E. D. McCoy, and H. R. Mushinsky. (1991). Habitat Structure: The Physical Arrangement of Objects in Space. Chapman and Hall, London.

Bergeron, Y., Leduc, A., Harvey, B.D., Gauthier, S. (2002). Natural fire regime: a guide for sustainable management of the Canadian boreal forest. Silva Fennica 36 (1), 81–95. https://doi.org/10.14214/sf.553

Berger, W.H., Parker F.L. (1970). Diversity of planktonic Foramenifera in deep sea sediments. Science 168:1345–137. http://dx.doi.org/10.1126/science.168.3937.1345

Bien, J., & Tibshirani, R. (2011). Hierarchical clustering with prototypes via minimax linkage. Journal of the American Statistical Association, 106(495), 1075-1084. https://doi.org/10.1198/jasa.2011.tm10183

Boudreau, J., Nelson, R. F., Margolis, H. A., Beaudoin, A., Guindon, L., & Kimes, D. S. (2008). Regional aboveground forest biomass using airborne and spaceborne LiDAR in Québec. Remote Sensing of Environment, 112(10), 3876-3890. https://doi.org/10.1016/j.rse.2008.06.003

Bouvier, M., Durrieu, S., Fournier, R. A., & Renaud, J. P. (2015). Generalizing predictive models of forest inventory attributes using an area-based approach with airborne LiDAR data. Remote Sensing of Environment, 156, 322-334. https://doi.org/10.1016/j.rse.2014.10.004

Breidenbach, J., Næsset, E. and Gobakken, T. (2012). Improving k-nearest neighbor predictions in forest inventories by combining high and low density airborne laser scanning data. Remote sensing of environment, 117, pp.358-365. https://doi.org/10.1016/j.rse.2011.10.010

Breiman, L., Friedman, J., Stone, C. J. and Olshen, R. A. (1984). Classification and regression trees. CRC press.

Brokaw, N. V. L., and Lent, R. A. (1999). Vertical structure. In Hunter, Malcolm L. (1999) Maintaining biodiversity in forest ecosystems. Cambridge, UK: Cambridge University Press. Pp. 373-399.

Cai, Y.M., Chatelet, D.S., Howlin, R.P., Wang, Z.Z. and Webb, J.S. (2019). A novel application of Gini coefficient for the quantitative measurement of bacterial aggregation. Scientific Reports, 9(1), pp.1-12. https://doi.org/10.1038/s41598-019-55567-z

Castel, T., Beaudoin, A., Stach, N., Stussi, N., Le Toan, T., & Durand, P. (2001). Sensitivity of space-borne SAR data to forest parameters over sloping terrain. Theory and experiment. International journal of remote sensing, 22(12), 2351-2376. https://doi.org/10.1080/01431160121407

Castel, T., Guerra, F., Caraglio, Y., & Houllier, F. (2002). Retrieval biomass of a large Venezuelan pine plantation using JERS-1 SAR data. Analysis of forest structure impact

on radar signature. Remote Sensing of Environment, 79(1), 30-41. https://doi.org/10.1016/S0034-4257(01)00236-X

Chave, J., Andalo, C., Brown, S., Cairns, M.A., Chambers, J.Q., Eamus, D., Fölster, H., Fromard, F., Higuchi, N., Kira, T. and Lescure, J.P. (2005). Tree allometry and improved estimation of carbon stocks and balance in tropical forests. Oecologia, 145(1), 87-99. https://doi.org/10.1007/s00442-005-0100-x

Chytrý, M., and Otýpková, Z. (2003). Plot sizes used for phytosociological sampling of European vegetation. Journal of Vegetation Science. 14: 563–570. https://doi.org/10.1111/j.1654-1103.2003.tb02183.x

Clark, D.B. and Clark, D.A. (2000). Landscape-scale variation in forest structure and biomass in a tropical rain forest. Forest Ecology and Management, 137(1-3), pp.185-198. https://doi.org/10.1016/S0378-1127(99)00327-8

Clifford, H.T., Stephenson W. (1975). An introduction to numerical classification. AcademicPress, London, UK. http://store.elsevier.com/product.jsp?locale=en_EU&isbn=9780323140508.[Cited 3 June 2020].

Colomina, I., and Molina P. (2014). Unmanned aerial systems for photogrammetry and remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing 92: 79–97. https://doi.org/10.1016/j.isprsjprs.2014.02.013

Coomes, D.A., and Allen, R.B. (2007a). Effects of size, competition and altitude on tree growth. Journal of Ecology. 95 (5), 1084–1097. https://doi.org/10.1111/j.1365-2745.2007.01280.x

Coomes, D.A., and Allen, R.B. (2007b). Mortality and tree-size distributions in natural mixed-age forests. Journal of Ecology. 95(1): 27–40. https://doi.org/10.1111/j.1365-2745.2006.01179.x

Cordonnier, T., and Kunstler, G. (2015). The Gini index brings asymmetric competition to light. Perspectives in Plant Ecology, Evolution and Systematics. 17(2): 107–115. https://doi.org/10.1016/j.ppees.2015.01.001

Curtis, R.O. (1982). A simple index of stand density for Douglas-fir. Forest Science, 28(1), pp.92-94. https://doi.org/10.1093/forestscience/28.1.92

Curtis, R.O. and Marshall, D.D. (2000). Why quadratic mean diameter? Western Journal of Applied Forestry, 15(3), pp.137-139. https://doi.org/10.1093/wjaf/15.3.137

de Almeida Papa, D., de Almeida, D.R.A., Silva, C.A., Figueiredo, E.O., Stark, S.C., Valbuena, R., Rodriguez, L.C.E. and d'Oliveira, M.V.N. (2020). Evaluating tropical forest classification and field sampling stratification from lidar to reduce effort and enable landscape monitoring. Forest Ecology and Management, 457, 117634. https://doi.org/10.1016/j.foreco.2019.117634

de Camino, R. (1976). Determinación de la homogeneidad de rodales. Bosque 1: 110–115. https://doi.org/10.4206/bosque.1976.v1n2-05

Diggle, P. J. (2003). Statistical analysis of spatial point patterns. 2nd ed. Edward Arnold, London.

Dodson, E. K., Ares, A., & Puettmann, K. J. (2012). Early responses to thinning treatments designed to accelerate late successional forest structure in young coniferous stands of western Oregon, USA. Canadian Journal of Forest Research, 42(2), 345-355. https://doi.org/10.1139/x11-188

Donato, D.C., Campbell, J.L., Franklin, J.F. (2012). Multiple successional pathways and precocity in forest development: can some forests be born complex? Journal of Vegetation Science. 23 (3), 576–584. https://doi.org/10.1111/j.1654-1103.2011.01362.x

Duduman, G. (2011). A forest management planning tool to create highly diverse uneven-aged stands. Forestry, 84(3), 301-314. https://doi.org/10.1093/forestry/cpr014

Echeverría, C., Newton, A.C., Lara, A., Benayas, J.M.R., Coomes, D.A. (2007). Impacts of forest fragmentation on species composition and forest structure in the temperate landscape of southern Chile. Global Ecology and Biogeography. 16 (4), 426–439. https://doi.org/10.1111/j.1466-8238.2007.00311.x

Eid, T. (2000). Use of uncertain inventory data in forestry scenario models and consequential incorrect harvest decisions. Silva Fennica. 34: 633. https://doi.org/10.14214/sf.633.

European Environmental Agency, (2020). Biogeographical regions. https://www.eea.europa.eu/data-and-maps/data/biogeographical-regions-europe-3 (accessed February 14, 2020).

Fahey, R.T., Atkins, J.W., Gough, C.M., Hardiman, B.S., Nave, L.E., Tallant, J.M., Nadehoffer, K.J., Vogel, C., Scheuermann, C.M., Stuart-Haëntjens, E. and Haber, L.T. (2019). Defining a spectrum of integrative trait-based vegetation canopy structural types. Ecology letters, 22(12), pp.2049-2059. https://doi.org/10.1111/ele.13388

Fassnacht, F. E., Mangold, D., Schäfer, J., Immitzer, M., Kattenborn, T., Koch, B., & Latifi, H. (2017). Estimating stand density, biomass and tree species from very high resolution stereo-imagery–towards an all-in-one sensor for forestry applications?. Forestry: An International Journal of Forest Research, 90(5), 613-631. https://doi.org/10.1093/forestry/cpx014

Foody, G. M., Cutler, M. E., McMorrow, J., Pelz, D., Tangki, H., Boyd, D. S., & Douglas, I. A. N. (2001). Mapping the biomass of Bornean tropical rain forest from remotely sensed data. Global Ecology and Biogeography, 10(4), 379-387. https://doi.org/10.1046/j.1466-822X.2001.00248.x

Franklin, J. (1986). Thematic Mapper analysis of coniferous forest structure and composition. International Journal of Remote Sensing, 7(10), pp.1287-1301. https://doi.org/10.1080/01431168608948931

Franco-Lopez, H., Ek, A. R., & Bauer, M. E. (2001). Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. Remote sensing of Environment, 77(3), 251-274. https://doi.org/10.1016/S0034-4257(01)00209-7

Frazer, G. W., Magnussen, S., Wulder, M. A., & Niemann, K. O. (2011). Simulated impact of sample plot size and co-registration error on the accuracy and uncertainty of LiDAR-derived estimates of forest stand biomass. Remote Sensing of Environment, 115(2), 636-649. https://doi.org/10.1016/j.rse.2010.10.008

Fritz, Ö. And Brunet, J. (2010). Epiphytic bryophytes and lichens in Swedish beech forests—effects of forest history and habitat quality. Ecological Bulletins, pp.95-108. www.jstor.org/stable/41442022

Gini, C. (1921) Measurement of inequality of incomes. Economic Journal 31:124–26. http://dx.doi.org/10.2307/2223319

Ginrich, S.F. (1967). Measuring and evaluating stocking and stand density in upland hardwood forests in the Central States. Forest Science, 13(1), pp.38-53. https://doi.org/10.1093/forestscience/13.1.38

Glasser, G. J. (1962). Variance formulas for the mean difference and coefficient of concentration. Journal of the American Statistical Association, 57(299), 648-654. https://doi.org/10.1080/01621459.1962.10500553

Gobakken, T., and Næsset, E. (2008). Assessing effects of laser point density, ground sampling intensity, and field sample plot size on biophysical stand properties derived from airborne laser scanner data. Canadian Journal of Forest Research. 38: 1095–1109. https://doi.org/10.1139/X07-219

Gobakken, T., & Næsset, E. (2009). Assessing effects of positioning errors and sample plot size on biophysical stand properties derived from airborne laser scanner data. Canadian Journal of Forest Research, 39(5), 1036-1052. https://doi.org/10.1139/X09-025

Gougeon, F.A., St-Onge, B.A., Wulder, M., and Leckie, D.G. (2001). Synergy of airborne laser altimetry and digital videography for individual tree crown delineation. In: Proceedings of the 23rd Canadian Symposium on Remote Sensing, August, pp. 21–24. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.713.8674&rep=rep1&type=pdf

Gove, J.H. (2004). Structural stocking guides: a new look at an old friend. Canadian Journal of Forest Research 34:1044–1056. http://dx.doi.org/10.1139/x03-272

Gove, J.H., Patil, G.P., and Taillie, C. (1995). A mathematical programming model for maintaining structural diversity in uneven-aged forest stands with implications to other formulations. Ecological Modeling. 79 (1–3), 11–19. https://doi.org/10.1016/0304-3800(94)00044-I

Häbel, H., Kuronen, M., Henttonen, H.M., Kangas, A. and Myllymäki, M. (2019). The effect of spatial structure of forests on the precision and costs of plot-level forest resource estimation. Forest Ecosystems, 6(1), p.8. https://doi.org/10.1186/s40663-019-0167-1

Hagar, J. C., Yost, A., & Haggerty, P. K. (2020). Incorporating LiDAR metrics into a structure-based habitat model for a canopy-dwelling species. Remote Sensing of Environment, 236, 111499. https://doi.org/10.1016/j.rse.2019.111499

Hansen, A. J., Phillips, L. B., Dubayah, R., Goetz, S., & Hofton, M. (2014). Regional-scale application of lidar: Variation in forest canopy structure across the southeastern US. Forest Ecology and Management, 329, 214-226. https://doi.org/10.1016/j.foreco.2014.06.009

Harch, B.D., Correll, R.L., Meech, W., Kirkby, C.A. and Pankhurst, C.E. (1997). Using the Gini coefficient with BIOLOG substrate 48tilization data to provide an alternative quantitative measure for comparing bacterial soil communities. Journal of Microbiological Methods, 30(1), pp.91-101. https://doi.org/10.1016/S0167-7012(97)00048-1

Harrell, P. A., Kasischke, E. S., Bourgeau-Chavez, L. L., Haney, E. M., & Christensen Jr, N. L. (1997). Evaluation of approaches to estimating aboveground biomass in southern pine forests using SIR-C data. Remote Sensing of Environment, 59(2), 223-233. https://doi.org/10.1016/S0034-4257(96)00155-1

Henttonen, H. M., & Kangas, A. (2015). Optimal plot design in a multipurpose forest inventory. Forest Ecosystems, 2(1), 31. https://doi.org/10.1186/s40663-015-0055-2

Hosking, J. R. M. (1989). Some theoretical results concerning L-moments. IBM Thomas J. Watson Research Division.

Hosking, J. R. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. Journal of the Royal Statistical Society: Series B (Methodological), 52(1), 105-124. https://doi.org/10.1111/j.2517-6161.1990.tb01775.x

Hudak, A.T., Crookston, N.L., Evans, J.S., Falkowski, M.J., Smith, A.M., Gessler, P.E. and Morgan, P. (2006). 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, 32(2), pp.126-138. https://doi.org/10.5589/m06-007

Humphrey, J.W., Newton, A., Peace, A.J., and Holden, E. (2000). The importance of conifer plantations in northern Britain as a habitat for native fungi. Biological Conservation. 96: 241–252. https://doi.org/10.1016/S0006-3207(00)00077-X

Hyde, P., Dubayah, R., Walker, W., Blair, J.B., Hofton, M. and Hunsaker, C. (2006). Mapping forest structure for wildlife habitat analysis using multi-sensor (LiDAR, SAR/InSAR, ETM+, Quickbird) synergy. Remote Sensing of Environment, 102(1-2), pp.63-73. https://doi.org/10.1016/j.rse.2006.01.021

Isenburg, M. (2016). LAStools — efficient tools for LiDAR processing (Version 160921,academic) [online]. Available from http://rapidlasso.com/LAStools.

Jakubowski, M. K., Guo, Q., & Kelly, M. (2013). Tradeoffs between lidar pulse density and forest measurement accuracy. Remote Sensing of Environment, 130, 245-253. https://doi.org/10.1016/j.rse.2012.11.024

Junttila, V., Finley, A. O., Bradford, J. B., & Kauranne, T. (2013). Strategies for minimizing sample size for use in airborne LiDAR-based forest inventory. Forest Ecology and Management, 292, 75-85. https://doi.org/10.1016/j.foreco.2012.12.019

Kaartinen, H., Hyyppä, J., Yu, X.; Vastaranta, M., Hyyppä, H., Kukko, A., Holopainen, M., Heipke, C., Hirschmugl, M., Morsdorf, F., Næsset, E., Pitkänen, J., Popescu, S., Solberg, S., Wolf, B.M. and Wu, J.C. (2012). An international comparison of individual tree detection and extraction using airborne laser scanning. Remote Sensing, 4(4), 950-974. https://doi.org/10.3390/rs4040950

Kandare, K., Ørka, H.O., Chan, J.C.-W., and Dalponte, M. (2016). Effects of forest structure and airborne laser scanning point cloud density on 3D delineation of individual tree crowns. European Journal of Remote Sensing. 49: 337–359. https://doi.org/10.5721/EuJRS20164919

Kankare, V., Vastaranta, M., Holopainen, M., Räty, M., Yu, X., Hyyppä, J., Hyyppä, H., Alho, P. and Viitala, R. (2013). Retrieval of forest aboveground biomass and stem volume with airborne scanning LiDAR. Remote Sensing, 5(5), pp.2257-2274. https://doi.org/10.3390/rs5052257

Kent, M. and P. Coker. (1992). The nature and properties of vegetation data. In Vegetation description and analysis: A practical approach. John Wiley & Sons, Chichester, West Sussex, UK. Pp. 77–111.

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

Knox, R.G., Peet R.K., and Christensen N.L. (1989). Population dynamics in loblolly pine stands: Changes in skewness and size inequality. Ecology 70: 1153–1167. http://dx.doi.org/10.2307/1941383

Kukkonen, M., Maltamo, M., Korhonen, L., & Packalen, P. (2019). Comparison of multispectral airborne laser scanning and stereo matching of aerial images as a single sensor solution to forest inventories by tree species. Remote Sensing of Environment, 231, 111208. https://doi.org/10.1016/j.rse.2019.05.027

Latifi, H. (2012). Characterizing forest structure by means of remote sensing: A review. In: Escalante-Ramírez B. (Ed.) Remote sensing-advanced techniques and platforms. InTech, Rijeka, Croatia. Pp. 3–28. http://dx.doi.org/10.5772/35143.

Le Toan, T., Quegan, S., Davidson, M.W.J., Balzter, H., Paillou, P., Papathanassiou, K., Plummer, S., Rocca, F., Saatchi, S., Shugart, H. and Ulander, L. (2011). The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. Remote sensing of environment, 115(11), 2850-2860. https://doi.org/10.1016/j.rse.2011.03.020

Lefsky, M.A., Cohen, W.B., Parker, G.G. and Harding, D.J. (2002). Lidar remote sensing for ecosystem studies: Lidar, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular interest to forest, landscape, and global ecologists. BioScience, 52(1), pp.19-30. https://doi.org/10.1641/0006-3568(2002)052[0019:LRSFES]2.0.CO;2

Lei, X., Wang, W., & Peng, C. (2009). Relationships between stand growth and structural diversity in spruce-dominated forests in New Brunswick, Canada. Canadian Journal of Forest Research, 39(10), 1835-1847. https://doi.org/10.1139/X09-089

Lelli, C., Bruun, H.H., Chiarucci, A., Donati, D., Frascaroli, F., Fritz, Ö., Goldberg, I., Nascimbene, J., Tøttrup, A.P., Rahbek, C. and Heilmann-Clausen, J. (2019). Biodiversity response to forest structure and management: Comparing species richness, conservation relevant species and functional diversity as metrics in forest conservation. Forest Ecology and Management, 432, pp.707-717. https://doi.org/10.1016/j.foreco.2018.09.057

Lexerød, N.L., and Eid, T. (2006). An evaluation of different diameter diversity indices based on criteria related to forest management planning. Forest Ecology and Management. 222: 17–28. https://doi:10.1016/j.foreco.2005.10.046

Lipovetsky, S. (2013). How good is best? Multivariate case of Ehrenberg-Weisberg analysis of residual errors in competing regressions. Journal of Modern Applied Statistical Methods, 12(2), 14. https://doi:10.22237/jmasm/1383279180

Liu, X., Zhang, Z., Peterson, J., and Chandra, S. (2007). The effect of LiDAR data density on DEM accuracy. In Proceedings of the International Congress on Modelling and Simulation (MODSIM07) 2007, Modelling and Simulation Society of Australia and New Zealand Inc. pp. 1363–1369. http://www.mssanz.org.au/MODSIM07/papers/21_s46/TheEffectOfL_s46_Liu_.pdf

Lombardi, F., Marchetti, M., Corona, P., Merlini, P., Chirici, G., Tognetti, R., Burrascano, S., Alivernini, A., and Puletti, N. 2015. Quantifying the effect of sampling plot size on the estimation of structural indicators in old-growth forest stands. Forest Ecology and Management. 346: 89–97. https://doi.org/10.1016/j.foreco.2015.02.011

MacArthur, R.H., MacArthur, J.W. (1961). On bird species diversity. Ecology 42: 594–598.

http://dx.doi.org/10.2307/1932254

Magurran, A.E., 2004. Measuring biological diversity. John Wiley & Sons. http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0632056339.html [cited 3 June 2020]

Magurran, A.E. (2005). Species abundance distributions: pattern or process?. Functional Ecology, 19(1), pp.177-181. https://doi.org/10.1111/j.0269-8463.2005.00930.x

Maltamo, M., Bollandsås, O. M., Gobakken, T. and Næsset, E. (2016). Large-scale prediction of aboveground biomass in heterogeneous mountain forests by means of airborne laser scanning. Canadian Journal of Forest Research, 46(9), 1138-1144. https://doi.org/10.1139/cjfr-2016-0086

Maltamo, M., Eerikäinen, K., Packalén, P., & Hyyppä, J. (2006). Estimation of stem volume using laser scanning-based canopy height metrics. Forestry, 79(2), 217-229. https://doi.org/10.1093/forestry/cpl007

Maltamo, M., Eerikäinen, K., Pitkänen, J., Hyyppä, J., Vehmas, M. (2004). Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions. Remote Sensing of Environment. 90 (3), 319–330. https://doi.org/10.1016/j.rse.2004.01.006

Maltamo, M., Hauglin, K.M., Næsset, E. and Gobakken, T. 2019. Estimating stand level stem diameter distribution utilizing accurately positioned tree-level harvester data and airborne laser scanning. Silva Fennica. 53(3), id 10075. https://doi.org/10.14214/sf.10075

Maltamo, M. and Kangas, A. (1998). Methods based on k-nearest neighbor regression in the prediction of basal area diameter distribution. Canadian Journal of Forest Research, 28(8), 1107-1115. https://doi.org/10.1139/x98-085

Maltamo, M., Mehtätalo, L., Vauhkonen, J., Packalén, P. (2012). Predicting and calibrating tree size and quality attributes by means of airborne laser scanning and field measurements. Canadian Journal of Forest Research. 42 (11), 1896–1907. https://doi.org/10.1139/x2012-134

Maltamo, M., Næsset, E., Vauhkonen J. (Eds.) (2014). Forestry applications of airborne laser scanning. Concepts and Case Studies. Managing Forest Ecosystems 27. Springer, 51tilizati. https://doi.org/10.1007/978-94-017-8663-8

Maltamo, M., Ørka, H.O., Bollandsås, O.M., Gobakken, T. and Næsset, E. (2015). Using pre-classification to improve the accuracy of species-specific forest attribute estimates from airborne laser scanner data and aerial images. Scandinavian Journal of Forest Research. 30: 336-345. https://doi.org/10.1080/02827581.2014.986520

Maltamo, M., Packalén, P., Yu, X., Eerikäinen, K., Hyyppä, J., Pitkänen, J. (2005). Identifying and quantifying structural characteristics of heterogeneous boreal forests using laser scanner data. Forest Ecology and Management. 216 (1–3), 41–50.

https://doi.org/10.1016/j.foreco.2005.05.034

Mandallaz, D. (2007). Sampling Techniques for Forest Inventories. CRC Press, Boca Raton. https://doi.org/10.1201/9781584889779

Marvin, D.C., Asner, G.P., Knapp, D.E., Anderson, C.B., Martin, R.E., Sinca, F., Tupayachi, R. (2014). Amazonian landscapes and the bias in field studies of forest structure and biomass. Proceedings of the National Academy of Sciences. 111 (48), E5224–E5232. https://doi.org/10.1073/pnas.1412999111

Mascaro, J., Detto, M., Asner, G.P., and Muller-Landau, H.C. (2011). Evaluating uncertainty in mapping forest carbon with airborne LiDAR. Remote Sensing of Environment. 115: 3770–3774. https://doi.org/10.1016/j.rse.2011.07.019

Matos, A. (2014). Effect of scale factor in estimation of Gini coefficient [online]. M.Sc. thesis, University of Eastern Finland. Available from http://epublications.uef.fi/pub/urn_nbn_fi_uef-20140718/index_en.html and http://www.oppi.uef.fi/opk/video/europeanforestry/a_matos_seminar.mp4 [accessed August 2016].

Mauro, F., Molina, I., García-Abril, A., Valbuena, R., & Ayuga-Téllez, E. (2016). Remote sensing estimates and measures of uncertainty for forest variables at different aggregation levels. Environmetrics, 27(4), 225-238. https://doi.org/10.1002/env.2387

McElhinny, C., Gibbons, P., Brack, C., & Bauhus, J. (2005). Forest and woodland stand structural complexity: its definition and measurement. Forest Ecology and Management, 218(1-3), 1-24. https://doi.org/10.1016/j.foreco.2005.08.034

McGaughey, R. 2015. FUSION/LDV: software for LIDAR data analysis and visualization. Version 3.50. United States Department of Agriculture (USDA) Forest Service, Pacific Northwest Research Station [accessed December 2015].

McInerney, D.O., Suárez J., Valbuena R., Nieuwenhuis M. (2010) Forest canopy height retrieval using lidar data, medium-resolution satellite imagery and kNN estimation in Aberfoyle, Scotland. Forestry 83: 195–206. http://dx.doi.org/10.1093/forestry/cpq001

McIntosh, R.P. (1967). An index of diversity and the relation of certain concepts to diversity. Ecology 48: 392–404. http://dx.doi.org/10.2307/1932674

Means, J.E., Acker, S.A., Fitt, B.J., Renslow, M., Emerson, L., Hendrix, C.J. (2000). Predicting forest stand characteristics with airborne scanning lidar. Photogrammetric Engineering and Remote Sensing. 66 (11), 1367–1372. https://www.asprs.org/wp-content/uploads/pers/2000journal/november/2000_nov_1367-1371.pdf

Meyer, D., Zeileis, A., Hornik, K., 2014. Vcd: Visualizing Categorical Data. R package version 1.3-2.

Mononen, L., Auvinen, A.P., Packalen, P., Virkkala, R., Valbuena, R., Bohlin, I., Valkama, J. and Vihervaara, P. (2018). Usability of citizen science observations together with airborne laser scanning data in determining the habitat preferences of forest birds. Forest Ecology and Management, 430, pp.498-508. https://doi.org/10.1016/j.foreco.2018.08.040

Montgomery, R.A. and Chazdon, R.L. (2001). Forest structure, canopy architecture, and light transmittance in tropical wet forests. Ecology, 82(10), pp.2707-2718. https://doi.org/10.1890/0012-9658(2001)082[2707:FSCAAL]2.0.CO;2

Moss, I. (2012). Stand structure classification, succession, and mapping using LiDAR (Doctoral dissertation). University of British Columbia. https://doi.org/10.1139/X09-025

Motz, K., Sterba, H., and Pommerening, A. (2010). Sampling measures of tree diversity. Forest Ecology and Management. 260: 1985–1996. https://doi.org/10.1016/j.foreco.2010.08.046

Müllner, D. (2013). fastcluster: Fast hierarchical, agglomerative clustering routines for R and Python. Journal of Statistical Software, 53(9), 1-18. http://dx.doi.org/10.18637/jss.v053.i09

Næsset, E. (2007). Airborne laser scanning as a method in operational forest inventory: Status of accuracy assessments accomplished in Scandinavia. Scandinavian Journal of Forest Research, 22(5), 433-442. https://doi.org/10.1080/02827580701672147

Næsset, E. (2002). Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote sensing of environment, 80(1), pp.88-99. https://doi.org/10.1016/S0034-4257(01)00290-5

Næsset, E., and Gobakken, T. (2008). Estimation of above-and below-ground biomass across regions of the boreal forest zone using airborne laser. Remote Sensing of Environment, 112(6), 3079-3090. https://doi.org/10.1016/j.rse.2008.03.004

Næsset, E., Gobakken, T., Solberg, S., Gregoire, T. G., Nelson, R., Ståhl, G., & Weydahl, D. (2011). Model-assisted regional forest biomass estimation using LiDAR and InSAR as auxiliary data: A case study from a boreal forest area. Remote Sensing of Environment, 115(12), 3599-3614. https://doi.org/10.1016/j.rse.2011.08.021

Näslund, M. (1936). Skogsförsöksanstaltens gallringsförsök i tallskog. Meddelanden från Statens Skogsförsöksanstalt, 29, 1-169 p. (In Swedish)

Nelson, R., Gobakken, T., Stahl, G. and Gregoire, T. G. (2008). Regional Forest Inventory using an Airborne Profiling LiDAR (< Special Issue> Silvilaser). Journal of Forest Planning, 13(Special_Issue), 287-294. https://doi.org/10.20659/jfp.13.Special_Issue_287

Neumann, M., Starlinger F. (2001). The significance of different indices for stand structure and diversity in forests. Forest Ecology and Management. 145: 91–106. http://dx.doi.org/10.1016/s0378-1127(00)00577-6

Newton, P. (1997). Stand density management diagrams: review of their development and utility in stand-level management planning. Forest Ecology and Management. 98 (3), 251–265. https://doi.org/10.1016/S0378-1127(97)00086-8

Nguyen, T. H., Jones, S., Soto-Berelov, M., Haywood, A., & Hislop, S. (2019, July). Estimate Forest Biomass Dynamics Using Multi-Temporal Lidar And Single-Date Inventory Data. In IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium (pp. 7338-7341). IEEE.

Ni, W., Liu, J., Zhang, Z., Sun, G., & Yang, A. (2015, July). Evaluation of UAV-based forest inventory system compared with LiDAR data. In 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) (pp. 3874-3877). IEEE. https://doi.org/10.1109/IGARSS.2015.7326670

Otypková, Z., and Chytry, M. (2006). Effects of plot size on the ordination of vegetation samples. Journal of Vegetation Science. 17: 465–472. https://doi.org/10.1111/j.1654-1103.2006.tb02467.x

O'Hara, K.L., Gersonde, R.F. (2004). Stocking control concepts in uneven-aged silviculture. Forestry 77 (2), 131–143. https://doi.org/10.1093/forestry/77.2.131

O'Hara K.L., Hasenauer H., Kindermann G. (2007). Sustainability in multi-aged stands: An analysis of long-term plenter systems. Forestry 80: 163–181. http://dx.doi.org/10.1093/forestry/cpl051

Packalén, P. and Maltamo, M. (2006). Predicting the plot volume by tree species using airborne laser scanning and aerial photographs. Forest Science, 52(6), pp.611-622. https://doi.org/10.1093/forestscience/52.6.611

Packalén, P., Mehtätalo, L. and Maltamo, M. (2011). ALS-based estimation of plot volume and site index in a eucalyptus plantation with a nonlinear mixed-effect model that accounts for the clone effect. Annals of Forest Science, 68(6), p.1085. https://doi.org/10.1007/s13595-011-0124-9

Packalen, P., Strunk, J., Packalen, T., Maltamo, M. and Mehtätalo, L. (2019). Resolution dependence in an area-based approach to forest inventory with airborne laser scanning. Remote sensing of environment, 224, pp.192-201. https://doi.org/10.1016/j.rse.2019.01.022

Packalen, P., Vauhkonen, J., Kallio, E., Peuhkurinen, J., Pitkänen, J., Pippuri, I., Strunk, J. and Maltamo, M. (2013). Predicting the spatial pattern of trees by airborne laser scanning. International journal of remote sensing, 34(14), 5154-5165. https://doi.org/10.1080/01431161.2013.787501

Päivinen, R (1987) Metsän inventoinnin suunnittelumalli. [A planning model for forest inventory, In Finnish]. 11th edn. University of Joensuu publications in Sciences, University of Joensuu, Joensuu.

Pascual, C., García-Abril, A., García-Montero, L.G., Martín-Fernández, S., Cohen, W.B. (2008). Object-based semi-automatic approach for forest structure characterization using lidar data in heterogeneous *Pinus sylvestris* stands. Forest Ecology and Management. 255 (11), 3677–3685. https://doi.org/10.1016/j.foreco.2008.02.055

Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C. and Coops, N.C. (2013). Essential biodiversity variables. Science, 339(6117), pp.277-278. https://doi.org/10.1126/science.1229931

Pippuri, I., Suvanto, A., Maltamo, M., Korhonen, K.T., Pitkänen, J. and Packalen, P. (2016). Classification of forest land attributes using multi-source remotely sensed data. International Journal of Applied Earth Observation and Geoinformation 44: 11-22. https://doi.org/10.1016/j.jag.2015.07.002

Pommerening, A. (2002). Approaches to quantifying forest structures. Forestry: An International Journal of Forest Research. 75 (3), 305–324. https://doi.org/10.1093/forestry/75.3.305

Pommerening, A., & Stoyan, D. (2006). Edge-correction needs in estimating indices of spatial forest structure. Canadian Journal of Forest Research, 36(7), 1723-1739. https://doi.org/10.1139/cjfr-2017-0084

Puliti S., Breidenbach J., Astrup R. (2020). Estimation of Forest Growing Stock Volume with UAV Laser Scanning Data: Can It Be Done without Field Data?. Remote Sensing 12(8): 1245. https://doi.org/10.3390/rs12081245

Rana, P., Korhonen, L., Gautam, B., & Tokola, T. (2014). Effect of field plot location on estimating tropical forest above-ground biomass in Nepal using airborne laser scanning data. ISPRS journal of photogrammetry and remote sensing, 94, 55-62. https://doi.org/10.1016/j.isprsjprs.2014.04.012

Ranson, K. J., Saatchi, S., & Sun, G. (1995). Boreal forest ecosystem characterization with SIR-C/XSAR. IEEE Transactions on Geoscience and Remote Sensing, 33(4), 867-876. https://doi.org/10.1109/36.406673

Räty, J., Packalen, P., Kotivuori, E. and Maltamo, M. (2020). Fusing diameter distributions predicted by an area-based approach and individual-tree detection in coniferous-dominated forests. Canadian Journal of Forest Research, 50(999), pp.113-125. http://dx.doi.org/10.1139/cjfr-2019-0102

Räty, J., Packalen, P., & Maltamo, M. (2018). Comparing nearest neighbor configurations in the prediction of species-specific diameter distributions. Annals of forest science, 75(1), 26. https://doi.org/10.1007/s13595-018-0711-0

Repola, J., (2008). Biomass equations for birch in Finland. Silva Fennica, 42(1): 605-624. https://doi.org/10.14214/sf.236

Repola, J., (2009). Biomass equations for Scots pine and Norway spruce in Finland. Silva Fennica 43(4): 625-647. https://doi.org/10.14214/sf.184

Rocha de Souza Pereira, F., Kampel, M., Gomes Soares, M. L., Estrada, G. C. D., Bentz, C., and Vincent, G. (2018). Reducing uncertainty in mapping of mangrove aboveground biomass using airborne discrete return lidar data. Remote Sensing, 10(4), 637. https://doi.org/10.3390/rs10040637

Rouvinen, S., Kuuluvainen T. (2005). Tree diameter distributions in natural and managed old pinus sylvestris-dominated forests. Forest Ecology and Management 208: 45–61. http://dx.doi.org/10.1016/j.foreco.2004.11.021

Ruiz, L.A., Hermosilla, T., Mauro, F., and Godino, M. (2014). Analysis of the influence of plot size and LiDAR density on forest structure attribute estimates. Forests, 5(5): 936–951. http://dx.doi.org/10.3390/f5050936

Sankey, T., Donager J., McVay J., Sankey J.B. (2017). UAV lidar and hyperspectral fusion for forest monitoring in the southwestern USA. Remote Sensing of Environment 195: 30–43. https://doi.org/10.1016/j.rse.2017.04.007

Shannon, C.E. (1948). A mathematical theory of communication. In: Shannon C.E., Weaver W. (Eds.) A mathematical theory of communication. University of Illinois Press. pp. 29–125. http://www.mast.queensu.ca/~math474/shannon1948.pdf [Cited 3 June 2020].

Shupe, S. M., and Marsh, S. E. (2004). Cover-and density-based vegetation classifications of the Sonoran Desert using Landsat TM and ERS-1 SAR imagery. Remote Sensing of Environment, 93(1-2), 131-149. https://doi.org/10.1016/j.rse.2004.07.002

Siipilehto J., (1999). Improving the accuracy of predicted basal-area diameter distribution in advanced stands by determining stem number. Silva Fennica, 33(4), 281-301. https://doi.org/10.14214/sf.650

Simpson, E.H. (1949). Measurement of diversity. Nature, 163: 688. http://dx.doi.org/10.1038/163688a0

Simpson, J. E., Smith, T. E., & Wooster, M. J. (2017). Assessment of errors caused by forest vegetation structure in airborne LiDAR-derived DTMs. Remote Sensing, 9(11), 1101. https://doi.org/10.3390/rs9111101

Smith, D. (1997). The practice of silviculture: applied forest ecology (9th ed.). New York, NY: John Wiley & Sons.

Smith, B., Wilson J.B. (1996). A consumer's guide to evenness indices. Oikos 76: 70–82. http://dx.doi.org/10.2307/3545749

Spies, T.A. and Franklin, J.F. (1991). The structure of natural young, mature, and old-growth Douglas-fir forests in Oregon and Washington. Wildlife and vegetation of unmanaged Douglas-fir forests, pp.91-109. https://pdfs.semanticscholar.org/71ef/bf19b3178748886854b3a7a638611e22b5ae.pdf

Stark, S.C., Leitold, V., Wu, J.L., Hunter, M.O., de Castilho, C.V., Costa, F.R., McMahon, S.M., Parker, G.G., Shimabukuro, M.T., Lefsky, M.A. and Keller, M. (2012). Amazon forest carbon dynamics predicted by profiles of canopy leaf area and light environment. Ecology letters, 15(12), pp.1406-1414. https://doi.org/10.1111/j.1461-0248.2012.01864.x

Staudhammer, C.L. and LeMay V.M. (2001). Introduction and evaluation of possible indices of stand structural diversity. Canadian Journal of Forest Research 31: 1105–1115. http://dx.doi.org/10.1139/cjfr-31-7-1105

Story, M., & Congalton, R. G. (1986). Accuracy assessment: a user's perspective. Photogrammetric Engineering and remote sensing, 52(3), 397-399. https://www.asprs.org/wp-content/uploads/pers/1986journal/mar/1986_mar_397-399.pdf

Straub, C., Tian, J., Seitz, R., & Reinartz, P. (2013). Assessment of Cartosat-1 and WorldView-2 stereo imagery in combination with a LiDAR-DTM for timber volume estimation in a highly structured forest in Germany. Forestry, 86(4), 463-473. https://doi.org/10.1093/forestry/cpt017

Su, Y., Guo, Q., Xue, B., Hu, T., Alvarez, O., Tao, S. and Fang, J. (2016). Spatial distribution of forest aboveground biomass in China: Estimation through combination of spaceborne lidar, optical imagery, and forest inventory data. Remote Sensing of Environment, 173, 187-199. https://doi.org/10.1016/j.rse.2015.12.002

Sun, T., Zhang, H., Wang, Y., Meng, X. and Wang, C. (2010). The application of environmental Gini coefficient (EGC) in allocating wastewater discharge permit: The case study of watershed total mass control in Tianjin, China. Resources, Conservation and Recycling, 54(9), pp.601-608. https://doi.org/10.1016/j.resconrec.2009.10.017

Tedeschi, L. O. (2006). Assessment of the adequacy of mathematical models. Agricultural systems, 89(2-3), 225-247. https://doi.org/10.1016/j.agsy.2005.11.004

Thomas, V., Treitz, P., McCaughey, J. H., & Morrison, I. (2006). Mapping stand-level forest biophysical variables for a mixedwood boreal forest using lidar: an examination of scanning density. Canadian Journal of Forest Research, 36(1), 34-47. https://doi.org/10.1139/x05-230

Le Toan, T., Quegan, S., Davidson, M.W.J., Balzter, H., Paillou, P., Papathanassiou, K., Plummer, S., Rocca, F., Saatchi, S., Shugart, H. and Ulander, L. (2011). The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle.

Remote sensing of environment, 115(11), pp.2850-2860. https://doi.org/10.1016/j.rse.2011.03.020

Tomppo, E., Kuusinen, N., Mäkisara, K., Katila, M., and McRoberts, R. E. (2017). Effects of field plot configurations on the uncertainties of ALS-assisted forest resource estimates. Scandinavian Journal of Forest Research, 32(6), 488-500. https://doi.org/10.1080/02827581.2016.1259425

Vauhkonen J., Ene L., Gupta S., Heinzel J., Holmgren J., Pitkänen J., Solberg S., Wang Y., Weinacker H., Hauglin KM., Lien V., Packalén P., Gobakken T., Koch B., Næsset E., Tokola T. and Maltamo, M. (2012). Comparative testing of single-tree detection algorithms under different types of forest. Forestry, 85(1), 27-40. https://doi.org/10.1093/forestry/cpr051

Valbuena, R. (2015). Forest structure indicators based on tree size inequality and their relationships to airborne laser scanning. Dissertation Forestales 1. pp. 11–21. http://dx.doi.org/10.14214/df.205

Valbuena, R., Eerikäinen, K., Packalen, P., Maltamo, M. (2016a). Gini coefficient predictions from airborne lidar remote sensing display the effect of management intensity on forest structure. Ecological Indicators. 60, 574–585. https://doi.org/10.1016/j.ecolind.2015.08.001

Valbuena, R., Hernando, A., Manzanera, J.A., Görgens, E.B., Almeida, D.R.A., Mauro, F., García-Abril, A. and Coomes, D.A., 2017a. Enhancing of accuracy assessment for forest above-ground biomass estimates obtained from remote sensing via hypothesis testing and overfitting evaluation. Ecological Modelling, 366, pp.15-26. https://doi.org/10.1016/j.ecolmodel.2017.10.009

Valbuena, R., Maltamo, M., Mehtätalo, L. and Packalen, P. (2017b). Key structural features of Boreal forests may be detected directly using L-moments from airborne lidar data. Remote Sensing of Environment, 194, 437-446. https://doi.org/10.1016/j.rse.2016.10.024

Valbuena, R., Maltamo, M. and Packalen, P. (2016b). Classification of multilayered forest development classes from low-density national airborne lidar datasets. Forestry: An International Journal of Forest Research, 89(4), 392-401. https://doi.org/10.1093/forestry/cpw010

Valbuena, R., Packalen, P., Mehtätalo, L., García-Abril, A., Maltamo, M. (2013). Characterizing forest structural types and shelterwood dynamics from Lorenz-based indicators predicted by airborne laser scanning. Canadian Journal of Forest Research. 43 (11), 1063–1074. http://dx.doi.org/10.1139/cjfr-2013-0147

Valbuena, R., Packalén, P., Martín-Fernández, S., and Maltamo, M. (2012). Diversity and equitability ordering profiles applied to study forest structure. Forest Ecology and Management. 276: 185–195. https://doi.org/10.1016/j.foreco.2012.03.036

Valbuena, R., Vauhkonen, J., Packalen, P., Pitkänen, J., Maltamo, M. (2014). Comparison of airborne laser scanning methods for estimating forest structure indicators based on Lorenz curves. ISPRS Journal of Photogrammetry and Remote Sensing. 95, 23–33. https://doi.org/10.1016/j.isprsjprs.2014.06.002

Van Aardt, J.A., Wynne, R.H. and Scrivani, J.A. (2008). Lidar-based mapping of forest volume and biomass by taxonomic group using structurally homogenous segments. Photogrammetric Engineering & Remote Sensing, 74(8), pp.1033-1044. https://doi.org/10.14358/PERS.74.8.1033

von Gadow K. Zhang C., Wehenkel C., Pommerening A., Corral-Rivas J., Korol M., Myklush S., Hui G., Kiviste A., Zhao X. (2012). Forest structure and diversity. In: Pukkala T., von Gadow K. (Eds.) Continuous Cover Forestry. Managing Forest Ecosystems 23. Springer, Dordrecht. Netherlands. pp. 29–83. http://dx.doi.org/10.1007/978-94-007-2202-6_2

Venables, W. N., & Ripley, B. D. (2001). Modern Applied Statistics with S-Plus. Third ed. Springer. https://www2.isye.gatech.edu/~brani/isyebayes/bank/VenRip3R.pdf

Vihervaara, P., Mononen, L., Auvinen, A., Virkkala, R., Lü, Y., Pippuri, I., Packalen, P., Valbuena, R., Valkama, J. (2015). How to integrate remotely sensed data and biodiversity for ecosystem assessments at landscape scale. Landscape Ecology. 30 (3), 501–516. https://doi.org/10.1016/j.gecco.2017.01.007

Vincent, G., Sabatier, D. and Rutishauser, E. (2014). Revisiting a universal airborne light detection and ranging approach for tropical forest carbon mapping: scaling-up from tree to stand to landscape. Oecologia, 175(2), pp.439-443. https://doi.org/10.1007/s00442-014-2913-y

Weiner, J., and Thomas, S.C. (1986). Size variability and competition in plant monocultures. Oikos, 47: 211–222. https://doi.org/10.2307/3566048

Weiner, J., and Solbrig, O.T. (1984). The meaning and measurement of size hierarchies in plant populations. Oecologia, 61: 334–336. https://doi.org/10.1007/BF00379630

Weltz, M. A., Ritchie, J. C. and Fox, H. D. (1994). Comparison of laser and field measurements of vegetation height and canopy cover. Water Resources Research, 30(5), 1311-1319. https://doi.org/10.1029/93WR03067

Whittaker, R.H. 1972. Evolution and measurement of species diversity. Taxon, 21: 213–251. https://doi.org/10.2307/1218190

Whittaker, R.H. (1977). Evolution of species diversity in land communities. In: Hecht M.K.,Steere W.C., Wallace B. (Eds). Evolutionary Biology 10. Plenum, New York, USA. pp.1–67.

Wulder, M. A., Bater, C. W., Coops, N. C., Hilker, T., & White, J. C. (2008). The role of LiDAR in sustainable forest management. The forestry chronicle, 84(6), 807-826. https://doi.org/10.5558/tfc2020-003

Yu, X., Hyyppä, J., Holopainen, M. and Vastaranta, M. (2010). Comparison of area-based and individual tree-based methods for predicting plot-level forest attributes. Remote Sensing, 2(6), pp.1481-1495. https://doi.org/10.3390/rs2061481

Zheng, X., Xia, T., Yang, X., Yuan, T. and Hu, Y. (2013). The land Gini coefficient and its application for land use structure analysis in China. PloS One, 8(10). https://dx.doi.org/10.1371%2Fjournal.pone.0076165

Zhang, Z., Liu, X., Peterson, J., & Wright, W. (2011). Cool temperate rainforest and adjacent forests classification using airborne LiDAR data. Area, 43(4), 438-448. https://doi.org/10.1111/j.1475-4762.2011.01035.x

Zimble, D.A., Evans, D.L., Carlson, G.C., Parker, R.C., Grado, S.C. and Gerard, P.D., 2003. Characterizing vertical forest structure using small-footprint airborne LiDAR. Remote sensing of Environment, 87(2-3), pp.171-182. https://doi.org/10.1016/S0034-4257(03)00139-1