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

SCHOLARONE™
Manuscripts

1   **Title:**

2   Effects of Plot Size, Stand Density and Scan Density on the Relationship between Airborne

3   Laser Scanning Metrics and the Gini Coefficient of Tree Size Inequality

4   **Authors:**

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

6

7   **Affiliations:**

8   (1) University of Eastern Finland. Faculty of Forest Sciences. PO Box 111 Joensuu,

9   Finland; adnan@uef.fi; matti.maltamo@uef.fi.

10   (2) University of Cambridge, Department of Plant Sciences. Forest Ecology and

11   Conservation. Downing Street, CB2 3EA Cambridge, UK. dac18@cam.ac.uk;

12   rv314@cam.ac.uk.

13   *Corresponding author.

14

15

16

17

18

19

20

21

22

23

24

25

26

27

1

28 **Abstract**

29 The estimation of Gini Coefficient (GC) of tree sizes using airborne laser scanning (ALS) can

30 provide maps of forest structure across the landscape, which can support sustainable forest

31 management. A challenge arises in determining the optimal spatial resolution that maximizes

32 the stability and precision of GC estimates, which in turn depends upon stand density or ALS

33 scan density. By subsampling different plot sizes within large field plots, we evaluated the

34 optimal spatial resolution by observing changes in GC estimation and in its correlation with

35 ALS metrics. We found that plot size had greater effects than either stand density or ALS

36 scan density in the relationship between GC and ALS metrics. Uncertainty in GC estimates

37 fell as plot size increased. Correlation with ALS metrics showed convex curves with maxima

38 at 250-450 m^2, which thus was considered the optimal plot size / spatial resolution. By

39 thinning the density of ALS point cloud, we deduced that at least 3 points·m^{-2} are needed for

40 reliable GC estimates. Many nationwide ALS scan densities are sparser than this, which may

41 be unreliable for GC estimation. Ours is a simple approach for evaluating the optimal spatial

42 resolution in remote sensing estimation of any forest attribute.

43 **Key words**

44 structural heterogeneity; spatial resolution optimization; sample size optimization; forest

45 structure; LiDAR

46

47

48

49

2

50   **1. Introduction**

51   *1.1 The Gini Coefficient as an Indicator of Forest Structural Heterogeneity*

52   Forest structural characteristics are widely used in the development of sustainable

53   management plans designed to protect habitats while carrying out forestry operations (Upton

54   and Fingleton, 1985; Pommerening, 2002; Motz et al., 2010; Vihervaara et al., 2015;

55   Valbuena et al., 2016). Management can be designed to emulate natural dynamics (Oliver and

56   Larson, 1990; Buongiorno et al., 1994; Lähde et al., 1999; Pukkala et al., 2016), by studying

57   how silvicultural operations affects forest structure locally (Humphrey et al., 2000; Valbuena

58   et al., 2013a; Robles et al., 2016).

59   Forest structure is often characterized by stem diameter distributions (O'Hara and Gersonde,

60   2004; McElhinny et al., 2005). If a single concise indicator of size inequality is desired, there

61   are many available, including Shannon or Simpson indices (Neumann and Starlinger, 2001;

62   Sterba and Ledermann, 2006; O'Hara et al., 2007; Lei et al., 2009) or variance-based metrics

63   (Staudhammer and LeMay, 2001). Recent research has highlighted the effectiveness of the

64   Gini coefficient (i.e. GC, Gini, 1921) for assessing the structural diversity (Lexerød and Eid,

65   2006a; O'hara et al. 2007; Duduman, 2009; Valbuena et al., 2012, 2013a). Originally

66   developed for evaluating inequality in income distributions (e.g., Hvistendahl, 2014), GC has

67   been applied to a variety of fields, such as healthcare (Asada, 2005) or land use (Zheng et al.,

68   2013). In plant sciences, it has been employed to evaluate size inequality (Weiner and

69   Solbrig, 1984). It has also been applied to forest ecosystems (Weiner and Thomas, 1986), to

70   quantify structural diversity (Knox and Peet, 1989), analyse competition (Lundqvist, 1994;

71   Cordonnier and Kunstler, 2015), or successional stages (Valbuena et al., 2013a). Comparative

72   studies indicate that GC is the best index for characterizing diameter distributions, providing a

73   logical ranking of different stand types (Lexerød and Eid, 2006a; Valbuena et al., 2012), so

74   that forest may be stratified according to their structure (Bollandsås and Næsset, 2007). It can

3

75   also be used to observe the effects of different management regimes (Bourdier et al., 2016;

76   Pukkala et al. 2016; Valbuena et al., 2016). For these reasons, estimation of GC is the focus

77   of this article.

78   When used in forest science, GC evaluates size inequality of trees growing in a vicinity

79   (Weiner, 1990). For a patch of forest containing n trees, within which the ith and j^{th} tree have

80   basal areas of g_i and g_j respectively, GC is calculated as (Glasser, 1962):

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

82   Therefore, GC is a statistical measure of relative dispersion, which is equivalent to half of the

83   relative mean absolute difference (Valbuena et al, 2017: appendix A3), and it ranges between

84   0-1, zero representing perfect equality and one being maximum inequality (Gini, 1921).

85   Hence GC describes the shape of tree-size distributions (Valbuena et al., 2016) and is

86   influenced by competitive interactions among trees (Cordonnier and Kunstler, 2015).

87   Valbuena et al. (2012) demonstrated that the $GC = 0.5$ can be considered as a boundary

88   between even-aged and uneven-aged stand structures. GC values far below 0.5 indicate a

89   unimodal "normally distributed" size structure, which are commonly found in even-aged

90   stands that are self-thinning (e.g. Coomes and Allen, 2007). Values close to 0.5 indicate

91   irregular size distributions (Duduman, 2009), while values much greater than 0.5 represent

92   "reverse-J" stand structures (Valbuena et al., 2013a).

93   *1.2 Influence of Plot Size in Measurements of Forest Structure*

94   Sample plots used for measuring plant communities are usually rectangular or circular in

95   shape (Whittaker, 1972; Kent and Coker, 1992), with a wide range of possible plot sizes from

96   fine to coarse scales (Chytrý and Otýpková, 2003). As the effects of plot size decrease with

97   increasing size of a plot (David and Mishriky, 1968; Barbeito et al., 2009), an optimal size

4

98   must be chosen compromising the acquisition of a field plot large enough to obtain a stable

99   measure of forest structure, but no larger than necessary because of the costs involved

100   (Otypková and Chytry, 2006). Structural diversity depends on the spatial resolution at which

101   an index is evaluated (Lexerød and Eid, 2006b). Varying the scale of observation may

102   therefore distort the information retrieved from an indicator (Chen and Crawford, 2012;

103   Mauro et al, 2016). As plot size increases, GC estimates may be more reliable, but also

104   fundamentally different stand conditions may aggregate (Coomes and Allen, 2007).

105   Therefore, interpretation of data analysed at different scales remains one of the most

106   challenging tasks in spatial statistics (Gotway and Young, 2002), as shown in the context of

107   agriculture (Smith, 1938), sociology (Hannan, 1971), and environmental sciences (Jelinski

108   and Wu, 1996). Also, the spatial distribution of trees has a practical effect on plot size, since

109   clustered patterns require larger plot sizes to obtain a same degree of confidence in estimates

110   (Upton and Fingleton, 1985; Pommerening, 2002; Kallimanis et al., 2008; Motz et al., 2010).

111   Recently, Magnussen et al. (2016) suggested a method of upscaling to a common plot size to

112   minimize scale effects in survey estimates, which achieved consistency among the quantiles

113   and proportions of sampling distributions of forest attributes.

114   *1.3 Influence of ALS Scan Density in Measurements of Forest Structure*

115   Airborne laser scanning (ALS) is recognised as a highly effective tool for regional

116   monitoring because it provides precise information about biophysical stand properties,

117   (Gobakken et al., 2006; Gobakken and Næsset, 2008). The GC may be calculated as a

118   descriptor of the distribution of ALS heights (Valbuena et al., 2017), or ALS metrics may be

119   related to GC of tree sizes (Valbuena et al., 2013b). The spatial resolution of ALS data used

120   in area-based methods has an effect on estimated values (Mascaro et al., 2011). In the context

121   of remote sensing-assisted forest estimations, spatial resolution refers not only to the size of

122   field plots but also to the size of pixels at which auxiliary variables are computed (Gobakken

5

123   and Næsset, 2008; Ruiz et al., 2014; Valbuena et al., 2016). In ALS-assisted estimations of

124   GC of tree size inequality, there is a lack of knowledge on the effects of varying plot size and

125   spatial resolution.

126   Scan density is one of the most important aspects of ALS datasets that affects both processing

127   and costs (Balsa-Barreiro and Lerma, 2014; Kandare et al., 2016). The importance of

128   optimizing ALS point density lays in its trade-offs against ALS swath width, and hence costs

129   (Baltsavias, 1999). Liu et al. (2007) observed that density reduction influenced the accuracy

130   of digital terrain models (DTM) due to the presence of empty space intervals between points.

131   A reduction in DTM accuracy may affect the calculation of metrics describing ALS height

132   (Ruiz et al., 2014; Singh et al., 2015), although it would be unlikely to affect metrics

133   describing their dispersion, such as GC. Gobakken and Næsset (2008) assessed the effect of

134   point density on biophysical stand properties, finding that maximum height was the least

135   affected metric and suggesting to avoid metrics most affected by point density. No previous

136   studies have yet determined how stand density and ALS scan densities affects GC estimates

137   from ALS.

138   *1.4 Objectives*

139   The aim of the study is to evaluate the effects of plot size and ALS scan density on field and

140   ALS-derived estimates of GC in the boreal forests of Finland. We developed a simple method

141   for selecting the optimal plot size for determining the GC of tree size inequality from field

142   data, and for predicting GC reliably using ALS metrics as auxiliary variables.

143   **2. Material and Methods**

144   *2.1 Study Area and Field Data Collection*

145   The study was carried out in a typical boreal managed forest located in Eastern Finland (62°

146   31′ N, 30° 10′ E). Scots Pine (*Pinus sylvestris* L.) is the dominant species which represents

147   73% of the total wood volume, while Norway spruce (*Picea abies* Karst.) represents 16%,

148   and deciduous species 11% of the total wood volume (Valbuena et al., 2014). The main

149   properties of the field data such as stand density (N), basal area (G) and quadratic mean

150   diameter (QMD) are shown in **Table 1**. The field data were collected in May-June 2010 and

151   consisted of 79 squared plots (henceforth "*original field plots*") of various dimensions

152   (20×20, 25×25 or 30×30 m, the smaller ones being in denser stands). With the intention of

153   representing the contrast between highly homogeneous even-aged areas and more

154   heterogeneous forest structures (Valbuena et al. 2016), forest stands were determined using

155   stratified random sampling, whereas plot locations were purposively selected. After choosing

156   the sampled stands, plots were located within the stands at a representative location. The

157   reason for doing this was to avoid plot locations at stand borders and the high cost and

158   measuring effort required to record the location of all individual stems within the plot. The

159   absolute positions of every individual tree with $dbh > 4$ cm and tree top height taller than 4 m

160   were mapped using an approach combining ALS and field surveying methods suggested by

161   Korpela et al. (2007). Before the field measurement, a map of individual tree positions was

162   generated from high density ALS data (see below) using an individual tree detection (ITD)

163   method (Packalen et al., 2013). Actual positions of trees defined by their longitude/latitude

164   coordinates (X_{actual}, Y_{actual}) were verified in the field, while the location of trees not

165   detected by the ITD method were measured relative to the ITD-derived ones (distances and

166   bearings) and least-square adjusted (Korpela et al., 2007).

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

168   *2.2 Simulation of Circular Plots*

7

169   Preliminary tasks for the simulation included transformations into relative coordinates, the

170   correction of edge effects and a sensitivity analysis to determine the number of simulations

171   needed. Then, within each *original field plot* we simulated circular plots at random positions.

172   Circular plots were chosen on the assumption that tree competition is the same in all spatial

173   directions. The radius of these *circular simulated plots* was increased in 1-m intervals,

174   generating concentric circles up to 15 m-radius. Since the position of individual trees were

175   available from the original field data, we could extract the trees located within each circular

176   simulated plots, computing an estimation of GC based on tree dbh. Likewise, the position of

177   individual ALS returns located within each simulated circular plots could be extracted, using

178   them to compute ALS metrics commonly employed in area-based estimation methods.

179   *2.2.1   Transformation to Relative Distances and Edge Correction*

180   Transformation of absolute tree coordinates into relative coordinates requires procedures of

181   plot rotation and translation (Matos, 2014). Since in the case of our study the edges of

182   original field plots were coincident with the UTM grid, there was no need to carry out plot

183   rotations. In plot translation absolute coordinates of original field plots were modified into

184   relative distances, by assigning the origin of axes (0, 0) to the south-western corner of the

185   original field plot. Absolute coordinates of south-western corner (X_{corner}, Y_{corner}) were

186   subtracted from the absolute coordinates of each tree (X_{abs}, Y_{abs}) to get their relative

187   coordinates (X_{rel}, Y_{rel}).

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

189   Moreover, Pommerening and Stoyan (2006) showed that edge effects play an important role

190   in spatial statistics. Because the immediate neighbour trees outside the boundary of the

191   *original field plots* were not measured, ignoring them would result in biased statistical

192   estimations. Thus, indices based on tree positions require an edge correction method to

193   reduce this bias. We chose a periodic boundary edge correction method (Diggle, 2003), since

194   Pommerening and Stoyan, (2006) found it to be superior to other alternatives. This method

195   consisted of replicating the same spatial pattern measured in the field around the *original*

196   *field plot* (**Fig. 1**). Concentric *circular simulated plots* randomly positioned at the edge of the

197   *original field plots* therefore also included the trees positioned out of the boundaries of the

198   *original field plots*.

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

200   *2.2.2   Plot Simulation and Sensitivity Analysis*

201   A pilot sensitivity analysis was done with the intention to identify the minimum number of

202   simulations within an original field plot which can guaranteed a stable and robust outcome

203   for the simulation. We selected the *original field plot* with highest GC, hence likely the one

204   most sensible to changes among different simulations, and repeated the analysis for 10, 100,

205   500, 700, 1000, 1500 and 2000 simulations. A position (X_{sim}, Y_{sim}) was randomly

206   simulated within the *original field plot*, and GC was calculated for each *circular simulated*

207   *plot* (see below) and for each plot radius (s ; m) (1-m intervals from 1 to 15 m) (**Fig. 1**). As

208   explained below, the standard error of the mean (SEM) of values obtained for GC at each

209   radius were considered in order to fix the minimum number of simulations at which no

210   considerable improvement was observed by adding further replications. After setting the

211   necessary number of simulations to a fixed number k based on the pilot sensitivity analysis,

212   the whole procedure was repeated for the remaining 78 original field plots. Relative and

213   absolute positions of all simulations were recorded so that they could later be used for

214   extracting their corresponding ALS returns as well.

215   2.3 *Gini Coefficient Estimation*

216   The target was to calculate sample estimations of the GC describing the size inequality of the

217   tree community represented at each *original field plot*. Its estimation (**Eq. 1**) was repeated for

218   every concentric *circular simulated plot* of radii 1-15 m, and for all the simulated

219   positions (X_{sim}, Y_{sim}). For this purpose, basal area (g; m^2) was calculated for each individual

220   stem. Differences in g were computed for each pair of trees within each circular simulated

221   plot. GC is the average of absolute differences relative to their mean ($\bar{g}$) (see detailed

222   descriptions of GC calculation in Lexerød and Eid (2006a) and Valbuena et al. (2013b)). The

223   reason of using g instead of dbh was to increase the influence of larger trees (Solomon and

224   Gove, 1999). The unbiased estimator by Glasser (1962) was employed because it is

225   appropriate for an estimation based on a finite number of trees n located within each circular

226   simulated plot (**Eq. 1**). The mean GC ($\overline{GC}$) and its corresponding SEM were computed for

227   each radius (from 1 to 15 m), and for each of the *original field plots*. SEM is a measure for

228   the accuracy of those means, accounting for the variability between the samples, according to

229   the number of simulations k and their sample standard deviation (SD). R statistical software

230   (R Development Core Team, 2016) was used for all these calculations and statistical

231   analyses.

232   We constructed a graph comparing $\overline{GC}$ results for increasing plot size s for all *original field*

233   *plots*. The GC value at *circular simulated plots* must necessarily approximate asymptotically

234   to the value of GC for the entire *original field plot* as the radius of circular simulated plots

235   increases (Matos, 2014). For this reason, the value of GC obtained by applying equation (1) to

236   the original field plot was used as a reference (GC_{ref}). In order to make all the simulated GC

237   values directly comparable, we calculated absolute GC differences (GC_{diff}) by subtracting

238   simulated GC values from the GC_{ref}:

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

240   This way, it was possible to analyse the difference of each simulated GC to its corresponding

241   asymptotic value, allowing to set a common criterion to evaluate all simulations based on the

242   stabilization of the estimated GC value (see below).

243   2.4 *Airborne Laser Scanning Data and Metric Computation*

244   ALS data was acquired on June 26, 2009 using ATM Gemini sensor (Optech, Canada) from

245   600-700 m above ground level with a 26° field of view. Scan swath was 320 m wide with a

246   55% side overlap between the strips. A high resolution dataset with 11.9 pulses·m^{-2} scan

247   density was produced from a pulse rate of 125 kHz. Details about the processing of ALS data

248   are described in Packalen et al. (2013). The last echoes were classified as ground and

249   interpolated into a DTM (Axelsson, 2000). The spatial resolution of DTM was 0.5 m based

250   on the scan density, and the height above ground of individual ALS returns was obtained by

251   subtraction of the DTM height beneath each of them. Echoes lower than 0.1 m from ground

252   level were eliminated, as they were considered to be reflected from ground.

253   Individual ALS returns of each circular simulated plot based on its absolute

254   coordinates (X_{sim}, Y_{sim}) were clipped, and area-based ALS metrics were computed from their

255   heights with the help of FUSION software (USDA Forest Service; McGaughey, 2015). ALS

256   metrics are statistics and descriptors of the distribution of ALS heights observed within a

257   given area, which are usually employed as auxiliary variables in ALS-assisted forest

258   estimations (**Table 2)**. Some of these metrics were common statistics as, for example, the

259   mean (*Mean*) standard deviation (*StdDev*) or the skewness (*Skew*) of the distribution of

260   heights above ground of ALS returns contained within each *circular simulated plot*. We also

261   computed the percentiles of their distribution, such as the 25th (*P25*), 50th (*P50*) or 99th (*P99*).

262   In addition, we calculated the so-called canopy cover metrics (McGaughey, 2015), such as

263   the proportion of returns backscattered from 0.1 m above the ground (*Cover*). Another metric

11

264   included in FUSION was the canopy relief ratio (*CRR*), which is the difference between

265   mean and minimum ALS return heights divided by a difference between maximum and

266   minimum heights (Pike and Wilson, 1971).

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

268   The effect of plot size in the relationship with GC was studied separately for each of these

269   ALS metrics. For each radius, we gathered all the simulations carried out at all the *original*

270   *field plots* and calculated all the ALS metrics listed in **Table 2**. They were used to calculate

271   Pearson correlation coefficients (r) using all the combinations of field GC against each ALS

272   metric. Then, we observed separately for each ALS metric the evolution of r when increasing

273   the plot size s of the *circular simulated plots*. Since we were only interested in the capacity of

274   the ALS metrics to explain variability in GC, regardless of whether their relationship was

275   direct or indirect, we considered the absolute value of the correlation coefficient $|r|$ in the

276   optimization, as explained below.

277   *2.5 Basic Relationships*

278   The plot size and spatial resolution at which an ALS-assisted estimation is carried out relates

279   intrinsically to the sample size used in all calculations. Sample size affects the relationship

280   between predictor and response, and therefore the accuracy of ALS estimation of any forest

281   attributes (Gotway and Young, 2002; Mascaro et al., 2010; Næsset et al., 2015; Magnussen et

282   al., 2016; Valbuena et al., 2016). In this context, sample size refers both to the number of

283   trees used to calculate a given forest attribute, GC in this case, but also to the number of ALS

284   returns involved in the computation of ALS metrics. The link between resolution and sample

285   size is employed on the empirical densities of the datasets, i.e. stand density (N; trees·ha^{-1}) or

286   ALS points density (d; points·m^{-2}) (Gobakken and Næsset, 2008; Motz et al., 2010;

287   Jakubowski et al., 2013). Therefore, the effects of plot size and spatial resolution of the ALS

288   estimated forest attributes also depend on N and d, and the combined effects of these two

289   factors may explain why plot sizes suitable for field surveys may be found sub-optimal for

290   ALS estimation (Næsset et al., 2015).

291   Hence, the relationship between the radius s of a circular plot and the number of trees (n)

292   contained within is tied to the N at the location of the plot.

293                              $$n = N\pi s^2$$ (4)

294   This begs the question on whether the optimization method should search for an optimal plot

295   radius (s^*; m) or an optimal sample size (n^*). In a forest environment of variable stand

296   density N (**Table 1**), does the relationship between GC and ALS metrics depends on the plot

297   size used, or on the number of trees surveyed? In order to research whether it makes a

298   difference, we repeated the same procedure for both s^* and n^* optimization. In other words,

299   we tested the results of optimization according to either plot radius or number of trees. In any

300   of the cases, the relationship in **eq. (4)** assures that the methodology can be replicated for

301   either dense or sparse forests, since s and n can always be deduced from one another by an

302   empirical N.

303   Likewise, a similar relationship holds between the size of that same circular plot and the

304   number of ALS returns backscattered from it, according to a given ALS scan density d. In

305   this context of estimation using auxiliary variables, the scale concerns both to the size of the

306   field plots and the spatial resolution of the pixel at which ALS metrics are calculated.

307   Therefore, the number of ALS points (p) relates to the spatial resolution / plot size used (s)

308   according to d :

309                              $$p = d\pi s^2$$ (5)

310   As before, the relationship in **eq. (5)** assures that the methodology can be replicated for any

311   range of ALS scan densities, since s and p are trivially deducted from one another by an

312   empirical d. As an overall conclusion, a given optimal plot size s^* necessarily implies

313   optimal sample sizes as well, both n^* and p^*. Keeping these relationships in mind is key to

314   demonstrating the validity of the optimization method for replication elsewhere according to

315   the N and d which may occur at any other study cases, and therefore the method is equally

316   valid for both dense and sparse forests and ALS surveys with low or high scan density.

317   *2.6 Plot Size Optimization*

318   To optimize the plot size which should be used for a reliable GC estimation, and thereby also

319   the optimal spatial resolution for an estimation of GC from ALS datasets, we determined two

320   criteria to be applied sequentially: (1) stabilization of GC as estimator of the population value

321   from the field information itself, and (2) maximizing the GC variability explained by ALS

322   metrics. Therefore, *Criterion I* considered the minimum plot radius at which the estimation of

323   GC remained stable to further increases in plot size. *Criterion II* was set to optimize the ALS-

324   assisted estimation, by observing changes in the correlation between the field GC and each

325   ALS metric among the simulated plot radii.

326   *Criterion I* was implemented by observing the evolution of $\overline{GC}_{diff}$ for increasing radii at

327   every original field plot. We set a maximum value of $\overline{GC}_{diff} = 0.05$ at which it was

328   considered that the estimation of GC was stable and representative of the population, and,

329   therefore, selected the minimum plot radius s as the smallest meeting the first criterion for all

330   the 79 original field plots.

331   *Criterion II* consisted in maximizing the explained variance in the GC values when predicted

332   from ALS metrics. To implement this criterion we combined all the GC/ALS metric pairs for

333   all the simulations carried out at all the original field plots, and grouped them according to

334   the different simulated radii. The optimal radius was set to be that one showing the maximum

335    $|r|$ value for a given metric. To make an overall decision, we put the focus on those metrics

336   showing higher correlations, and decided a range of optimal sizes accordingly (since the

337   empirical maximum may differ for different ALS metrics). As a summary, the final optimal

338   plot size s^* for a given metrics was:

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

340   *2.7 Sample Size Optimization*

341   For sample size optimization, seeking to deduct what is the minimum number of trees needed

342   to obtain a reliable GC estimation, and the optimum for its ALS prediction, we applied the

343   same two sequential criteria employed for plot size optimization (section 2.6). Therefore, the

344   simulations were similar as before, but they increased the size of simulated circular plots

345   according to the resulting number of trees n instead of plot radii. Thus, for implementing

346   *Criterion I*, the evolution of $\overline{GC}_{diff}$ was observed for increasing number of trees n, also

347   setting a maximum value of $\overline{GC}_{diff} = 0.05$. As before, we selected the minimum n as the

348   smallest meeting *Criterion I* for all 79 original field plots. *Criterion II* also consisted in

349   maximizing the absolute correlation between the GC values and each of the ALS metrics.

350   New values of $|r|$ were obtained for increasing values of n, and the final optimal sample size

351   (n^*) for each given ALS metric was then set as:

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

353   Finally, we compared which alternative, **eq. (6)** or **(7)**, would be more convenient for a

354   practical plot size optimization, discussing the results obtained by either method.

355   *2.8 Reduction of ALS Point Density*

356   Once deduced an optimal spatial resolution s^*, we also investigated the effects of varying

357   ALS scan density d. The original point density was reduced to 0.5, 0.75, 1, 3, 5, 7.5, and 10

358   points·m^{-2}. A common option to reduce point density is by moving a 1 m window and

359   selecting random points from the point cloud to reach the desired point density (e.g.,

360   Magnussen et al. 2010). We calculated a correct thinning factor for each desired point density

361   d (Ruiz et al., 2014), following the method detailed by Jakubowski et al. (2013) which

362   incorporates routines included in *LAStools* (RapidLasso GmbH Inc.; Isenburg, 2016). New

363   ALS metrics over each of the k simulated circular plot positions and their correlations against

364   the GC values obtained from the field information were calculated, and the entire procedure

365   was repeated for all the reduced densities. In a similar manner as it was done for s and n, the

366   effects of varying ALS scan density were studied by observing the changes in $|r|$, i.e. the

367   effects in the relationship between the GC of tree size inequality and the ALS metrics with

368   more explanatory capacity towards this given forest attribute.

369   **3. Results**

370   *3.1 Establishing the Number of Simulations*

371   **Figure 2** shows the results of sensitivity analysis carried out to select the minimum number

372   of simulations that would yield a robust estimation of GC for increasing simulated plot radii.

373   As expected, the GC value estimated from few simulations fluctuated considerably, and this

374   fluctuation decreased as the number of simulations increased (**Fig. 2a**). The expected general

375   trend toward the asymptotic value obtained by the entire population (GC_{ref}) was generally

376   observed in **Fig. 2a**. Very little variation in GC estimates were observed when the number of

377   simulations increased from 700. Similarly, the SEM decreased as the number of simulations

378   increased (**Fig. 2b**), remaining virtually unchanged from 700 to 2000 simulations.

16

379   Consequently, we decided to carry out the analysis using $k = 700$ simulations of 15

380   concentric circular simulated plots located within each 79 original field plots.

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

382   *3.2 Plot Size Optimization*

383   **Figure 3a** shows the resulting $\overline{GC}_{diff}$ for each of the 79 original plots, and **Table 3** is a

384   summary of these results which was used for establishing *Criterion I*, which set the minimum

385   plot size that would provide a reliable GC estimation for the population. Circular simulated

386   plots of small sizes provided GC estimates that differed considerably from the population

387   values as considered by GC_{ref}. Nonetheless, once the estimation reached stabilization, an

388   increase in the radius of a circular plot (and hence the sampling effort) would not necessarily

389   imply a considerable change in the estimation of GC (**Fig. 3a**). Our results showed that only

390   few of the original field plots (probably very homogeneous stands) obtained stable GC

391   estimations from very small circular simulated plots (**Table 3**). On the other hand, for larger

392   circular simulated plots the differences against the original field plots representing the

393   population became negligible. We observed that stabilization of the GC estimation started

394   beyond of simulated plot radius $s = 6$ m, from which all the original field plots fell within

395   the $\overline{GC}_{diff} < 0.05$ limit. We therefore established that the smallest plot size required for a

396   reliable GC estimation should be set at areas sizing around 113 m^2.

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

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

399   With regards to *Criterion II*, the evolution of $|r|$ with increasing plot size was observed for

400   all ALS metrics included in FUSION. Results showed that changes in the relationship

401   between the field GC of tree sizes and metrics describing the distribution of ALS return

17

402   followed some general trends and patterns. For this reason and for simplifying, we chose to

403   show only few ALS metrics in **Fig. 4a**, which we considered representatives of the general

404   trends observed. These ALS metrics were the described *P25, P50, P99, Skew, StdDev, Cover*

405   and *CRR* (**Table 2**). **Fig. 4a** showed an erratic fluctuation for the values of $|r|$ obtained for

406   plot sizes smaller than a radius $s = 5$ m , which was possibly caused by the instability

407   observed in the GC estimation at smaller plot sizes (**Fig. 3**). For this reason, we shadowed this

408   area in grey colour in **Fig. 4**, denoting that such small plot sizes were already dismissed under

409   *Criterion I*. Once GC estimation reached stabilization, its correlation to ALS metrics often

410   yielded a convex curve as plot size increased (**Fig. 4a**). Therefore, the optimal plot size was

411   possible to determine via maximization of $|r|$. This tendency was more clearly marked for

412   those ALS metrics showing higher values of $|r|$, i.e. more correlated to the GC of tree sizes

413   (**eq. 2**), such as *Skew, Cover* or *CRR*. For other ALS metrics less related to GC, like return

414   height percentiles (*P25, P50* or *P99*) or *StdDev*, this tendency was less marked (**Fig. 4a**). For

415   the optimization of plot size, we selected those metrics showing highest correlation

416   against GC, since in practice they would be those more involved in its estimation. **Table 4**

417   shows that the maximum $|r|$ for ALS metrics *Skew, Cover* or *CRR* ranged $s^* = 9\text{-}12$ m plot

418   radius (the quality of histograms and scatterplots between variables involved can be checked

419   in the **Supplementary Material**). It can be observed in **Fig. 4a** that beyond a circular

420   simulated plot of 12 m the correlation showed a decreasing trend for most ALS metrics. Also,

421   local maxima may be found for some ALS metrics for very small plot sizes, which is

422   probably an artefact due to the above-mention instability in GC estimation at very small plot

423   sizes (**Fig. 3**). This proved the necessity of imposing *Criterion I* as a prior step to correlation

424   maximization. As a conclusion, under the established combined *Criteria I* and *II*, we

425   determined that any plot radius $s < 6$ m (113 m^2 area) should be avoided (denoted by grey

426   colour in **Fig. 4a**), and the optimal plot size for an ALS-assisted estimation of GC must be

427　carried out using scales sizing 250-450 m^2, which concerns to both the size of the field plot

428　and the pixel of the grid employed for ALS estimation.

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

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

431　*3.3 Sample Size Optimization (Stand Density Effect)*

432　On the other hand, **Figure 3b** shows the evolution of $\overline{GC}_{diff}$ for increasing sample sizes

433　(number of trees n) at each of the 79 original field plots. It is worth mentioning the **Figs. 3a**

434　and **3b** relate to one another according to **eq. (4)**. As a consequence, a similar tendency can

435　be found for both of them. **Table 3** expresses the number of trees that correspond on average

436　to a given sample size. Therefore, the minimum value obtained for *Criterion I* in plot size

437　optimization, $s = 6$, corresponds to stating that a minimum number of $n = 15$ trees are

438　required for a stable GC estimation (shaded area in **Fig. 4b**). We nevertheless further

439　postulated that this minimum number of trees may be dependent on the heterogeneity of the

440　forest itself, being possibly larger in the presence of higher inequality of tree sizes. This

441　presumption was demonstrably true, as it can be observed in a scatterplot comparing the

442　minimum number of trees required for a stable GC estimation at each of the 79 original plots

443　against their overall value of tree size inequality observed (GC_{ref}; **Fig. 5**). Such relationship

444　was not so straightforward if *Criterion I* was imposed using s instead (results not shown),

445　which demonstrates the effect of varying forest stand density N. Hence, obtaining a stable GC

446　estimation is more dependent of measuring a minimum number of trees than imposing a

447　given size for the field plot used.

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

449   The case for *Criterion II* was different, as it can be deducted when observing the same ALS

450   metrics employed to optimize s – *P25, P50, P99, Skew, StdDev, Cover* and *CRR* –, but trying

451   to optimize n instead (**Fig. 4b**). Again, a similar tendency can be found since **Figs.4a-b** are

452   also related by **eq. (4)**. Results were therefore very similar whether optimization was carried

453   out according to plot size (**eq. 6**) or sample size (**eq. 7**). The values of $|r|$ also followed a

454   convex curve when increasing the number of trees measured, and an optimal sample size n^*

455   could be reliably determined via $|r|$ maximization. Our results showed that a number of trees

456   approximately ranging $n^* = 30\text{-}60$ (**Table 4**) should be involved in the computation of GC,

457   in order to maximize the efficiency of its estimation using ALS. Since the value of $|r|$

458   involves both the field GC and the ALS metrics, its changes are determined by both N and d

459   (**eqs. 4-5**), and both may cause a change in the correlation between the two variables.

460   *3.4 Effect of Point Density on the Relationship of GC*

461   According to the previous results, we set the optimal plot size to $s^* = 9$ m in order to further

462   analyse the possible effects due to varying scan density. Among all the ALS metrics (**Table**

463   **2**), we selected those same ones employed previously – *P25, P50, P99, Skew, StdDev, Cover*

464   and *CRR* – to allow direct comparison. **Fig. 6** shows the evolution in $|r|$ for increasing ALS

465   point density d. No considerable changes were observed in the correlation between the field

466   GC and the ALS metrics, which suggests that d has no major effects on their relationship.

467   However, a decreasing trend in $|r|$ could generally be observed when point densities

468   decreased below $d < 3$ points·m^{-2} (**Fig. 6**). Overall, these results therefore suggest that the

469   relationship between GC and ALS metrics is mainly dependent on the plot size employed,

470   and rather independent of stand density and ALS scan density

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

20

472   **4. Discussion**

473   In this study we evaluated the effects of plot size and sample size on the GC of tree size

474   inequality, and on its practical estimation using remote sensing methods based on ALS.

475   Sample size refers to the number of individual elements (trees or ALS returns) included

476   within a given sample area, which is therefore determined by the spatial resolution employed

477   for evaluating a given forest attribute. We also analysed the effects of ALS scan density and,

478   overall, we observed that plot size had greater effects on the relationship between GC and

479   ALS metrics than either of the other two criteria considered. The motivation for studying

480   these effects is grounded on the fact that inappropriate plot sizes may provide unreliable

481   estimates and lead to sub-optimal forest management decisions (Eid, 2000; Mauro et al.,

482   2010). Valbuena et al. (2013a) pointed out that the estimation of GC is affected by the area at

483   which it is evaluated. Results in **Fig. 3** illustrate how the $\overline{GC}_{diff}$ decreases when increasing

484   the size of circular plots and, and hence their corresponding sample size. $\overline{GC}_{diff}$ values

485   markedly dropped for smaller plot radii and sample sizes. This decrease smooths from bigger

486   sizes, which indicates stabilization of the estimation (*Criterion I*). **Fig. 2a** also shows an

487   example of this tendency to asymptotically approach the population value, which was also

488   observed by George (2003), Barbeito et al. (2009), or Matos (2014). Based on *Criterion I*

489   ($\overline{GC}_{diff} < 0.05$), the circular plot should be large enough ($s \geq 6$ m) to have minimum

490   sample size of $n \geq 15$ trees (**Fig. 3**). Although the minimum plot size also depends on the

491   stand density of an area, **eq. (4)** can be used to adjust the method to any forest areas, whether

492   sparsely or densely forested. This conclusion may therefore be partly extended to other forest

493   types, as it can be for example deduced (via **eq. 4**) that minimum radius of $s \geq 12$ m would

494   be needed in sparsely forested area of only 300 stems·ha[-1] (Lombardi et al., 2015). **Eq. (4)**

495   therefore brings generality to the method, since plot sizes may hence be tailored to forest

496   areas of differing stand densities.

497   In this article we also postulated that maximizing the explained variability between the GC

498   estimated from the field and ALS metrics could be a valid criterion to optimize the reliability

499   of ALS-assisted estimations of GC (*Criterion II*). Results in **Fig. 4a** showed that our

500   presumption was correct, since the $|r|$ values between GC and most ALS metrics, especially

501   the most correlated ones, followed a convex curve with a maximum that could be searched to

502   reach an optimal plot size / spatial resolution for the estimation. On the other hand, once the

503   GC reached some stabilization, the correlation between them remains largely unchanged.

504   Therefore, a lower plot size limit is to be imposed to avoid local minima that could appear as

505   an artefact of the unstable estimation of GC at low sample sizes. We shaded this area in grey

506   colour in **Fig. 4 (a, b)**, denoting the area that was already dismissed as a result of *Criterion I*

507   (**Fig. 3**; George, 2003). In larger plots the sample size was more representative of the total

508   population. Combining both criteria, we found in our study area that an optimal circular plot

509   radius of $s^* = 9\text{-}12$ m, which corresponds to a spatial resolution of sampling units sizing

510   250-450 m^2 (**Fig. 4a**), would be suitable for ALS-assisted GC estimation. Since plot size and

511   sample size are interdependent (**eq. 4**), this result may be suitable for any area with a similar

512   average number of trees ($N \cong 1300$ stems·ha^{-1}; **Table 2**). According to these results,

513   therefore, most forest datasets commonly acquired in operational inventories would be

514   acceptable for an ALS-assisted estimation of the GC of tree sizes. Lombardi et al. (2015)

515   deduced a larger optimal plot radius $s^* = 13\text{-}15$ m for other forest attributes, most likely due

516   to lower N in the forest areas considered. For studies dealing with differing plot sizes, one

517   possibility could be to upscale GC to a common plot size (Kent and Coker, 1992; Magnussen

518   et al., 2016).

519   Some of the reflexions raised in this article affect all other types of forest attributes and

520   remotely sensed auxiliary variables that may be used in forest estimations (Jelinski and Wu,

521   1996). However, different forest attributes are differently affected by varying plot sizes

22

522   (Chytrý and Otýpková, 2003). Some forest variables such as stand density or biomass would

523   show an averaging effect as plot size increases (Jelinski and Wu, 1996; Gotway and Young,

524   2002; Ruiz et al., 2014), which in turn derives in improved model efficiency when using

525   larger scales in remote sensing estimations (Næsset et al., 2015; Mauro et al., 2016). But

526   there is a trade-off between model accuracy and spatial resolution, and root mean squared

527   errors increase from10-15% for 1000-4000 m^2 to 20-25% for 200-250 m^2 (Næsset, 2002,

528   2004, 2007). However, this averaging effect is not applicable to forest attributes describing

529   structural diversity and heterogeneity (Coomes and Allen, 2007). In fact, many variables

530   necessarily augment when the plot size increases, for instance species richness and diversity

531   (e.g., Humphrey et al. 2000; Otypková and Chytry, 2006; Kallimanis et al., 2008; Fibich et

532   al., 2016) as traditionally assessed through rarefaction (Kent and Coker, 1992). A similar

533   effect can be observed in other measures of forest heterogeneity (Barbeito et al., 2009; Motz

534   et al., 2010; McRoberts et al., 2012), and thus in the GC (Valbuena et al., 2013a, Matos,

535   2014), since increasing the size of a plot increases the probability of finding an additional

536   differently-sized tree (Chen and Crawford, 2012; Valbuena et al., 2012). This is why

537   estimated GC values in **Fig. 3** asymptotically approach the value of the larger original field

538   plot (George, 2003; Matos, 2014), which is never exceeded. Imposing a criterion defining

539   which of the plausible plot sizes should be used is therefore not a trivial question to tackle.

540   Matos (2014) employed a number of different criteria based on the field information only –

541   stabilization of the estimate, stabilization of certainty of the estimate and convergence with

542   GC_{ref} –, none of them resulting fully satisfactory and definitive as they all ultimately rely on

543   a subjective assumption (Cressie, 1993). For this reason, in this article we approached the

544   question of plot size from the viewpoint of its practical estimation using ALS remote sensing.

545   The convex curves obtained in **Fig. 4a** proved this approach to be highly beneficial, since

546   maximization of correlation $|r|$ between GC values and selected ALS predictors provides

23

547   with a more objective method for determining the optimal plot size for the assessment of GC

548   of tree size inequality. Still, due to the very high uncertainty observed in the estimation of GC

549   when using very small plot sizes (**Fig. 3b**; Smith, 1938; Lombardi et al., 2015), we deducted

550   that a criterion avoiding great divergence with GC_{ref} may be imposed as a prior step to

551   maximization (Motz et al. (2010) referred to it as minimum grid spacing). Further research

552   could focus on modelling GC from ALS metrics and investigate how the interaction among

553   many ALS metrics in a same model may play a relevant role in the optimization of plot size

554   and spatial resolution.

555   The analyses carried out with reduced point densities revealed that lowering point density

556   barely affects the correlation between GC and ALS metrics, unless using a very sparse scan

557   density $d < 3$ points·m^{-2}. Previous studies such as Maltamo et al. (2006), Ruiz et al. (2014)

558   or Singh et al. (2015) also indicate that reducing the point density is not affecting the

559   accuracy of volume prediction and demonstrate that the effects of varying scan densities can

560   be eluded in practical applications. It must be taken into account, however, that the DTM

561   used in this study was based on original point density, and the errors in DTM determination

562   at sparser densities (Liu et al., 2007; Ruiz et al., 2014) may induce to further uncertainty,

563   although this presumably has a lesser effect on those metrics most related to GC.

564   Furthermore, since ALS datasets from national programmes are currently surveying entire

565   countries at densities typically between 0.5-1 points·m^{-2} (Artuso et al., 2003), it must also be

566   pointed out the relevance of results in **Fig. 6** which render most of these nation-wide ALS

567   datasets unsuitable for reliably estimating GC (Kandare et al. 2016). In line with results in

568   Valbuena et al. (2017), who postulated that the low densities incur in critical omission of

569   understorey development, our results demonstrate that indeed there is a need for increasing

570   point densities up to $d = 3$ points·m^{-2}. This result is very concurrent with those obtained by

571   Ruiz et al. (2014) and Watt et al. (2014) for different forest attributes in different stand types,

24

572   and therefore the case seems clear that ALS datasets obtained for forest applications should

573   reach this minimum density requirement.

574   **5. Conclusion**

575   In this study we studied how changing spatial resolution can affect the relationship between

576   GC and ALS metrics. We used three criteria for optimization: plot size, stand density and

577   ALS scan density. The effects of stand and scan densities are intimately interrelated to plot

578   size, since they together determine the sample size employed in calculations. Amongst those

579   three criteria, we found plot size to predominantly affect the relationship between GC and

580   ALS metrics.

581   We observed that the estimation of GC is strongly affected by the size of the forest plot

582   surveyed. Very small sample size and plot radii are more sensitive to GC variations,

583   unrepresentative of the total population, producing unstable and unreliable GC estimations.

584   The GC estimation stabilizes as the size of plots and samples increases, as larger plots contain

585   a more appropriate number of observations (sample size) representing the population. We

586   determined that, in a boreal managed forest, a minimum number of 15 trees ought to be

587   measured for a reliable GC estimation, regardless of the stand density present at each forest

588   stand.

589   We developed a method for plot size optimization based on a combination of two criteria: (1)

590   imposing a minimum of number of 15 trees measured, and (2) maximizing the absolute

591   correlation between field GC and ALS metrics. The plot level correlation between ALS

592   metrics and field GC showed a convex tendency for increasing plot sizes. Our results showed

593   that 9-12 m-radius plots produced the maximum correlation thus they are suitable for ALS-

594   assisted GC estimation. Basic relationships between plot size and sample size may be used to

595   accommodate the method to forested environments of varying stand densities.

596   With regards to the effects of ALS scan density, we observed that it can barely have any

597   effects unless lowered under 3 points m^{-2}. This however may be relevant for the practical

598   application of low-density national datasets, and therefore we would recommend increasing

599   their scan densities with the intention to render nation-wide datasets useful for studying forest

600   heterogeneity.

601   **Acknowledgements**

607   **References**

608   Asada, Y., 2005. Assessment of the health of Americans: the average health-related quality of
609   life and its inequality across individuals and groups. Population Health Metrics 3, 7-7.

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

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

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

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

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

623 Bollandsås, O.M., Næsset, E., 2007. Estimating percentile-based diameter distributions in
624 uneven-sized Norway spruce stands using airborne laser scanner data. Scand. J. For. Res. 22,
625 33-47.

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

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

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

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

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

637 Cordonnier, T., and Kunstler, G. 2015. The Gini index brings asymmetric competition to
638 light. Perspectives in Plant Ecology, Evolution and Systematics, 17 (2), 107-115.

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

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

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

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

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

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

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

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

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

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

660   Gobakken, T., Næsset, E., 2008. Assessing effects of laser point density, ground sampling
661   intensity, and field sample plot size on biophysical stand properties derived from airborne
662   laser scanner data. Canadian Journal of Forest Research 38, 1095-1109.

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

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

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

668   Hvistendahl, M., 2014. While emerging economies boom, equality goes bust. Science, 344,
669   832-835

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

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

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

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

679   Kandare, K., Ørka, H.O., Chan, J.C.-W., Dalponte, M., 2016. Effects of forest structure and
680   airborne laser scanning point cloud density on 3D delineation of individual tree crowns.
681   European Journal of Remote Sensing, 49, pp. 337-359.

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

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

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

28

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

690   Lei X., Wang W., Peng, C., 2009. Relationships between stand growth and structural
691   diversity in spruce-dominated forests in New Brunswick, Canada. Canadian Journal of Forest
692   Research 39, 1835–1847.

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

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

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

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

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

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

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

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

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

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

721   Mauro F., Molina I., García-Abril A., Valbuena R. & Ayuga-Téllez E. (2016) Remote
722   Sensing Estimates and Measures of Uncertainty for Forest Variables at Different Aggregation
723   Levels. Environmetrics 27(4): 225-238

29

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

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

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

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

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

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

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

739   Næsset, E., Bollandsås, O.M., Gobakken, T., Solberg, S., McRoberts, R.E., 2015. The effects
740   of field plot size on model-assisted estimation of aboveground biomass change using
741   multitemporal interferometric SAR and airborne laser scanning data. Remote Sens. Environ.
742   168, 252-264.

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

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

747   O'Hara, K.L., Hasenauer, H., Kindermann, G., 2007. Sustainability in multi-aged stands: an
748   analysis of long-term plenter systems. Forestry 80, 163-181.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

781   Valbuena, R., Packalén, P., Martín-Fernández, S., Maltamo, M., 2012. Diversity and
782   equitability ordering profiles applied to study forest structure. For. Ecol. Manage. 276, 185-
783   195.

784   Valbuena, R., Packalen, P., Mehtätalo, L., García-Abril, A., Maltamo, M., 2013a.
785   Characterizing forest structural types and shelterwood dynamics from Lorenz-based
786   indicators predicted by airborne laser scanning. Canadian Journal of Forest Research 43,
787   1063-1074.

788   Valbuena, R., Maltamo, M., Martín-Fernández, S., Packalen, P., pascual, C. and Nabuurs, G.,
789   2013b. Patterns of covariance between airborne laser scanning metrics and Lorenz curve
790   descriptors of tree size inequality. Canadian Journal of Remote Sensing, 39, pp. S18-S31.

791   Valbuena, R., Vauhkonen, J., Packalen, P., Pitkänen, J., Maltamo, M., 2014. Comparison of
792   airborne laser scanning methods for estimating forest structure indicators based on Lorenz
793   curves. ISPRS Journal of Photogrammetry and Remote Sensing 95, 23-33.

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

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

800 Vihervaara P., Mononen L., Auvinen A.P., Virkkala R., Lü Y., Pippuri I., Packalen P.,
801 Valbuena R., Valkama J., 2015. How to Integrate Remotely Sensed Data and Biodiversity for
802 Ecosystem Assessments at Landscape Scale. Landscape Ecology 30 (3): 501-516.

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

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

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

810 Weiner J., Thomas S., 1986. Size variability and competition in plant monocultures. Oikos
811 47: 211–222.

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

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

815

816 **Table Titles**

817 **Table 1**. Properties of the study area.

818

819 **Table 2**. Summary of ALS metrics computed with FUSION and used in this research

820 (McGaughey, 2015).

821

822 **Table 3**. For each radii, proportion of the total number of original field plots within the

823 $\overline{GC}_{diff} < 0.05$ limit (*Criterion I*), and average number of trees contained within those plots.

32

824

825   **Table 4**. Maximum absolute correlation between field GC and ALS predictors (*Criterion II*).

826   See **Table 2** for description of ALS metrics.

827

828

829   **Figure Captions**

830

831   **Figure 1**. Reproduction of tree positions (dots) within an original field plot (red rectangle)

832   surrounded by edge correction i.e. translation method (i.e. periodic boundary), and a sample

833   of 10 random realizations of simulated concentric circular plots with radii sizing 1-3 m (for

834   simplicity). Axes show both absolute (above) and relative (below) coordinates (respectively

835   X_{abs}, Y_{abs} and X_{rel}, Y_{rel} in **Eq. 2**).

836

837   **Figure 2**. Results of sensitivity analysis to select minimum numbers of simulations.

838   Evolution for increasing radii of (a) mean $\widehat{GC}$ values and (b) their standard errors for $k =$

839   10-2000 simulations.

840

841   **Figure 3**. *Criterion I.* Asymptotic representation showing the evolution of $\overline{GC}_{diff}$ (at each of

842   the 79 original field plots) for increasing (a) plot sizes $s = 1$-15 m radius (corresponding area

843   also shown in upper axis) (b) and sample size $n = 1$-50 number of trees (shortened to

844   enhance visualization).

845

33

846   **Figure 4.** *Criterion II.* Absolute of correlation $|r|$ between GC values and selected ALS

847   predictors (see legend, and explanations of ALS metrics in **Table 1** and section 2.4) for

848   increasing (a) plot size $s = 1\text{-}15$ m radius (corresponding area also shown in upper axis) (b)

849   and sample size $n = 1\text{-}90$ number of trees.

850

851   **Figure 5.** Minimum number of trees (sample size) to reach GC stabilisation in relation to the

852   reference GC value obtained from the original field plot (GC_{ref}).

853

854   **Figure 6.** Changes due to varying ALS scan densities in the absolute of correlation $|r|$

855   between GC values and ALS predictors. See explanations of ALS metrics in **Table 1** (section

856   2.4).

857

858   **Supplementary Materials**

859   **Supplementary Figure 1.** Histograms showing the distribution of the response variable –

860    $\overline{GC}$ (vertical bars) – and the predictor variables – *Skewness, Cover, CRR, P99, StdDev, P50*

861   and *P25* (horizontal bars) –. The resulting scatterplots between each response-predictor pair

862   are also shown. For simplicity, only results for the optimal plot radius $s^* = 9$ m are shown.

863

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.

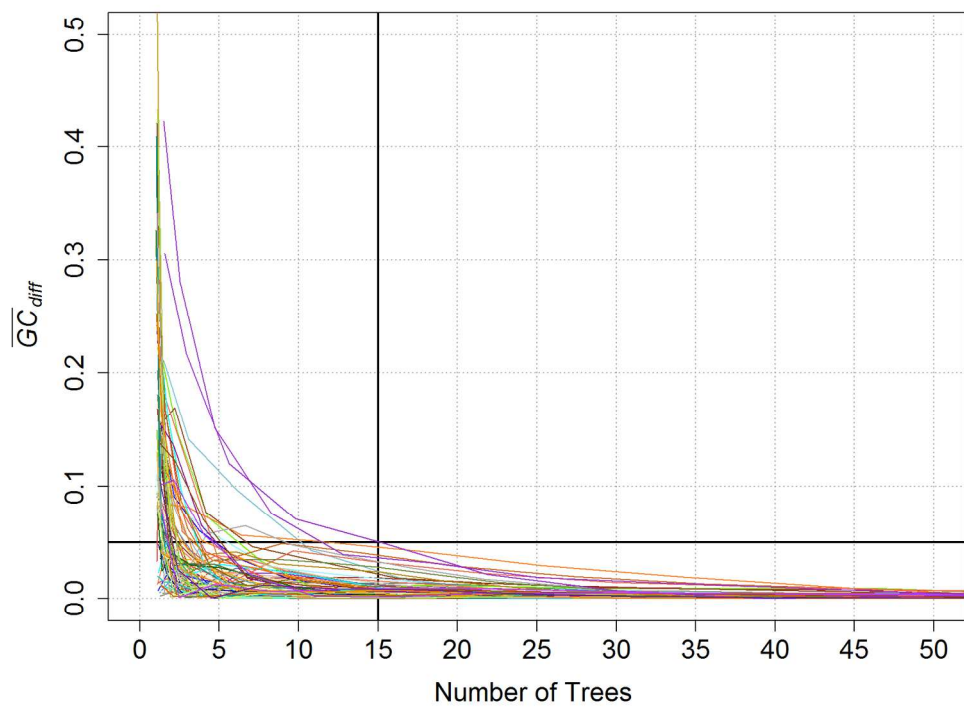| ALS metric	Maximum correlation max$\|r\|$	Optimal plot radius (s^*; m)	Optimal number of trees (n^*; m)
Skew	0.58	10	41
Cover	0.45	12	59
CRR	0.42	9	33
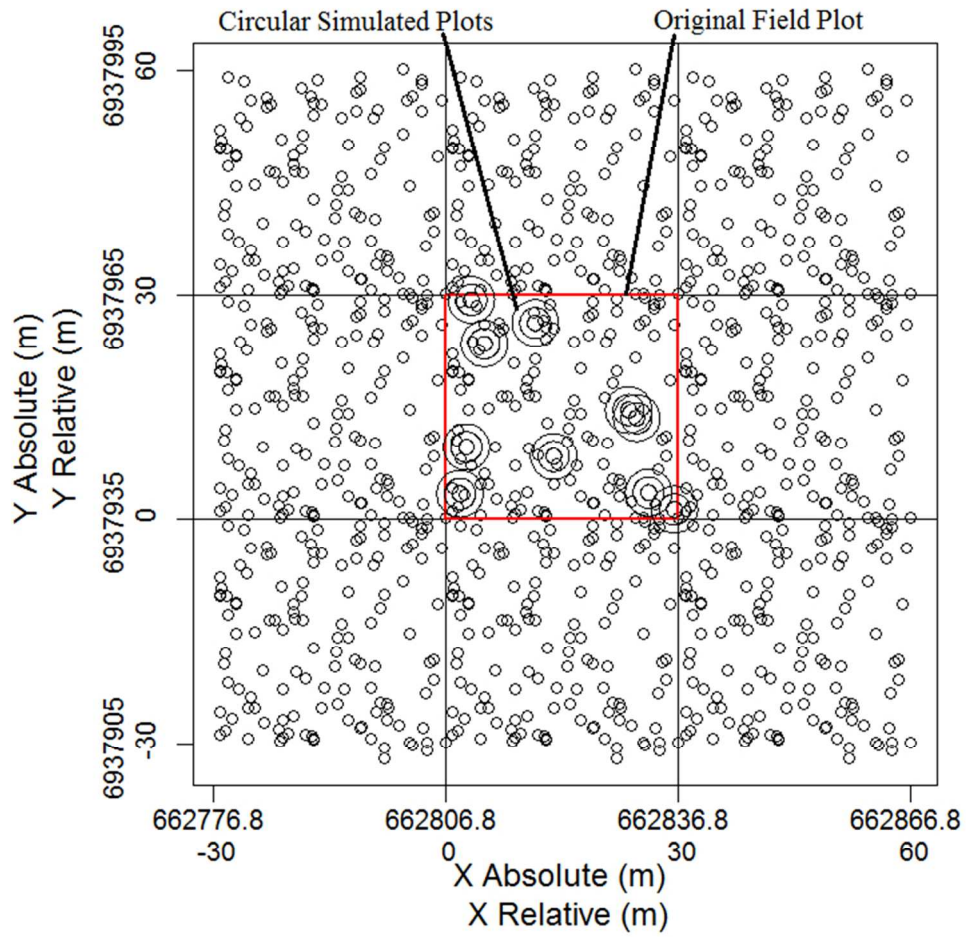
Figure 1

158x158mm (120 x 120 DPI)
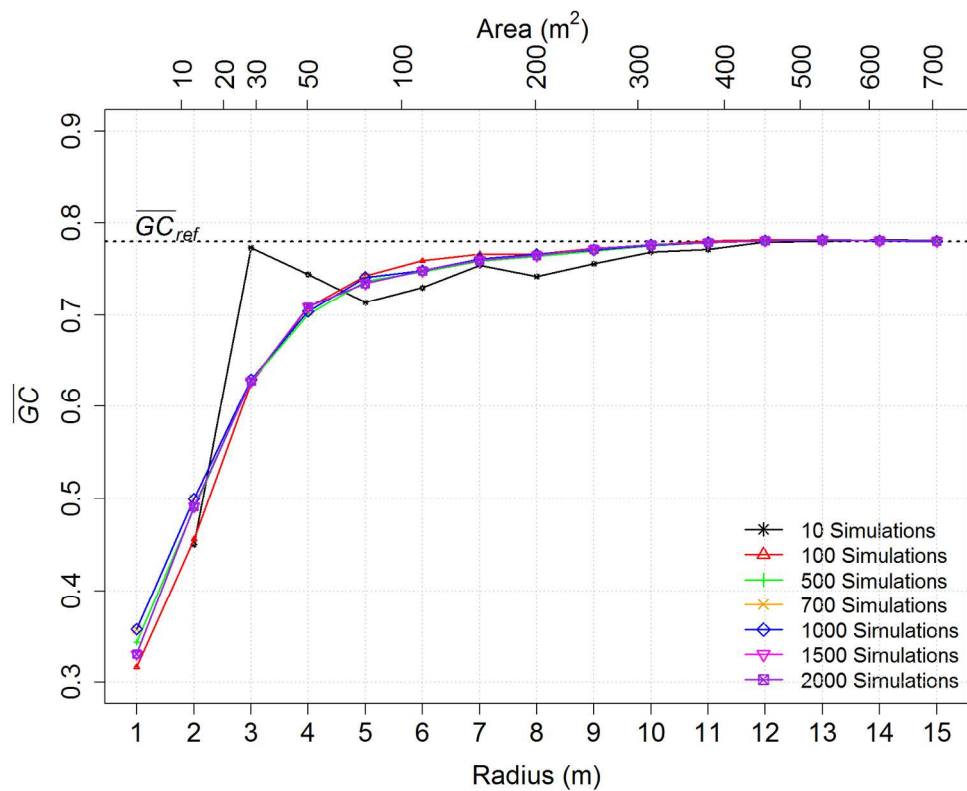
Figure 2(a)

241x190mm (200 x 200 DPI)
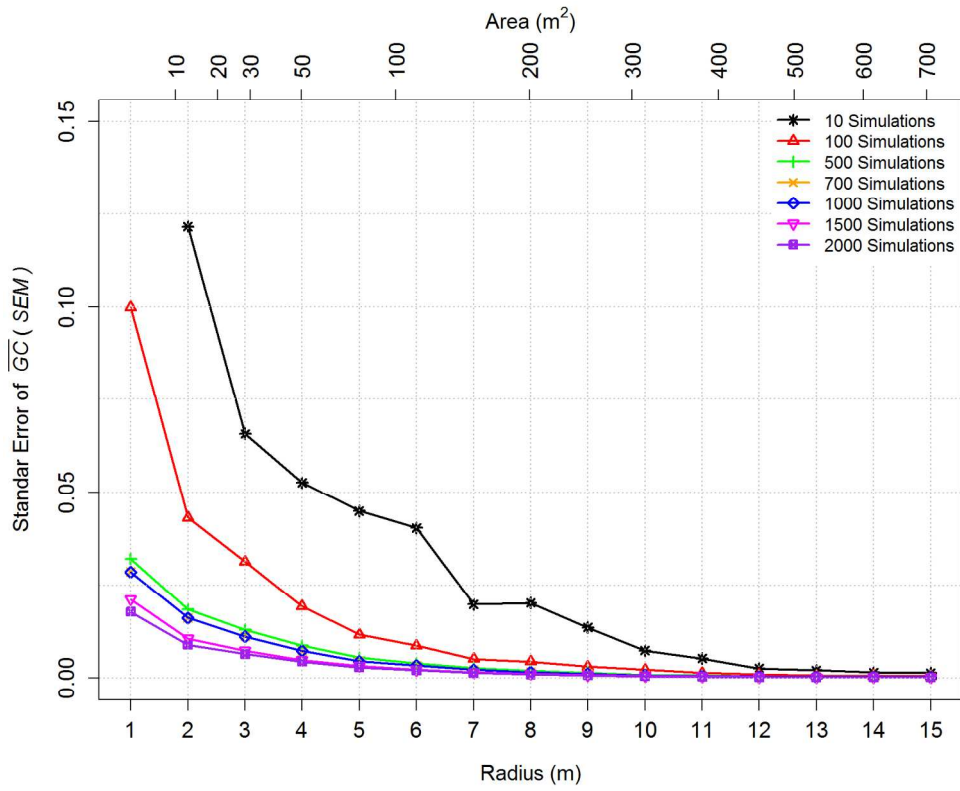
Figure 2(b)

241x190mm (200 x 200 DPI)
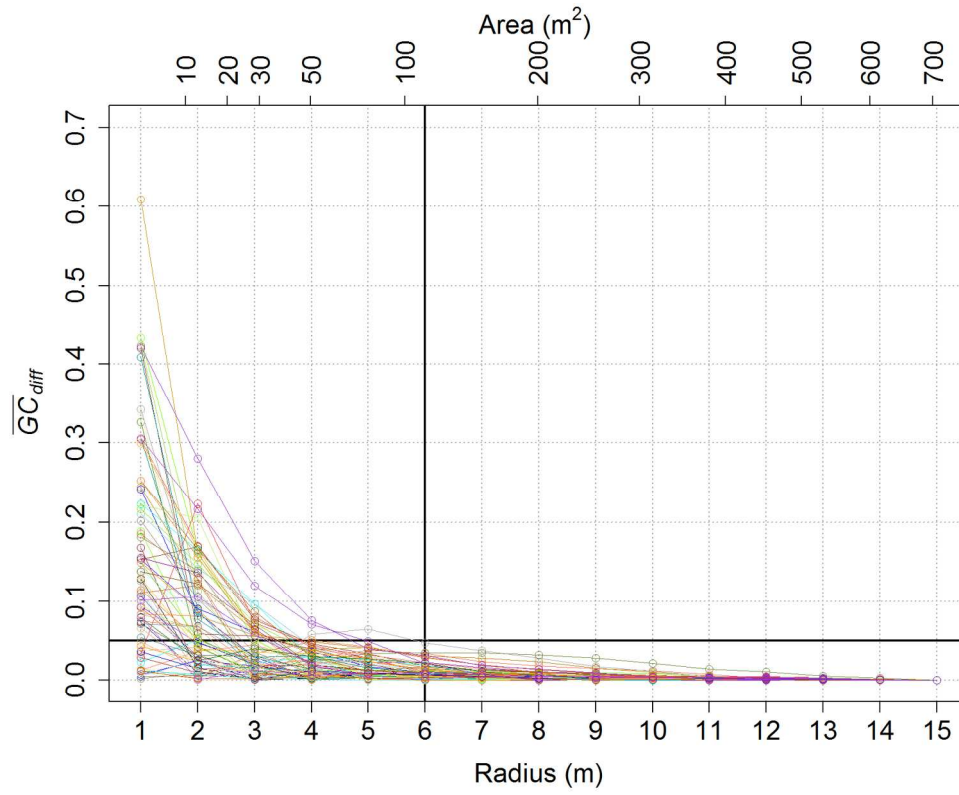
Figure 3(a)

241x190mm (200 x 200 DPI)

Figure 3(b)

241x190mm (200 x 200 DPI)
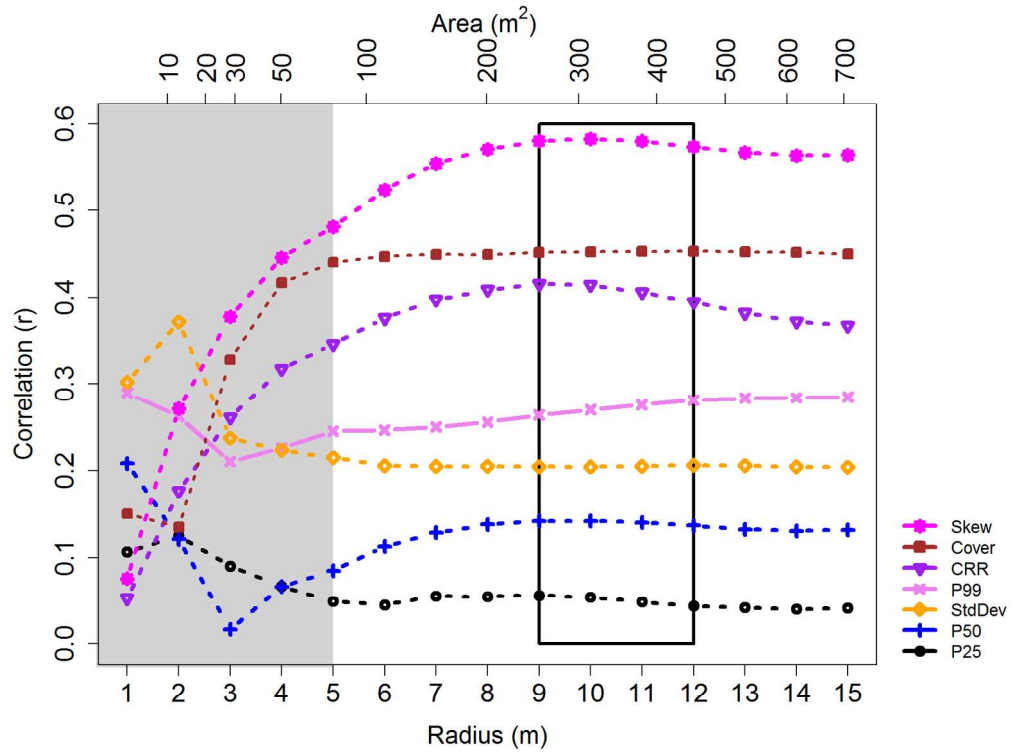
Figure 4(a)

241x190mm (200 x 200 DPI)
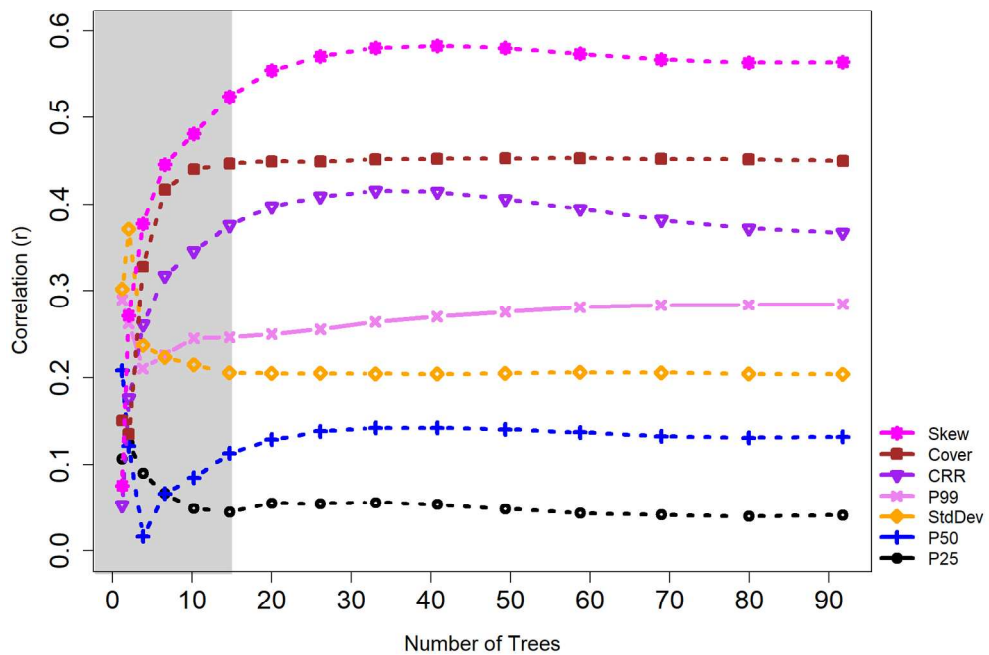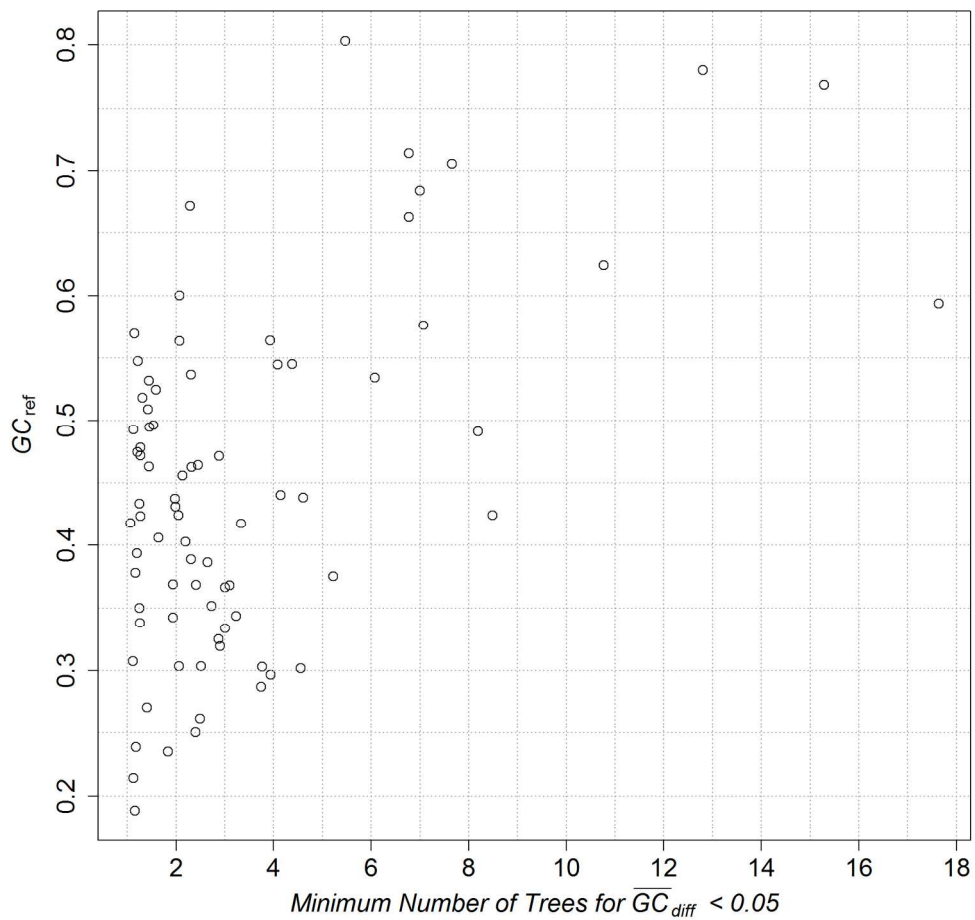
Figure 4(b)

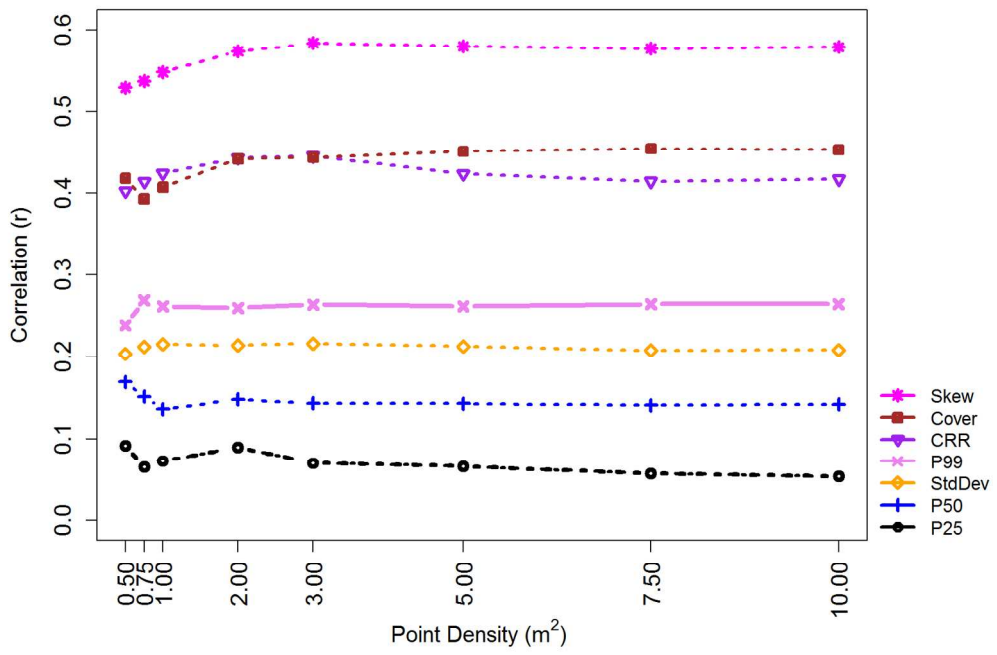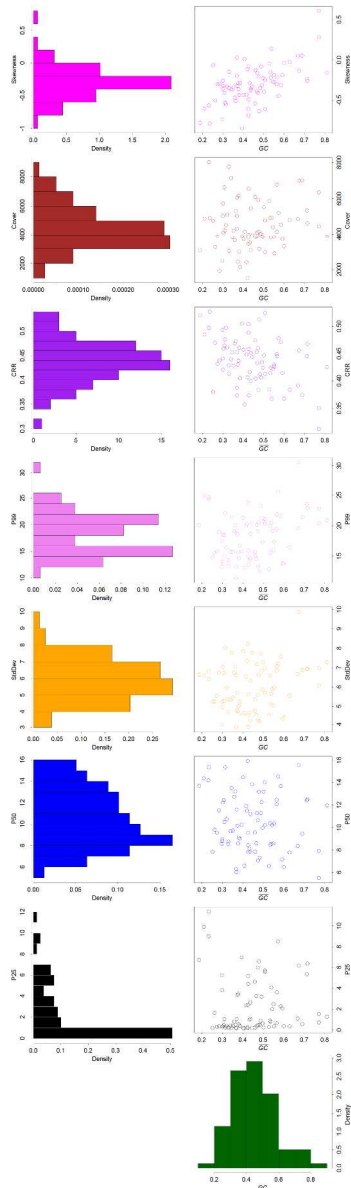241x190mm (200 x 200 DPI)

Figure 5

254x254mm (200 x 200 DPI)

Figure 6

228x165mm (200 x 200 DPI)

558x1625mm (100 x 100 DPI)