Canadian Journal of Forest Research
Revue canadienne de recherche forestière

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

Journal:	*Canadian Journal of Forest Research*
Manuscript ID	cjfr-2017-0084.R2
Manuscript Type:	Article
Date Submitted by the Author:	27-Sep-2017
Complete List of Authors:	Adnan, Syed; Ita-Suomen yliopisto Luonnontieteiden ja metsatieteiden tiedekunta, Forest Sciences Maltamo, Matti; University of Eastern Finland, School of Forest Sciences Coomes, David; Department of Plant Sciences Valbuena, Ruben; University of Cambridge. , Department of Plant Sciences
Keyword:	Sample Size Optimization, Forest Structure, LiDAR, Structural Heterogeneity, Spatial Resolution Optimization
Is the invited manuscript for consideration in a Special Issue? :	N/A

Title:

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

Authors:

Syed Adnan *(1), Matti Maltamo (1), David Coomes (2), Rubén Valbuena (2)

Affiliations:

(1) University of Eastern Finland. Faculty of Forest Sciences. PO Box 111 Joensuu, Finland; adnan@uef.fi; matti.maltamo@uef.fi.

(2) University of Cambridge, Department of Plant Sciences. Forest Ecology and Conservation. Downing Street, CB2 3EA Cambridge, UK. dac18@cam.ac.uk; rv314@cam.ac.uk.

*Corresponding author.

Abstract

The estimation of Gini Coefficient (GC) of tree sizes using airborne laser scanning (ALS) can provide maps of forest structure across the landscape, which can support sustainable forest management. A challenge arises in determining the optimal spatial resolution that maximizes the stability and precision of GC estimates, which in turn depends upon stand density or ALS scan density. By subsampling different plot sizes within large field plots, we evaluated the optimal spatial resolution by observing changes in GC estimation and in its correlation with ALS metrics. We found that plot size had greater effects than either stand density or ALS scan density in the relationship between GC and ALS metrics. Uncertainty in GC estimates fell as plot size increased. Correlation with ALS metrics showed convex curves with maxima at 250-450 m^2, which thus was considered the optimal plot size / spatial resolution. By thinning the density of ALS point cloud, we deduced that at least 3 $points \cdot m^{-2}$ are needed for reliable GC estimates. Many nationwide ALS scan densities are sparser than this, which may be unreliable for GC estimation. Ours is a simple approach for evaluating the optimal spatial resolution in remote sensing estimation of any forest attribute.

Key words

structural heterogeneity; spatial resolution optimization; sample size optimization; forest structure; LiDAR

1. Introduction

1.1 The Gini Coefficient as an Indicator of Forest Structural Heterogeneity

Forest structural characteristics are widely used in the development of sustainable management plans designed to protect habitats while carrying out forestry operations (Upton and Fingleton, 1985; Pommerening, 2002; Motz et al., 2010; Vihervaara et al., 2015; Valbuena et al., 2016). Management can be designed to emulate natural dynamics (Oliver and Larson, 1990; Buongiorno et al., 1994; Lähde et al., 1999; Pukkala et al., 2016), by studying how silvicultural operations affects forest structure locally (Humphrey et al., 2000; Valbuena et al., 2013a; Robles et al., 2016).

Forest structure is often characterized by stem diameter distributions (O'Hara and Gersonde, 2004; McElhinny et al., 2005). If a single concise indicator of size inequality is desired, there are many available, including Shannon or Simpson indices (Neumann and Starlinger, 2001; Sterba and Ledermann, 2006; O'Hara et al., 2007; Lei et al., 2009) or variance-based metrics (Staudhammer and LeMay, 2001). Recent research has highlighted the effectiveness of the Gini coefficient (i.e. GC, Gini, 1921) for assessing the structural diversity (Lexerød and Eid, 2006a; O'hara et al. 2007; Duduman, 2009; Valbuena et al., 2012, 2013a). Originally developed for evaluating inequality in income distributions (e.g., Hvistendahl, 2014), GC has been applied to a variety of fields, such as healthcare (Asada, 2005) or land use (Zheng et al., 2013). In plant sciences, it has been employed to evaluate size inequality (Weiner and Solbrig, 1984). It has also been applied to forest ecosystems (Weiner and Thomas, 1986), to quantify structural diversity (Knox and Peet, 1989), analyse competition (Lundqvist, 1994; Cordonnier and Kunstler, 2015), or successional stages (Valbuena et al., 2013a). Comparative studies indicate that GC is the best index for characterizing diameter distributions, providing a logical ranking of different stand types (Lexerød and Eid, 2006a; Valbuena et al., 2012), so that forest may be stratified according to their structure (Bollandsås and Næsset, 2007). It can also be used to observe the effects of different management regimes (Bourdier et al., 2016; Pukkala et al. 2016; Valbuena et al., 2016). For these reasons, estimation of GC is the focus of this article.

When used in forest science, GC evaluates size inequality of trees growing in a vicinity (Weiner, 1990). For a patch of forest containing n trees, within which the i^{th} and j^{th} tree have basal areas of g_i and g_j respectively, GC is calculated as (Glasser, 1962):

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

Therefore, GC is a statistical measure of relative dispersion, which is equivalent to half of the relative mean absolute difference (Valbuena et al, 2017: appendix A3), and it ranges between 0-1, zero representing perfect equality and one being maximum inequality (Gini, 1921). Hence GC describes the shape of tree-size distributions (Valbuena et al., 2016) and is influenced by competitive interactions among trees (Cordonnier and Kunstler, 2015). Valbuena et al. (2012) demonstrated that the $GC = 0.5$ can be considered as a boundary between even-aged and uneven-aged stand structures. GC values far below 0.5 indicate a unimodal "normally distributed" size structure, which are commonly found in even-aged stands that are self-thinning (e.g. Coomes and Allen, 2007). Values close to 0.5 indicate irregular size distributions (Duduman, 2009), while values much greater than 0.5 represent "reverse-J" stand structures (Valbuena et al., 2013a).

1.2 Influence of Plot Size in Measurements of Forest Structure

Sample plots used for measuring plant communities are usually rectangular or circular in shape (Whittaker, 1972; Kent and Coker, 1992), with a wide range of possible plot sizes from fine to coarse scales (Chytrý and Otýpková, 2003). As the effects of plot size decrease with increasing size of a plot (David and Mishriky, 1968; Barbeito et al., 2009), an optimal size must be chosen compromising the acquisition of a field plot large enough to obtain a stable measure of forest structure, but no larger than necessary because of the costs involved (Otypková and Chytry, 2006). Structural diversity depends on the spatial resolution at which an index is evaluated (Lexerød and Eid, 2006b). Varying the scale of observation may therefore distort the information retrieved from an indicator (Chen and Crawford, 2012; Mauro et al, 2016). As plot size increases, GC estimates may be more reliable, but also fundamentally different stand conditions may aggregate (Coomes and Allen, 2007). Therefore, interpretation of data analysed at different scales remains one of the most challenging tasks in spatial statistics (Gotway and Young, 2002), as shown in the context of agriculture (Smith, 1938), sociology (Hannan, 1971), and environmental sciences (Jelinski and Wu, 1996). Also, the spatial distribution of trees has a practical effect on plot size, since clustered patterns require larger plot sizes to obtain a same degree of confidence in estimates (Upton and Fingleton, 1985; Pommerening, 2002; Kallimanis et al., 2008; Motz et al., 2010). Recently, Magnussen et al. (2016) suggested a method of upscaling to a common plot size to minimize scale effects in survey estimates, which achieved consistency among the quantiles and proportions of sampling distributions of forest attributes.

1.3 Influence of ALS Scan Density in Measurements of Forest Structure

Airborne laser scanning (ALS) is recognised as a highly effective tool for regional monitoring because it provides precise information about biophysical stand properties, (Gobakken et al., 2006; Gobakken and Næsset, 2008). The GC may be calculated as a descriptor of the distribution of ALS heights (Valbuena et al., 2017), or ALS metrics may be related to GC of tree sizes (Valbuena et al., 2013b). The spatial resolution of ALS data used in area-based methods has an effect on estimated values (Mascaro et al., 2011). In the context of remote sensing-assisted forest estimations, spatial resolution refers not only to the size of field plots but also to the size of pixels at which auxiliary variables are computed (Gobakken and Næsset, 2008; Ruiz et al., 2014; Valbuena et al., 2016). In ALS-assisted estimations of GC of tree size inequality, there is a lack of knowledge on the effects of varying plot size and spatial resolution.

Scan density is one of the most important aspects of ALS datasets that affects both processing and costs (Balsa-Barreiro and Lerma, 2014; Kandare et al., 2016). The importance of optimizing ALS point density lays in its trade-offs against ALS swath width, and hence costs (Baltsavias, 1999). Liu et al. (2007) observed that density reduction influenced the accuracy of digital terrain models (DTM) due to the presence of empty space intervals between points. A reduction in DTM accuracy may affect the calculation of metrics describing ALS height (Ruiz et al., 2014; Singh et al., 2015), although it would be unlikely to affect metrics describing their dispersion, such as GC. Gobakken and Næsset (2008) assessed the effect of point density on biophysical stand properties, finding that maximum height was the least affected metric and suggesting to avoid metrics most affected by point density. No previous studies have yet determined how stand density and ALS scan densities affects GC estimates from ALS.

1.4 Objectives

The aim of the study is to evaluate the effects of plot size and ALS scan density on field and ALS-derived estimates of GC in the boreal forests of Finland. We developed a simple method for selecting the optimal plot size for determining the GC of tree size inequality from field data, and for predicting GC reliably using ALS metrics as auxiliary variables.

2. Material and Methods

2.1 Study Area and Field Data Collection

The study was carried out in a typical boreal managed forest located in Eastern Finland (62˚ 31′ N, 30˚ 10′ E). Scots Pine (*Pinus sylvestris* L.) is the dominant species which represents 73% of the total wood volume, while Norway spruce (*Picea abies* Karst.) represents 16%, and deciduous species 11% of the total wood volume (Valbuena et al., 2014). The main properties of the field data such as stand density (N), basal area (G) and quadratic mean diameter (QMD) are shown in **Table 1**. The field data were collected in May-June 2010 and consisted of 79 squared plots (henceforth "*original field plots*") of various dimensions (20×20, 25×25 or 30×30 m, the smaller ones being in denser stands). With the intention of representing the contrast between highly homogeneous even-aged areas and more heterogeneous forest structures (Valbuena et al. 2016), forest stands were determined using stratified random sampling, whereas plot locations were purposively selected. After choosing the sampled stands, plots were located within the stands at a representative location. The reason for doing this was to avoid plot locations at stand borders and the high cost and measuring effort required to record the location of all individual stems within the plot. The absolute positions of every individual tree with dbh > 4 cm and tree top height taller than 4 m were mapped using an approach combining ALS and field surveying methods suggested by Korpela et al. (2007). Before the field measurement, a map of individual tree positions was generated from high density ALS data (see below) using an individual tree detection (ITD) method (Packalen et al., 2013). Actual positions of trees defined by their longitude/latitude coordinates (X_{actual}, Y_{actual}) were verified in the field, while the location of trees not detected by the ITD method were measured relative to the ITD-derived ones (distances and bearings) and least-square adjusted (Korpela et al., 2007).

****approximate position of Table 1*****

2.2 Simulation of Circular Plots

Preliminary tasks for the simulation included transformations into relative coordinates, the correction of edge effects and a sensitivity analysis to determine the number of simulations needed. Then, within each *original field plot* we simulated circular plots at random positions. Circular plots were chosen on the assumption that tree competition is the same in all spatial directions. The radius of these *circular simulated plots* was increased in 1-m intervals, generating concentric circles up to 15 m-radius. Since the position of individual trees were available from the original field data, we could extract the trees located within each circular simulated plots, computing an estimation of GC based on tree dbh. Likewise, the position of individual ALS returns located within each simulated circular plots could be extracted, using them to compute ALS metrics commonly employed in area-based estimation methods.

2.2.1 Transformation to Relative Distances and Edge Correction

Transformation of absolute tree coordinates into relative coordinates requires procedures of plot rotation and translation (Matos, 2014). Since in the case of our study the edges of original field plots were coincident with the UTM grid, there was no need to carry out plot rotations. In plot translation absolute coordinates of original field plots were modified into relative distances, by assigning the origin of axes (0, 0) to the south-western corner of the original field plot. Absolute coordinates of south-western corner (X_{corner}, Y_{corner}) were subtracted from the absolute coordinates of each tree (X_{abs}, Y_{abs}) to get their relative coordinates (X_{rel}, Y_{rel}).

$$(X_{rel}, Y_{rel}) = (X_{abs}, Y_{abs}) - (X_{corner}, Y_{corner}) \quad (2)$$

Moreover, Pommerening and Stoyan (2006) showed that edge effects play an important role in spatial statistics. Because the immediate neighbour trees outside the boundary of the *original field plots* were not measured, ignoring them would result in biased statistical estimations. Thus, indices based on tree positions require an edge correction method to reduce this bias. We chose a periodic boundary edge correction method (Diggle, 2003), since Pommerening and Stoyan, (2006) found it to be superior to other alternatives. This method consisted of replicating the same spatial pattern measured in the field around the *original field plot* (**Fig. 1**). Concentric *circular simulated plots* randomly positioned at the edge of the *original field plots* therefore also included the trees positioned out of the boundaries of the *original field plots*.

****approximate position of Figure 1*****

2.2.2 Plot Simulation and Sensitivity Analysis

A pilot sensitivity analysis was done with the intention to identify the minimum number of simulations within an original field plot which can guaranteed a stable and robust outcome for the simulation. We selected the *original field plot* with highest GC, hence likely the one most sensible to changes among different simulations, and repeated the analysis for 10, 100, 500, 700, 1000, 1500 and 2000 simulations. A position (X_{sim}, Y_{sim}) was randomly simulated within the *original field plot*, and GC was calculated for each *circular simulated plot* (see below) and for each plot radius (s ; m) (1-m intervals from 1 to 15 m) (**Fig. 1**). As explained below, the standard error of the mean (SEM) of values obtained for GC at each radius were considered in order to fix the minimum number of simulations at which no considerable improvement was observed by adding further replications. After setting the necessary number of simulations to a fixed number k based on the pilot sensitivity analysis, the whole procedure was repeated for the remaining 78 original field plots. Relative and absolute positions of all simulations were recorded so that they could later be used for extracting their corresponding ALS returns as well.

2.3 *Gini Coefficient Estimation*

The target was to calculate sample estimations of the GC describing the size inequality of the tree community represented at each *original field plot*. Its estimation (**Eq. 1**) was repeated for every concentric *circular simulated plot* of radii 1-15 m, and for all the simulated positions (X_{sim}, Y_{sim}). For this purpose, basal area (g; m^2) was calculated for each individual stem. Differences in g were computed for each pair of trees within each circular simulated plot. GC is the average of absolute differences relative to their mean ($\bar{g}$) (see detailed descriptions of GC calculation in Lexerød and Eid (2006a) and Valbuena et al. (2013b)). The reason of using g instead of dbh was to increase the influence of larger trees (Solomon and Gove, 1999). The unbiased estimator by Glasser (1962) was employed because it is appropriate for an estimation based on a finite number of trees n located within each circular simulated plot (**Eq. 1**). The mean GC ($\overline{GC}$) and its corresponding SEM were computed for each radius (from 1 to 15 m), and for each of the *original field plots*. SEM is a measure for the accuracy of those means, accounting for the variability between the samples, according to the number of simulations k and their sample standard deviation (SD). R statistical software (R Development Core Team, 2016) was used for all these calculations and statistical analyses.

We constructed a graph comparing $\overline{GC}$ results for increasing plot size s for all *original field plots*. The GC value at *circular simulated plots* must necessarily approximate asymptotically to the value of GC for the entire *original field plot* as the radius of circular simulated plots increases (Matos, 2014). For this reason, the value of GC obtained by applying equation (1) to the original field plot was used as a reference (GC_{ref}). In order to make all the simulated GC values directly comparable, we calculated absolute GC differences (GC_{diff}) by subtracting simulated GC values from the GC_{ref}:

$$\overline{GC}_{diff} = |GC_{ref} - \overline{GC}| \quad (3)$$

This way, it was possible to analyse the difference of each simulated GC to its corresponding asymptotic value, allowing to set a common criterion to evaluate all simulations based on the stabilization of the estimated GC value (see below).

2.4 *Airborne Laser Scanning Data and Metric Computation*

ALS data was acquired on June 26, 2009 using ATM Gemini sensor (Optech, Canada) from 600-700 m above ground level with a 26° field of view. Scan swath was 320 m wide with a 55% side overlap between the strips. A high resolution dataset with 11.9 pulses·m^{-2} scan density was produced from a pulse rate of 125 kHz. Details about the processing of ALS data are described in Packalen et al. (2013). The last echoes were classified as ground and interpolated into a DTM (Axelsson, 2000). The spatial resolution of DTM was 0.5 m based on the scan density, and the height above ground of individual ALS returns was obtained by subtraction of the DTM height beneath each of them. Echoes lower than 0.1 m from ground level were eliminated, as they were considered to be reflected from ground.

Individual ALS returns of each circular simulated plot based on its absolute coordinates (X_{sim}, Y_{sim}) were clipped, and area-based ALS metrics were computed from their heights with the help of FUSION software (USDA Forest Service; McGaughey, 2015). ALS metrics are statistics and descriptors of the distribution of ALS heights observed within a given area, which are usually employed as auxiliary variables in ALS-assisted forest estimations (**Table 2**). Some of these metrics were common statistics as, for example, the mean (*Mean*) standard deviation (*StdDev*) or the skewness (*Skew*) of the distribution of heights above ground of ALS returns contained within each *circular simulated plot*. We also computed the percentiles of their distribution, such as the 25^{th} (*P25*), 50^{th} (*P50*) or 99^{th} (*P99*). In addition, we calculated the so-called canopy cover metrics (McGaughey, 2015), such as the proportion of returns backscattered from 0.1 m above the ground (*Cover*). Another metric included in FUSION was the canopy relief ratio (*CRR*), which is the difference between mean and minimum ALS return heights divided by a difference between maximum and minimum heights (Pike and Wilson, 1971).

****approximate position of Table 2*****

The effect of plot size in the relationship with GC was studied separately for each of these ALS metrics. For each radius, we gathered all the simulations carried out at all the *original field plots* and calculated all the ALS metrics listed in **Table 2**. They were used to calculate Pearson correlation coefficients (r) using all the combinations of field GC against each ALS metric. Then, we observed separately for each ALS metric the evolution of r when increasing the plot size s of the *circular simulated plots*. Since we were only interested in the capacity of the ALS metrics to explain variability in GC, regardless of whether their relationship was direct or indirect, we considered the absolute value of the correlation coefficient $|r|$ in the optimization, as explained below.

2.5 Basic Relationships

The plot size and spatial resolution at which an ALS-assisted estimation is carried out relates intrinsically to the sample size used in all calculations. Sample size affects the relationship between predictor and response, and therefore the accuracy of ALS estimation of any forest attributes (Gotway and Young, 2002; Mascaro et al., 2010; Næsset et al., 2015; Magnussen et al., 2016; Valbuena et al., 2016). In this context, sample size refers both to the number of trees used to calculate a given forest attribute, GC in this case, but also to the number of ALS returns involved in the computation of ALS metrics. The link between resolution and sample size is employed on the empirical densities of the datasets, i.e. stand density (N; trees·ha^{-1}) or ALS points density (d; points·m^{-2}) (Gobakken and Næsset, 2008; Motz et al., 2010; Jakubowski et al., 2013). Therefore, the effects of plot size and spatial resolution of the ALS estimated forest attributes also depend on N and d, and the combined effects of these two factors may explain why plot sizes suitable for field surveys may be found sub-optimal for ALS estimation (Næsset et al., 2015).

Hence, the relationship between the radius s of a circular plot and the number of trees (n) contained within is tied to the N at the location of the plot.

$$n = N\pi s^2 \tag{4}$$

This begs the question on whether the optimization method should search for an optimal plot radius (s^*; m) or an optimal sample size (n^*). In a forest environment of variable stand density N (**Table 1**), does the relationship between GC and ALS metrics depends on the plot size used, or on the number of trees surveyed? In order to research whether it makes a difference, we repeated the same procedure for both s^* and n^* optimization. In other words, we tested the results of optimization according to either plot radius or number of trees. In any of the cases, the relationship in **eq. (4)** assures that the methodology can be replicated for either dense or sparse forests, since s and n can always be deduced from one another by an empirical N.

Likewise, a similar relationship holds between the size of that same circular plot and the number of ALS returns backscattered from it, according to a given ALS scan density d. In this context of estimation using auxiliary variables, the scale concerns both to the size of the field plots and the spatial resolution of the pixel at which ALS metrics are calculated. Therefore, the number of ALS points (p) relates to the spatial resolution / plot size used (s) according to d :

$$p = d\pi s^2 \tag{5}$$

As before, the relationship in **eq. (5)** assures that the methodology can be replicated for any range of ALS scan densities, since s and p are trivially deducted from one another by an empirical d. As an overall conclusion, a given optimal plot size s^* necessarily implies optimal sample sizes as well, both n^* and p^*. Keeping these relationships in mind is key to demonstrating the validity of the optimization method for replication elsewhere according to the N and d which may occur at any other study cases, and therefore the method is equally valid for both dense and sparse forests and ALS surveys with low or high scan density.

2.6 Plot Size Optimization

To optimize the plot size which should be used for a reliable GC estimation, and thereby also the optimal spatial resolution for an estimation of GC from ALS datasets, we determined two criteria to be applied sequentially: (1) stabilization of GC as estimator of the population value from the field information itself, and (2) maximizing the GC variability explained by ALS metrics. Therefore, *Criterion I* considered the minimum plot radius at which the estimation of GC remained stable to further increases in plot size. *Criterion II* was set to optimize the ALS-assisted estimation, by observing changes in the correlation between the field GC and each ALS metric among the simulated plot radii.

Criterion I was implemented by observing the evolution of $\overline{GC}_{diff}$ for increasing radii at every original field plot. We set a maximum value of $\overline{GC}_{diff} = 0.05$ at which it was considered that the estimation of GC was stable and representative of the population, and, therefore, selected the minimum plot radius s as the smallest meeting the first criterion for all the 79 original field plots.

Criterion II consisted in maximizing the explained variance in the GC values when predicted from ALS metrics. To implement this criterion we combined all the GC/ALS metric pairs for all the simulations carried out at all the original field plots, and grouped them according to the different simulated radii. The optimal radius was set to be that one showing the maximum $|r|$ value for a given metric. To make an overall decision, we put the focus on those metrics showing higher correlations, and decided a range of optimal sizes accordingly (since the empirical maximum may differ for different ALS metrics). As a summary, the final optimal plot size s^* for a given metrics was:

$$s^* = \max|r| \;\middle|\; \overline{GC}_{diff} < 0.05 \tag{6}$$

2.7 Sample Size Optimization

For sample size optimization, seeking to deduct what is the minimum number of trees needed to obtain a reliable GC estimation, and the optimum for its ALS prediction, we applied the same two sequential criteria employed for plot size optimization (section 2.6). Therefore, the simulations were similar as before, but they increased the size of simulated circular plots according to the resulting number of trees n instead of plot radii. Thus, for implementing *Criterion I*, the evolution of $\overline{GC}_{diff}$ was observed for increasing number of trees n, also setting a maximum value of $\overline{GC}_{diff} = 0.05$. As before, we selected the minimum n as the smallest meeting *Criterion I* for all 79 original field plots. *Criterion II* also consisted in maximizing the absolute correlation between the GC values and each of the ALS metrics. New values of $|r|$ were obtained for increasing values of n, and the final optimal sample size (n^*) for each given ALS metric was then set as:

$$n^* = \max|r| \;\middle|\; \overline{GC}_{diff} < 0.05 \tag{7}$$

Finally, we compared which alternative, **eq. (6)** or **(7)**, would be more convenient for a practical plot size optimization, discussing the results obtained by either method.

2.8 Reduction of ALS Point Density

Once deducted an optimal spatial resolution s^*, we also investigated the effects of varying ALS scan density d. The original point density was reduced to 0.5, 0.75, 1, 3, 5, 7.5, and 10 points·m^{-2}. A common option to reduce point density is by moving a 1 m window and selecting random points from the point cloud to reach the desired point density (e.g., Magnussen et al. 2010). We calculated a correct thinning factor for each desired point density d (Ruiz et al., 2014), following the method detailed by Jakubowski et al. (2013) which incorporates routines included in *LAStools* (RapidLasso GmbH Inc.; Isenburg, 2016). New ALS metrics over each of the k simulated circular plot positions and their correlations against the GC values obtained from the field information were calculated, and the entire procedure was repeated for all the reduced densities. In a similar manner as it was done for s and n, the effects of varying ALS scan density were studied by observing the changes in $|r|$, i.e. the effects in the relationship between the GC of tree size inequality and the ALS metrics with more explanatory capacity towards this given forest attribute.

3. Results

3.1 Establishing the Number of Simulations

Figure 2 shows the results of sensitivity analysis carried out to select the minimum number of simulations that would yield a robust estimation of GC for increasing simulated plot radii. As expected, the GC value estimated from few simulations fluctuated considerably, and this fluctuation decreased as the number of simulations increased (**Fig. 2a**). The expected general trend toward the asymptotic value obtained by the entire population (GC_{ref}) was generally observed in **Fig. 2a**. Very little variation in GC estimates were observed when the number of simulations increased from 700. Similarly, the SEM decreased as the number of simulations increased (**Fig. 2b**), remaining virtually unchanged from 700 to 2000 simulations.

Consequently, we decided to carry out the analysis using $k = 700$ simulations of 15 concentric circular simulated plots located within each 79 original field plots.

****approximate position of Figure 2*****

3.2 Plot Size Optimization

Figure 3a shows the resulting $\overline{GC}_{diff}$ for each of the 79 original plots, and **Table 3** is a summary of these results which was used for establishing *Criterion I*, which set the minimum plot size that would provide a reliable GC estimation for the population. Circular simulated plots of small sizes provided GC estimates that differed considerably from the population values as considered by GC_{ref}. Nonetheless, once the estimation reached stabilization, an increase in the radius of a circular plot (and hence the sampling effort) would not necessarily imply a considerable change in the estimation of GC (**Fig. 3a**). Our results showed that only few of the original field plots (probably very homogeneous stands) obtained stable GC estimations from very small circular simulated plots (**Table 3**). On the other hand, for larger circular simulated plots the differences against the original field plots representing the population became negligible. We observed that stabilization of the GC estimation started beyond of simulated plot radius $s = 6$ m, from which all the original field plots fell within the $\overline{GC}_{diff} < 0.05$ limit. We therefore established that the smallest plot size required for a reliable GC estimation should be set at areas sizing around 113 m^2.

****approximate position of Figure 3*****

****approximate position of Table 3*****

With regards to *Criterion II*, the evolution of $|r|$ with increasing plot size was observed for all ALS metrics included in FUSION. Results showed that changes in the relationship between the field GC of tree sizes and metrics describing the distribution of ALS return followed some general trends and patterns. For this reason and for simplifying, we chose to show only few ALS metrics in **Fig. 4a**, which we considered representatives of the general trends observed. These ALS metrics were the described *P25, P50, P99, Skew, StdDev, Cover* and *CRR* (**Table 2**). **Fig. 4a** showed an erratic fluctuation for the values of $|r|$ obtained for plot sizes smaller than a radius $s = 5$ m, which was possibly caused by the instability observed in the GC estimation at smaller plot sizes (**Fig. 3**). For this reason, we shadowed this area in grey colour in **Fig. 4**, denoting that such small plot sizes were already dismissed under *Criterion I*. Once GC estimation reached stabilization, its correlation to ALS metrics often yielded a convex curve as plot size increased (**Fig. 4a**). Therefore, the optimal plot size was possible to determine via maximization of $|r|$. This tendency was more clearly marked for those ALS metrics showing higher values of $|r|$, i.e. more correlated to the GC of tree sizes (**eq. 2**), such as *Skew*, *Cover* or *CRR*. For other ALS metrics less related to GC, like return height percentiles (*P25, P50* or *P99*) or *StdDev*, this tendency was less marked (**Fig. 4a**). For the optimization of plot size, we selected those metrics showing highest correlation against GC, since in practice they would be those more involved in its estimation. **Table 4** shows that the maximum $|r|$ for ALS metrics *Skew*, *Cover* or *CRR* ranged $s^* = 9$-12 m plot radius (the quality of histograms and scatterplots between variables involved can be checked in the **Supplementary Material**). It can be observed in **Fig. 4a** that beyond a circular simulated plot of 12 m the correlation showed a decreasing trend for most ALS metrics. Also, local maxima may be found for some ALS metrics for very small plot sizes, which is probably an artefact due to the above-mention instability in GC estimation at very small plot sizes (**Fig. 3**). This proved the necessity of imposing *Criterion I* as a prior step to correlation maximization. As a conclusion, under the established combined *Criteria I* and *II*, we determined that any plot radius $s < 6$ m (113 m^2 area) should be avoided (denoted by grey colour in **Fig. 4a**), and the optimal plot size for an ALS-assisted estimation of GC must be carried out using scales sizing 250-450 m^2, which concerns to both the size of the field plot and the pixel of the grid employed for ALS estimation.

****approximate position of Figure 4*****

****approximate position of Table 4*****

3.3 Sample Size Optimization (Stand Density Effect)

On the other hand, **Figure 3b** shows the evolution of $\overline{GC}_{diff}$ for increasing sample sizes (number of trees n) at each of the 79 original field plots. It is worth mentioning the **Figs. 3a** and **3b** relate to one another according to **eq. (4**). As a consequence, a similar tendency can be found for both of them. **Table 3** expresses the number of trees that correspond on average to a given sample size. Therefore, the minimum value obtained for *Criterion I* in plot size optimization, $s = 6$, corresponds to stating that a minimum number of $n = 15$ trees are required for a stable GC estimation (shaded area in **Fig. 4b**). We nevertheless further postulated that this minimum number of trees may be dependent on the heterogeneity of the forest itself, being possibly larger in the presence of higher inequality of tree sizes. This presumption was demonstrably true, as it can be observed in a scatterplot comparing the minimum number of trees required for a stable GC estimation at each of the 79 original plots against their overall value of tree size inequality observed (GC_{ref}; **Fig. 5**). Such relationship was not so straightforward if *Criterion I* was imposed using s instead (results not shown), which demonstrates the effect of varying forest stand density N. Hence, obtaining a stable GC estimation is more dependent of measuring a minimum number of trees than imposing a given size for the field plot used.

****approximate position of Figure 5*****

The case for *Criterion II* was different, as it can be deducted when observing the same ALS metrics employed to optimize s – *P25, P50, P99, Skew, StdDev, Cover* and *CRR* –, but trying to optimize n instead (**Fig. 4b**). Again, a similar tendency can be found since **Figs.4a-b** are also related by **eq. (4**). Results were therefore very similar whether optimization was carried out according to plot size (**eq. 6**) or sample size (**eq. 7**). The values of $|r|$ also followed a convex curve when increasing the number of trees measured, and an optimal sample size n^* could be reliably determined via $|r|$ maximization. Our results showed that a number of trees approximately ranging $n^* =$ 30-60 (**Table 4**) should be involved in the computation of GC, in order to maximize the efficiency of its estimation using ALS. Since the value of $|r|$ involves both the field GC and the ALS metrics, its changes are determined by both N and d (**eqs. 4-5**), and both may cause a change in the correlation between the two variables.

3.4 Effect of Point Density on the Relationship of GC

According to the previous results, we set the optimal plot size to $s^* = 9$ m in order to further analyse the possible effects due to varying scan density. Among all the ALS metrics (**Table 2**), we selected those same ones employed previously – *P25, P50, P99, Skew, StdDev, Cover* and *CRR* – to allow direct comparison. **Fig. 6** shows the evolution in $|r|$ for increasing ALS point density d. No considerable changes were observed in the correlation between the field GC and the ALS metrics, which suggests that d has no major effects on their relationship. However, a decreasing trend in $|r|$ could generally be observed when point densities decreased below $d < 3$ points·m^{-2} (**Fig. 6**). Overall, these results therefore suggest that the relationship between GC and ALS metrics is mainly dependent on the plot size employed, and rather independent of stand density and ALS scan density

****approximate position of Figure 6*****

4. Discussion

In this study we evaluated the effects of plot size and sample size on the GC of tree size inequality, and on its practical estimation using remote sensing methods based on ALS. Sample size refers to the number of individual elements (trees or ALS returns) included within a given sample area, which is therefore determined by the spatial resolution employed for evaluating a given forest attribute. We also analysed the effects of ALS scan density and, overall, we observed that plot size had greater effects on the relationship between GC and ALS metrics than either of the other two criteria considered. The motivation for studying these effects is grounded on the fact that inappropriate plot sizes may provide unreliable estimates and lead to sub-optimal forest management decisions (Eid, 2000; Mauro et al., 2010). Valbuena et al. (2013a) pointed out that the estimation of GC is affected by the area at which it is evaluated. Results in **Fig. 3** illustrate how the $\overline{GC}_{diff}$ decreases when increasing the size of circular plots and, and hence their corresponding sample size. $\overline{GC}_{diff}$ values markedly dropped for smaller plot radii and sample sizes. This decrease smooths from bigger sizes, which indicates stabilization of the estimation (*Criterion I*). **Fig. 2a** also shows an example of this tendency to asymptotically approach the population value, which was also observed by George (2003), Barbeito et al. (2009), or Matos (2014). Based on *Criterion I* ($\overline{GC}_{diff} < 0.05$), the circular plot should be large enough ($s \geq 6$ m) to have minimum sample size of $n \geq 15$ trees (**Fig. 3**). Although the minimum plot size also depends on the stand density of an area, **eq. (4)** can be used to adjust the method to any forest areas, whether sparsely or densely forested. This conclusion may therefore be partly extended to other forest types, as it can be for example deduced (via **eq. 4**) that minimum radius of $s \geq 12$ m would be needed in sparsely forested area of only 300 stems·ha^{-1} (Lombardi et al., 2015). **Eq. (4)** therefore brings generality to the method, since plot sizes may hence be tailored to forest areas of differing stand densities.

In this article we also postulated that maximizing the explained variability between the GC estimated from the field and ALS metrics could be a valid criterion to optimize the reliability of ALS-assisted estimations of GC (*Criterion II*). Results in **Fig. 4a** showed that our presumption was correct, since the $|r|$ values between GC and most ALS metrics, especially the most correlated ones, followed a convex curve with a maximum that could be searched to reach an optimal plot size / spatial resolution for the estimation. On the other hand, once the GC reached some stabilization, the correlation between them remains largely unchanged. Therefore, a lower plot size limit is to be imposed to avoid local minima that could appear as an artefact of the unstable estimation of GC at low sample sizes. We shaded this area in grey colour in **Fig. 4 (a, b)**, denoting the area that was already dismissed as a result of *Criterion I* (**Fig. 3**; George, 2003). In larger plots the sample size was more representative of the total population. Combining both criteria, we found in our study area that an optimal circular plot radius of $s^* = 9\text{-}12$ m, which corresponds to a spatial resolution of sampling units sizing 250-450 m^2 (**Fig. 4a**), would be suitable for ALS-assisted GC estimation. Since plot size and sample size are interdependent (**eq. 4**), this result may be suitable for any area with a similar average number of trees ($N \cong 1300$ stems·ha^{-1}; **Table 2**). According to these results, therefore, most forest datasets commonly acquired in operational inventories would be acceptable for an ALS-assisted estimation of the GC of tree sizes. Lombardi et al. (2015) deduced a larger optimal plot radius $s^* = 13\text{-}15$ m for other forest attributes, most likely due to lower N in the forest areas considered. For studies dealing with differing plot sizes, one possibility could be to upscale GC to a common plot size (Kent and Coker, 1992; Magnussen et al., 2016).

Some of the reflexions raised in this article affect all other types of forest attributes and remotely sensed auxiliary variables that may be used in forest estimations (Jelinski and Wu, 1996). However, different forest attributes are differently affected by varying plot sizes

(Chytrý and Otýpková, 2003). Some forest variables such as stand density or biomass would show an averaging effect as plot size increases (Jelinski and Wu, 1996; Gotway and Young, 2002; Ruiz et al., 2014), which in turn derives in improved model efficiency when using larger scales in remote sensing estimations (Næsset et al., 2015; Mauro et al., 2016). But there is a trade-off between model accuracy and spatial resolution, and root mean squared errors increase from10-15% for 1000-4000 m^2 to 20-25% for 200-250 m^2 (Næsset, 2002, 2004, 2007). However, this averaging effect is not applicable to forest attributes describing structural diversity and heterogeneity (Coomes and Allen, 2007). In fact, many variables necessarily augment when the plot size increases, for instance species richness and diversity (e.g., Humphrey et al. 2000; Otypková and Chytry, 2006; Kallimanis et al., 2008; Fibich et al., 2016) as traditionally assessed through rarefaction (Kent and Coker, 1992). A similar effect can be observed in other measures of forest heterogeneity (Barbeito et al., 2009; Motz et al., 2010; McRoberts et al., 2012), and thus in the GC (Valbuena et al., 2013a, Matos, 2014), since increasing the size of a plot increases the probability of finding an additional differently-sized tree (Chen and Crawford, 2012; Valbuena et al., 2012). This is why estimated GC values in **Fig. 3** asymptotically approach the value of the larger original field plot (George, 2003; Matos, 2014), which is never exceeded. Imposing a criterion defining which of the plausible plot sizes should be used is therefore not a trivial question to tackle. Matos (2014) employed a number of different criteria based on the field information only – stabilization of the estimate, stabilization of certainty of the estimate and convergence with GC_{ref} –, none of them resulting fully satisfactory and definitive as they all ultimately rely on a subjective assumption (Cressie, 1993). For this reason, in this article we approached the question of plot size from the viewpoint of its practical estimation using ALS remote sensing. The convex curves obtained in **Fig. 4a** proved this approach to be highly beneficial, since maximization of correlation $|r|$ between GC values and selected ALS predictors provides with a more objective method for determining the optimal plot size for the assessment of GC of tree size inequality. Still, due to the very high uncertainty observed in the estimation of GC when using very small plot sizes (**Fig. 3b**; Smith, 1938; Lombardi et al., 2015), we deducted that a criterion avoiding great divergence with GC_{ref} may be imposed as a prior step to maximization (Motz et al. (2010) referred to it as minimum grid spacing). Further research could focus on modelling GC from ALS metrics and investigate how the interaction among many ALS metrics in a same model may play a relevant role in the optimization of plot size and spatial resolution.

The analyses carried out with reduced point densities revealed that lowering point density barely affects the correlation between GC and ALS metrics, unless using a very sparse scan density $d < 3$ points·m^{-2}. Previous studies such as Maltamo et al. (2006), Ruiz et al. (2014) or Singh et al. (2015) also indicate that reducing the point density is not affecting the accuracy of volume prediction and demonstrate that the effects of varying scan densities can be eluded in practical applications. It must be taken into account, however, that the DTM used in this study was based on original point density, and the errors in DTM determination at sparser densities (Liu et al., 2007; Ruiz et al., 2014) may induce to further uncertainty, although this presumably has a lesser effect on those metrics most related to GC. Furthermore, since ALS datasets from national programmes are currently surveying entire countries at densities typically between 0.5-1 points·m^{-2} (Artuso et al., 2003), it must also be pointed out the relevance of results in **Fig. 6** which render most of these nation-wide ALS datasets unsuitable for reliably estimating GC (Kandare et al. 2016). In line with results in Valbuena et al. (2017), who postulated that the low densities incur in critical omission of understorey development, our results demonstrate that indeed there is a need for increasing point densities up to $d = 3$ points·m^{-2}. This result is very concurrent with those obtained by Ruiz et al. (2014) and Watt et al. (2014) for different forest attributes in different stand types, and therefore the case seems clear that ALS datasets obtained for forest applications should reach this minimum density requirement.

5. Conclusion

In this study we studied how changing spatial resolution can affect the relationship between GC and ALS metrics. We used three criteria for optimization: plot size, stand density and ALS scan density. The effects of stand and scan densities are intimately interrelated to plot size, since they together determine the sample size employed in calculations. Amongst those three criteria, we found plot size to predominantly affect the relationship between GC and ALS metrics.

We observed that the estimation of GC is strongly affected by the size of the forest plot surveyed. Very small sample size and plot radii are more sensitive to GC variations, unrepresentative of the total population, producing unstable and unreliable GC estimations. The GC estimation stabilizes as the size of plots and samples increases, as larger plots contain a more appropriate number of observations (sample size) representing the population. We determined that, in a boreal managed forest, a minimum number of 15 trees ought to be measured for a reliable GC estimation, regardless of the stand density present at each forest stand.

We developed a method for plot size optimization based on a combination of two criteria: (1) imposing a minimum of number of 15 trees measured, and (2) maximizing the absolute correlation between field GC and ALS metrics. The plot level correlation between ALS metrics and field GC showed a convex tendency for increasing plot sizes. Our results showed that 9-12 m-radius plots produced the maximum correlation thus they are suitable for ALS-assisted GC estimation. Basic relationships between plot size and sample size may be used to accommodate the method to forested environments of varying stand densities.

With regards to the effects of ALS scan density, we observed that it can barely have any effects unless lowered under 3 points m^{-2}. This however may be relevant for the practical application of low-density national datasets, and therefore we would recommend increasing their scan densities with the intention to render nation-wide datasets useful for studying forest heterogeneity.

Acknowledgements

This research was carried out in the framework of the project LORENZLIDAR: Classification of Forest Structural Types with LiDAR remote sensing applied to study tree size-density scaling theories, funded by a Marie S. Curie Individual Fellowship H202-MSCA-IF-2014 (project 658180). Syed Adnan's PhD is funded by the National University of Sciences and Technology (NUST), Pakistan under FDP 2014-15.

References

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, 7-7.

Artuso, R., Bovet, S., Streilein, A. 2003. Practical Methods for the Verification of Countrywide Terrain and Surface Models. ISPRS WG III/3 Workshop. 3-D reconstruction from airborne laser scanner and InSAR data. Dresden, Germany 8-10 October 2003. In: The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXIV, part 3/WG13.

Axelsson, P., 2000. DEM generation from laser scanner data using adaptive TIN models. International Archives of Photogrammetry and Remote Sensing 33, 111-118.

Balsa-Barreiro, J., Lerma, J.L., 2014. Empirical study of variation in lidar point density over different land covers. Int. J. Remote Sens. 35, 3372-3383.

Baltsavias, E.P., 1999. Airborne laser scanning: basic relations and formulas. ISPRS Journal of photogrammetry and remote sensing 54, 199-214.

Barbeito, I., Cañellas, I., Montes, F., 2009. Evaluating the behaviour of vertical structure indices in Scots pine forests. Ann. For. Sci. 66, 710-710.

Bollandsås, O.M., Næsset, E., 2007. Estimating percentile-based diameter distributions in uneven-sized Norway spruce stands using airborne laser scanner data. Scand. J. For. Res. 22, 33-47.

Bourdier T., Cordonnier T., Kunstler G., Piedallu C., Lagarrigues G., Courbaud B. 2016. Tree size inequality reduces forest productivity: an analysis combining inventory data for ten european species and a light competition model. PLoS ONE 11(3), e0151852.

Buongiorno, J., Dahir, S., Lu, H., Lin, C., 1994. Tree Size Diversity and Economic Returns in Uneven-Aged Forest Stands. For. Sci. 40, 83-103.

Chen, Z., Crawford, C.A.G., 2012. The role of geographic scale in testing the income inequality hypothesis as an explanation of health disparities. Soc. Sci. Med. 75, 1022-1031.

Chytrý, M., Otýpková, Z., 2003. Plot sizes used for phytosociological sampling of European vegetation. Journal of Vegetation Science 14, 563-570.

Coomes, D.A. and Allen, R.B., 2007. Mortality and tree-size distributions in natural mixed-age forests. Journal of Ecology, 95(1), pp.27-40.

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.

Cressie, N., 1993. Aggregation in Geostatistical Problems, in Soares, A. (Ed.), Geostatistics Troia'92: Volume 1. Springer Netherlands, Dordrecht, pp. 25-36.

David, H.A., Mishriky, R.S., 1968. Order Statistics for Discrete Populations and for Grouped Samples. Journal of the American Statistical Association 63, 1390-1398.

Diggle, P.J., 2003. Statistical Analysis of Spatial Point Patterns, 2nd ed. Arnold, London.

Duduman, G., 2009. An ecological approach for establishing the allowable cut in forests where single tree selection system is applied. Editura Universităţii Suceava.[In Romanian] .

Eid, T., 2000. Use of uncertain inventory data in forestry scenario models and consequential incorrect harvest decisions. Silva Fenn. 34, 89-100.

Fibich, P., Lepš, J., Novotný, V., Klimeš, P., Těšitel, J., Molem, K., Damas, K., Weiblen, G.D., 2016. Spatial patterns of tree species distribution in New Guinea primary and secondary lowland rain forest. Journal of Vegetation Science 27, 328-339.

George, D., 2003. The Small-Sample Bias of the Gini Coefficient: Results and Implications for Empirical Research. Review of Economics and Statistics 85, 226-234.

Gini, C., 1921. Measurement of inequality of incomes. The Economic Journal 31, 124-126.

Glasser, G.J., 1962. Variance formulas for the mean difference and coefficient of concentration. Journal of the American Statistical Association 57, 648-654.

Gobakken, T., Næsset, E., Nelson, R., 2006. Developing regional forest inventory procedures based on scanning LiDAR. Proceedings of the International Conference Silvilaser 2006, Matsuyama, Japan. Japan Society of Forest Planning, Forestry and Forest Products Research Institute and Ehime University. pp. 99-104.

Gobakken, T., 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.

Gotway, C.A., Young, L.J., 2002. Combining incompatible spatial data. Journal of the American Statistical Association 97, 632-648.

Hannan, M.T., 1971. Aggregation and Disaggregation in Sociology. Lexington Books.

Humphrey, J., Newton, A., Peace, A., Holden, E., 2000. The importance of conifer plantations in northern Britain as a habitat for native fungi. Biol. Conserv. 96, 241-252.

Hvistendahl, M., 2014. While emerging economies boom, equality goes bust. Science, 344, 832-835

Isenburg, M., 2016. “LAStools–efficient tools for LiDAR processing”. (Version 160921, academic). Retrieved from http://rapidlasso.com/LAStools.

Jakubowski, M.K., Guo, Q., Kelly, M., 2013. Tradeoffs between lidar pulse density and forest measurement accuracy. Remote Sens. Environ. 130, 245-253.

Jelinski, D., Wu, J., 1996. The modifiable areal unit problem and implications for landscape ecology. Landscape Ecol. 11, 129-140.

Kallimanis, A.S., Halley, J.M., Vokou, D. & Sgardelis, S.P., 2008. The scale of analysis determines the spatial pattern of woody species in the Mediterranean environment. Plant Ecology 196: 143–151.

Kandare, K., Ørka, H.O., Chan, J.C.-W., 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, pp. 337-359.

Kent, M., Coker, P., 1992. Vegetation description and analysis: a practical approach. Chichester: Wiley.

Knox, R.G., Peet, R.K., Christensen, N.L., 1989. Population dynamics in loblolly pine stands: changes in skewness and size inequality. Ecology 70, 1153-1167.

Korpela, I., Tuomola, T., Välimäki, E., 2007. Mapping forest plots: an efficient method combining photogrammetry and field triangulation. Silva Fenn. 41.

Lähde E., Laiho O., Norokorpi Y., 1999. Diversity-oriented silviculture in the boreal zone of Europe. Forest Ecology and Management 118: 223–243.

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, 1835–1847.

Lexerød, N., Eid, T., 2006a. An evaluation of different diameter diversity indices based on criteria related to forest management planning. For. Ecol. Manage. 222, 17-28.

Lexerød, N., Eid, T., 2006b. Assessing suitability for selective cutting using a stand level index. Forest Ecology and Management, 237(1–3), pp. 503-512.

Liu, X., Zhang, Z., Peterson, J., Chandra, S., 2007. The effect of LiDAR data density on DEM accuracy. Proceedings of the International Congress on Modelling and Simulation (MODSIM07) 2007, Modelling and Simulation Society of Australia and New Zealand Inc., pp. 1363-1369.

Lombardi, F., Marchetti, M., Corona, P., Merlini, P., Chirici, G., Tognetti, R., Burrascano, S., Alivernini, A., Puletti, N., 2015. Quantifying the effect of sampling plot size on the estimation of structural indicators in old-growth forest stands. For. Ecol. Manage. 346, 89-97.

Lundqvist, L., 1994. Growth and Competition in Partially Cut Sub-Alpine Norway Spruce Forests in Northern Sweden. For. Ecol. Manage. 65, 115-122.

Magnussen, S., Mandallaz, D., Lanz, A., Ginzler, C., Næsset, E., Gobakken, T., 2016. Scale effects in survey estimates of proportions and quantiles of per unit area attributes. For. Ecol. Manage. 364, 122-129.

Maltamo, M., Eerikäinen, K., Packalén, P., Hyyppä, J., 2006. Estimation of stem volume using laser scanning-based canopy height metrics. Forestry 79, 217-229.

Mascaro, J., Detto, M., Asner, G.P., Muller-Landau, H.C., 2011. Evaluating uncertainty in mapping forest carbon with airborne LiDAR. Remote Sens. Environ. 115, 3770-3774.

Matos, A., 2014. Effect of scale factor in estimation of Gini coefficient. Master's thesis. University of Eastern Finland.
http://epublications.uef.fi/pub/urn_nbn_fi_uef-20140718/index_en.html
http://www.oppi.uef.fi/opk/video/europeanforestry/a_matos_seminar.mp4
(accessed August 2016)

Mauro, F., Valbuena, R., Manzanera, J., García-Abril, A., 2010. Influence of Global Navigation Satellite System errors in positioning inventory plots for tree-height distribution studies Canadian journal of forest research 41, 11-23.

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

McElhinny, C., Gibbons, P., Brack, C., Bauhus, J., 2005. Forest and woodland stand structural complexity: Its definition and measurement. For. Ecol. Manage. 218, 1-24.

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

McRoberts, R.E., Winter, S., Chirici, G., LaPoint, E., 2012. Assessing Forest Naturalness. For. Sci. 58, 294-309.

Motz, K., Sterba, H., Pommerening, A., 2010. Sampling measures of tree diversity. For. Ecol. Manage. 260, 1985-1996.

Næsset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sens. Environ. 80, 88-99.

Næsset, E., 2004. Accuracy of forest inventory using airborne laser scanning: evaluating the first Nordic full-scale operational project. Scand. J. For. Res. 19, 554-557.

Næsset, E., 2007. Airborne laser scanning as a method in operational forest inventory: Status of accuracy assessments accomplished in Scandinavia. Scand. J. For. Res. 22, 433-442.

Næsset, E., Bollandsås, O.M., Gobakken, T., Solberg, S., McRoberts, R.E., 2015. The effects of field plot size on model-assisted estimation of aboveground biomass change using multitemporal interferometric SAR and airborne laser scanning data. Remote Sens. Environ. 168, 252-264.

Neumann, M., Starlinger, F., 2001. The significance of different indices for stand structure and diversity in forests. For. Ecol. Manage. 145, 91-106.

O'Hara, K.L., Gersonde, R.F., 2004. Stocking control concepts in uneven-aged silviculture. Forestry 77, 131-143.

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.

Oliver, C.D., Larson, B.C., 1990. Forest Stand Dynamics. McGraw-Hill, Inc.

Otypková, Z., Chytry, M., 2006. Effects of plot size on the ordination of vegetation samples. Journal of Vegetation Science 17, 465-472.

Packalen, P., Vauhkonen, J., Kallio, E., Peuhkurinen, J., Pitkänen, J., Pippuri, I., Strunk, J., Maltamo, M., 2013. Predicting the spatial pattern of trees by airborne laser scanning. Int. J. Remote Sens. 34, 5154-5165.

Pike, R.J., Wilson, S.E., 1971. Elevation-relief ratio, hypsometric integral, and geomorphic area-altitude analysis. Geological Society of America Bulletin 82, 1079-1084.

Pommerening, A., 2002. Approaches to quantifying forest structures. Forestry 75, 305-324.

Pommerening, A., Stoyan, D., 2006. Edge-correction needs in estimating indices of spatial forest structure. Canadian Journal of Forest Research 36, 1723-1739.

Pukkala, T., Laiho, O., Lähde, E., 2016. Continuous cover management reduces wind damage. For. Ecol. Manage. 372, 120-127.

R Development Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Robles, A., Rodríguez, M.A., Alvarez-Taboada, F. 2016. Characterization of wildland-urban interfaces using LiDAR data to estimate the risk of wildfire damage. Revista de Teledeteccion, 45: 57-69.

Ruiz, L.A., Hermosilla, T., Mauro, F., Godino, M. 2014. Analysis of the influence of plot size and LiDAR density on forest structure attribute estimates. Forests, 5 (5): 936-951

Singh, K.K., Chen, G., McCarter, J.B., Meentemeyer, R.K. 2015. Effects of LiDAR point density and landscape context on estimates of urban forest biomass. ISPRS Journal of Photogrammetry and Remote Sensing, 101: 310-322

Smith, H.F., 1938. An empirical law describing heterogeneity in the yields of agricultural crops. The Journal of Agricultural Science 28, 1-23.

Staudhammer, C.L., LeMay, V.M., 2001. Introduction and evaluation of possible indices of stand structural diversity. Canadian journal of forest research 31, 1105-1115.

Sterba, H., Ledermann, T., 2006. Inventory and modelling for forests in transition from even-aged to uneven-aged management. For. Ecol. Manage. 224, 278-285.

Upton, G., Fingleton, B., 1985. Spatial Data Analysis by Example. Volume 1: Point Pattern and Quantitative Data. John Wiley & Sons Ltd.

Valbuena, R., Packalén, P., Martín-Fernández, S., Maltamo, M., 2012. Diversity and equitability ordering profiles applied to study forest structure. For. Ecol. Manage. 276, 185-195.

Valbuena, R., Packalen, P., Mehtätalo, L., García-Abril, A., Maltamo, M., 2013a. Characterizing forest structural types and shelterwood dynamics from Lorenz-based indicators predicted by airborne laser scanning. Canadian Journal of Forest Research 43, 1063-1074.

Valbuena, R., Maltamo, M., Martín-Fernández, S., Packalen, P., pascual, C. and Nabuurs, G., 2013b. Patterns of covariance between airborne laser scanning metrics and Lorenz curve descriptors of tree size inequality. Canadian Journal of Remote Sensing, 39, pp. S18-S31.

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.

Valbuena, R., Eerikainen, K., Packalen, P., Maltamo, M., 2016. Gini coefficient predictions from airborne lidar remote sensing display the effect of management intensity on forest structure. Ecol. Ind. 60, 574-585.

Valbuena, R., Maltamo, M., Mehtätalo, L., Packalen, P., 2017. Key structural features of boreal forests may be detected directly using l-moments from airborne lidar data. Remote Sensing of Environment 194: 437-446.

Vihervaara P., Mononen L., Auvinen A.P., 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.

Watt, M.S., Meredith, A., Watt, P., Gunn, A., 2014. The influence of LiDAR pulse density on the precision of inventory metrics in young unthinned Douglas-fir stands during initial and subsequent LiDAR acquisitions. New Zealand Journal of Forestry Science, 44 (1), 9 p.

Weiner, J., 1990. Asymmetric competition in plant populations. Trends in Ecology & Evolution, 5(11), pp. 360-364.

Weiner, J., Solbrig, O.T., 1984. The meaning and measurement of size hierarchies in plant populations. Oecologia 61, 334-336.

Weiner J., Thomas S., 1986. Size variability and competition in plant monocultures. Oikos 47: 211–222.

Whittaker, R.H., 1972. Evolution and measurement of species diversity. Taxon , 213-251.

Zheng, X., Xia, T., Yang, X., Yuan, T., Hu, Y., 2013. The Land Gini Coefficient and Its Application for Land Use Structure Analysis in China. PLoS ONE 8, 1-10.

Table Titles

Table 1. Properties of the study area.

Table 2. Summary of ALS metrics computed with FUSION and used in this research (McGaughey, 2015).

Table 3. For each radii, proportion of the total number of original field plots within the $\overline{GC}_{diff} < 0.05$ limit (*Criterion I*), and average number of trees contained within those plots.

Table 4. Maximum absolute correlation between field GC and ALS predictors (*Criterion II*). See **Table 2** for description of ALS metrics.

Figure Captions

Figure 1. Reproduction of tree positions (dots) within an original field plot (red rectangle) surrounded by edge correction i.e. translation method (i.e. periodic boundary), and a sample of 10 random realizations of simulated concentric circular plots with radii sizing 1-3 m (for simplicity). Axes show both absolute (above) and relative (below) coordinates (respectively X_{abs} , Y_{abs} and X_{rel} , Y_{rel} in **Eq. 2**).

Figure 2. Results of sensitivity analysis to select minimum numbers of simulations. Evolution for increasing radii of (a) mean $\widehat{GC}$ values and (b) their standard errors for $k =$ 10-2000 simulations.

Figure 3. *Criterion I*. Asymptotic representation showing the evolution of $\overline{GC}_{diff}$ (at each of the 79 original field plots) for increasing (a) plot sizes $s =$ 1-15 m radius (corresponding area also shown in upper axis) (b) and sample size $n =$ 1-50 number of trees (shortened to enhance visualization).

Figure 4. *Criterion II*. Absolute of correlation $|r|$ between GC values and selected ALS predictors (see legend, and explanations of ALS metrics in **Table 1** and section 2.4) for increasing (a) plot size s = 1-15 m radius (corresponding area also shown in upper axis) (b) and sample size n = 1-90 number of trees.

Figure 5. Minimum number of trees (sample size) to reach GC stabilisation in relation to the reference GC value obtained from the original field plot (GC_{ref}).

Figure 6. Changes due to varying ALS scan densities in the absolute of correlation $|r|$ between GC values and ALS predictors. See explanations of ALS metrics in **Table 1** (section 2.4).

Supplementary Materials

Supplementary Figure 1. Histograms showing the distribution of the response variable – $\overline{GC}$ (vertical bars) – and the predictor variables – *Skewness, Cover, CRR, P99, StdDev, P50* and *P25* (horizontal bars) –. The resulting scatterplots between each response-predictor pair are also shown. For simplicity, only results for the optimal plot radius s^{*} = 9 m are shown.

Title:

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

Authors:

Syed Adnan *(1), Matti Maltamo (1), David Coomes (2), Rubén Valbuena (2)

Affiliations:

(1) University of Eastern Finland. Faculty of Forest Sciences. PO Box 111 Joensuu, Finland; adnan@uef.fi; matti.maltamo@uef.fi.

(2) University of Cambridge, Department of Plant Sciences. Forest Ecology and Conservation. Downing Street, CB2 3EA Cambridge, UK. dac18@cam.ac.uk; rv314@cam.ac.uk.

*Corresponding author.

Abstract

The estimation of Gini Coefficient (GC) of tree sizes using airborne laser scanning (ALS) can provide maps of forest structure across the landscape, which can support sustainable forest management. A challenge arises in determining the optimal spatial resolution that maximizes the stability and precision of GC estimates, which in turn depends upon stand density or ALS scan density. By subsampling different plot sizes within large field plots, we evaluated the optimal spatial resolution by observing changes in GC estimation and in its correlation with ALS metrics. We found that plot size had greater effects than either stand density or ALS scan density in the relationship between GC and ALS metrics. Uncertainty in GC estimates fell as plot size increased. Correlation with ALS metrics showed convex curves with maxima at 250-450 m^2, which thus was considered the optimal plot size / spatial resolution. By thinning the density of ALS point cloud, we deduced that at least 3 points·m^{-2} are needed for reliable GC estimates. Many nationwide ALS scan densities are sparser than this, which may be unreliable for GC estimation. Ours is a simple approach for evaluating the optimal spatial resolution in remote sensing estimation of any forest attribute.

Key words

structural heterogeneity; spatial resolution optimization; sample size optimization; forest structure; LiDAR

1. Introduction

1.1 The Gini Coefficient as an Indicator of Forest Structural Heterogeneity

Forest structural characteristics are widely used in the development of sustainable management plans designed to protect habitats while carrying out forestry operations (Upton and Fingleton, 1985; Pommerening, 2002; Motz et al., 2010; Vihervaara et al., 2015; Valbuena et al., 2016). Management can be designed to emulate natural dynamics (Oliver and Larson, 1990; Buongiorno et al., 1994; Lähde et al., 1999; Pukkala et al., 2016), by studying how silvicultural operations affects forest structure locally (Humphrey et al., 2000; Valbuena et al., 2013a; Robles et al., 2016).

Forest structure is often characterized by stem diameter distributions (O'Hara and Gersonde, 2004; McElhinny et al., 2005). If a single concise indicator of size inequality is desired, there are many available, including Shannon or Simpson indices (Neumann and Starlinger, 2001; Sterba and Ledermann, 2006; O'Hara et al., 2007; Lei et al., 2009) or variance-based metrics (Staudhammer and LeMay, 2001). Recent research has highlighted the effectiveness of the Gini coefficient (i.e. GC, Gini, 1921) for assessing the structural diversity (Lexerød and Eid, 2006a; O'hara et al. 2007; Duduman, 2009; Valbuena et al., 2012, 2013a). Originally developed for evaluating inequality in income distributions (e.g., Hvistendahl, 2014), GC has been applied to a variety of fields, such as healthcare (Asada, 2005) or land use (Zheng et al., 2013). In plant sciences, it has been employed to evaluate size inequality (Weiner and Solbrig, 1984). It has also been applied to forest ecosystems (Weiner and Thomas, 1986), to quantify structural diversity (Knox and Peet, 1989), analyse competition (Lundqvist, 1994; Cordonnier and Kunstler, 2015), or successional stages (Valbuena et al., 2013a). Comparative studies indicate that GC is the best index for characterizing diameter distributions, providing a logical ranking of different stand types (Lexerød and Eid, 2006a; Valbuena et al., 2012), so that forest may be stratified according to their structure (Bollandsås and Næsset, 2007). It can also be used to observe the effects of different management regimes (Bourdier et al., 2016; Pukkala et al. 2016; Valbuena et al., 2016). For these reasons, estimation of GC is the focus of this article.

When used in forest science, GC evaluates size inequality of trees growing in a vicinity (Weiner, 1990). For a patch of forest containing n trees, within which the i^{th} and j^{th} tree have basal areas of g_i and g_j respectively, GC is calculated as (Glasser, 1962):

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

Therefore, GC is a statistical measure of relative dispersion, which is equivalent to half of the relative mean absolute difference (Valbuena et al, 2017: appendix A3), and it ranges between 0-1, zero representing perfect equality and one being maximum inequality (Gini, 1921). Hence GC describes the shape of tree-size distributions (Valbuena et al., 2016) and is influenced by competitive interactions among trees (Cordonnier and Kunstler, 2015). Valbuena et al. (2012) demonstrated that the $GC = 0.5$ can be considered as a boundary between even-aged and uneven-aged stand structures. GC values far below 0.5 indicate a unimodal "normally distributed" size structure, which are commonly found in even-aged stands that are self-thinning (e.g. Coomes and Allen, 2007). Values close to 0.5 indicate irregular size distributions (Duduman, 2009), while values much greater than 0.5 represent "reverse-J" stand structures (Valbuena et al., 2013a).

1.2 Influence of Plot Size in Measurements of Forest Structure

Sample plots used for measuring plant communities are usually rectangular or circular in shape (Whittaker, 1972; Kent and Coker, 1992), with a wide range of possible plot sizes from fine to coarse scales (Chytrý and Otýpková, 2003). As the effects of plot size decrease with increasing size of a plot (David and Mishriky, 1968; Barbeito et al., 2009), an optimal size must be chosen compromising the acquisition of a field plot large enough to obtain a stable measure of forest structure, but no larger than necessary because of the costs involved (Otypková and Chytry, 2006). Structural diversity depends on the spatial resolution at which an index is evaluated (Lexerød and Eid, 2006b). Varying the scale of observation may therefore distort the information retrieved from an indicator (Chen and Crawford, 2012; Mauro et al, 2016). As plot size increases, GC estimates may be more reliable, but also fundamentally different stand conditions may aggregate (Coomes and Allen, 2007). Therefore, interpretation of data analysed at different scales remains one of the most challenging tasks in spatial statistics (Gotway and Young, 2002), as shown in the context of agriculture (Smith, 1938), sociology (Hannan, 1971), and environmental sciences (Jelinski and Wu, 1996). Also, the spatial distribution of trees has a practical effect on plot size, since clustered patterns require larger plot sizes to obtain a same degree of confidence in estimates (Upton and Fingleton, 1985; Pommerening, 2002; Kallimanis et al., 2008; Motz et al., 2010). Recently, Magnussen et al. (2016) suggested a method of upscaling to a common plot size to minimize scale effects in survey estimates, which achieved consistency among the quantiles and proportions of sampling distributions of forest attributes.

1.3 Influence of ALS Scan Density in Measurements of Forest Structure

Airborne laser scanning (ALS) is recognised as a highly effective tool for regional monitoring because it provides precise information about biophysical stand properties, (Gobakken et al., 2006; Gobakken and Næsset, 2008). The GC may be calculated as a descriptor of the distribution of ALS heights (Valbuena et al., 2017), or ALS metrics may be related to GC of tree sizes (Valbuena et al., 2013b). The spatial resolution of ALS data used in area-based methods has an effect on estimated values (Mascaro et al., 2011). In the context of remote sensing-assisted forest estimations, spatial resolution refers not only to the size of field plots but also to the size of pixels at which auxiliary variables are computed (Gobakken and Næsset, 2008; Ruiz et al., 2014; Valbuena et al., 2016). In ALS-assisted estimations of GC of tree size inequality, there is a lack of knowledge on the effects of varying plot size and spatial resolution.

Scan density is one of the most important aspects of ALS datasets that affects both processing and costs (Balsa-Barreiro and Lerma, 2014; Kandare et al., 2016). The importance of optimizing ALS point density lays in its trade-offs against ALS swath width, and hence costs (Baltsavias, 1999). Liu et al. (2007) observed that density reduction influenced the accuracy of digital terrain models (DTM) due to the presence of empty space intervals between points. A reduction in DTM accuracy may affect the calculation of metrics describing ALS height (Ruiz et al., 2014; Singh et al., 2015), although it would be unlikely to affect metrics describing their dispersion, such as GC. Gobakken and Næsset (2008) assessed the effect of point density on biophysical stand properties, finding that maximum height was the least affected metric and suggesting to avoid metrics most affected by point density. No previous studies have yet determined how stand density and ALS scan densities affects GC estimates from ALS.

1.4 Objectives

The aim of the study is to evaluate the effects of plot size and ALS scan density on field and ALS-derived estimates of GC in the boreal forests of Finland. We developed a simple method for selecting the optimal plot size for determining the GC of tree size inequality from field data, and for predicting GC reliably using ALS metrics as auxiliary variables.

2. Material and Methods

2.1 Study Area and Field Data Collection

The study was carried out in a typical boreal managed forest located in Eastern Finland (62° 31′ N, 30° 10′ E). Scots Pine (*Pinus sylvestris* L.) is the dominant species which represents 73% of the total wood volume, while Norway spruce (*Picea abies* Karst.) represents 16%, and deciduous species 11% of the total wood volume (Valbuena et al., 2014). The main properties of the field data such as stand density (N), basal area (G) and quadratic mean diameter (QMD) are shown in **Table 1**. The field data were collected in May-June 2010 and consisted of 79 squared plots (henceforth "*original field plots*") of various dimensions (20×20, 25×25 or 30×30 m, the smaller ones being in denser stands). With the intention of representing the contrast between highly homogeneous even-aged areas and more heterogeneous forest structures (Valbuena et al. 2016), forest stands were determined using stratified random sampling, whereas plot locations were purposively selected. After choosing the sampled stands, plots were located within the stands at a representative location. The reason for doing this was to avoid plot locations at stand borders and the high cost and measuring effort required to record the location of all individual stems within the plot. The absolute positions of every individual tree with dbh > 4 cm and tree top height taller than 4 m were mapped using an approach combining ALS and field surveying methods suggested by Korpela et al. (2007). Before the field measurement, a map of individual tree positions was generated from high density ALS data (see below) using an individual tree detection (ITD) method (Packalen et al., 2013). Actual positions of trees defined by their longitude/latitude coordinates (X_{actual}, Y_{actual}) were verified in the field, while the location of trees not detected by the ITD method were measured relative to the ITD-derived ones (distances and bearings) and least-square adjusted (Korpela et al., 2007).

****approximate position of Table 1*****

2.2 Simulation of Circular Plots

Preliminary tasks for the simulation included transformations into relative coordinates, the correction of edge effects and a sensitivity analysis to determine the number of simulations needed. Then, within each *original field plot* we simulated circular plots at random positions. Circular plots were chosen on the assumption that tree competition is the same in all spatial directions. The radius of these *circular simulated plots* was increased in 1-m intervals, generating concentric circles up to 15 m-radius. Since the position of individual trees were available from the original field data, we could extract the trees located within each circular simulated plots, computing an estimation of GC based on tree dbh. Likewise, the position of individual ALS returns located within each simulated circular plots could be extracted, using them to compute ALS metrics commonly employed in area-based estimation methods.

2.2.1 Transformation to Relative Distances and Edge Correction

Transformation of absolute tree coordinates into relative coordinates requires procedures of plot rotation and translation (Matos, 2014). Since in the case of our study the edges of original field plots were coincident with the UTM grid, there was no need to carry out plot rotations. In plot translation absolute coordinates of original field plots were modified into relative distances, by assigning the origin of axes (0, 0) to the south-western corner of the original field plot. Absolute coordinates of south-western corner (X_{corner}, Y_{corner}) were subtracted from the absolute coordinates of each tree (X_{abs}, Y_{abs}) to get their relative coordinates (X_{rel}, Y_{rel}).

$$(X_{rel}, Y_{rel}) = (X_{abs}, Y_{abs}) - (X_{corner}, Y_{corner}) \quad (2)$$

Moreover, Pommerening and Stoyan (2006) showed that edge effects play an important role in spatial statistics. Because the immediate neighbour trees outside the boundary of the *original field plots* were not measured, ignoring them would result in biased statistical estimations. Thus, indices based on tree positions require an edge correction method to reduce this bias. We chose a periodic boundary edge correction method (Diggle, 2003), since Pommerening and Stoyan, (2006) found it to be superior to other alternatives. This method consisted of replicating the same spatial pattern measured in the field around the *original field plot* (**Fig. 1**). Concentric *circular simulated plots* randomly positioned at the edge of the *original field plots* therefore also included the trees positioned out of the boundaries of the *original field plots*.

****approximate position of Figure 1*****

2.2.2 Plot Simulation and Sensitivity Analysis

A pilot sensitivity analysis was done with the intention to identify the minimum number of simulations within an original field plot which can guaranteed a stable and robust outcome for the simulation. We selected the *original field plot* with highest GC, hence likely the one most sensible to changes among different simulations, and repeated the analysis for 10, 100, 500, 700, 1000, 1500 and 2000 simulations. A position (X_{sim}, Y_{sim}) was randomly simulated within the *original field plot*, and GC was calculated for each *circular simulated plot* (see below) and for each plot radius (s ; m) (1-m intervals from 1 to 15 m) (**Fig. 1**). As explained below, the standard error of the mean (SEM) of values obtained for GC at each radius were considered in order to fix the minimum number of simulations at which no considerable improvement was observed by adding further replications. After setting the necessary number of simulations to a fixed number k based on the pilot sensitivity analysis, the whole procedure was repeated for the remaining 78 original field plots. Relative and absolute positions of all simulations were recorded so that they could later be used for extracting their corresponding ALS returns as well.

2.3 *Gini Coefficient Estimation*

The target was to calculate sample estimations of the GC describing the size inequality of the tree community represented at each *original field plot*. Its estimation (**Eq. 1**) was repeated for every concentric *circular simulated plot* of radii 1-15 m, and for all the simulated positions (X_{sim}, Y_{sim}). For this purpose, basal area (g; m^2) was calculated for each individual stem. Differences in g were computed for each pair of trees within each circular simulated plot. GC is the average of absolute differences relative to their mean ($\bar{g}$) (see detailed descriptions of GC calculation in Lexerød and Eid (2006a) and Valbuena et al. (2013b)). The reason of using g instead of dbh was to increase the influence of larger trees (Solomon and Gove, 1999). The unbiased estimator by Glasser (1962) was employed because it is appropriate for an estimation based on a finite number of trees n located within each circular simulated plot (**Eq. 1**). The mean GC ($\overline{GC}$) and its corresponding SEM were computed for each radius (from 1 to 15 m), and for each of the *original field plots*. SEM is a measure for the accuracy of those means, accounting for the variability between the samples, according to the number of simulations k and their sample standard deviation (SD). R statistical software (R Development Core Team, 2016) was used for all these calculations and statistical analyses.

We constructed a graph comparing $\overline{GC}$ results for increasing plot size s for all *original field plots*. The GC value at *circular simulated plots* must necessarily approximate asymptotically to the value of GC for the entire *original field plot* as the radius of circular simulated plots increases (Matos, 2014). For this reason, the value of GC obtained by applying equation (1) to the original field plot was used as a reference (GC_{ref}). In order to make all the simulated GC values directly comparable, we calculated absolute GC differences (GC_{diff}) by subtracting simulated GC values from the GC_{ref}:

$$\overline{GC}_{diff} = |GC_{ref} - \overline{GC}| \qquad (3)$$

This way, it was possible to analyse the difference of each simulated GC to its corresponding asymptotic value, allowing to set a common criterion to evaluate all simulations based on the stabilization of the estimated GC value (see below).

2.4 *Airborne Laser Scanning Data and Metric Computation*

ALS data was acquired on June 26, 2009 using ATM Gemini sensor (Optech, Canada) from 600-700 m above ground level with a 26^{o} field of view. Scan swath was 320 m wide with a 55% side overlap between the strips. A high resolution dataset with 11.9 pulses·m^{-2} scan density was produced from a pulse rate of 125 kHz. Details about the processing of ALS data are described in Packalen et al. (2013). The last echoes were classified as ground and interpolated into a DTM (Axelsson, 2000). The spatial resolution of DTM was 0.5 m based on the scan density, and the height above ground of individual ALS returns was obtained by subtraction of the DTM height beneath each of them. Echoes lower than 0.1 m from ground level were eliminated, as they were considered to be reflected from ground.

Individual ALS returns of each circular simulated plot based on its absolute coordinates (X_{sim}, Y_{sim}) were clipped, and area-based ALS metrics were computed from their heights with the help of FUSION software (USDA Forest Service; McGaughey, 2015). ALS metrics are statistics and descriptors of the distribution of ALS heights observed within a given area, which are usually employed as auxiliary variables in ALS-assisted forest estimations (**Table 2**). Some of these metrics were common statistics as, for example, the mean (*Mean*) standard deviation (*StdDev*) or the skewness (*Skew*) of the distribution of heights above ground of ALS returns contained within each *circular simulated plot*. We also computed the percentiles of their distribution, such as the 25th (*P25*), 50th (*P50*) or 99th (*P99*). In addition, we calculated the so-called canopy cover metrics (McGaughey, 2015), such as the proportion of returns backscattered from 0.1 m above the ground (*Cover*). Another metric included in FUSION was the canopy relief ratio (*CRR*), which is the difference between mean and minimum ALS return heights divided by a difference between maximum and minimum heights (Pike and Wilson, 1971).

****approximate position of Table 2*****

The effect of plot size in the relationship with GC was studied separately for each of these ALS metrics. For each radius, we gathered all the simulations carried out at all the *original field plots* and calculated all the ALS metrics listed in **Table 2**. They were used to calculate Pearson correlation coefficients (r) using all the combinations of field GC against each ALS metric. Then, we observed separately for each ALS metric the evolution of r when increasing the plot size s of the *circular simulated plots*. Since we were only interested in the capacity of the ALS metrics to explain variability in GC, regardless of whether their relationship was direct or indirect, we considered the absolute value of the correlation coefficient $|r|$ in the optimization, as explained below.

2.5 Basic Relationships

The plot size and spatial resolution at which an ALS-assisted estimation is carried out relates intrinsically to the sample size used in all calculations. Sample size affects the relationship between predictor and response, and therefore the accuracy of ALS estimation of any forest attributes (Gotway and Young, 2002; Mascaro et al., 2010; Næsset et al., 2015; Magnussen et al., 2016; Valbuena et al., 2016). In this context, sample size refers both to the number of trees used to calculate a given forest attribute, GC in this case, but also to the number of ALS returns involved in the computation of ALS metrics. The link between resolution and sample size is employed on the empirical densities of the datasets, i.e. stand density (N; trees·ha^{-1}) or ALS points density (d; points·m^{-2}) (Gobakken and Næsset, 2008; Motz et al., 2010; Jakubowski et al., 2013). Therefore, the effects of plot size and spatial resolution of the ALS estimated forest attributes also depend on N and d, and the combined effects of these two factors may explain why plot sizes suitable for field surveys may be found sub-optimal for ALS estimation (Næsset et al., 2015).

Hence, the relationship between the radius s of a circular plot and the number of trees (n) contained within is tied to the N at the location of the plot.

$$n = N\pi s^2 \tag{4}$$

This begs the question on whether the optimization method should search for an optimal plot radius (s^*; m) or an optimal sample size (n^*). In a forest environment of variable stand density N (**Table 1**), does the relationship between GC and ALS metrics depends on the plot size used, or on the number of trees surveyed? In order to research whether it makes a difference, we repeated the same procedure for both s^* and n^* optimization. In other words, we tested the results of optimization according to either plot radius or number of trees. In any of the cases, the relationship in **eq. (4)** assures that the methodology can be replicated for either dense or sparse forests, since s and n can always be deduced from one another by an empirical N.

Likewise, a similar relationship holds between the size of that same circular plot and the number of ALS returns backscattered from it, according to a given ALS scan density d. In this context of estimation using auxiliary variables, the scale concerns both to the size of the field plots and the spatial resolution of the pixel at which ALS metrics are calculated. Therefore, the number of ALS points (p) relates to the spatial resolution / plot size used (s) according to d :

$$p = d\pi s^2 \tag{5}$$

As before, the relationship in **eq. (5)** assures that the methodology can be replicated for any range of ALS scan densities, since s and p are trivially deducted from one another by an empirical d. As an overall conclusion, a given optimal plot size s^* necessarily implies optimal sample sizes as well, both n^* and p^*. Keeping these relationships in mind is key to demonstrating the validity of the optimization method for replication elsewhere according to the N and d which may occur at any other study cases, and therefore the method is equally valid for both dense and sparse forests and ALS surveys with low or high scan density.

2.6 Plot Size Optimization

To optimize the plot size which should be used for a reliable GC estimation, and thereby also the optimal spatial resolution for an estimation of GC from ALS datasets, we determined two criteria to be applied sequentially: (1) stabilization of GC as estimator of the population value from the field information itself, and (2) maximizing the GC variability explained by ALS metrics. Therefore, *Criterion I* considered the minimum plot radius at which the estimation of GC remained stable to further increases in plot size. *Criterion II* was set to optimize the ALS-assisted estimation, by observing changes in the correlation between the field GC and each ALS metric among the simulated plot radii.

Criterion I was implemented by observing the evolution of $\overline{GC}_{diff}$ for increasing radii at every original field plot. We set a maximum value of $\overline{GC}_{diff} = 0.05$ at which it was considered that the estimation of GC was stable and representative of the population, and, therefore, selected the minimum plot radius s as the smallest meeting the first criterion for all the 79 original field plots.

Criterion II consisted in maximizing the explained variance in the GC values when predicted from ALS metrics. To implement this criterion we combined all the GC/ALS metric pairs for all the simulations carried out at all the original field plots, and grouped them according to the different simulated radii. The optimal radius was set to be that one showing the maximum $|r|$ value for a given metric. To make an overall decision, we put the focus on those metrics showing higher correlations, and decided a range of optimal sizes accordingly (since the empirical maximum may differ for different ALS metrics). As a summary, the final optimal plot size s^* for a given metrics was:

$$s^* = \max|r| \;\Big|\; \overline{GC}_{diff} < 0.05 \qquad (6)$$

2.7 Sample Size Optimization

For sample size optimization, seeking to deduct what is the minimum number of trees needed to obtain a reliable GC estimation, and the optimum for its ALS prediction, we applied the same two sequential criteria employed for plot size optimization (section 2.6). Therefore, the simulations were similar as before, but they increased the size of simulated circular plots according to the resulting number of trees n instead of plot radii. Thus, for implementing *Criterion I*, the evolution of $\overline{GC}_{diff}$ was observed for increasing number of trees n, also setting a maximum value of $\overline{GC}_{diff} = 0.05$. As before, we selected the minimum n as the smallest meeting *Criterion I* for all 79 original field plots. *Criterion II* also consisted in maximizing the absolute correlation between the GC values and each of the ALS metrics. New values of $|r|$ were obtained for increasing values of n, and the final optimal sample size (n^*) for each given ALS metric was then set as:

$$n^* = \max|r| \;\Big|\; \overline{GC}_{diff} < 0.05 \qquad (7)$$

Finally, we compared which alternative, **eq. (6)** or **(7)**, would be more convenient for a practical plot size optimization, discussing the results obtained by either method.

2.8 Reduction of ALS Point Density

Once deducted an optimal spatial resolution s^*, we also investigated the effects of varying ALS scan density d. The original point density was reduced to 0.5, 0.75, 1, 3, 5, 7.5, and 10 points·m^{-2}. A common option to reduce point density is by moving a 1 m window and selecting random points from the point cloud to reach the desired point density (e.g., Magnussen et al. 2010). We calculated a correct thinning factor for each desired point density d (Ruiz et al., 2014), following the method detailed by Jakubowski et al. (2013) which incorporates routines included in *LAStools* (RapidLasso GmbH Inc.; Isenburg, 2016). New ALS metrics over each of the k simulated circular plot positions and their correlations against the GC values obtained from the field information were calculated, and the entire procedure was repeated for all the reduced densities. In a similar manner as it was done for s and n, the effects of varying ALS scan density were studied by observing the changes in $|r|$, i.e. the effects in the relationship between the GC of tree size inequality and the ALS metrics with more explanatory capacity towards this given forest attribute.

3. Results

3.1 Establishing the Number of Simulations

Figure 2 shows the results of sensitivity analysis carried out to select the minimum number of simulations that would yield a robust estimation of GC for increasing simulated plot radii. As expected, the GC value estimated from few simulations fluctuated considerably, and this fluctuation decreased as the number of simulations increased (**Fig. 2a**). The expected general trend toward the asymptotic value obtained by the entire population (GC_{ref}) was generally observed in **Fig. 2a**. Very little variation in GC estimates were observed when the number of simulations increased from 700. Similarly, the SEM decreased as the number of simulations increased (**Fig. 2b**), remaining virtually unchanged from 700 to 2000 simulations.

Consequently, we decided to carry out the analysis using $k = 700$ simulations of 15 concentric circular simulated plots located within each 79 original field plots.

****approximate position of Figure 2*****

3.2 Plot Size Optimization

Figure 3a shows the resulting $\overline{GC}_{diff}$ for each of the 79 original plots, and **Table 3** is a summary of these results which was used for establishing *Criterion I*, which set the minimum plot size that would provide a reliable GC estimation for the population. Circular simulated plots of small sizes provided GC estimates that differed considerably from the population values as considered by GC_{ref}. Nonetheless, once the estimation reached stabilization, an increase in the radius of a circular plot (and hence the sampling effort) would not necessarily imply a considerable change in the estimation of GC (**Fig. 3a**). Our results showed that only few of the original field plots (probably very homogeneous stands) obtained stable GC estimations from very small circular simulated plots (**Table 3**). On the other hand, for larger circular simulated plots the differences against the original field plots representing the population became negligible. We observed that stabilization of the GC estimation started beyond of simulated plot radius $s = 6$ m, from which all the original field plots fell within the $\overline{GC}_{diff} < 0.05$ limit. We therefore established that the smallest plot size required for a reliable GC estimation should be set at areas sizing around 113 m^2.

****approximate position of Figure 3*****

****approximate position of Table 3*****

With regards to *Criterion II*, the evolution of $|r|$ with increasing plot size was observed for all ALS metrics included in FUSION. Results showed that changes in the relationship between the field GC of tree sizes and metrics describing the distribution of ALS return followed some general trends and patterns. For this reason and for simplifying, we chose to show only few ALS metrics in **Fig. 4a**, which we considered representatives of the general trends observed. These ALS metrics were the described *P25, P50, P99, Skew, StdDev, Cover* and *CRR* (**Table 2**). **Fig. 4a** showed an erratic fluctuation for the values of $|r|$ obtained for plot sizes smaller than a radius $s = 5$ m , which was possibly caused by the instability observed in the GC estimation at smaller plot sizes (**Fig. 3**). For this reason, we shadowed this area in grey colour in **Fig. 4**, denoting that such small plot sizes were already dismissed under *Criterion I*. Once GC estimation reached stabilization, its correlation to ALS metrics often yielded a convex curve as plot size increased (**Fig. 4a**). Therefore, the optimal plot size was possible to determine via maximization of $|r|$. This tendency was more clearly marked for those ALS metrics showing higher values of $|r|$, i.e. more correlated to the GC of tree sizes (**eq. 2**), such as *Skew*, *Cover* or *CRR*. For other ALS metrics less related to GC, like return height percentiles (*P25, P50* or *P99*) or *StdDev*, this tendency was less marked (**Fig. 4a**). For the optimization of plot size, we selected those metrics showing highest correlation against GC, since in practice they would be those more involved in its estimation. **Table 4** shows that the maximum $|r|$ for ALS metrics *Skew*, *Cover* or *CRR* ranged $s^* = 9$-12 m plot radius (the quality of histograms and scatterplots between variables involved can be checked in the **Supplementary Material**). It can be observed in **Fig. 4a** that beyond a circular simulated plot of 12 m the correlation showed a decreasing trend for most ALS metrics. Also, local maxima may be found for some ALS metrics for very small plot sizes, which is probably an artefact due to the above-mention instability in GC estimation at very small plot sizes (**Fig. 3**). This proved the necessity of imposing *Criterion I* as a prior step to correlation maximization. As a conclusion, under the established combined *Criteria I* and *II*, we determined that any plot radius $s < 6$ m (113 m^2 area) should be avoided (denoted by grey colour in **Fig. 4a**), and the optimal plot size for an ALS-assisted estimation of GC must be carried out using scales sizing 250-450 m^2, which concerns to both the size of the field plot and the pixel of the grid employed for ALS estimation.

****approximate position of Figure 4*****

****approximate position of Table 4*****

3.3 Sample Size Optimization (Stand Density Effect)

On the other hand, **Figure 3b** shows the evolution of $\overline{GC}_{diff}$ for increasing sample sizes (number of trees n) at each of the 79 original field plots. It is worth mentioning the **Figs. 3a** and **3b** relate to one another according to **eq. (4**). As a consequence, a similar tendency can be found for both of them. **Table 3** expresses the number of trees that correspond on average to a given sample size. Therefore, the minimum value obtained for *Criterion I* in plot size optimization, $s = 6$, corresponds to stating that a minimum number of $n = 15$ trees are required for a stable GC estimation (shaded area in **Fig. 4b**). We nevertheless further postulated that this minimum number of trees may be dependent on the heterogeneity of the forest itself, being possibly larger in the presence of higher inequality of tree sizes. This presumption was demonstrably true, as it can be observed in a scatterplot comparing the minimum number of trees required for a stable GC estimation at each of the 79 original plots against their overall value of tree size inequality observed (GC_{ref}; **Fig. 5**). Such relationship was not so straightforward if *Criterion I* was imposed using s instead (results not shown), which demonstrates the effect of varying forest stand density N. Hence, obtaining a stable GC estimation is more dependent of measuring a minimum number of trees than imposing a given size for the field plot used.

****approximate position of Figure 5*****

The case for *Criterion II* was different, as it can be deducted when observing the same ALS metrics employed to optimize s – *P25, P50, P99, Skew, StdDev, Cover* and *CRR* –, but trying to optimize n instead (**Fig. 4b**). Again, a similar tendency can be found since **Figs.4a-b** are also related by **eq. (4**). Results were therefore very similar whether optimization was carried out according to plot size (**eq. 6**) or sample size (**eq. 7**). The values of $|r|$ also followed a convex curve when increasing the number of trees measured, and an optimal sample size n^* could be reliably determined via $|r|$ maximization. Our results showed that a number of trees approximately ranging $n^* = 30\text{-}60$ (**Table 4**) should be involved in the computation of GC, in order to maximize the efficiency of its estimation using ALS. Since the value of $|r|$ involves both the field GC and the ALS metrics, its changes are determined by both N and d (**eqs. 4-5**), and both may cause a change in the correlation between the two variables.

3.4 Effect of Point Density on the Relationship of GC

According to the previous results, we set the optimal plot size to $s^* = 9$ m in order to further analyse the possible effects due to varying scan density. Among all the ALS metrics (**Table 2**), we selected those same ones employed previously – *P25, P50, P99, Skew, StdDev, Cover* and *CRR* – to allow direct comparison. **Fig. 6** shows the evolution in $|r|$ for increasing ALS point density d. No considerable changes were observed in the correlation between the field GC and the ALS metrics, which suggests that d has no major effects on their relationship. However, a decreasing trend in $|r|$ could generally be observed when point densities decreased below $d < 3$ points·m^{-2} (**Fig. 6**). Overall, these results therefore suggest that the relationship between GC and ALS metrics is mainly dependent on the plot size employed, and rather independent of stand density and ALS scan density

****approximate position of Figure 6*****

4. Discussion

In this study we evaluated the effects of plot size and sample size on the GC of tree size inequality, and on its practical estimation using remote sensing methods based on ALS. Sample size refers to the number of individual elements (trees or ALS returns) included within a given sample area, which is therefore determined by the spatial resolution employed for evaluating a given forest attribute. We also analysed the effects of ALS scan density and, overall, we observed that plot size had greater effects on the relationship between GC and ALS metrics than either of the other two criteria considered. The motivation for studying these effects is grounded on the fact that inappropriate plot sizes may provide unreliable estimates and lead to sub-optimal forest management decisions (Eid, 2000; Mauro et al., 2010). Valbuena et al. (2013a) pointed out that the estimation of GC is affected by the area at which it is evaluated. Results in **Fig. 3** illustrate how the $\overline{GC}_{diff}$ decreases when increasing the size of circular plots and, and hence their corresponding sample size. $\overline{GC}_{diff}$ values markedly dropped for smaller plot radii and sample sizes. This decrease smooths from bigger sizes, which indicates stabilization of the estimation (*Criterion I*). **Fig. 2a** also shows an example of this tendency to asymptotically approach the population value, which was also observed by George (2003), Barbeito et al. (2009), or Matos (2014). Based on *Criterion I* ($\overline{GC}_{diff} < 0.05$), the circular plot should be large enough ($s \geq 6$ m) to have minimum sample size of $n \geq 15$ trees (**Fig. 3**). Although the minimum plot size also depends on the stand density of an area, **eq. (4)** can be used to adjust the method to any forest areas, whether sparsely or densely forested. This conclusion may therefore be partly extended to other forest types, as it can be for example deduced (via **eq. 4**) that minimum radius of $s \geq 12$ m would be needed in sparsely forested area of only 300 stems·ha^{-1} (Lombardi et al., 2015). **Eq. (4)** therefore brings generality to the method, since plot sizes may hence be tailored to forest areas of differing stand densities.

In this article we also postulated that maximizing the explained variability between the GC estimated from the field and ALS metrics could be a valid criterion to optimize the reliability of ALS-assisted estimations of GC (*Criterion II*). Results in **Fig. 4a** showed that our presumption was correct, since the $|r|$ values between GC and most ALS metrics, especially the most correlated ones, followed a convex curve with a maximum that could be searched to reach an optimal plot size / spatial resolution for the estimation. On the other hand, once the GC reached some stabilization, the correlation between them remains largely unchanged. Therefore, a lower plot size limit is to be imposed to avoid local minima that could appear as an artefact of the unstable estimation of GC at low sample sizes. We shaded this area in grey colour in **Fig. 4 (a, b)**, denoting the area that was already dismissed as a result of *Criterion I* (**Fig. 3**; George, 2003). In larger plots the sample size was more representative of the total population. Combining both criteria, we found in our study area that an optimal circular plot radius of $s^* = 9\text{-}12$ m, which corresponds to a spatial resolution of sampling units sizing 250-450 m^2 (**Fig. 4a**), would be suitable for ALS-assisted GC estimation. Since plot size and sample size are interdependent (**eq. 4**), this result may be suitable for any area with a similar average number of trees ($N \cong 1300$ stems·ha^{-1}; **Table 2**). According to these results, therefore, most forest datasets commonly acquired in operational inventories would be acceptable for an ALS-assisted estimation of the GC of tree sizes. Lombardi et al. (2015) deduced a larger optimal plot radius $s^* = 13\text{-}15$ m for other forest attributes, most likely due to lower N in the forest areas considered. For studies dealing with differing plot sizes, one possibility could be to upscale GC to a common plot size (Kent and Coker, 1992; Magnussen et al., 2016).

Some of the reflexions raised in this article affect all other types of forest attributes and remotely sensed auxiliary variables that may be used in forest estimations (Jelinski and Wu, 1996). However, different forest attributes are differently affected by varying plot sizes

(Chytrý and Otýpková, 2003). Some forest variables such as stand density or biomass would show an averaging effect as plot size increases (Jelinski and Wu, 1996; Gotway and Young, 2002; Ruiz et al., 2014), which in turn derives in improved model efficiency when using larger scales in remote sensing estimations (Næsset et al., 2015; Mauro et al., 2016). But there is a trade-off between model accuracy and spatial resolution, and root mean squared errors increase from10-15% for 1000-4000 m^2 to 20-25% for 200-250 m^2 (Næsset, 2002, 2004, 2007). However, this averaging effect is not applicable to forest attributes describing structural diversity and heterogeneity (Coomes and Allen, 2007). In fact, many variables necessarily augment when the plot size increases, for instance species richness and diversity (e.g., Humphrey et al. 2000; Otypková and Chytry, 2006; Kallimanis et al., 2008; Fibich et al., 2016) as traditionally assessed through rarefaction (Kent and Coker, 1992). A similar effect can be observed in other measures of forest heterogeneity (Barbeito et al., 2009; Motz et al., 2010; McRoberts et al., 2012), and thus in the GC (Valbuena et al., 2013a, Matos, 2014), since increasing the size of a plot increases the probability of finding an additional differently-sized tree (Chen and Crawford, 2012; Valbuena et al., 2012). This is why estimated GC values in **Fig. 3** asymptotically approach the value of the larger original field plot (George, 2003; Matos, 2014), which is never exceeded. Imposing a criterion defining which of the plausible plot sizes should be used is therefore not a trivial question to tackle. Matos (2014) employed a number of different criteria based on the field information only – stabilization of the estimate, stabilization of certainty of the estimate and convergence with GC_{ref} –, none of them resulting fully satisfactory and definitive as they all ultimately rely on a subjective assumption (Cressie, 1993). For this reason, in this article we approached the question of plot size from the viewpoint of its practical estimation using ALS remote sensing. The convex curves obtained in **Fig. 4a** proved this approach to be highly beneficial, since maximization of correlation $|r|$ between GC values and selected ALS predictors provides with a more objective method for determining the optimal plot size for the assessment of GC of tree size inequality. Still, due to the very high uncertainty observed in the estimation of GC when using very small plot sizes (**Fig. 3b**; Smith, 1938; Lombardi et al., 2015), we deducted that a criterion avoiding great divergence with GC_{ref} may be imposed as a prior step to maximization (Motz et al. (2010) referred to it as minimum grid spacing). Further research could focus on modelling GC from ALS metrics and investigate how the interaction among many ALS metrics in a same model may play a relevant role in the optimization of plot size and spatial resolution.

The analyses carried out with reduced point densities revealed that lowering point density barely affects the correlation between GC and ALS metrics, unless using a very sparse scan density $d < 3$ points·m^{-2}. Previous studies such as Maltamo et al. (2006), Ruiz et al. (2014) or Singh et al. (2015) also indicate that reducing the point density is not affecting the accuracy of volume prediction and demonstrate that the effects of varying scan densities can be eluded in practical applications. It must be taken into account, however, that the DTM used in this study was based on original point density, and the errors in DTM determination at sparser densities (Liu et al., 2007; Ruiz et al., 2014) may induce to further uncertainty, although this presumably has a lesser effect on those metrics most related to GC. Furthermore, since ALS datasets from national programmes are currently surveying entire countries at densities typically between 0.5-1 points·m^{-2} (Artuso et al., 2003), it must also be pointed out the relevance of results in **Fig. 6** which render most of these nation-wide ALS datasets unsuitable for reliably estimating GC (Kandare et al. 2016). In line with results in Valbuena et al. (2017), who postulated that the low densities incur in critical omission of understorey development, our results demonstrate that indeed there is a need for increasing point densities up to $d = 3$ points·m^{-2}. This result is very concurrent with those obtained by Ruiz et al. (2014) and Watt et al. (2014) for different forest attributes in different stand types, and therefore the case seems clear that ALS datasets obtained for forest applications should reach this minimum density requirement.

5. Conclusion

In this study we studied how changing spatial resolution can affect the relationship between GC and ALS metrics. We used three criteria for optimization: plot size, stand density and ALS scan density. The effects of stand and scan densities are intimately interrelated to plot size, since they together determine the sample size employed in calculations. Amongst those three criteria, we found plot size to predominantly affect the relationship between GC and ALS metrics.

We observed that the estimation of GC is strongly affected by the size of the forest plot surveyed. Very small sample size and plot radii are more sensitive to GC variations, unrepresentative of the total population, producing unstable and unreliable GC estimations. The GC estimation stabilizes as the size of plots and samples increases, as larger plots contain a more appropriate number of observations (sample size) representing the population. We determined that, in a boreal managed forest, a minimum number of 15 trees ought to be measured for a reliable GC estimation, regardless of the stand density present at each forest stand.

We developed a method for plot size optimization based on a combination of two criteria: (1) imposing a minimum of number of 15 trees measured, and (2) maximizing the absolute correlation between field GC and ALS metrics. The plot level correlation between ALS metrics and field GC showed a convex tendency for increasing plot sizes. Our results showed that 9-12 m-radius plots produced the maximum correlation thus they are suitable for ALS-assisted GC estimation. Basic relationships between plot size and sample size may be used to accommodate the method to forested environments of varying stand densities.

With regards to the effects of ALS scan density, we observed that it can barely have any effects unless lowered under 3 points m^{-2}. This however may be relevant for the practical application of low-density national datasets, and therefore we would recommend increasing their scan densities with the intention to render nation-wide datasets useful for studying forest heterogeneity.

Acknowledgements

This research was carried out in the framework of the project LORENZLIDAR: Classification of Forest Structural Types with LiDAR remote sensing applied to study tree size-density scaling theories, funded by a Marie S. Curie Individual Fellowship H202-MSCA-IF-2014 (project 658180). Syed Adnan's PhD is funded by the National University of Sciences and Technology (NUST), Pakistan under FDP 2014-15.

References

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, 7-7.

Artuso, R., Bovet, S., Streilein, A. 2003. Practical Methods for the Verification of Countrywide Terrain and Surface Models. ISPRS WG III/3 Workshop. 3-D reconstruction from airborne laser scanner and InSAR data. Dresden, Germany 8-10 October 2003. In: The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXIV, part 3/WG13.

Axelsson, P., 2000. DEM generation from laser scanner data using adaptive TIN models. International Archives of Photogrammetry and Remote Sensing 33, 111-118.

Balsa-Barreiro, J., Lerma, J.L., 2014. Empirical study of variation in lidar point density over different land covers. Int. J. Remote Sens. 35, 3372-3383.

Baltsavias, E.P., 1999. Airborne laser scanning: basic relations and formulas. ISPRS Journal of photogrammetry and remote sensing 54, 199-214.

Barbeito, I., Cañellas, I., Montes, F., 2009. Evaluating the behaviour of vertical structure indices in Scots pine forests. Ann. For. Sci. 66, 710-710.

Bollandsås, O.M., Næsset, E., 2007. Estimating percentile-based diameter distributions in uneven-sized Norway spruce stands using airborne laser scanner data. Scand. J. For. Res. 22, 33-47.

Bourdier T., Cordonnier T., Kunstler G., Piedallu C., Lagarrigues G., Courbaud B. 2016. Tree size inequality reduces forest productivity: an analysis combining inventory data for ten european species and a light competition model. PLoS ONE 11(3), e0151852.

Buongiorno, J., Dahir, S., Lu, H., Lin, C., 1994. Tree Size Diversity and Economic Returns in Uneven-Aged Forest Stands. For. Sci. 40, 83-103.

Chen, Z., Crawford, C.A.G., 2012. The role of geographic scale in testing the income inequality hypothesis as an explanation of health disparities. Soc. Sci. Med. 75, 1022-1031.

Chytrý, M., Otýpková, Z., 2003. Plot sizes used for phytosociological sampling of European vegetation. Journal of Vegetation Science 14, 563-570.

Coomes, D.A. and Allen, R.B., 2007. Mortality and tree-size distributions in natural mixed-age forests. Journal of Ecology, 95(1), pp.27-40.

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.

Cressie, N., 1993. Aggregation in Geostatistical Problems, in Soares, A. (Ed.), Geostatistics Troia'92: Volume 1. Springer Netherlands, Dordrecht, pp. 25-36.

David, H.A., Mishriky, R.S., 1968. Order Statistics for Discrete Populations and for Grouped Samples. Journal of the American Statistical Association 63, 1390-1398.

Diggle, P.J., 2003. Statistical Analysis of Spatial Point Patterns, 2nd ed. Arnold, London.

Duduman, G., 2009. An ecological approach for establishing the allowable cut in forests where single tree selection system is applied. Editura Universităţii Suceava.[In Romanian] .

Eid, T., 2000. Use of uncertain inventory data in forestry scenario models and consequential incorrect harvest decisions. Silva Fenn. 34, 89-100.

Fibich, P., Lepš, J., Novotný, V., Klimeš, P., Těšitel, J., Molem, K., Damas, K., Weiblen, G.D., 2016. Spatial patterns of tree species distribution in New Guinea primary and secondary lowland rain forest. Journal of Vegetation Science 27, 328-339.

George, D., 2003. The Small-Sample Bias of the Gini Coefficient: Results and Implications for Empirical Research. Review of Economics and Statistics 85, 226-234.

Gini, C., 1921. Measurement of inequality of incomes. The Economic Journal 31, 124-126.

Glasser, G.J., 1962. Variance formulas for the mean difference and coefficient of concentration. Journal of the American Statistical Association 57, 648-654.

Gobakken, T., Næsset, E., Nelson, R., 2006. Developing regional forest inventory procedures based on scanning LiDAR. Proceedings of the International Conference Silvilaser 2006, Matsuyama, Japan. Japan Society of Forest Planning, Forestry and Forest Products Research Institute and Ehime University. pp. 99-104.

Gobakken, T., 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.

Gotway, C.A., Young, L.J., 2002. Combining incompatible spatial data. Journal of the American Statistical Association 97, 632-648.

Hannan, M.T., 1971. Aggregation and Disaggregation in Sociology. Lexington Books.

Humphrey, J., Newton, A., Peace, A., Holden, E., 2000. The importance of conifer plantations in northern Britain as a habitat for native fungi. Biol. Conserv. 96, 241-252.

Hvistendahl, M., 2014. While emerging economies boom, equality goes bust. Science, 344, 832-835

Isenburg, M., 2016. "LAStools–efficient tools for LiDAR processing". (Version 160921, academic). Retrieved from http://rapidlasso.com/LAStools.

Jakubowski, M.K., Guo, Q., Kelly, M., 2013. Tradeoffs between lidar pulse density and forest measurement accuracy. Remote Sens. Environ. 130, 245-253.

Jelinski, D., Wu, J., 1996. The modifiable areal unit problem and implications for landscape ecology. Landscape Ecol. 11, 129-140.

Kallimanis, A.S., Halley, J.M., Vokou, D. & Sgardelis, S.P., 2008. The scale of analysis determines the spatial pattern of woody species in the Mediterranean environment. Plant Ecology 196: 143–151.

Kandare, K., Ørka, H.O., Chan, J.C.-W., 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, pp. 337-359.

Kent, M., Coker, P., 1992. Vegetation description and analysis: a practical approach. Chichester: Wiley.

Knox, R.G., Peet, R.K., Christensen, N.L., 1989. Population dynamics in loblolly pine stands: changes in skewness and size inequality. Ecology 70, 1153-1167.

Korpela, I., Tuomola, T., Välimäki, E., 2007. Mapping forest plots: an efficient method combining photogrammetry and field triangulation. Silva Fenn. 41.

Lähde E., Laiho O., Norokorpi Y., 1999. Diversity-oriented silviculture in the boreal zone of Europe. Forest Ecology and Management 118: 223–243.

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, 1835–1847.

Lexerød, N., Eid, T., 2006a. An evaluation of different diameter diversity indices based on criteria related to forest management planning. For. Ecol. Manage. 222, 17-28.

Lexerød, N., Eid, T., 2006b. Assessing suitability for selective cutting using a stand level index. Forest Ecology and Management, 237(1–3), pp. 503-512.

Liu, X., Zhang, Z., Peterson, J., Chandra, S., 2007. The effect of LiDAR data density on DEM accuracy. Proceedings of the International Congress on Modelling and Simulation (MODSIM07) 2007, Modelling and Simulation Society of Australia and New Zealand Inc., pp. 1363-1369.

Lombardi, F., Marchetti, M., Corona, P., Merlini, P., Chirici, G., Tognetti, R., Burrascano, S., Alivernini, A., Puletti, N., 2015. Quantifying the effect of sampling plot size on the estimation of structural indicators in old-growth forest stands. For. Ecol. Manage. 346, 89-97.

Lundqvist, L., 1994. Growth and Competition in Partially Cut Sub-Alpine Norway Spruce Forests in Northern Sweden. For. Ecol. Manage. 65, 115-122.

Magnussen, S., Mandallaz, D., Lanz, A., Ginzler, C., Næsset, E., Gobakken, T., 2016. Scale effects in survey estimates of proportions and quantiles of per unit area attributes. For. Ecol. Manage. 364, 122-129.

Maltamo, M., Eerikäinen, K., Packalén, P., Hyyppä, J., 2006. Estimation of stem volume using laser scanning-based canopy height metrics. Forestry 79, 217-229.

Mascaro, J., Detto, M., Asner, G.P., Muller-Landau, H.C., 2011. Evaluating uncertainty in mapping forest carbon with airborne LiDAR. Remote Sens. Environ. 115, 3770-3774.

Matos, A., 2014. Effect of scale factor in estimation of Gini coefficient. Master's thesis. University of Eastern Finland.
http://epublications.uef.fi/pub/urn_nbn_fi_uef-20140718/index_en.html
http://www.oppi.uef.fi/opk/video/europeanforestry/a_matos_seminar.mp4
(accessed August 2016)

Mauro, F., Valbuena, R., Manzanera, J., García-Abril, A., 2010. Influence of Global Navigation Satellite System errors in positioning inventory plots for tree-height distribution studies Canadian journal of forest research 41, 11-23.

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

McElhinny, C., Gibbons, P., Brack, C., Bauhus, J., 2005. Forest and woodland stand structural complexity: Its definition and measurement. For. Ecol. Manage. 218, 1-24.

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

McRoberts, R.E., Winter, S., Chirici, G., LaPoint, E., 2012. Assessing Forest Naturalness. For. Sci. 58, 294-309.

Motz, K., Sterba, H., Pommerening, A., 2010. Sampling measures of tree diversity. For. Ecol. Manage. 260, 1985-1996.

Næsset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sens. Environ. 80, 88-99.

Næsset, E., 2004. Accuracy of forest inventory using airborne laser scanning: evaluating the first Nordic full-scale operational project. Scand. J. For. Res. 19, 554-557.

Næsset, E., 2007. Airborne laser scanning as a method in operational forest inventory: Status of accuracy assessments accomplished in Scandinavia. Scand. J. For. Res. 22, 433-442.

Næsset, E., Bollandsås, O.M., Gobakken, T., Solberg, S., McRoberts, R.E., 2015. The effects of field plot size on model-assisted estimation of aboveground biomass change using multitemporal interferometric SAR and airborne laser scanning data. Remote Sens. Environ. 168, 252-264.

Neumann, M., Starlinger, F., 2001. The significance of different indices for stand structure and diversity in forests. For. Ecol. Manage. 145, 91-106.

O'Hara, K.L., Gersonde, R.F., 2004. Stocking control concepts in uneven-aged silviculture. Forestry 77, 131-143.

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.

Oliver, C.D., Larson, B.C., 1990. Forest Stand Dynamics. McGraw-Hill, Inc.

Otypková, Z., Chytry, M., 2006. Effects of plot size on the ordination of vegetation samples. Journal of Vegetation Science 17, 465-472.

Packalen, P., Vauhkonen, J., Kallio, E., Peuhkurinen, J., Pitkänen, J., Pippuri, I., Strunk, J., Maltamo, M., 2013. Predicting the spatial pattern of trees by airborne laser scanning. Int. J. Remote Sens. 34, 5154-5165.

Pike, R.J., Wilson, S.E., 1971. Elevation-relief ratio, hypsometric integral, and geomorphic area-altitude analysis. Geological Society of America Bulletin 82, 1079-1084.

Pommerening, A., 2002. Approaches to quantifying forest structures. Forestry 75, 305-324.

Pommerening, A., Stoyan, D., 2006. Edge-correction needs in estimating indices of spatial forest structure. Canadian Journal of Forest Research 36, 1723-1739.

Pukkala, T., Laiho, O., Lähde, E., 2016. Continuous cover management reduces wind damage. For. Ecol. Manage. 372, 120-127.

R Development Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Robles, A., Rodríguez, M.A., Alvarez-Taboada, F. 2016. Characterization of wildland-urban interfaces using LiDAR data to estimate the risk of wildfire damage. Revista de Teledeteccion, 45: 57-69.

Ruiz, L.A., Hermosilla, T., Mauro, F., Godino, M. 2014. Analysis of the influence of plot size and LiDAR density on forest structure attribute estimates. Forests, 5 (5): 936-951

Singh, K.K., Chen, G., McCarter, J.B., Meentemeyer, R.K. 2015. Effects of LiDAR point density and landscape context on estimates of urban forest biomass. ISPRS Journal of Photogrammetry and Remote Sensing, 101: 310-322

Smith, H.F., 1938. An empirical law describing heterogeneity in the yields of agricultural crops. The Journal of Agricultural Science 28, 1-23.

Staudhammer, C.L., LeMay, V.M., 2001. Introduction and evaluation of possible indices of stand structural diversity. Canadian journal of forest research 31, 1105-1115.

Sterba, H., Ledermann, T., 2006. Inventory and modelling for forests in transition from even-aged to uneven-aged management. For. Ecol. Manage. 224, 278-285.

Upton, G., Fingleton, B., 1985. Spatial Data Analysis by Example. Volume 1: Point Pattern and Quantitative Data. John Wiley & Sons Ltd.

Valbuena, R., Packalén, P., Martín-Fernández, S., Maltamo, M., 2012. Diversity and equitability ordering profiles applied to study forest structure. For. Ecol. Manage. 276, 185-195.

Valbuena, R., Packalen, P., Mehtätalo, L., García-Abril, A., Maltamo, M., 2013a. Characterizing forest structural types and shelterwood dynamics from Lorenz-based indicators predicted by airborne laser scanning. Canadian Journal of Forest Research 43, 1063-1074.

Valbuena, R., Maltamo, M., Martín-Fernández, S., Packalen, P., pascual, C. and Nabuurs, G., 2013b. Patterns of covariance between airborne laser scanning metrics and Lorenz curve descriptors of tree size inequality. Canadian Journal of Remote Sensing, 39, pp. S18-S31.

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.

Valbuena, R., Eerikainen, K., Packalen, P., Maltamo, M., 2016. Gini coefficient predictions from airborne lidar remote sensing display the effect of management intensity on forest structure. Ecol. Ind. 60, 574-585.

Valbuena, R., Maltamo, M., Mehtätalo, L., Packalen, P., 2017. Key structural features of boreal forests may be detected directly using l-moments from airborne lidar data. Remote Sensing of Environment 194: 437-446.

Vihervaara P., Mononen L., Auvinen A.P., 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.

Watt, M.S., Meredith, A., Watt, P., Gunn, A., 2014. The influence of LiDAR pulse density on the precision of inventory metrics in young unthinned Douglas-fir stands during initial and subsequent LiDAR acquisitions. New Zealand Journal of Forestry Science, 44 (1), 9 p.

Weiner, J., 1990. Asymmetric competition in plant populations. Trends in Ecology & Evolution, 5(11), pp. 360-364.

Weiner, J., Solbrig, O.T., 1984. The meaning and measurement of size hierarchies in plant populations. Oecologia 61, 334-336.

Weiner J., Thomas S., 1986. Size variability and competition in plant monocultures. Oikos 47: 211–222.

Whittaker, R.H., 1972. Evolution and measurement of species diversity. Taxon , 213-251.

Zheng, X., Xia, T., Yang, X., Yuan, T., Hu, Y., 2013. The Land Gini Coefficient and Its Application for Land Use Structure Analysis in China. PLoS ONE 8, 1-10.

Table Titles

Table 1. Properties of the study area.

Table 2. Summary of ALS metrics computed with FUSION and used in this research (McGaughey, 2015).

Table 3. For each radii, proportion of the total number of original field plots within the $\overline{GC}_{diff} < 0.05$ limit (*Criterion I*), and average number of trees contained within those plots.

Table 4. Maximum absolute correlation between field GC and ALS predictors (*Criterion II*). See **Table 2** for description of ALS metrics.

Figure Captions

Figure 1. Reproduction of tree positions (dots) within an original field plot (red rectangle) surrounded by edge correction i.e. translation method (i.e. periodic boundary), and a sample of 10 random realizations of simulated concentric circular plots with radii sizing 1-3 m (for simplicity). Axes show both absolute (above) and relative (below) coordinates (respectively X_{abs} , Y_{abs} and X_{rel} , Y_{rel} in **Eq. 2**).

Figure 2. Results of sensitivity analysis to select minimum numbers of simulations. Evolution for increasing radii of (a) mean $\widehat{GC}$ values and (b) their standard errors for $k =$ 10-2000 simulations.

Figure 3. *Criterion I.* Asymptotic representation showing the evolution of $\overline{GC}_{diff}$ (at each of the 79 original field plots) for increasing (a) plot sizes $s =$ 1-15 m radius (corresponding area also shown in upper axis) (b) and sample size $n =$ 1-50 number of trees (shortened to enhance visualization).

Figure 4. *Criterion II.* Absolute of correlation $|r|$ between GC values and selected ALS predictors (see legend, and explanations of ALS metrics in **Table 1** and section 2.4) for increasing (a) plot size $s = 1\text{-}15$ m radius (corresponding area also shown in upper axis) (b) and sample size $n = 1\text{-}90$ number of trees.

Figure 5. Minimum number of trees (sample size) to reach GC stabilisation in relation to the reference GC value obtained from the original field plot (GC_{ref}).

Figure 6. Changes due to varying ALS scan densities in the absolute of correlation $|r|$ between GC values and ALS predictors. See explanations of ALS metrics in **Table 1** (section 2.4).

Supplementary Materials

Supplementary Figure 1. Histograms showing the distribution of the response variable – $\overline{GC}$ (vertical bars) – and the predictor variables – *Skewness, Cover, CRR, P99, StdDev, P50* and *P25* (horizontal bars) –. The resulting scatterplots between each response-predictor pair are also shown. For simplicity, only results for the optimal plot radius $s^* = 9$ m are shown.

Table 1. Properties of the study area.

Parameter	Minimum	Mean	Maximum	SD
N (stems·ha^{-1})	467	1298	3025	594
G (m^2·ha^{-1})	14	25	44	7
QMD (cm)	10	17	29	4

N: stand density; G: basal area; QMD: quadratic mean diameter; SD: standard deviation.

Table 2. Summary of ALS metrics computed with FUSION and used in this research (McGaughey, 2015).

Symbol	Description	Forest Characteristics
P50	Median (i.e. 50th percentile)	Average tree height
StdDev	Standard deviation	Variation in tree heights
Skew	Skewness	Tree dominance
P25	1st quartile (i.e. 25th percentile)	Presence of understorey
P99	99th percentile	Dominant height
CRR	Canopy relief ratio = (Mean – Min) / (Max – Min)	Vertical structure
Cover	Percentage of all returns above 2 m	Canopy cover

Table 3. For each radii, proportion of the total number of original field plots within the $\overline{GC}_{diff} <$ 0.05 limit (*Criterion I*), and average number of trees contained within those plots.

Plot radius (m)	Ratio of original field plots reaching stabilization (%)	Average sample size of trees based on simulations
1	25.3	1.1
2	41.1	2.0
3	70.8	3.7
4	94.9	6.5
5	91.4	10.2
6	100	14.6
7	100	19.9
8	100	26.1
9	100	33.0
10	100	40.7
11	100	49.3
12	100	58.7
13	100	68.9
14	100	79.9
15	100	91.7

Table 4. Maximum absolute correlation between field GC and ALS predictors (*Criterion II*). See **Table 2** for description of ALS metrics.

ALS metric	Maximum correlation $\max\|r\|$	Optimal plot radius (s^*; m)	Optimal number of trees (n^*; m)
Skew	0.58	10	41
Cover	0.45	12	59
CRR	0.42	9	33

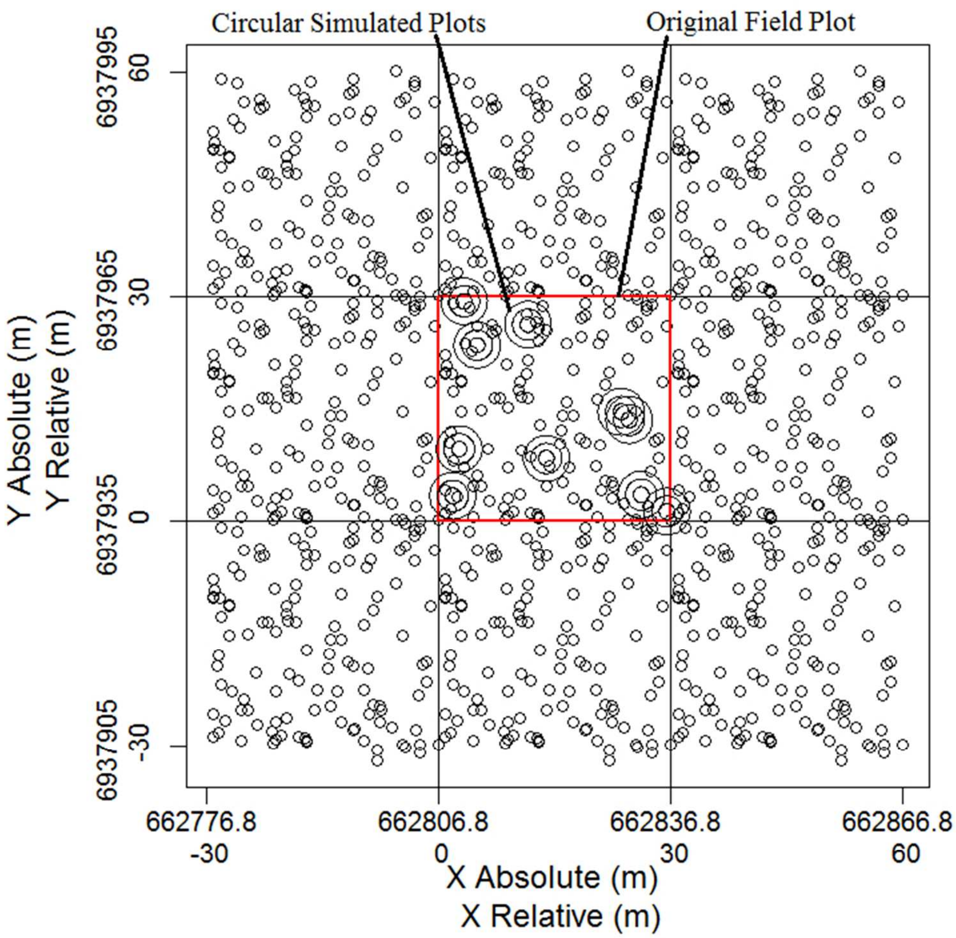


Figure 1

158x158mm (120 x 120 DPI)

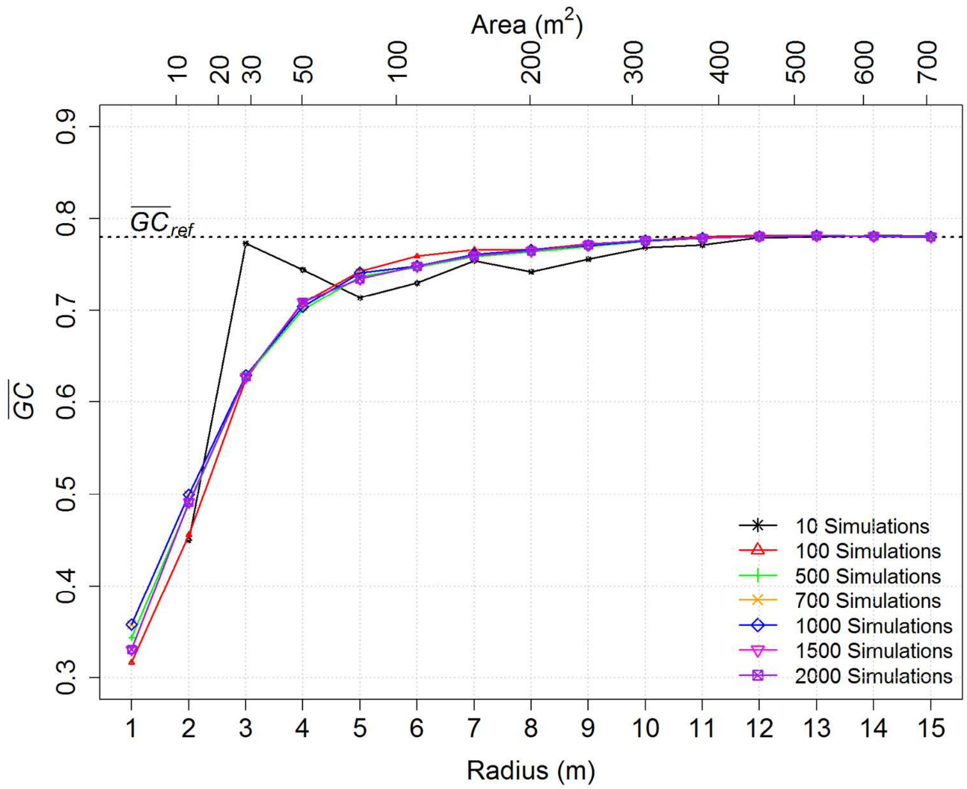


Figure 2(a)

241x190mm (200 x 200 DPI)

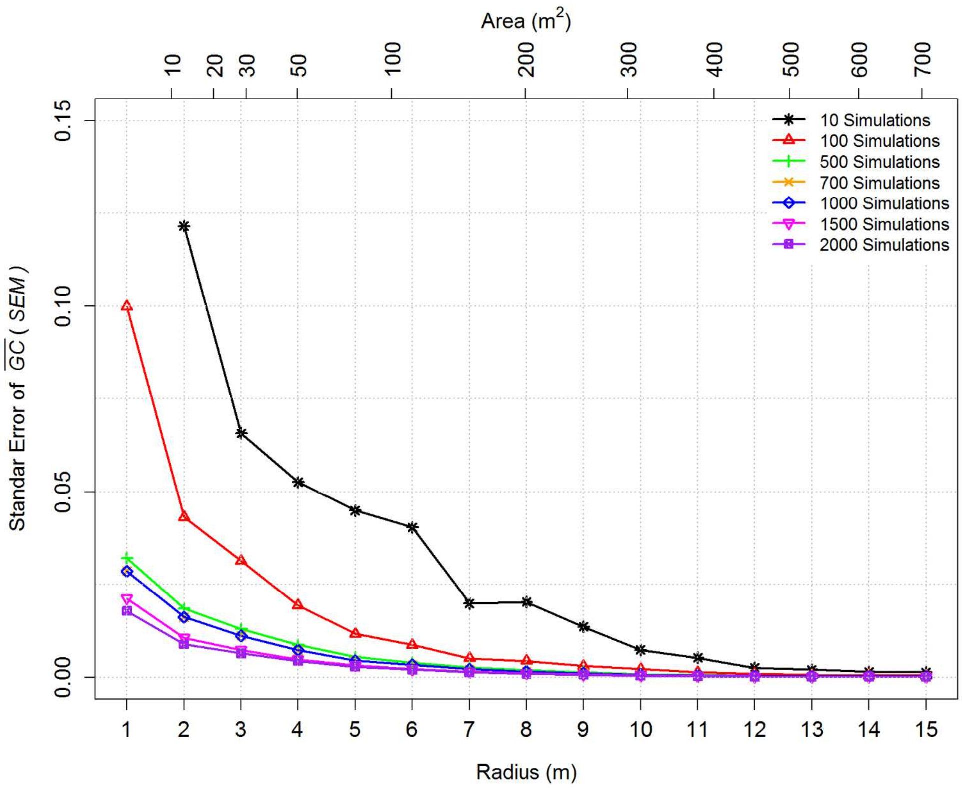


Figure 2(b)

241x190mm (200 x 200 DPI)

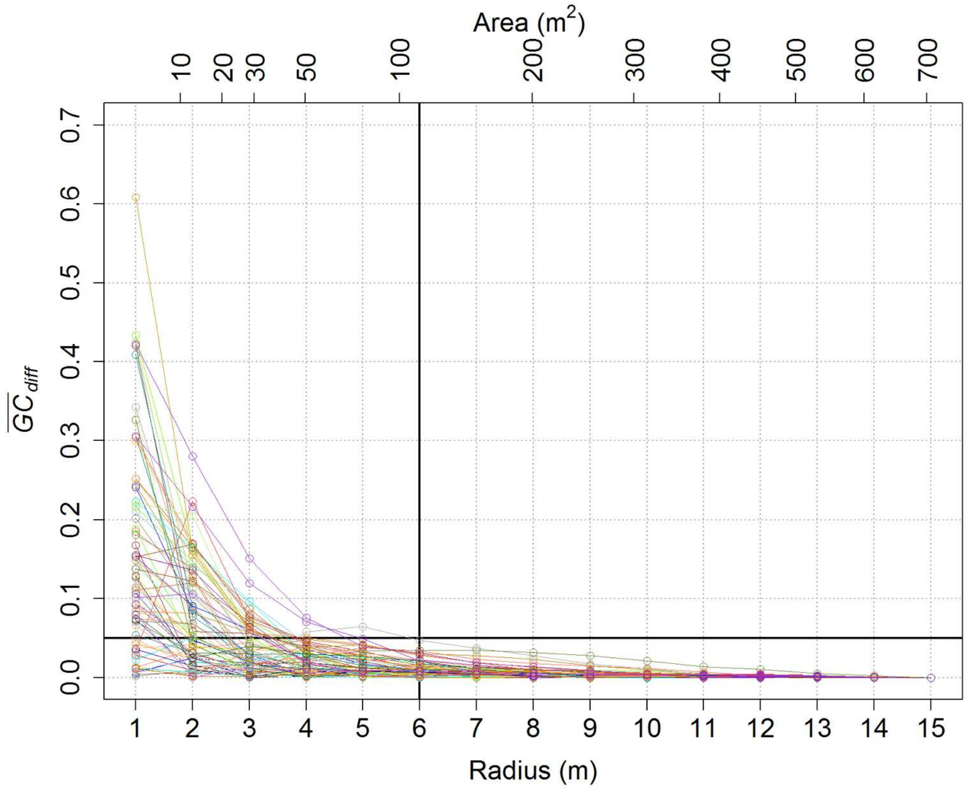


Figure 3(a)

241x190mm (200 x 200 DPI)

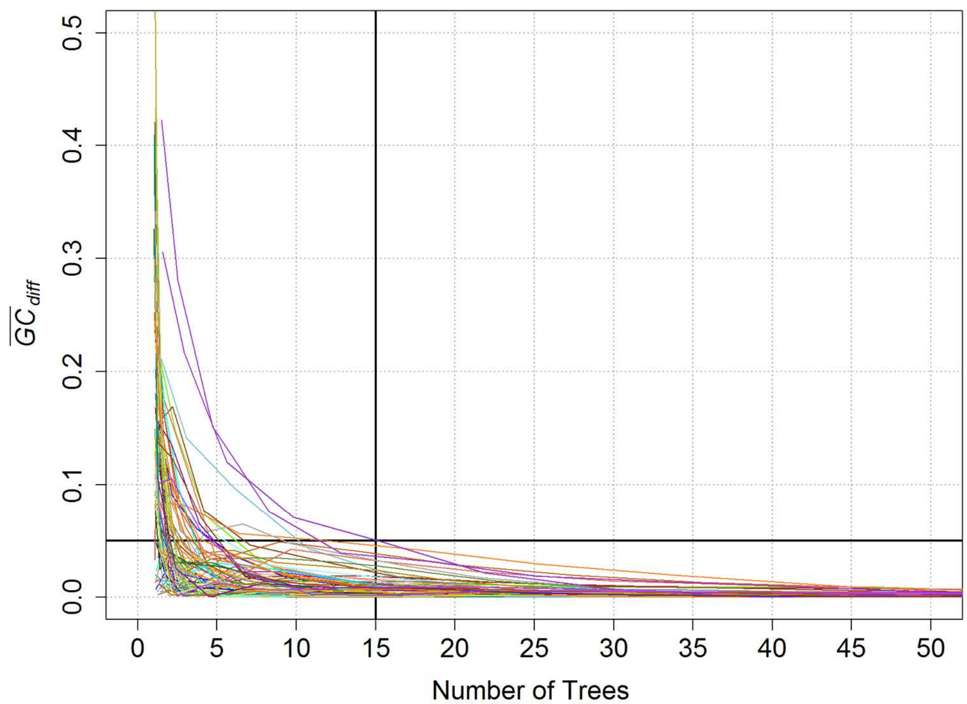


Figure 3(b)

241x190mm (200 x 200 DPI)

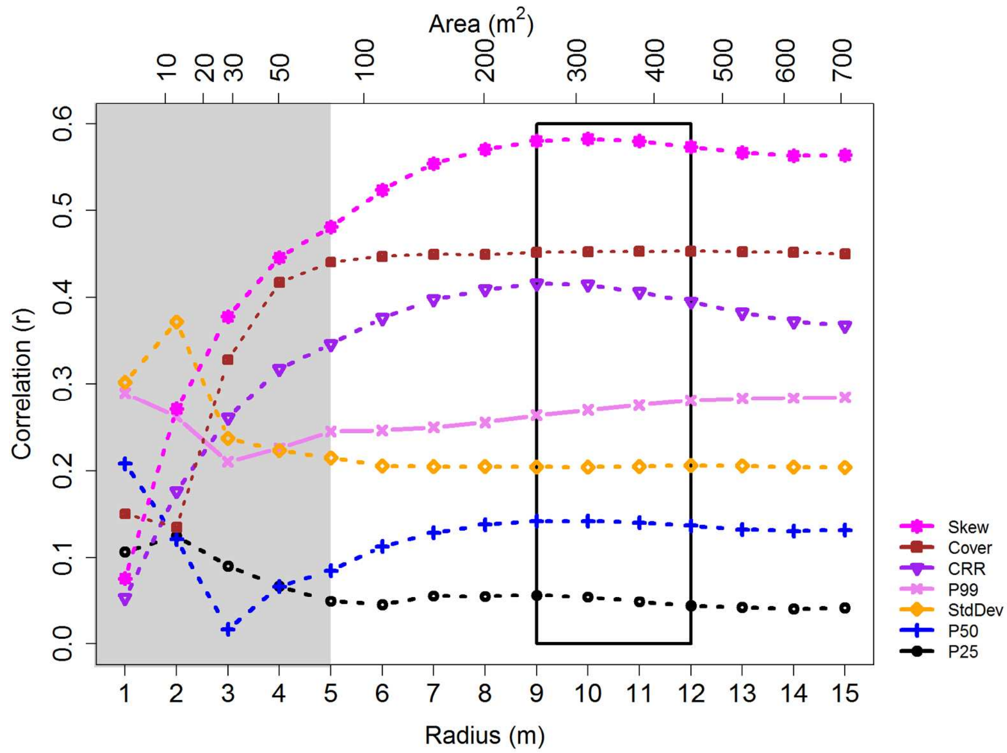


Figure 4(a)

241x190mm (200 x 200 DPI)

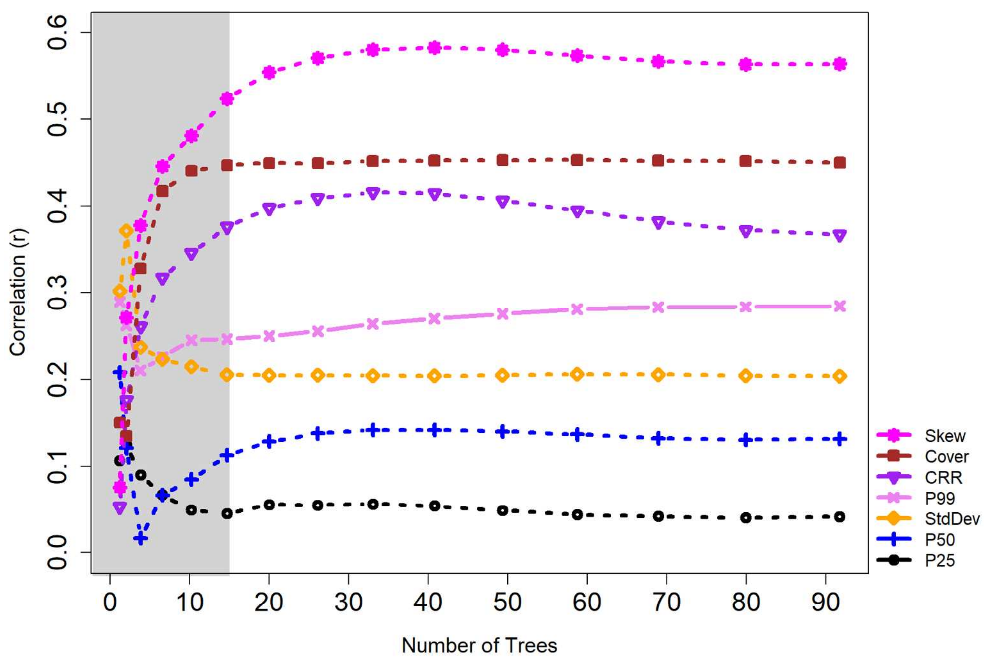


Figure 4(b)

241x190mm (200 x 200 DPI)

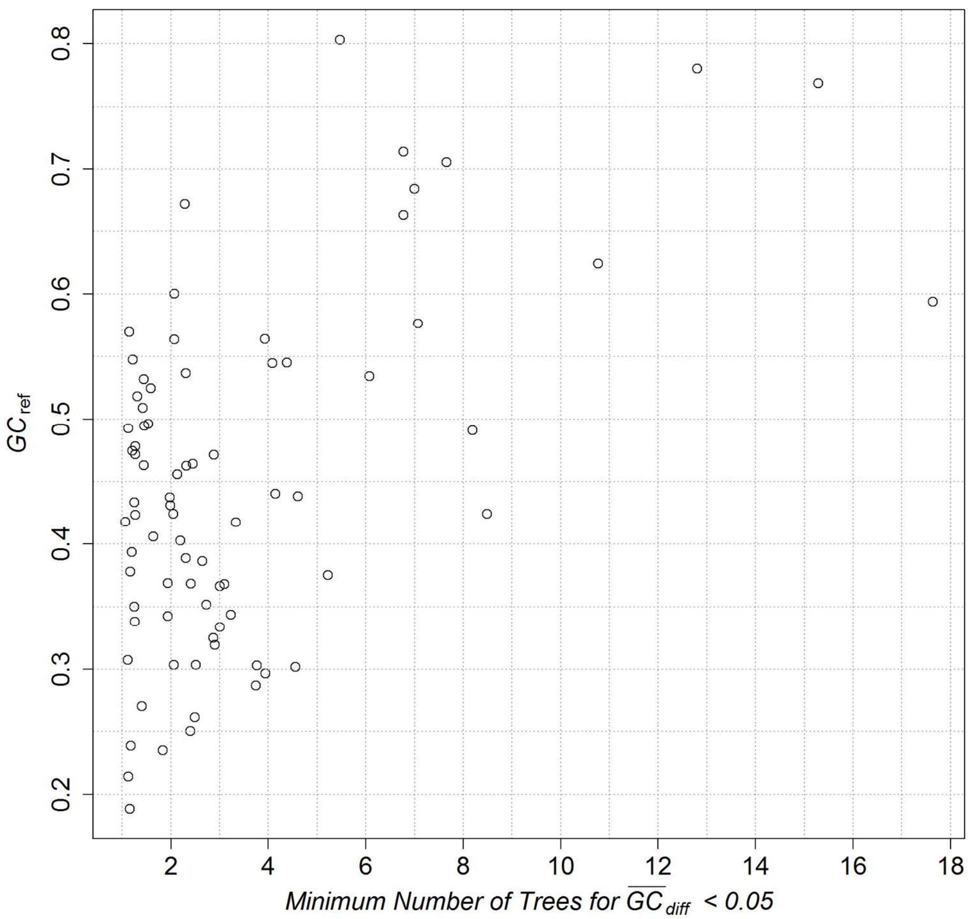


Figure 5

254x254mm (200 x 200 DPI)

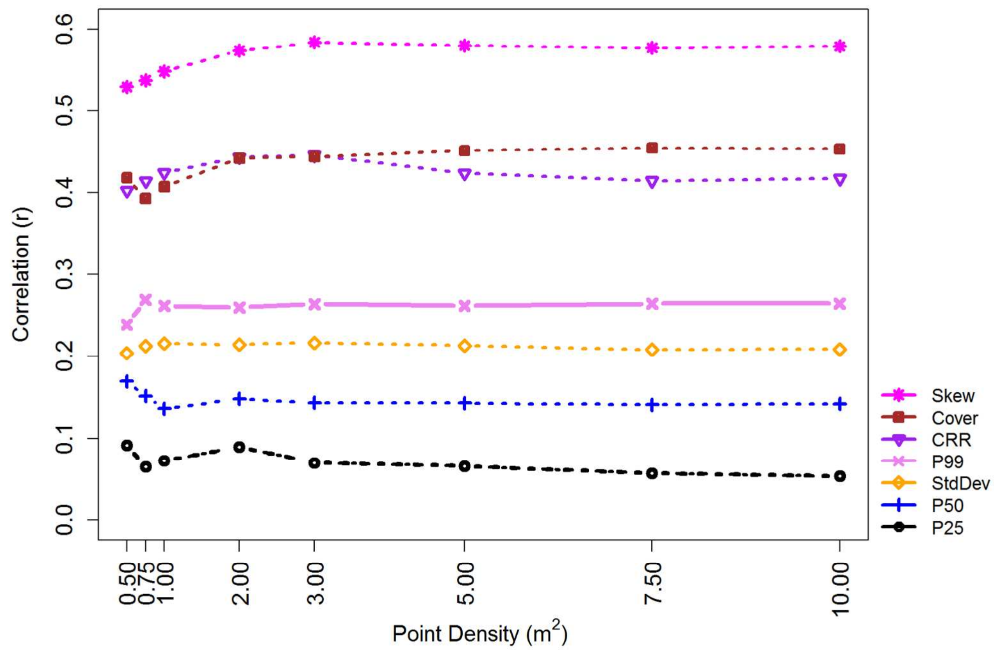
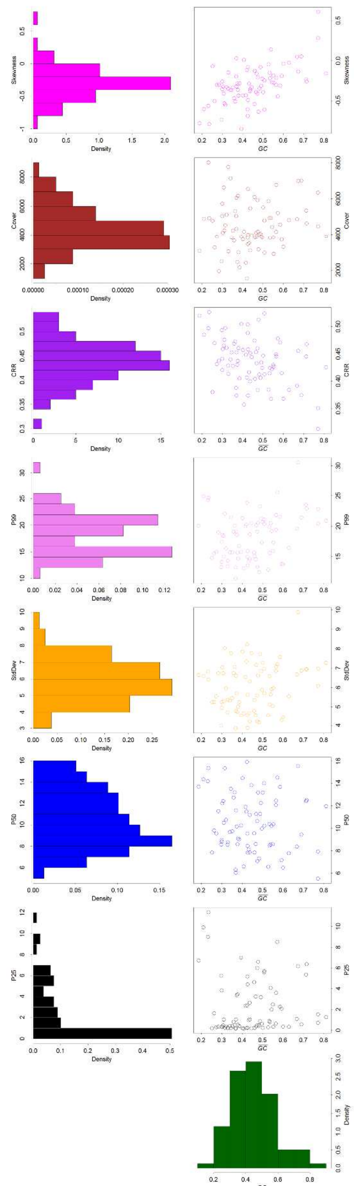


Figure 6

228x165mm (200 x 200 DPI)



558x1625mm (100 x 100 DPI)