

Estimation of woodland canopy structure with terrestrial LiDAR: expanded methods

John L. Godlee

Contents

1	Introduction	2
2	Sampling	2
3	Field measurements	4
3.1	Trees	4
3.2	Grass biomass	5
3.3	Hemispherical photography	6
3.4	Stand structure	7
4	Terrestrial LiDAR	11
4.1	Registration	12
4.2	Voxelisation	12
4.3	Noise reduction	13
4.4	LiDAR analysis	14
4.4.1	Foliage density profiles	14
4.4.2	Canopy cover	17
4.4.3	Grass biomass estimation	19
4.4.4	Canopy rugosity	20
5	Statistical analysis	22
5.1	Foliage density profiles	22
5.2	Grass biomass	22

5.3 Canopy rugosity	23
-------------------------------	----

1 1 Introduction

2 This document provides detailed field and analytical methods for the study of tree canopy struc-
3 ture in southern African woodlands. The study aimed to understand the effects of tree species di-
4 versity and stand structure on tree canopy structure and grass biomass. Chapter XXX contains
5 the same methods in brief.

6 2 Sampling

7 Fieldwork was conducted at two sites, the first in Bicuar National Park, southwest Angola ($S15.1^\circ$,
8 $E14.8^\circ$), and the second in and around Mtarure Forest Reserve, southeast Tanzania ($S9.0^\circ$, $E39.0^\circ$).
9 Fieldwork was conducted during the peak growth period of each site, in order to capture the high-
10 est foliage volume in the canopy and grass volume in the understorey.

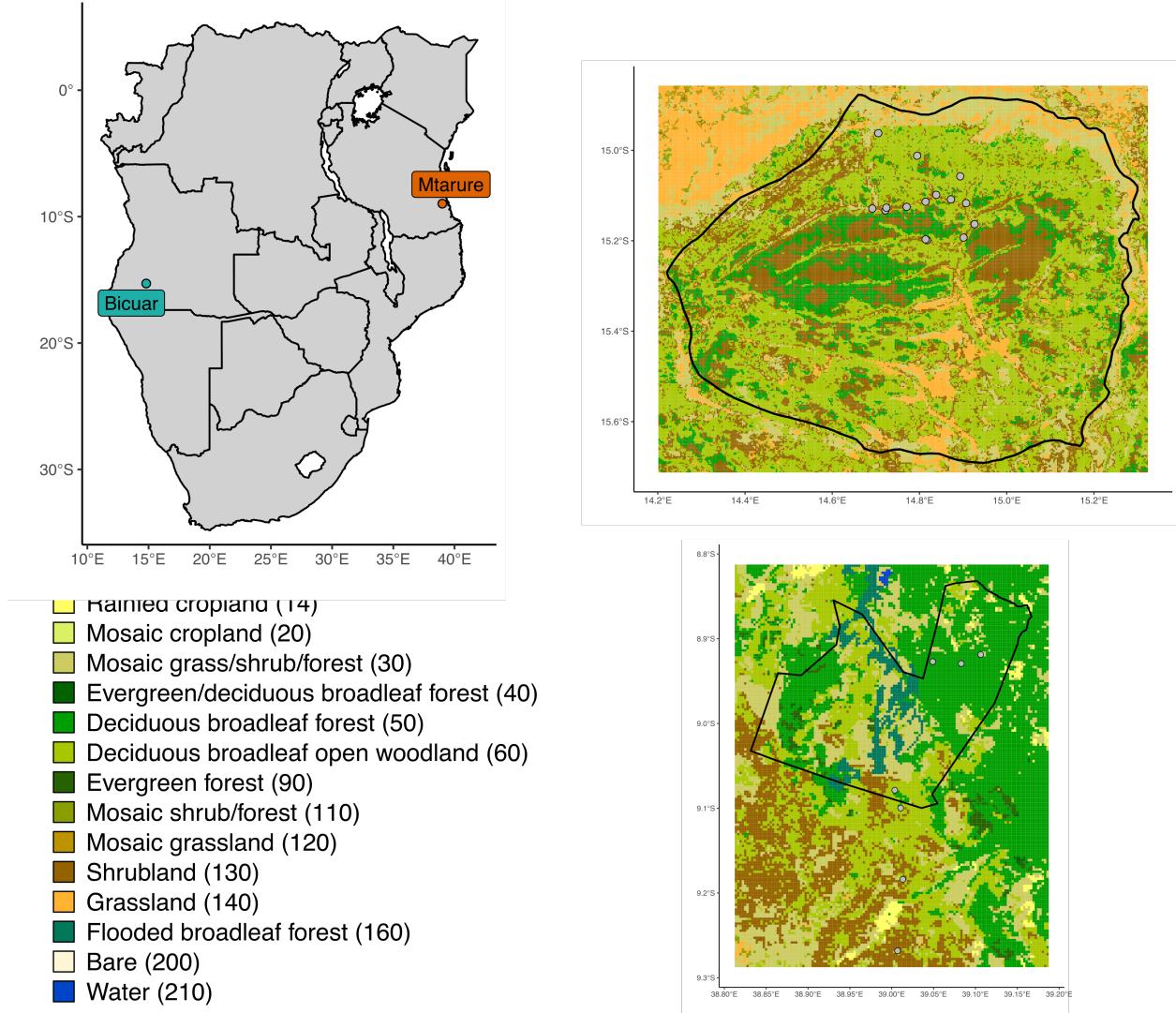


Figure 1: Location of study sites within southern Africa (a), and of 1 ha plots within each site. The black polygons denote the boundaries of protected areas which encompass the majority of study sites, Bicuar National Park in Angola (b), and Mtarure Forest Reserve in Tanzania (c). Each site map is coloured according to the GlobCover global land cover classification.

Site	MAT (°C)	MAP (mm y ⁻¹)	Temp. range (°C)	CWD
				(mm y ⁻¹)
Bicuar	20.8 (0.70)	825.9 (52.01)	24.5 (0.90)	-844.8 (44.29)
Mtarure	25.7 (0.24)	958.4 (25.19)	12.0 (0.33)	-739.6 (8.06)

Table 1: Climatic data for each site, extracted from WorldClim at 2.5 minute resolution. Values are the mean and standard deviation (in brackets) of all pixels intersecting each protected area.

11 At each site, a number of 1 ha permanent plots were sampled. In Angola, 15 plots were sampled,
12 while in Tanzania, only seven were sampled, following the curtailment of fieldwork due to COVID-

13 19 travel restrictions. Permanent plots were located in areas of homogeneous vegetation type,
 14 away from roads and undisturbed by humans. Plots were established following the SEOSAW pro-
 15 tocol (version 3.0, SEOSAW 2020). Plots were located quasi-randomly by first locating areas from
 16 satellite imagery expected to comprise savanna woodland vegetation. At each site, plots were de-
 17 liberately located along a gradient of stem density.
 18 Each permanent plot was further subdivided into nine 10 m diameter circular subplots arranged in
 19 a regular grid, with a buffer from the plot edge (Figure 2).

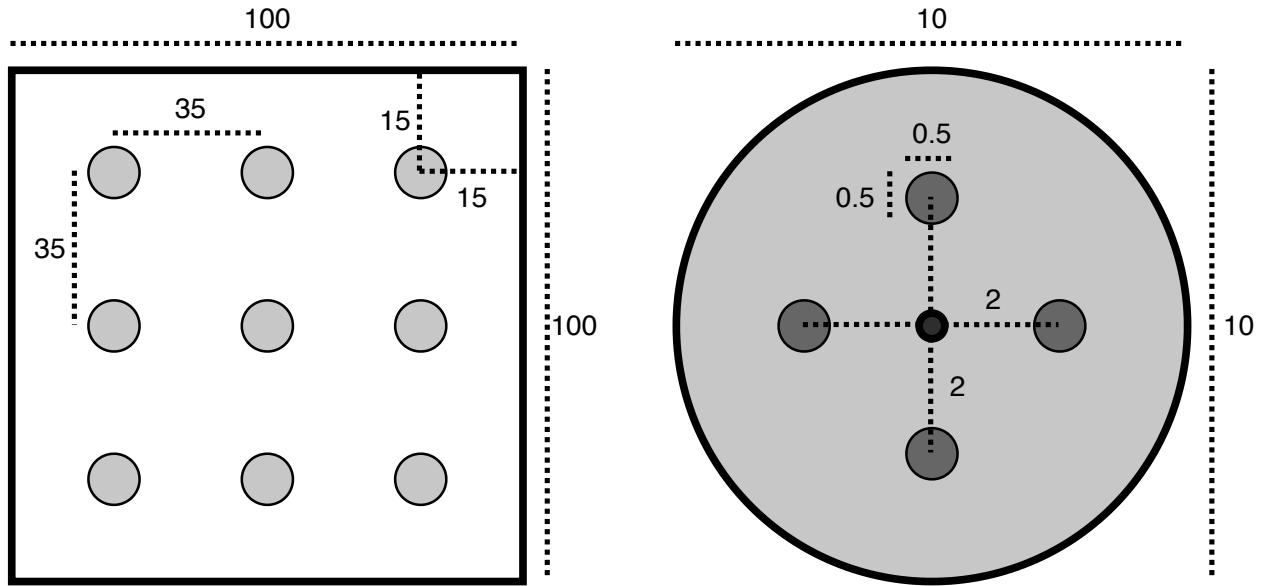


Figure 2: The layout of 10 m diameter subplots within each 1 ha square plot. Each subplot is situated inside a 15 m buffer from the plot edge, with 35 m between subplot centres. Subplots are arranged in a 3x3 grid. Disc-pasture measurements and biomass samples are located in cardinal directions 2 m from the centre of the subplot. All distances are in metres.

20 3 Field measurements

21 3.1 Trees

22 For each subplot, we measured all woody stems >5 cm stem diameter with canopy material inside
 23 the subplot. For each stem we recorded:

- 24 • Tree identity
- 25 • Stem diameter (diameter at breast height - 1.3 m)

26 For each tree, which may be composed of multiple stems joined at the base, we recorded:

- 27 • Species

- 28 • Height to top of canopy
- 29 • Canopy area, ellipse from two perpendicular measurements (Figure 3)
- 30 • Distance from subplot centre
- 31 • Compass direction from subplot centre

Plot edges

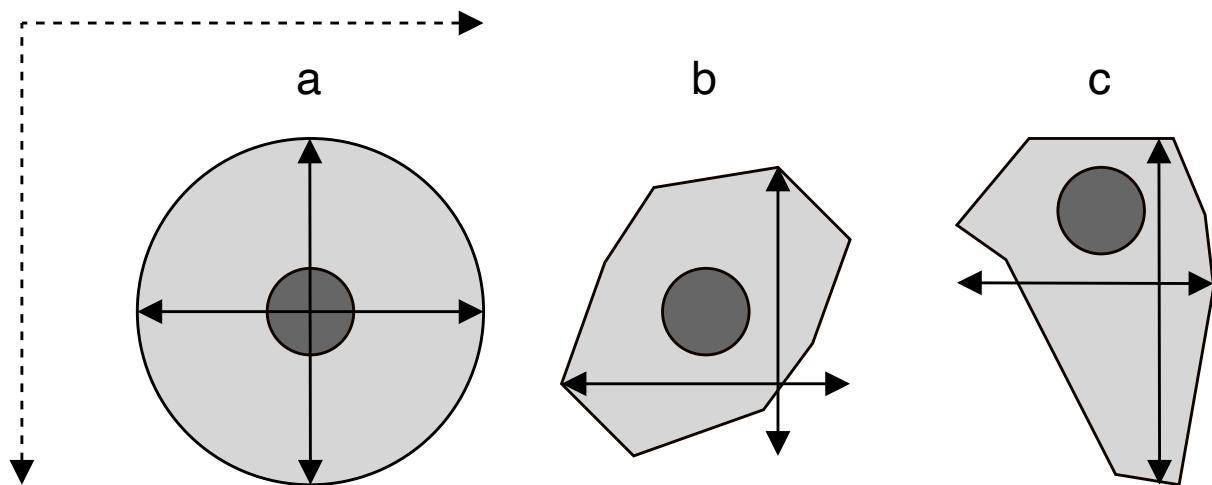


Figure 3: Examples of tree crowns as viewed from above to demonstrate how crown extent measurements are located. Darker grey circles show the main stem while pale grey polygons show the maximum extent of the crown. Extent measurements are taken parallel to the plot edges. a) shows a perfectly circular tree crown, b) and c) show irregular tree crowns, demonstrating that maximum crown extent in a given orientation be offset from the stem.

3.2 Grass biomass

32 Grass volume and biomass within each subplot was estimated from four sample points located 2
 33 m from the subplot centre in cardinal directions (Figure 2). At each point, a disc-pasture meter
 34 measurement was taken with a 45.8 cm radius disc weighing exactly 1.5 kg (Bransby and Tain-
 35 ton, 1977). Small woody stems were removed from disc-pasture sample points before the disc-
 36 pasture measurement was taken. The location of the sample point was moved if the designated
 37 point intersected with coarse woody debris, rocks, shrubs, or standing trees. Within each 1 ha
 38 plot, biomass harvesting was conducted at nine randomly allocated disc-pasture sample points.
 39 Tree leaf litter was removed from biomass samples. Biomass harvesting involved clipping all grass
 40 material within the 45.8 cm radius to ground level, taking care not to include roots. Grass sam-
 41 ples from Angola were dried until the mass remained constant (± 5 g) for >48 hours, then weighed
 42

43 to ascertain the grass biomass. Grass samples from Tanzania could not be processed due to cur-
44 tailment of fieldwork due to COVID-19 travel restrictions.

45 **3.3 Hemispherical photography**

46 At the centre of each subplot a single photograph was taken with a Nikon D750 full-frame DSLR
47 camera, with a circular fisheye lens. The lens had an equisolid (equal area) projection, which avoids
48 image distortion. The projection function is given by:

$$R = 2f \sin(\theta/2) \quad (1)$$

49 Where R is the radial position of a point on the image on the sensor, f is the focal length of the
50 lens, and θ is the angle in radians of the desired angular radius of the cropped image.

51 The photo was taken facing directly to zenith, with the top of the camera facing magnetic north,
52 at a height of 1.3 m or above understorey vegetation, whichever was higher. Table 2 shows de-
53 scribes the camera settings for each hemispherical photo.

Table 2: Description of camera settings used for each hemispherical photo. Note that the values of shutter speed and ISO are deliberately variable within sensible thresholds to adapt to light conditions.

Setting	Value
Camera model	Nikon D750
Lens model	Sigma 8 mm f/3.5 EX DG Circular Fisheye
Pixel pitch	5.95 µm
Sensor resolution	24.3 MP
Shutter speed	>1/60s
Aperture	5-7
ISO	100-200
Exposure compensation	-0.7 (Brusa and Bunker, 2014)
Focus	∞ (Hu and Zhu, 2009; Frazer et al., 2001)
Image size	Large Fine JPEG - circular image 4016x4016 px
Orientation	Landscape

54 Photos were captured under uniform light conditions as much as possible, either under overcast

55 skies or early in the day before direct sunlight could be seen on the photo.
 56 ImageJ (Fiji version 2.1.0/1.53c) was used to binarize hemispherical photos (), to separate plant
 57 material from sky. We first split each image into red, green and blue channels. We used the Huang
 58 algorithm to automatically threshold images, using the blue channel only, under the assump-
 59 tion that plant material reflects little blue light, while the sky reflects much more (). Images were
 60 saved as PNG at the original pixel resolution.

61 **3.4 Stand structure**

62 From the stem measurements we calculated a number of indices to characterise whole-plot and
 63 subplot stand structure.

64 We calculated the spatial mingling index (M) according to von Gadow and Hui (2002) at the plot
 65 level, with the adjustment for available species pool suggested by Hui, Zhao, et al. (2011). The
 66 spatial mingling index is a spatially explicit estimate of the degree to which species are spatially
 67 mixed within a plot:

$$M = \sum_{i=1}^N \left(\frac{S_i}{n_{\max}} \frac{1}{k} \sum_{j=1}^k v_j \right) \quad (2)$$

$$\text{with } v_j = \begin{cases} 0, & \text{neighbour } j \text{ same species as reference } i \\ 1, & \text{otherwise} \end{cases} \quad (3)$$

(4)

68 where k is the number of nearest neighbours considered for each reference tree, S_i is the number
 69 of species found among the k nearest neighbours of tree i , n_{\max} is the potential number of species
 70 in the neighbourhood, i.e. $k + 1$, and N is the total number of trees in the plot. In our case we
 71 used the conventional value of $k = 4$ (von Gadow and Hui, 2002; Hui and Albert, 2004; Hui, von
 72 Gadow, et al., 2007). The value of M_i increases with greater mixing of species, and all else being
 73 equal will increase with number of species within the plot (Figure 4).

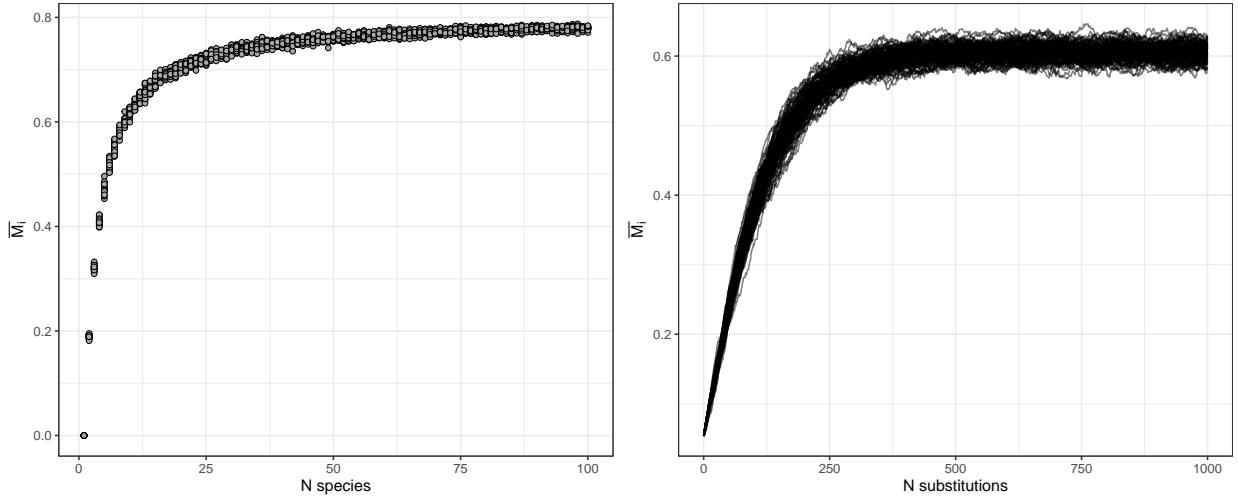


Figure 4: The behaviour of M_i with increasing number of species (left), and increasing spatial mixing of species. The left panel was generated by randomly assigning different numbers of species, in equal proportions, to an evenly spaced grid of individuals. 20 replicates were conducted for each number of species. The right panel was generated by randomly swapping pairs of individuals in a plot with 9 species arranged in mono-specific square blocks in an evenly spaced grid. Each line shows a single replicate, where individuals were swapped in an additive fashion, with 100 total.

⁷⁴ We also calculated the Winkelmass W according to von Gadow and Hui (2002) at the plot level.

⁷⁵ The Winkelmass estimates the degree of spatial uniformity in stem spatial distribution:

$$W = \sum_{i=1}^N \frac{1}{k} \sum_{j=1}^k v_j \quad (5)$$

$$\text{with } v_j = \begin{cases} 0, & \alpha_j \leq \alpha_0 \\ 1, & \text{otherwise} \end{cases} \quad (6)$$

$$(7)$$

⁷⁶ where α_j is the angle between consecutive neighbours and α_0 is the principal angle, where $\alpha_0 =$
⁷⁷ $360/k$. The value of the Winkelmass increases with increasing spatial clumping (decreasing spatial
⁷⁸ regularity) of individuals (Figure 6), and in a plot with random tree distribution will increase as
⁷⁹ more neighbours are considered (Figure 7).

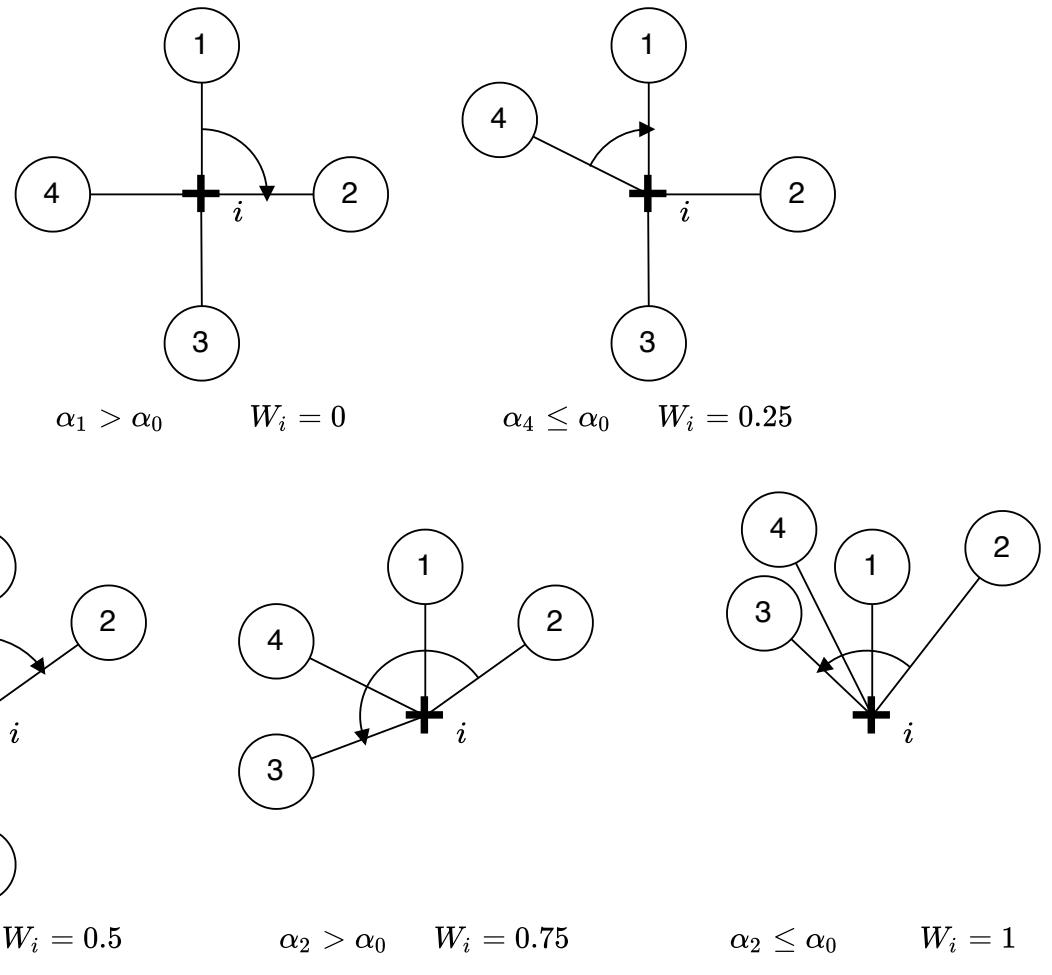


Figure 5: Possible values of W_i at a sample point i , denoted by a cross. Neighbours are represented as circles numbered sequentially from 1 to 4, where $k = 4$. The angles of arrows in each example are given below, along with the Winkelmass for that example.

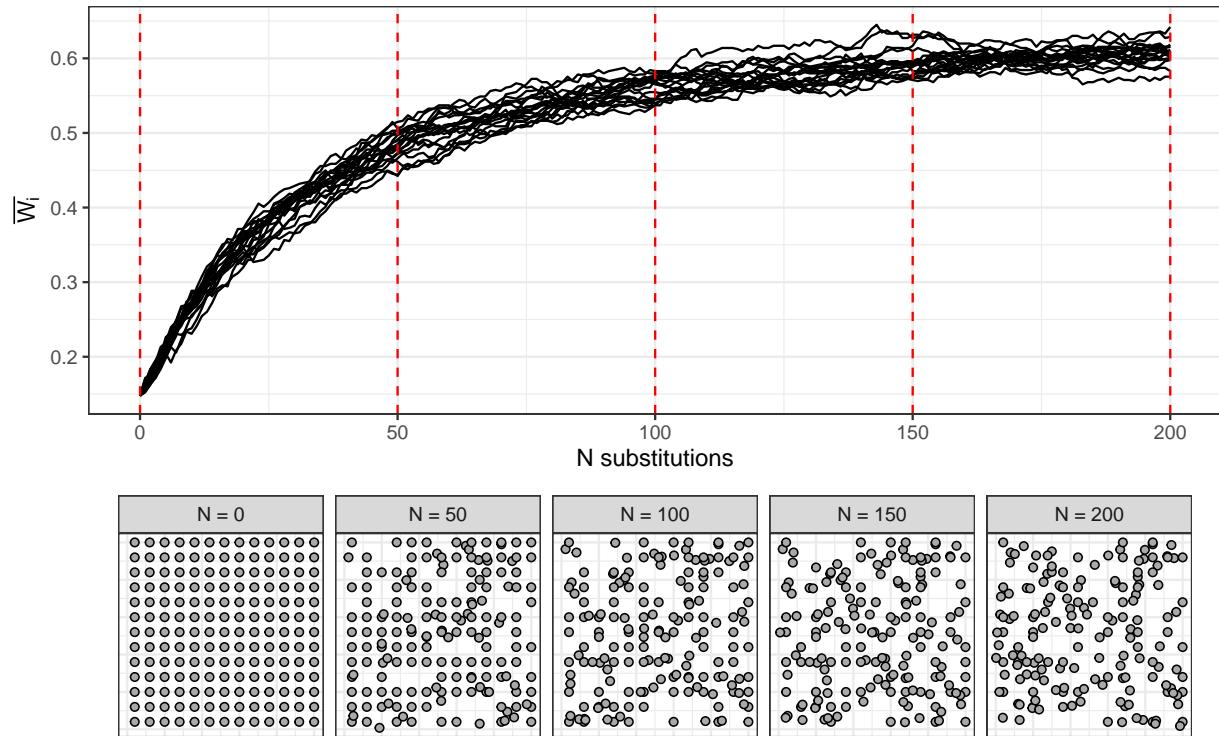


Figure 6: Variation in Winkelmass with increasing spatial irregularity of individuals. The top panel shows variation of Winkelmass in 20 plots as individuals are sequentially moved to a random location within the plot. Red dotted lines correspond to the panels below which show the spatial distribution of individuals after a given number of random individual movements.

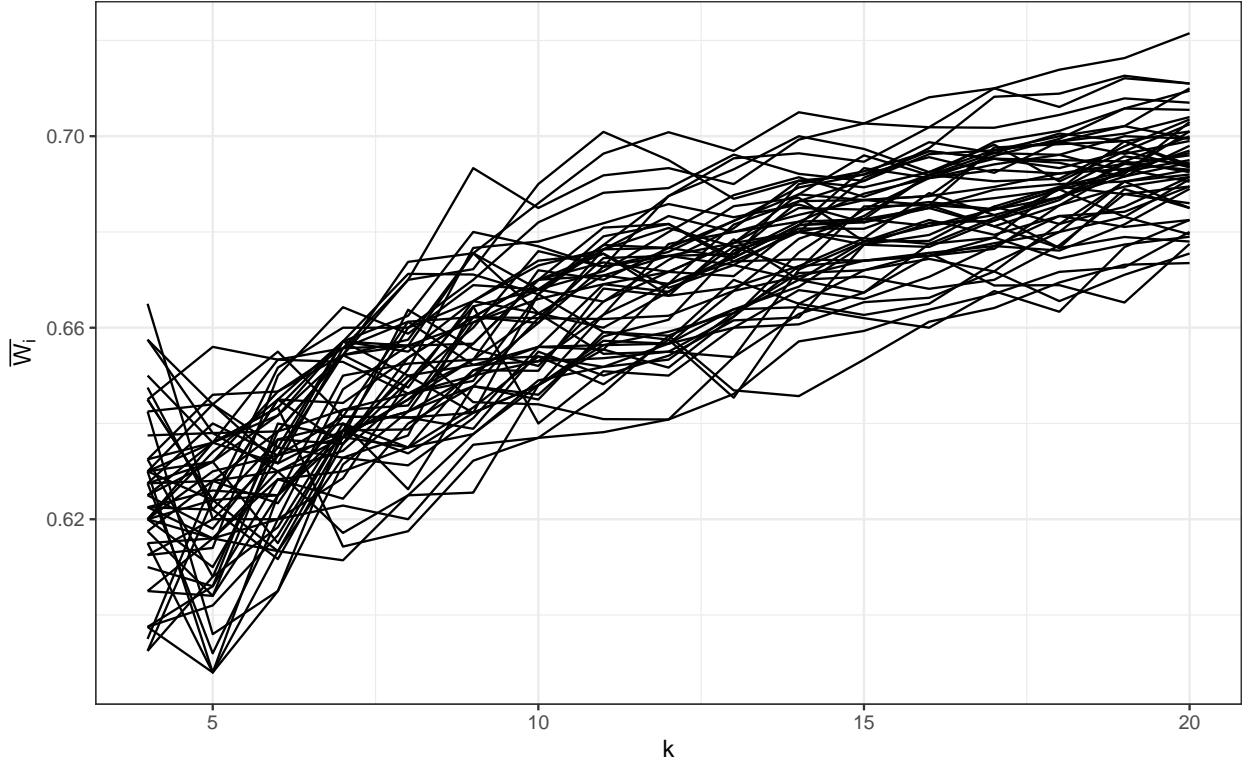


Figure 7: Variation in Winkelmass with increasing number of neighbours k considered in the calculation. 50 replicate plots were used, each with 100 individuals randomly distributed in space.

80 To estimate tree spatial structure in subplots we used an adapted version of the Iterative Hegyi
 81 index (H_i) (Hegyi, 1974). Our adapted formula allows the index to be based on a point rather
 82 than a focal tree, transforming it from a tree-centric competition index to a point-centric crowding
 83 index:

$$H_i = \log \sum_{j=1}^n \left(\frac{1}{L_{ij}} D_j \right) \quad (8)$$

84 where D_j is the stem diameter of neighbour tree j and L_j is the distance of the neighbour from
 85 the subplot centre.

86 4 Terrestrial LIDAR

87 Within each subplot, a variable number of scans were recorded using a Leica HDS6100 phase-
 88 shift terrestrial laser scanner (TLS). The number and position of scans within a subplot was de-
 89 termined by the arrangement and density of canopy material in the subplot. Scan positions were
 90 arranged to minimise shadows within the canopy, and to maximise canopy penetration. Number of
 91 scans per subplot ranged between one and five in both Angola and Tanzania (Table 3).

92 Five Leica 6" planar tilt and turn targets were used at each subplot to align scans. To allow reg-
 93 istration of scans among subplots, the location of each target was registered using a Leica VIVA
 94 GS10 GNSS unit, set up in post-processed kinematic (PPK) configuration with a base-station lo-
 95 cated ~100 m from the edge of each 1 ha plot. The location of each target was measured for at
 96 least 4 minutes. Further, we used the TrimbleRTX GNSS post-processing service to precisely lo-
 97 cate each target (Chen et al., 2011). When registering scans we discarded targets with location
 98 accuracy of >3 cm.

Table 3: Description of scan settings used for each scan.

Setting	Value
Scanner model	Leica HDS6100
Wavelength	650-690 nm
Spot size at exit	3 mm
Beam divergence	0.22 mrad
Range	79 m @90%; 50 m @18% albedo
Azimuth range	0-360°
Zenith range	0-155°
Increments	0.018°
Point spacing over 25 m	7.9 mm
Pixels per line	20000
Lines	10000
Compressed file size	~800 MB
Duration of scan	6 minutes 44 seconds

99 4.1 Registration

100 Scan registration for each subplot was conducted in Leica Cyclone (version 9.1). Targets from
 101 each scan were aligned using Cyclone's automatic target acquisition.
 102 After registration, scan scenes were exported from Cyclone as PTX files, one per subplot.

103 4.2 Voxellation

104 PTX files were converted to compressed LAZ files using PDAL (). The exact code used to ex-
 105 tract and apply the PTX rotation matrix to each point in the PTX file can be found IN THIS

106 APPENDIX HERE.

107 LAZ files were voxelised to different voxel sizes depending on the application of the data. For
108 grass biomass estimation, we used 2 cm^3 cubic voxels, while for subplot height profile estimation
109 we used 5 cm^3 voxels, and for whole plot canopy rugosity we used 10 cm^3 voxels. WHY THO

110 **4.3 Noise reduction**

111 Outlier detection and noise reduction was conducted in PDAL using the `filters.outlier` filter,
112 using the “statistical method” (sensu Rusu et al. 2008), with $k = 8$ (mean number of neighbours),
113 and $m = 1.96$ (standard deviation threshold, approximating a 95% confidence interval):

$$\bar{\mu} = \frac{1}{N} \sum_{i=1}^N \mu_i \quad (9)$$

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (\mu_i - \bar{\mu})^2} \quad (10)$$

$$t = \mu + m\sigma \quad (11)$$

$$\text{outlier}_i = \begin{cases} \text{true}, & \text{if } \mu_i \geq t \\ \text{false}, & \text{otherwise} \end{cases} \quad (12)$$

114 where N is the number of points in the scene, $\bar{\mu}$ is the mean distance to nearest neighbour points,
115 and σ is the standard deviation of these distances. t is the threshold distance used to define an
116 outlier.

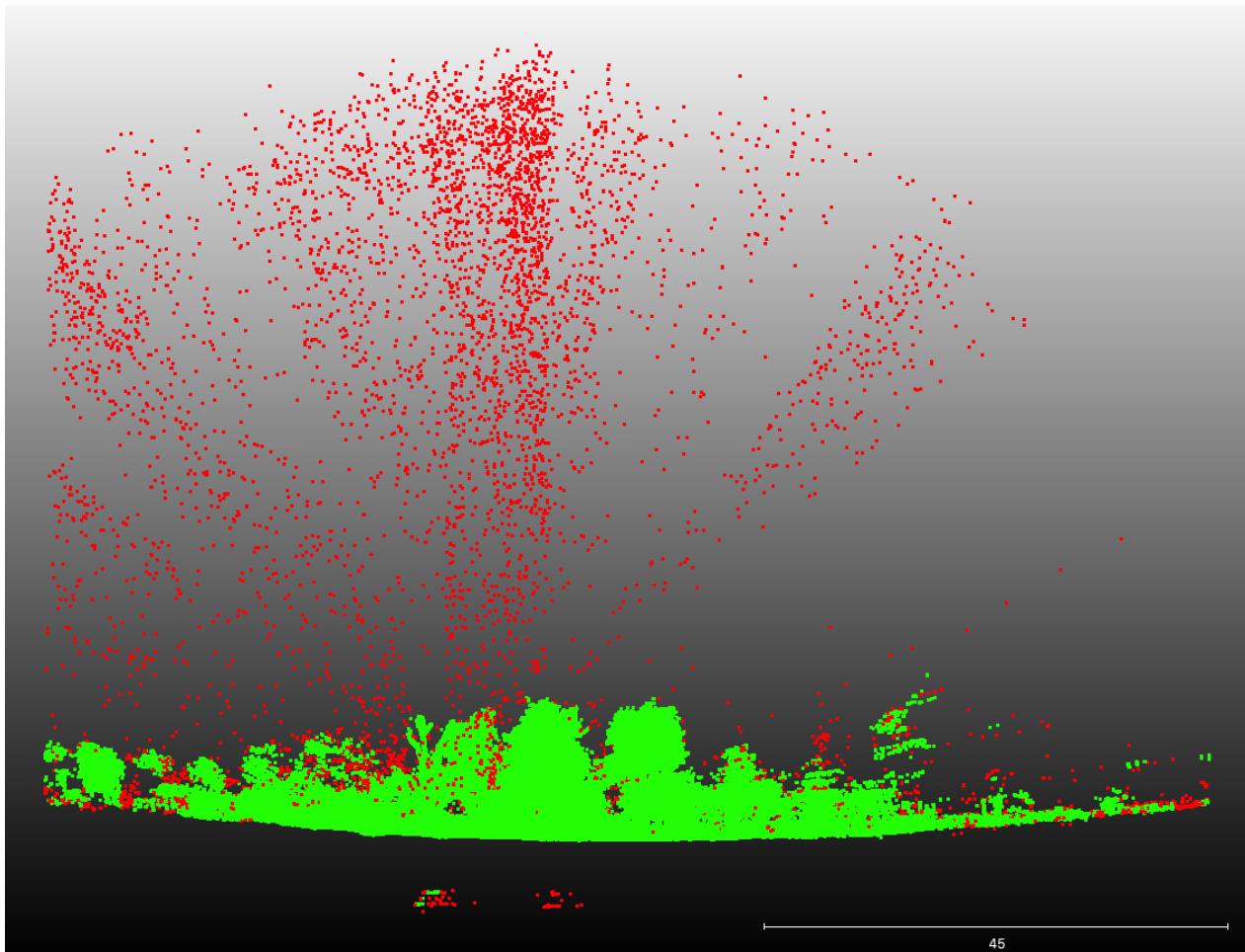


Figure 8: 2 m deep cross section of subplot showing the efficacy of the noise reduction and voxelisation process. Red points are points excluded by this cleaning process, while green points are used in further analysis.

¹¹⁷ **4.4 LiDAR analysis**

¹¹⁸ **4.4.1 Foliage density profiles**

¹¹⁹ To estimate subplot foliage density profiles, first the point cloud was cropped to a 10 m diameter cylinder of infinite height. Then the `filters.pmf` (Progressive Morphological Filter - PMF) ¹²⁰ PDAL function was used to identify ground points (*sensu* Zhang et al. 2003). The `filters.hag_nn` ¹²¹ (Nearest Neighbour) PDAL function was used to generate height above ground of each point within ¹²² the cylinder. Points below ground level were then discarded. Height profile points were exported ¹²³ to a XYZ file then imported into R for further processing. ¹²⁴

¹²⁵ We excluded points above the 99.9th percentile of height, under the assumption that these often ¹²⁶ constituted noise that had not been adequately removed by PDAL.

¹²⁷ In R, within each 5 cm width vertical layer, we calculated the foliage density as the proportion of

128 filled 5 cm³ voxels. We filtered the point cloud data to the tree canopy, excluding grass. We iden-
 129 tified the breakpoint between the grass understorey and the tree canopy as the first local minima
 130 above 1.3 m from the ground.

131 We extracted statistics from the foliage density profile for use in statistical analysis. We first smoothed
 132 the density profile using a loess model with a span of 0.1. We then calculated the number of local
 133 maxima and minima along the profile. We defined local maxima and minima as points where the
 134 foliage density of the surrounding 50 cm of 5 cm bins was lower or higher, respectively.

135 We calculated the effective number of layers (ENL), using the true-numbers equivalent of the Shan-
 136 non diversity index (*sensu* (Ehbrecht et al., 2016)). We also calculated the conventional Shannon
 137 diversity index on the foliage density of 50 cm bins:

$$H' = - \sum_{i=1}^N p_i \ln p_i \quad (13)$$

138 Where N is the number of 50 cm bins in the height profile, and p_i is the proportion of filled voxels
 139 in layer i (foliage density).

140 We calculated the area under the curve of foliage density using trapezoid estimation.

141 We extracted the height of the maximum foliage density peak, and calculated the difference be-
 142 tween the highest and lowest local maxima. We also extracted the maximum canopy height within
 143 the subplot.

144 We calculated the coefficient of variation of the point cloud height distribution.

145 To describe the uniformity of the foliage density distribution we used Ripley's L function, which is
 146 more commonly used in describing spatial variation across a 2 dimensional surface. Ripley's L is
 147 an adjustment to Ripley's K, defined as:

$$\hat{K}(t) = \lambda^{-1} \sum_{i \neq j} \frac{I(d_{ij} < t)}{n} \quad (14)$$

$$\hat{L}(t) = \left(\frac{\hat{K}(t)}{\pi} \right)^{1/2} \quad (15)$$

148 We also used the standard error of a linear model of foliage density and height as a simple single
 149 number method of describing the uniformity of foliage density. Under a completely even distri-
 150 bution of foliage material through the canopy, the standard error should be zero, while clumping
 151 causes deviations from this uniform distribution and increases the standard error.

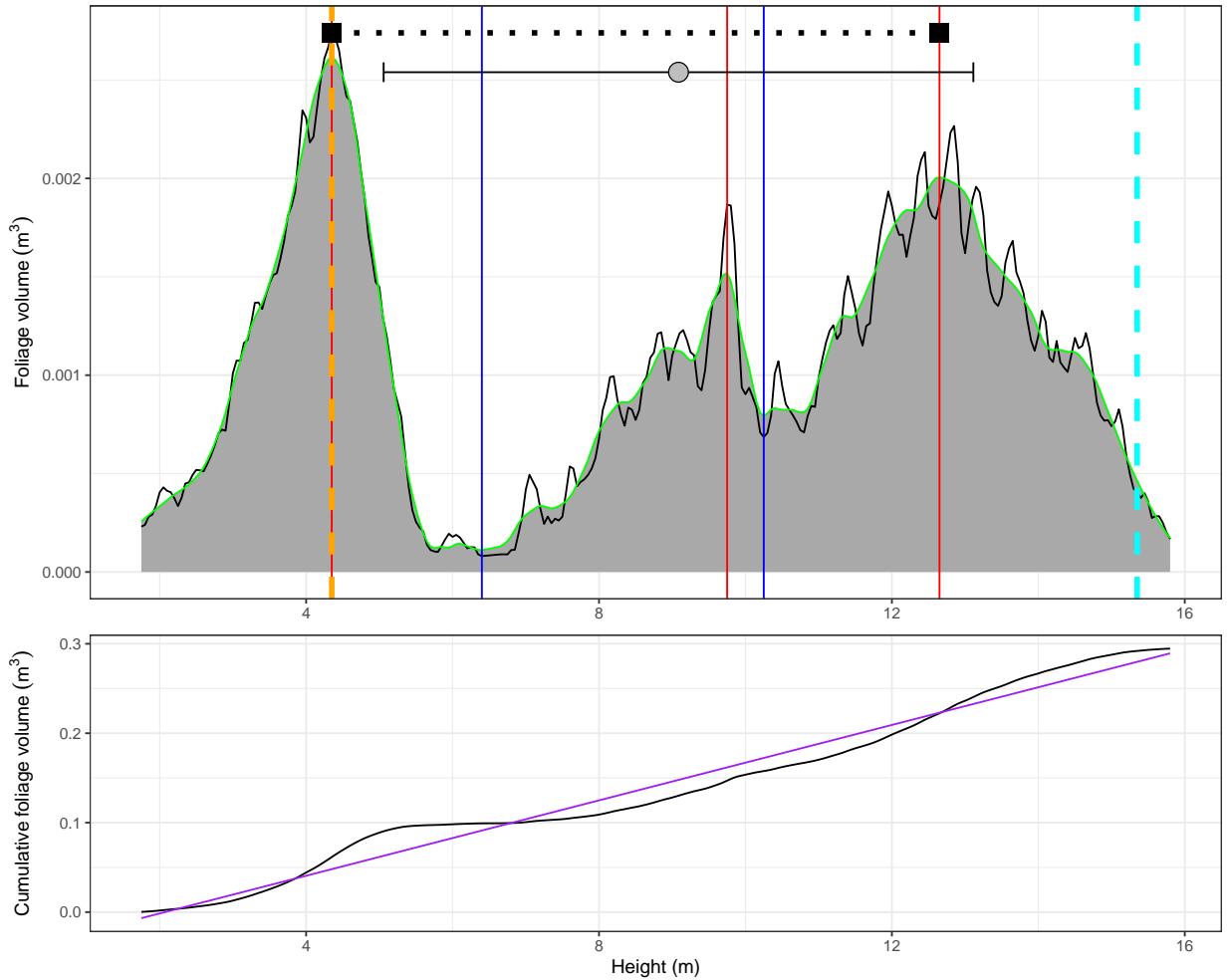


Figure 9: Subplot foliage volume height profile (top) and cumulative foliage volume profile (bottom) for a subplot in Bicuar National Park, to illustrate some of the canopy structure metrics extracted from each height profile. Starting with the top panel: the red lines denote peaks in the distribution, while blue lines represent troughs. The dashed orange line shows the height of the highest peak of foliage density. The dashed cyan line shows the 99th percentile of canopy height, used here as a measure of canopy top height across the subplot. The black squares connected by a dotted line show the layer differentiation, the height difference between the lowest peak and highest peak of foliage density *sensu* Palace et al. (2015). The grey point with interval lines shows the mean foliage volume height ± 1 standard deviation. The black trace shows the foliage density height profile, and the green trace shows the loess model fitted to the data, with the area under the canopy shaded grey. The bottom panel: the black trace shows the cumulative foliage volume through the canopy, taken from the loess fit in the top panel. The purple line shows the line of best fit of a linear model through this data. Not illustrated are the effective number of layers, calculated from the Shannon entropy of foliage volume in 0.5 m height bins.

152 **4.4.2 Canopy cover**

153 Due to terrestrial LiDAR measurement locations being spread over the subplot to avoid occlusion
154 of canopy material, we simulated a scan position at the centre of the subplot using the point cloud
155 data from all scans per subplot. Similar to the processing chain for the foliage density profiles,
156 PDAL was used to crop the point cloud to a 20 m cylinder around the subplot centre, then used
157 `filters.hag_nn` to classify ground points and recalculate height above ground. We cropped the
158 point cloud to points above 1.3 m, with a 50 cm exclusion sphere around the scan position at 1.3
159 m above the ground. The point cloud was converted to a POV-Ray object, where each point was
160 transformed to a 1 cm³ cube. POV-Ray was then used to produce a ray-traced image. As with
161 the hemispherical photos, we used a fisheye lens with an equisolid projection and a view angle of
162 180°, located at the subplot centre, at the same height as the hemispherical photo, with the top
163 of the camera facing magnetic north and the camera facing straight up. Each cube was set as a
164 non-reflective object, and the sky had an equal gamma of 1.0. POV-Ray produced an image of
165 4016x4016 px, identical to the cropped circular dimensions of the images produced by the hemi-
166 spherical photos.

167 Simple canopy cover as seen from the ground was measured using two methods: 1) hemispheri-
168 cal photography and 2) terrestrial LiDAR. Hemiphot () was used to estimate cover from both the
169 hemispherical photos and the TLS POV-Ray simulation. Hemiphot calculates canopy cover in 90
170 evenly sized concentric rings. To obtain the total cover of an image:

$$C_\alpha = 1 - G_{\text{tot}} = \sum_{\alpha=0.5}^{\alpha=89.5} (G_\alpha A_\alpha / A_{\text{tot}}) \quad (16)$$

171 Where G_α is the fraction of unfilled pixels in ring α , A_α is the sky area of the ring segment, and
172 A_{tot} is the total sky area of the hemisphere.

173 We compared canopy cover estimates from both the TLS and hemispherical photo using a linear
174 mixed model which accounted for variation among plots and between the two sites. While plots in
175 Mtarure had a marginally steeper slope, this difference was not significant. We found that hemi-
176 spherical photography almost exclusively under-estimated canopy cover, except in the most open
177 subplots. Additionally, at lower canopy cover the under-estimation of canopy cover by hemispheri-
178 cal photography was larger (Figure 10).

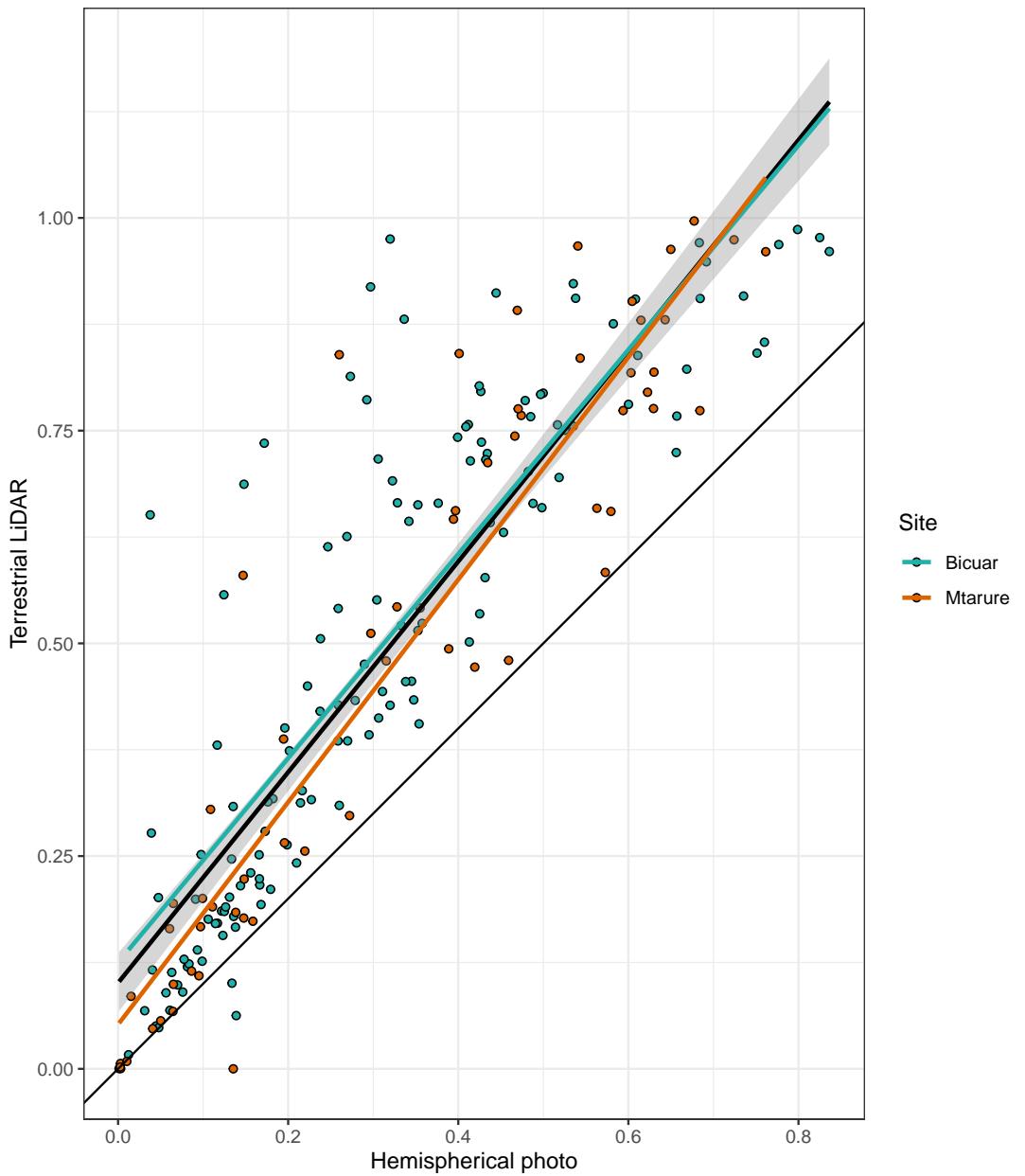


Figure 10: Comparison of canopy cover estimation from TLS and hemispherical photography. The black line of best fit is a linear model of all points ± 1 standard error, while the coloured lines are site specific linear models.



Figure 11: Comparison of hemispherical images for a single subplot in Bicuar National Park. The left image is generated from a hemispherical photo, while the right image is generated from multiple laser scans modelled as cubic voxels with POV-Ray (right).

179 **4.4.3 Grass biomass estimation**

180 An allometric model was developed to estimate grass biomass at every disc-pasture sample point
181 using the grass biomass sample masses. This model was only developed for Angola where grass
182 biomass samples were weighed. The model consisted of a linear mixed effects regression testing
183 the relationship between disc-pasture height (independent) and grass biomass (dependent), with a
184 random slope term for each 1 ha plot.

185 grass volume was measured from TLS point cloud data following the methodology of. First the
186 point cloud was cropped to points below 2 m. The point cloud was then aggregated to cubic vox-
187 els of 2 cm^3 . Within each vertical 2 cm^2 column, the mean height of points was calculated, then
188 the volume below the mean was assumed to be entirely filled with grass material.

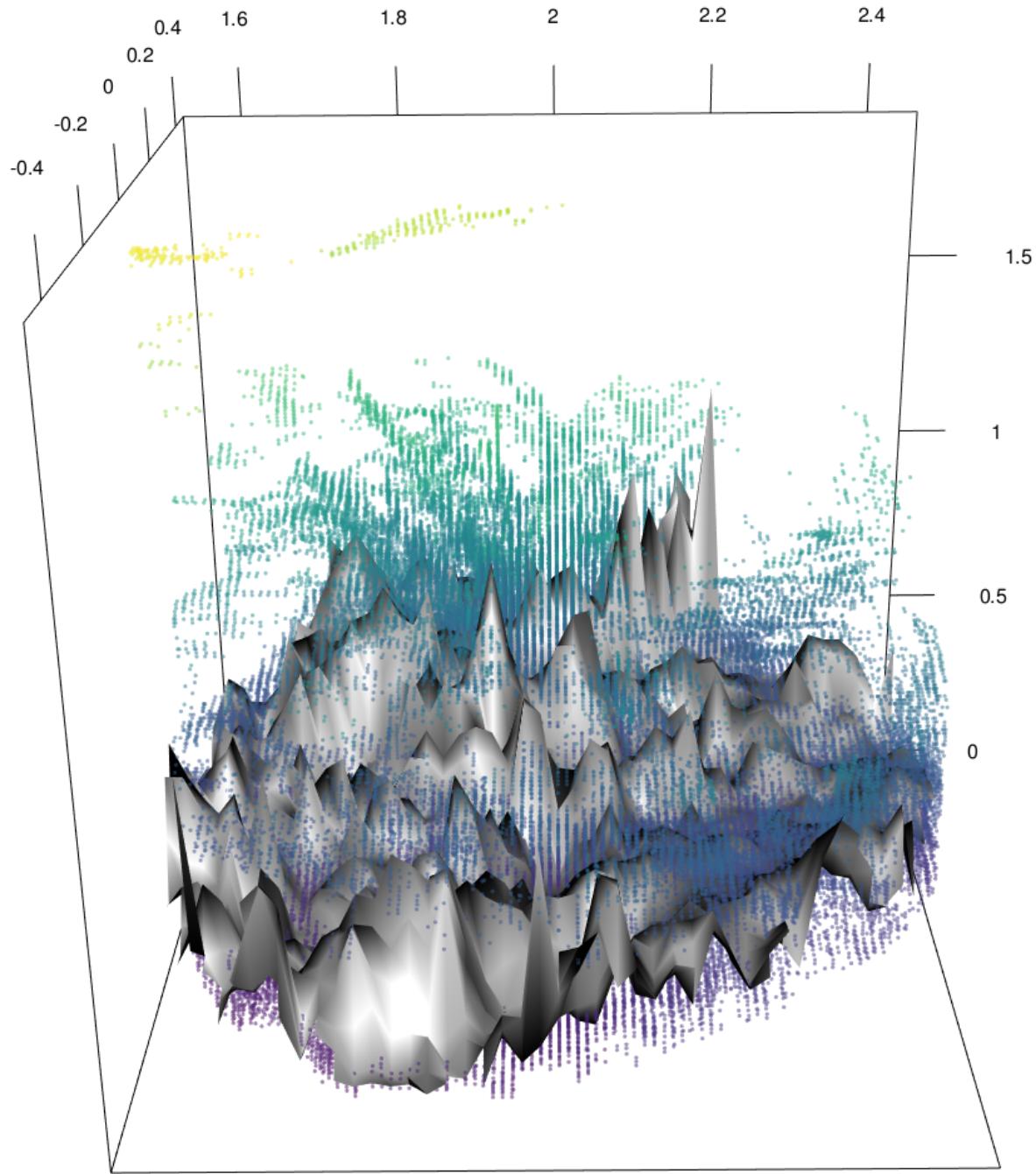


Figure 12: Point cloud with mean heights for each 2 cm^2 column labelled and the estimated grass volume below.

¹⁸⁹ **4.4.4 Canopy rugosity**

¹⁹⁰ The canopy rugosity of each 1 ha plot was estimated. All scans from each plot were merged to a
¹⁹¹ single point cloud, and noise reduction was performed as described above and the cloud was vox-
¹⁹² elised to 10 cm^3 cubic voxels. The point cloud was cropped to the plot boundaries, which were
¹⁹³ located with dGPS similar to the LiDAR targets.

¹⁹⁴ A canopy height model was produced to describe the upper canopy surface. The 99th percentile of

195 height in each 10 cm^2 vertical column was extracted. The maximum height was not used as this
 196 occasionally constituted a severe outlier which skewed further canopy height model smoothing.
 197 We used the pit-filling algorithm described in Khosravipour et al. (2014) to smooth the canopy
 198 height profile by removing gaps within trees caused by incomplete penetration of the LiDAR beam
 199 (Figure 13).

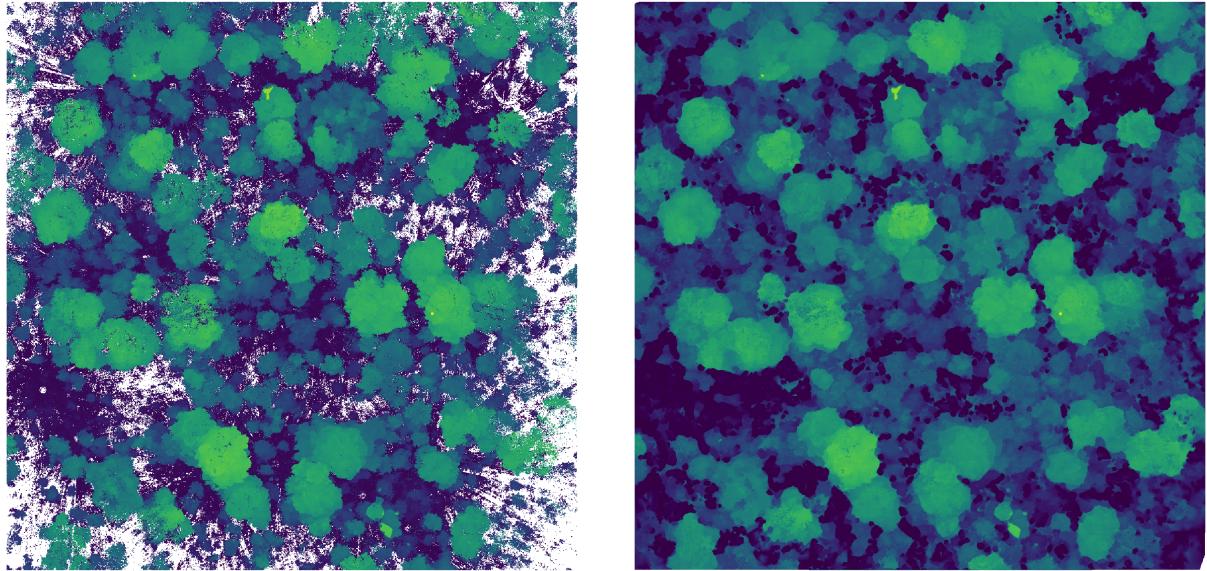


Figure 13: Top-down view of a 1 ha plot in Bicuar National Park. a) shows the point cloud after voxelisation, noise reduction, and taking the 99th percentile of stem height in each 5 cm vertical bin. b) shows the same point cloud after pit filling to generate a smooth canopy height profile. Points are coloured according to point height from the ground.

200 From the canopy height profile we extracted a number of statistics for use in statistical modelling.
 201 We calculated the mean and coefficient of variation of canopy height across the plot (canopy ru-
 202 gosity), following (Parker and Russ, 2004). We calculated the Topographic Ruggedness Index
 203 (TRI) as the mean of absolute differences between the heights of each column and the height of
 204 its eight surrounding cells (Wilson et al., 2007). From this we estimated the plot level mean TRI
 205 and coefficient of variation.
 206 We also calculated a second measure of canopy rugosity (R_c) following Hardiman et al. (2011),
 207 using all point cloud data rather than just the top surface:

$$R_c = \sigma(\sigma G_z)_x \quad (17)$$

208 Where G_z is the vertical height axis z , x is the horizontal axis, and σ is the standard deviation.

209 **5 Statistical analysis**

210 All linear mixed effects models were conducted using the `{lmer}` package in R version 4.0.2 (R
211 Core Team, 2020).

212 **5.1 Foliage density profiles**

213 We conducted a number of linear mixed effects models to assess the effects of tree diversity and
214 stand structure on various aspects of canopy structure measured at the 10 m subplot scale. Lin-
215 ear mixed effects models were used to account for the non-independence of samples caused by
216 the nested sampling structure of subplots within plots, and plots within sites. For each subplot
217 canopy structure measure, we created a linear mixed effects model with fixed effects of subplot
218 species richness, and tree spatial structure using the adapted Hegyi index (H_i) and the coefficient
219 of variation of stem diameter. We compared the standardized effect sizes of each fixed effect to
220 understand the relative effect of species richness and spatial structure on canopy structure. We
221 compared models with all combinations of fixed effects to understand which combination of fixed
222 effects best explained variation in each subplot canopy structure measure. We also compared mod-
223 els to a null model including only random effects of plot and site to evaluate whether this ‘best’
224 model explained real variation in canopy structure.

225 **5.2 Grass biomass**

226 To estimate the correlation between grass volume estimated by TLS and grass biomass estimated
227 from the allometry of DPM height and grass biomass samples, we conducted a linear mixed ef-
228 fects model of grass biomass vs. grass volume, with nested random slope terms for each 1 ha plot
229 nested within site.

230 We conducted a linear mixed effects model to assess the effects of canopy structure on grass vol-
231 ume, with random slope terms for each 1 ha plot nested within site. We began with a maximal
232 model which included fixed effects of subplot tree species richness, stem density, TLS canopy cover,
233 layer diversity, height of maximum foliage density, standard deviation of the foliage density pro-
234 file, and our simple measure of foliage density uniformity. We re-fitted the model with all possible
235 combinations of fixed and random effects and compared AIC, BIC, and log-likelihood to determine
236 which combination of explanatory variables best accounted for variation in grass volume. Once
237 this ‘best model’ had been identified we extracted standardized effect sizes for each fixed effect to
238 compare their relative contribution to the model. We also compared random effects for each fixed
239 effect to understand how the relationship differed between the two sites.

240 **5.3 Canopy rugosity**

241 To understand the effect of species composition and stand structure on whole-plot canopy rugos-
242 ity, we conducted a linear mixed effects model with fixed effects of tree species shannon diver-
243 sity index, stem density, spatial mingling index and winkelmann, with random intercept terms for
244 each site. We extracted slopes for each fixed effect to compare their effect sizes and compared our
245 model with a null model which consisted only of the random effect of site and the fixed effect of
246 stem density.

247 **References**

- 248 Bransby, D.I. and N.M. Tainton (1977). “The disc pasture meter : Possible applications in grazing
249 management”. In: *Proceedings of the Annual Congresses of the Grassland Society of Southern
250 Africa* 12.1, pp. 115–118. DOI: 10.1080/00725560.1977.9648818.
- 251 Brusa, A. and D. E. Bunker (2014). “Increasing the precision of canopy closure estimates from
252 hemispherical photography: Blue channel analysis and under-exposure”. In: *Agricultural and
253 Forest Meteorology* 195–196, pp. 102–107. DOI: 10.1016/j.agrformet.2014.05.001.
- 254 Chen, X. et al. (Sept. 2011). *Trimble RTX, an innovative new approach for network RTK*. Tech.
255 rep. Portland OR, USA: International Technical Meeting of the Satellite Division of the Insti-
256 tute of Navigation, ION GNSS, pp. 2214–2219.
- 257 Ehbrecht, Martin et al. (2016). “Effective number of layers: A new measure for quantifying three-
258 dimensional stand structure based on sampling with terrestrial LiDAR”. In: *Forest Ecology and
259 Management* 380, pp. 212–223. DOI: 10.1016/j.foreco.2016.09.003.
- 260 Frazer, Gordon W. et al. (Sept. 2001). “A comparison of digital and film fisheye photography for
261 analysis of forest canopy structure and gap light transmission”. In: *Agricultural and Forest Mete-
262 orology* 109.4, pp. 249–263. DOI: 10.1016/s0168-1923(01)00274-x. URL: <https://doi.org/10.1016%2Fs0168-1923%2801%2900274-x>.
- 264 Hardiman, Brady S et al. (2011). “The role of canopy structural complexity in wood net primary
265 production of a maturing northern deciduous forest”. In: *Ecology* 92.9, pp. 1818–1827. DOI: 10.
266 1890/10-2192.1.
- 267 Hegyi, F. (1974). “A simulation model for managing jack-pine stands”. In: *Royal College of Forestry,
268 editor*. Stockholm, Sweden: Royal College of Forestry, pp. 74–90.
- 269 Hu, Lile and Jiaojun Zhu (May 2009). “Determination of the tridimensional shape of canopy gaps
270 using two hemispherical photographs”. In: *Agricultural and Forest Meteorology* 149.5, pp. 862–
271 872. DOI: 10.1016/j.agrformet.2008.11.008. URL: <https://doi.org/10.1016%2Fj.agrformet.2008.11.008>.

- 273 Hui, G. and M. Albert (2004). "Stichprobensimulationen zur Schätzung nachbarschaftsbezogener
274 Strukturparameter in Waldbeständen [Simulation studies for estimating neighborhood-based
275 structural parameters in forest stands]". In: *Allgemeine Forst und Jagdzeitung* 175, pp. 10–11.
- 276 Hui, G., K. von Gadow, et al. (2007). *Structure-based forest management*. Beijing, China: China
277 Forestry Publishing House.
- 278 Hui, G., X. Zhao, et al. (2011). "Evaluating tree species spatial diversity based on neighbourhood
279 relationships". In: *Forest Science* 57.4, pp. 292–300. DOI: <http://dx.doi.org/10.1093/forestscience/57.4.292>.
- 281 Khosravipour, Anahita et al. (2014). "Generating Pit-free Canopy Height Models from Airborne
282 Lidar". In: *Photogrammetric Engineering & Remote Sensing* 80.9, pp. 863–872. DOI: 10.14358/
283 pers.80.9.863.
- 284 Palace, Michael W. et al. (2015). "Estimating forest structure in a tropical forest using field mea-
285 surements, a synthetic model and discrete return lidar data". In: *Remote Sensing of Environ-
286 ment* 161, pp. 1–11. DOI: 10.1016/j.rse.2015.01.020.
- 287 Parker, Geoffrey G and Mary E Russ (2004). "The canopy surface and stand development: assess-
288 ing forest canopy structure and complexity with near-surface altimetry". In: *Forest Ecology and
289 Management* 189.1-3, pp. 307–315. DOI: 10.1016/j.foreco.2003.09.001.
- 290 R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation
291 for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.
- 292 Rusu, Radu Bogdan et al. (2008). "Towards 3D Point cloud based object maps for household envi-
293 ronments". In: *Robotics and Autonomous Systems* 56.11, pp. 927–941. DOI: 10.1016/j.robot.
294 2008.08.005.
- 295 SEOSAW (2020). "A network to understand the changing socio-ecology of the southern African
296 woodlands (SEOSAW): Challenges, benefits, and methods". In: *PLANTS, PEOPLE, PLANET*.
297 DOI: 10.1002/ppp3.10168.
- 298 von Gadow, K. and G. Hui (2002). *Characterising forest spatial structure and diversity*. Ed. by L.
299 Bjoerk. Lund, Sweden, pp. 20–30.
- 300 Wilson, Margaret F. J. et al. (2007). "Multiscale Terrain Analysis of Multibeam Bathymetry Data
301 for Habitat Mapping on the Continental Slope". In: *Marine Geodesy* 30.1-2, pp. 3–35. DOI: 10.
302 1080/01490410701295962.
- 303 Zhang, Keqi et al. (2003). "A progressive morphological filter for removing nonground measure-
304 ments from airborne LIDAR data". In: *IEEE Transactions on Geoscience and Remote Sensing*
305 41.4, pp. 872–882. DOI: 10.1109/tgrs.2003.810682.