



<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper published in *Methods in Ecology and Evolution*. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the original published paper (version of record):

Ståhl, G., Ekström, M., Dahlgren, J., Esseen, P-A., Grafström, A. et al. (2020)
Presence-absence sampling for estimating plant density using survey data with variable plot size
Methods in Ecology and Evolution, 11(4): 580-590
<https://doi.org/10.1111/2041-210X.13348>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

This is the peer reviewed version of the following article: Ståhl, G, Ekström, M, Dahlgren, J, Esseen, PA, Grafström, A, Jonsson, BG. Presence–absence sampling for estimating plant density using survey data with variable plot size. *Methods Ecol Evol*. 2020; 11: 580– 590. which has been published in final form at <https://doi.org/10.1111/2041-210X.13348>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-168880>

1 Estimating density from presence/absence data in clustered populations

2

3 M. Ekström^{a,b,1}, S. Sandring^b, A. Grafström^b, P.-A. Esseen^c, B. G. Jonsson^d, G.
4 Ståhl^b

5 ^a Department of Statistics, USBE, Umeå University, SE-901 87 Umeå, Sweden

6 ^b Department of Forest Resource Management, Swedish University of Agricultural Sciences, SE-901 83 Umeå,
7 Sweden

8 ^c Department of Ecology and Environmental Science, Umeå University, SE-901 87 Umeå, Sweden

9 ^d Department of Natural Sciences, Mid Sweden University, SE-851 70 Sundsvall, Sweden

10

11 Summary

12 1. Inventories of plant populations are fundamental in ecological research and
13 monitoring, but such surveys are often prone to field assessment errors. Pres-
14 ence/absence (P/A) sampling may have advantages over plant cover assess-
15 ments for reducing such errors. However, the linking between P/A data and
16 plant density depends on model assumptions for plant spatial distributions.
17 Previous studies have shown how that plant density can be estimated under
18 e.g. Poisson model assumptions on the plant locations. In this study new
19 methods are developed and evaluated for linking P/A data with plant density
20 assuming that plants occur in clustered spatial patterns.

21 2. New theory was derived for estimating plant density under Neyman-Scott type
22 cluster models such as the Matérn and Thomas cluster processes. Suggested
23 estimators, corresponding confidence intervals, and a proposed goodness of fit
24 test were evaluated in a Monte-Carlo simulation study assuming a Matérn
25 cluster process. Further, the estimators were applied to plant data from envi-
26 ronmental monitoring in Sweden to demonstrate their empirical application.

27 3. The simulation study showed that our methods work well for large enough
28 sample sizes. The judgment of what is “large enough” is often difficult, but

¹Corresponding author. *Email address:* Magnus.Ekstrom@umu.se (M. Ekström)

29 simulations indicate that a sample size is large enough when the sampling dis-
30 tributions of the parameter estimators are symmetric or mildly skewed. Boot-
31 strap may be used to check whether this is true. The empirical results suggests
32 that the derived methodology may be useful for estimating density of plants
33 such as *Leucanthemum vulgare* and *Scorzonera humilis*.

34 4. By developing estimators of plant density from P/A data under realistic model
35 assumptions about plants' spatial distributions, P/A sampling will become a
36 more useful tool for inventories of plant populations. Our new theory is an
37 important step in this direction.

38 **Key-words:** independent cluster process, intensity, Matérn cluster process, plant
39 monitoring, point pattern, sample plots, spatial models, Thomas cluster process,
40 vegetation survey

41 1 | INTRODUCTION

42 Inventories of plant communities are known to pose several challenges (Bonham
43 2013). Although broad-scale surveys of vegetation patterns may be based on remote
44 sensing data (Groom, Mücher, Ihse, & Wrbka, 2006), more detailed information
45 about species occurrences, vegetation cover, or plant densities rely on data from field-
46 based inventories. A common approach is to assess vegetation cover by species or
47 species groups on plots through visual inspection (Bråkenhielm & Liu, 1995; Bonham,
48 2013). However, this method is prone to surveyor judgment and the variability
49 among surveyors in assessing vegetation cover on a plot may be substantial (Gallegos-
50 Torell & Glimskär, 2009; Morrison, 2016). Presence/absence (P/A) sampling is
51 an alternative where only the presence or absence of a set of species on a plot is
52 registered. This sampling method is less prone to surveyor judgment than cover
53 assessments (Kercher, Frieswyk, & Zedler, 2003; Ringvall, Petersson, Ståhl, & Lämås,
54 2005; Milberg et al., 2008).

55 Normal outputs from inventories of plant communities include the abundance

56 of species in terms of plant density, cover, or biomass (Bonham, 2013). In P/A
57 sampling, occurrence proportions are obtained, but such proportions are difficult to
58 interpret since they depend on the used plot sizes (Ståhl et al., 2017). To obtain
59 more easily interpreted outputs from P/A inventories, results need to re-expressed
60 in terms of e.g. plant density. Such outputs need to be based on model assumptions
61 regarding the spatial distribution of plants.

62 A commonly adopted assumption is that plant locations follow a homogeneous
63 Poisson point process (HPPP) model (Bonham, 2013). This model possesses the
64 property of complete spatial randomness, meaning that the events of a pattern are
65 equally likely to occur anywhere and do not interact with each other. With such a
66 model, recalculations from occurrence proportion to plant density is fairly straight-
67 forward (Fisher, 1934; Bartlett, 1935; Ståhl et al., 2017). It should be noted that if
68 the positions of plants follow a HPPP, they show neither positive spatial dependence
69 (clustering) nor negative spatial dependence (regularity). The HPPP assumption is
70 therefore seldom satisfied because plants are typically aggregated into clusters of dif-
71 ferent size and distribution across the landscape (Bonham, 2013; Ståhl et al., 2017).
72 The closely related binomial point process arises from the HPPP by conditioning on
73 the total number of plants in an area of interest. Arrhenius (1921) considers P/A
74 data under such a model, and Royle & Nichols (2003) and He & Reed (2006) show
75 how recalculations from occurrence proportion to plant density can be made.

76 The HPPP implies that the species abundance in a plot follows a Poisson distribu-
77 tion, while the binomial point process implies that it follows a binomial distribution.
78 Another popular model for plot abundance is the negative binomial distribution,
79 which is regarded useful in applications where a clustering alternative is preferred to
80 the HPPP (He & Gaston, 2000, 2007; Hwang & He, 2011). However, only two known
81 homogeneous point processes give the negative binomial distribution for plot abun-
82 dances, and both are extreme cases (Daley & Vere-Jones, 2008). This highlights the
83 need for more elaborate and realistic models for linking P/A data with plant density
84 in clustered populations.

85 Although we recognize the possibility of using inhomogeneous models, where the

86 expected number of plants per area unit is spatially varying, we restrict the discussion
87 in this paper to homogeneous models. We refer to, e.g., Baddeley, Rubak, & Turner
88 (2016) and the references therein for a discussion on inhomogeneous Poisson process
89 models and Gelfand & Shirota (2018) for fusion of P/A data with presence-only data
90 using inhomogeneous log-Gaussian Cox processes.

91 Our objective was to represent a set of locations of plants in a landscape as
92 a point pattern generated by general Neyman-Scott type cluster models, and to
93 propose and evaluate a method for estimating the parameters in the assigned point
94 process model, using data from P/A sampling. A particular objective was to derive
95 an estimator of the intensity of the process (expected number of plants per area unit),
96 and evaluate this estimator using both Monte Carlo simulations and empirical data
97 from environmental monitoring. The intensity of a point process will henceforth be
98 called the plant density, or simply density.

99 2 | MATERIAL AND METHODS

100 2.1 | Theoretical background

101 A clustered pattern can be constructed from a mechanism where “offspring” points
102 are scattered around their respective “parent” points, e.g. young plants cluster
103 around parent plants, where the offsprings arise from seeds or clonal growth
104 (ramets) from the parent plant. To formalize the above, let X be a finite point
105 process on \mathbb{R}^2 . Conditioned on X , let Y_x be a finite point process centered at
106 $x \in X$. If the processes Y_x , $x \in X$, are independent of one another given X , then
107 $Y = \bigcup_{x \in X} Y_x$ is known as an *independent cluster process* (e.g. Lawson & Denison,
108 2002). The data consist of a realization of $Z = Y \cap S$, where $S \subset \mathbb{R}^2$ is a compact set.

109

110 **Assumption P:** The (parent) process X is a HPPP with density τ and the number
111 of (offspring) points in Y_x is Poisson distributed, with mean λ . The points in Y_x
112 are independently generated from $f(t - x|\gamma)$, where f is the density function of a

113 continuous random variable in \mathbb{R}^2 parameterized by γ .

114

115 Under Assumption P, the process $Y = \bigcup_{x \in X} Y_x$ is of Neyman-Scott type (Lawson
116 & Denison, 2002; Baddeley, Rubak, & Turner, 2016). Its density is $\tau\lambda$. By specifying
117 the offspring probability density $f(t-x|\gamma)$ in Assumption P, some well-known point
118 process models of clustering are obtained:

- 119 • If $f(t-x|\gamma)$ in Assumption P is a uniform density in a disc of radius γ centered
120 around the parent x , then the point process is a *Matérn cluster process* (Matérn,
121 1960, 1986). See Fig. 1.
- 122 • If $f(t-x|\gamma)$ in Assumption P is an isotropic bivariate normal density centered
123 around the parent x , with variance γ in the “x” and “y” directions, then the
124 point process is a (modified) *Thomas cluster process* (Thomas, 1949; Diggle,
125 1978).

126 Baddeley, Rubak, & Turner (2016) provide additional examples of point processes
127 that satisfy Assumption P, such as the Cauchy cluster process and the variance-
128 gamma cluster process.

129 The parameter vector $\boldsymbol{\theta} = (\tau, \lambda, \gamma)$ is unknown and needs to be estimated from
130 observed data. In the current paper we will derive estimators of $\boldsymbol{\theta}$ using P/A data
131 from sample plots. Let $N(B)$ denote the number of points that fall in $B \subseteq S$, i.e.,
132 $N(B) = \{z : z \in Z \cap B\}$. Note, $\{N(B) > 0\}$ is the event that at least one point is
133 present in B , and $\{N(B) = 0\}$ denotes absence of points in B .

134 Let

$$135 \quad H(B|\boldsymbol{\theta}) = \exp\left(-\tau \int \left(1 - \exp\left(-\lambda \int_B f(t-x|\gamma)dt\right)\right) dx\right), \quad B \subseteq S. \quad (1)$$

136 For deriving maximum likelihood estimators of $\boldsymbol{\theta}$ under Assumption P and various
137 sample plot designs, the following theorem is of fundamental importance. Among
138 other things, the theorem establishes that $H(B|\boldsymbol{\theta})$ is the probability of absence
139 of points in $B \subseteq S$, given that Assumption P holds true. More generally, given
140 disjoint sets B_1, \dots, B_m , the theorem gives a formula for the probability of absence

141 of points in e.g. the first few of these sets and presence in the remaining ones. The
 142 theorem is essential for defining the likelihood function, which is used after data
 143 are available to describe plausibility of a parameter vector $\boldsymbol{\theta}$. Any parameter vector
 144 that maximizes the likelihood function (or, equivalently, its logarithm) is known as
 145 a maximum likelihood estimator, and intuitively it is the value of $\boldsymbol{\theta}$ that make the
 146 observed data most probable.

147

148 **Theorem 1.** *Let B_i , $i \in M = \{1, \dots, m\}$, be disjoint sets in S , $M_s \subseteq M$, and
 149 $M_s^c = M \setminus M_s$. If Assumption P is valid, then*

$$\begin{aligned}
 150 \quad & P\{N(B_i) > 0, i \in M_s, \text{ and } N(B_i) = 0, i \in M_s^c\} \\
 151 \quad & = H\left(\bigcup_{i \in M_s^c} B_i \mid \boldsymbol{\theta}\right) - \sum_{i \in M_s} H\left(B_i \cup \left[\bigcup_{j \in M_s^c} B_j\right] \mid \boldsymbol{\theta}\right) \\
 152 \quad & + \sum_{i_1, i_2 \in M_s, i_1 < i_2} H\left(B_{i_1} \cup B_{i_2} \cup \left[\bigcup_{j \in M_s^c} B_j\right] \mid \boldsymbol{\theta}\right) - \dots + (-1)^{m_s} H\left(\bigcup_{i \in M} B_i \mid \boldsymbol{\theta}\right),
 \end{aligned}$$

153 where m_s is the number of elements in the set M_s .

154

155 The proof of Theorem 1 is given in Appendix S1, Supporting Information. Usage of
 156 Theorem 1 is illustrated in the next two examples.

157

158 **Example 1.** Consider a concentric plot design, in which the j th innermost circle C_j
 159 has a radius r_j , $j = 1, \dots, k$ (Fig. 2). Let $B_1 = C_1$ and $B_j = C_j \setminus C_{j-1}$, $j = 2, \dots, k$.
 160 We assume that the surveyer starts with the innermost circle and move outwards,
 161 until the first plant (point) is observed. Thus, if no plants are present in B_1, \dots, B_{j-1} ,
 162 and at least one plant is present in B_j , where $j \leq k$, or if no plants are present in
 163 $C_k = \bigcup_{j=1}^k B_j$, then the surveyer is done, and moves on to the next set of concentric
 164 circular plots. Thus, we observe whether the following events are true or false,

$$165 \quad A_0 = \{\text{absence in } C_k\} = \{N(C_k) = 0\},$$

$$166 \quad A_1 = \{\text{presence in } C_1\} = \{N(C_1) > 0\},$$

$$167 \quad A_j = \{\text{presence in } B_j \text{ but not in } C_{j-1}\} = \{N(C_{j-1}) = 0 \text{ and } N(B_j) > 0\}.$$

168 The corresponding probabilities are obtained from Theorem 1,

169
$$\pi_0 = P\{A_0\} = H(C_k|\boldsymbol{\theta}),$$

170
$$\pi_1 = P\{A_1\} = 1 - H(C_1|\boldsymbol{\theta}),$$

171
$$\pi_j = P\{A_j\} = H(C_{j-1}|\boldsymbol{\theta}) - H(C_j|\boldsymbol{\theta}), \quad j = 2, \dots, k.$$

172 **Example 2.** In this example we consider a sample plot design used for monitoring
 173 of biodiversity in Sweden. For a list of plant species, P/A is recorded in subplots
 174 grouped into sets of nine 0.25 m² circular plots (Fig. 3). With such a subplot layout,
 175 C_j , $j = 1, \dots, 9$, we define $B_0 = C_1 \cup C_2 \cup C_3$, $B_1 = C_4 \cup C_5$, $B_2 = C_6 \cup C_7$, and
 176 $B_3 = C_8 \cup C_9$. To reduce complexity we consider events defined using the B_i 's rather
 177 than the C_j 's. For notational convenience, let $B_{j:k} = \cup_{i=j}^k B_i$. The events that we
 178 consider are

179
$$A_0 = \{\text{absence in } B_{0:3}\},$$

180
$$A_1 = \{\text{presence in } B_0 \text{ but not in } B_{1:3}\},$$

181
$$A_2 = \{\text{absence in } B_0 \text{ and presence in exactly one of } B_1, B_2, \text{ and } B_3\},$$

182
$$A_3 = \{\text{presence in } B_0 \text{ and presence in exactly one of } B_1, B_2, \text{ and } B_3\},$$

183
$$A_4 = \{\text{absence in } B_0 \text{ and presence in exactly two of } B_1, B_2, \text{ and } B_3\},$$

184
$$A_5 = \{\text{presence in } B_0 \text{ and presence in exactly two of } B_1, B_2, \text{ and } B_3\},$$

185
$$A_6 = \{\text{absence in } B_0 \text{ and presence in each of } B_1, B_2, \text{ and } B_3\},$$

186
$$A_7 = \{\text{presence in each of } B_0, B_1, B_2, \text{ and } B_3\}.$$

187 The corresponding probabilities, $\pi_j = P\{A_j\}$, $j = 0, \dots, 7$, are obtained using Theo-

188 rem 1 and the fact that the process is invariant under rotations and reflections,

189 $\pi_0 = P\{N(B_{0:3}) = 0\} = H(B_{0:3}|\boldsymbol{\theta}),$

190 $\pi_1 = P\{N(B_{1:3}) = 0 \text{ and } N(B_0) > 0\} = H(B_{1:3}|\boldsymbol{\theta}) - H(B_{0:3}|\boldsymbol{\theta}),$

191 $\pi_2 = 3P\{N(B_{0:2}) = 0 \text{ and } N(B_3) > 0\} = 3(H(B_{0:2}|\boldsymbol{\theta}) - H(B_{0:3}|\boldsymbol{\theta})),$

192 $\pi_3 = 3P\{N(B_{2:3}) = 0, N(B_0) > 0, \text{ and } N(B_1) > 0\}$

193 $= 3(H(B_{2:3}|\boldsymbol{\theta}) - H(B_{0:2}|\boldsymbol{\theta}) - H(B_{1:3}|\boldsymbol{\theta}) + H(B_{0:3}|\boldsymbol{\theta})),$

194 $\pi_4 = 3P\{N(B_{0:1}) = 0, N(B_2) > 0, \text{ and } N(B_3) > 0\}$

195 $= 3(H(B_{0:1}|\boldsymbol{\theta}) - 2H(B_{0:2}|\boldsymbol{\theta}) + H(B_{0:3}|\boldsymbol{\theta})),$

196 $\pi_5 = 3P\{N(B_3) = 0, N(B_0) > 0, N(B_1) > 0, \text{ and } N(B_2) > 0\}$

197 $= 3(H(B_3|\boldsymbol{\theta}) - 2H(B_{2:3}|\boldsymbol{\theta}) - H(B_{0:1}|\boldsymbol{\theta}) + H(B_{1:3}|\boldsymbol{\theta}) + 2H(B_{0:2}|\boldsymbol{\theta}) - H(B_{0:3}|\boldsymbol{\theta})),$

198 $\pi_6 = P\{N(B_0) = 0, N(B_1) > 0, N(B_2) > 0, \text{ and } N(B_3) > 0\}$

199 $= H(B_0|\boldsymbol{\theta}) - 3H(B_{1:2}|\boldsymbol{\theta}) + 3H(B_{0:2}|\boldsymbol{\theta}) - H(B_{0:3}|\boldsymbol{\theta}),$

200 $\pi_7 = P\{N(B_0) > 0, N(B_1) > 0, N(B_2) > 0, \text{ and } N(B_3) > 0\} = 1 - \sum_{j=0}^6 \pi_j.$

201 2.2 | Estimation and hypothesis testing

202 The basis for our study is to link P/A registrations with plant density through
 203 Neyman-Scott type cluster models of plant occurrence. More specifically, focus will
 204 be on data collected according to the sample plot designs described in Examples 1
 205 and 2, but our methodology can also be applied to many other sample plot designs.

206 In Example 1, assume that there are n sets of concentric circular plots, C_{ij} ,
 207 $i = 1, \dots, n, j = 1, \dots, k$, or, in Example 2, assume that there are n sets of circular
 208 subplots, $C_{ij}, i = 1, \dots, n, j = 1, \dots, k$, where $k = 9$. Suppose that the $C_{i\bullet} = \cup_{j=1}^k C_{ij}$,
 209 $i = 1, \dots, n$, are so far apart that it is not unreasonable to assume that the point
 210 patterns $Z_{i'} = Y \cap C_{i'\bullet}$ and $Z_{i''} = Y \cap C_{i''\bullet}$ are independent for all $i' \neq i''$. Let I_{ij} be
 211 the indicator of the event $A_{ij}, i = 1, \dots, n, j = 0, \dots, m$, where $m = k$ in Example 1
 212 and $m = 7$ in Example 2. Note that $\pi_j = \pi_j(\boldsymbol{\theta}), j = 0, \dots, m$, may be regarded as the
 213 probabilities in the $m + 1$ cells of a multinomial distribution, and that $n_j = \sum_{i=1}^n I_{ij}$,
 214 $j = 0, \dots, m$, are the observed frequencies in these cells.

215 Denote the true value of $\boldsymbol{\theta}$ by $\boldsymbol{\theta}_0$. The objective is to estimate $\boldsymbol{\theta}_0$ on the basis of
 216 the observed frequencies, n_j , $j = 0, \dots, m$. Under Assumption P, the log-likelihood
 217 function for this problem is proportional to

$$218 \quad l(\boldsymbol{\theta}) = \sum_{j=0}^m n_j \log \pi_j(\boldsymbol{\theta}), \quad (2)$$

219 and the maximum likelihood estimator of $\boldsymbol{\theta}_0$, denoted $\hat{\boldsymbol{\theta}} = (\hat{\tau}, \hat{\lambda}, \hat{\gamma})$, is defined as a
 220 $\boldsymbol{\theta}$ -value in $\Theta = \{\boldsymbol{\theta} = (\tau, \lambda, \gamma) : \tau, \lambda, \gamma > 0\}$ that maximizes $l(\boldsymbol{\theta})$. Sufficient conditions
 221 under which the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ is consistent and asymptotically
 222 normally distributed are given in Rao (1973, Section 5e.2). It should be noted,
 223 however, that these conditions may be violated if $H(B|\boldsymbol{\theta})$ in (1) is not smooth enough
 224 as a function of γ ; see Rao (1973) for details. For example, for asymptotic normality,
 225 $H(B|\boldsymbol{\theta})$ is not smooth enough if it fails to have first-order partial derivatives which
 226 are continuous at $\boldsymbol{\theta}_0$.

227 The maximum likelihood estimator of the density of the process is $\hat{\tau}\hat{\lambda}$, and for
 228 constructing a confidence interval for the density we argue as follows. Assuming that
 229 the information matrix $I(\boldsymbol{\theta}) = (i_{rs}(\boldsymbol{\theta}))$, given by

$$230 \quad i_{rs}(\boldsymbol{\theta}) = \sum_{j=0}^m \frac{1}{\pi_j(\boldsymbol{\theta})} \frac{\partial \pi_j(\boldsymbol{\theta})}{\partial \theta_r} \frac{\partial \pi_j(\boldsymbol{\theta})}{\partial \theta_s}$$

231 where $\theta_1 = \tau$, $\theta_2 = \lambda$, and $\theta_3 = \gamma$, is non-singular at $\boldsymbol{\theta}_0 = (\tau_0, \lambda_0, \gamma_0)$, let $i^{rs}(\boldsymbol{\theta}_0)$, $r, s =$
 232 $1, 2, 3$, denote the elements of the inverse to the matrix $I(\boldsymbol{\theta}_0)$. By the asymptotic
 233 normality of $\hat{\boldsymbol{\theta}}$, i.e., that

$$234 \quad \sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{D} N(0, [I(\boldsymbol{\theta}_0)]^{-1}),$$

235 and the delta method (e.g. Lehmann, 1999), we have

$$236 \quad \sqrt{n} \left(\log \hat{\tau} + \log \hat{\lambda} - \log \tau_0 - \log \lambda_0 \right) \xrightarrow{D} N \left(0, \frac{i^{11}(\boldsymbol{\theta}_0)}{\tau_0^2} + \frac{i^{22}(\boldsymbol{\theta}_0)}{\lambda_0^2} + \frac{2i^{12}(\boldsymbol{\theta}_0)}{\tau_0 \lambda_0} \right),$$

237 and this result together with yet another application of the delta method yield

$$238 \quad \sqrt{n} \left(\hat{\tau}\hat{\lambda} - \tau_0\lambda_0 \right) \xrightarrow{D} N \left(0, i^{11}(\boldsymbol{\theta}_0)\lambda_0^2 + i^{22}(\boldsymbol{\theta}_0)\tau_0^2 + 2i^{12}(\boldsymbol{\theta}_0)\tau_0\lambda_0 \right).$$

239 Thus, an approximate 95% confidence interval for the density $\tau_0\lambda_0$ of the cluster
 240 process is given by

$$241 \quad \hat{\tau}\hat{\lambda} \pm 1.96\sqrt{\frac{i^{11}(\hat{\boldsymbol{\theta}})\hat{\lambda}^2 + i^{22}(\hat{\boldsymbol{\theta}})\hat{\tau}^2 + 2i^{12}(\hat{\boldsymbol{\theta}})\hat{\tau}\hat{\lambda}}{n}}. \quad (3)$$

242 Corresponding approximate 95% confidence intervals for the individual parameters
 243 are given by

$$244 \quad \hat{\theta}_r \pm 1.96\sqrt{\frac{i^{rr}(\hat{\boldsymbol{\theta}})}{n}}, \quad r = 1, 2, 3, \quad (4)$$

245 where, again, $\theta_1 = \tau$, $\theta_2 = \lambda$, and $\theta_3 = \gamma$.

246 The above results assume that Assumption P is valid. For this reason it is of
 247 interest to assess whether or not our cluster model assumption holds true. For doing
 248 this, one may use the χ^2 goodness of fit statistic for a multinomial distribution (e.g.
 249 Bishop, Fienberg, & Holland 2007). The statistic is defined as

$$250 \quad \chi^2 = n \sum_{j=0}^m \frac{(p_j - \hat{\pi}_j)^2}{\hat{\pi}_j} \quad (5)$$

251 where $p_j = n_j/n$ and $\hat{\pi}_j = \pi_j(\hat{\boldsymbol{\theta}})$. Under the null hypothesis that the cluster process
 252 model is valid, the statistic is asymptotically χ^2 -distributed with $m - 3$ degrees of
 253 freedom (Bishop, Fienberg, & Holland 2007). If the statistic is improbably large
 254 according to that χ^2 distribution, then one rejects the null hypothesis.

255 2.3 | Computational issues

256 Analytic expressions for maximum likelihood estimators in complex models are usu-
 257 ally not easily available, and numerical methods are needed for maximizing log-
 258 likelihood functions. In addition, numerical methods are needed for computing
 259 the $H(B|\boldsymbol{\theta})$ function in (1), on which the probabilities $\pi_j(\boldsymbol{\theta})$ and the likelihood
 260 functions are based. For the Thomas process, the inner integral in $H(B|\boldsymbol{\theta})$, i.e.
 261 $F_{\gamma,x}(B) = \int_B f(t - x|\gamma)dt$, may be computed using an efficient numerical method
 262 described in DiDonato & Jarnagin (1961), which is implemented in, for example,
 263 the pmvnE11 function in the package shotGroups (Wollschlaeger, 2017) written

264 for use in R (R Core Team, 2019). If the point process is a Matérn cluster process,
265 $F_{\gamma,x}(B)$ may be computed analytically (Appendix S2).

266 For computing the outer integral in $H(B|\boldsymbol{\theta})$ we used the `polyCub.SV` function
267 in the R package `polyCub` (Meyer & Held, 2014, Supplement B), which is based
268 on the product Gauss cubature as proposed by Sommariva & Vianello (2007). In
269 `polyCub.SV`, the number of cubature points may be modified via the argument
270 `nGQ`. It defaults to 20. Increasing the number of points increases the accuracy of
271 the computation of the log-likelihood value but also increases the computation time.

272 In R, there are several numerical procedures for maximizing log-likelihood func-
273 tions. We used the general-purpose optimization routine `constrOptim`, which im-
274 plements, among others, the Nelder-Mead and the BFGS algorithms, and with which
275 one may maximize the log-likelihood subject to the constraints that $\tau, \lambda, \gamma > 0$. The
276 BFGS algorithm, which is a quasi-Newton method, uses both log-likelihood function
277 values and gradients to build up a picture of the three-dimensional surface to be
278 maximized, while the Nelder-Mead algorithm uses only values of the log-likelihood
279 function. We have tried both algorithms and found that BFGS is somewhat faster
280 and therefore preferred for computing estimates.

281 2.4 | Case examples

282 2.4.1 | A Monte Carlo study

283 Since the inner integral of $H(B|\boldsymbol{\theta})$ in (1) may be computed analytically for
284 the Matérn cluster process, we considered this particular process in our Monte
285 Carlo study. Realisations of the Matérn cluster process were generated with the
286 `rMatClust` algorithm in the `spatstat` package (Baddeley, Rubak, & Turner 2016)
287 and maximum likelihood estimates of $\boldsymbol{\theta}_0$ were obtained based on concentric plot de-
288 sign data with $r_j = 0.1$, $j = 1, \dots, k$, and $k = 10$ (see Example 1).

289 In total, we studied eight different cases, where the cases refer to various parame-
290 ter setups. For each case, we generated 1000 replications of the process, and for each
291 such replication we computed the maximum likelihood estimate of $\boldsymbol{\theta}_0$ (Appendix S3),

292 performed the χ^2 goodness of fit test (5), and computed the confidence intervals (3)
293 and (4). Based on the replicate estimates of θ_0 , we estimated the median and the
294 mean of the estimators of the individual parameters (τ , λ , and γ) and the density $\tau\lambda$
295 of the Matérn cluster process, for each case considered. Based on the same replicate
296 estimates, we computed actual confidence levels (ACLs) and median lengths of the
297 confidence intervals, as well as actual significance levels (ASLs) of the χ^2 goodness
298 of fit test. In this study, the nominal confidence level and the nominal significance
299 level were taken to be 95% and 5%, respectively.

300 2.4.2 | P/A data from environmental monitoring

301 The National Inventory of Landscapes (NILS) is a nation-wide environmental moni-
302 toring programme with 631 permanent sample units ($5 \times 5 \text{ km}^2$) that form a random
303 systematic grid across Sweden (Esseen, Glimskär, Ståhl, & Sundquist 2007). The
304 programme started in 2003 and includes field inventory (and aerial photo interpre-
305 tation) of permanent sample plots in all types of terrestrial environments. Field
306 sampling is conducted every fifth year in circular plots of different sizes depending
307 on the measured parameters (Ståhl et al., 2011). NILS provides an infrastructure
308 for other monitoring and research programmes that need basic landscape data. Data
309 for this study were obtained from three monitoring projects associated with NILS.
310 These projects use the same method of collecting P/A-data of plants in 9 subplots
311 (Fig. 3), whereas the original NILS methodology only includes 3 subplots per plot.

312 The first part of the data was obtained from a monitoring programme on semi-
313 natural grassland, pastures and meadows, where data were collected in randomly
314 selected grasslands within NILS sample units that earlier have been identified in
315 a national inventory (Jordbruksverket, 2005). The second part was obtained from
316 monitoring of terrestrial habitats (MOTH) under the European Habitats Directive
317 (Gardfjell, Hagner, Adler, & Forsman, unpubl.), and the third part from regional
318 monitoring of grasslands and wetlands (Rygne, 2009). All data were collected dur-
319 ing 2009-2013. From the combined data set only plots classified as pastures and
320 grasslands were included. To minimize variation in conditions further, the sample

321 was restricted to strata 1-5 (Fig. 4), where most grassland plants have their main
322 distribution in Sweden. Only subplots with a tree cover less than 50% were used. Fi-
323 nally, only plots with a complete set of P/A data for all nine subplots were included
324 for analysis ($n = 2109$).

325 As in Ståhl et al. (2017), the theory assumes that plant occurrences on a subplot
326 are registered whenever a predetermined reference point of a plant is located on the
327 subplot. However, registrations of presences were made if any part of a plant was
328 located on a subplot, and therefore we made a correction by adding a presumed
329 average plant radius to each subplot radius in the calculations. The presumed radius
330 of a plant was set to 10 cm, except for *Scorzonera humilis*, where it was set to 12
331 cm.

332 3 | RESULTS

333 3.1 | The Monte Carlo study

334 Following the setup of the Monte Carlo study of the concentric plot design for the
335 Matérn cluster process described in Section 2.4.1, we studied eight different cases.
336 In most cases (Cases 1 to 6), the estimators showed no or very little bias, except for
337 the mean cluster size λ and the density $\tau\lambda$ of the Matérn cluster process, where the
338 estimators tended to have a small upward mean-bias (Table 1). Also, in all these
339 cases, the ACLs and ASLs were close or quite close to their respective nominal levels
340 (Tables 1 and 2), and, as illustrated in Fig. 5 for Case 6, the estimators tended to
341 be approximately normally distributed. The standard errors of the estimates of τ ,
342 λ , and γ and the median lengths of the corresponding confidence intervals increased
343 with increasing values of the respective corresponding true parameters (Table 1).

344 In the last two cases (Cases 7 and 8), the density τ of the parent process and the
345 cluster radius γ were relatively large, and the estimators of γ and $\tau\lambda$ showed only a
346 small upward mean-bias (Table 1). The estimators of τ and λ were, however, more
347 heavily mean-biased (and median-biased). In addition, the ACLs for λ and γ were

348 notably lower than the nominal level. This was noticed also for the ASLs (Table 2).
 349 In both Cases 7 and 8, the estimators had notably skewed distributions, except for
 350 the estimator of the density $\tau\lambda$ (the histograms in Fig. 6 illustrate this for Case 7).
 351 In comparison with Cases 1-6, the sample size n in Cases 7 and 8 needed to be larger
 352 before the asymptotic properties “kicked in.” For these latter two cases, results for
 353 $n = 10,000$ are presented in Tables 3-4 and Fig. 7. The histograms for $\hat{\lambda}$ and $\hat{\gamma}$
 354 for Case 7 (Fig. 7) still show some skewness and some of the estimators in Table 3
 355 still have some small upward mean-biases, but in comparison with the corresponding
 356 results for $n = 2000$ (Tables 1-2 and Fig. 6) the results were much improved.

357 3.2 | P/A data from environmental monitoring

358 In Table 5, the empirical results based on monitoring data are presented for three
 359 different plant species. The p -value for the goodness of fit test of the Matérn cluster
 360 process assumption is given for each species. It can be observed that two of the
 361 species, *Leucanthemum vulgare* and *Scorzonera humilis*, passed the goodness of fit
 362 test. For the chi-square approximation to be valid, a common rule of thumb is that
 363 (estimated) expected frequencies, $n\hat{\pi}_i$, $i = 0, \dots, 7$, should be at least 5. Therefore,
 364 when we performed the goodness of fit test for *L. vulgare* and *S. humilis*, category
 365 $i = 4$ was merged with $i = 6$ and category $i = 5$ with $i = 7$, and, for *Pimpinella*
 366 *saxifraga*, category $i = 4$ was merged with $i = 6$.

367 The Monte Carlo study in the previous subsection suggests that the proposed
 368 estimation method works well when the sampling distributions of the parameter
 369 estimators are symmetric or mildly skewed. To check whether this holds true or not
 370 for the *L. vulgare* data, we applied the bootstrap (e.g. Davison & Hinkley, 1997).
 371 That is, bootstrap samples of size n , with replacement, were drawn from the original
 372 sample of n sets of subplots, and estimates of parameters were computed for each
 373 bootstrap sample. The resulting histograms are shown in Fig. 8. The “bootstrap
 374 distributions” for the density of the parent process, the mean cluster size, and the
 375 density of the Matérn cluster process had only mild skewness, suggesting that the
 376 estimators $\hat{\tau}$, $\hat{\lambda}$, and $\widehat{\tau\lambda}$ are nearly unbiased. The same conclusion was drawn for *S.*

377 *humilis*.

378 4 | DISCUSSION

379 Elzinga, Salzer, & Willoughby (1998) argue that the key advantages of P/A
380 sampling are “that no special skills are required (anyone who can recognize the species
381 can do the monitoring) and that the monitoring requires very little time.” On the
382 other hand, a significant drawback of the method is that it does not generally provide
383 information on plant density, although some authors have studied this problem under
384 simple point pattern models such as the HPPP model (e.g. Fisher, 1934; Ståhl et
385 al., 2017). In this study, we develop new theory for linking P/A data with plant
386 density, and extend previous work to Neyman-Scott type cluster models such as the
387 Matérn and Thomas cluster processes. For practical purposes, this is of importance,
388 since plants typically form clusters of varying scales of patterns across the landscape
389 (Bonham, 2013), which can not be modeled using HPPP models.

390 In addition to deriving a maximum likelihood estimator of plant density, we
391 suggest a corresponding confidence interval for the plant density. Both the estimator
392 and the confidence interval rely on model assumptions, and may fail when the model
393 is incorrect. For this reason we propose a χ^2 goodness of fit test for testing if
394 the P/A data fits the assigned cluster process model. A simulation study shows
395 that the suggested estimator, confidence interval, and test work well when using a
396 suitable plot design together with a large enough sample size n for various clustered
397 populations. Our simulations indicate that a sample size is large enough when the
398 sampling distributions of the parameter estimators are symmetric or mildly skewed.
399 To check whether this holds true or not in a practical application, bootstrap may be
400 used to estimate the sampling distributions (e.g. Davison & Hinkley, 1997).

401 Although the proposed approach for estimating plant density may be imple-
402 mented for a large range of species, we recognize that this may imply significant
403 analytical work. Hence, we believe that a good starting point is to focus on a few
404 focal species, such as invasive species or threatened species. For these, the popula-

405 tion size (density) is of particular interest to estimate and follow. We recommend
406 using a Matérn cluster model initially, unless the nature of the data clearly suggests
407 another choice. The main reason is that its implementation requires less numerical
408 integration than for other Neyman-Scott type cluster models.

409 The impact of deviations from the model assumptions is an important topic for
410 further studies, as well as extensions to inhomogeneous cluster point processes that
411 allow the density of the process to be location dependent. The latter may be obtained
412 by allowing model parameters to depend on covariate information. Of particular
413 interest here are the cleverly constructed inhomogeneous Neyman-Scott processes in
414 Waagepetersen (2007), with special cases such as the inhomogeneous Matérn and
415 Thomas cluster processes (Baddeley, Rubak, & Turner, 2016). Stratified approaches
416 may also be used. Here the strata may be those defined in the sampling design, or
417 post-strata based on land use or land cover categories, or more advanced schemes
418 employing several sources of information available wall-to-wall for the study area
419 (e.g. Saarela et al., 2015).

420 Another important topic for further studies is to explore different P/A sampling
421 designs and to find designs and plot sizes that will yield estimators of plant density
422 with as high precision as possible, given that the design is cost-efficient, reliable, and
423 good enough for practical purposes. For example, a plot design with relatively small
424 plot sizes suitable for one species may not be appropriate for another species with
425 different density. Both theoretical and empirical studies in this direction are needed.
426 A promising candidate for P/A sampling that enables modeling of cluster point
427 processes is the concentric plot design discussed in this paper. Another appealing
428 possibility is P/A sampling of equally sized quadratic field plots, grouped into sets
429 of 2×2 contiguous quadrats (cf. Morrison, Le Brocq, & Clarke, 1995).

430 | AUTHORS' CONTRIBUTIONS

431 M.E. conceived the idea, designed the analysis methodology, and conducted the
432 analyses; G.S. contributed with expertise in field inventories and P/A sampling; S.S.

433 retrieved the data from projects integrated with the National Inventory of Land-
434 scapes (NILS) and contributed to the analysis with NILS knowledge; P.A.E. and
435 B.G.J. contributed with the ecological perspectives underlying the analyses; A.G.
436 contributed to the statistical methodology. All authors contributed to writing the
437 article and the literature review. The final version of the article has been approved
438 by all authors.

439 | ACKNOWLEDGEMENTS

440 We acknowledge the financial support from the Swedish Research Council. The
441 NILS programme is mainly funded by the Swedish Environmental Protection Agency.

442 | DATA ACCESSIBILITY

443 Upon acceptance of the paper, we intend to archive our empirical data at the
444 Dryad Digital Repository.

445 | References

- 446 Arrhenius, O. (1921). Species and area. *Journal of Ecology*, 9, 95–99. doi: 10.2307/2255763
- 447 Baddeley, A., Rubak, E., & Turner, R. (2016). *Spatial point patterns: Methodology and*
448 *applications with R*. Boca Raton, FL: CRC Press.
- 449 Bartlett, M.S. (1935) Appendix in Blackman, G.E. 1935. A study by statistical methods
450 of the distribution of species in grassland associations. *Annals of Botany*, 49, 749–777.
451 doi: 10.1093/oxfordjournals.aob.a090534
- 452 Bishop, Y. M., Fienberg, S. E., & Holland, P. W. (2007). *Discrete multivariate analysis:*
453 *Theory and practice*. New York, NY: Springer Science & Business Media.
- 454 Bonham, C. D. (2013). *Measurements for terrestrial vegetation* (2nd ed.). New York, NY:
455 John Wiley & Sons.

- 456 Bråkenhielm, S., & Liu, Q. (1995). Comparison of field methods in vegetation monitoring.
457 *Water, Air and Soil Pollution*, 79, 75–87. doi: 10.1007/BF01100431
- 458 Daley, D. J., & Vere-Jones, D. (2008). An introduction to the theory of point processes
459 (2nd ed.). New York, NY: Springer-Verlag.
- 460 Davison, A. C., & Hinkley, D. V. (1997). Bootstrap methods and their application. Cam-
461 bridge: Cambridge University Press.
- 462 DiDonato, A. R., & Jarnagin, M. P. (1961). Integration of the general bivariate Gaussian
463 distribution over an offset circle. *Mathematics of Computation*, 15, 375–382. doi:
464 10.1090/S0025-5718-1961-0129116-8
- 465 Diggle, P. J. (1978). On parameter estimation for spatial point processes. *Journal of the*
466 *Royal Statistical Society, Series B*, 40, 178–181.
- 467 Elzinga, C. L., Salzer, D. W., & Willoughby, J. W. (1998). Measuring and monitoring
468 plant populations. BLM Technical Reference 1730-1. BLM National Applied Resource
469 Sciences Center. Denver, Colorado.
- 470 Esseen, P.-A., Glimskär, A., Ståhl, G., & Sundquist, S. (2007). Field instruction for the
471 national inventory of the landscape in Sweden, NILS. Swedish University of Agricultural
472 Sciences, Department of Forest Resource Management, Umeå, SE.
- 473 Fisher, R. A. (1934). *Statistical Methods for research workers* (5th ed.). Edinburgh: Oliver
474 & Boyd.
- 475 Gallegos-Torell, Å., & Glimskär, A. (2009). Computer-aided calibration for visual es-
476 timation of vegetation cover. *Journal of Vegetation Science*, 20, 973–983. doi:
477 10.1111/j.1654-1103.2009.01111.x
- 478 Gardfjell, H., Hagner, Å., Adler, S., & Forsman, H. (unpublished). Habitat Inven-
479 tory by Aerial Photo Interpretation in MOTH – Terrestrial and Seashore Inventory.
480 www.slu.se/en/Collaborative-Centres-and-Projects/moth/.
- 481 Gelfand, A. E., & Shirota, S. (2018). Preferential sampling for presence/absence data
482 and for fusion of presence/absence data with presence-only data. arXiv:1809.01322v1
483 [stat.ME].
- 484 Groom, G., Múcher, C. A., Ihse, M., & Wrבka, T. (2006). Remote sensing in landscape
485 ecology: experiences and perspectives in a European context. *Landscape Ecology*, 21,

486 391–408. doi: 10.1007/s10980-004-4212-1

487 He, F., & Gaston, K. J. (2000) Estimating species abundance from occurrence. *The Amer-*
488 *ican Naturalist*, 156, 553–559. doi: 10.1086/303403

489 He, F., & Gaston, K. J. (2007). Estimating abundance from occurrence: an underdeter-
490 mined problem. *The American Naturalist*, 170, 655–659. doi: 10.1086/521340

491 He, F., & Reed, W. (2006). Downscaling abundance from the distribution of species:
492 occupancy theory and applications. In J. Wu, K. B. Jones, H. Li, & O. L. Loucks
493 (Eds.), *Scaling and uncertainty analysis in ecology: Methods and applications* (pp. 89–
494 108). Dordrecht: Springer.

495 Hwang, W.-H., & He, F. (2011). Estimating abundance from presence/absence maps.
496 *Methods in Ecology and Evolution*, 2, 550–559. doi: 10.1111/j.2041-210X.2011.00105.x

497 Jordbruksverket (2005). Ängs- och betesmarksinventeringen 2002-2004 [Result of the Sur-
498 vey of semi-natural pastures and meadows]. Jordbruksverket, Rapport 1, 44 p. (in
499 Swedish with English summary).

500 Kercher, S. M., Frieswyk, C. B., & Zedler, J. B. (2003). Effects of sampling teams and
501 estimation methods on the assessment of plant cover. *Journal of Vegetation Science*,
502 14, 899–906. doi: 10.1111/j.1654-1103.2003.tb02223.x

503 Lawson, A. B., & Denison, D. G. T., editors (2002). *Spatial cluster modelling*. Boca
504 Raton, FL: Chapman & Hall/CRC Press.

505 Lehmann, E. L. (1999). *Elements of large-sample theory*. New York, NY: Springer.

506 Matérn, B. (1960). Spatial variation: stochastic models and their application to some
507 problems in forest surveys and other sampling investigations. *Meddelanden från Statens*
508 *Skogsforskningsinstitut*, 49(5), 1–144.

509 Matérn, B. (1986). *Spatial Variation*. *Lecture Notes in Statistics* 36. New York, NY:
510 Springer Verlag.

511 Meyer, S., & Held, L. (2014). Power-law models for infectious disease spread. *Annals of*
512 *Applied Statistics*, 8, 1612–1639. doi: 10.1214/14-AOAS743

513 Milberg, P., Bergstedt, J., Fridman, J., Odell, G., & Westerberg, L. (2008). Observer bias
514 and random variation in vegetation monitoring data. *Journal of Vegetation Science*,
515 19, 633–644. doi: 10.3170/2008-8-18423

- 516 Morrison, L. W. (2016). Observer error in vegetation surveys: a review. *Journal of Plant*
517 *Ecology* 9, 367–379. doi: 10.1093/jpe/rtv077
- 518 Morrison, D. A., Le Brocq, A. F., & Clarke, P. J. (1995). An assessment of some
519 improved techniques for estimating the abundance (frequency) of sedentary organisms.
520 *Vegetation*, 120, 131–145. doi: 10.1007/BF00034343
- 521 Rao, C. R. (1957). Maximum likelihood estimation for multinomial distribution. *Sankhyā*,
522 18, 139–148.
- 523 Rao, C. R. (1973). *Linear statistical inference and its applications*. New York, NY: John
524 Wiley & Sons.
- 525 R Core Team (2019). *R: A language and environment for statistical computing*. R Foun-
526 dation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- 527 Ringvall, A., Petersson, H., Ståhl, G., & Lämås, T. (2005). Surveyor consistency in P/A
528 sampling for monitoring vegetation in a boreal forest. *Forest Ecology and Management*,
529 212, 109–117. doi: 10.1016/j.foreco.2005.03.002
- 530 Royle, J. A., & Nichols, J. D. (2003). Estimating abundance from repeated presence-
531 absence data or point counts. *Ecology*, 84, 777–790. doi: 10.1890/0012-
532 9658(2003)084[0777:EAFRPA]2.0.CO;2
- 533 Rygne, H. (ed.) (2009). *Metodutveckling för regional miljöövervakning och miljömål-*
534 *suppföljning via NILS. Länsstyrelsen i Örebro län, 2009:25* (in Swedish).
- 535 Saarela, S., Grafström, A., Ståhl, G., Kangas, A., Holopainen, M., Tuominen, S., ...
536 Hyypä, J. (2015). Model-assisted estimation of growing stock volume using different
537 combinations of LiDAR and Landsat data as auxiliary information. *Remote Sensing of*
538 *Environment*, 158, 431–440. doi: 10.1016/j.rse.2014.11.020
- 539 Sommariva, A., & Vianello, M. (2007). Product Gauss cubature over polygons based
540 on Green’s integration formula. *BIT Numerical Mathematics*, 47, 441–453. doi:
541 10.1007/s10543-007-0131-2
- 542 Ståhl, G., Allard, A., Esseen, P.-A., Glimskär, A., Ringvall, A., Svensson, J., ... Inghe, O.
543 (2011). National Inventory of Landscapes in Sweden (NILS) – scope, design, and expe-
544 riences from establishing a multiscale biodiversity monitoring system. *Environmental*
545 *monitoring and assessment*, 173, 579–595. doi: 10.1007/s10661-010-1406-7

546 Ståhl, G., Ekström, M., Dahlgren, J., Esseen, P.-A., Grafström, A., & Jonsson, B. G.
547 (2017). Informative plot sizes in presence-absence sampling of forest floor vegetation.
548 *Methods in Ecology and Evolution*, 8, 1284–1291. doi: 10.1111/2041-210X.12749

549 Thomas, M. (1949). A generalisation of Poisson’s binomial limit for use in ecology.
550 *Biometrika*, 36, 18–25. doi: 10.2307/2332526

551 Waagepetersen, R. (2007). An estimating function approach to inference for inhomogeneous
552 Neyman-Scott processes. *Biometrics*, 63, 252–258. doi: 10.1111/j.1541-
553 0420.2006.00667.x

554 Wollschlaeger, D. (2017). shotGroups: Analyze shot group data. R package version 0.7.3.
555 <https://CRAN.R-project.org/package=shotGroups>

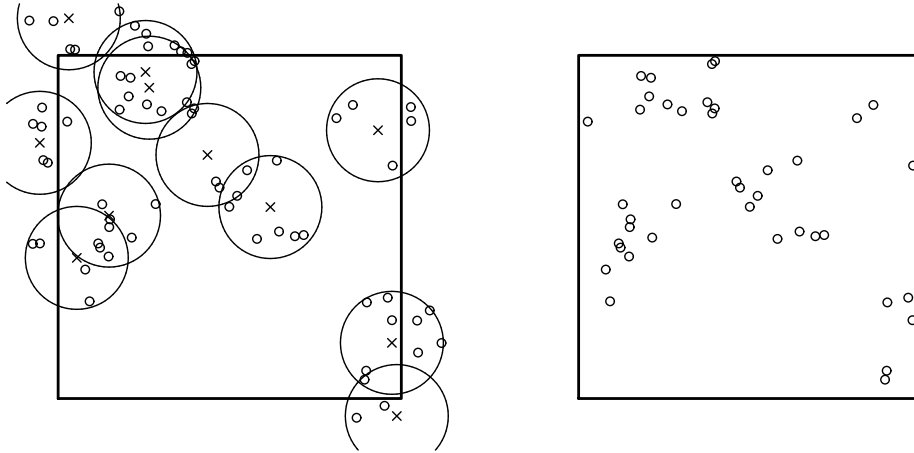


Fig. 1: A Matérn cluster process with parent density $\tau = 6$, mean cluster size $\lambda = 5$, and cluster radius $\gamma = 0.15$. The left panel shows parents (crosses), cluster regions (with radius γ), and offsprings (small open circles). The right panel shows the offsprings that constitute the Matérn cluster process in a square field S .

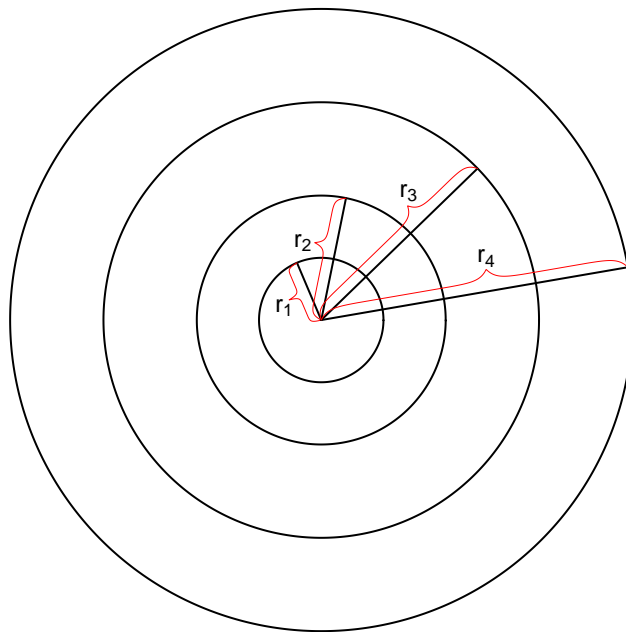


Fig. 2: Plot design with concentric circular sample plots with radii r_1, \dots, r_4 .

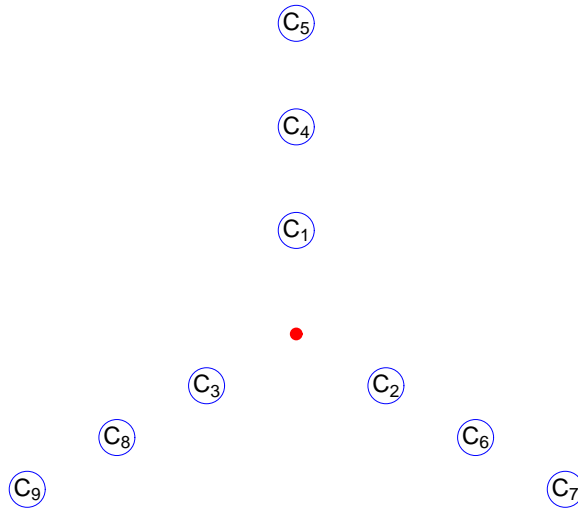


Fig. 3: Field subplot layout in Example 2. The distance from the centre (the red solid circle) to the centre of C_i , $i = 1, 2, 3$, is 3 m. The corresponding distances to C_i , $i = 4, 6, 8$, and to C_i , $i = 5, 7, 9$, are 5 and 7 m, respectively. The area of each C_i is 0.25 m^2 .

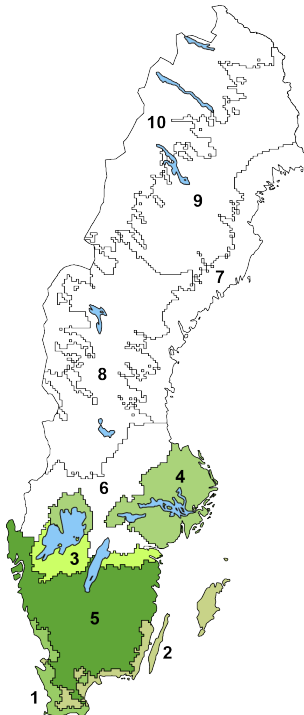


Fig. 4: Map of Sweden showing 10 strata used in NLS. Data from strata 1–5 were selected for the study.

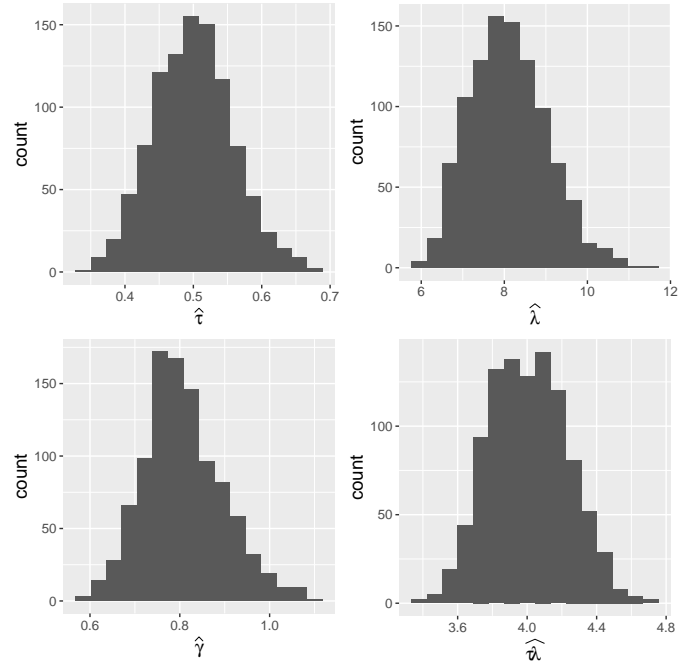


Fig. 5: Histograms of estimates: Case 6 with $n = 2000$.

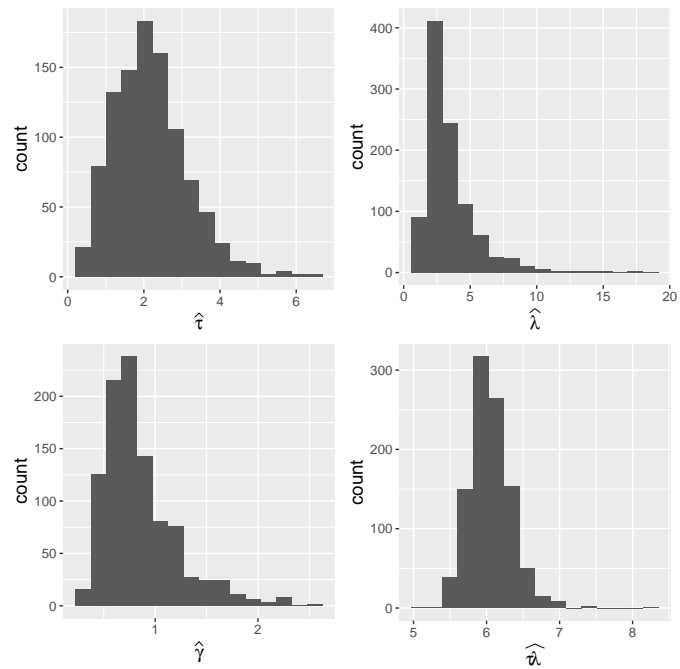


Fig. 6: Histograms of estimates: Case 7 with $n = 2000$.

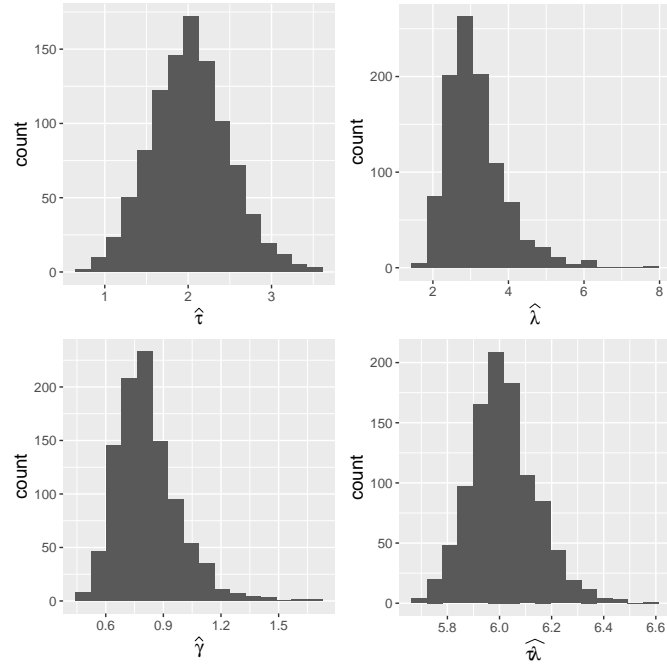


Fig. 7: Histograms of estimates: Case 7 with $n = 10,000$.

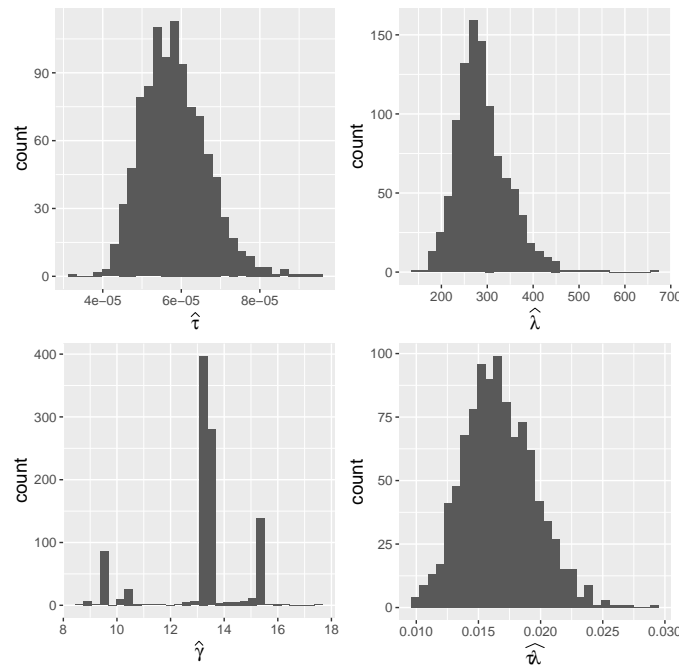


Fig. 8: Histograms of 1000 bootstrap replicates of estimates for the *Leucanthemum vulgare* data.

Table 1: Medians, means and standard errors (SEs) of estimates, and actual confidence levels (ACLs) and median lengths (MedLs) of the associated confidence intervals. The sample size is $n = 2000$.

	Parameter	True value	Median	Mean	SE	ACL (%)	MedL
Case 1	τ	0.50	0.50	0.50	0.04	96.2	0.13
	λ	3.00	3.01	3.12	0.71	95.8	2.35
	γ	0.30	0.30	0.31	0.08	96.2	0.26
	$\tau\lambda$	1.50	1.50	1.56	0.32	94.8	0.98
Case 2	τ	0.50	0.50	0.50	0.02	95.0	0.10
	λ	8.00	7.98	8.09	1.25	94.8	4.48
	γ	0.30	0.30	0.30	0.03	96.8	0.13
	$\tau\lambda$	4.00	3.98	4.06	0.63	94.1	2.24
Case 3	τ	2.00	1.99	2.00	0.16	96.2	0.62
	λ	3.00	3.05	3.05	0.35	96.5	1.42
	γ	0.30	0.30	0.30	0.05	95.5	0.18
	$\tau\lambda$	6.00	6.03	6.08	0.58	94.9	2.14
Case 4	τ	2.00	2.01	2.01	0.15	95.3	0.58
	λ	8.00	8.04	8.06	0.74	95.8	2.91
	γ	0.30	0.30	0.30	0.03	95.3	0.11
	$\tau\lambda$	16.00	16.03	16.17	1.43	95.5	5.22
Case 5	τ	0.50	0.50	0.50	0.07	96.2	0.28
	λ	3.00	3.04	3.11	0.51	97.2	1.70
	γ	0.80	0.80	0.82	0.17	95.2	0.56
	$\tau\lambda$	1.50	1.50	1.51	0.10	94.2	0.39
Case 6	τ	0.50	0.50	0.50	0.06	93.8	0.21
	λ	8.00	8.04	8.11	0.92	95.1	3.38
	γ	0.80	0.80	0.81	0.09	94.4	0.32
	$\tau\lambda$	4.00	4.00	4.01	0.23	95.5	0.88
Case 7 ¹	τ	2.00	2.09	2.18	0.98	93.6	3.82
	λ	3.00	2.89	3.52	2.14	86.5	5.51
	γ	0.80	0.78	0.85	0.36	88.5	1.18
	$\tau\lambda$	6.00	6.02	6.05	0.29	96.1	1.10
Case 8	τ	2.00	2.17	2.44	1.43	94.9	4.64
	λ	8.00	7.36	8.56	4.68	86.2	15.43
	γ	0.80	0.76	0.84	0.83	88.3	1.09
	$\tau\lambda$	16.00	16.07	16.13	0.67	97.1	2.76

¹ The results shown are based on the 999 (out of 1000) replications that converged.

Table 2: Actual significance levels (ASLs) for the goodness of fit test of cases presented in Table 1. The sample size is $n = 2000$.

Case	ASL (%)
1	5.0
2	5.6
3	5.5
4	6.4
5	5.7
6	5.3
7 ¹	3.9
8	2.8

¹ The results shown are based on the 999 (out of 1000) replications that converged.

Table 3: Medians, means and standard errors (SEs) of estimates, and actual confidence levels (ACLs) and median lengths (MedLs) of the associated confidence intervals. The sample size is $n = 10,000$.

	Parameter	True value	Median	Mean	SE	ACL (%)	MedL
Case 7	τ	2.00	2.01	2.02	0.47	93.6	1.78
	λ	3.00	3.00	3.15	0.80	92.3	2.62
	γ	0.80	0.79	0.82	0.16	92.8	0.56
	$\tau\lambda$	6.00	6.01	6.01	0.13	94.4	0.50
Case 8	τ	2.00	2.07	2.10	0.52	94.4	2.08
	λ	8.00	7.77	8.10	2.03	91.6	7.75
	γ	0.80	0.79	0.80	0.13	93.1	0.53
	$\tau\lambda$	16.00	16.00	16.02	0.31	95.2	1.21

Table 4: Actual significance levels (ASLs) for the goodness of fit test of cases presented in Table 3. The sample size is $n = 10,000$.

Case	ASL (%)
7	4.6
8	4.7

Table 5: Estimated parameters of the Matérn cluster process (the estimated density $\hat{\tau}$ of the parent process (parent plants per m^2), estimated mean cluster size $\hat{\lambda}$, estimated cluster radius $\hat{\gamma}$ (m), and estimated density $\hat{\tau\lambda}$ of the Matérn cluster process (plants per m^2)) and the p -value of the goodness of fit test.

Species	$\hat{\tau}$	$\hat{\lambda}$	$\hat{\gamma}$	$\hat{\tau\lambda}$	p -value
<i>Leucanthemum vulgare</i> (oxeye daisy)	0.000063	271.8	12.1	0.017	0.055
<i>Pimpinella saxifraga</i> (burnet-saxifrage)	0.000089	648.0	13.6	0.058	0.00013
<i>Scorzonera humilis</i> (viper's-grass)	0.0000054	1843.3	39.1	0.010	0.68