Citation for the published paper:
Göran Ståhl, Magnus Ekström, Jonas Dahlgren, Per-Anders Esseen, Anton Grafström & Bengt-Gunnar Jonsson. (2017) Informative plot sizes in presence-absence sampling of forest floor vegetation. *Methods in Ecology and Evolution*. Volume: 8, Number: 10, pp 1284-1291.
http://dx.doi.org/10.1111/2041-210X.12749.

2017-02-19

Informative plot sizes in presence-absence sampling of forest floor vegetation

Göran Ståhl[1]
Magnus Ekström[2]
Jonas Dahlgren[1]
Per-Anders Esseen[3]
Anton Grafström[1]
Bengt-Gunnar Jonsson[4]

Ståhl, G., Ekström, M., Dahlgren, J., Esseen, P,-A., Grafström, A. & Jonsson, B.G. 2017. Informative plot sizes in presence-absence sampling of forest floor vegetation. Methods in Ecology and Evolution, DOI: 10.1111/2041-210X.12749,

1) Swedish University of Agricultural Sciences, Dept. of Forest Resource Management, SE 901 83 Umeå, Sweden
2) Umeå University, Dept. of Statistics, SE 901 87 Umeå, Sweden
3) Umeå University, Dept. of Ecology and Environmental Science, SE 901 87 Umeå, Sweden
4) Mid-Sweden University, Dept. of Natural Sciences, SE 851 70 Sundsvall

Abstract

1. Plant communities are attracting increased interest in connection with forest and landscape inventories due to society's interest in ecosystem services. However, the acquisition of accurate information about plant communities poses several methodological challenges. Here we investigate the use of presence-absence sampling with the aim to monitor state and change of plant density. We study what plot sizes are informative, i.e. the estimators should have as high precision as possible.
2. Plant occurrences were modeled through different Poisson processes and tests were developed for assessing the plausibility of the model assumptions. Optimum plot sizes were determined by minimizing the variance of the estimators. While state estimators of similar kind as ours have been proposed in previous studies, our tests and change estimation procedures are new.
3. We found that the most informative plot size for state estimation is 1.6 divided by the plant density, i.e. if the true density is 1 plant per square meter the optimum plot size is 1.6 square meters. This is in accordance with previous findings. More importantly, the most informative plot size for change estimation was smaller and depended on the change patterns. We provide theoretical results as well as some empirical results based on data from the Swedish National Forest Inventory.
4. Use of too small or too large plots resulted in poor precision of the density (and density change) estimators. As a consequence, a range of different plot sizes would be required for jointly monitoring both common and rare plants using presence-absence sampling in monitoring programmes.

Keywords: vegetation survey, sample plots, optimum plot size, plant monitoring, vegetation change, Poisson model, point pattern, density, intensity

Introduction

The demand for accurate information about ecosystem state and change is increasing, not least due to the reporting requirements from international agreements. For example, the United Nations Framework Convention on Climate Change (UNFCCC) and its Kyoto Protocol require regular reporting of greenhouse gas emissions and removals from different land use categories (IPCC 2003). Further, the Convention on Biological Diversity (Secretariat of the CBD, 2005) has the ambition to mainstream the concept of ecosystem services into the management of biological natural resources and programs for mapping and monitoring ecosystem services have been launched.

In perspective of the increasing and changing reporting requirements, existing monitoring programmes may need to be adapted or reoriented in order to deliver the information required (e.g., Cienciala *et al.* 2008; Fridman *et al.* 2014). One important issue concerns the state and change of plant species and communities. For example, national forest inventories (Tomppo *et al.* 2010) have traditionally focused on properties of tree populations. However, non-tree vegetation such as shrubs and forest floor species is obviously important for several ecosystem services, like forage for ungulates, biodiversity, soil protection, and commodities such as berries (e.g., Gamfeldt *et al.* 2013).

The inventory of plant communities poses several challenges (e.g. Greig-Smith 1983; Bonham 2013). Often, the individual plants cannot easily be counted or measured, since it may be difficult to determine the extent of an individual plant. Instead, inventories focus on vegetation cover, biomass, or proportion presences in sample plots (Greig-Smith 1983; Bråkenhielm and Liu 1995; Elzinga *et al.* 1998; Ringvall *et al. 2005;* Godinez-Alvarez *et al.* 2009; Bonham 2013). Assessment of vegetation cover is a normal practice since this measure is intuitively straightforward to interpret. Normally the survey protocols prescribe that cover should be assessed through visual inspection (e.g., Bråkenhielm and Liu 1995; Bonham 2013). One or more surveyors assess the plots and assign cover percentages for the target species or species groups. Such procedures are quick and thus cheap, and they work fairly well for a wide range of plot sizes and shapes. A disadvantage is that data from this type of assessment typically contain both systematic and random measurement errors, due to the subjective judgments (Milberg *et al.* 2008; Bergstedt *et al*. 2009; Morrison 2016). Thus, in large-scale monitoring programmes the surveyors are often trained in order to harmonize the assessments and mitigate assessment errors (e.g., Gallegos-Torell and Glimskär 2009). Despite the potential problems, cover assessment through visual inspection is an important method in large-scale vegetation surveys (e.g., Milberg *et al.* 2008; Ståhl *et al.* 2011).

Point-frequency methods (e.g., Greig-Smith 1983; Bråkenhielm and Liu 1995; Bonham 2013) can be used in order to avoid some of the subjectivity linked to visual cover assessments. With such methods, a needle device (or similar) is located at random locations (often in grid patterns), and is vertically lowered into the vegetation until it either hits one or more layers of vegetation or a non-vegetated area. However, it is difficult to determine exactly where a sampling point is located, especially in multi-layered vegetation, as thus this approach remains partly subjective and systematic errors are sometimes observed. Further, statistically sound cover assessment through the point-frequency method requires large sample sizes and thus the method is expensive to apply.

Presence-absence sampling means that registrations are made only in terms of whether or not a target species or group of species is present in a plot (e.g., Greig-Smith 1983; Elzinga *et al.* 1998; He and Reed 2006; Bonham 2013). The method is simple to conduct, but the interpretation of results is difficult as it depends on plot size and species occurrence patterns. Further, subjective elements remain; some studies indicate that surveyors do not find all target species on a plot during operational surveys (e.g., Ringvall *et al.* 2005). However, especially when fairly small plots are used the method appears to be less prone to surveyor assessment errors compared to the cover assessment method (ibid.). Also, it is relatively cheap to conduct as it requires only the registration of presence or absence of the target species, or species group, on a plot.

It can be anticipated that surveys of plant communities will become even more important in the future in connection with forest and landscape inventories, as a wider range of ecosystem services will be acknowledged and addressed by society. It is likely that presence-absence sampling will often be applied due to its simplicity and low cost. However, interest in changes typically is linked to changes in population density (number of plants per area unit), biomass, or cover, rather than presence proportions (e.g., He and Reed 2006). Thus, inventory estimates based on presence-absence sampling need to be recalculated in order to be meaningfully interpreted. This can be achieved through modeling based on assumptions about the underlying random process that has generated the population and its change.

A straightforward and commonly adopted assumption is a Poisson model (e.g., Bonham 2013), which stipulates a given plant density (typically called intensity in connection with the Poisson process) and entirely random locations of individuals. However, many species occur in more clustered patterns and thus other spatial processes have been explored. Use of the Bionomial distribution for the number of occurrences of species on plots is closely related to the Poisson model assumption. Using this assumption, Royle and Nichols (2003), He and Reed (2006), and Heywood and DeBacker (2007), have shown how occurrence frequencies can be recalculated to plant densities. He and Gaston (2000) used the negative binomial distribution for this recalculation. This distribution has advantages over the standard binomial distribution, since it can account for clustered occurrences of plants. However, some simplifying assumptions had to be applied in order to infer plant density from this distribution. As a further improvement, the Gamma-Poisson model was applied by Hwang and He (2011).

Studies on optimal plot sizes for presence-absence surveys have been conducted by, e.g., Critchley and Poulton (1998) and Heywood and DeBacker (2007). However, as pointed out by Heywood and DeBacker (2007), in many cases informal reasoning rather than statistical analysis has been applied in the search for the best plot sizes (but see Bartlett (1935) for an exception).

The objective of this study was to evaluate what sample plot sizes in presence-absence sampling are most informative when monitoring plant species (or groups of species) in case interest is linked to state and/or change in plant density. By informative plot sizes we mean plot sizes from which state and change in frequency can be recalculated and expressed in terms of precise estimates of state and change in plant density. The focus of our study was change estimation, expanding on previous findings for the case of state estimation (Bartlett 1935). Although a simplification of reality (cf. Hwang and He 2011) the Poisson model was used as a basis for the modeling of plant distributions, since it remains as an important reference for spatial modeling of plant populations (Bonham 2013).

Materials and Methods

Theoretical background

Data in the form of a set of points, irregularly distributed within a region, arise in many different contexts, e.g. the location of plants in a landscape. Such datasets are referred to as spatial point patterns; a (spatial) point process is a stochastic mechanism that generates a countable set of locations of points in the plane.

A homogeneous Poisson point process Λ is characterized by the following two properties:

1. *Poisson distribution*. The number of points of Λ in any finite region of area A follows a Poisson distribution with mean $A\lambda$ for some constant λ.

2. *Independence*. The numbers of points of Λ in *n* disjoint planar regions form *n* independent random variables, for arbitrary *n*.

The number λ is known as the intensity of Λ, and equals the expected number of points to be found in a region of unit area. Although the term *intensity* typically is used in connection with studies using the Poisson process, in this article we will instead use the synonym term *density* since it is commonly used in vegetation ecology.

Estimating population density

The basis for our study was to link presence-absence registrations with plant density through homogeneous Poisson point process models of plant occurrence. We use *n* disjoint sample plots of size *a* within a large study area of size *A*. Under our Poisson model, the probability, *p*, that at least one plant will occur on a given plot is:

$$p = 1 - e^{-a\lambda}. \qquad (1)$$

For a sample survey using *n* plots, *p* can be estimated as $\hat{p} = n^{-1} \sum_{i=1}^{n} I_i$, where I_i is an indicator variable that takes the value 1 if the species is present on plot *i* and 0 otherwise. Note that $I_1, \ldots, I_n$ may be regarded as *n* independent Bernoulli random variables with success probability *p*, and that $\hat{p}$ is the maximum likelihood estimator of *p*. We assume that a reference point, such as the plants' attachment to the ground, is available for determining whether or not a plant individual (ramet) is located within a plot.

Rearranging (1), we can estimate the plant density from the proportion plots with plant occurrences, i.e.

$$\hat{\lambda} = -\frac{\ln(1-\hat{p})}{a}, \qquad (2)$$

which may be regarded as the maximum likelihood estimator of λ. Since $\hat{p}$ and $\hat{\lambda}$ are not linearly related, the estimator can be improved upon by noting that it will be slightly positively biased and thus the expectation of a second order Taylor linearization (plugging in the estimated proportion) may be used as an alternative estimator, i.e. $\hat{\lambda}_{\text{alt}} = -\frac{\ln(1-\hat{p})}{a} - \frac{\hat{V}(\hat{p})}{2a(1-\hat{p})^2}$, with $\hat{V}(\hat{p}) = \hat{p}(1-\hat{p})/n$. Another option to reduce bias would be to utilize the corrected logarithmic transformation presented by Walter (1975, 1976) and Pettigrew *et al.* (1986). For simplicity, we use the non-corrected estimator (2) in the subsequent work.

In order to evaluate what plot size is most informative, i.e. what plot size leads to the smallest variance of the estimator of λ, given the model assumptions, we minimize the variance of $\hat{\lambda}$, for a given density λ. The variance of the estimator (2) is obtained through Taylor linearization as:

$$V(\hat{\lambda}) = \frac{1}{a^2}\frac{1}{(1-p)^2}\frac{p(1-p)}{n} = \frac{1}{na^2}\frac{p}{(1-p)} = \frac{1}{na^2}\frac{1-e^{-a\lambda}}{e^{-a\lambda}}. \qquad (3)$$

If needed, a variance estimator can be obtained by plugging in the estimated proportions in the second last expression for the variance in (3). The optimum plot size can be found by evaluating a range of alternative plot sizes using (3) and identifying what plot size leads to minimum variance. Alternatively, we differentiate $V(\hat{\lambda})$, or $\ln V(\hat{\lambda})$, with respect to *a*, set the derivative to zero, and solve the resulting equation. The derivative of $\ln V(\hat{\lambda})$ with respect to *a* is, after some simplifications,

$$\frac{\partial \ln V(\hat{\lambda})}{\partial a} = -\frac{2}{a} + \frac{\lambda}{1-e^{-a\lambda}}. \tag{4}$$

Setting the above derivative to zero leads to the recursion formula $(a\lambda)_{i+1} = 2\left(1 - e^{-(a\lambda)_i}\right)$, for which $\lim_{i\to\infty}(a\lambda)_i \approx 1.5936$. Here, *i* denotes the iteration step of the recursion. Thus, for a given plant density, λ, the optimum plot size is given by

$$a \approx \frac{1.5936}{\lambda}, \tag{5}$$

and by (1) this corresponds to a probability of presence of about 0.8. Although derived in a slightly different manner, this result is identical to the result presented by Bartlett (1935).

However, this result assumes that the Poisson model assumption is realistic. For this reason it is of interest to assess whether or not the Poisson assumption holds true, and thus a statistical test was developed and applied to our empirical data. The details of the development and application of this test are provided in Appendix 1.

Change estimation

In vegetation surveys, state estimation – as according to formula (2) – would often be of interest since it provides information about the density of different plant species. However, the interest in change estimation, i.e. whether or not the population is increasing or decreasing, typically is even larger. The choice of plot size is very important in this context as substantial changes in population density may not be detected if too large or too small plots are used. In the following sections we develop novel methods for determining informative plot sizes for change estimation.

We address change estimation between two arbitrary time points, t_2 and t_1; the indices 2 and 1 are attached to the different parameters in order to distinguish between the two time points. Our objective is to estimate $\Delta = \lambda_2 - \lambda_1$, given a Poisson model assumption, using the estimator

$$\hat{\Delta} = -\frac{\ln(1-\hat{p}_2)}{a} + \frac{\ln(1-\hat{p}_1)}{a}. \tag{6}$$

Like the state estimator, the change estimator can be slightly improved upon by applying any of the methods previously described; however, this is not pursued here, not least since any bias of the individual state estimates would tend to cancel in the change estimator. Similar to the state estimation case we wish to find the plot size that minimizes the variance of the change estimator.

In case the homogeneous Poisson point processes at time point 2 and time point 1 are independent, utilizing the same number of plots and the same plot size (for simplicity), the variance of (6) follows straightforwardly from (3) as

$$V(\hat{\Delta}) = \frac{1}{na^2}\left[\frac{p_2}{1-p_2} + \frac{p_1}{1-p_1}\right] = \frac{1}{na^2}\left[\frac{e^{-a\lambda_1}+e^{-a\lambda_2}-2e^{-a(\lambda_1+\lambda_2)}}{e^{-a(\lambda_1+\lambda_2)}}\right]. \qquad (7)$$

If $\Delta = \lambda_2 - \lambda_1 = 0$, then the optimum plot size, minimizing $V(\hat{\Delta})$ for a given value of $\lambda = \lambda_1 = \lambda_2$, is identical to the plot size (5) that minimizes the variance given in (3).

However, typically the assumption of independence of the two Poisson point processes is not reasonable, especially when the time between the surveys is short in relation to the life length of individual plants, and we can elaborate on the theory as follows. Many biological processes involve mortality, and so-called thinned point processes provide a way of modeling the mortality of plants. The simplest form of thinning is π-thinning where each point (plant) in the point process is deleted with probability $1-\pi$ and the deletion is independent of the location of the point as well as the deletion or non-deletion of other points. If Λ is a homogeneous Poisson point process with density λ, then the π-thinned Λ is a homogeneous Poisson point process with density $\pi\lambda$.

Poisson processes have a nice "superposition" property: The superposition (union) of independent Poisson processes is again a Poisson process, whose density is the sum of the densities of the individual processes (Cressie 1991). Thus, rather than assuming independent homogeneous Poisson point processes at time point 2 and time point 1, we make the following assumptions: At time point 1, the locations of plants follow a homogeneous Poisson point process Λ_1 with density λ_1. At time point 2, the locations of plants follow the superposition Λ_2 of the independent processes Λ_1^* and Λ_2^*, where the former is a π-thinning of Λ_1 and the latter is a homogeneous Poisson point process of newly regenerated plants with density λ_2^*. Thus, the density of Λ_2 is $\lambda_2 = \pi\lambda_1 + \lambda_2^*$. If $\lambda_2^* = (1-\pi)\lambda_1$, then $\lambda_1 = \lambda_2$, i.e., the density is unchanged from time point 1 to 2. For notational convenience, we define $\lambda_3 = (1-\pi)\lambda_1$, which is the density of plants that have disappeared between the two time points.

The parameters λ_1, $\lambda_2 = \pi\lambda_1 + \lambda_2^*$, and $\lambda_3 = (1-\pi)\lambda_1$ are unknown and need to be estimated from data. Assume that we have presence-absence data from *n* permanent field plots of size *a*, i.e., each of the *n* sample plots from time point 1 are revisited at time point 2. Following the assumptions provided above we obtain the following probabilities for individual permanent plots,

$$\pi_{00} = P(\text{absence of plants at both time points}) = e^{-a(\lambda_2+\lambda_3)}, \qquad (8a)$$
$$\pi_{11} = P(\text{presence of plants at both time points}) = 1 - e^{-a\lambda_1} - e^{-a\lambda_2} + e^{-a(\lambda_2+\lambda_3)}. \qquad (8b)$$
$$\pi_{01} = P(\text{absence at time point 1 and presence at time point 2}) = e^{-a\lambda_1} - e^{-a(\lambda_2+\lambda_3)}, \qquad (8c)$$
$$\pi_{10} = P(\text{presence at time point 1 and absence at time point 2}) = e^{-a\lambda_2} - e^{-a(\lambda_2+\lambda_3)}. \qquad (8d)$$

The details of the derivations are provided in Appendix 2.

Let $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)'$. Note that π_{ij}, $i, j = 0, 1$, which depend on the vector $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)'$, may be regarded as the probabilities in the $k = 4$ cells of a multinomial distribution. Assume that n_{ij}, $i, j = 0, 1$, are the observed frequencies in the k cells of the multinomial distribution, e.g., n_{00} is the observed frequency of sample plots with absence of plants at both time points. The likelihood equations, see e.g. Rao (1973, Section 5e.2), for this multinomial distribution can be written as

$$\sum_{i=0}^{1}\sum_{j=0}^{1}\frac{n_{ij}}{\pi_{ij}}\frac{\partial\pi_{ij}}{\partial\lambda_r}=0, \qquad r=1,2,3. \tag{9}$$

If $\lambda_2+\lambda_3>\lambda_1>\lambda_3>0$, then by Rao (1973)[1] there exists a consistent root $\hat{\boldsymbol{\lambda}}$ of the likelihood equations such that $\sqrt{n}(\hat{\boldsymbol{\lambda}}-\boldsymbol{\lambda})$ is asymptotically normal with (vector) mean zero and covariance matrix $I(\boldsymbol{\lambda})^{-1}$, where $I(\boldsymbol{\lambda})=\left(I_{ij}(\boldsymbol{\lambda})\right)$ is the 3×3 Fisher information matrix with elements

$$I_{rs}(\boldsymbol{\lambda})=\sum_{i=0}^{1}\sum_{j=0}^{1}\frac{1}{\pi_{ij}}\frac{\partial\pi_{ij}}{\partial\lambda_r}\frac{\partial\pi_{ij}}{\partial\lambda_s} \tag{10}$$

Thus, the estimated difference, $\hat{\lambda}_2-\hat{\lambda}_1$, is approximately normally distributed with mean $\lambda_2-\lambda_1$and variance

$$\frac{1}{n}([I(\boldsymbol{\lambda})^{-1}]_{11}+[I(\boldsymbol{\lambda})^{-1}]_{22}-2[I(\boldsymbol{\lambda})^{-1}]_{12}), \tag{11}$$

where $[I(\boldsymbol{\lambda})^{-1}]_{ij}$ is the ijth element of the inverse of $I(\boldsymbol{\lambda})$, and the variance (11) can be consistently estimated by replacing the true unknown $\boldsymbol{\lambda}$ with $\hat{\boldsymbol{\lambda}}$. This result may be used for testing the hypothesis $H_0\colon\lambda_1=\lambda_2$ vs. $H_1\colon\lambda_1\neq\lambda_2$, i.e., the null hypothesis that there is no density change against the alternative, i.e. a change. Alternatively, one may use a chi-square test or a likelihood ratio test. Under the null hypothesis, the chi-square test statistic and (minus) twice the logarithm of the likelihood ratio are approximately chi-square distributed (e.g., Rao 1973, Lehmann and Romano 2005). The usual rule of thumb is that the smallest estimated $n_{ij}\pi_{ij}$ under the null hypothesis should be five or more for the tests to be valid (e.g., Sokal and Rohlf 1995). All three tests were evaluated on the empirical data and they were found to provide almost identical results. However, numerical results are provided only for the likelihood ratio test.

For a given $\boldsymbol{\lambda}$, the optimum plot size *a* that minimizes the variance in (11) can be obtained using standard numerical optimization routines. For example, if there is no change in density and $\lambda=\lambda_1=\lambda_2$, then the optimum plot size a is $1.4771/\lambda$, $1.2876/\lambda$, and $1.1066/\lambda$ for π equal to 0.2, 0.5, and 0.8, respectively, and if π is nearly 0, then the optimum plot size is nearly equal to the one given in (5). If instead π is close

---

[1] We have proved the aforementioned result by verifying that Rao's (1973, Section 5e.2, Result *iv*) three conditions for asymptotic normality hold under our multinomial model. It is easy to see that Rao's weak identifiability condition is fulfilled, and that the parametrizations (8a)-(8d) are smooth enough to have continuous first-order partial derivatives. Finally, by tedious calculations not presented here, we have shown that the Fisher information matrix is non-singular.

to 1, then the optimum plot size is approximately $1/\lambda$. Thus, given that $\lambda = \lambda_1 = \lambda_2$ and that plants survive from time point 1 to time point 2 with a positive probability, the optimum plot size is smaller than the one given in (5).

Evaluations

We evaluated what plot sizes are most informative, i.e. lead to minimum variance of the estimators of density and density change, based on (i) hypothetical assumptions about the density of a plant species and its change between two time periods and (ii) data from the Swedish National Forest Inventory (NFI; Fridman *et al.* 2014), in order to provide empirical results.

The NFI is a sparse sample plot inventory that every year covers the entire country with sample plots. It comprises both temporary and permanent plots; the permanent plots are revisited with a five year interval whereas the temporary plots are visited only once. Forest floor vegetation is assessed every 10 years on permanent plots. At each permanent plot two circular 0.25 m^2 subplots are laid out with a 4 m distance. On these plots the presence or absence of a predetermined set of species is registered. An overview of the species studied and the data is provided in Table 1. For purposes of comparison and for making the Poisson model assumption more plausible, we studied two separate regions (Figure 1) and used only plots from an intermediate forest age class (20-60 years at t_1). (The NFI also registers plants on 100 m^2 plots (Fridman *et al.* 2014) but these data were not utilised in this study.)

Empirical data were acquired for the years 2003-2004 (t_1) and 2013-2014 (t_2). The dataset is available at the Dryad digital repository (Ståhl *et al.* 2017). We estimated the density of the selected species at time point t_1 and the change in density between t_2 and t_1. In addition, we estimated the variance of the estimators as well as what plot size would have minimized the variance of the estimators, assuming that the estimated densities corresponded to the true densities.

The theory assumes that plant occurrences on a plot are registered whenever a predetermined reference point of a plant (such as its rooting) is located on the plot. Since NFI registrations of presences are made if any part of a plant is located on a plot we made a correction by adding an assumed average plant radius to the plot radius in the calculations. The assumed radius of a plant was set to 10 cm, except for the species *Trientalis europaea* where it was set to 3 cm. Statistical tests of the Poisson model assumption, following the theory presented in Appendix 1, were applied and reported (Table 2) as well as tests of significant changes in plant density between the two time points (Table 3).

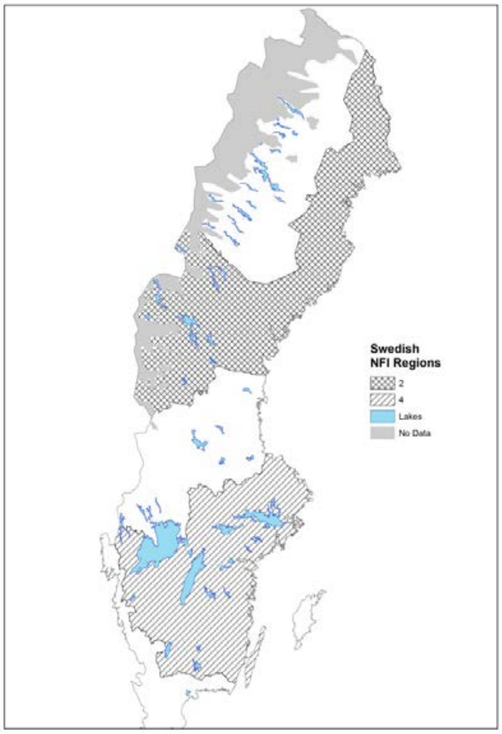

Figure 1. *A map of Sweden and the locations of the NFI regions 2 and 4, which were selected for the study*

Table 1. *An overview of the species studied and the number of plots with recorded presence (P) and absence (A) in regions 2 and 4 and forest age class 21-60 years for the two time periods,* t_1 *(2003-2004) and* t_2 *(2013-2014).*

		Occurrence combination, t_1 and t_2			
Species	Region	P/P	P/A	A/P	A/A
Vaccinium myrtillus (dwarf shrub)	2	166	6	22	26
	4	104	7	28	66
Deschampsia flexuosa (grass)	2	117	24	19	60
	4	105	18	12	70
Luzula pilosa (rush)	2	25	28	24	143
	4	16	7	30	152
Trientalis europaea (forb)	2	22	19	33	146
	4	3	18	11	173
Linnaea borealis (dwarf shrub)	2	25	12	39	144
	4	2	1	5	197

Results

In Figure 2, the most informative plot size for estimating population density, at different actual plant densities, are shown. The calculations were performed according to Eq 5. It can be observed that the optimum plot sizes vary considerably between different densities, according to $a \approx 1.5936/\lambda$.

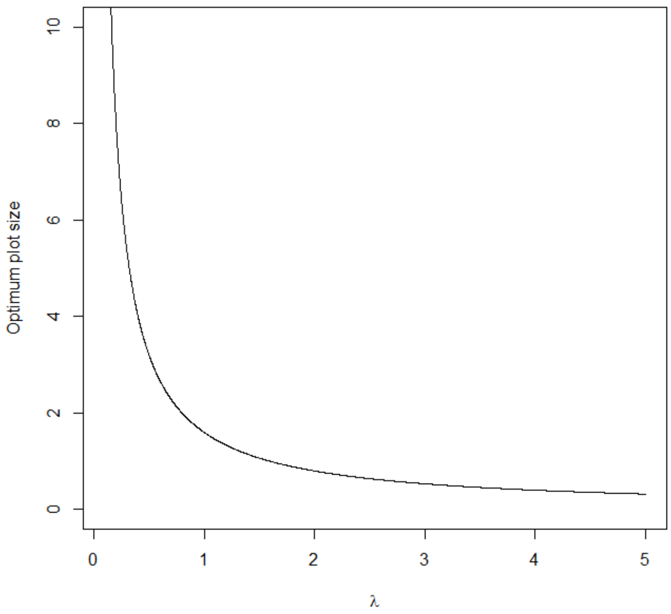

Figure 2. *The most informative plot size for different plant densities, λ. (The area unit is arbitrary as long as the same unit is used for plot area and plant density.)*

In Figure 3, the most informative plot sizes for estimating population change are presented. The results are similar to the results in Figure 2, although it can be noted that the optimum plot sizes for estimating change are smaller than the optimum plot sizes for estimating state.

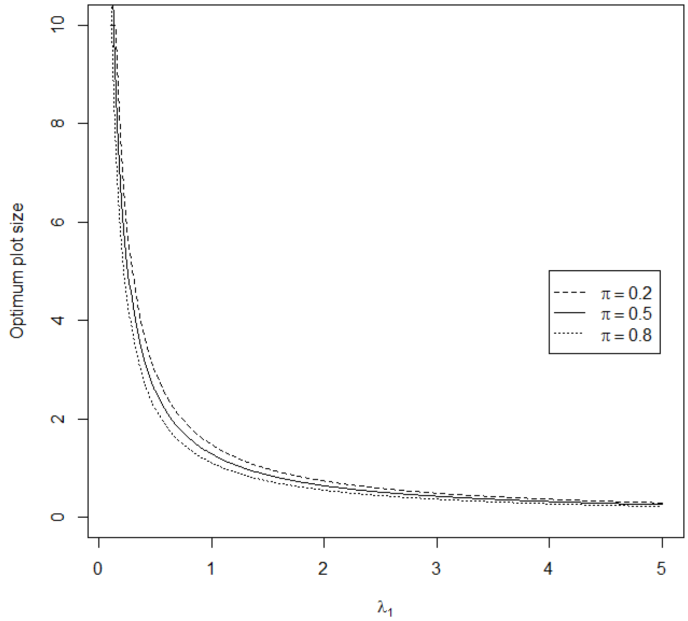

Figure 3. The most informative plot size for estimating change in population density, for a given density λ_1 *and three different values of* π*, where* $\lambda_2^* = (1-\pi)\lambda_1$*, implying that* $\lambda_1 = \lambda_2$*. (The area unit is arbitrary as long as the same unit is used for plot area and plant density.)*

In Tables 2 and 3 the empirical results based on NFI data are presented. In Table 2 the estimated population density of the species, the most informative plot size for the species (assuming that the estimated density corresponds to the true density), and the p-value for the test of the Poisson model assumption are presented for different species and regions.

It can be observed that only few species-region combinations passed the Poisson test. Further, for estimating current density the most informative plot sizes appear to be slightly larger than the plot size actually applied in the NFI.

In Table 3 the changes between the two time points are presented. Note that in this table the p-value reports the probability of the null hypothesis, i.e. that there is no change in population density between the two time points.

Table 2. Estimated density $\hat{\lambda}$ (plants per m^2), estimated optimum plot size $\hat{a}$ (m^2), and p-value of the Poisson test for five vascular plants in two regions.

Species	Region	$\hat{\lambda}$	$\hat{a}$	p-value
Vaccinium myrtillus (dwarf shrub)	2	3.16	0.5	0.00
	4	1.50	1.1	0.00
Deschampsia flexuosa (grass)	2	1.61	1.0	0.00
	4	1.40	1.1	0.00
Luzula pilosa (rush)	2	0.31	5.1	0.16
	4	0.34	4.7	0.02
Trientalis europaea (forb)	2	0.62	2.6	0.00
	4	0.13	12.1	0.41
Linnaea borealis (dwarf shrub)	2	0.46	3.5	0.02
	4	0.05	29.2	0.12

Table 3. Estimated densities $\hat{\lambda}_1, \hat{\lambda}_2$ (plants per m^2 at time points 1 and 2), estimated optimum plot size $\hat{a}$ (m^2), and p-value of the likelihood ratio test for five vascular plants on two regions. The p-value is reported only if the rule of thumb was satisfied, see methods.

Species	Region	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{a}$	p-value
Vaccinium myrtillus (dwarf shrub)	2	2.56	3.16	0.4	0.00
	4	1.18	1.50	0.8	0.00
Deschampsia flexuosa (grass)	2	1.76	1.61	0.7	0.16
	4	1.56	1.40	0.8	0.07
Luzula pilosa (rush)	2	0.37	0.31	4.0	0.23
	4	0.16	0.34	4.5	0.00
Trientalis europaea (forb)	2	0.43	0.62	2.6	0.02
	4	0.23	0.13	7.6	-
Linnaea borealis (dwarf shrub)	2	0.28	0.46	3.4	0.00
	4	0.02	0.05	28.2	-

Under the Poisson model assumption, significant changes in population density were observed for several species-region combinations. For example, the density of *Vaccinium myrtillus* (bilberry) increased significantly in both regions studied. The optimum plot sizes for estimating change were found to be slightly smaller than the optimum plot sizes for estimating state.

Discussion

In this study we expand on previous theory, notably by Bartlett (1935), regarding what plot sizes should ideally be used in presence-absence sampling of plants in order to estimate state and change in plant population density. We denote these plot sizes the informative plot sizes, meaning that they are the plot sizes that maximize the precision of the estimators. Further, we apply the theory to empirical data from the Swedish National Forest Inventory.

Although we used a slightly different approach than Bartlett (1935) we obtained exactly the same result regarding what plot sizes are informative for state estimation; the informative plot size is 1.6 divided by the plant density of the target species, provided that the occurrence of the species follows a Poisson spatial model. For example, for the plant densities 0.1, 1 and 10 plants per square metre the informative plot sizes are 16, 1.6, and 0.16 square metres, respectively. We also observed that large deviations from the informative plot size lead to considerable decreases of the precision of the estimators. Consequently, a single plot size will never fit all species in presence-absence sampling. Instead, a range of different plot sizes would be needed in monitoring programmes addressing several species (with different densities). This can be achieved in different ways, i.e. through the common practice of registering plants in several small co-located quadrats (e.g. Bonham 2013) or by registering species in concentric circular plots of increasing size (e.g., Fridman *et al.* 2014). In both cases only additional species to those already found need to be registered when assessing a new quadrat or a larger plot size at a given sample location.

The main focus of our study was to develop and apply novel estimators for estimating change in plant density, based on presence-absence data, and evaluate what plot sizes are informative in the case of change estimation. Interestingly, in the case of change estimation the informative plot sizes normally are smaller than the informative plot sizes for state estimation. They were also found to be dependent on the patterns of change, especially to what degree plant occurrences at a given site are stable over time. The reason for the smaller informative plot size in this case is the increased chance of actually registering a change on a plot and thus being able to observe a density change. The theoretical results thus are intuitive. However, in case the plant locations between the two time points change entirely at random, the informative plot size for change estimation was found to be identical to the informative plot size for state estimation.

Our results are valid for plant spatial patterns following the Poisson model. While this is an important reference model for modeling plant occurrences (e.g., Bonham 2013) several studies on presence-absence sampling apply different models in order to capture clustered patterns of plant occurrences (e.g., Hwang and He 2011). To make a crude assessment of the impact of deviations from the Poisson model on the results a simulation study was performed using different clustered populations; it is presented in Appendix 3. The conclusion was that small to moderate population model deviations had minor impact on the results in terms of estimated density, whereas substantial population model deviation led to severely biased results. Thus, an important topic for further studies is to explore different cluster processes for linking plant frequencies and densities, and assessing to what extent the informative plot sizes in those cases differ from the ones obtained with the Poisson model assumption. Inventory costs and impacts of measurement errors might also be of interest to study in future analyses, as well as our assumption that a reference point of plant individuals should be possible to distinguish on plants for judging whether or not they should be counted.

The empirical study revealed several interesting results. Firstly, only few species-region combinations passed our test of the Poisson model assumption (*Luzula pilosa* in region 2, *Trientalis europaea* in region 4, and *Linnaea borealis* in region 4), which provides further evidence that the Poisson model is not realistic in many cases. However, ignoring that the Poisson assumption was not adequate in several cases, the informative plot sizes mostly were slightly larger than the plot size actually applied in the Swedish National Forest Inventory (0.25 m^2). However, for *Vaccinium myrtillus* and *Deschampsia flexuosa* the difference between the actual and the most informative plot sizes were fairly small.

Given the model assumptions, significant changes in plant density were observed for almost all five studied species. The changes corresponded well to what might be expected for plants in Swedish forests in the age class 20-60 years at the first time point (i.e. the age interval studied), where the regeneration phase characterized by small trees and fairly open areas gradually is substituted by larger trees and denser forests where shade-tolerant plant species substitute light-demanding species (e.g. Hedwall and Brunet 2016, Tonteri *et al.* 2016). For example, we found the density of *Vaccinium myrtillus* to increase while *Deschampsia flexuosa* decreased. Note that the changes (Table 3) are valid only for the specific regions and forest age class studied, and thus they should not be taken as general trends for these species in Swedish forests. A similar but more comprehensive empirical study based on presence-absence data from Swedish forests was conducted by Odell and Ståhl (1998). In this study the occurrence frequencies of several common forest species were found to increase from the 1980s to the 1990s, but no recalculation to density change estimates was conducted. Recent studies show that many forest floor species have changed their frequencies and cover in boreal and temperate forests in Fennoscandia (Hedwall and Brunet 2016; Tonteri *et al.* 2016).

Conclusion

Use of presence-absence sampling in monitoring plant communities has a potential to overcome some of the problems associated with cover assessments. The present study describes methods for determining optimum plot sizes for estimating plant density state and change based on presence-absence data and Poisson model assumptions. The optimum plot sizes for state estimation were known from previous studies (Bartlett 1935), but our methods for change estimation are novel. For change estimation, the optimum plot sizes were found to be smaller than the optimum plot sizes for state estimation.

The results points to a need for incorporating a system of nested plot sizes in monitoring programmes, since no single plot size will be suitable when several species or species groups are addressed. Based on the current set of species (ranging from dominant to moderately common) it seems likely that informative plot sizes in Fennoscandian forests range from some tenth of a square metre (for dominant species) up to potentially 100 m^2 (for low frequent species). However, further theoretical and empirical studies are needed, taking the spatial distribution and degree of rarity into account, before sampling schemes for different forest ecosystems can be established.

In applications, statistical considerations regarding plot sizes must also be coupled with practical considerations regarding time consumption and measurement errors (e.g. Ringvall et al 2005; Bonham 2013). Our study did not take these issues into account, but assumed that any measurement errors and time consumptions are independent of plot size. Finally, presence-absence sampling in nested plot systems may optionally be combined with cover estimates for all or a set of species to provide information on local plant abundance.

Acknowledgement

We acknowledge the financial support from the Swedish Research Council and the constructive comments from the Associate Editor and two reviewers.

Data accessibility

The data used in this study were obtained from the database of the Swedish National Forest Inventory. They are available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.218n0.

Author contributions

Göran Ståhl conceived the idea and designed the analysis methodology together with Magnus Ekström, who conducted a major part of the analyses. Jonas Dahlgren retrieved the National Forest Inventory (NFI) data and contributed to the analysis with NFI knowledge. Per-Anders Esseen and Bengt Gunnar Jonsson contributed with the ecological perspectives underlying the analyses. Anton Grafström contributed to the statistical methodology. All authors contributed to writing the article and the literature review. The final version of the article has been approved by all authors.

References

Bartlett, M.S. (1935) *Appendix* in Blackman, G.E. 1935. A study by statistical methods of the distribution of species in grassland associations. *Annals of Botany*, **49**, 749–777.

Bergstedt, J., Westerberg, L. & Milberg, P. (2009) In the eye of the beholder: bias and stochastic variation in cover estimates. *Plant Ecology*, **204**, 271–283.

Bonham, C.D. (2013) *Measurements for Terrestrial Vegetation* (2nd ed.). John Wiley & Sons, New York.

Bråkenhielm, S. & Liu, Q. (1995) Comparison of field methods in vegetation monitoring. *Water, Air and Soil Pollution*, **79**, 75–87.

Cienciala, E., Tomppo, E., Snorrason, A., Broadmeadow, M., Colin, A., Dunger, K., Exnerova, Z., Lasserre, B., Petersson, H., Priwitzer, T., Sanchez, G. & Ståhl, G. (2008) Preparing emission reporting from forests: use of National Forest Inventories in European countries. *Silva Fennica*, **42**, 73–88.

Cressie, N. (1991) *Statistics for Spatial Data*. Wiley, New York.

Critchley, C.N.R. & Poulton, S.M.C. (1998) A method to optimize precision and scale in grassland monitoring. *Journal of Vegetation Science*, **9**, 837–846.

Elzinga, C. L., Salzer, D.W. & Willoughby, J.W. (1998). Measuring and monitoring plant populations. BLM Technical Reference 1730-1. BLM National Applied Resource Sciences Center. Denver, Colorado.

Fridman, J., Holm, S., Nilsson, M., Ringvall, A. & Ståhl, G. (2014) Adapting National Forest Inventories to changing requirements – the case of the Swedish National Forest Inventory at the turn of the 20th century. *Silva Fennica,* 48. http://dx.doi.org/10.14214/sf.1095

Gallegos-Torell, Å. & Glimskär, A. (2009) Computer-aided calibration for visual estimation of vegetation cover. *Journal of Vegetation Science,* **20**, 973–983.

Gamfeldt, L., Snäll, T., Bagchi, R., Jonsson, M., Gustafsson, L., Kjellander, P., Ruiz-Jaen, M., Fröberg, M., Stendahl, J., Philipson, C.D., Mikusinski, G., Andersson, E., Westerlund, B., Andrén, H., Moberg, F., Moen, J. & Bengtsson, J. (2013) Higher levels of multiple ecosystem services are found in forests with more tree species. *Nature Communications,* **4**, 1340.

Godinez-Alvarez, H., Herrick, J.E., Mattocks, M., Toledo, D. & Van Zee, J. (2009) Comparison of three vegetation monitoring methods: their relative utility for ecological assessment and monitoring. *Ecological Indicators,* **9**, 1001–1008.

Greig-Smith, P. (1983) *Quantitative Plant Ecology*. Blackwell, Oxford.

He, F. and Gaston, K.J. (2000) Estimating species abundance from occurrence. *The American naturalist,* **156**, 553–559.

He, F. & Reed, W. (2006) Downscaling abundance from the distribution of species: occupancy theory and applications. In: Wu, J., Jones, K.B., Li., H. & Loucks, O.L. (eds). *Scaling and Uncertainty Analysis in Ecology: Methods and Applications*. Springer.

Hedwall, P.-O. & Brunet, J. (2016) Trait variations of ground flora species disentangle the effects of global change and altered land-use in Swedish forests during 20 years. *Global Change Biology*, **22**, 4038–4047.

Heywood, J.S. & DeBacker, M.D. (2007) Optimal sampling designs for monitoring plant frequency. *Rangeland Ecology and Management*, **60**, 426–434.

Hwang, W.-H. & He, F. (2011) Estimating abundance from presence/absence maps. *Methods in Ecology and Evolution*, **2**, 550–559.

IPCC. (2003) *Good Practice Guidance for Land Use, Land-use Change and Forestry*. ISBN 4-88788-003-0. 593 pp.

Lehmann, E.L. & Romano, J.P. (2005) *Testing Statistical Hypotheses*, 3rd Ed. Springer, New York.

Milberg, P., Bergstedt, J., Fridman, J., Odell, G. & Westerberg, L. (2008) Observer bias and random variation in vegetation monitoring data. *Journal of Vegetation Science*, **19**, 633–644.

Morrison, L.W. (2016) Observer error in vegetation surveys: a review. *Journal of Plant Ecology* **9**, 367–379.

Odell, G. & Ståhl, G. (1998) Vegetationsförändringar i svensk skogsmark från 1980-talet till 1990-talet. Resultat från den landsomfattande Ståndortskarteringen. [*Vegetation changes in Swedish forests from the 1980s to the 1990s. Results from the National Survey of Forest Soils and Vegetation*]. *Svensk Botanisk Tidskrift,* **92**, 227–232 (in Swedish).

Pettigrew, H.M., Gart, J.J. & Thomas, D.G. (1986) The bias and higher cumulants of the logarithm of a binomial variate. *Biometrika*, **73**, 425–435.

Rao, C.R. (1973) *Linear Statistical Inference and Its Applications*. John Wiley & Sons, New York.

Ringvall, A., Petersson, H., Ståhl, G. & Lämås, T. (2005) Surveyor consistency in presence/absence sampling for monitoring vegetation in a boreal forest. *Forest Ecology and Management,* **212**, 109–117.

Royle, J.A. & Nichols, J.D. (2003) Estimating abundance from repeated presence-absence data or point counts. *Ecology*, **84**, 777–790.

Secretariat of the Convention on Biological Diversity (2005) *Handbook of the Convention on Biological Diversity Including its Cartagena Protocol on Biosafety*, 3rd edition, (Montreal, Canada).

Sokal, R.R. & Rohlf, F.J. (1995) *Biometry: The Principles and Practices of Statistics in Biological Research*, 3rd Ed. W. H. Freeman and Company, New York.

Ståhl, G., Allard, A., Esseen, P.-A., Glimskär, A., Ringvall, A., Svensson, J., Sundquist, S., Christensen, P., Gallegos Torell, Å., Högström, M., Lagerqvist, K., Marklund, L., Nilsson, B. & Inghe, O. (2011) National Inventory of Landscapes in Sweden (NILS) – scope, design, and experiences from establishing a multiscale biodiversity monitoring system. *Environmental Monitoring and Assessment,* **173**, 579–595.

Ståhl, G., Ekström, M., Dahlgren, J., Esseen, P.-A., Grafström, A., & Jonsson, B.G. (2017). Data from: Informative plot sizes in presence-absence sampling of forest floor vegetation. Dryad Digital Repository. doi:10.5061/dryad.218n0

Tomppo, E., Gschwantner, T., Lawrence, M. & McRoberts, R.E. (eds). (2010) *National Forest Inventories. Pathways for Common Reporting*. Springer.

Tonteri, T., Salemaa, M., Rautio, P., Hallikainen, V., Korpela, L. & Merilä, P. (2016) Forest management regulates temporal change in the cover of boreal plants species. *Forest Ecology and Management*, **381**, 115–124.

Walter, S.D. (1975) The distribution of Levin's measure of attributable risk. *Biometrika*, **62**, 371–374.

Walter, S. D. (1976) The estimation and interpretation of attributable risk in health research. *Biometrics,* **32**, 829–849.

Appendix 1: Testing the hypothesis of a Poisson distribution

The asymptotic normality of the maximum likelihood estimator $\hat{p}$ together with the delta method (e.g., Lehmann 1999) imply that the estimator $\hat{\lambda}$ in (2) is asymptotically normally distributed, i.e.

$$\sqrt{n}\left(\hat{\lambda} - \lambda\right) \to N\left(0, \sigma_{\lambda}^{2}\right),$$

where $\sigma_{\lambda}^{2} = nV\left(\hat{\lambda}\right) = \frac{1}{a^2}\frac{1-e^{-a\lambda}}{e^{-a\lambda}}$.

Suppose that each plot of size *a* is split into two subplots of equal size. The probability, *p_{sub}*, that at least one plant will occur on a given subplot is $p_{sub} = 1 - e^{-\frac{a}{2}\lambda} = 1 - (1-p)^{1/2}$, and p_{sub} can be estimated as $\hat{p}_{sub} = n^{-1}\sum_{i=1}^{n}\left(I_i^{(1)} + I_i^{(2)}\right)/2$, where $I_i^{(j)}$ is an indicator variable that takes the value 1 if the species is present on subplot *j* of plot *i* and 0 otherwise. Let $\boldsymbol{p} = (p, p_{sub})'$ and $\hat{\boldsymbol{p}} = (\hat{p}, \hat{p}_{sub})'$. By the multivariate central limit theorem,

$$\sqrt{n}(\hat{\boldsymbol{p}} - \boldsymbol{p}) \to N(\boldsymbol{0}, \Sigma),$$

where

$$\Sigma = \begin{pmatrix} p(1-p) & p_{sub}(1-p) \\ p_{sub}(1-p) & p_{sub}(1-p_{sub})/2 \end{pmatrix}.$$

Similarly as in (2), we may estimate λ by $\hat{\lambda}_{sub} = -(2/a)\ln(1-\hat{p}_{sub})$. By the asymptotic normality of $\hat{\boldsymbol{p}}$ and the delta method we find that

$$\sqrt{n}\left(\hat{\lambda} - \hat{\lambda}_{sub}\right) \to N\left(0, \frac{\left(1 - e^{-\frac{a}{2}\lambda}\right)^{2}}{a^{2}e^{-a\lambda}}\right),$$

which implies that

$$\frac{\sqrt{n}ae^{-\frac{a}{2}\lambda}\left(\hat{\lambda}-\hat{\lambda}_{sub}\right)}{1-e^{-\frac{a}{2}\lambda}} \to N(0,1)$$

and, by Slutsky's theorem (e.g., Lehmann 1999), that

$$\frac{\sqrt{n}ae^{-\frac{a}{2}\hat{\lambda}_{sub}}\left(\hat{\lambda}-\hat{\lambda}_{sub}\right)}{1-e^{-\frac{a}{2}\hat{\lambda}_{sub}}} \to N(0,1).$$

This result can be used as a basis for tests of whether or not the plants in a given area are distributed according to a homogeneous Poisson process.

Appendix 2: Derivation of the multinomial cell probabilities in Eqs. 8a-8d

At time point 1, the locations of points follow a homogeneous Poisson point process Λ_1 with density λ_1. The points of Λ_1 are deleted with probability $1 - \pi$ (and retained with probability π), where the deletion is independent of the location of the point as well as the deletion or non-deletion of other points. Let Λ_1^* and Λ_1^{**} denote the resulting processes of retained and deleted points, respectively. The point processes Λ_1^* and Λ_1^{**} are independent, and Λ_1 is, by definition, the superposition of Λ_1^* and Λ_1^{**}. At time point 2, the locations of points follow the superposition Λ_2 of the independent processes Λ_1^* and Λ_2^*, where the latter is a homogeneous Poisson point process of newly regenerated plants with density λ_2^*. Thus, the density of Λ_2 is $\lambda_2 = \pi\lambda_1 + \lambda_2^*$. For a point process Λ, let $\Lambda(A)$ denote the number of points in plot A of size a.

We get

$$\begin{aligned}\pi_{01} &= P\{(\Lambda_1(A) = 0) \cap (\Lambda_2(A) \geq 1)\}\\ &= P\{\Lambda_1(A) = 0\}P\{\Lambda_2(A) \geq 1 \mid \Lambda_1(A) = 0\}\\ &= P\{\Lambda_1(A) = 0\}P\{\Lambda_2^*(A) \geq 1\,\}\\ &= e^{-a\lambda_1}\left(1 - e^{-a\lambda_2^*}\right),\end{aligned}$$

$$\begin{aligned}\pi_{00} &= P\{(\Lambda_1(A) = 0) \cap (\Lambda_2(A) = 0)\}\\ &= P\{\Lambda_1(A) = 0\}P\{\Lambda_2^*(A) = 0\}\\ &= e^{-a\lambda_1}e^{-a\lambda_2^*},\end{aligned}$$

$$\begin{aligned}\pi_{10} &= P\{(\Lambda_1(A) \geq 1) \cap (\Lambda_2(A) = 0)\}\\ &= P\{(\Lambda_1^{**}(A) \geq 1) \cap (\Lambda_1^*(A) = 0) \cap (\Lambda_2^*(A) = 0)\}\\ &= \left(1 - e^{-a(1-\pi)\lambda_1}\right)e^{-a\pi\lambda_1}e^{-a\lambda_2^*},\end{aligned}$$

and

$$\pi_{11} = 1 - \pi_{01} - \pi_{00} - \pi_{10} = 1 - e^{-a\lambda_1} - e^{-a(\pi\lambda_1 + \lambda_2^*)} + e^{-a(\lambda_1 + \lambda_2^*)}.$$

Finally, the result in (8a-d) follows from the reparameterization λ_1, $\lambda_2 = \pi\lambda_1 + \lambda_2^*$, and $\lambda_3 = (1 - \pi)\lambda_1$.

Appendix 3: clustered populations - a simulation study

The results in the paper are obtained under a Poisson model. In this appendix, we study the performance of the estimator of density $\hat{\lambda}$, defined in (2), and the corresponding approximate variance $V(\hat{\lambda})$, given in (3), for clustered plant populations. i.e. we wish to assess how sensitive our results are to the Poisson model assumption.

Assume that a clustered point process is formed by first taking a pattern of "parent" points, generated according to a homogeneous Poisson point process with density τ. Then, for each parent point, a random number of "daughter" points are generated, where the number of daughters of each parent is a Poisson random variable with mean μ, and the locations of the daughter points of one parent are independent and bivariate normally distributed around the parent point with covariance matrix $\Sigma = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$. A clustered point process of this kind, consisting of the daughter points, is known as a (modified) Thomas process (Thomas, 1949; Illian et al., 2008). The density of the Thomas process is $\lambda = \tau\mu$.

Design of the simulation experiment:

- i) Generate a point pattern according to a given point process (a homogeneous Poisson point process or a Thomas process).
- ii) Conduct presence-absence sampling using *n* = 200 disjoint circular field plots of size *a*, where the plots are so far apart that it is not unreasonable to assume that the point patterns in plots *i* and *j* are independent for each *i* ≠ *j*.
- iii) Estimate the density of the process using $\hat{\lambda}$, according to (2). That is, the estimate of λ is obtained assuming that the underlying process is a homogeneous Poisson point process.
- iv) Repeat the above steps 1000 times. Calculate the mean and the variance, over the 1000 replications, of the obtained estimates of λ.

The results are presented in Table A3.1. For each process, the true density is $\lambda = 1$. We used two different sample plot sizes in the simulations, $a = 1$ and $a = 2$, and according to Figure 2 the former plot size is too small and the latter too large for being optimal in the case of a homogeneous Poisson point process. For $a = 1$, formula (3) yields $V(\hat{\lambda}) \approx 0.0086$ and for $a = 2$, $V(\hat{\lambda}) \approx 0.0080$. From Table A3.1, we see that a Poisson approximation to the Thomas process works reasonably well when either i) the mean number of daughter points per parent is small, or ii) the daughter points are not tightly scattered around the parent.

Table A3.1. Means and variances of $\hat{\lambda}$.

Process	$a = 1$		$a = 2$	
	mean($\hat{\lambda}$)	var($\hat{\lambda}$)	mean($\hat{\lambda}$)	var($\hat{\lambda}$)
Poisson, $\lambda = 1$	0.998	0.0088	1.005	0.0084
Thomas, $\tau = 0.1, \mu = 10, \sigma = 0.5$	0.449	0.0029	0.338	0.0013
Thomas, $\tau = 0.1, \mu = 10, \sigma = 5$	0.987	0.0080	0.975	0.0074
Thomas, $\tau = 1, \mu = 1, \sigma = 0.5$	0.889	0.0073	0.842	0.0057
Thomas, $\tau = 1, \mu = 1, \sigma = 5$	1.002	0.0083	1.000	0.0089
Thomas, $\tau = 10, \mu = 0.1, \sigma = 0.5$	0.993	0.0087	0.988	0.0075
Thomas, $\tau = 10, \mu = 0.1, \sigma = 5$	1.008	0.0088	1.007	0.0084

References to appendices

Illian, J., Penttinen, A., Stoyan, H. and Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. John Wiley & Sons, Chichester.

Lehmann, E.L. (1999) *Elements of Large-Sample Theory*. Springer, New York.
Thomas, M. (1949). A generalisation of Poisson's binomial limit for use in ecology. *Biometrika* **36**, 18-25.