



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Knights, K;McCarthy, MA;Camac, J;Guillera-Arroita, G

Title:

Efficient effort allocation in line-transect distance sampling of high-density species: When to walk further, measure less-often and gain precision

Date:

2021-06-01

Citation:

Knights, K., McCarthy, M. A., Camac, J. & Guillera-Arroita, G. (2021). Efficient effort allocation in line-transect distance sampling of high-density species: When to walk further, measure less-often and gain precision. *Methods in Ecology and Evolution*, 12 (6), pp.962-970. <https://doi.org/10.1111/2041-210X.13589>.

Persistent Link:

<https://hdl.handle.net/11343/337969>

1

2 MS KATHRYN AINA KNIGHTS (Orcid ID : 0000-0002-9918-9573)

3 DR GURUTZETA GUILLERA-ARROITA (Orcid ID : 0000-0002-8387-5739)

4

5

6 Article type : Research Article

7

8

9 Handling editor: Dr Chris Sutherland

10

11 **Title:** Efficient effort allocation in line-transect distance sampling of high-density species: when to
12 walk further, measure less-often and gain precision

13 **Running title:** Distance sampling optimisation

14 **Authors:** Kathryn Knights, Michael A. McCarthy, James Camac and Gurutzeta Guillera-Arroita

15 **Authors' Institutions:** School of BioSciences, University of Melbourne, Parkville, VIC 3010, Australia

16 **Corresponding author:** Kathryn Knights, kaknights@student.unimelb.edu.au, School of BioSciences,
17 University of Melbourne, Parkville, VIC 3010, Australia

18 **Abstract:**

19 1. Line-transect distance sampling is widely used to estimate population densities by using
20 distances of observed targets from transect lines to model detectability. When the target
21 taxa are high-density, the frequent measuring of distances may make the method seem
22 impractical. We present a method that improves the efficiency of distance sampling when
23 the target species occurs at high-density. Only a proportion of targets are measured to
24 model the detection function, and the time saved on the survey is then used to cover a
25 longer total length of transect and accrue a larger 'count only' sample. This approach can

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/2041-210X.13589](https://doi.org/10.1111/2041-210X.13589)

This article is protected by copyright. All rights reserved

26 improve the precision of the population density estimate when the cost of measuring the
27 distance to a detected target is more than half the cost of walking to the next target.

28 2. We find the optimal proportion of distances to measure that minimises the variance of the
29 density estimate for a fixed survey budget. We quantify how much this optimised strategy
30 increases the precision of the density estimate compared with conventional line-transect
31 distance sampling. We then use simulated distance sampling data to test our expressions,
32 and illustrate circumstances under which the optimised approach would be beneficial using
33 distance sampling data on high-density plants.

34 3. The simulations indicate that the optimised method delivers benefits in precision, but the
35 magnitude of the benefit is lower than predicted from our expressions, which are based on
36 an asymptotic approximation of the variance. We apply an adjustment to the predicted
37 benefit equation to account for this difference, and show that, in all three plant case studies,
38 the optimised approach could improve the precision gained from a distance sampling survey
39 between 20-50%.

40 4. This new approach could broaden the ecological contexts in which distance sampling is
41 applied, to include estimation of densities of abundant taxa where plots are conventionally
42 used. The method may have interesting applications for other survey types, including
43 multispecies surveys or those using cues or signs that occur at high-density.

44 **Keywords:** Distance sampling, high-density species, line-transect, optimisation, population density
45 estimation, variance approximation, wildlife monitoring

46

47 Introduction

48 Wildlife monitoring goals may be broad and varied, but resources are invariably limited. Cost-
49 efficient monitoring is increasingly important, owing to funding shortfalls for wildlife management
50 and conservation (McCarthy et al., 2012). Consequently, comparing performance and costs are a
51 staple of literature on monitoring methods (e.g. Alves et al., 2013; Clark, 2016; Ducey et al., 2013;
52 Fewster, 2011; Field et al., 2005; Fragoso et al., 2016; Keeping et al., 2018; Pollock et al., 2002).
53 Optimising the design of monitoring programs improves cost-efficiency; either by reducing costs
54 while obtaining the same amount of information, or alternatively, by obtaining more information for
55 the same cost. Here, we propose a method to optimise line-transect distance sampling of high-
56 density species, wherein the re-allocation of survey effort can increase the precision of a density
57 estimate for no increase in the overall effort to complete a survey.

58 Distance sampling (Buckland et al., 2001) is extensively applied in wildlife management and research
59 contexts to estimate abundance (Kral et al., 2018; Marshall et al., 2008; Plumptre, 2000). In distance
60 sampling, distances from the line or point of observation to each detected target are measured to
61 estimate detectability by modelling the decline in detections with increasing distance from the line
62 or point. Distance methods have broad appeal because they offer unbiased estimation of
63 population density and abundance, if assumptions are met and the sample size is sufficient. These
64 methods can be applied to a wide variety of species and contexts. For example, detection function
65 models – the models describing the decrease in detectability with increasing distance – can take a
66 range of shapes, and estimates can be reliable even when detectability varies across observers,
67 habitat type and individual target characteristics (Buckland et al., 2001).

68 However, the method tends not to be used for taxa that occur at high densities, such as many
69 plants. Instead, direct count methods – generally based on a strip, plot or plotless design that
70 assumes perfect detection – are preferred (Bonar et al., 2010; Carpenter et al., 1999; Hill, 2005;
71 Shiver & Borders, 1996). In the authors' experience, reluctance to use distance methods is
72 influenced by the perception that the method is impractical: the standard implementation requires
73 measuring distance to all observed targets, which means very frequent measuring when detections
74 are common. Accruing distance measurements to estimate the detection function is the focus of
75 survey effort when few detections are available. In contrast, these distance measurements would
76 be relatively easy to acquire with frequent detections, meaning that the focus of survey effort can be
77 on estimating the encounter rate and its variance – the more difficult variance component to
78 estimate (Fewster et al., 2009). Assigning distance measures to bins rather than measuring
79 distances precisely ('binning') has been previously suggested as one way to increase the efficiency of
80 distance sampling (Buckland et al., 2001). Here, we focus on scenarios with many detections and
81 offer an alternative survey design strategy that improves distance sampling efficiency for the line
82 transect protocol. We exploit the possibility of measuring the distance to only a subset of detections
83 to model the detection function, while still counting all detections to estimate population density.
84 We derive the optimal proportion of targets to measure by explicitly considering the trade-off
85 between time taken to measure and accrual of detected targets. Henceforth we refer to the line
86 transect distance sampling method where distances are measured accurately and to all detected
87 targets as LTS (conventional 'line transect sampling'), and our proposed approach, where distances
88 are measured accurately but only to a subset of targets, as the 'optimised' method.

89 Our sampling approach follows the general strategy proposed by Pollock et al. (2002). They suggest
90 combining an index of abundance, collected over a large area, with an estimate of detectability

91 drawn from an intensively sampled subset of sites. This strategy may be applied where different
92 survey methods are used to collect an index of abundance and to estimate detectability, or, as in our
93 approach, where the survey protocol, and hence the detection process, is the same for all individuals
94 regardless of whether additional information is collected (here distance measurements) to model
95 detectability. Pollock et al. (2002) further suggest that this survey design strategy should be applied
96 to any large-scale monitoring study where there is a significant cost discrepancy between the index
97 and the intensive method. They, however, note that the additional cost of measuring the distance
98 to detected targets (as opposed to just counting them) is minor, compared with much more costly
99 methods of estimating detectability, such as capture-recapture. While we agree this applies to many
100 distance sampling monitoring settings where detection events are rare, effort spent on measuring
101 can indeed be high in the scenarios on which our study focuses. We believe the application to
102 distance sampling of a design strategy along the line suggested by Pollock et al. (2002) warrants
103 further investigation: we identify the point at which the cost of measuring is sufficiently large,
104 relative to searching for a new target, such that measuring only a sub-sample of distances is justified.
105 This happens often when detections occur rapidly. The increase in the number of detections, given
106 the ability to survey a greater total length of transect, improves the precision of the density estimate
107 (the component of the variance related to the encounter rate is reduced).

108 While there are many features of a survey protocol that practitioners may evaluate to determine
109 what best meets their needs – such as ease of application, availability of expertise and experience,
110 or reproducibility – we focus on the crucial aspects of bias and precision in the context of cost. We
111 organize our paper as follows: First, we determine the optimal proportion of detected targets for
112 which distances should be measured based on the statistical theory behind distance sampling
113 (Buckland et al., 2001). We then simulate data collection to evaluate how well our theoretical
114 optimisation performs in practice. Finally, we identify the circumstances under which our optimised
115 method outperforms the LTS sampling design and illustrate our approach with case studies of
116 distance sampling of Australian plants, *Brassica sp.* and *Grevillea australis*.

117 **Materials and Methods**

118 In the following sections, we develop the theory behind our proposed survey optimisation, and
119 describe the methods used to assess its performance. The statistical theory of line transect distance
120 sampling, as described in Buckland et al. (2001) underpins the mathematical derivation. We focus
121 on the half-normal detection function, which is one of the three key functions recommended by
122 Buckland et al. (2001), and one that tends to yield simpler calculations and closed-form expressions,
123 as shown below.

124 **Optimal proportion of distances to measure: theoretical derivation**

125 To identify the optimal proportion of distances to measure, we seek to minimise the variance of the
126 density estimator \hat{D} . In line transect distance sampling, density is calculated as:

127
$$\hat{D} = \frac{n}{2wL\hat{p}}$$

128 where:

129 n is the total number of detection events (for our purposes this is both events where distance is
130 measured, n_1 , and those that are unmeasured or counted only, n_0 , such that $n_1 + n_0 = n$); w is the
131 half strip width, which is the farthest perpendicular distance from the transect line to which
132 detections are recorded (or the farthest distance used in analysis if there is truncation), therefore
133 $2w$ is the complete width of the area sampled; L is the total length of transect covered, and \hat{p} is the
134 probability of detection of objects in the sampled area. The calculation of \hat{D} is sometimes expressed
135 in terms of the 'effective strip width', $\hat{\mu}$, where $\hat{\mu} = w\hat{p}$ is the area under the distance detection
136 function. The 'effective strip width' is also equal to $1/\hat{f}(0)$, where \hat{f} has the same shape as the
137 distance detection function but rescaled to be a probability density function (i.e. it integrates to 1,
138 instead of $\hat{\mu}$). With this, we can express \hat{D} as:

139
$$\hat{D} = \frac{n}{2L\hat{\mu}} = \frac{n}{2L}\hat{f}(0).$$

140 The variance of the estimator \hat{D} can be approximated asymptotically using the delta method (Seber,
141 1982; Buckland et al., 2001, p. 52), with components of variance due to the observed sample size
142 and the detection function (Buckland et al., 2001, pp. 76-77), as follows:

143
$$\text{var}(\hat{D}) = \hat{D}^2 \left[\frac{\text{var}(n)}{n^2} + \frac{\text{var}(\hat{f}(0))}{(\hat{f}(0))^2} \right] \text{eqn 1.}$$

144 The literature often refers to the variance due to encounter rate (n/L , which we refer to as E) rather
145 than observed sample size (Buckland et al., 2001, p. 79). Here, we consider the idealised scenario
146 where there is no spatial heterogeneity in density, for the purposes of a proof of concept. The
147 following assumptions apply: we assume that the targets are located randomly in space, and that
148 sampling lines are placed independent of the distribution of the targets, such that the distances
149 between targets and the transect are uniformly distributed, and the number of detections follows a
150 Poisson distribution. We recognise that many taxa exhibit some degree of aggregation or clustering
151 that may violate these assumptions, and we address this in Appendix 1. If targets are Poisson

152 distributed, encounter rate variance – the much greater contributor to overall variance (Fewster et
 153 al., 2009; Glass et al., 2015), and that which we are seeking to minimise – will be relatively low, so
 154 real monitoring scenarios may see greater benefits than we predict here. Appendix 1 shows that,
 155 when the assumptions are violated, our proposed optimisation method still achieves efficiencies
 156 that are mostly equal to or greater than those predicted in the following analysis. A more thorough
 157 analysis of this is planned for future research and will not be addressed further here.

158 For a Poisson distribution $var(n) = E(n)$. Henceforth we use n as a plug-in estimate of the expected
 159 value, so the first term in equation 1 simplifies to $1/n$. If the detection function is half-normal with
 160 no truncation, the variance of $\hat{f}(0)$ is:

$$161 \quad var(\hat{f}(0)) = \frac{(\hat{f}(0))^2}{2n},$$

162 so the second term in equation 1 simplifies to $1/2n$ (Buckland et al., 2001, pp. 65-66). Substituting
 163 these two expressions back into equation 1, allowing for the possibility that distances are only
 164 measured to a proportion α of the n detected targets, gives:

$$165 \quad var(\hat{D}) = \hat{D}^2 \left[\frac{n}{n^2} + \frac{(f(0))^2}{2\alpha n (f(0))^2} \right]$$

$$166 \quad = \hat{D}^2 \left[\frac{1}{n} \left(1 + \frac{1}{2\alpha} \right) \right] \text{ eqn 2.}$$

167 For a given survey budget B , decreasing the proportion α implies that a longer total length of
 168 transect can be surveyed, hence accumulating a greater number of detected targets n . To identify
 169 the optimal α , we need to re-express the variance of \hat{D} considering the effect of α on n , as distinct
 170 from n_{LTS} , which is the total number of detected targets where all targets are measured and
 171 therefore $\alpha = 1$. We start by equating the overall budget B to the sum of the expected total costs of
 172 walking and measuring distances:

$$173 \quad B = LE(C_w + \alpha C_m) \text{ eqn 3,}$$

174 where ‘cost’ is a proxy for time or person-hours expended in conducting a survey, and ‘walking’ may
 175 mean any method used to move along the line transect; C_w is the cost of walking and counting
 176 without measuring, per detection; C_m is the cost of measuring the distance to a detected target; L is
 177 the total transect length in the survey; E is the mean number of detections per unit length (i.e. the
 178 encounter rate; $E = 2\mu D$); and B , C_m and C_w are all expressed in the same generic ‘budget units’.

179 We can rearrange equation 3 to obtain the following expression for the expected transect length L :

$$180 \quad L = \frac{B}{E(C_w + \alpha C_m)} \text{ eqn 4.}$$

181 We can now substitute this expression into equation 2, as the term $1/n = 1/EL$, giving:

$$182 \quad \text{var}(\hat{D}) = \hat{D}^2 \left[\left(\frac{C_w + \alpha C_m}{B} \right) \left(1 + \frac{1}{2\alpha} \right) \right] \text{ eqn 5.}$$

183 Finally, differentiating equation 5 with respect to α , equating to 0 and solving, we obtain the optimal
184 value of α , which is a function of the ratio of costs per detection:

$$185 \quad \alpha^* = \sqrt{\frac{C_w}{2C_m}} \text{ eqn 6.}$$

186 This gives the value of α where the function is minimised (various shapes for precision as a function
187 of α are shown in Appendix 2). Note that α^* depends only on the ratio $C_w:C_m$, and is independent of
188 the budget, B . We can see with equation 6 that $\alpha^* < 1$ requires that $C_m > \frac{1}{2}C_w$. Thus, for
189 abundant species where new targets are encountered rapidly relative to the time taken to measure
190 distances to them (i.e. C_w is low compared to C_m), measuring distances to only a fraction of the
191 detected targets is optimal. We assume unbiased selection of the subset that is measured, with no
192 correlation between measurement and detection distance. We discuss later practical strategies to
193 ensure this.

194 *Predicted benefit of the optimisation*

195 We refer to the proportional increase in precision due to the optimisation as its 'benefit', b :

$$196 \quad b = \frac{\text{var}(\hat{D}_{\text{LTS}})}{\text{var}(\hat{D}_{\text{opt}})} \text{ eqn 7.}$$

197 The optimisation is beneficial when this ratio is > 1 , meaning that the variance of the estimator of D
198 under LTS is higher than when measuring distance to a subset of detected targets. Introducing our
199 derived expression for the optimal α^* (equation 6) into equation 2, the variance of the density
200 estimator for the optimised scenario, i.e. the denominator in equation 7, is:

$$201 \quad \text{var}(\hat{D}_{\text{opt}}) = D^2 \left[\frac{1}{n} \left(1 + \sqrt{\frac{C_m}{2C_w}} \right) \right] \text{ eqn 8.}$$

202 Similarly, we can express the variance of \hat{D}_{LTS} , i.e. the numerator in equation 7, by using equation 2
 203 now setting $\alpha = 1$:

$$204 \quad \text{var}(\hat{D}_{LTS}) = D^2 \frac{3}{2n_{LTS}} \text{ eqn 9.}$$

205 Here n_{LTS} is the number of detections using the LTS method when the budget of survey effort (B) is
 206 the same as that required to detect n targets when the optimisation is applied. Given that detecting
 207 and measuring each target costs C_w and C_m respectively, we have that:

$$208 \quad n_{LTS} = \frac{B}{C_w + C_m}.$$

209 In the optimised method, the total time taken to survey is equal to the total time walking (nC_w) plus
 210 the total time measuring ($n\alpha C_m$) which, when substituting in the optimal value of α , is:

$$211 \quad B = n \left(C_w + \sqrt{\frac{C_w C_m}{2}} \right).$$

212 We can express n_{LTS} in terms of n as follows:

$$213 \quad n_{LTS} = n \frac{C_w + \sqrt{\frac{C_w C_m}{2}}}{C_w + C_m} \text{ eqn 10.}$$

214 Combining equations 8, 9 and 10, the predicted benefit of the optimisation, b , is:

$$215 \quad \frac{\text{var}(\hat{D}_{LTS})}{\text{var}(\hat{D}_{opt})} = \frac{3}{2} \frac{1 + R_c}{\left(1 + \sqrt{\frac{R_c}{2}}\right)^2} \text{ eqn 11,}$$

216 where $R_c = C_m/C_w$, i.e. R_c is the ratio of costs (cost of measuring per detection to cost of walking per
 217 detection). Thus, the benefit of optimisation is an increasing function of R_c , for $R_c > \frac{1}{2}$. As

218 mentioned above, below this threshold the optimal proportion of targets to measure is 1, i.e. LTS is
 219 the optimal strategy in those circumstances.

220 *Simulation to assess optimisation performance*

221 We used simulated distance data to test how well the derived optimisation performs, considering
 222 that the delta method only approximates the variance. We ran simulations for 61 sets of parameters
 223 that represent different 'scenarios' where the optimal $\alpha < 1$ (Table 1 and Appendix 3). Cost ratios

224 between ~ 0.5 and ~ 25 were generated from all unique combinations of six unique population
225 densities, four unique values for the scale parameter of the half-normal detection function (σ), and
226 four different costs of measuring. We expressed costs in terms of generic 'budget units', where one
227 unit is the cost of walking one unit-length of transect. For each value of σ , we set the half-strip width
228 w to 2σ , which resulted in probability of detection at distance w of 0.135. For each scenario, we used
229 equation 6 to calculate α^* and equation 4 to calculate the corresponding optimal length of transect
230 that would be covered, assuming a fixed budget of 10,000 units. We chose to use a substantial
231 budget to avoid very small sample sizes in the lower density, lower detectability scenarios. We ran
232 20,000 simulations for each scenario; 10,000 using the optimised method, and 10,000 using LTS.

233 Simulations were run in R version 3.5.0 (R Core Team, 2019). The simulated distance sampling
234 scenarios were equivalent to data collection along a single long transect. We generated distance
235 sampling data by drawing the number of individuals in the sampling area from a Poisson distribution
236 $N \sim \text{Pois}(E(n))$ (where $E(n) = 2wLD$, the expected number of individuals in the sampling area) and
237 their corresponding perpendicular distances from the transect were drawn from a uniform
238 distribution $Y \sim \text{Unif}(0, w)$. We determined whether each individual was detected in the survey as
239 the outcome of a Bernoulli trial, with the probability of success being the probability of detection at
240 its distance according to the half-normal detection function; $X \sim \text{Bernoulli}(\exp(-\frac{y^2}{2\sigma^2}))$. The set of
241 distances for individuals that were detected under each scenario constitutes our full distance
242 sampling dataset. To test the optimised method, this dataset was then subset by the relevant value
243 of α^* . The n_1 subset was selected using even intervals across the length of the dataset, e.g. if α^*
244 $= 0.2$, the subset consists of every 5th detection distance. Model fitting was conducted using the
245 function `ds` in the R package `Distance` (Miller et al., 2019), specifying the half-normal detection
246 function, as in data generation, therefore an assessment of model fit was not necessary. Density
247 estimates for the optimised method were scaled up to account for the total number of detected
248 individuals in the sample: $\hat{D}_{\text{opt}} = d_1(n/n_1)$, where d_1 is the density estimated from the subsample.
249 This allowed us to use the `Distance` package function to analyse data under the optimised method:
250 only the subset of measured distances was used to fit the detection function, and the density
251 estimate obtained was then corrected to account for all detections. We calculate variance
252 empirically from 10,000 density estimates ($\text{var}(\hat{D}_{\text{opt}})$ and $\text{var}(\hat{D}_{\text{LTS}})$).

253 *Case studies*

254 We use real distance sampling datasets to demonstrate the utility and benefit of the optimised
255 method in real-world surveys of high-density plant species. These datasets were collected as pilot-

256 stage data and so are not extensive in the length of transects, but sample sizes are adequate to
 257 explore the application of the optimisation. A *Brassica sp.* was surveyed on 100 m of transect in
 258 Royal Park, Melbourne, Australia, where the survey species was moderately abundant and
 259 detectable, and the habitat is open. *Grevillea australis* adults and seedlings were surveyed in
 260 separate sites (unburnt and burnt, respectively) in an alpine heathland habitat, on the Bogong High
 261 Plains in the Australian Alps, Victoria, Australia. Both adults and seedlings were surveyed on 20 m of
 262 transect. Seedlings were highly abundant but had low detectability, whereas the adults were much
 263 larger and more detectable but less abundant. Transects were delineated by a forestry measuring
 264 tape, and perpendicular distances were measured also using a forestry measuring tape. Two
 265 observers conducted the *Brassica* survey, and one observer conducted both *Grevillea* surveys.
 266 Average time to walk to the next detected target and average time to measure a target were
 267 estimated, using field notes, by the principal researcher running the surveys (details in Appendix 4).

268 Results

269 *Performance of the optimisation*

270 Our simulated distance data confirms that the proposed optimisation can improve the precision of
 271 the density estimate, and that this benefit scales according to the cost ratio R_c (Fig. 1). The observed
 272 benefit is typically lower than the predicted benefit based on the theoretical derivation, ranging
 273 from a mean of 0.95 (i.e. slight loss of precision when using the optimisation; scenario number 16 in
 274 Appendix 3) to a mean of 1.62 (benefit for scenario 96, Appendix 3). In contrast, according to the
 275 predictions, the benefit should be >1 for all scenarios tested, and be as high as 1.90. The difference
 276 between the observed and predicted benefits is likely due to our predicted variances differing from
 277 the true variances, particularly as the variances increase (Appendix 5). Adjusting the predicted
 278 benefit brings the prediction more in line with the results observed in simulations (Fig. 1); a good
 279 adjustment is provided by Equation 12 while being a simple expression.

$$280 \quad \text{adjusted } b = 0.8(\text{predicted } b) + \frac{0.19}{(\text{predicted } b)} \text{ eqn12.}$$

281 For context, the observed benefit of the optimisation is around 1.2 for $R_c = 5$. Achieving a 20%
 282 improvement in precision using LTS would require 20% more sampling, as the variance of \hat{D} is
 283 proportional to $\frac{1}{n_{LTS}}$ (i.e. $b_{LTS} \propto \frac{n_{LTS, \text{Increased}}}{n_{LTS}}$, where b_{LTS} is the same benefit as the optimisation,
 284 achieved using LTS). This means that survey effort, and therefore costs, would need to be increased
 285 by 20% to meet the same performance as our optimisation.

286 There is a small bias in estimation of density with both the LTS and optimised methods (mean error
287 of all simulations is 0.15% for LTS and 0.18% for the optimised method), but generally the estimates
288 cluster around the real values (Fig. 2). That the slight bias is consistent between the methods
289 indicates that the optimisation has not introduced bias to the estimation of density.

290 *Case studies for optimising distance sampling surveys of high-density plants*

291 The optimisation seems beneficial for all our case study targets (adjusted benefits ranging from
292 ~20% for the *Brassica sp.* survey to ~50% for *G. australis* seedlings, Table 2). As the adjusted benefit
293 is more likely to be realistic according to our simulations, we refer to this as the primary result. As
294 expected, the difference in R_c between targets accounts for the differences in the magnitude of the
295 benefit. The combination of density and detectability determines the encounter rate for the survey
296 target, which is lowest for the *Brassica sp.*, rising for *G. australis* adults and highest for *G. australis*
297 seedlings (Table 2). When the method of measurement remains consistent between target species
298 (in this case manual measurement with a forestry tape over short distances), the more rapid the
299 encounter rate, the lower the cost of walking, leading to a higher R_c .

300 We used the estimated total survey time for each of the pilot surveys to calculate the transect length
301 and number of detected targets that would be expected in a survey of 8 hours (Table 3; we used 8
302 hours of survey time to ensure large sample sizes for all targets). For each target species, the
303 expected number of detected targets measured under the optimised method (n_1) is smaller than the
304 total number of detections in the standard method (n_{LTS}), while the total number of detected
305 targets under the optimised method (n) is much larger than n_{LTS} , given the longer transect lengths
306 covered under the optimisation.

307 **Discussion**

308 We present a modified distance sampling methodology for line transect sampling, allocating the
309 overall survey effort such that population parameter estimates are more precise than when using
310 standard methods, with negligible difference in bias. The gains we observed are unlikely to be
311 realised with the point sampling protocol, as the greater contributor to overall variance for points is
312 from estimating the detection function, so reducing the number of distance measurements available
313 for estimation may have a greater impact on precision.

314 Our proposed method is one of several options for practitioners to improve the utility of the
315 conventional line transect method. 'Binning' is an alternative strategy that simplifies the distance
316 sampling survey process when detections occur at a high frequency (Buckland et al., 2001) and

317 makes the method useful in situations when precise measurements of distance may be unfeasible.
318 Our approach avoids issues associated with inaccurate distance measurements in bins, such as lower
319 precision compared with LTS, and problems identifying lack of model fit, when using few bins (T.
320 Marques, pers. comm., 2018). More recent developments include approaches that integrate
321 multiple monitoring methods and data sources within a modelling framework, enabling spatially
322 explicit analysis of population processes. Although the analogy may not be explicit, these
323 approaches follow the general form suggested by Pollock et al. (2002), in that they estimate key
324 parameters using intensive methods (density, detectability), and extrapolate across broader spatial
325 scales using data that are cheaper to obtain (e.g. presence-only data; Farr et al., 2021, and
326 environmental covariates; Farr et al., 2021 and Miller et al., 2013). Rather than being solely an
327 alternative option, it may be possible to incorporate optimised distance sampling within such
328 frameworks if the encounter rate for the target species is sufficiently high. Similarly, other
329 approaches that potentially utilise distance sampling, such as (Fewster, 2011) could also integrate
330 optimised distance sampling. There is, however, the caveat that the optimised method is only
331 relevant to improve precision of estimated encounter rate, and not where distance sampling only
332 provides an estimate of detectability.

333 It is notable that the observed benefit from optimisation is largely attributable to the value of R_c
334 (Fig. 1). Although the overall effort of a survey is usually reported in the literature (the length of
335 transect, person-hours etc.), the relative costs that we have identified as critical inputs to the
336 optimisation calculations are not usually recorded. This limits the capacity to identify existing real-
337 world examples where this method would be beneficial. However, as illustrated by our case studies,
338 we can expect this to happen when targets are encountered rapidly such that the time to detect a
339 new target is small compared to the time taken to accurately measure distances (and hence the cost
340 ratio of measuring to walking per detection is high). Surveys in rugged and complex terrain, or where
341 w is very wide could also qualify as high-cost measuring, even if the encounter rate is lower, as long
342 as the time taken to measure is much greater than the time taken to reach the next target. In
343 contrast, technology such as accurate laser rangefinders can significantly reduce the time to
344 measure distances, which would mean the R_c would not be high and the optimised method would
345 therefore likely not be beneficial. Emerging technologies such as drone-mounted cameras and
346 other types of remote sensing might be out-pacing methods like distance sampling in wildlife
347 monitoring. Although remote sensing is useful for surveying animal populations in open habitat, and
348 many characteristics of vegetation – coverage and structure, for example, can be ascertained from
349 LiDAR – some taxa and characteristics are more suited to non-remote methods. This may be the
350 case when identifying species is key, which can be done more easily in-person if species are cryptic,

351 look similar to other species, or are small or difficult to detect, etc. In addition, distance sampling
352 could remain a viable and efficient option for monitoring where visibility and manoeuvrability
353 present significant barriers for remote sensing. For example, remote sensing and even laser
354 rangefinders are of limited use in heavily vegetated areas, such as rainforests.

355 Although distance methods are used for a broad array of taxa, certain taxa, such as plants, tend to
356 be sampled using total counts from plots of various designs. Methods for efficient distance sampling
357 of plants have been developed (Buckland et al., 2007), but the effort involved may be prohibitive if
358 the sample area is large. The main drawback of plot-based direct count methods for abundance and
359 density estimation is that they assume perfect detection, which may not always be true e.g. for
360 surveys of a rainforest tree species (Kissa & Sheil, 2012). In Kissa and Sheil (2012), the species in
361 question is low density, and so the undercount in a strip transect had a notable impact on the
362 density estimate compared with LTS. Systematic undercounting (as a result of not accounting for
363 detectability) or overcounting (as a result of double counting individuals) in plots may occur in either
364 high-density or low-density populations. Error rates in estimated density in our simulations, for both
365 LTS and the optimised method, are very low; it follows that error rates in plots would also have to be
366 very low to be of comparable accuracy. Further study is needed to ascertain whether optimised line-
367 transect distance sampling offers improved cost-efficiency for a target level of precision over
368 traditional plot-based methods.

369 It is worth noting that our case studies are based in a setting with higher detection than a rainforest
370 or other area with dense vegetation. Plot-based methods may perform well for plants in grassland
371 or alpine heathland, or similar contexts, compared with distance methods, but determination of
372 whether distance-based or alternative methods are preferable for density estimation will likely differ
373 by context, and is outside the scope of this paper. If line-transect distance sampling is applied, it
374 may be worthwhile to assess whether the survey can be optimised using our approach.

375 If distances to a subset of detected targets are to be measured, as in the optimisation approach,
376 then the selection of that subset might introduce bias to the density estimate if the measurements
377 used to fit the detection function are biased, e.g. toward shorter distances or a particular part of the
378 transect. We have assumed unbiased selection in our analysis, but in practice, surveyors are subject
379 to perception bias, and there are further sources of bias that can affect which of the available
380 targets are detected. Using every n th detection in the field is a starting point for minimising the
381 effect of observer bias, but this may still be subject to unintentional filtering, and a selection that is
382 not representative of the true detection function may result. Field testing strategies to minimise

383 unintentional filtering are a potential future step in the development of this optimised method of
384 distance sampling.

385 The optimised method has implications for survey design. A pilot study or some knowledge of the
386 target population is essential to ascertain the expected encounter rate and relative costs. Pilot
387 studies routinely provide most of this information, and are generally good practice to evaluate the
388 efficacy of the study design for its purpose and integrate knowledge about unexpected aspects of
389 fieldwork (Burns et al., 2014; Spellerberg, 2005). Adding the collection of cost data to the
390 preliminaries should not be a demanding addition to the pilot stage of a monitoring program. Long-
391 term monitoring programs may already have data that can indicate the expected encounter rates,
392 and experienced surveyors could estimate the relative cost of measuring. Established monitoring
393 programmes with access to this information could use the expressions presented in this paper to
394 determine if their survey can be optimised and the expected benefit. Even where costs of measuring
395 are not as high as those presented in our case studies, the benefit in precision that can be gained at
396 no extra cost could still be equivalent to substantial savings on large-scale surveys.

397 There are further potential advantages of the optimised method in the context of multispecies
398 surveys of density and abundance. Among several target species there is likely to be a range of
399 detectabilities, population densities and expected encounter rates from common through to rare. If
400 LTS is applied, the maximum feasible length of transect for a given budget will be limited primarily by
401 the abundances of the common species, with time absorbed by measuring distances to those
402 frequent detections. The limit on transect length may prevent the collection of a sufficient sample
403 size to gain a density estimate for the rarer species (60-80 detections is a recommended minimum
404 (Buckland et al., 2001)), which may be of greater interest or importance due to their rarity. If the
405 optimised approach is incorporated into a multispecies survey, surveyors can set the total length of
406 transect such that they can expect ≥ 60 detection events of the rare species. The equations we
407 provide for the optimised method offer a way to re-allocate survey effort to reach this sufficient
408 number of detections of the rarer species; measuring distances to only a proportion of detections of
409 the more common species could save the time needed to cover the longer total length of transect.
410 Even if the proportion measured of the common species is not exactly at the optimal point (i.e. it is
411 lower than optimal), the precision could still be better than using LTS with the same budget. The
412 precision of the density estimate tends to decrease rapidly when the proportion measured is lower
413 than optimal (urging caution with measuring a lower than optimal proportion), but the decrease in
414 precision is not as dramatic and reaches a maximum at $\alpha = 1$ (i.e. LTS) when the proportion
415 measured is above optimal (Appendix 2). Therefore, there is scope to adjust the proportion

416 measured for the more common species either above or below the optimal, and consider the effect
417 on the precision of the estimate and whether it is an acceptable compromise according to the goals
418 of the monitoring program or survey.

419 We have addressed in our simulations and case studies the simple case of observations of single
420 individuals to test the optimised method. Distance sampling is common for species that occur in
421 clusters, or that are observed indirectly through cues and signs, such as vocalisations or dung piles.
422 Glass et al. (2015) demonstrate that cluster size variance is a relatively small contributor to overall
423 uncertainty around the density estimate (much smaller than the detection function). As such we can
424 postulate that the additional source of variance from estimating cluster size should not swamp the
425 benefit from reducing uncertainty using the optimised approach. Similarly, in cases where the
426 contribution to the variance from conversion of density of signs to density of individuals is
427 sufficiently low, the optimisation could deliver benefits in surveys of species that are not themselves
428 high density, but produce signs that would yield a sufficiently high encounter rate. Examples of
429 species for which this may be possible is largely speculative at this stage, as R_c is a key input to the
430 optimisation calculations and is generally not reported. However, mammals such as forest elephants
431 (*Loxodonta cyclotis*, Walsh and White, 1999) and deer (e.g. sika deer, *Cervus nippon*, Marques et al.,
432 2001), are commonly surveyed by distance sampling using dung counts, and could be candidates to
433 test if the measuring costs and encounter rates would reach the level necessary for the optimised
434 method to deliver benefits in precision.

435 Our optimisation method offers a way to improve the efficiency of line transect distance sampling
436 for high density species, increasing precision of the density estimate at no extra cost. The method
437 should be most beneficial for species for which the encounter rate and R_c are both high (such as the
438 *G. australis* seedlings in our case study), but there are benefits across a range of densities,
439 detectabilities and values of R_c . The use of our approach to optimise distance sampling for high-
440 density species could further widen the ecological monitoring contexts and range of taxa for which
441 distance sampling can be applied as a monitoring tool.

442 **Data availability:** Data used in the case studies and code for all simulations and analyses presented
443 here are publicly available via Zenodo (Knights & Camac, 2021).

444 **Acknowledgements:** This work was supported by a Melbourne Research Scholarship from the
445 University of Melbourne, and the ARC Centre of Excellence for Environmental Decisions (CEED). The
446 authors would like to thank Joanne Potts and Tiago Marques for invaluable comments.

447 **Authors contributions:** GGA conceived the original idea and conducted the mathematical
448 derivations for the optimisation; KK designed the simulation and analysed the data with advice from
449 GGA and MM; KK led the writing of the manuscript; JC designed and collected the pilot study data on
450 Australian plants. All authors contributed critically to the drafts and gave final approval for
451 publication.

452 References

- 453 Alves, J., da Silva, A. A., Soares, A. M. V. M., & Fonseca, C. (2013). Pellet group count methods to
454 estimate red deer densities: Precision, potential accuracy and efficiency. *Mammalian*
455 *Biology*, 78(2), 134–141. <https://doi.org/10.1016/j.mambio.2012.08.003>
- 456 Bonar, S. A., Fehmi, J. S., & Mercado-Silva, N. (2010). An overview of sampling issues in species
457 diversity and abundance surveys. In A. E. Magurran & B. J. McGill (Eds.), *Biological Diversity:*
458 *Frontiers in Measurement and Assessment*. Oxford University Press.
- 459 Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., & Thomas, L. (2001).
460 *Introduction to Distance Sampling: Estimating Abundance of Biological Populations*. Oxford
461 University Press.
- 462 Buckland, S. T., Borchers, D. L., Johnston, A., Henrys, P. A., & Marques, T. A. (2007). Line transect
463 methods for plant surveys. *Biometrics*, 63(4), 989–998. [https://doi.org/10.1111/j.1541-](https://doi.org/10.1111/j.1541-0420.2007.00798.x)
464 [0420.2007.00798.x](https://doi.org/10.1111/j.1541-0420.2007.00798.x)
- 465 Burns, E., Lindenmayer, D., Tennant, P., Dickman, C., Green, P., Hanigan, I., Hoffmann, A., Keith, D.,
466 Metcalfe, D., Nolan, K., Russell-Smith, J., Wardle, G., Welsh, A., Williams, R., & Yates, C.
467 (2014). *Making ecological monitoring successful: Insights and lessons from the Long Term*
468 *Ecological Research Network*. LTERN. www.tern.org.au/ltern
- 469 Carpenter, A. T., Elzinga, C. L., Salzer, D. W., & Willoughby, J. W. (1999). Measuring and Monitoring
470 Plant Populations. *Journal of Range Management*, 52(5), 544.
471 <https://doi.org/10.2307/4003786>
- 472 Clark, R. G. (2016). Statistical Efficiency in Distance Sampling. *Plos One*, 11(3), e0149298.
473 <https://doi.org/10.1371/journal.pone.0149298>
- 474 Ducey, M. J., Williams, M. S., Gove, J. H., Roberge, S., & Kenning, R. S. (2013). Distance-limited
475 perpendicular distance sampling for coarse woody debris: Theory and field results. *Forestry*,
476 86(1), 119–128. <https://doi.org/10.1093/forestry/cps059>
- 477 Farr, M. T., Green, D. S., Holekamp, K. E., & Zipkin, E. F. (2021). Integrating distance sampling and
478 presence-only data to estimate species abundance. *Ecology*, 102(1), e03204.
479 <https://doi.org/10.1002/ecy.3204>

- 480 Fewster, R. M. (2011). Variance Estimation for Systematic Designs in Spatial Surveys. *Biometrics*,
481 67(4), 1518–1531. <https://doi.org/10.1111/j.1541-0420.2011.01604.x>
- 482 Fewster, Rachel M., Buckland, S. T., Burnham, K. P., Borchers, D. L., Jupp, P. E., Laake, J. L., &
483 Thomas, L. (2009). Estimating the Encounter Rate Variance in Distance Sampling. *Biometrics*,
484 65(1), 225–236. <https://doi.org/10.1111/j.1541-0420.2008.01018.x>
- 485 Field, S. A., Tyre, A. J., Possingham, H. P., & Lubow. (2005). Optimizing allocation of monitoring effort
486 under economic and observational constraints. *Journal of Wildlife Management*, 69(2), 473–
487 482. [https://doi.org/10.2193/0022-541X\(2005\)069\[0473:OAOMEU\]2.0.CO;2](https://doi.org/10.2193/0022-541X(2005)069[0473:OAOMEU]2.0.CO;2)
- 488 Fragoso, J. M. V., Levi, T., Oliveira, L. F. B., Luzar, J. B., Overman, H., Read, J. M., & Silvius, K. M.
489 (2016). Line Transect Surveys Underdetect Terrestrial Mammals: Implications for the
490 Sustainability of Subsistence Hunting. *PLoS ONE*, 11(4).
491 <https://doi.org/10.1371/journal.pone.0152659>
- 492 Glass, R., Forsyth, D. M., Coulson, G., & Festa-Bianchet, M. (2015). Precision, accuracy and bias of
493 walked line-transect distance sampling to estimate eastern grey kangaroo population size.
494 *Wildlife Research*, 42(8), 633–641. <https://doi.org/10.1071/WR15029>
- 495 Hill, D. A. (Ed.). (2005). *Handbook of biodiversity methods: Survey, evaluation and monitoring*.
496 Cambridge University Press.
- 497 Keeping, D., Burger, J. H., Keitsile, A. O., Gielen, M.-C., Mudongo, E., Wallgren, M., Skarpe, C., &
498 Foote, A. L. (2018). Can trackers count free-ranging wildlife as effectively and efficiently as
499 conventional aerial survey and distance sampling? Implications for citizen science in the
500 Kalahari, Botswana. *Biological Conservation*, 223, 156–169.
501 <https://doi.org/10.1016/j.biocon.2018.04.027>
- 502 Kissa, D. O., & Sheil, D. (2012). Visual detection based distance sampling offers efficient density
503 estimation for distinctive low abundance tropical forest tree species in complex terrain.
504 *Forest Ecology and Management*, 263, 114–121.
505 <https://doi.org/10.1016/j.foreco.2011.09.020>
- 506 Knights, K. A., & Camac, J. (2021). *Distance sampling optimisation data and code (v1.0)*. *Zenodo*.
507 <https://doi.org/10.5281/zenodo.4563353>
- 508 Kral, K., Harmon, J., Limb, R., & Hovick, T. (2018). Improving our science: The evolution of butterfly
509 sampling and surveying methods over time. *Journal of Insect Conservation*, 22(1), 1–14.
510 <https://doi.org/10.1007/s10841-018-0046-z>
- 511 Marques, F. F. C., Buckland, S. T., Goffin, D., Dixon, C. E., Borchers, D. L., Mayle, B. A., & Peace, A. J.
512 (2001). Estimating deer abundance from line transect surveys of dung: Sika deer in southern

- 513 Scotland. *Journal of Applied Ecology*, 38(2), 349–363. <https://doi.org/10.1046/j.1365->
514 2664.2001.00584.x
- 515 Marshall, A. R., Lovett, J. C., & White, P. C. L. (2008). Selection of line-transect methods for
516 estimating the density of group-living animals: Lessons from the primates. *American Journal*
517 *of Primatology*, 70(5), 452–462. <https://doi.org/10.1002/ajp.20516>
- 518 McCarthy, D. P., Donald, P. F., Scharlemann, J. P. W., Buchanan, G. M., Balmford, A., Green, J. M. H.,
519 Bennun, L. A., Burgess, N. D., Fishpool, L. D. C., Garnett, S. T., Leonard, D. L., Maloney, R. F.,
520 Morling, P., Schaefer, H. M., Symes, A., Wiedenfeld, D. A., & Butchart, S. H. M. (2012).
521 Financial Costs of Meeting Global Biodiversity Conservation Targets: Current Spending and
522 Unmet Needs. *SCIENCE*, 338(6109), 946–949. <https://doi.org/10.1126/science.1229803>
- 523 Miller, D. L., Rexstad, E., Thomas, L., Marshall, L., & Laake, J. L. (2019). Distance Sampling in R.
524 *Journal of Statistical Software*, 89(1). <https://doi.org/10.18637/jss.v089.i01>
- 525 Plumptre, A. J. (2000). Monitoring mammal populations with line transect techniques in African
526 forests. (Vol 37, pg 356, 2000). *Journal of Applied Ecology*, 37(4), 698–698.
- 527 Pollock, K. H., Nichols, J. D., Simons, T. R., Farnsworth, G. L., Bailey, L. L., & Sauer, J. R. (2002). Large
528 scale wildlife monitoring studies: Statistical methods for design and analysis. *Environmetrics*,
529 13(2), 105–119. <https://doi.org/10.1002/env.514>
- 530 R Core Team, 2019. R: A language and environment for statistical computing. R Foundation for
531 Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>
- 532 Seber, G. A. F. (1982). *The estimation of animal abundance: And related parameters* (2nd ed). Griffin.
- 533 Shiver, B. D., & Borders, B. E. (1996). *Sampling techniques for forest resource inventory*. Wiley.
- 534 Spellerberg, I. F. (2005). *Monitoring Ecological Change*. Cambridge University Press.
- 535 Walsh, P. D., & White, L. J. T. (1999). What it will take to monitor forest elephant populations.
536 *Conservation Biology*, 13(5), 1194–1202. <https://doi.org/10.1046/j.1523-1739.1999.98148.x>
- 537
- 538
- 539 Tables

540 Table 1 Parameter values used in the simulations. The α^* was calculated for each unique
 541 combination of density (D), σ and cost of measuring (C_m), and scenarios where it was less than one
 542 were retained for the study (61 out of the total 96). Here, the cost of measuring is shown in generic
 543 'budget units'. These values for cost, in combination with D and σ values, yield values for R_c
 544 between ~ 0.5 and ~ 25 . See Appendix 3 for the full list of parameters and calculated variables.

Parameter	Values
D (density (indiv/m ²))	0.005, 0.01, 0.02, 0.05, 0.075, 0.1
σ (detection function model scale parameter (m))	1, 3, 5, 10
Cost of measuring one detected target in generic 'budget units'.	1.1, 2, 5, 10

545

546

547 Table 2 Calculations for case study species; encounter rates (E) indicate number of encounters per
 548 meter; ratio of costs R_c is based on estimated time to complete the survey and estimated time to
 549 measure distance per detection (Appendix 4); α^* is the optimal proportion of detected targets to
 550 measure based on Equation 6, the predicted benefit b and adjusted predicted benefit are calculated
 551 using Equations 11 and 12.

Survey target	E	R_c	α^*	b	Adjusted b
<i>Brassica sp.</i>	1.3	4.5	0.33	1.32	1.19
<i>Grevillea australis</i> (seedling)	11.7	17.9	0.17	1.78	1.51
<i>Grevillea australis</i> (adult)	2.7	13.3	0.19	1.67	1.44

552

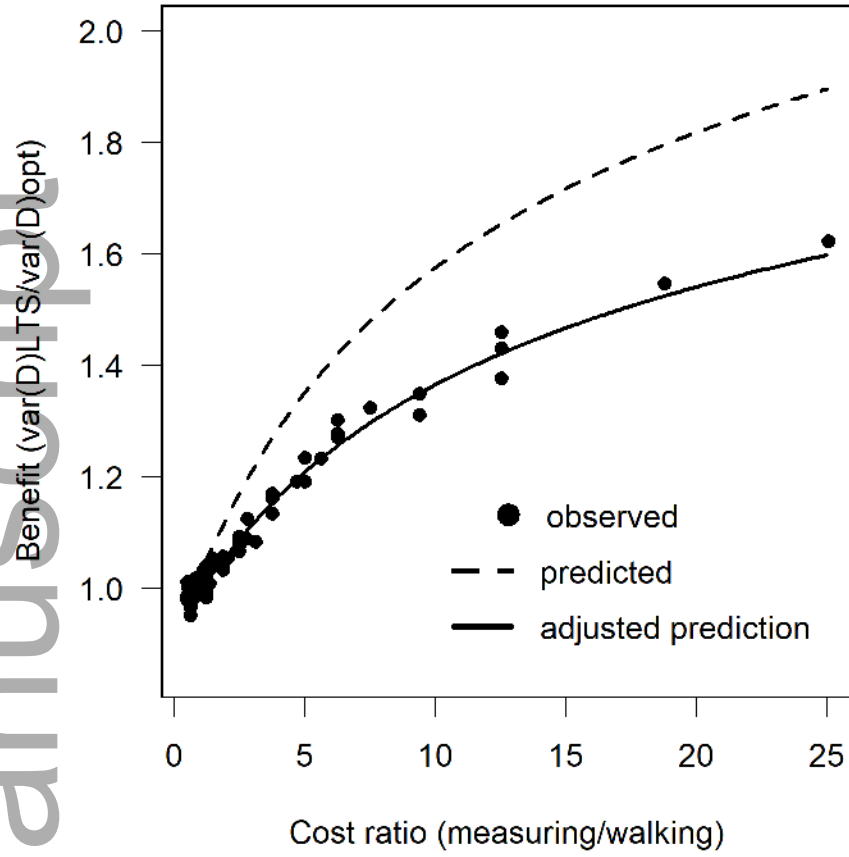
553

554 Table 3 Expected transect lengths (L) and sample sizes (n) for surveys of the three target species with
555 a survey budget of 8 hours, scaled up from the case study surveys (Appendix 4).

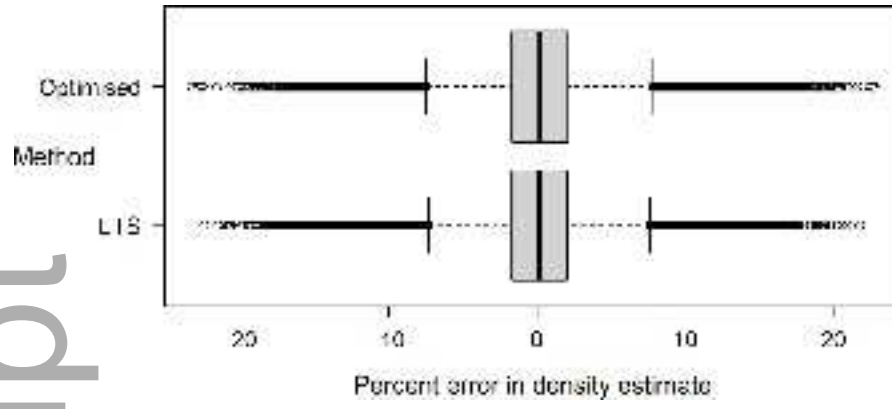
Survey Target	L_{LTS} (m)	n_{LTS}	L_{Opt} (m)	n	n_1
<i>Brassica sp.</i>	879	1178	1933	2591	864
<i>Grevillea australis</i> (seedling)	117	1364	554	6460	1079
<i>Grevillea australis</i> (adult)	337	893	1343	3559	691

556

Author Manuscript



mee3_13589_f1.png



mee3_13589_f2.png