Ecological Informatics 18 (2013) 35–39

Contents lists available at ScienceDirect

Ecological Informatics

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

# Hierarchical and spatially aggregated plant cover data

Christian Damgaard *

*Department of Bioscience, Aarhus University, Vejlsøvej 25, 8600 Silkeborg, Denmark*

ARTICLE INFO

*Article history:*
Received 5 February 2013
Received in revised form 28 May 2013
Accepted 2 June 2013
Available online 10 June 2013

*Keywords:*
Distribution of plant abundance
Hierarchical models
Plant ecology
Point-intercept
Statistical power

ABSTRACT

Most plant species are spatially aggregated and here the importance of taking the spatial variation into account when analyzing plant cover data is demonstrated in a general stochastic model where both the within-site and the among-site spatial variation of species cover data are parameterized. Using a generalised binomial distribution (or Pólya–Eggenberger distribution), where the among-site variation in mean cover is modeled by a zero-inflated beta distribution, it is possible to adequately analyze hierarchical plant cover data and link the estimates to the underlying ecological processes. The model is demonstrated in a case-study of pin-point cover data of *Erica tetralix* from 1148 wet heathland plots at 84 Danish sites, and it is shown that both parameter estimates and the conclusions of hypotheses testing critically depend on the correct modeling of the observed spatial variation. Finally, statistical power simulations of plant cover measurements are presented, which will be useful for planning ecological experiments and monitoring programs.



## 1. Introduction

The abundance of species is controlled by both local and regional environmental and spatial processes (e.g. Cottenie, 2005; Hanski, 1998), and detailed quantitative knowledge on the distribution of plant species is a prerequisite for understanding the role of different abiotic and biotic factors in the distribution and metapopulation dynamics of species abundance. Consequently, it is an important and often performed task in basic and applied plant ecological research to describe and compare plant abundance among different plant communities, among different treatments, or along environmental gradients. This task, which typically is complicated by non-normally distributed abundance data including zero values, has been solved by a number of different statistical methods, including analysis of variance and other parametric and non-parametric methods (e.g. Bolker et al., 2009; Eskelson et al., 2011; Strien and Meelis, 1990; Warton and Hui, 2011) and, as will be demonstrated in this paper, the conclusions of the analysis depend critically on the method used in the analysis. It will, consequently, be valuable to establish a set of best practices when analyzing plant abundance data.

Plant abundance may be described by the cover, i.e. the relative projected area covered by a species. Plant cover takes the size of individuals into account and is an important and often measured characteristic of the composition of plant communities (Kent and Coker, 1992). The most common way to measure plant cover in herbal plant communities is to make a visual assessment of the relative area covered by the different species in a small circle or quadrate (Kent and Coker, 1992). However, an alternative more objective methodology, called the pin-point method (or point-intercept method), has been widely employed (Kent and Coker, 1992; Levy and Madden, 1933). In a pin-point analysis, a pin is inserted vertically through one of the grid points in a frame with a fixed grid pattern into the vegetation. The different species that are hit by the pin are recorded and the cover of the plant species is defined as the relative number of hits.

In a previous paper the analysis of longitudinal pin-point cover data was described (Damgaard, 2012) and here I will present a method for investigating the spatial variation of hierarchical pin-point cover data. The cover of many plant species have been shown to have an aggregated spatial pattern due to e.g. the size of the plant, clonal growth, and limited seed dispersal (Chen et al., 2006, 2008; Herben et al., 2000; Pacala and Levin, 1997; Stoll and Weiner, 2000), and pin-point plant cover data at the local scale will, typically, be over-dispersed relative to the binomial distribution (Damgaard, 2008, 2009). Here, the within-site spatial aggregation in pin-point plant cover data will be modeled using the Pólya–Eggenberger distribution. This model allows for an augmented variance compared to a binomial distribution and has previously been shown to fit the cover of many herbal species well (Damgaard et al., 2011).

Furthermore, if plant cover data are sampled using a hierarchical sampling procedure where plots are sampled from a number of different sites, then the possible among-site variation in plant cover must be taken into account as well. The among-site variation in mean cover is thought to arise by different plant ecological processes: i) plant species do not occur everywhere possible due to metacommunity dynamics, i.e. in some sites a plant species may be absent due to random extinction events and/or limited possibility of the plant to colonise the habitat (Cordonnier et al., 2006; Leibold et al., 2004; MacArthur and Wilson, 1967; Rees et al., 2001), or ii) the mean plant cover at a site may vary due to an underlying variation in abiotic and biotic factors or due to random stochastic perturbations of species cover (e.g. Adler et al., 2007;

* Tel.: +45 30183153.
*E-mail address:* cfd@dmu.dk.

1574-9541/$ – see front matter 
http://dx.doi.org/10.1016/j.ecoinf.2013.06.001

Hubbell, 2001). Here, the among-site variation in mean plant cover will be modeled by a zero-inflated beta distribution (Ospina and Ferrari, 2010).

Generally, I will demonstrate the importance of taking the spatial variation of plant species into account when analyzing species cover data by modeling the ecological processes at both the within- and among-site spatial scale. Furthermore, a set of best practices when analyzing plant cover data will be proposed, and power simulations that will be useful for planning plant ecological experiments and monitoring programs will be presented. The importance of taking the within-site spatial aggregation of plant species into account and the need to use a set of best practices will be illustrated by a case-study of pin-point cover data of *Erica tetralix* on Danish wet heathlands.

## 2. Materials and methods

### 2.1. The distribution of pin-point plant cover data

The measurement of pin-point cover data is a binomial process where a pin in the pin-point frame either hits or does not hit the plant species. A discrete stochastic variable $Y$ may be defined as the number of pins out of $n$ possible pins in the pin-point frame that hit the plant species, and since the distribution of plant species at a site typically is spatially aggregated, $Y$ is assumed to be generated by a generalised binomial distribution (or Pólya–Eggenberger distribution) with probability parameter $q$ and intra-plot correlation parameter $\delta$ (Damgaard, 2008, 2009, 2012):

$$f(y;n,q,\delta) = \binom{n}{y} \frac{\varphi(q(\frac{1}{\delta}-1),y)\varphi\left(\frac{(1-q)(1-\delta)}{\delta},n-y\right)}{\varphi(\frac{1}{\delta}-1,n)} \quad (1)$$

where $\varphi$ is the Pochhammer function, $\varphi(x,n) = \Gamma(x+n)/\Gamma(x) = x(x+1)\ldots(x+n-1)$. The mean of the generalised binomial distribution is independent of $\delta$, $E(Y) = nq$, and $Var(Y) = n(1-q)q(1-\delta(1-n))$.

The probability density function of the Pólya–Eggenberger distribution (Eq. (1)) is equal to the beta-binomial distribution $f(y;n,\nu) = Binomial(n,\nu)$, where $\nu \sim Beta(q/\delta - q, (1-q)(1-\delta)/\delta)$, but is somewhat more general in that negative intra-plot correlation is allowed (Qu et al., 1993).

If more sites are sampled, then most likely there will be some variation in the mean cover ($q$) among the sites. This among-site variation is here modeled by a zero-inflated beta distribution, i.e., a mixture distribution that with probability $\gamma$ draws from a stochastic zero process that models the probability that the species is absent from the site, and with probability $(1-\gamma)$ draws from a re-parameterised beta distribution that model the among-site variation in mean cover if the species is present at the site. The chosen re-parameterised beta distribution has the mean cover as one of the two parameters, which I find to be practical for many applied and testing purposes. Consequently the overall distribution of the pin-point data at both spatial levels is modeled by $g(y;n,c,\mu,\upsilon,\gamma,\delta) = f(y;n,q,\delta)$, where the mean cover at the site ($q$) is modeled by:

$$q \sim Mixture((\gamma, 1-\gamma),(Uniform(0,c), Beta(\mu\nu,(1-\mu)\nu)) \quad (2)$$

where $\mu$ is the mean cover of all sites where the species is present, $\nu$ is the scale parameter in the re-parameterised beta distribution, and $c$ is a small constant that ensures that $q$ is a continuous variable. The used parameters are summarised in Table 1.

### 2.2. Two-step fitting procedure

The generalised binomial distribution model of pin-point data, where the mean cover of the sites is modeled by a zero-inflated beta distribution (Eq. (2)) was fitted using a two-step procedure: first, the parameter of the zero-process ($\gamma$) and the scale parameter of the re-parameterised beta distribution ($\nu$) were estimated from the mean cover at the different sites by the weighted likelihood function of the mixture distribution (Eq. (2)), where the mean cover was weighted by the sample size. Using the resulting maximum likelihood estimates of $\gamma$ and $\nu$, the mean cover ($\mu$) and the intra-plot correlation parameter ($\delta$) is estimated using generalised binomial distribution (Eq. (1)) from the pin-point data at the plot level.

**Table 1**
The parameters used in Eqs. (1) and (2).

| Parameter | Description |
|---|---|
| $y$ | Number of times that the plant species is hit in a pin point frame. |
| $n$ | Number of grid points in the grid point frame |
| $q$ | Mean plant cover |
| $\gamma$ | The probability that the species absent from a site. |
| $\mu$ | The site mean cover if the species is present at the site. |
| $\nu$ | The scale parameter of the distribution of the site mean cover if the species is present at the site assuming the site mean cover are distributed according to a re-parameterized beta distribution. |
| $\delta$ | The intra-plot correlation parameter that measures the within-site spatial aggregation of the plant species as the correlation between the outcomes of successive Bernoulli trials (Qu et al., 1993). The parameter is bounded between $-min\left(\frac{q}{n-1-q}, \frac{1-q}{n-1}\right)$ and 1. For plant species that tend to be spatially aggregated, the number of hits within a pin-point frame is positively correlated. In this case ($\delta > 0$), the variance of the number of hits will be augmented relative to the binomial distribution. The hypothesis of no correlation binomial distributed hits may be tested in a likelihood ratio test. |

### 2.3. Example: the cover of E. tetralix on wet heathlands

In order to illustrate the importance of including the effect of spatial aggregation in the analysis of plant cover data, the distribution of pin-point data of *E. tetralix* in wet heathlands was fitted to the above model and a subset of the data was regressed to the measured nitrogen content in the leaves.

*E. tetralix* is a perennial subshrub, which is a characteristic plant species of humid, peaty or semi-peaty heaths (Northern Atlantic wet heaths (4010), EU, 2003). Only data from vegetation plots that were classified as wet heathland (4010), according to the habitat classification system used for the European Habitat Directive (EU, 2003) by the classification method outlined in Nygaard et al. (2009), was used in the analyses.

Pin-point cover data was sampled from 1148 wet heathland plots at 84 Danish sites using a square frame of 16 grid points that were equally spaced by 10 cm (Nielsen et al., 2012) in an unbalanced design in the period from 2004 to 2009, i.e. at some sites there were relatively many plots and at some sites there were relatively few plots. The data from these plots were used to fit the distribution of the cover of *E. tetralix* to the model where both the among-site and the within site variation was taken into account.

In 165 of the 1148 wet heathland plots, the nitrogen content in the leaves in a neighboring dwarf shrub plant (*E. tetralix* or *Calluna vulgaris*) had been determined (measured as percent of dry matter), and the data from these 165 plots were used in the regression analysis. The plant cover data was regressed to the nitrogen content by maximising the likelihood functions of the data assuming they were distributed according to the generalised binomial distribution (Eq. (1)). In the regression analysis the cover ($q$) was assumed to be a function of the nitrogen content in the leaves and modeled by a monotonic sigmoid model (Damgaard, 2008).

The investigated data are a small subset of the ecological data that is collected within the Danish monitoring program NOVANA (Nielsen et al., 2012).