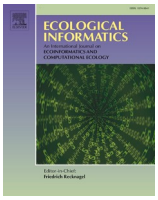
Ecological Informatics 80 (2024) 102377

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf



Estimation of plant density based on presence/absence data using hybrid inference

Léna Gozé [a,*], Magnus Ekström [a,b], Saskia Sandring [a], Bengt-Gunnar Jonsson [c], Jörgen Wallerman [a], Göran Ståhl [a]

[a] Department of Forest Resource Management, Swedish University of Agricultural Sciences, Skogsmarksgränd, 901 83 Umeå, Sweden
[b] Department of Statistics, USBE, Umeå University, Stastistics, 901 87 Umeå, Sweden
[c] Department of Natural Sciences, Design and Sustainable Development, Mid Sweden University, 851 70 Sundsvall, Sweden

ARTICLE INFO

Keywords:
Binary regression
Forest inventory data
Inhomogeneous Poisson point processes
Plant monitoring
Vegetation survey

ABSTRACT

Monitoring of plant populations has become more and more important, especially in the current context of environmental change. In this paper, we propose methods to estimate plant density from presence/absence surveys, wherein the presence or absence of each species is recorded on sample plots. Presence/absence sampling is a useful and relatively simple method for monitoring state and change of plant communities. Moreover, it has advantages compared to traditional plant cover assessment, the latter being more prone to observer bias. We present a hybrid estimation framework, that combines model- and design-based inference features, in which a generalised linear model (for binary presence/absence data) and an inhomogeneous Poisson model (for plant locations) are used to estimate plant density in a region of interest. We look at two different cases, the first one with a known area and the second one where the area is unknown and must be estimated. Our methods are applied to real data on *Vaccinium vitis-idaea* from the Swedish National Forest Inventory as well as simulated data to assess the performance of our estimators of plant density and corresponding variance estimators. The results obtained are promising and indicate that this method has a potential to add considerable analytic strength to monitoring programmes that collect presence/absence data.

1. Introduction

Collecting data on ground vegetation in forests is an important part of environmental monitoring, e.g., as part of initiatives for assessing trends in biodiversity (e.g., Pain et al. 2020; CBD 2002) or reporting within international agreements, such as the EU's Habitats Directive (Commission of the European Communities 2003). The demands for such monitoring programmes are currently increasing (e.g. O'Connor et al. 2020). However, monitoring plant populations is far from trivial. The methods applied should preferably be cost-efficient, easy to apply, and use protocols that avoid assessment errors. Methods based on assessing plant cover fulfil the first two requirements, but they tend to be prone to observer bias and variability due to phenology (e.g., Gallegos Torell & Glimskär 2009; Futschik et al. 2020; Kennedy & Addison 1987; Kercher et al. 2003).

In some cases, especially if the sample plots are not too large, methods based on presence/absence (P/A) sampling are less prone to errors of the kinds mentioned above (e.g., Ringvall et al. 2005; Kercher et al. 2003), since only the presence or absence of target species within plots needs to be registered. Some studies also suggest that P/A-data could be more useful than cover data in characterizing plant communities (e.g., Bastow Wilson 2012). On the other hand, whereas state and change in terms of vegetation cover or plant density are straightforward to interpret, state and change in terms of presence or absence frequencies are vaguer measures, which depend on sample plot size (e.g., Ståhl et al. 2017). However, if plant spatial occurrences are modelled, large-area estimates in terms of state and change of plant density or vegetation cover can be derived from P/A data (Ekström et al. 2020; Ståhl 2003) through application of model-based inference (e.g., Cassel et al. 1977; Warton et al. 2015). In addition, if a model for the probability that at least one plant will occur on a given plot (or pixel) depends on one or more auxiliary variables, then the model-based inferential framework assumes the availability of wall-to-wall auxiliary variables (cf. Fortin et al. 2023).

Auxiliary information is becoming increasingly available through different remote sensing techniques (e.g., Olsson 2020; Baena et al.

* Corresponding author.
E-mail address: lena.goze@slu.se (L. Gozé).

https://doi.org/10.1016/j.ecoinf.2023.102377
Received 12 July 2023; Received in revised form 10 November 2023; Accepted 11 November 2023
Available online 21 November 2023

2018; Dubayah et al. 2022) and so are data about presence of species through citizen science data collection programs (e.g., the Species Observation System in Sweden (Artdatabanken 2022) or the Atlas of Living Australia and its citizen science data portal (Belbin 2011)), which can be combined with P/A data (Fithian et al. 2015). Thus, opportunities for modelling plant occurrence are much better today compared to some decades ago. This type of modelling, with the availability of wall-to-wall auxiliary information from, e.g., remote sensing, can offer information in terms of both estimates and maps. Estimates are needed, e.g., for trend analysis and reporting to agreements such as the Habitats Directive mentioned above. Maps are useful for implementing management plans related to preserving threatened species (Baena et al. 2018) or limiting the impact of invasive species.

As the degree of detail in the auxiliary data increases, it will be possible to develop better models for plant occurrences, thus facilitating model-based estimation of plant density with higher precision. Dense networks of field plots from National Forest Inventories (NFI, e.g., Fridman et al. 2014; Tomppo et al. 2010) could provide such auxiliary data, because very detailed descriptions of biotic and abiotic conditions, including soil variables, are made on such plots. However, with sample plot data alone, i.e. without wall-to-wall data, it is not possible to apply the standard theory of model-based inference. Instead, hybrid inference can be an alternative (e.g., Corona et al. 2014; Ståhl et al. 2016), where features of model-based and design-based inference are combined.

Examples of applications of hybrid inference include biomass surveys based on LiDAR sample data in Norway (Ståhl et al. 2011) and North America (Margolis et al. 2015), biomass prediction for temperate and pan-tropical regions in the context of the Global Ecosystem Dynamics Investigation project (Saarela et al. 2022), comparison of forest biomass estimates based on coarse and fine resolution data in the USA (McRoberts et al. 2019), and estimation of growing stock volume in Italy (Corona et al. 2014), Finland (Saarela et al. 2015), and Spain (Condés and McRoberts 2017). It has been applied to a broad variety of models, such as mixed-effect models (Fortin et al. 2016) and more complex models where variance estimation requires resampling methods such as the parametric bootstrap (Fortin et al. 2018).

Using conventional model-based inference, Ekström et al. (Unpublished results) investigated the use of P/A data for regional estimation of plant density for a selection of plant species occurring mainly in forests. The main components of the study were inhomogeneous Poisson point processes for modelling the spatial locations of plants and generalised linear models (GLMs) with a complementary log-log link function for associating P/A data with the intensity of the point process, taking auxiliary remotely sensed data into account. As will be described in detail later, a similar modelling approach is used in the present study, with the important difference that auxiliary data were obtained from a large probability sample rather than from wall-to-wall remote sensing. A GLM with a complementary log-log link function for modelling P/A data has also been used in other studies, such as Yee & Mitchell (1991), Royle & Dorazio (2008), Lindenmayer et al. (2009), Baddeley et al. (2010) or Fithian et al. (2015). However, contrary to these articles, which focus on pixel-wise estimation for, e.g., producing maps, our study focuses on obtaining large-area estimates of plant density based on data collected exclusively from sample plots. To our knowledge, no previous studies that make use of hybrid inference have been conducted based on GLMs.

A complementary log-log link function has also been used for modelling of presence-only data (e.g., Phillips et al. (2017); Wan et al. (2017); Sreekumar & Nameer (2022)), although none of them make use of hybrid inference. In addition, it should be mentioned that the standard logit link is frequently used in studies analysing P/A data of species occurrences (e.g., Foody 2008; Ekström et al. 2018; Esseen et al. 2022; Esseen & Ekström 2023). However, for the case where the locations of plants are regarded as a realisation of an inhomogeneous Poisson point process, Baddeley et al. (2010) provide an explanation of why the complementary log-log link function should be preferred for modelling P/A data.

The objective of this study is to assess the usefulness of hybrid inference for estimating plant density, where GLMs estimated from a small sample of P/A data (and auxiliary data) were applied to a large sample of auxiliary data from the Swedish NFI. An important part of the study is to develop formal plant density estimators, variances, and variance estimators for this approach, because no previous studies are available where hybrid inference has been applied in this modelling context. The performance of our estimators and corresponding variance estimators was examined through Monte Carlo simulations and the use of empirical NFI data on a common dwarf shrub, *Vaccinium vitis-idaea*.

We choose to focus our study on estimating the expected plant density (we refer to (13) for a precise definition) rather than on predicting the actual plant density (which is a random quantity in our study setting). The main reason is that this approach simplifies the analyses to some extent meanwhile, for large-area surveys, the relative difference between actual plant density and its expected value is very small, if the models used are approximately correct (cf. Ståhl et al. 2016). The motivation for studying plant density rather than the absolute number of plants is that density is a more relevant measure for plants with large populations (in contrast to many animals), and because the measure allows for comparison between regions of different size.

2. Methods

In this section, we first explain the necessary basis for our derivations, then propose estimators of the expected number of plants in a region of interest U, where U can be, e.g., a municipality, a province or a country. Furthermore, we develop variance formulas and corresponding variance estimators. The estimator of the expected density, defined as the expected number of plants per unit area, is thereafter obtained via the estimator of the expected number of plants and is presented for two cases: one with known area a_U of U and one with unknown area. We also look at the case where we want to estimate the expected density for a specific domain within U, for example the forested part of U. Two different sampling designs are considered. In the first design, plot centres are sampled according to some joint probability density function on U, or rather the union of U and a so-called "buffer" for handling edge effects (Subsections 2.2–2.4). In the second design, centres of clusters of plots are sampled rather than individual plot centres (Subsection 2.5).

2.1. Models

Assume that the plant population is generated by an inhomogeneous Poisson point process with intensity

$$\lambda_{\boldsymbol{\beta}}(\boldsymbol{u}) = \exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}(\boldsymbol{u})\right), \boldsymbol{u} \in U \subset \mathbb{R}^2 \tag{1}$$

(Baddeley et al. 2010), where $\boldsymbol{\beta} \in \mathbb{R}^q$ denotes the vector of model parameters and $\boldsymbol{x}(\boldsymbol{u})$ denotes a covariate vector (of length q) at point $\boldsymbol{u}$. The expected number of plants in U is then given by

$$\Lambda(\boldsymbol{\beta}) = \int_U \lambda_{\boldsymbol{\beta}}(\boldsymbol{u}) d\boldsymbol{u}. \tag{2}$$

We consider plots $C(\boldsymbol{u}_i)$, where index i designates plot i, and where the plot centres $\{\boldsymbol{u}_i\}$ are selected according to some specified sampling design. Let N_i denote the number of plants in $C(\boldsymbol{u}_i) \cap U$. Our assumptions imply that N_i is Poisson distributed, and then

$$\mathbb{E}(N_i) = \int_{C(\boldsymbol{u}_i)\cap U} \lambda_{\boldsymbol{\beta}}(\boldsymbol{u}) d\boldsymbol{u} = \int_{C(\boldsymbol{u}_i)\cap U} \exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}(\boldsymbol{u})\right) d\boldsymbol{u}.$$

Unless stated otherwise, we assume, as an approximation, that $\boldsymbol{x}(\boldsymbol{u})$ is constant in a sample plot, and thus $\boldsymbol{x}(\boldsymbol{u}) = \boldsymbol{x}(\boldsymbol{u}_i) = \boldsymbol{x}_i$ for all $\boldsymbol{u} \in C(\boldsymbol{u}_i)$, and

$$\mathbb{E}(N_i) = a_i \exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_i\right), \tag{3}$$

with a_i being the area of the intersection of plot $C(\boldsymbol{u}_i)$ and the region of interest U (cf. Baddeley et al. 2010). Since N_i is Poisson-distributed, the probability of presence can be expressed by

$$p_i = 1 - P(N_i = 0) = 1 - \exp\big(-a_i \exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_i)\big) \tag{4}$$

so that the loglikelihood for the binary response variables (i.e. P/A data from $C(\boldsymbol{u}_i) \cap U$) becomes the loglikelihood of a complementary log-log regression with an offset equal to the log of the plot area, i.e. of the binary regression model given by

$$g(p_i) = \log(a_i) + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_i, \quad \text{where} \quad g(p) = \log(-\log(1-p)). \tag{5}$$

According to Baddeley et al. (2010), the corresponding likelihood may be regarded as an approximation of the likelihood that would have been obtained without the assumption of constant covariate data in a plot.

2.2. Estimation of the expected number of plants in U

Hybrid inference can be used when covariate information is not available everywhere in the region of interest but only at sample plot level, for example for budgetary reasons (Ståhl et al. 2016). As stated in the introduction, this hybrid method includes aspects of both design-based and model-based inference. As in, amongst others, the papers by Ståhl et al. (2011), Nelson et al. (2012), Corona et al. (2014), Saarela et al. (2015) or Saarela et al. (2022) on hybrid inference, we utilise two samples that are readily available, for instance in monitoring programme databases. Our first sample S_1 of size n_1 contains plot centre locations for plots with both binary response data and covariate data, while our second sample S_2 of size n_2 contains plot centre locations for plots with only covariate data. Typically, n_2 is much larger than n_1. Sample S_1 is used only to establish a model and estimate the vector of model coefficients in a GLM (as opposed to, e.g., Ståhl et al. (2011), where a standard linear model is used). Thereafter, the fitted GLM and covariate information from S_2 are used to predict expected numbers of plants on all plots with centres in S_2, and subsequently the expected plant density in the region of interest, using design-based estimation and Horvitz-Thompson-like estimators. Sample plots with centre locations in S_1 and S_2 do not necessarily need to have the same size, and the sampling designs used to obtain the data in S_1 and S_2 are allowed to differ.

When sampling from a finite population, the well-known Horvitz-Thompson estimator (Horvitz & Thompson 1952) is often used for obtaining estimates of population parameters. However, in our case the population is not finite but a continuous set of locations, and therefore we use Cordy's continuous analogue of the Horvitz-Thompson estimator (Cordy 1993), which we introduce next.

Let f be the joint probability density function (pdf) for sample $S_2 = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_{n_2}\}$, and $f_i(\boldsymbol{u})$ the marginal pdf for point $\boldsymbol{u}_i$. The inclusion density function is

$$\pi(\boldsymbol{u}) = \sum_{i=1}^{n_2} f_i(\boldsymbol{u}), \tag{6}$$

and it can intuitively be considered as a local measure of the number of sample points to be selected per unit area (Cordy 1993). If, for example, the points in S_2 are independent and identically distributed (iid), this means that $\pi(\boldsymbol{u}) = n_2 f_1(\boldsymbol{u})$.

The inclusion zone for a point $\boldsymbol{u} \in U$ consists of all points in the frame that would result in the inclusion of $\boldsymbol{u}$ if they were selected to the sample. It may be formally written as $K(\boldsymbol{u}) = \{\boldsymbol{u}' \in U : \boldsymbol{u} \in C(\boldsymbol{u}')\}$, where $C(\boldsymbol{u}')$ is a plot centred around point $\boldsymbol{u}'$. For simplicity purposes, we assume from here on that all plots $C(\boldsymbol{u}_i)$, $i \in S_2$, are circular and have the same area a. The area of the inclusion zone of $\boldsymbol{u} \in U$ is $\widetilde{a}_{\boldsymbol{u}} = \int_U I(\boldsymbol{u} \in C(\boldsymbol{u}'))d\boldsymbol{u}'$. If point $\boldsymbol{u}$ is sufficiently into the interior of U, then its inclusion zone will have the same shape and size as each of the circular plots. On the other hand, if $\boldsymbol{u}$ is close enough to the boundary of U, then its inclusion zone will have a smaller size than a. The Horvitz–Thompson-type estimator presented below has the ability to take this into account, but would require the inclusion zone area to be determined for each point $\boldsymbol{u}_i \in S_2$ near the edge (cf. Gregoire & Valentine 2007). A less labour-intensive way to solve this problem is to use the so-called buffer method, which applies to both the single-plot and cluster-plot designs. Thus, we suppose that a buffer at least as large as the plot radius is used around U (Gregoire & Valentine 2007). This allows sample points $\boldsymbol{u}_i$ to fall outside U, i.e. in some larger region $U^{\bullet}$, defined as the union of U and the buffer. The use of a buffer impacts the definitions of $\widetilde{a}_{\boldsymbol{u}}$ and $K(\boldsymbol{u})$, in which U needs to be replaced by $U^{\bullet}$. The introduction of a buffer implies that all points in U have the same inclusion zone area, and thus $\widetilde{a}_{\boldsymbol{u}} = a_{\boldsymbol{u}} = a$ for all $\boldsymbol{u} \in U$, where $a_{\boldsymbol{u}}$ denotes the area of $C(\boldsymbol{u})$. In this setting, we set $\lambda_{\boldsymbol{\beta}}(\boldsymbol{u}) = 0$ for all $\boldsymbol{u} \in U$ (cf. Gregoire & Valentine 2007).

The "generalised" Horvitz-Thompson estimator of the expected number of plants in U is then given by

$$\widehat{\Lambda}(\boldsymbol{\beta}) = \sum_{i=1}^{n_2} \frac{\lambda(\boldsymbol{u}_i)}{\pi(\boldsymbol{u}_i)}, \tag{7}$$

where $\pi(\boldsymbol{u})$ is given by (6) and

$$\lambda(\boldsymbol{u}) = \int_{C(\boldsymbol{u})} \frac{\lambda_{\boldsymbol{\beta}}(\boldsymbol{u}')}{a_{\boldsymbol{u}'}} d\boldsymbol{u}', \boldsymbol{u} \in U^{\bullet},$$

is the average intensity over $C(\boldsymbol{u}_i)$, where $a_{\boldsymbol{u}'} = a$ by our assumptions (Cordy 1993, Grafström et al. 2017). Note that

$$\begin{aligned} \int_{U^{\bullet}} \lambda(\boldsymbol{u}) d\boldsymbol{u} &= \int_{U^{\bullet}} \int_{C(\boldsymbol{u})} \frac{\lambda_{\boldsymbol{\beta}}(\boldsymbol{u}')}{a_{\boldsymbol{u}'}} d\boldsymbol{u}' d\boldsymbol{u} = \int_{U^{\bullet}} \frac{\lambda_{\boldsymbol{\beta}}(\boldsymbol{u}')}{a_{\boldsymbol{u}'}} \int_{U^{\bullet}} I(\boldsymbol{u}' \in C(\boldsymbol{u})) d\boldsymbol{u} d\boldsymbol{u}' \\ &= \int_{U} \frac{\lambda_{\boldsymbol{\beta}}(\boldsymbol{u}')}{a_{\boldsymbol{u}'}} \int_{U^{\bullet}} I(\boldsymbol{u}' \in C(\boldsymbol{u})) d\boldsymbol{u} d\boldsymbol{u}' = \int_{U} \lambda_{\boldsymbol{\beta}}(\boldsymbol{u}') d\boldsymbol{u}' = \Lambda(\boldsymbol{\beta}) \end{aligned} \tag{8}$$

and, according to Theorem 1 in Cordy (1993), this implies that the Horvitz-Thompson estimator of $\Lambda(\boldsymbol{\beta})$ is unbiased if $\pi(\boldsymbol{u}) > 0$ for all $\boldsymbol{u} \in U^{\bullet}$. Hence, with a buffer for handling edge effects, we obtain an unbiased estimator of $\Lambda(\boldsymbol{\beta})$. The price to be paid is that the buffer method tends to inflate the variance of the estimator (Gregoire & Valentine 2007). If the area of the buffer is small relative to the area of U, this increase in variance can be expected to be small. Using (3), $\lambda(\boldsymbol{u}_i)$ can be rewritten as

$$\lambda(\boldsymbol{u}_i) = \int_{C(\boldsymbol{u}_i) \cap U} \frac{\exp(\boldsymbol{\beta}^{\mathrm{T}} x(\boldsymbol{u}))}{a_{\boldsymbol{u}}} d\boldsymbol{u} = \frac{a_i}{a} \exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_i) = r_i \exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_i) = \widetilde{\lambda}_{\boldsymbol{\beta}}(\boldsymbol{u}_i),$$

where r_i is the ratio of the area a_i of $C(\boldsymbol{u}_i) \cap U$ and the area of $C(\boldsymbol{u}_i)$. With $\widetilde{\lambda}_{\boldsymbol{\beta}}(\boldsymbol{u}_i)$ defined as above, note that if $C(\boldsymbol{u}_i) \subseteq U$, then $\widetilde{\lambda}_{\boldsymbol{\beta}}(\boldsymbol{u}_i) = \lambda_{\boldsymbol{\beta}}(\boldsymbol{u}_i)$. This implies that

$$\widehat{\Lambda}(\boldsymbol{\beta}) = \sum_{i=1}^{n_2} \frac{\widetilde{\lambda}_{\boldsymbol{\beta}}(\boldsymbol{u}_i)}{\pi(\boldsymbol{u}_i)} = \sum_{i=1}^{n_2} \frac{r_i \exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_i)}{\pi(\boldsymbol{u}_i)}.$$

$\widehat{\Lambda}(\boldsymbol{\beta})$ can also be regarded as a natural predictor of the actual number of plants, given the available information and in the context of the inhomogeneous Poisson point process. As $\boldsymbol{\beta}$ is usually unknown, we will use $\widehat{\Lambda}(\widehat{\boldsymbol{\beta}})$ as our estimator of the expected number of plants, where $\widehat{\boldsymbol{\beta}}$ is an

estimator of $\boldsymbol{\beta}$ obtained using model (5) based on data from S_1.

2.3. Variance estimation

To estimate the variance of the estimator $\widehat{\Lambda}(\boldsymbol{\beta})$ of $\Lambda(\boldsymbol{\beta})$, we use the Sen-Yates-Grundy variance formula defined in Cordy (1993),

$$\mathrm{Var}(\widehat{\Lambda}(\boldsymbol{\beta})) = \frac{1}{2}\int_{U^\bullet}\int_{U^\bullet}\Delta(\boldsymbol{u},\boldsymbol{u}')\left(\frac{\lambda(\boldsymbol{u})}{\pi(\boldsymbol{u})}-\frac{\lambda(\boldsymbol{u}')}{\pi(\boldsymbol{u}')}\right)^2 d\boldsymbol{u}d\boldsymbol{u}',$$

where

$$\Delta(\boldsymbol{u},\boldsymbol{u}') = \pi(\boldsymbol{u})\pi(\boldsymbol{u}')-\pi(\boldsymbol{u},\boldsymbol{u}') \quad \text{and} \quad \pi(\boldsymbol{u},\boldsymbol{u}') = \sum_{i\in I_n}\sum_{j\in J_{n,i}} f_{ij}(\boldsymbol{u},\boldsymbol{u}'), \tag{9}$$

the latter being the pairwise inclusion density function with $I_n = \{1,\dots,n_2\}$, $J_{n,i} = \{1,\dots,n_2\}\backslash\{i\}$, and f_{ij} the joint marginal pdf of $\boldsymbol{u}_i$ and $\boldsymbol{u}_j$. As advised by, e.g., Tillé (2006), the Sen-Yates-Grundy formula should be used in case a fixed sample size is used. By Cordy (1993), if $\pi(\boldsymbol{u})$ and $\pi(\boldsymbol{u},\boldsymbol{u}')$ are strictly positive for all $(\boldsymbol{u},\boldsymbol{u}') \in U^\bullet$, an unbiased estimator of the Sen-Yates-Grundy variance is given by

$$\begin{aligned}\widehat{\mathrm{Var}}(\widehat{\Lambda}(\boldsymbol{\beta})) &= \frac{1}{2}\sum_{i\in I_n}\sum_{j\in J_{n,i}}\frac{\Delta(\boldsymbol{u}_i,\boldsymbol{u}_j)}{\pi(\boldsymbol{u}_i,\boldsymbol{u}_j)}\left(\frac{\lambda(\boldsymbol{u}_i)}{\pi(\boldsymbol{u}_i)}-\frac{\lambda(\boldsymbol{u}_j)}{\pi(\boldsymbol{u}_j)}\right)^2 \\ &= \frac{1}{2}\sum_{i\in I_n}\sum_{j\in J_{n,i}}\frac{\Delta(\boldsymbol{u}_i,\boldsymbol{u}_j)}{\pi(\boldsymbol{u}_i,\boldsymbol{u}_j)}\left(\frac{r_i\exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_i)}{\pi(\boldsymbol{u}_i)}-\frac{r_j\exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_j)}{\pi(\boldsymbol{u}_j)}\right)^2,\end{aligned} \tag{10}$$

and that is in effect the part of the variance due to sampling of the plot centres in S_2, treating the model coefficients as known. With unknown $\boldsymbol{\beta}$, i.e. where $\boldsymbol{\beta}$ needs to be estimated by $\widehat{\boldsymbol{\beta}}$, an estimate of the variance of $\widehat{\Lambda}(\widehat{\boldsymbol{\beta}})$ can be expressed as

$$\begin{aligned}\widehat{\mathrm{Var}}(\widehat{\Lambda}(\widehat{\boldsymbol{\beta}})) &= \frac{1}{2}\sum_{i\in I_n}\sum_{j\in J_{n,i}}\frac{\Delta(\boldsymbol{u}_i\boldsymbol{u}_j)}{\pi(\boldsymbol{u}_i\boldsymbol{u}_j)}\left(\frac{r_i\exp(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{x}_i)}{\pi(\boldsymbol{u}_i)}-\frac{r_j\exp(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{x}_j)}{\pi(\boldsymbol{u}_j)}\right)^2 \\ &\quad + \sum_{k=1}^{q}\sum_{l=1}^{q}\widehat{\mathrm{Cov}}_{S_1}(\widehat{\beta}_k\widehat{\beta}_l)\widehat{v}_k\widehat{v}_l,\end{aligned} \tag{11}$$

with

$$\widehat{v}_k = \sum_{i=1}^{n_2}\frac{1}{\pi(\boldsymbol{u}_i)}\widetilde{\lambda}_{\widehat{\boldsymbol{\beta}}}^{(k)}(\boldsymbol{u}_i), \tag{12}$$

where $\widehat{\beta}_k$ denotes the kth component of the $\widehat{\boldsymbol{\beta}}$ vector, and

$$\widetilde{\lambda}_{\widehat{\boldsymbol{\beta}}}^{(k)}(\boldsymbol{u}_i) = \frac{\partial\widetilde{\lambda}_{\widehat{\boldsymbol{\beta}}}(\boldsymbol{u}_i)}{\partial\widehat{\beta}_k} = r_i x_{ik}\exp(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{x}_i)$$

with x_{ik} denoting the kth component of $\boldsymbol{x}_i$. The different $\widehat{\mathrm{Cov}}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)$ terms can be obtained from statistical software, for example using the *glm* function in R. The derivation of (11) can be found in Appendix A. Another case, where S_2 is a sample of centres of plot clusters, is considered in Subsection 2.5.

2.4. Estimation of the expected plant density

In this section, we utilise our estimator of the total number of plants for estimating the expected plant density. First, we assume that the area of the region of interest is known. In this case, the expected density $R(\boldsymbol{\beta})$ is defined as the expected number of plants in the region divided by the area a_U of U,

$$R(\boldsymbol{\beta}) = \frac{\Lambda(\boldsymbol{\beta})}{a_U}, \tag{13}$$

where $\Lambda(\boldsymbol{\beta})$ is defined in (2). This quantity can be estimated by

$$\widehat{R}(\widehat{\boldsymbol{\beta}}) = \frac{\widehat{\Lambda}(\widehat{\boldsymbol{\beta}})}{a_U}, \tag{14}$$

where $\widehat{\Lambda}(\boldsymbol{\beta})$ is defined in (7). Its corresponding variance estimator is given by

$$\widehat{\mathrm{Var}}(\widehat{R}(\widehat{\boldsymbol{\beta}})) = \frac{\widehat{\mathrm{Var}}(\widehat{\Lambda}(\widehat{\boldsymbol{\beta}}))}{a_U}, \tag{15}$$

where $\widehat{\mathrm{Var}}(\widehat{\Lambda}(\widehat{\boldsymbol{\beta}}))$ is the same as in (11).

However, information about the area of the region of interest may not be available, or we may wish to estimate expected plant density in a subregion of unknown area, for example in the forested area of a region. In such cases, the area has to be estimated. Thus, $\Lambda(\boldsymbol{\beta})$ needs to be modified as

$$\Lambda^\star(\boldsymbol{\beta}) = \int_U \lambda_{\boldsymbol{\beta}}(\boldsymbol{u})I_u d\boldsymbol{u},$$

with I_u being an indicator function taking the value 1 if $\boldsymbol{u}$ is situated in the target part of the landscape and 0 otherwise; I_u is set to 0 outside of U. The area of the target part of the landscape in U can be written as

$$A = \int_U I_u d\boldsymbol{u}$$

and the expected plant density in the area of interest is given by

$$R^\star(\boldsymbol{\beta}) = \frac{\Lambda^\star(\boldsymbol{\beta})}{A}. \tag{16}$$

This quantity can be estimated by

$$\widehat{R}^\star(\boldsymbol{\beta}) = \frac{\widehat{\Lambda}^\star(\boldsymbol{\beta})}{\widehat{A}}, \tag{17}$$

where $\widehat{\Lambda}^\star(\boldsymbol{\beta})$ is defined as

$$\widehat{\Lambda}^\star(\boldsymbol{\beta}) = \sum_{i=1}^{n_2}\frac{\lambda^\star(\boldsymbol{u}_i)}{\pi(\boldsymbol{u}_i)}, \tag{18}$$

where

$$\lambda^\star(\boldsymbol{u}_i) = \int_{C(\boldsymbol{u}_i)}\frac{\lambda_{\boldsymbol{\beta}}(\boldsymbol{u})I_u}{a_u}d\boldsymbol{u},$$

and

$$\widehat{A} = \sum_{i=1}^{n_2}\frac{z(\boldsymbol{u}_i)}{\pi(\boldsymbol{u}_i)} \tag{19}$$

is an estimator of the area A, with

$$z(\boldsymbol{u}_i) = \int_{C(\boldsymbol{u}_i)}\frac{I_u}{a_u}d\boldsymbol{u}.$$

Note that, if we adopt a reasoning similar to the one in (8), $\widehat{A}$ is an unbiased estimator of A if $\pi(\boldsymbol{u}) > 0$ for all $\boldsymbol{u} \in U^\bullet$ (Cordy 1993).

In Appendix A, the following estimator of the variance of $\widehat{R}^\star(\widehat{\boldsymbol{\beta}})$ is derived:

$$\widehat{\mathrm{Var}}(\widehat{R}^{\star}(\widehat{\boldsymbol{\beta}}))=\frac{1}{2\widehat{A}^2}\sum_{i\in I_n}\sum_{j\in J_{n,i}}\frac{\Delta(\boldsymbol{u}_i,\boldsymbol{u}_j)}{\pi(\boldsymbol{u}_i,\boldsymbol{u}_j)}\left(\frac{\widehat{\lambda}^{\star}(\boldsymbol{u}_i)-\widehat{R}^{\star}(\widehat{\boldsymbol{\beta}})z(\boldsymbol{u}_i)}{\pi(\boldsymbol{u}_i)}-\frac{\widehat{\lambda}^{\star}(\boldsymbol{u}_j)-\widehat{R}^{\star}(\widehat{\boldsymbol{\beta}})z(\boldsymbol{u}_j)}{\pi(\boldsymbol{u}_j)}\right)^2$$
$$+\frac{1}{\widehat{A}^2}\sum_{k=1}^{q}\sum_{l=1}^{q}\widehat{Cov}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\sum_{i\in I_n}\frac{\widehat{d}_{1,k}(\boldsymbol{u}_i)-z(\boldsymbol{u}_i)\widehat{d}_{2,k}/\widehat{A}}{\pi(\boldsymbol{u}_i)}\sum_{j\in I_n}\frac{\widehat{d}_{1,l}(\boldsymbol{u}_j)-z(\boldsymbol{u}_j)\widehat{d}_{2,l}/\widehat{A}}{\pi(\boldsymbol{u}_j)} \tag{20}$$
$$+\frac{2}{\widehat{A}^2}\sum_{k=1}^{q}\sum_{l=1}^{q}\widehat{\mathrm{Cov}}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\widehat{d}_{2,l}\sum_{i\in I_n}\frac{\widehat{d}_{1,k}(\boldsymbol{u}_i)}{\pi(\boldsymbol{u}_i)}-\frac{1}{\widehat{A}^2}\sum_{k=1}^{q}\sum_{l=1}^{q}\widehat{\mathrm{Cov}}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\widehat{d}_{2,k}\widehat{d}_{2,l},$$

where

$$\widehat{\lambda}^{\star}(\boldsymbol{u}_i)=\int_{C(\boldsymbol{u}_i)}\frac{\lambda_{\hat{\boldsymbol{\beta}}}(\boldsymbol{u})I_{\boldsymbol{u}}}{a_{\boldsymbol{u}}}d\boldsymbol{u},$$

$$\widehat{d}_{1,k}(\boldsymbol{u}_i)=\int_{C(\boldsymbol{u}_i)}\frac{I_{\boldsymbol{u}}}{a_{\boldsymbol{u}}}\lambda_{\hat{\boldsymbol{\beta}}}^{(k)}(\boldsymbol{u})d\boldsymbol{u},\quad \widehat{d}_{2,k}=\sum_{i=1}^{n_2}\frac{I_{\boldsymbol{u}_i}\lambda_{\hat{\boldsymbol{\beta}}}^{(k)}(\boldsymbol{u}_i)}{\pi(\boldsymbol{u}_i)}, \tag{21}$$

$$\widehat{\Lambda}(\boldsymbol{\beta})=\sum_{j\in I_n}\frac{\lambda(\boldsymbol{u}_j)}{\pi(\boldsymbol{u}_j)}=\sum_{j\in I_n}\frac{1/k_j\sum_{i=1}^{k_j}r_i\exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_i^j)}{\pi(\boldsymbol{u}_j)}. \tag{23}$$

Using the same reasoning that led us to (11), we obtain the following variance estimators for $\widehat{\Lambda}(\widehat{\boldsymbol{\beta}})$;

$$\widehat{\mathrm{Var}}(\widehat{\Lambda}(\widehat{\boldsymbol{\beta}}))=\frac{1}{2}\sum_{j\in I_n}\sum_{j\in J_{n,i}}\frac{\Delta(\boldsymbol{u}_i,\boldsymbol{u}_j)}{\pi(\boldsymbol{u}_i,\boldsymbol{u}_j)}\left(\frac{1}{\pi(\boldsymbol{u}_j)k_j}\sum_{i=1}^{k_j}r_i\exp(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{x}_i^j)-\frac{1}{\pi(\boldsymbol{u}_j)k_j}\sum_{l=1}^{k_j}r_l\exp(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{x}_l^j)\right)^2+\sum_{k=1}^{p}\sum_{k'=1}^{p}\widehat{Cov}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_{k'})\widehat{v}_k\widehat{v}_{k'}, \tag{24}$$

and

$$\lambda_{\hat{\boldsymbol{\beta}}}^{(k)}(\boldsymbol{u}_i)=\frac{\partial\lambda_{\hat{\boldsymbol{\beta}}}(\boldsymbol{u}_i)}{\partial\widehat{\beta}_k}=x_{ik}\exp(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{x}_i).$$

It can happen that sample plots are divided into several parts, for example if one part of the plot is in forests and other parts are in other landscape categories. In such cases, some adjustments of the above estimators of the expected plant density and variance are needed. See Appendix B.

2.5. Cluster sampling case

It is also of interest to consider the case where S_2 is a sample of centres of clusters (sometimes called tracts) of plots rather than a sample of centres of individual plots. Indeed, this sampling procedure is used in, e.g., the Swedish NFI (Anon 2014). In this case, $C(\boldsymbol{u}_j)$ denotes a cluster j of k_j plots centred around $\boldsymbol{u}_j$, and we denote the area of the plots within the cluster by $a_{\mathbf{u}_j}=k_j s$, where s is the area of a single plot (all plots are assumed to have the same area). A buffer is also used in this case, although it will be larger (at least as large as the radius of the tract, see Grafström et al. 2017). We can still use the Horvitz-Thompson estimator (7) to get our estimator of the expected number of plants in U; the resulting expression will just be slightly different.

Using approximation (3) and if no plot is divided,

$$\lambda(\boldsymbol{u}_j)=\int_{C(\boldsymbol{u}_j)\cap U}\frac{\exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}(\boldsymbol{u}))}{a_{\boldsymbol{u}}}d\boldsymbol{u}=\frac{1}{k_j}\sum_{i=1}^{k_j}r_i\exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_i^j), \tag{22}$$

where $\mathbf{x}_i^j$ denotes the (constant) covariate information in plot i of cluster j, and r_i is the ratio of the area of the intersection of plot i n cluster j and U to the area of a single plot. Then, the Horvitz-Thompson estimator $\widehat{\Lambda}(\boldsymbol{\beta})$ may be written as

with

$$\widehat{v}_k=\sum_{j\in I_n}\frac{1}{\pi(\boldsymbol{u}_j)}\frac{1}{k_j}\sum_{i=1}^{k_j}r_i x_{ik}^j\exp(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{x}_i^j),$$

where x_{ik}^j denotes the kth component of vector $\boldsymbol{x}_i^j$. Similar changes are made in case we want to estimate expected plant density in (sub)regions with unknown area.

2.6. Statistical testing

The estimates of the expected plant density and corresponding variance estimators rely on the condition that the binary regression model (5) is realistic. For this reason, it is of importance to assess whether said model, used to estimate $\boldsymbol{\beta}$, holds true. In order to do that, we use a parametric bootstrap test suggested by Ekström et al. (Unpublished results). It should be noted that if model (5) is incorrect, then so is the underlying Poisson model assumption. Details on how to perform the test are given in Appendix C.

3. Real data study

The Swedish NFI (Fridman et al. 2014) is a field sample plot inventory of Swedish forests that consists of both temporary and permanent tracts, each composed of several plots. The temporary plots (which have a radius of 7 m) are only inventoried once, while the permanent plots are inventoried once every 5 years. Moreover, the permanent tracts are separated into two subcategories, "C_1", where both terrain and vegetation inventories are conducted, and "C_2", which denotes all other tracts. At each permanent "C_1" plot, P/A data for a set of plant species are recorded on each of two small circular "vegetation plots"; those small vegetation plots have an area of 0.25 m^2 each and are separated by 5 m and located 2.5 m from the main plot centre, the main plot having a radius of 10 m. Those registrations are not made during each visit, but rather once every two visits (i.e. every tenth year). Vegetation

registrations are not made on temporary plots. The covariates are registered at main plot level for both temporary and permanent plots. Thus, values of the covariates are always the same in each pair of small vegetation plots. The registrations are performed by experienced field workers on plots for which the positions are defined in advance according to the given sampling design.

We chose to study Lingonberry (*Vaccinium vitis-idaea*) data in the Norrbotten Lappmarken region (in northern Sweden) during the years 2008–2012. According to the Swedish NFI, region Norrbotten Lappmarken has a known area of 7,785,748 ha. The particular landscape category we chose for the estimation of $R^\star$ and its corresponding variance is productive forestland (i.e. land that can produce on average at least 1 m^3 of wood per hectare and per year and that is not significantly used for other purposes, according to Anon (2014)), whose area is unknown.

Sample S_1 consists of the centres of the small vegetation plots included in permanent "C_1" plots, in Norrbotten Lappmarken during 2008–2012. Sample S_1 has size $n_1 = 724$, corresponding to 362 pairs of vegetation plots that were used for the parametric bootstrap test. Cluster sampling was used to obtain sample S_2. It originally consists of the centres of the tracts of temporary circular plots. This sample has a size of $n_2 = 111$ tract centres, which corresponds to 1132 sample plots in total. There are one to twelve plots with available data in each (quadratic) tract, and the plots are separated by at least 600 m (Anon 2014).

In Table 1, the fitted binary regression model for *Vaccinium vitis-idaea* is presented for productive forestland in Norrbotten Lappmarken for years 2008–2012. The model was not rejected by the parametric bootstrap test (p-value = 0.184). Its explanatory variables are a transformation of the number of tree stems per hectare, multiplied by 100, and an indicator variable stating whether the soil is humid/wet. It can be seen that *Vaccinium vitis-idaea* seem less likely to be found on humid/wet soil, compared to dry soils. On the other hand, the model suggests that the more tree stems per hectare, the higher the probability of presence of *Vaccinium vitis-idaea*.

Table 2 contains estimated expected densities in two different cases. The first case is cluster sampling, where centres of clusters of plots were assumed to be sampled independently and uniformly on $U^\bullet$. In the second case, the computations were made by (incorrectly) assuming that centres of individual plots were sampled rather than centres of clusters. The densities were estimated using two different estimators (expected density estimator with known area (14) and unknown area (17), and their cluster sampling case counterparts). The corresponding variance estimates, (15) and (20) respectively (as well as their cluster sampling case counterparts), are also given. In both cases, the variance estimate of the expected density estimator in productive forestland is almost twice as high as the variance estimate using the whole region. It can be explained by the relatively small amount of plots that are situated in productive forestland in Norrbotten Lappmarken in the Swedish NFI data (approximately 50% of the total).

Table 1
Estimated model coefficients $\widehat{\boldsymbol{\beta}}$ for *Vaccinium vitis-idaea* in productive forestland in Norrbotten Lappmarken. The intercept was offset-adjusted. $\mathbf{1}_{wet}$ is an indicator variable stipulating whether a plot is humid/wet or not. $((\text{No.stems/ha} + 0.6)/1000)^{-0.5}$ is a non-linear transformation of the "number of tree stems per hectare" (in hundreds per hectare) covariate, found by using the mfp R package (Ambler & Benner 2015), which applies multivariable fractional polynomials (Sauerbrei & Royston 1999).

Species	Estimated parameters ($\widehat{\boldsymbol{\beta}}$)	
Vaccinium vitis-idaea (Lingonberry)	Offset-adjusted Intercept	2.423
	$\mathbf{1}_{wet}$	−0.667
	$((\text{No.stems/ha} + 0.6)/1000)^{-0.5}$	−0.025

Table 2
Estimated expected plant densities in m^{-2} and corresponding estimates of variance for *Vaccinium vitis-idaea* in Norrbotten Lappmarken. Two cases were considered: one where the computations were made assuming cluster sampling and another where it was (incorrectly) assumed that single plots were sampled. $\widehat{R}(\widehat{\boldsymbol{\beta}})$ and $\widehat{\text{Var}}(\widehat{R}(\widehat{\boldsymbol{\beta}}))$ are computed for the whole Norrbotten Lappmarken region, while $\widehat{R}^\star(\widehat{\boldsymbol{\beta}})$ and $\widehat{\text{Var}}(\widehat{R}^\star(\widehat{\boldsymbol{\beta}}))$ are computed for the productive forestland area of Norrbotten Lappmarken only.

Case	$\widehat{R}(\widehat{\boldsymbol{\beta}})$	$\widehat{R}^\star(\widehat{\boldsymbol{\beta}})$	$\widehat{\text{Var}}(\widehat{R}(\widehat{\boldsymbol{\beta}}))$	$\widehat{\text{Var}}(\widehat{R}^\star(\widehat{\boldsymbol{\beta}}))$
Tracts	7.61	9.72	0.205	0.406
Single plots	7.49	9.73	0.209	0.411

4. Monte Carlo study

The aim of the Monte Carlo study was to evaluate our estimators of expected plant density and variance estimators and assess whether they performed well. The simulations, all performed in **R** (R Core Team 2022), were conducted as follows.

- We created a quadratic grid of 1024 cells that corresponds to our area frame U, as well as a buffer zone around U. Each grid cell had an area of 1 ha and artificial covariates.
- The created covariates were based on the ones included in the model for *Vaccinium vitis-idaea*. The indicator variable stipulating whether a plot is humid/wet or not was built on actual data in the Norrbotten Lappmarken region between 2008 and 2012, which had approximately 16.85% of plots being considered as humid/wet. This particular covariate was created as realisations of a Bernoulli distribution with parameter $p = 0.1685$ in each cell. As for the number of stems per hectare, we used fitted Weibull distributions as described below. Two cases were considered:
  1. In the first case, we assumed that the whole grid was productive forestland, and the area of the area frame (the cell grid) was assumed to be known. In that case, we supposed that the number of stems per hectare varied only depending on whether the soil was humid/wet or dry. Based on Swedish NFI data in productive forestland, Weibull distributions were fitted using the *fitdist* function from the fitdistrplus package (Delignette-Muller & Dutang 2015). On humid/wet grid cells, the fitted distribution was a Weibull distribution with shape parameter $k = 1.047$ and scale parameter $\lambda = 3898.3$. For the dry grid cells, a two-step procedure was used since 4% of the original data had values equal to 0. Therefore, a random number between 0 and 1 was generated for each grid cell; if this number was smaller than 0.04, the number of stems per hectare for that grid cell was set to 0; otherwise it was a realisation of a Weibull-distributed random variable with parameters $k = 0.903$ and $\lambda = 2076.5$.
  2. In the second case, we created an indicator variable which was assigned the value 1 if the cell was in productive forestland, and 0 otherwise. As 49.8% of the original sample plots are in productive forestland, each cell was assigned the value 1 with a probability of 0.498. The number of stems per hectare was supposed to vary according to both humidity of the soil and type of landscape (productive forestland or not), which means that four different subcases had to be considered. The area of productive forestland in the grid was estimated by (19). The covariates were generated exclusively for the cells that are situated in productive forestland (which means in two of the subscases), and in such case were generated exactly as in case 1.
- Each Monte Carlo simulation consisted of 2000 replicates; P/A data were generated from an inhomogeneous Poisson point process with the *rpoispp* function from the spatstat package (Baddeley et al. 2016) in each replicate; plot centres in S_2 were sampled independently according to a uniform distribution over $U^\bullet$, while a two-step generation procedure was used for S_1: first, plot centres for the

Table 3
Actual expected plant densities $R(\boldsymbol{\beta})$ (resp. $R^{\star}(\boldsymbol{\beta})$), estimated mean values of the estimated expected densities $\widehat{\mathbb{E}}(\widehat{R}(\widehat{\boldsymbol{\beta}}))$ (resp. $\widehat{\mathbb{E}}(\widehat{R}^{\star}(\widehat{\boldsymbol{\beta}}))$), estimated mean value of the variance estimates $\widehat{\mathbb{E}}(\widehat{\mathrm{Var}}(\widehat{R}(\widehat{\boldsymbol{\beta}})))$ (resp. $\widehat{\mathbb{E}}(\widehat{\mathrm{Var}}(\widehat{R}^{\star}(\widehat{\boldsymbol{\beta}})))$) and s^2, the sample variance of the $\widehat{R}(\widehat{\boldsymbol{\beta}})$ (resp. $\widehat{R}^{\star}(\widehat{\boldsymbol{\beta}})$), for simulated *Vaccinium vitis-idaea* data in a grid of 1024 cells, each cell having an area of 1 ha. In the known area case, the area is a_U, the area of the grid. In the unknown area case, the area is estimated according to (19). The variances were estimated using formulas (15) and (20). "/" means that the formula does not apply to the specific case.

Case	$R(\boldsymbol{\beta})$	$R^{\star}(\boldsymbol{\beta})$	$\widehat{\mathbb{E}}(\widehat{R}(\widehat{\boldsymbol{\beta}}))$	$\widehat{\mathbb{E}}(\widehat{R}^{\star}(\widehat{\boldsymbol{\beta}}))$	$\widehat{\mathbb{E}}(\widehat{\mathrm{Var}}(\widehat{R}(\widehat{\boldsymbol{\beta}})))$	$\widehat{\mathbb{E}}(\widehat{\mathrm{Var}}(\widehat{R}^{\star}(\widehat{\boldsymbol{\beta}})))$	s^2
Known area	9.740	/	9.606	/	0.191	/	0.196
Unknown area	/	9.715	/	9.657	/	0.187	0.189

permanent plots were sampled independently according to a uniform distribution over U, and then the small vegetation plots in S_1 were created for each permanent plot as described in Section 3. The value of the vector of coefficients $\boldsymbol{\beta}$ was set equal to the one from the fitted model for *Vaccinium vitis-idaea* in Norrbotten Lappmarken in years 2008–2012 (Table 1). Estimated model coefficients $\widehat{\boldsymbol{\beta}}$ were computed for every replicate using the S_1 data, while the estimated expected plant density and its corresponding variance estimate were computed for every replicate using the S_2 data. The sample sizes were $n_1 = 1500$ and $n_2 = 1500$. The same plot radii as in the Swedish NFI were used (see Section 3). The plots in S_2 were divided when they overlapped different grid cells (see details in Appendix B). In accordance with the Swedish NFI (Jonas Dahlgren, personal communication), the small vegetation plots within S_1 were not divided.

The results for the simulation study are presented in Table 3. The estimator $\widehat{R}(\widehat{\boldsymbol{\beta}})$ was used for Case 1 and $\widehat{R}^{\star}(\widehat{\boldsymbol{\beta}})$ was used for Case 2. In Case 1, the estimator $\widehat{R}(\widehat{\boldsymbol{\beta}})$ was on average close to but a little lower than the real expected plant density. In Case 2, the estimator $\widehat{R}^{\star}(\widehat{\boldsymbol{\beta}})$ was even closer to the true value, but even in that case a slight negative bias occurred. The two variance estimators seem to have a very small bias and have low values. Based on these observations, we can conclude that our estimators performed quite well.

5. Discussion

In this study, we show how P/A data can be used for modelling and monitoring plant population densities. We argue that this approach offers advantages over methods based on visual assessment of vegetation cover, since studies indicate that P/A sampling may not be as prone to observer bias as methods based on assessing vegetation cover, and since P/A sampling is a rapid and thus cheap method to apply (e.g., Ringvall et al. 2005).

Since the auxiliary modelling data are available for both considered samples, but the binary response data are available for only one sample, we apply methods from hybrid inference (e.g., Corona et al. 2014) for estimating the expected value of plant density and the corresponding variance. This concerns taking into account both modelling and sampling uncertainty, and to our knowledge, our study is the first one that involves GLMs in hybrid inference. This type of inference is important in this context since, in many cases, detailed descriptions of environmental conditions, needed for the modelling, may not be available wall-to-wall but only from sampling locations, e.g., from sample plots within environmental monitoring programmes. In this article, we extend the already existing theory on hybrid inference to GLMs with binary response data.

Our method is most suitable when n_2, the sample size of S_2, is much larger than n_1, the sample size of S_1. Indeed, the main purpose in applying this method is to gather a minimum of information to develop a reliable model on the smallest sample possible (principally due to budgetary reasons), to then apply this model in connection with covariates that come from a larger sample whose units do not contain the desired response data. However, with our available data, n_2 was only a little larger than n_1. This shows that our method works even in that particular case.

In regions with high perimeter-to-area ratios, a large or very large proportion of the sampling plots will extend beyond the region's boundary. In such cases, our suggested methodology, which uses a "buffer" to address edge effects, may be unsuitable and could result, for example, in estimators with larger variances than desired.

An important part of the study involves making the proposed hybrid inference framework available for practical application in monitoring programmes, in which case we need to take into account that sample plots are often allocated in clusters and that the area of the domain of study is unknown (e.g., Fridman et al. 2014). This introduces several additional details to the general framework, which are important for the usefulness of the framework in practice.

The Monte Carlo simulations we performed show that our framework for estimating the expected plant density provides accurate estimates when the modelling assumptions are valid. In the study based on empirical data from the Swedish NFI, we obtained estimates of expected Lingonberry (*Vaccinium vitis-idaea*) densities in Northern Sweden that appear to be realistic, although we cannot check them since no reference data are available.

For the sake of simplicity, we assumed that the sampling design of S_1 was non-informative (see Appendix A), i.e. the design was not taken into account during model parameter estimation. Ignoring an informative sampling design may yield biased estimates of regression coefficients. For handling informative designs, methods using probability weighting may be used (e.g., Heeringa et al. 2010; Ekström et al. 2018).

It is possible to generalise the considered hybrid inference framework to other types of GLMs. Instead of P/A data as a response variable, one could use a continuous variable (such as biomass) or a discrete variable such as a count variable (number of trees, birds etc.). The main requirement is to have two samples; one to estimate model coefficients, with both covariate and response data, and another one, with only covariate data, for estimation of, e.g., expected biomass per hectare or expected plant density based on the estimated model coefficients. As long as this requirement is met, then hybrid inference should work, in principle, with any kind of response variable. The statistical developments would, however, be different from the ones derived in the present paper; although with counts instead of P/A, the difference would not be that significant (in both cases, it would be possible to use an inhomogeneous Poisson model). With count data that are not subject to too many errors, it should be possible to obtain better estimators than the ones obtained from P/A data. However, the survey would be more expensive to conduct.

There is one key condition for the developed technique to be applicable; the underlying point process should be, at least approximately, an inhomogeneous Poisson point process. We estimate models that utilise a combination of P/A and auxiliary data to estimate expected plant density, assuming that the spatial distribution of plants follow an inhomogeneous Poisson process, i.e. the plant densities vary due to the environmental conditions. In the article, we check the suitability of the binary regression model implied by the underlying inhomogeneous Poisson point process through a statistical test specifically developed for the purpose (cf. Appendix C). Recognising that plants can occur in clustered spatial patterns, extensions from inhomogeneous Poisson point processes to inhomogeneous cluster point processes serve as an

important topic for further studies. However, if we would like to use a similar methodology as in Ekström et al. (2020), we would need to gather data on more than two subplots for each main plot.

In our paper, the intensity of the inhomogeneous Poisson point process is determined via a log-linear model that involves a number of covariates. This model cannot be fitted directly, since no observed point pattern or observed values of counts of points in plots are available. This problem is circumvented by making use of observable P/A variables. Given that the pattern is a realisation of an inhomogeneous Poisson point process (whose intensity on the ith cell is given by (1)), it follows that the P/A variables satisfy a binary GLM, with complementary log-log link and an offset, with the same parameter vector as that which appears in the intensity of the inhomogeneous Poisson point process. Thus, for extending the current approach to other inhomogeneous point processes than the Poisson, the parameters of their intensities must be estimable from P/A data and corresponding covariate data at plot level. In addition, estimates of covariance matrices of estimators of parameters are also needed. One possibility to achieve this is to extend the intensity estimator in Ekström et al. (2020) from homogeneous cluster point processes such as the Matérn and Thomas processes to corresponding heterogeneous processes, whose intensities are functions of on one or more covariates (Waagepetersen 2007).

When the point pattern is generated by an inhomogeneous Poisson point process, the binary GLM model in (5) will have independent binary (P/A) response variables conditional on the covariates. For other point processes, responses cannot be expected to fulfill this property of conditional independence. Then, instead of using a standard GLM, other estimation methods such as generalised estimating equations (Albert & McShane 1995; Gotway & Stroup 1997) and a composite likelihood approach for spatial binary data (Heagerty & Lele 1998) can be used. However, as mentioned, this is not enough for extending the current approach to more general point processes. Most importantly, the estimable unknown parameters in the regression model for the P/A data must also include all unknown parameters in the intensity function of the point process model.

For a Poisson point process with a homogeneous intensity λ, the species abundance N in a plot C of area a follows a Poisson distribution with mean $a\lambda$, and the probability of presence of at least one plant in the plot C equals $p = 1 - \exp(-a\lambda)$. Rearranging this equation, we can estimate the intensity (plant density) λ from the proportion $\widehat{p}$ of plots with plant occurrences, i.e., by $\widehat{\lambda} = -a^{-1}\log(1-\widehat{p})$ (e.g., Ståhl et al. 2017). A homogeneous spatial Poisson process is synonymous with complete spatial randomness. However, in nature, individuals of many species are typically aggregated (Pielou 1977; He & Gaston 2000). For plot abundance N, the model most commonly used to describe such aggregation is the negative binomial distribution (He et al. 2002), which implies the following relationship between the presence probability p and plant density λ, $p = 1 - \left(1 + k^{-1}\lambda\right)^{-k}$, where k is referred to as a "clumping" parameter, with small $k > 0$ representing strong aggregation (Wright 1991; He & Gaston 2000; He et al. 2002). Under this model, Conlisk et al. (2007) specify the likelihood function and conclude that the clumping parameter cannot be estimated from P/A data, i.e., that it must be specified from outside the model. The suitability of the negative binomial distribution has also been much debated (Holt et al. 2002; Gaston et al. 2011) and only two known homogeneous point processes give the negative binomial distribution for plot abundances, and both are extreme cases (Daley & Vere-Jones 2003). For some further developments of the negative binomial distribution model, we refer to Solow & Smith (2010), Hwang & Huggins (2016), Huggins et al. (2018), Hwang et al. (2022), and Stoklosa et al. (2022). For other suggested models than those based on the Poisson and the negative binomial distributions for describing the relationship between the presence probability p and plant density λ, see, e.g., Holt et al. (2002), He et al. (2002), and the references therein. Extensions of the negative binomial model and other related models to an inhomogeneous setting would be useful for extending the approach presented in the current article to more general settings.

Many monitoring and citizen science programmes already have large amounts of P/A data in their databases (e.g., the Norwegian Biodiversity Information Center in Norway (Hoem 2022); the Global Biodiversity Information Facility GBIF (GBIF 2022)). Therefore, the techniques and estimators developed in the present study can be applied to already available data, especially since new fine-scaled covariate data are becoming increasingly common in such databases. Promising results were obtained in this study, which means that the proposed framework for monitoring plant population density through P/A sampling and modelling holds promise for future practical application, e.g., in national reporting of trends in declining species.

Author contributions

LG wrote the main draft, performed the analyses and simulations and contributed to the theoretical developments; ME conceived the idea, contributed to the theoretical developments and contributed critically to the drafts; SS, BGJ, JW and GS contributed critically to the drafts.

Funding

This work was supported by Kempestiftelserna (SMK-1955).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data will be made available upon request.

Acknowledgements

We thank Jonas Dahlgren for having provided the data used in Section 3.

Appendix A. Theoretical developments in the case of single plots

A.1. Case with known area

For simplicity, we assume that the sampling design of S_1 is non-informative, i.e. the vector of model parameters is estimated without taking this sampling design into account. Under this assumption, for large samples and under mild conditions (see for example Sen & Singer 1993),

$$\sqrt{n_2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim \mathcal{N}\left(0, I^{-1}(\boldsymbol{\beta})\right), \tag{A.1}$$

where $I(\boldsymbol{\beta})$ denotes the Fisher information matrix and can be estimated by

$$\widehat{I}(\widehat{\boldsymbol{\beta}}) = \frac{1}{n_2}\sum_{i\in I_n}\frac{1}{[g'(p_i(\widehat{\boldsymbol{\beta}}))]^2 v_i(\widehat{\boldsymbol{\beta}})}\ \boldsymbol{x}_i\boldsymbol{x}_i^{'}, \tag{A.2}$$

with $\widehat{\boldsymbol{\beta}}$ being the estimate of $\boldsymbol{\beta}$, g defined by (5), p_i defined by (4), and $v_i(\boldsymbol{\beta}) = \mathrm{Var}(Y_i) = p_i(1-p_i)$, where $Y_i = 1$ if there is presence of plants in plot i, and $Y_i = 0$ otherwise.

Using a similar reasoning as in [Ståhl et al. 2011], we start with the decomposition

$$\widehat{\Lambda}(\widehat{\boldsymbol{\beta}}) - \Lambda(\boldsymbol{\beta}) = \sum_{i\in I_n}\frac{\widetilde{\lambda}_{\widehat{\beta}}(\boldsymbol{u}_i)}{\pi(\boldsymbol{u}_i)} - \Lambda = D_1 + D_2, \tag{A.3}$$

where

$$D_1 = \sum_{i\in I_n}\frac{\widetilde{\lambda}_{\boldsymbol{\beta}}(\boldsymbol{u}_i)}{\pi(\boldsymbol{u}_i)} - \Lambda \quad \text{and} \quad D_2 = \sum_{i\in I_n}\frac{\widetilde{\lambda}_{\widehat{\beta}}(\boldsymbol{u}_i) - \widetilde{\lambda}_{\boldsymbol{\beta}}(\boldsymbol{u}_i)}{\pi(\boldsymbol{u}_i)}.$$

Our objective is to compute the variance

$$\mathrm{Var}(D_1 + D_2) = \mathrm{Var}(D_1) + \mathrm{Var}(D_2) + 2\ \mathrm{Cov}(D_1, D_2).$$

Using the Sen-Yates-Grundy formula presented in Cordy (1993), an unbiased estimator of $\mathrm{Var}(D_1)$ is given by (10). If $\boldsymbol{\beta}$ is unknown, we estimate this variance with

$$\widehat{\mathrm{Var}}(D_1) = \frac{1}{2}\sum_{i\in I_n}\sum_{j\in J_{n,i}}\frac{\Delta(\boldsymbol{u}_i,\boldsymbol{u}_j)}{\pi(\boldsymbol{u}_i,\boldsymbol{u}_j)}\left(\frac{r_i\exp(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{x}_i)}{\pi(\boldsymbol{u}_i)} - \frac{r_j\exp(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{x}_j)}{\pi(\boldsymbol{u}_j)}\right)^2. \tag{A.4}$$

The law of total variance is used in order to compute $\mathrm{Var}(D_2)$, i.e

$$\mathrm{Var}(D_2) = \mathrm{Var}_{S_2}[\mathbb{E}_{S_1}(D_2|S_2)] + \mathbb{E}_{S_2}[\mathrm{Var}_{S_1}(D_2|S_2)]. \tag{A.5}$$

For non-linear models, a Taylor approximation can be applied, i.e.

$$\widetilde{\lambda}_{\widehat{\beta}}(\boldsymbol{u}) \approx \widetilde{\lambda}_{\boldsymbol{\beta}}(\boldsymbol{u}) + \sum_{k=1}^{q}(\widehat{\beta}_k - \beta_k)\widetilde{\lambda}_{\boldsymbol{\beta}}^{(k)}(\boldsymbol{u}), \tag{A.6}$$

where

$$\widetilde{\lambda}_{\boldsymbol{\beta}}^{(k)}(\boldsymbol{u}_i) = r_i x_{ik}\exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_i).$$

Then,

$$D_2 \approx \sum_{i\in I_n}\sum_{k=1}^{q}\frac{(\widehat{\beta}_k - \beta_k)}{\pi(\boldsymbol{u}_i)}\ \widetilde{\lambda}_{\boldsymbol{\beta}}^{(k)}(\boldsymbol{u}_i) = \sum_{k=1}^{q}(\widehat{\beta}_k - \beta_k)v_k,$$

where

$$v_k = \sum_{i\in I_n}\frac{1}{\pi(\boldsymbol{u}_i)}\widetilde{\lambda}_{\boldsymbol{\beta}}^{(k)}(\boldsymbol{u}_i)$$

and q being the number of model coefficients. Conditioned on S_2, v_k is a constant. Then, by (A.1), $\mathbb{E}_{S_1}(D_2|S_2) \approx \sum_{k=1}^{q}\mathbb{E}_{S_1}(\widehat{\beta}_k - \beta_k|S_2)v_k \approx 0$ for large samples, and thus $Var_{S_2}[\mathbb{E}_{S_1}(D_2|S_2)] \approx 0$. Furthermore,

$$
\begin{aligned}
\mathrm{Var}_{S_1}(D_2|S_2) &\approx \mathrm{Var}_{S_1}\left(\sum_{k=1}^{q}(\widehat{\beta}_k-\beta_k)v_k\,|S_2\right)\\
&\approx \sum_{k=1}^{q}\sum_{l=1}^{q}\mathrm{Cov}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)v_k v_l\\
&= \sum_{i\in I_n}\sum_{j\in I_n}\frac{1}{\pi(\boldsymbol{u}_i)\pi(\boldsymbol{u}_j)}\sum_{k=1}^{q}\sum_{l=1}^{q}\mathrm{Cov}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)r_i\,r_j\,x_{ik}\,x_{jl}\exp\left(\boldsymbol{\beta}^{\mathrm{T}}(\boldsymbol{x}_i+\boldsymbol{x}_j)\right)\\
&= \sum_{k=1}^{q}\sum_{l=1}^{q}\mathrm{Cov}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\sum_{i\in I_n}\frac{r_i^2}{\pi(\boldsymbol{u}_i)^2}x_{ik}x_{il}\exp\left(2\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_i\right)\\
&+\sum_{k=1}^{q}\sum_{l=1}^{q}\mathrm{Cov}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\sum_{i\in I_n}\sum_{j\in J_{n,i}}\frac{r_i r_j}{\pi(\boldsymbol{u}_i)\pi(\boldsymbol{u}_j)}x_{ik}\,x_{jl}\exp\left(\boldsymbol{\beta}^{\mathrm{T}}(\boldsymbol{x}_i+\boldsymbol{x}_j)\right).
\end{aligned}
$$

From the arguments in the proof of Theorem 2 in Cordy (1993), we get

$$
\begin{aligned}
\mathrm{Var}(D_2) &\approx \mathbb{E}_{S_2}[\mathrm{Var}_{S_1}(D_2|S_2)]\\
&= \sum_{k=1}^{q}\sum_{l=1}^{q}\mathrm{Cov}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\int_{U^\star}\frac{r_{\boldsymbol{u}}^2}{\pi(\boldsymbol{u})}\boldsymbol{x}^k(\boldsymbol{u})\boldsymbol{x}^l(\boldsymbol{u})\exp\left(2\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}(\boldsymbol{u})\right)d\boldsymbol{u}\\
&+\sum_{k=1}^{q}\sum_{l=1}^{q}\mathrm{Cov}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\int_{U^\star}\int_{U^\star}\frac{\pi(\boldsymbol{u},\boldsymbol{u}')}{\pi(\boldsymbol{u})\pi(\boldsymbol{u}')}r_{\boldsymbol{u}}r_{\boldsymbol{u}}\boldsymbol{x}^k(\boldsymbol{u})\boldsymbol{x}^l(\boldsymbol{u})\exp\left(\boldsymbol{\beta}^{\mathrm{T}}(\boldsymbol{x}(\boldsymbol{u})+\boldsymbol{x}(\boldsymbol{u}'))\right)d\boldsymbol{u}d\boldsymbol{u}',
\end{aligned}
$$

where $\boldsymbol{x}^k(\boldsymbol{u})$ denotes the kth component of the $\boldsymbol{x}$ vector and $r_{\boldsymbol{u}}$ is the ratio of the area of $C(\boldsymbol{u})\cap U$ and the area of $C(\boldsymbol{u})$. Thus, $Var(D_2)$ can be estimated by

$$
\begin{aligned}
\widehat{\mathrm{Var}}(D_2) &= \sum_{k=1}^{q}\sum_{l=1}^{q}\widehat{\mathrm{Cov}}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\sum_{i\in I_n}\frac{r_i^2}{\pi(\boldsymbol{u}_i)^2}\,x_{ik}x_{il}\exp\left(2\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{x}_i\right)\\
&+\sum_{k=1}^{q}\sum_{l=1}^{q}\widehat{\mathrm{Cov}}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\sum_{i\in I_n}\sum_{j\in J_{n,i}}\frac{r_i r_j}{\pi(\boldsymbol{u}_i)\pi(\boldsymbol{u}_j)}\,x_{ik}x_{jl}\exp\left(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}(\boldsymbol{x}_i+\boldsymbol{x}_j)\right)\\
&= \sum_{k=1}^{q}\sum_{l=1}^{q}\widehat{\mathrm{Cov}}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\sum_{i\in I_n}\sum_{j\in I_n}\frac{r_i r_j}{\pi(\boldsymbol{u}_i)\pi(\boldsymbol{u}_j)}\,x_{ik}x_{jl}\exp\left(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}(\boldsymbol{x}_i+\boldsymbol{x}_j)\right)\\
&= \sum_{k=1}^{q}\sum_{l=1}^{q}\widehat{\mathrm{Cov}}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\widehat{v}_k\widehat{v}_l,
\end{aligned}
\tag{A.7}
$$

where $\widehat{v}_k$ is defined in (12).

The next step is to compute the covariance between D_1 and D_2. According to the law of total covariance,

$$
\mathrm{Cov}(D_1,D_2)=\mathbb{E}_{S_2}[\mathrm{Cov}_{S_1}(D_1,D_2|S_2)]+\mathrm{Cov}_{S_2}[\mathbb{E}_{S_1}(D_1|S_2),\mathbb{E}_{S_1}(D_2|S_2)]. \tag{A.8}
$$

It can be deduced that $\mathrm{Cov}_{S_2}[\mathbb{E}_{S_1}(D_1|S_2),\mathbb{E}_{S_1}(D_2|S_2)]\approx 0$ because, as argued before, $\mathbb{E}_{S_1}(D_2|S_2)\approx 0$. Then, as the stochastic nature of D_1 is determined by sample S_2 and not by sample S_1, $\mathbb{E}_{S_1}(D_1D_2|S_2)=D_1\mathbb{E}_{S_1}(D_2|S_2)\approx 0$. Because of the latter, $\mathbb{E}_{S_2}[\mathrm{Cov}_{S_1}(D_1,D_2|S_2)]\approx 0$. Thus, $\mathrm{Cov}(D_1,D_2)\approx 0$ and we just need to add the variances of D_1 and D_2 to get an approximate variance of D_1+D_2. As a result, setting (A.4) and (A.7) together, the estimate becomes

$$
\begin{aligned}
\widehat{\mathrm{Var}}(\widehat{\Lambda}(\widehat{\boldsymbol{\beta}})) &= \widehat{\mathrm{Var}}(D_1)+\widehat{\mathrm{Var}}(D_2)+2\widehat{\mathrm{Cov}}(D_1,D_2)\\
&= \frac{1}{2}\sum_{i\in I_n}\sum_{j\in J_{n,i}}\frac{\Delta(\boldsymbol{u}_i,\boldsymbol{u}_j)}{\pi(\boldsymbol{u}_i,\boldsymbol{u}_j)}\left(\frac{r_i\exp(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{x}_i)}{\pi(\boldsymbol{u}_i)}-\frac{r_j\exp(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{x}_j)}{\pi(\boldsymbol{u}_j)}\right)^2+\sum_{k=1}^{q}\sum_{l=1}^{q}\widehat{\mathrm{Cov}}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\widehat{v}_k\widehat{v}_l,
\end{aligned}
$$

with $\widehat{v}_k$ defined in (12).

A.2. Expected density estimator in a specific area of the landscape

Suppose we want to estimate the number of plants exclusively in a certain landscape category, for example forests. Then, the parameter vector $\boldsymbol{\beta}$ will be estimated only from the plots that are situated in this landscape category.

As in Result 5.6.2 in Särndal et al. (1992), for estimating the variance of $\widehat{R}^\star(\boldsymbol{\beta})$ we use a Taylor linearisation by introducing $\widehat{R}_0^\star(\boldsymbol{\beta})$, that is related to $\widehat{R}^\star(\boldsymbol{\beta})$ by the relation

$$
\widehat{R}^\star(\boldsymbol{\beta})\approx\widehat{R}_0^\star(\boldsymbol{\beta})=R^\star(\boldsymbol{\beta})+\frac{1}{A}\sum_{i\in I_n}\frac{\lambda^\star(\boldsymbol{u}_i)-R^\star(\boldsymbol{\beta})z(\boldsymbol{u}_i)}{\pi(\boldsymbol{u}_i)}. \tag{A.9}
$$

Remember that the estimator of $\boldsymbol{\beta}$, $\widehat{\boldsymbol{\beta}}$, is approximately normally distributed with mean $\boldsymbol{\beta}$ (see (A.1)). We estimate $R^{\star}(\boldsymbol{\beta})$ with $\widehat{R}^{\star}(\widehat{\boldsymbol{\beta}})$. The goal here is to derive an estimate of the variance of $\widehat{R}^{\star}(\widehat{\boldsymbol{\beta}})$, or equivalently the variance of $\widehat{R}^{\star}(\widehat{\boldsymbol{\beta}}) - R^{\star}(\boldsymbol{\beta})$, which by the arguments in the proof of Result 5.6.2 in Särndal et al. (1992) is approximately the same as the one for

$$D(\widehat{\boldsymbol{\beta}}) = \widehat{R}_0^{\star}(\widehat{\boldsymbol{\beta}}) - R^{\star}(\widehat{\boldsymbol{\beta}}) = \frac{1}{A}\sum_{i\in I_n}\frac{\widehat{\lambda}^{\star}(\boldsymbol{u}_i) - R^{\star}(\widehat{\boldsymbol{\beta}})z(\boldsymbol{u}_i)}{\pi(\boldsymbol{u}_i)},$$

where

$$\widehat{\lambda}^{\star}(\boldsymbol{u}) = \int_{C(\mathbf{u})}\frac{\lambda_{\widehat{\boldsymbol{\beta}}}(\boldsymbol{u}')I_{\boldsymbol{u}'}}{a_{\boldsymbol{u}'}}d\boldsymbol{u}', \boldsymbol{u}\in U^{\bullet}. \tag{A.10}$$

We can write

$$\widehat{R}_0^{\star}(\widehat{\boldsymbol{\beta}}) - R^{\star}(\boldsymbol{\beta}) = \left(\widehat{R}_0^{\star}(\widehat{\boldsymbol{\beta}}) - R^{\star}(\widehat{\boldsymbol{\beta}})\right) + \left(R^{\star}(\widehat{\boldsymbol{\beta}}) - R^{\star}(\boldsymbol{\beta})\right) = D(\widehat{\boldsymbol{\beta}}) + D_*(\widehat{\boldsymbol{\beta}}),$$

where $D_*(\widehat{\boldsymbol{\beta}}) = R^{\star}(\widehat{\boldsymbol{\beta}}) - R^{\star}(\boldsymbol{\beta})$. By the following Taylor approximation

$$\lambda_{\widehat{\boldsymbol{\beta}}}(\boldsymbol{u}) \approx \lambda_{\boldsymbol{\beta}}(\boldsymbol{u}) + \sum_{k=1}^{q}(\widehat{\beta}_k - \beta_k)\lambda_{\boldsymbol{\beta}}^{(k)}(\boldsymbol{u}), \tag{A.11}$$

where

$$\lambda_{\boldsymbol{\beta}}^{(k)}(\boldsymbol{u}) = \frac{\partial\lambda_{\boldsymbol{\beta}}(\boldsymbol{u})}{\partial\beta_k},$$

we obtain

$$\begin{aligned}\mathbb{E}[R^{\star}(\widehat{\boldsymbol{\beta}})] &= \mathbb{E}_{S_1}[R^{\star}(\widehat{\boldsymbol{\beta}})] = \frac{1}{A}\mathbb{E}_{S_1}[\Lambda^{\star}(\widehat{\boldsymbol{\beta}})] = \frac{1}{A}\int_U \mathbb{E}_{S_1}[\lambda_{\widehat{\boldsymbol{\beta}}}(\boldsymbol{u})]I_{\boldsymbol{u}}d\boldsymbol{u} \\ &\approx \frac{1}{A}\int_U \lambda_{\boldsymbol{\beta}}(\boldsymbol{u})I_{\boldsymbol{u}}d\boldsymbol{u} + \frac{1}{A}\sum_{k=1}^{q}\mathbb{E}_{S_1}[\widehat{\beta}_k - \beta_k]\int_U \lambda_{\boldsymbol{\beta}}^{(k)}(\boldsymbol{u})I_{\boldsymbol{u}}d\boldsymbol{u} \approx \frac{1}{A}\int_U \lambda_{\boldsymbol{\beta}}(\boldsymbol{u})I_{\boldsymbol{u}}d\boldsymbol{u} = R^{\star}(\boldsymbol{\beta})\end{aligned} \tag{A.12}$$

and

$$\begin{aligned}\mathbb{E}\left[(R^{\star}(\widehat{\boldsymbol{\beta}}))^2\right] &= \mathbb{E}_{S_1}\left[(R^{\star}(\widehat{\boldsymbol{\beta}}))^2\right] = \frac{1}{A^2}\mathbb{E}_{S_1}\left[(\Lambda^{\star}(\widehat{\boldsymbol{\beta}}))^2\right] \\ &= \frac{1}{A^2}\int_U\int_U \mathbb{E}_{S_1}[\lambda_{\widehat{\boldsymbol{\beta}}}(\boldsymbol{u})\lambda_{\widehat{\boldsymbol{\beta}}}(\boldsymbol{u}')]I_{\boldsymbol{u}}I_{\boldsymbol{u}'}d\boldsymbol{u}d\boldsymbol{u}' \\ &\approx \frac{1}{A^2}\int_U\int_U \lambda_{\boldsymbol{\beta}}(\boldsymbol{u})\lambda_{\boldsymbol{\beta}}(\boldsymbol{u}')I_{\boldsymbol{u}}I_{\boldsymbol{u}'}d\boldsymbol{u}d\boldsymbol{u}' + \frac{1}{A^2}\sum_{k=1}^{q}\sum_{l=1}^{q}\mathrm{Cov}_{S_1}(\widehat{\beta}_k, \widehat{\beta}_l)d_{2,k}d_{2,l} \\ &= (R^{\star}(\boldsymbol{\beta}))^2 + \frac{1}{A^2}\sum_{k=1}^{q}\sum_{l=1}^{q}\mathrm{Cov}_{S_1}(\widehat{\beta}_k, \widehat{\beta}_l)d_{2,k}d_{2,l},\end{aligned}$$

where

$$d_{2,k} = \int_U I_{\boldsymbol{u}}\lambda_{\boldsymbol{\beta}}^{(k)}(\boldsymbol{u})d\boldsymbol{u} = \frac{\partial\Lambda^{\star}(\boldsymbol{\beta})}{\partial\beta_k}.$$

Thus,

$$\mathbb{E}[D_*(\widehat{\boldsymbol{\beta}})] = \mathbb{E}_{S_1}[D_*(\widehat{\boldsymbol{\beta}})] \approx 0 \tag{A.13}$$

and

$$\mathrm{Var}(D_*(\widehat{\boldsymbol{\beta}})) = \mathrm{Var}_{S_1}(D_*(\widehat{\boldsymbol{\beta}})) \approx \frac{1}{A^2}\sum_{k=1}^{q}\sum_{l=1}^{q}\mathrm{Cov}_{S_1}(\widehat{\beta}_k, \widehat{\beta}_l)d_{2,k}d_{2,l}. \tag{A.14}$$

Let us go further with $D(\widehat{\boldsymbol{\beta}})$. We have

$$\mathrm{Var}(D(\widehat{\boldsymbol{\beta}})) = \mathrm{Var}_{S_2}[\mathbb{E}_{S_1}(D(\widehat{\boldsymbol{\beta}})|S_2)] + \mathbb{E}_{S_2}[\mathrm{Var}_{S_1}(D(\widehat{\boldsymbol{\beta}})|S_2)]. \tag{A.15}$$

We see that

$$\mathbb{E}_{S_1}(D(\widehat{\boldsymbol{\beta}})\,|S_2) = \frac{1}{A}\sum_{i\in I_n} \frac{\mathbb{E}_{S_1}(\widehat{\lambda}^{\star}(\mathbf{u}_i)\,|S_2) - \mathbb{E}_{S_1}(R^{\star}(\widehat{\boldsymbol{\beta}})\,|S_2)z(\boldsymbol{u}_i)}{\pi(\boldsymbol{u}_i)}$$

and, by (A.11), we obtain

$$\begin{aligned}\mathbb{E}_{S_1}[\widehat{\lambda}^{\star}(\boldsymbol{u})\,|S_2] &= \int_{C(\boldsymbol{u})}\frac{1}{a_{\boldsymbol{u}'}}E_{S_1}[\lambda_{\widehat{\boldsymbol{\beta}}}(\boldsymbol{u}')]I_{\boldsymbol{u}'}d\boldsymbol{u}'\\ &\approx \int_{C(\boldsymbol{u})}\frac{1}{a_{\boldsymbol{u}'}}\lambda_{\boldsymbol{\beta}}(\boldsymbol{u}')I_{\boldsymbol{u}'}d\boldsymbol{u}' + \sum_{k=1}^{q}E_{S_1}[\widehat{\beta}_k-\beta_k]\int_{C(\boldsymbol{u})}\frac{1}{a_{\boldsymbol{u}'}}I_{\boldsymbol{u}'}\lambda_{\boldsymbol{\beta}}^{(k)}(\boldsymbol{u}')d\boldsymbol{u}'\\ &\approx \int_{C(\boldsymbol{u})}\frac{1}{a_{\boldsymbol{u}'}}\lambda_{\boldsymbol{\beta}}(\boldsymbol{u}')I_{\boldsymbol{u}'}d\boldsymbol{u}' = \lambda^{\star}(\boldsymbol{u}).\end{aligned}$$

Thus,

$$\mathbb{E}_{S_1}(D(\widehat{\boldsymbol{\beta}})\,|S_2) \approx \frac{1}{A}\sum_{i\in I_n}\frac{\lambda^{\star}(\boldsymbol{u}_i) - R^{\star}(\boldsymbol{\beta})z(\boldsymbol{u}_i)}{\pi(\boldsymbol{u}_i)} = \widehat{R}_0^{\star}(\boldsymbol{\beta}) - R^{\star}(\boldsymbol{\beta}) \tag{A.16}$$

and, from the Sen-Yates-Grundy formula presented in Cordy (1993),

$$\begin{aligned}&\mathrm{Var}_{S_2}[\mathbb{E}_{S_1}(D(\widehat{\boldsymbol{\beta}})\,|S_2)] \approx \mathrm{Var}_{S_2}\left(\widehat{R}_0^{\star}(\boldsymbol{\beta})\right)\\ &= \frac{1}{2A^2}\int_{U^{\bullet}}\int_{U^{\bullet}}\Delta(\boldsymbol{u}_i,\boldsymbol{u}_j)\left(\frac{\lambda^{\star}(\boldsymbol{u}) - R^{\star}(\boldsymbol{\beta})z(\boldsymbol{u})}{\pi(\boldsymbol{u})} - \frac{\lambda^{\star}(\boldsymbol{u}') - R^{\star}(\boldsymbol{\beta})z(\boldsymbol{u}')}{\pi(\boldsymbol{u}')}\right)^2.\end{aligned} \tag{A.17}$$

Then, we can look closer at

$$\mathrm{Var}_{S_1}(D(\widehat{\boldsymbol{\beta}})\,|S_2) = \mathbb{E}_{S_1}\left(D^2(\widehat{\boldsymbol{\beta}})\,|S_2\right) - (\mathbb{E}_{S_1}(D(\widehat{\boldsymbol{\beta}})\,|S_2))^2,$$

which is a part of (A.15), where

$$\mathbb{E}_{S_1}\left(D^2(\widehat{\boldsymbol{\beta}})\,|S_2\right) = \frac{1}{A^2}\sum_{i\in I_n}\sum_{j\in I_n}\mathbb{E}_{S_1}\left[\left(\frac{\widehat{\lambda}^{\star}(\boldsymbol{u}_i) - R^{\star}(\widehat{\boldsymbol{\beta}})z(\boldsymbol{u}_i)}{\pi(\boldsymbol{u}_i)}\right)\left(\frac{\widehat{\lambda}^{\star}(\boldsymbol{u}_j) - R^{\star}(\widehat{\boldsymbol{\beta}})z(\boldsymbol{u}_j)}{\pi(\boldsymbol{u}_j)}\right)\middle|S_2\right]. \tag{A.18}$$

From (A.11), we see that

$$\begin{aligned}\mathbb{E}_{S_1}[\widehat{\lambda}^{\star}(\boldsymbol{u})\widehat{\lambda}^{\star}(\boldsymbol{u}')] &\approx \int_{C(\boldsymbol{u})}\int_{C(\boldsymbol{u}')}a_{\boldsymbol{v}}^{-1}a_{\boldsymbol{v}'}^{-1}\lambda_{\boldsymbol{\beta}}(\boldsymbol{v})\lambda_{\boldsymbol{\beta}}(\boldsymbol{v}')I_{\boldsymbol{v}}I_{\boldsymbol{v}'}d\boldsymbol{v}'d\boldsymbol{v}\\ &+\sum_{k=1}^{q}\sum_{l=1}^{q}\mathrm{Cov}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)d_{1,k}(\boldsymbol{u})d_{1,l}(\boldsymbol{u}')\\ &= \lambda^{\star}(\boldsymbol{u})\lambda^{\star}(\boldsymbol{u}') + \sum_{k=1}^{q}\sum_{l=1}^{q}\mathrm{Cov}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)d_{1,k}(\boldsymbol{u})d_{1,l}(\boldsymbol{u}'),\end{aligned} \tag{A.19}$$

where

$$d_{1,k}(\boldsymbol{u}) = \int_{C(\boldsymbol{u})}a_{\boldsymbol{u}'}^{-1}I_{\boldsymbol{u}'}\lambda_{\boldsymbol{\beta}}^{(k)}(\boldsymbol{u}')d\boldsymbol{u}' = \int_{C(\boldsymbol{u})}a_{\boldsymbol{u}'}^{-1}I_{\boldsymbol{u}'}x(\boldsymbol{u}')_k\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}(\boldsymbol{u}')\right)d\boldsymbol{u}',$$

and that

$$\begin{aligned}&\mathbb{E}_{S_1}[\widehat{\lambda}^{\star}(\boldsymbol{u})R^{\star}(\widehat{\boldsymbol{\beta}})] = \frac{1}{A}\int_{C(\boldsymbol{u})}a_{\boldsymbol{v}}^{-1}\mathbb{E}_{S_1}[\lambda_{\widehat{\boldsymbol{\beta}}}(\boldsymbol{v})\Lambda^{\star}(\widehat{\boldsymbol{\beta}})]I_{\boldsymbol{v}}d\boldsymbol{v} = \frac{1}{A}\int_{U}\int_{C(\boldsymbol{u})}a_{\boldsymbol{v}}^{-1}\mathbb{E}_{S_1}[\lambda_{\widehat{\boldsymbol{\beta}}}(\boldsymbol{v})\lambda_{\widehat{\boldsymbol{\beta}}}(\boldsymbol{v}')]I_{\boldsymbol{v}}I_{\boldsymbol{v}'}d\boldsymbol{v}d\boldsymbol{v}'\\ &\approx \frac{1}{A}\int_{U}\int_{C(\boldsymbol{u})}a_{\boldsymbol{v}}^{-1}\lambda_{\boldsymbol{\beta}}(\boldsymbol{v})\lambda_{\boldsymbol{\beta}}(\boldsymbol{v}')I_{\boldsymbol{v}}I_{\boldsymbol{v}'}d\boldsymbol{v}d\boldsymbol{v}'\\ &\quad+\frac{1}{A}\sum_{k=1}^{q}\sum_{l=1}^{q}Cov_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\int_{U}\int_{C(\boldsymbol{u})}a_{\boldsymbol{v}}^{-1}x(\boldsymbol{v})_k\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}(\boldsymbol{v})\right)x(\boldsymbol{v}')_l\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}(\boldsymbol{v}')\right)I_{\boldsymbol{v}}I_{\boldsymbol{v}'}d\boldsymbol{v}d\boldsymbol{v}'\\ &= \lambda^{\star}(\boldsymbol{u})R^{\star}(\boldsymbol{\beta}) + \frac{1}{A}\sum_{k=1}^{q}\sum_{l=1}^{q}Cov_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)d_{1,k}(\boldsymbol{u})d_{2,l}.\end{aligned} \tag{A.20}$$

From (A.18), (A.19) and (A.20), we obtain

$$\begin{aligned}\mathbb{E}_{S_1}\left[D^2(\widehat{\boldsymbol{\beta}})\,|S_2\right] &\approx \frac{1}{A^2}\sum_{i\in I_n}\sum_{j\in I_n}\left(\frac{\lambda^\star(\boldsymbol{u}_i)-R^\star(\boldsymbol{\beta})z(\boldsymbol{u}_i)}{\pi(\boldsymbol{u}_i)}\right)\left(\frac{\lambda^\star(\boldsymbol{u}_j)-R^\star(\boldsymbol{\beta})z(\boldsymbol{u}_j)}{\pi(\boldsymbol{u}_j)}\right)\\ &+\frac{1}{A^2}\sum_{k=1}^{q}\sum_{l=1}^{q}\mathrm{Cov}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\sum_{i\in I_n}\sum_{j\in I_n}\left(\frac{d_{1,k}(\boldsymbol{u}_i)-z(\boldsymbol{u}_i)d_{2,k}/A}{\pi(\boldsymbol{u}_i)}\right)\left(\frac{d_{1,l}(\boldsymbol{u}_j)-z(\boldsymbol{u}_j)d_{2,l}/A}{\pi(\boldsymbol{u}_j)}\right)\\ &=\left(\widehat{R}_0^\star(\boldsymbol{\beta})-R^\star(\boldsymbol{\beta})\right)^2\\ &\quad+\frac{1}{A^2}\sum_{k=1}^{q}\sum_{l=1}^{q}\mathrm{Cov}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\sum_{i\in I_n}\sum_{j\in I_n}\left(\frac{d_{1,k}(\boldsymbol{u}_i)-z(\boldsymbol{u}_i)d_{2,k}/A}{\pi(\boldsymbol{u}_i)}\right)\left(\frac{d_{1,l}(\boldsymbol{u}_j)-z(\boldsymbol{u}_j)d_{2,l}/A}{\pi(\boldsymbol{u}_j)}\right).\end{aligned}$$

This, together with (A.16), gives

$$\begin{aligned}&\mathrm{Var}_{S_1}(D(\widehat{\boldsymbol{\beta}})\,|S_2)\approx\frac{1}{A^2}\sum_{k=1}^{q}\sum_{l=1}^{q}\mathrm{Cov}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\sum_{i\in I_n}\left(\frac{d_{1,k}(\boldsymbol{u}_i)-z(\boldsymbol{u}_i)d_{2,k}/A}{\pi(\boldsymbol{u}_i)}\right)\sum_{j\in I_n}\left(\frac{d_{1,l}(\boldsymbol{u}_j)-z(\boldsymbol{u}_j)d_{2,l}/A}{\pi(\boldsymbol{u}_j)}\right)\\ &=\frac{1}{A^2}\sum_{k=1}^{q}\sum_{l=1}^{q}\mathrm{Cov}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\sum_{i\in I_n}\frac{1}{\pi(\boldsymbol{u}_i)^2}\left(d_{1,k}(\boldsymbol{u}_i)-z(\boldsymbol{u}_i)d_{2,k}\frac{1}{A}\right)\left(d_{1,l}(\boldsymbol{u}_i)-z(\boldsymbol{u}_i)d_{2,l}\frac{1}{A}\right)\\ &\quad+\frac{1}{A^2}\sum_{k=1}^{q}\sum_{l=1}^{q}\mathrm{Cov}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\sum_{i\in I_n}\sum_{j\in J_{n,i}}\frac{1}{\pi(\boldsymbol{u}_i)\pi(\boldsymbol{u}_j)}\left(d_{1,k}(\boldsymbol{u}_i)-z(\boldsymbol{u}_i)d_{2,k}\frac{1}{A}\right)\left(d_{1,l}(\boldsymbol{u}_j)-z(\boldsymbol{u}_j)d_{2,l}\frac{1}{A}\right).\end{aligned}$$

It follows that

$$\begin{aligned}&\mathbb{E}_{S_2}[\mathrm{Var}_{S_1}(D(\widehat{\boldsymbol{\beta}})\,|S_2)\,]\approx\frac{1}{A^2}\sum_{k=1}^{q}\sum_{l=1}^{q}\mathrm{Cov}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\int_{U^\bullet}\frac{1}{\pi(\boldsymbol{u})}\left(d_{1,k}(\boldsymbol{u})-z(\boldsymbol{u})d_{2,k}\frac{1}{A}\right)\left(d_{1,l}(\boldsymbol{u})-z(\boldsymbol{u})d_{2,l}\frac{1}{A}\right)d\boldsymbol{u}\\ &+\frac{1}{A^2}\sum_{k=1}^{q}\sum_{l=1}^{q}\mathrm{Cov}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\int_{U^\bullet}\int_{U^\bullet}\frac{\pi(\boldsymbol{u},\boldsymbol{u}^{'})}{\pi(\boldsymbol{u})\pi(\boldsymbol{u}^{'})}\left(d_{1,k}(\boldsymbol{u})-z(\boldsymbol{u})d_{2,k}\frac{1}{A}\right)\left(d_{1,l}(\boldsymbol{u}^{'})-z(\boldsymbol{u}^{'})d_{2,l}\frac{1}{A}\right)d\boldsymbol{u}d\boldsymbol{u}^{'}.\end{aligned}\tag{A.21}$$

If we put (A.17) and (A.21) together, we obtain

$$\begin{aligned}&\mathrm{Var}(D(\widehat{\boldsymbol{\beta}})\,)=\mathrm{Var}_{S_2}[\mathbb{E}_{S_1}(D(\widehat{\boldsymbol{\beta}})\,|S_2)\,]+\mathbb{E}_{S_2}[\mathrm{Var}_{S_1}(D(\widehat{\boldsymbol{\beta}})\,|S_2)\,]\\ &\approx\frac{1}{2A^2}\int_{U^\bullet}\int_{U^\bullet}\Delta(\boldsymbol{u}_i\boldsymbol{u}_j)\left(\frac{\lambda^\star(\boldsymbol{u})-R^\star(\boldsymbol{\beta})z(\boldsymbol{u})}{\pi(\boldsymbol{u})}-\frac{\lambda^\star(\boldsymbol{u}^{'})-R^\star(\boldsymbol{\beta})z(\boldsymbol{u}^{'})}{\pi(\boldsymbol{u}^{'})}\right)^2d\boldsymbol{u}d\boldsymbol{u}^{'}\\ &+\frac{1}{A^2}\sum_{k=1}^{q}\sum_{l=1}^{q}\mathrm{Cov}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\int_{U^\bullet}\frac{1}{\pi(\boldsymbol{u})}\left(d_{1,k}(\boldsymbol{u})-z(\boldsymbol{u})d_{2,k}\frac{1}{A}\right)\left(d_{1,l}(\boldsymbol{u})-z(\boldsymbol{u})d_{2,l}\ \frac{1}{A}\right)d\boldsymbol{u}\\ &+\frac{1}{A^2}\sum_{k=1}^{q}\sum_{l=1}^{q}\mathrm{Cov}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\int_{U^\bullet}\int_{U^\bullet}\frac{\pi(\boldsymbol{u}\boldsymbol{u}^{'})}{\pi(\boldsymbol{u})\pi(\boldsymbol{u}^{'})}\left(d_{1,k}(\boldsymbol{u})-z(\boldsymbol{u})d_{2,k}\frac{1}{A}\right)\left(d_{1,l}(\boldsymbol{u}^{'})-z(\boldsymbol{u}^{'})d_{2,l}\ \frac{1}{A}\right)d\boldsymbol{u}d\boldsymbol{u}^{'}.\end{aligned}\tag{A.22}$$

Furthermore,

$$\mathrm{Cov}(D(\widehat{\boldsymbol{\beta}})\,,D_*(\widehat{\boldsymbol{\beta}})\,)=\mathrm{Cov}_{S_2}[\mathbb{E}_{S_1}(D(\widehat{\boldsymbol{\beta}})\,|S_2)\,,\mathbb{E}_{S_1}(D_*(\widehat{\boldsymbol{\beta}})\,|S_2)\,]+\mathbb{E}_{S_2}[\mathrm{Cov}_{S_1}(D(\widehat{\boldsymbol{\beta}})\,,D_*(\widehat{\boldsymbol{\beta}})\,|S_2)\,].$$

From earlier calculations, we know that $\mathbb{E}_{S_1}(D(\widehat{\boldsymbol{\beta}})\,|S_2)\approx\widehat{R}_0^\star(\boldsymbol{\beta})-R^\star(\boldsymbol{\beta})$ and $\mathbb{E}_{S_1}(D_*(\widehat{\boldsymbol{\beta}})\,|S_2)\approx 0$, and thus $\mathrm{Cov}_{S_2}[\mathbb{E}_{S_1}(D(\widehat{\boldsymbol{\beta}})\,|S_2)\,,\mathbb{E}_{S_1}(D_*(\widehat{\boldsymbol{\beta}})\,|S_2)\,]\approx 0$. In addition, using (A.16),

$$\begin{aligned}\mathrm{Cov}_{S_1}(D(\widehat{\boldsymbol{\beta}})\,,D_*(\widehat{\boldsymbol{\beta}})\,|S_2) &\approx\mathbb{E}_{S_1}(D(\widehat{\boldsymbol{\beta}})D_*(\widehat{\boldsymbol{\beta}})\,|S_2)=\mathbb{E}_{S_1}(D(\widehat{\boldsymbol{\beta}})R^\star(\widehat{\boldsymbol{\beta}})\,|S_2)-R^\star(\boldsymbol{\beta})\mathbb{E}_{S_1}(D(\widehat{\boldsymbol{\beta}})\,|S_2)\\ &\approx\mathbb{E}_{S_1}(D(\widehat{\boldsymbol{\beta}})R^\star(\widehat{\boldsymbol{\beta}})\,|S_2)-R^\star(\boldsymbol{\beta})\left(\widehat{R}_0^\star(\boldsymbol{\beta})-R^\star(\boldsymbol{\beta})\right)\end{aligned}$$

and, from (A.12) and (A.20),

$$\mathbb{E}_{S_1}(D(\widehat{\boldsymbol{\beta}})R^\star(\widehat{\boldsymbol{\beta}})\,|S_2) = \frac{1}{A}\mathbb{E}_{S_1}\left(\sum_{i\in I_n}\frac{\widehat{\lambda}^\star(\boldsymbol{u}_i)R^\star(\widehat{\boldsymbol{\beta}}) - (R^\star(\widehat{\boldsymbol{\beta}}))^2 z(\boldsymbol{u}_i)}{\pi(\boldsymbol{u}_i)}\middle| S_2\right)$$

$$\approx R^\star(\boldsymbol{\beta})\frac{1}{A}\sum_{i\in I_n}\frac{\lambda^\star(\boldsymbol{u}_i) - R^\star(\boldsymbol{\beta})z(\boldsymbol{u}_i)}{\pi(\boldsymbol{u}_i)}$$

$$+\frac{1}{A^2}\sum_{k=1}^{q}\sum_{l=1}^{q}Cov_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\sum_{i\in I_n}\frac{1}{\pi(\boldsymbol{u}_i)}d_{1,k}(\boldsymbol{u}_i)d_{2,l}$$

$$-\frac{1}{A^3}\sum_{k=1}^{q}\sum_{l=1}^{q}Cov_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\sum_{i\in I_n}\frac{z(\boldsymbol{u}_i)}{\pi(\boldsymbol{u}_i)}d_{2,k}d_{2,l}.$$

As a consequence,

$$\begin{aligned}
\mathrm{Cov}(D(\widehat{\boldsymbol{\beta}}), D_*(\widehat{\boldsymbol{\beta}})) &\approx \mathbb{E}_{S_2}\left(R^\star(\boldsymbol{\beta})(\widehat{R}_0^\star(\boldsymbol{\beta}) - R^\star(\boldsymbol{\beta}))\right) + \frac{1}{A^2}\sum_{k=1}^{q}\sum_{l=1}^{q}\mathrm{Cov}_{S_1}\left(\widehat{\beta}_k,\widehat{\beta}_l\right)\sum_{i\in I_n}\frac{1}{\pi(\boldsymbol{u}_i)}d_{1,k}(\boldsymbol{u}_i)d_{2,l}\\
&\quad -\frac{1}{A^3}\sum_{k=1}^{q}\sum_{l=1}^{q}\mathrm{Cov}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\sum_{i\in I_n}\frac{z(\boldsymbol{u}_i)}{\pi(\boldsymbol{u}_i)}d_{2,k}d_{2,l} - R^\star(\boldsymbol{\beta})(\widehat{R}_0^\star(\boldsymbol{\beta}) - R^\star(\boldsymbol{\beta}))\Big)\\
&= \frac{1}{A^2}\sum_{k=1}^{q}\sum_{l=1}^{q}\mathrm{Cov}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\mathbb{E}_{S_2}\left(\sum_{i\in I_n}\frac{1}{\pi(\boldsymbol{u}_i)}d_{1,k}(\boldsymbol{u}_i)d_{2,l}\right) - \frac{1}{A^3}\sum_{k=1}^{p}\sum_{l=1}^{p}\mathrm{Cov}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\mathbb{E}_{S_2}\left(\sum_{i\in I_n}\frac{z(\boldsymbol{u}_i)}{\pi(\boldsymbol{u}_i)}d_{2,k}d_{2,l}\right)\\
&= \frac{1}{A^2}\sum_{k=1}^{q}\sum_{l=1}^{q}\mathrm{Cov}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)d_{2,l}\int_{U^\bullet}d_{1,k}(\boldsymbol{u})d\boldsymbol{u} - \frac{1}{A^3}\sum_{k=1}^{q}\sum_{l=1}^{q}\mathrm{Cov}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)d_{2,k}d_{2,l}\int_{U}z(\boldsymbol{u})d\boldsymbol{u}\\
&= \frac{1}{A^2}\sum_{k=1}^{q}\sum_{l=1}^{q}\mathrm{Cov}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)d_{2,l}\int_{U^\bullet}d_{1,k}(\boldsymbol{u})d\boldsymbol{u} - \frac{1}{A^2}\sum_{k=1}^{q}\sum_{l=1}^{q}\mathrm{Cov}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)d_{2,k}d_{2,l}.
\end{aligned} \tag{A.23}$$

Finally, putting (A.22), (A.14) and (A.23) together,

$$\begin{aligned}
&\mathrm{Var}(\widehat{R}_0^\star(\widehat{\boldsymbol{\beta}}) - R^\star(\boldsymbol{\beta})) = \mathrm{Var}(D(\widehat{\boldsymbol{\beta}})) + \mathrm{Var}(D_*(\widehat{\boldsymbol{\beta}})) + 2\,\mathrm{Cov}(D(\widehat{\boldsymbol{\beta}}), D_*(\widehat{\boldsymbol{\beta}}))\\
&\approx \frac{1}{2A^2}\int_{U^\bullet}\int_{U^\bullet}\Delta(\boldsymbol{u}_i,\boldsymbol{u}_j)\left(\frac{\lambda^\star(\boldsymbol{u}) - R^\star(\boldsymbol{\beta})z(\boldsymbol{u})}{\pi(\boldsymbol{u})} - \frac{\lambda^\star(\boldsymbol{u}') - R^\star(\boldsymbol{\beta})z(\boldsymbol{u}')}{\pi(\boldsymbol{u}')}\right)^2\\
&\quad +\frac{1}{A^2}\sum_{k=1}^{q}\sum_{l=1}^{q}\mathrm{Cov}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\int_{U^\bullet}\frac{1}{\pi(\boldsymbol{u})}\left(d_{1,k}(\boldsymbol{u}) - z(\boldsymbol{u})d_{2,k}\frac{1}{A}\right)\left(d_{1,l}(\boldsymbol{u}) - z(\boldsymbol{u})d_{2,l}\frac{1}{A}\right)d\boldsymbol{u}\\
&\quad +\frac{1}{A^2}\sum_{k=1}^{q}\sum_{l=1}^{q}\mathrm{Cov}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\int_{U^\bullet}\int_{U^\bullet}\frac{\pi(\boldsymbol{u},\boldsymbol{u}')}{\pi(\boldsymbol{u})\pi(\boldsymbol{u}')}\left(d_{1,k}(\boldsymbol{u}) - z(\boldsymbol{u})d_{2,k}\frac{1}{A}\right)\left(d_{1,l}(\boldsymbol{u}') - z(\boldsymbol{u}')d_{2,l}\frac{1}{A}\right)d\boldsymbol{u}d\boldsymbol{u}'\\
&\quad +\frac{2}{A^2}\sum_{k=1}^{q}\sum_{l=1}^{q}\mathrm{Cov}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)d_{2,l}\int_{U^\bullet}d_{1,k}(\boldsymbol{u})d\boldsymbol{u} - \frac{1}{A^2}\sum_{k=1}^{q}\sum_{l=1}^{q}\mathrm{Cov}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)d_{2,k}d_{2,l}.
\end{aligned}$$

By using Theorem 1 and the variance estimator based on the Sen-Yates-Grundy formula in Cordy (1993), this variance can be estimated by

$$\begin{aligned}
\widehat{\mathrm{Var}}(\widehat{R}_0^\star(\widehat{\boldsymbol{\beta}}) - R^\star(\boldsymbol{\beta})) &= \frac{1}{2\widehat{A}^2}\sum_{i\in I_n}\sum_{j\in J_{n,i}}\frac{\Delta(\boldsymbol{u}_i,\boldsymbol{u}_j)}{\pi(\boldsymbol{u}_i,\boldsymbol{u}_j)}\left(\frac{\widehat{\lambda}^\star(\boldsymbol{u}_i) - \widehat{R}^\star(\widehat{\boldsymbol{\beta}})z(\boldsymbol{u}_i)}{\pi(\boldsymbol{u}_i)} - \frac{\widehat{\lambda}^\star(\boldsymbol{u}_j) - \widehat{R}^\star(\widehat{\boldsymbol{\beta}})z(\boldsymbol{u}_j)}{\pi(\boldsymbol{u}_j)}\right)^2\\
&\quad +\frac{1}{\widehat{A}^2}\sum_{k=1}^{q}\sum_{l=1}^{q}\widehat{\mathrm{Cov}}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\sum_{i\in I_n}\frac{1}{\pi(\boldsymbol{u}_i)^2}\left(\widehat{d}_{1,k}(\boldsymbol{u}_i) - z(\boldsymbol{u}_i)\widehat{d}_{2,k}\frac{1}{\widehat{A}}\right)\left(\widehat{d}_{1,l}(\boldsymbol{u}_i) - z(\boldsymbol{u}_i)\widehat{d}_{2,l}\frac{1}{\widehat{A}}\right)\\
&\quad +\frac{1}{\widehat{A}^2}\sum_{k=1}^{q}\sum_{l=1}^{q}\widehat{\mathrm{Cov}}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\sum_{i\in I_n}\sum_{j\in I_{i,n}}\frac{1}{\pi(\boldsymbol{u}_i)\pi(\boldsymbol{u}_j)}\left(\widehat{d}_{1,k}(\boldsymbol{u}_i) - z(\boldsymbol{u}_i)\widehat{d}_{2,k}\frac{1}{\widehat{A}}\right)\left(\widehat{d}_{1,l}(\boldsymbol{u}_j) - z(\boldsymbol{u}_j)\widehat{d}_{2,l}\frac{1}{\widehat{A}}\right)\\
&\quad +\frac{2}{\widehat{A}^2}\sum_{k=1}^{q}\sum_{l=1}^{q}\widehat{\mathrm{Cov}}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\widehat{d}_{2,l}\sum_{i\in I_n}\frac{\widehat{d}_{1,k}(\boldsymbol{u}_i)}{\pi(\boldsymbol{u}_i)} - \frac{1}{\widehat{A}^2}\sum_{k=1}^{q}\sum_{l=1}^{q}\widehat{\mathrm{Cov}}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\widehat{d}_{2,k}\widehat{d}_{2,l}\\
&= \frac{1}{2\widehat{A}^2}\sum_{i\in I_n}\sum_{j\in J_{n,i}}\frac{\Delta(\boldsymbol{u}_i,\boldsymbol{u}_j)}{\pi(\boldsymbol{u}_i,\boldsymbol{u}_j)}\left(\frac{\widehat{\lambda}^\star(\boldsymbol{u}_i) - \widehat{R}^\star(\widehat{\boldsymbol{\beta}})z(\boldsymbol{u}_i)}{\pi(\boldsymbol{u}_i)} - \frac{\widehat{\lambda}^\star(\boldsymbol{u}_j) - \widehat{R}^\star(\widehat{\boldsymbol{\beta}})z(\boldsymbol{u}_j)}{\pi(\boldsymbol{u}_j)}\right)^2\\
&\quad +\frac{1}{\widehat{A}^2}\sum_{k=1}^{q}\sum_{l=1}^{q}\widehat{\mathrm{Cov}}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\sum_{i\in I_n}\frac{\widehat{d}_{1,k}(\boldsymbol{u}_i) - z(\boldsymbol{u}_i)\widehat{d}_{2,k}/\widehat{A}}{\pi(\boldsymbol{u}_i)}\sum_{j\in I_n}\frac{\widehat{d}_{1,l}(\boldsymbol{u}_j) - z(\boldsymbol{u}_j)\widehat{d}_{2,l}/\widehat{A}}{\pi(\boldsymbol{u}_j)}\\
&\quad +\frac{2}{\widehat{A}^2}\sum_{k=1}^{q}\sum_{l=1}^{q}\widehat{\mathrm{Cov}}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\widehat{d}_{2,l}\sum_{i\in I_n}\frac{\widehat{d}_{1,k}(\boldsymbol{u}_i)}{\pi(\boldsymbol{u}_i)} - \frac{1}{\widehat{A}^2}\sum_{k=1}^{q}\sum_{l=1}^{q}\widehat{\mathrm{Cov}}_{S_1}(\widehat{\beta}_k,\widehat{\beta}_l)\widehat{d}_{2,k}\widehat{d}_{2,l},
\end{aligned} \tag{A.24}$$

where $\widehat{d}_{1,k}$ and $\widehat{d}_{2,k}$ are as defined in (21).

Appendix B. Case with divided plots

It can happen that sample plots are divided into several parts, for example if one part of the plot is in forests and other parts are in other landscape categories, or if the plot overlaps borders between different regions, strata or forest stands (for example in the Swedish NFI, Anon. 2014). In such cases, the covariate information is not the same in different parts of the plot. Let us consider a case where we want to study expected plant densities in forests, and consider a particular plot $C(\boldsymbol{u}_i)$. Then let $I_{\mathbf{u}}$ be equal to 1 if $\boldsymbol{u}$ is in a forested area in U, and 0 otherwise. If the plot is divided and no part of the plot is in a forested area in U,

$$\lambda^\star(\boldsymbol{u}_i) = \int_{C(\boldsymbol{u}_i)} \frac{\lambda_{\boldsymbol{\beta}}(\boldsymbol{u}) I_{\boldsymbol{u}}}{a_{\boldsymbol{u}}} d\boldsymbol{u} = 0 \quad \text{and} \quad \widehat{\lambda}^\star(\boldsymbol{u}_i) = 0. \tag{B.1}$$

If only one part of the plot is in a forested area in U, and if we denote the area of this part by $a_i^{(s)}$,

$$\lambda^\star(\boldsymbol{u}_i) = \int_{C(\boldsymbol{u}_i)} \frac{\lambda_{\boldsymbol{\beta}}(\boldsymbol{u}) I_{\boldsymbol{u}}}{a_{\boldsymbol{u}}} d\boldsymbol{u} = \lambda_{\boldsymbol{\beta}}(\boldsymbol{u}'_i) \frac{a_i^{(s)}}{a} \quad \text{and} \quad \widehat{\lambda}^\star(\boldsymbol{u}_i) = \lambda_{\widehat{\boldsymbol{\beta}}}(\boldsymbol{u}'_i) \frac{a_i^{(s)}}{a}, \tag{B.2}$$

where $\lambda_{\widehat{\boldsymbol{\beta}}}(\boldsymbol{u}_i) = \exp\big(\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{x}(\boldsymbol{u}_i)\big) = \exp\big(\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{x}_i\big)$ and $\boldsymbol{u}'_i$ is an arbitrary point in the forested part of $C(\boldsymbol{u}_i) \cap U$. If $C(\boldsymbol{u}_i)$ has two parts that are in forests within U (with areas $a_i^{(s_1)}$ and $a_i^{(s_2)}$ respectively), then

$$\lambda^\star(\boldsymbol{u}_i) = \lambda_{\boldsymbol{\beta}}(\boldsymbol{u}'_i) \frac{a_i^{(s_1)}}{a} + \lambda_{\boldsymbol{\beta}}(\boldsymbol{u}''_i) \frac{a_i^{(s_2)}}{a} \quad \text{and} \quad \widehat{\lambda}^\star(\boldsymbol{u}_i) = \lambda_{\widehat{\boldsymbol{\beta}}}(\boldsymbol{u}'_i) \frac{a_i^{(s_1)}}{a} + \lambda_{\widehat{\boldsymbol{\beta}}}(\boldsymbol{u}''_i) \frac{a_i^{(s_2)}}{a} \tag{B.3}$$

where $\boldsymbol{u}'_i$ is an arbitrary point in the first forest part of $C(\boldsymbol{u}_i) \cap U$ and $\boldsymbol{u}''_i$ is an arbitrary point in the second forest part of $C(\boldsymbol{u}_i) \cap U$. And so on with three or more forest parts. Thus, the change of expression of $\widehat{\lambda}^\star(\boldsymbol{u}_i)$ will imply changes when applying formulas (16) and (A.24) for estimating the expected density and its variance estimator. Similar changes need to be done in the cluster sampling case presented in Section 2.5.

Appendix C. Details of the proposed goodness-of-fit test

Assume that there are two disjoint vegetation plots, A_{i1} and A_{i2}, contained in each (main) plot i, where all A_{ij} are of size a_A, $i = 1, \ldots, n$. Each vegetation plot A_{i1} and A_{i2} in a pair is separated by the same distance d. In each A_{ij}, the presence or absence of the plant species of interest is registered. Let M_i be the number of plants in plot A_i, $i = 1, \ldots, n$. Let Y_{ij} be 1 if presence in A_{ij}, and 0 otherwise, $i = 1, \ldots, n$, $j = 1, 2$. In our case, the M_i are not observed, contrary to the Y_{ij}, hence the necessity to develop a test based on the latter. Based on the sample of Y_{ij} data and corresponding covariate data x_i (assumed to be fixed in plot i), an estimator $\widehat{\boldsymbol{\beta}}$ of the parameter vector $\boldsymbol{\beta}$ is obtained using a binary regression with a complementary log-log link function (5). Let Y_i be 1 if there is at least one point in the union of A_{i1} and A_{i2}, and 0 otherwise. Based on a binary regression with a complementary log-log link function, offset $\log(2a_A)$, and the data $\{Y_i, x_i\}$, $i = 1, \ldots, n$, another estimator of $\boldsymbol{\beta}$ is constructed, denoted by $\widetilde{\boldsymbol{\beta}}$.

If the inhomogeneous Poisson point process model assumption is correct, then so is the model for the Y_{ij}. The reverse is not necessarily true. However, if the model for the Y_{ij} is incorrect, then so is the Poisson model for the M_i.

If the inhomogeneous Poisson point process model is correct, Y_{i1} and Y_{i2} will be independent conditional on the covariates, and binary regression model (5) implies the binary regression model based on the data $\{Y_i, x_i\}$. In this case, $\widehat{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\beta}}$ will be close for large n. On the other hand, if Y_{i1} and Y_{i2} are not independent conditional on the covariates, then this implication will not hold and $\widehat{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\beta}}$ will likely differ even if n is large. Based on this idea, Ekström et al. (Unpublished results) suggested the test statistic

$$S = (\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}})^{\mathrm{T}} \ \widehat{\boldsymbol{\Sigma}}^{-1} (\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}), \tag{C.1}$$

where $\widehat{\boldsymbol{\Sigma}}$ is an estimate of the covariance matrix of $\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}$ given by

$$\widehat{\boldsymbol{\Sigma}} = n\big(\widehat{\boldsymbol{I}}_1^{-1}(\widehat{\boldsymbol{\beta}}) + \widehat{\boldsymbol{I}}_2^{-1}(\widehat{\boldsymbol{\beta}}) - 2\widehat{\boldsymbol{I}}_1^{-1}(\widehat{\boldsymbol{\beta}}) \widehat{\boldsymbol{C}}(\widehat{\boldsymbol{\beta}}) \widehat{\boldsymbol{I}}_2^{-1}(\widehat{\boldsymbol{\beta}})\big),$$

where

$$\widehat{\boldsymbol{I}}_1(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^{n} \frac{2}{[g'(q_{i1}(\boldsymbol{\beta}))]^2 t_{i1}(\boldsymbol{\beta})} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}},$$

$$\widehat{\boldsymbol{I}}_2(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{[g'(q_i(\boldsymbol{\beta}))]^2 t_i(\boldsymbol{\beta})} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}},$$

$$\widehat{\boldsymbol{C}}(\boldsymbol{\beta}) = \frac{2}{n} \sum_{i=1}^{n} \frac{1}{g'(q_i(\boldsymbol{\beta})) t_i(\boldsymbol{\beta})} \frac{1}{g'(q_{i1}(\boldsymbol{\beta})) t_{i1}(\boldsymbol{\beta})} q_{i1}(\boldsymbol{\beta})(1 - q_i(\boldsymbol{\beta})) \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}},$$

$q_{ij}(\boldsymbol{\beta}) = 1 - \exp\big(-a_A \exp(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{x}_i)\big)$, $t_{ij}(\boldsymbol{\beta}) = q_{ij}\Big(1 - q_{ij}\Big)$, $q_i(\boldsymbol{\beta}) = 1 - \exp\big(-2a_A \exp(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{x}_i)\big)$, $t_i(\boldsymbol{\beta}) = q_i(1 - q_i)$, and $g(p) = \log(-\log(1 - p))$.

If the Poisson model is valid, S is asymptotically distributed according to a chi-squared distribution with q degrees of freedom, where q is the length of $\boldsymbol{\beta}$. The binary model (5),and hence the Poisson model, is rejected if S is improbably large according to this chi-squared distribution. For small or

moderately large sample sizes, a better option might be to use parametric bootstrap (Davison and Hinkley, 1997). The bootstrap algorithm for computing the p-value of the test is given below.

For $b = 1, \ldots, B$, where B is a large integer:

i) For A_{ij}, generate points according to a Poisson point process with log intensity $\log\widehat{\lambda}_i = \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{x}_i$, $i = 1, \ldots, n, j = 1, 2$.

ii) Based on the point data obtained in i), let Y^*_{ijb} be 1 if presence in A_{ij} and 0 otherwise, and let $Y^*_{ib} = max\{Y^*_{i1b}, Y^*_{i2b}\}$, $i = 1, \ldots, n$.

iii) Let S^* be defined as in (C.1), but based on $\left\{Y^*_{ijb}\right\}$ and $\left\{Y^*_{ib}\right\}$ rather than $\left\{Y_{ij}\right\}$ and $\{Y_i\}$.

The p-value of the test is given by the proportion of times S^* is larger than or equal to S.

References

Albert, P.S., McShane, L.M., 1995. A Generalized Estimating Equations Approach for Spatially Correlated Binary Data: Applications to the Analysis of Neuroimaging Data. Biometrics 51 (2), 627–638. https://doi.org/10.2307/2532950.

Ambler, G., Benner, A., 2015. mfp: Multivariable Fractional Polynomials. R Pack. Vers. 1 (5), 2. https://CRAN.R-project.org/package=mfp.

Anon, 2014. Fältinstruktion 2014, RIS, Riksinventeringen av skog [Field Instructions for the Swedish National Forest Inventory and the Swedish Forest Soil Inventory]. The Swedish University of Agricultural Sciences, Umeå, Sweden (In Swedish).

Artdatabanken, 2022. Artportalen. https://www.artdatabanken.se/sok-art-och-miljodata/artportalen/. Retrieved 17 August 2022.

Baddeley, A., Berman, M., Fisher, N.I., Hardegen, A., Milne, R.K., Schuhmacher, D., Turner, R., 2010. Spatial logistic regression and change-of-support for Poisson point processes. Electron. J. Stat. 4, 1151–1201. https://doi.org/10.1214/10-EJS581.

Baddeley, A., Rubak, E., Turner, R., 2016. Spatial Point Patterns: Methodology and Applications with R. CRC Press, Boca Raton. https://doi.org/10.1201/b19708.

Baena, S., Boyd, D.S., Moat, J., 2018. UAVs in pursuit of plant conservation-real world experiences. Eco. Inform. 47, 2–9. https://doi.org/10.1016/j.ecoinf.2017.11.001.

Bastow Wilson, J., 2012. Species presence/absence sometimes represents a plant community as well as species abundances do, or better. J. Veg. Sci. 23 (6), 1013–1023. https://doi.org/10.1111/j.1654-1103.2012.01430.x.

Belbin, L., 2011. The Atlas of Living Australia's Spatial Portal. In: Jones, M.B., Gries, C. (Eds.), Proceedings of the Environmental Information Management Conference 2011 (EIM 2011), pp. 39–43 (Santa Barbara).

Cassel, C., Särndal, C.E., Wretman, J.H., 1977. Foundations of inference in survey sampling. Wiley, New York. https://doi.org/10.2307/3314835.

CBD, 2002. Global Strategy for Plant Conservation. The Secretariat of the Convention on Biological Diversity, Montreal, Canada.

Commission of the European Communities, 2003. Council directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora. Official Journal of the European Union 1. 236 99 23.9.2003, Brussels. European Commission 1992/95/2003.

Condés, S., McRoberts, R.E., 2017. Updating national forest inventory estimates of growing stock volume using hybrid inference. For. Ecol. Manag. 400, 48–57. https://doi.org/10.1016/j.foreco.2017.04.046.

Conlisk, E., Conlisk, J., Harte, J., 2007. The impossibility of estimating a negative binomial clustering parameter from presence-absence data: a comment on He and Gaston. Am. Nat. 170 (4), 651–659. https://doi.org/10.1086/521339.

Cordy, C.B., 1993. An extension of the Horvitz-Thompson theorem to point sampling from a continuous universe. Stat. Probab. Lett. 18 (5), 353–362. https://doi.org/10.1016/0167-7152(93)90028-H.

Corona, P., Fattorini, L., Franseschi, S., Scrinzi, G., Torresan, C., 2014. Estimation of standing wood volume in forest compartments by exploiting airborne laser scanning information: model-based, design-based and hybrid perspectives. Can. J. For. Res. 44 (11), 1303–1311. https://doi.org/10.1139/cjfr-2014-0203.

Daley, D.J., Vere-Jones, D., 2003. An introduction to the theory of point processes: volume I: elementary theory and methods. Springer. https://doi.org/10.1007/b97277.

Davison, A., Hinkley, D., 1997. Bootstrap Methods and their Application (Cambridge Series in Statistical and Probabilistic Mathematics). Cambridge University Press. https://doi.org/10.1017/CBO9780511802843.

Delignette-Muller, M.-L., Dutang, C., 2015. fitdistrplus: An R Package for Fitting Distributions. J. Stat. Softw. 64 (4), 1–34. https://doi.org/10.18637/jss.v064.i04.

Dubayah, R., Armston, J., Healey, S.P., Bruening, J.M., Patterson, P.L., Kellner, J.R., Duncanson, L., Saarela, S., Ståhl, G., Yang, Z., Tang, H., Bryan Blair, J., Fatoyinbo, L., Goetz, S., Hancock, S., Hansen, M., Hofton, M., Hurtt, G., Luthcke, S., 2022. GEDI launches a new era of biomass inference from space. Environ. Res. Lett. 17 (9), 095001 https://doi.org/10.1088/1748-9326/ac8694.

Ekström, M., Esseen, P.-A., Westerlund, B., Grafström, A., Jonsson, B.G., Ståhl, G., 2018. Logistic regression for clustered data from environmental monitoring programs. Eco. Inform. 43, 165–173. https://doi.org/10.1016/j.ecoinf.2017.10.006.

Ekström, M., Sandring, S., Grafström, A., Esseen, P.-A., Jonsson, B.G., Ståhl, G., 2020. Estimating density from presence-absence data in clustered populations. Methods Ecol. Evol. 11 (3), 390–402. https://doi.org/10.1111/2041-210X.13347.

Ekström, M., Gozé, L., Wallerman, J., Dahlgren, J., Jonsson, B.-G., Sandring, S., Ståhl, G., 2023. Model-based estimation and mapping of plant density based on remote sensing and presence/absence data.

Esseen, P.-A., Ekström, M., 2023. Influence of canopy structure and microclimate on three-dimensional distribution of the iconic lichen Usnea longissimi. For. Ecol. Manag. 529, 120667 https://doi.org/10.1016/j.foreco.2022.120667.

Esseen, P.-A., Ekström, M., Grafström, A., Jonsson, B.G., Palmqvist, K., Westerlund, B., Ståhl, G., 2022. Multiple drivers of large-scale lichen decline in boreal forest canopies. Glob. Chang. Biol. 28 (10), 3293–3309. https://doi.org/10.1111/gcb.16128.

Fithian, W., Elith, J., Hastie, T., Keith, D.A., 2015. Bias correction in species distribution models: pooling survey and collection data for multiple species. Methods Ecol. Evol. 6 (4), 424–438. https://doi.org/10.1111/2041-210X.12242.

Foody, G.M., 2008. Refining predictions of climate change impacts on plant species distribution through the use of local statistics. Eco. Inform. 3 (3), 228–236. https://doi.org/10.1016/j.ecoinf.2008.02.002.

Fortin, M., Manso, R., Calama, R., 2016. Hybrid estimation based on mixed-effects models in forest inventories. Can. J. For. Res. 46 (11), 1310–1319. https://doi.org/10.1139/cjfr-2016-0298.

Fortin, M., Manso, R., Schneider, R., 2018. Parametric bootstrap estimators for hybrid inference in forest inventories. Forestry 91 (3), 354–365. https://doi.org/10.1093/forestry/cpx048.

Fortin, M., Lier, O.V., Côté, J.-F., 2023. Combining forest growth models and remotely sensed data through a hierarchical model-based inferential framework. Can. J. For. Res. 53, 1–13. https://doi.org/10.1139/cjfr-2022-0168.

Fridman, J., Holm, S., Nilsson, M., Nilsson, P., Ringvall, A.H., 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 (3). https://doi.org/10.14214/sf.1095.

Futschik, A., Winkler, M., Steinbauer, K., Lamprecht, A., Rumpf, S.B., Barančok, P., Palaj, A., Gottfried, M., Pauli, H., 2020. Disentangling observer error and climate change effects in long-term monitoring of alpine plant species composition and cover. J. Veg. Sci. 31 (1), 14–25. https://doi.org/10.1111/jvs.12822.

Gallegos Torell, Å., Glimskär, A., 2009. Computer-aided calibration for visual estimation of vegetation cover. J. Veg. Sci. 20 (6), 973–983. https://doi.org/10.1111/j.1654-1103.2009.01111.x.

Gaston, K.J., He, F., Maguran, A., McGill, B., 2011. Species occurrence and occupancy. Biol. Divers. 141–151.

GBIF, 2022. What Is GBIF? Available from. https://www.gbif.org/what-is-gbif.

Gotway, C.A., Stroup, W.W., 1997. A Generalized Linear Model Approach to Spatial Data Analysis and Prediction. J. Agric. Biol. Environ. Stat. 2 (2), 157–178. https://doi.org/10.2307/1400401.

Grafström, A., Schnell, S., Saarela, S., Hubbell, S.P., Condit, R., 2017. The continuous population approach to forest inventories and use of information in the design. Environmetrics 28 (8). https://doi.org/10.1002/env.2480.

Gregoire, T.G., Valentine, H.T., 2007. Sampling Strategies for Natural Resources and the Environment. Chapman & Hall/CRC. https://doi.org/10.1201/9780203498880.

He, F., & Gaston, K.J. (2000). Estimating species abundance from occurrence. Am. Nat., 156 (5). 553–559. ISSN 0003–0147.

He, F., Gaston, K., Wu, J., 2002. On species occupancy-abundance models. Écoscience 9 (1), 119–126. https://doi.org/10.1080/11956860.2002.11682698.

Heagerty, P.J., Lele, S.R., 1998. A Composite Likelihood Approach to Binary Spatial Data. J. Am. Stat. Assoc. 93 (443), 1099–1111. https://doi.org/10.2307/2669853.

Heeringa, S.G., West, B.T., Berglund, P.A., 2010. Applied Survey Data Analysis. Chapman and Hall/CRC, Boca Raton. https://doi.org/10.1201/9781420080674.

Hoem, S., 2022. Norwegian Biodiversity Information Centre - Other Datasets. Version 13.236. The Norwegian Biodiversity Information Centre (NBIC). https://doi.org/10.15468/tm56sc. Occurrence dataset.

Holt, A.R., Gaston, K.J., He, F., 2002. Occupancy-abundance relationships and spatial distribution: a review. Basic Appl. Ecol. 3 (1), 1–13. https://doi.org/10.1078/1439-1791-00083.

Horvitz, D.G., Thompson, D.J., 1952. A generalization of sampling without replacement from a finite universe. J. Am. Stat. Assoc. 47 (260), 663–685. https://doi.org/10.1080/01621459.1952.10483446.

Huggins, R., Hwang, W.-H., Stoklosa, J., 2018. Estimation of abundance from presence–absence maps using cluster models. Environ. Ecol. Stat. 25, 495–522. https://doi.org/10.1007/s10651-018-0415-5.

Hwang, W.-H., Huggins, R., 2016. Estimating abundance from presence–absence maps via a paired Negative-Binomial Model. Scand. J. Stat. 43 (2), 573–586. https://doi.org/10.1111/sjos.12192.

Hwang, W.-H., Huggins, R., Stoklosa, J., 2022. A model for analyzing clustered occurrence data. Biometrics 78 (2), 598–611. https://doi.org/10.1111/biom.13435.

Kennedy, K.A., Addison, P.A., 1987. Some considerations for the use of visual estimates of plant cover in biomonitoring. J. Ecol. 151-157 https://doi.org/10.2307/2260541.
Kercher, S.M., Frieswyk, C.B., Zedler, J.B., 2003. Effects of sampling teams and estimation methods on the assessment of plant cover. J. Veg. Sci. 14 (6), 899–906. https://doi.org/10.1111/j.1654-1103.2003.tb02223.x.
Lindenmayer, D.B., Welsh, A., Donnelly, C., Crane, M., Michael, D., Macgregor, C., McBurney, L., Montague-Drake, R., Gibbons, P., 2009. Are nestboxes a viable alternative source of cavities for hollow-dependent animals? Long-term monitoring of nest box occupancy, pest use and attrition. Biol. Conserv. 142, 33–42. https://doi.org/10.1016/j.biocon.2008.09.026.
Margolis, H., Nelson, R., Montesano, P., Beaudoin, A., Sun, G., Andersen, H.-E., Wulder, M., 2015. Combining Satellite lidar, Airborne lidar and Ground Plots to Estimate the Amount and Distribution of Aboveground Biomass in the Boreal Forest of North America. Can. J. For. Res. 45 (7), 838–855. https://doi.org/10.1139/cjfr-2015-0006.
McRoberts, R.E., Næsset, E., Liknes, G.C., Chen, Q., Walters, B.F., Saatchi, S., Herold, M., 2019. Using a Finer Resolution Biomass Map to Assess the Accuracy of a Regional, Map-Based Estimate of forest Biomass. Surv. Geophys. 40 (4), 1001–1015. https://doi.org/10.1007/s10712-019-09507-1.
Nelson, R., Gobakken, T., Næsset, E., Gregoire, T.G., Ståhl, G., Holm, S., Flewelling, J., 2012. Lidar sampling - using an airborne profiler to estimate forest biomass in Hedmark County, Norway. Remote Sens. Environ. 123, 563–578. https://doi.org/10.1016/j.rse.2011.10.036.
O'Connor, B., Bojinski, S., Röösli, C., Schaepman, M.E., 2020. Monitoring global changes in biodiversity and climate essential as ecological crisis intensifies. Eco. Inform. 55, 101033 https://doi.org/10.1016/j.ecoinf.2019.101033.
Olsson, B., 2020. National Land Cover Database. Swedish Environmental Protection Agency, Stockholm. https://www.naturvardsverket.se/en/services-and-permits/maps-and-map-services/national-land-cover-database/.
Pain, D.J., Bardin, P., Hutchinson, N., Pénzesné Kónya, E., Krause, M., 2020. A review of European progress towards the Global Strategy for Plant Conservation 2011-2020. Planta Eur. Plantlife Int. XXXpp.
Phillips, S.J., Anderson, R.P., Dudík, M., Schapire, R.E., Blair, M.E., 2017. Opening the black box: an open-source release of Maxent. Ecography 40, 887–893. https://doi.org/10.1111/ecog.03049.
Pielou, E.C., 1977. Mathematical Ecology. Wiley.
R Core Team, 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL. https://www.R-project.org/.
Ringvall, A., Petersson, H., Ståhl, G., Lämås, T., 2005. Surveyor consistency in presence/absence sampling for monitoring vegetation in a boreal forest. For. Ecol. Manag. 212 (1–3), 109–117. https://doi.org/10.1016/j.foreco.2005.03.002.
Royle, J.A., Dorazio, R.M., 2008. Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities. Academic Press, London. https://doi.org/10.1016/B978-0-12-374097-7.50001-5.
Saarela, S., Schnell, S., Grafström, A., Tuominen, S., Nordkvist, K., Hyyppä, J., Kangas, A., Ståhl, G., 2015. Effects of sample size and model form on the accuracy of model-based estimators of growing stock volume. Can. J. For. Res. 45 (11), 1524–1534. https://doi.org/10.1139/cjfr-2015-0077.
Saarela, S., Holm, S., Healey, S.P., Patterson, P.L., Yang, Z., Andersen, H.-E., Dubayah, R. O., Qi, W., Duncanson, L.I., Armston, J.D., Gobakken, T., Næsset, E., Ekström, M., Ståhl, G., 2022. Comparing frameworks for biomass prediction for the global ecosystem dynamics investigation. Remote Sens. Environ. 278 https://doi.org/10.1016/j.rse.2022.113074.
Särndal, C.-E., Swensson, B., Wretman, J., 1992. Model Assisted Survey Sampling. Springer. https://doi.org/10.1007/978-1-4612-4378-6.
Sauerbrei, W., Royston, P., 1999. Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials. J. R. Stat. Soc. Ser. A 162 (1), 71–94. https://doi.org/10.1111/1467-985X.00122.
Sen, P.K., Singer, J.M., 1993. Large Sample Methods in Statistics: An Introduction with Applications. Chapman & Hall, New York.
Solow, A.R., Smith, W.K., 2010. On predicting abundance from occupancy. Am. Nat. 176 (1), 96–98. https://doi.org/10.1086/653077.
Sreekumar, E.R., Nameer, P.O., 2022. A MaxEnt modelling approach to understand the climate change effects on the distributional range of white-bellied Sholakili Sholicola albiventris (Blanford, 1868) in the Western Ghats, India. Ecol. Inform. 70 https://doi.org/10.1016/j.ecoinf.2022.101702.
Ståhl, G., 2003. Presence/absence sampling as a substitute for cover assessment in vegetation monitoring. In: Corona, P., Köhl, M., Marchetti, M. (Eds.), Advances in Forest Inventory for Sustainable Forest Management and Biodiversity Monitoring. Kluwer Academic Publishers, pp. 137–142. https://doi.org/10.1007/978-94-017-0649-0_11.
Ståhl, G., Holm, S., Gregoire, T.G., Gobakken, T., Næsset, E., Nelson, R., 2011. Model-based inference for biomass estimation in a LiDAR sample survey in Hedmark County, Norway. Can. J. For. Res. 41 (1), 96–107. https://doi.org/10.1139/X10-161.
Ståhl, G., Saarela, S., Schnell, S., Holm, S., Breidenbach, J., Healey, S.P., Patterson, P.L., Magnussen, S., Næsset, E., McRoberts, R.E., Gregoire, T.G., 2016. Use of models in large-area forest surveys: comparing model-assisted, model-based and hybrid estimation. Forest Ecosyst. 3 (5) https://doi.org/10.1186/s40663-016-0064-9.
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 Ecol. Evol. 8 (10), 1284–1291. https://doi.org/10.1111/2041-210X.12749.
Stoklosa, J., Blakey, R.V., Hui, F.K.C., 2022. An Overview of Modern Applications of Negative Binomial Modelling in Ecology and Biodiversity. Diversity 14 (5), 320. https://doi.org/10.3390/d14050320.
Tillé, Y., 2006. Sampling Algorithms. Springer, New York. ISBN 0-387-30814-8.
Tomppo, E., Gschwantner, T., Lawrence, M., McRoberts, R.E. (Eds.), 2010. National Forest Inventories. Pathways for Common Reporting, 1. European Science Foundation, pp. 541–553. https://doi.org/10.1007/978-90-481-3233-1.
Waagepetersen, P., 2007. An estimating function approach to inference for inhomogeneous Neyman–Scott processes. Biometrics 63, 252–258. https://doi.org/10.1111/j.1541-0420.2006.00667.x.
Wan, J.-Z., Wang, C.-J., Yu, F.-H., 2017. Wind effects on habitat distributions of wind-dispersed invasive plants across different biomes on a global scale: assessment using six species. Eco. Inform. 42, 38–45. https://doi.org/10.1016/j.ecoinf.2017.09.002.
Warton, D.I., Foster, S.D., De'ath, G., Stoklosa, J., Dunstan, P.K., 2015. Model-based thinking for community ecology. Plant Ecol. 216, 669–682. https://doi.org/10.1007/s11258-014-0366-3.
Wright, D.H., 1991. Correlations Between Incidence and Abundance are Expected by Chance. J. Biogeogr. 18 (4), 463–466. https://doi.org/10.2307/2845487.
Yee, T.W., Mitchell, N.D., 1991. Generalized additive models in plant ecology. J. Veg. Sci. 2, 587–602. https://doi.org/10.2307/3236170.