

# Estimation of woodland canopy structure with terrestrial LiDAR: expanded methods

John L. Godlee

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Sampling</b>	<b>2</b>
<b>3</b>	<b>Field measurements</b>	<b>5</b>
3.1	Trees . . . . .	5
3.2	Grass biomass . . . . .	6
3.3	Hemispherical photography . . . . .	6
3.4	Stand structure . . . . .	8
<b>4</b>	<b>Terrestrial LiDAR</b>	<b>10</b>
4.1	Registration . . . . .	11
4.2	Voxelisation . . . . .	11
4.3	Noise reduction . . . . .	11
4.4	LiDAR analysis . . . . .	12
4.4.1	Foliage density profiles . . . . .	12
4.4.2	Canopy cover . . . . .	15
4.4.3	Grass biomass estimation . . . . .	17
4.4.4	Canopy rugosity . . . . .	18
<b>5</b>	<b>Statistical analysis</b>	<b>20</b>
5.1	Foliage density profiles . . . . .	20
5.2	Grass biomass . . . . .	20

5.3 Canopy rugosity . . . . .	21
-------------------------------	----

## **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.

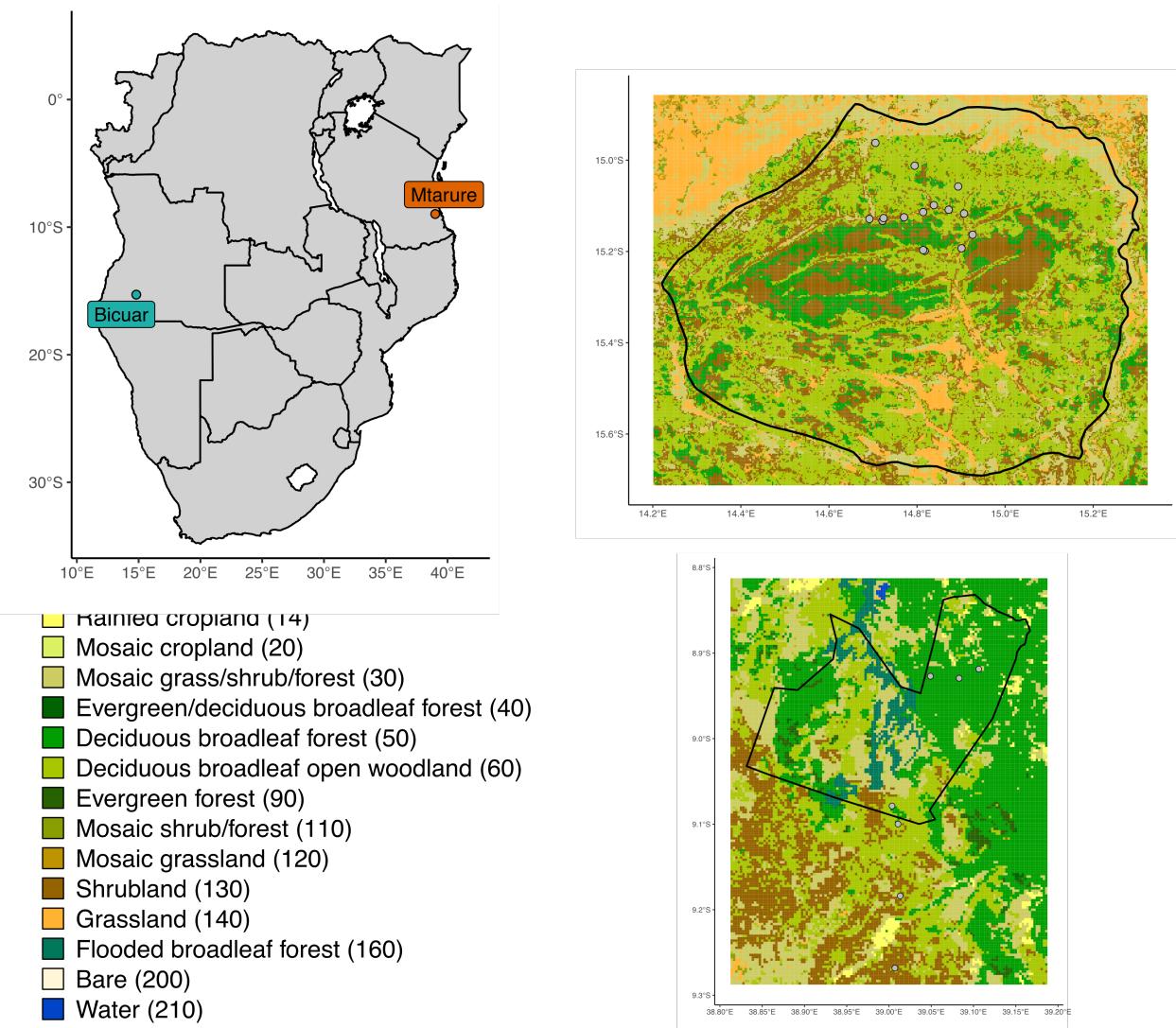


Figure 1: Location of study sites within southern Africa (a), and of 1 ha plots within each site. The black polygons denote the boundaries of protected areas which encompass the majority of study sites, Bicuar National Park in Angola (b), and Mtarure Forest Reserve in Tanzania (c). Each site map is coloured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

Site	MAT	MAP	Temp. range	CWD
	(°C)	(mm y <sup>-1</sup> )	(°C)	(mm y <sup>-1</sup> )
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. MAT = Mean Annual Temperature. MAP = Mean Annual Precipitation. Temp. range = Temperature range, calculated as the mean of annual difference between highest temperature of hottest month and lowest temperature of coldest month. CWD = Climatic Water Deficit, calculated as the sum of the difference between monthly rainfall and monthly evapotranspiration when the difference is negative, sensu Chave et al. (2014).

- 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 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).

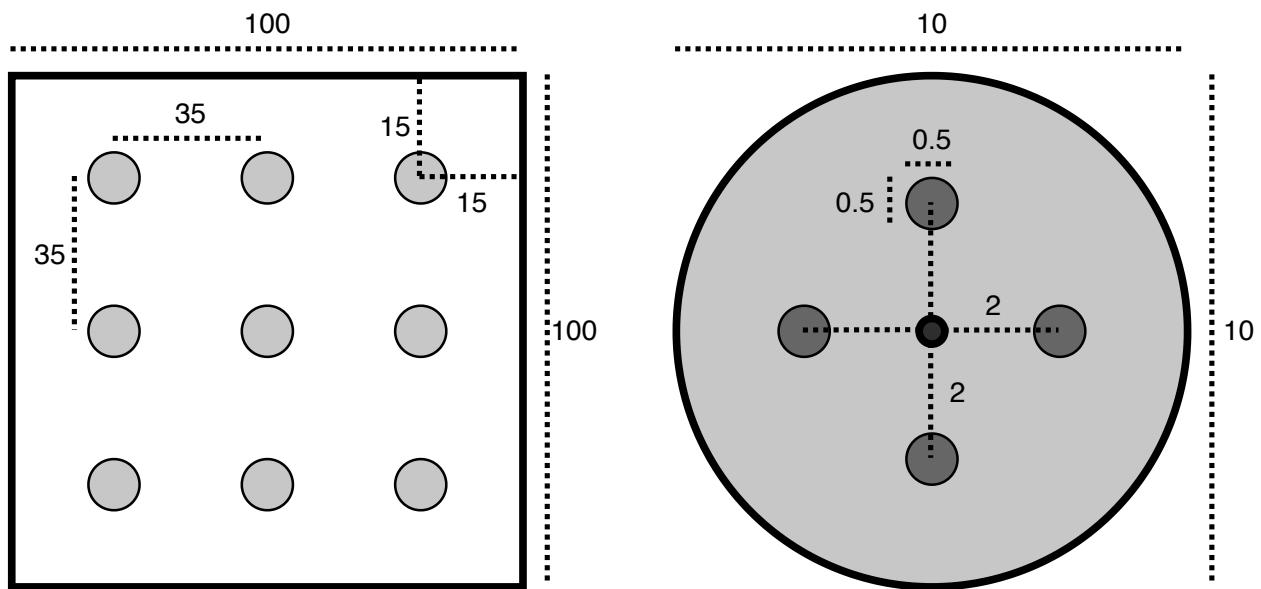


Figure 2: The layout of 10 m diameter subplots within each 1 ha square plot. Each subplot is situated inside a 15 m buffer from the plot edge, with 35 m between subplot centres. Subplots are arranged in a 3x3 grid. Disc-pasture measurements and biomass samples are located in cardinal directions 2 m from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

### 3 Field measurements

#### 3.1 Trees

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

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

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

- Species
- Height to top of canopy
- Canopy area, ellipse from two perpendicular measurements (Figure 3)
- Distance from subplot centre
- Compass direction from subplot centre

# Plot edges

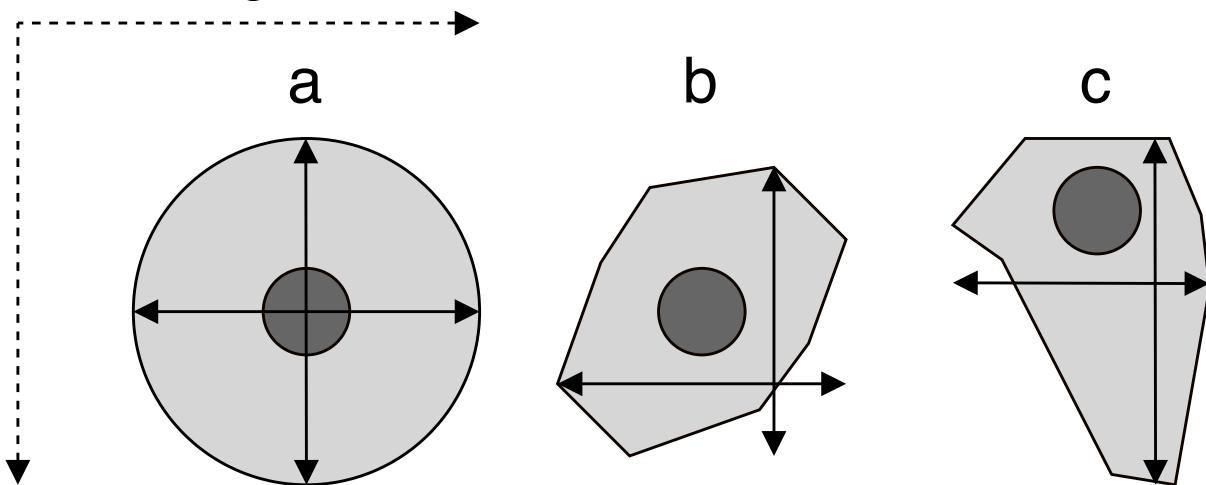


Figure 3: Examples of tree crowns as viewed from above to demonstrate how crown extent measurements are located. Darker grey circles show the main stem while pale grey polygons show the maximum extent of the crown. Extent measurements are taken parallel to the plot edges. a) shows a perfectly circular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

## <sup>32</sup> 3.2 Grass biomass

<sup>33</sup> Grass volume and biomass within each subplot was estimated from four sample points located 2  
<sup>34</sup> m from the subplot centre in cardinal directions (Figure 2). At each point, a disc-pasture meter  
<sup>35</sup> measurement was taken with a 45.8 cm radius disc weighing exactly 1.5 kg (Bransby and Tain-  
<sup>36</sup> ton, 1977). Small woody stems were removed from disc-pasture sample points before the disc-  
<sup>37</sup> pasture measurement was taken. The location of the sample point was moved if the designated  
<sup>38</sup> point intersected with coarse woody debris, rocks, shrubs, or standing trees. Within each 1 ha  
<sup>39</sup> plot, biomass harvesting was conducted at nine randomly allocated disc-pasture sample points.  
<sup>40</sup> Tree leaf litter was removed from biomass samples. Biomass harvesting involved clipping all grass  
<sup>41</sup> material within the 45.8 cm radius to ground level, taking care not to include roots. Grass sam-  
<sup>42</sup> ples from Angola were dried until the mass remained constant ( $\pm 5$  g) for  $>48$  hours, then weighed  
<sup>43</sup> to ascertain the grass biomass. Grass samples from Tanzania could not be processed due to cur-  
<sup>44</sup> tailment of fieldwork due to COVID-19 travel restrictions.

## <sup>45</sup> 3.3 Hemispherical photography

<sup>46</sup> At the centre of each subplot a single photograph was taken with a Nikon D750 full-frame DSLR  
<sup>47</sup> 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  $\theta$  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 $\mu\text{m}$
Sensor resolution	24.3 MP
Shutter speed	>1/60s
Aperture	5-7
ISO	100-200
Exposure compensation	-0.7 (Brusa and Bunker, 2014)
Focus	$\infty$ (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. The spatial mingling index is a spatially explicit estimate of the degree to which species are  
 66 spatially mixed within a plot:

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

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

(4)

67 where  $k$  is the number of nearest neighbours considered for each reference tree,  $S_i$  is the number  
 68 of species found among the  $k$  nearest neighbours of tree  $i$ ,  $S$  is the total number of species in the  
 69 plot, and  $N$  is the total number of trees in the plot. In our case we used the conventional value of  
 70  $k = 4$ .

71 We also calculated the Winkelmass  $W$  according to von Gadow and Hui (2002) at the plot level.  
 72 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)

73 where  $\alpha_j$  is the angle between consecutive neighbours and  $\alpha_0$  is the principal angle, where  $\alpha_0 =$   
 74  $360/k$ .

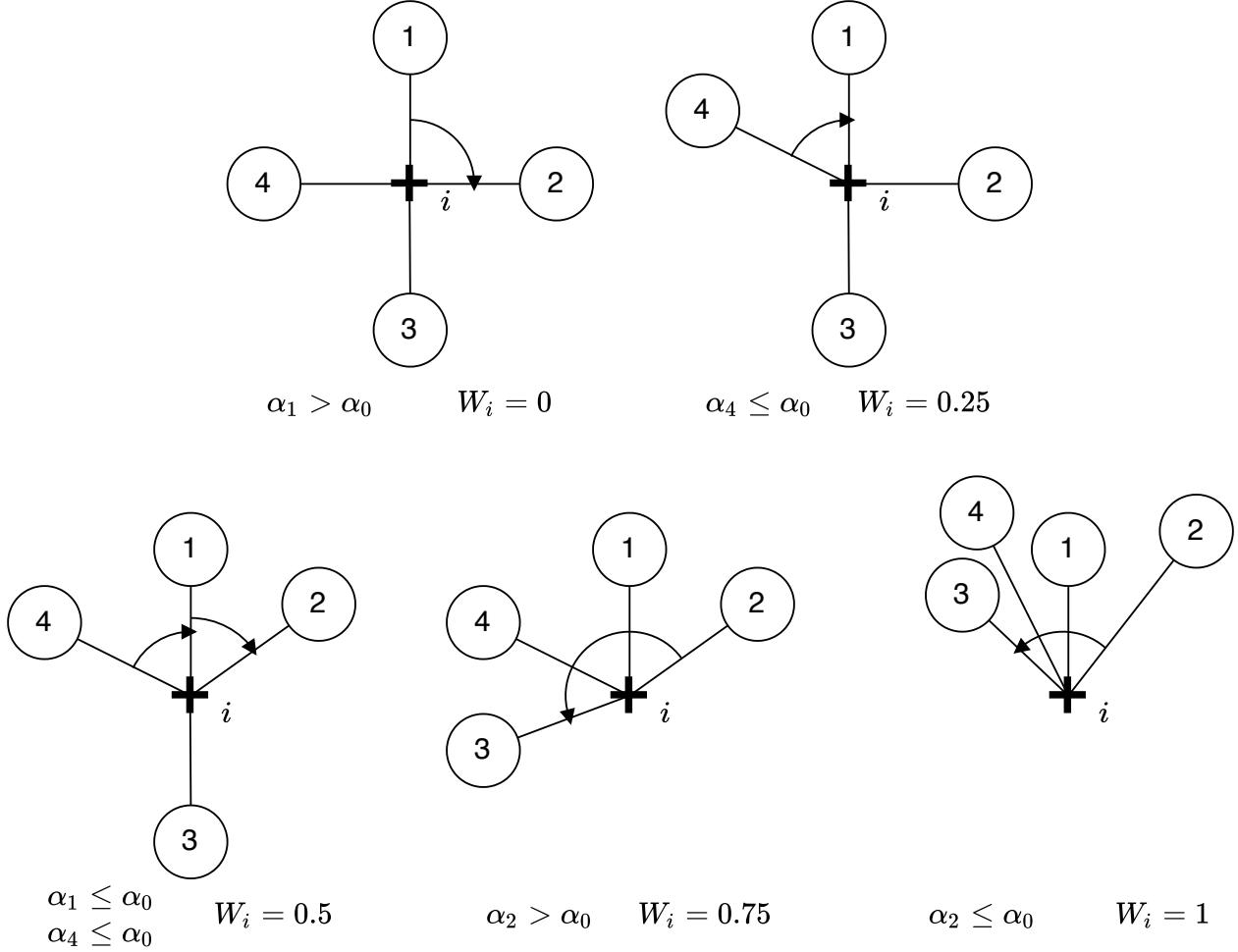


Figure 4: Possible values of  $W_i$  at a sample point  $i$ , denoted by a cross. Neighbours are represented as circles numbered sequentially from 1 to 4, where  $k = 4$ . The angles of arrows in each example are given below, along with the Winkelmaass for that example.

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

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

79 where  $D_j$  is the stem diameter of neighbour tree  $j$  and  $L_j$  is the distance of the neighbour from  
 80 the subplot centre.

## 81 4 Terrestrial LIDAR

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

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

94 **4.1 Registration**

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

98 **4.2 Voxellation**

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

102 LAZ files were voxelised to different voxel sizes depending on the application of the data. For  
103 grass biomass estimation, we used  $2 \text{ cm}^3$  cubic voxels, while for subplot height profile estimation  
104 we used  $5 \text{ cm}^3$  voxels, and for whole plot canopy rugosity we used  $10 \text{ cm}^3$  voxels. WHY THO

105 **4.3 Noise reduction**

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

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

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

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

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

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

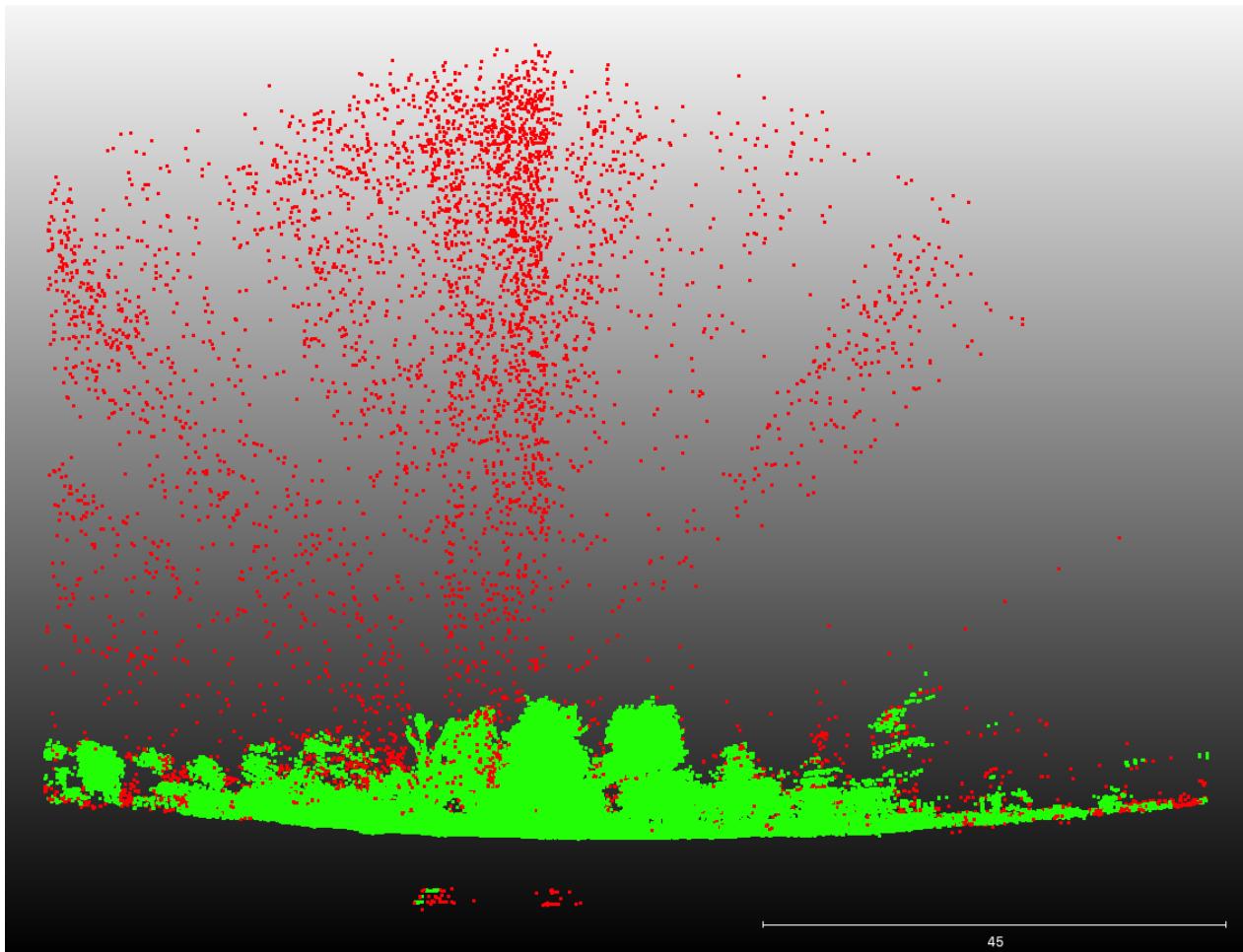


Figure 5: 2 m deep cross section of subplot showing the efficacy of the noise reduction and voxelisation process. Red points are points excluded by this cleaning process, while green points are used in further analysis.

## 112 4.4 LiDAR analysis

### 113 4.4.1 Foliage density profiles

114 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) 115 PDAL function was used to identify ground points (sensu Zhang et al. 2003). The `filters.hag_nn` 116 (Nearest Neighbour) PDAL function was used to generate height above ground of each point within 117 the cylinder. Points below ground level were then discarded. Height profile points were exported 118 to a XYZ file then imported into R for further processing.

119 We excluded points above the 99.9th percentile of height, under the assumption that these often 120 constituted noise that had not been adequately removed by PDAL.

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

123 filled 5 cm<sup>3</sup> voxels. We filtered the point cloud data to the tree canopy, excluding grass. We iden-  
 124 tified the breakpoint between the grass understorey and the tree canopy as the first local minima  
 125 above 1.3 m from the ground.

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

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

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

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

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

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

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

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

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

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

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

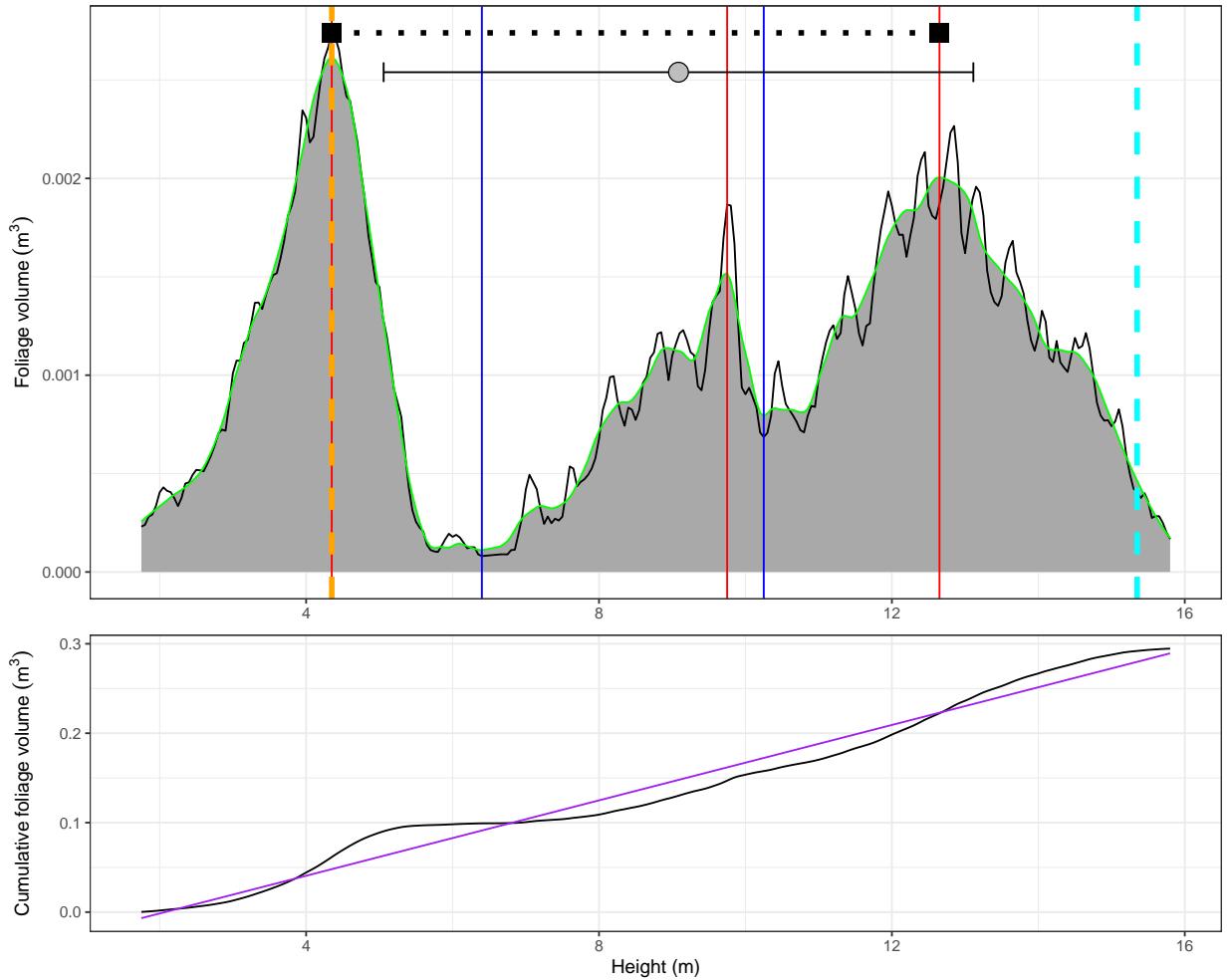


Figure 6: Subplot foliage volume height profile (top) and cumulative foliage volume profile (bottom) for a subplot in Bicuar National Park, to illustrate some of the canopy structure metrics extracted from each height profile. Starting with the top panel: the red lines denote peaks in the distribution, 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  $\pm 1$  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.

147 **4.4.2 Canopy cover**

148 Due to terrestrial LiDAR measurement locations being spread over the subplot to avoid occlusion  
149 of canopy material, we simulated a scan position at the centre of the subplot using the point cloud  
150 data from all scans per subplot. Similar to the processing chain for the foliage density profiles,  
151 PDAL was used to crop the point cloud to a 20 m cylinder around the subplot centre, then used  
152 `filters.hag_nn` to classify ground points and recalculate height above ground. We cropped the  
153 point cloud to points above 1.3 m, with a 50 cm exclusion sphere around the scan position at 1.3  
154 m above the ground. The point cloud was converted to a POV-Ray object, where each point was  
155 transformed to a 1 cm<sup>3</sup> cube. POV-Ray was then used to produce a ray-traced image. As with  
156 the hemispherical photos, we used a fisheye lens with an equisolid projection and a view angle of  
157 180°, located at the subplot centre, at the same height as the hemispherical photo, with the top  
158 of the camera facing magnetic north and the camera facing straight up. Each cube was set as a  
159 non-reflective object, and the sky had an equal gamma of 1.0. POV-Ray produced an image of  
160 4016x4016 px, identical to the cropped circular dimensions of the images produced by the hemi-  
161 spherical photos.

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

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

166 Where  $G_\alpha$  is the fraction of unfilled pixels in ring  $\alpha$ ,  $A_\alpha$  is the sky area of the ring segment, and  
167  $A_{\text{tot}}$  is the total sky area of the hemisphere.

168 We compared canopy cover estimates from both the TLS and hemispherical photo using a linear  
169 mixed model which accounted for variation among plots and between the two sites. While plots in  
170 Mtarure had a marginally steeper slope, this difference was not significant. We found that hemi-  
171 spherical photography almost exclusively under-estimated canopy cover, except in the most open  
172 subplots. Additionally, at lower canopy cover the under-estimation of canopy cover by hemispheri-  
173 cal photography was larger (Figure 7).

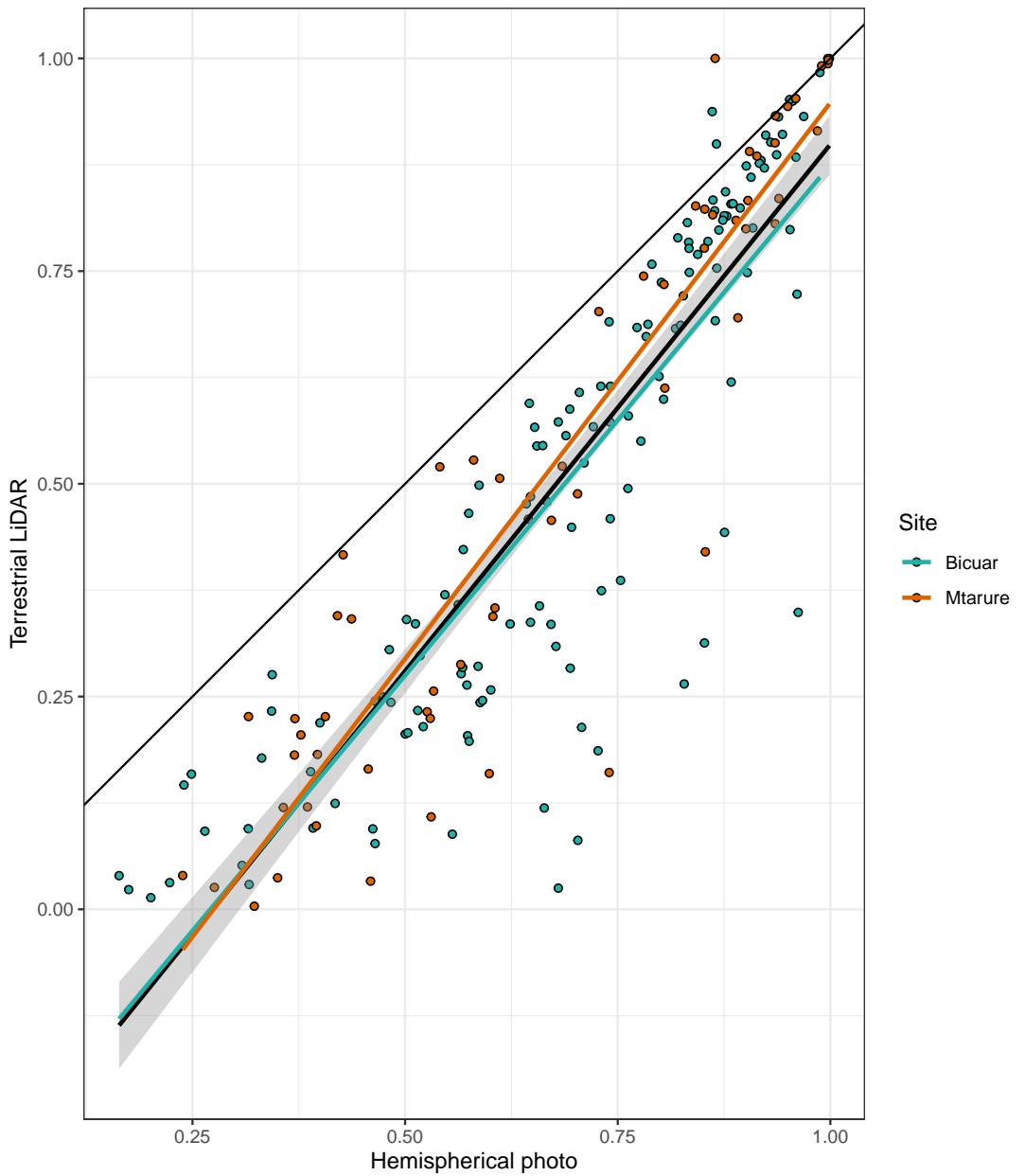


Figure 7: Comparison of canopy cover estimation from TLS and hemispherical photography. The black line of best fit is a linear model of all points  $\pm 1$  standard error, while the coloured lines are site specific linear models.



Figure 8: 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).

174 **4.4.3 Grass biomass estimation**

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

180 grass volume was measured from TLS point cloud data following the methodology of. First the  
181 point cloud was cropped to points below 2 m. The point cloud was then aggregated to cubic vox-  
182 els of  $2 \text{ cm}^3$ . Within each vertical  $2 \text{ cm}^2$  column, the mean height of points was calculated, then  
183 the volume below the mean was assumed to be entirely filled with grass material.

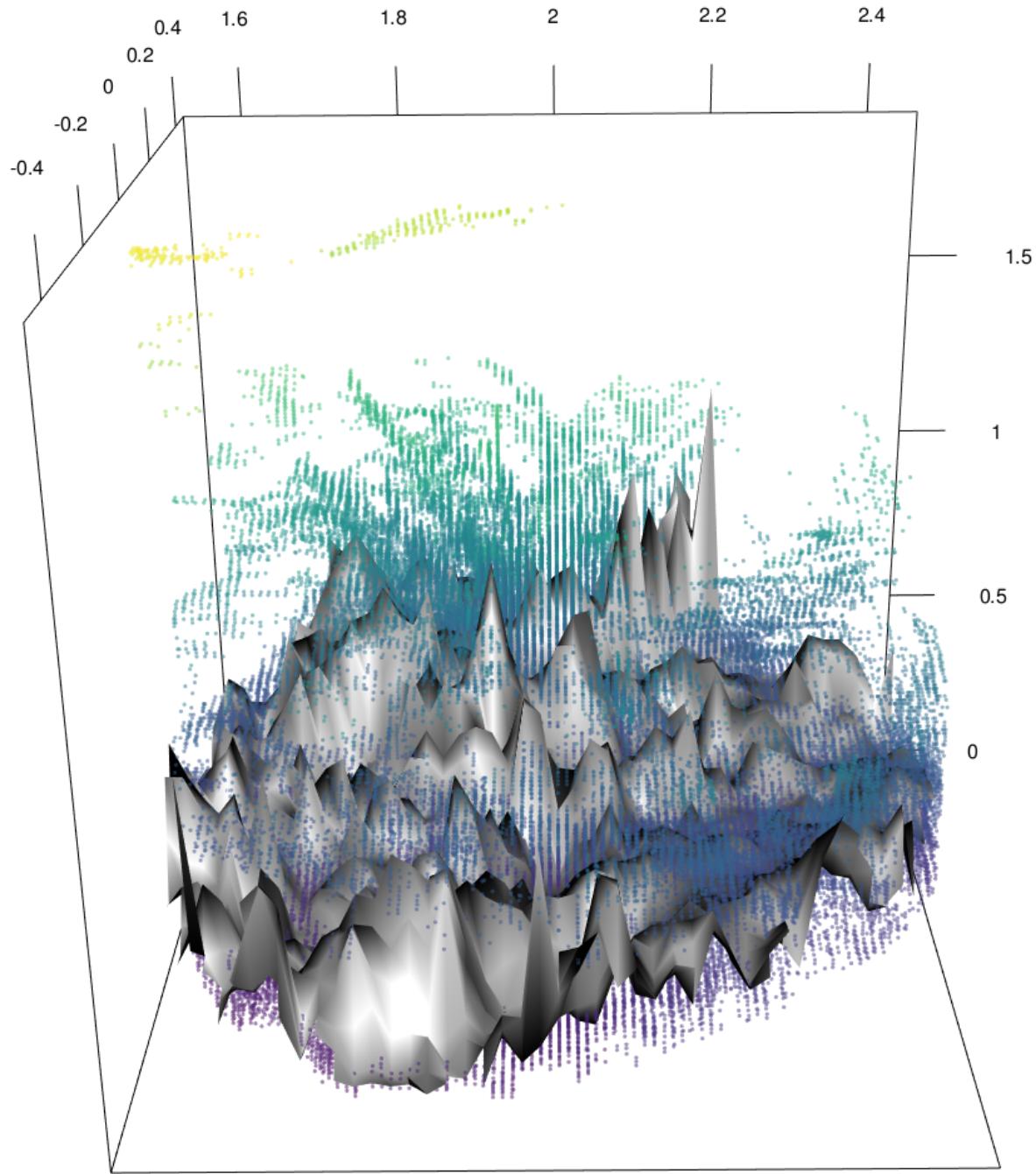


Figure 9: Point cloud with mean heights for each  $2 \text{ cm}^2$  column labelled and the estimated grass volume below.

<sup>184</sup> **4.4.4 Canopy rugosity**

<sup>185</sup> The canopy rugosity of each 1 ha plot was estimated. All scans from each plot were merged to a  
<sup>186</sup> single point cloud, and noise reduction was performed as described above and the cloud was vox-  
<sup>187</sup> elised to  $10 \text{ cm}^3$  cubic voxels. The point cloud was cropped to the plot boundaries, which were  
<sup>188</sup> located with dGPS similar to the LiDAR targets.

<sup>189</sup> A canopy height model was produced to describe the upper canopy surface. The 99th percentile of

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

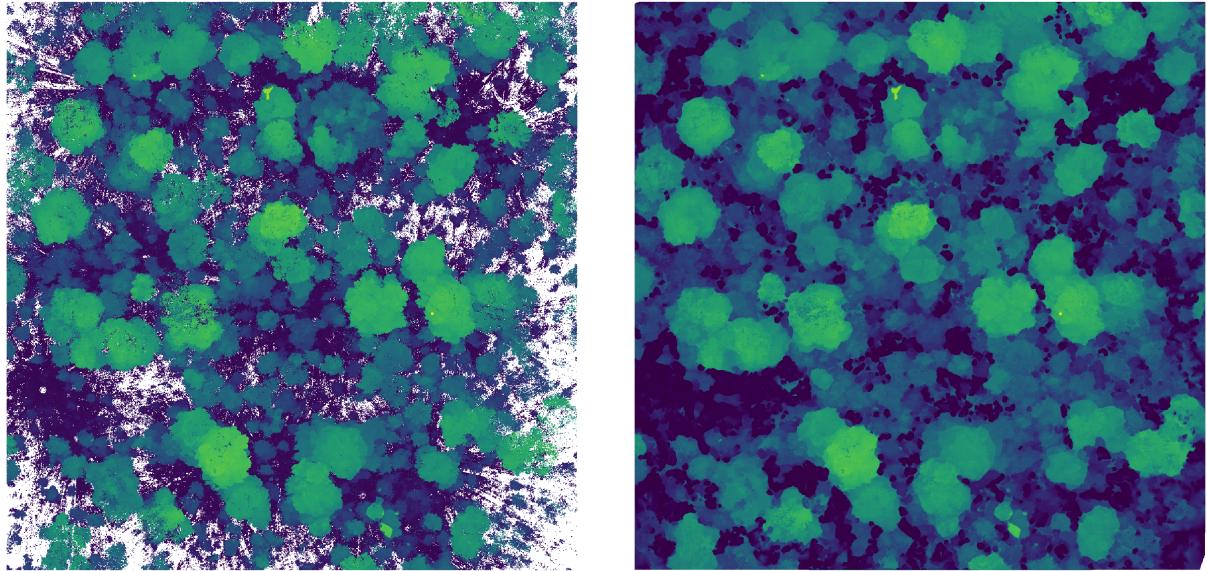


Figure 10: Top-down view of a 1 ha plot in Bicuar National Park. a) shows the point cloud after voxelisation, noise reduction, and taking the 99th percentile of stem height in each 5 cm vertical bin. b) shows the same point cloud after pit filling to generate a smooth canopy height profile. Points are coloured according to point height from the ground.

From the canopy height profile we extracted a number of statistics for use in statistical modelling. We calculated the mean and coefficient of variation of canopy height across the plot (canopy rugosity), following (Parker and Russ, 2004). We calculated the Topographic Ruggedness Index (TRI) as the mean of absolute differences between the heights of each column and the height of its eight surrounding cells (Wilson et al., 2007). From this we estimated the plot level mean TRI and coefficient of variation.

We also calculated a second measure of canopy rugosity ( $R_c$ ) following Hardiman et al. (2011), using all point cloud data rather than just the top surface:

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

Where  $G_z$  is the vertical height axis  $z$ ,  $x$  is the horizontal axis, and  $\sigma$  is the standard deviation.

204 **5 Statistical analysis**

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

207 **5.1 Foliage density profiles**

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

220 **5.2 Grass biomass**

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

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

235 **5.3 Canopy rugosity**

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

242 **References**

- 243 Bransby, D.I. and N.M. Tainton (1977). “The disc pasture meter : Possible applications in grazing  
244 management”. In: *Proceedings of the Annual Congresses of the Grassland Society of Southern  
245 Africa* 12.1, pp. 115–118. DOI: 10.1080/00725560.1977.9648818.
- 246 Brusa, A. and D. E. Bunker (2014). “Increasing the precision of canopy closure estimates from  
247 hemispherical photography: Blue channel analysis and under-exposure”. In: *Agricultural and  
248 Forest Meteorology* 195–196, pp. 102–107. DOI: 10.1016/j.agrformet.2014.05.001.
- 249 Chave, J. et al. (2014). “Improved allometric models to estimate the aboveground biomass of trop-  
250 ical trees”. In: *Global Change Biology* 20.10, pp. 3177–3190. DOI: <http://dx.doi.org/10.1111/gcb.12629>.
- 252 Chen, X. et al. (Sept. 2011). *Trimble RTX, an innovative new approach for network RTK*. Tech.  
253 rep. Portland OR, USA: International Technical Meeting of the Satellite Division of the Insti-  
254 tute of Navigation, ION GNSS, pp. 2214–2219.
- 255 Ehbrecht, Martin et al. (2016). “Effective number of layers: A new measure for quantifying three-  
256 dimensional stand structure based on sampling with terrestrial LiDAR”. In: *Forest Ecology and  
257 Management* 380, pp. 212–223. DOI: 10.1016/j.foreco.2016.09.003.
- 258 Frazer, Gordon W. et al. (Sept. 2001). “A comparison of digital and film fisheye photography for  
259 analysis of forest canopy structure and gap light transmission”. In: *Agricultural and Forest Mete-  
260 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>.
- 262 Hardiman, Brady S et al. (2011). “The role of canopy structural complexity in wood net primary  
263 production of a maturing northern deciduous forest”. In: *Ecology* 92.9, pp. 1818–1827. DOI: 10.  
264 1890/10-2192.1.
- 265 Hegyi, F. (1974). “A simulation model for managing jack-pine stands”. In: *Royal College of Forestry,  
266 editor*. Stockholm, Sweden: Royal College of Forestry, pp. 74–90.

- 267 Hu, Lile and Jiajun Zhu (May 2009). “Determination of the tridimensional shape of canopy gaps  
268 using two hemispherical photographs”. In: *Agricultural and Forest Meteorology* 149.5, pp. 862–  
269 872. DOI: 10.1016/j.agrformet.2008.11.008. URL: <https://doi.org/10.1016%2Fj.agrformet.2008.11.008>.
- 271 Khosravipour, Anahita et al. (2014). “Generating Pit-free Canopy Height Models from Airborne  
272 Lidar”. In: *Photogrammetric Engineering & Remote Sensing* 80.9, pp. 863–872. DOI: 10.14358/  
273 pers.80.9.863.
- 274 Palace, Michael W. et al. (2015). “Estimating forest structure in a tropical forest using field mea-  
275 surements, a synthetic model and discrete return lidar data”. In: *Remote Sensing of Environ-  
276 ment* 161, pp. 1–11. DOI: 10.1016/j.rse.2015.01.020.
- 277 Parker, Geoffrey G and Mary E Russ (2004). “The canopy surface and stand development: assess-  
278 ing forest canopy structure and complexity with near-surface altimetry”. In: *Forest Ecology and  
279 Management* 189.1-3, pp. 307–315. DOI: 10.1016/j.foreco.2003.09.001.
- 280 R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation  
281 for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.
- 282 Rusu, Radu Bogdan et al. (2008). “Towards 3D Point cloud based object maps for household envi-  
283 ronments”. In: *Robotics and Autonomous Systems* 56.11, pp. 927–941. DOI: 10.1016/j.robot.  
284 2008.08.005.
- 285 SEOSAW (2020). “A network to understand the changing socio-ecology of the southern African  
286 woodlands (SEOSAW): Challenges, benefits, and methods”. In: *PLANTS, PEOPLE, PLANET*.  
287 DOI: 10.1002/ppp3.10168.
- 288 von Gadow, K. and G. Hui (2002). *Characterising forest spatial structure and diversity*. Ed. by L.  
289 Bjoerk. Lund, Sweden, pp. 20–30.
- 290 Wilson, Margaret F. J. et al. (2007). “Multiscale Terrain Analysis of Multibeam Bathymetry Data  
291 for Habitat Mapping on the Continental Slope”. In: *Marine Geodesy* 30.1-2, pp. 3–35. DOI: 10.  
292 1080/01490410701295962.
- 293 Zhang, Keqi et al. (2003). “A progressive morphological filter for removing nonground measure-  
294 ments from airborne LIDAR data”. In: *IEEE Transactions on Geoscience and Remote Sensing*  
295 41.4, pp. 872–882. DOI: 10.1109/tgrs.2003.810682.