Ecological Informatics 26 (2015) 18–26

Contents lists available at ScienceDirect

Ecological Informatics

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

Analyzing plant cover class data quantitatively: Customized zero-inflated cumulative beta distributions show promising results

Basile Herpigny, Frédéric Gosselin *

IRSTEA, UR EFNO, Centre de Nogent-sur-Vernisson, Domaine des Barres, F-45290 Nogent-sur-Vernisson, France

ARTICLE INFO

Article history:
Received 9 September 2014
Received in revised form 12 December 2014
Accepted 14 December 2014
Available online 19 December 2014

Keywords:
Ordered multinomial models
Cumulative logit
Zero inflation
Bayesian statistics
Plant abundance
Braun-Blanquet

ABSTRACT

Although parametric statistical methods have several advantages over ordination methods, understory plant cover class data are traditionally more often analyzed with ordination techniques than with parametric ones. Among the latter, only the cumulative logit model can take into account all the peculiarities of cover data: bounded between 0 and 100%, asymmetric classes, high proportion of zeroes. However, results provided by the cumulative logit model are difficult to interpret. We tested ten Bayesian models based on a zero-inflated cumulative beta probability distribution which is bounded, can assume various shapes and accounts for zeroes. Some of these models also make results easier to interpret by allowing the user to directly estimate the mean and variance of data underlying cover class observations, much as in generalized linear models (GLMs). We applied our new models and the cumulative logit model to real data, then compared their performance using the Deviance Information Criterion (DIC) and sampled posterior *p*-values.

Four of the Bayesian beta models performed better (lower DIC), as well or rarely worse (depending on species) than the cumulative logit model and showed an ease of interpretation similar to that of GLMs. They therefore provide promising alternatives to existing parametric methods for modeling plant cover class data.

1. Introduction

Estimating plant abundance through visual assessment of cover is widely used in ecology to compare communities at the species, life form or functional group level. It is a fast and non-destructive method, which requires far less time than other non-destructive methods such as the pin-point method (Kent and Cocker, 1992; Levy and Madden, 1933) and less energy than harvesting and measuring biomass (Chen et al., 2008). One of the most commonly used methods of plant cover visual estimation is based on the Braun-Blanquet scale (Braun-Blanquet, 1932). Eight abundance classes are distinguished: 0, r, +, 1, 2, 3, 4 and 5—respectively corresponding to absence of the species, negligible cover, less than 0.1% cover, between 0.1 and 5% cover, between 5 and 25% cover, between 25 and 50% cover, between 50 and 75% cover, and more than 75% cover (Braun-Blanquet, 1932). Braun-Blanquet's cover data can provide plant cover estimates similar to those obtained with the pin-point method (Damgaard, 2014).

In the literature, a whole range of statistical techniques have been used to analyze Braun-Blanquet's cover data. Most studies rely on various forms of ordination such as Correspondance Analysis and its variations (Cilliers and Bredenkamp, 2000; Hardtle et al., 2005; Islebe and Velazquez, 1994; Lepš and Hadincová, 1992; Peinado et al., 1998; Pysek, 1994; Velazquez and Islebe, 1995; Wolf, 1993). There are many fewer analyses of Braun-Blanquet cover data with parametric statistical models (Chen et al., 2006; Eskelson et al., 2011; Van Couwenberghe et al., 2013). This is because Braun-Blanquet data are complex in several ways. First, Braun-Blanquet cover data form discrete ordered classes of different sizes. Moreover, due to the way plant species are distributed, the data will often include an important amount of zeroes, i.e. sites where the species is observed as absent, either because of random extinction events, of the inability of the species to colonize the site (Damgaard, 2009), or simply because the species was not detected. Due to their complexity, cover data have often been replaced by presence–absence data in biodiversity studies, even though this process results in the loss of useful information (van der Maarel, 1979). The classes on the Braun-Blanquet scale are also often transformed into mean cover indices to facilitate the analysis of the data (van der Maarel, 1979; Van Der Maarel, 2007). This transformation switches the data from an ordinal to a metric scale. Furthermore, the transformation is arbitrary and introduces a distortion between classes (for example, 2.5% will designate all percentages between 0.1 and 5%, whereas 87.5% will designate a much larger array of percentages, namely 75 to 100%) (Podani, 2006).

Even though they are complex, using parametric models rather than ordination to analyze Braun-Blanquet data has several advantages. First, modeling under complex conditions (heteroscedasticity, interdependent data) is more complicated when using non-parametric models (Gosselin, 2011b; Laara, 2009). Second, parametric models provide the magnitude of the effects and a level of uncertainty thus allowing

* Corresponding author. Tel.: +33 2 38 95 03 58.
E-mail address: frederic.gosselin@irstea.fr (F. Gosselin).

http://dx.doi.org/10.1016/j.ecoinf.2014.12.002
1574-9541/

comparisons between studies (Anderson et al., 2001; Richard, 2004; Yoccoz, 1999), whereas non-parametric models provide only the probability of the significance of the effects (Harrell, 2001). Finally, even though parametric models make assumptions about data distribution, non-parametric models also make statistical assumptions (often assuming independence, for example). If these statistical assumptions are not valid, non-parametric models may turn out to be less robust than their parametric counterparts (Johnson, 1995; McArdle and Anderson, 2004; Richard, 2004). To sum up, parametric models allow for confirmatory analysis (hypothesis testing) and estimation of the effects, whereas ordination techniques are better suited for exploratory analysis (hypotheses generation) (Gosselin and Gosselin, 2004; Richard, 2004).

The cumulative logit model (Liu and Agresti, 2005) can take into account all the peculiarities of cover class data: bounded between 0 and 100%, asymmetric classes, high proportion of zeroes. However it is rather difficult to interpret and has seldom been used in ecology (cf Section 2.2.2). As an alternative, beta binomial and beta distributions have already been used to model count based or continuous cover data (Chen et al., 2006; Chen et al., 2008). To address the often high proportion of zeroes in cover data, a zero-inflated beta distribution has also been used by Damgaard (2009) and Damgaard (2013) to model different kinds of cover data. The model consists in a dual process: first modeling the number of zeroes, then modeling the rest of the data with a beta distribution. These studies used a non-regression setting to model cover data. Continuous cover data have also been modeled with a beta distribution in a regression setting in response to ecological variables (e.g. tree cover) (Eskelson et al., 2011). In this case, however, data were transformed so as to be in the] 0,1 [interval, therefore eluding the need to use a zero-inflated model.

Our objective is to present a new method for modeling the response of Braun-Blanquet data (or other ordered class data) to ecological variables that addresses the specificities of such vegetation data. This method is based on a probability distribution mixing a Dirac distribution on zero and a cumulative beta distribution. It builds on the work of Chen et al. (2006), Damgaard (2009), and Eskelson et al. (2011), but it introduces the following novelties: (i) modeling the zero-inflation probability as a function of the mean predicted cover, (ii) permitting the estimation of the limits of the latent cover classes corresponding to the Braun Blanquet classes, and (iii) proposing and testing different model parameterizations. The originality of our approach is that we modeled cover data in a regression setting, with several models of zero-inflated beta distribution, while including an ecological component in the statistical model to assess the response of specific cover to dendrometric variables under different environmental conditions.

In the following, we start by presenting the zero-inflated cumulative beta distribution and the reference model to which it will be compared: the cumulative logit. These models are then applied on real data through Bayesian methods and compared with DIC (Deviance Information Criterion) and sampled posterior p-values—a goodness of fit p-value (Gosselin, 2011a).

2. Material and methods

2.1. Data

The models were compared using the same data set as in Zilliox and Gosselin (2014). The data were collected from 2006 to 2010 by the French National Forest Inventory (NFI; cf. Morneau et al., 2008) on plots dominated by Norway spruce and silver fir (*Picea abies* (L.) Karst and *Abies alba* Mill.) in the Alps and the Jura regions. We excluded data from winter measurements or when snow or frost covered the soil (inappropriate conditions for floristic surveys) and data from simplified plots (i.e. reduced-size plots for which the size reduction was unknown). After data removal, 475 plots were left. For each plot, the available data include floristic, dendrometric and environmental measurements.

2.1.1. Floristic data

On each plot the NFI floristic data consists in the cover class of each vascular plant species detected within a 15 m-radius disk (extended to 25 m-radius for tree species that include a tree with a diameter at breast height greater than 7.5 cm). The cover data distinguished only 6 cover classes: 0, 1, 2, 3, 4 and 5 (absence, less than 5%, between 5 and 25%, between 25 and 50%, between 50 and 75%, more than 75% cover). We used only these classes in our models. We analyzed cover data only for the seventeen most abundant species of the study area (Table 1). Most of them presented a mean cover between 5 and 10% in plots of the study area, with the notable exceptions of *A. alba* Mill. and *P. abies* (L.) Karst (30 and 40% respectively). The distribution of cover classes was similar for most species, with class 0 or class 1 being the most frequent class, and class 5 being very rare. Again, *A. alba* Mill. and *P. abies* (L.) Karst are exceptions, with class 4 being the most frequent for both species.

2.1.2. Environmental data

The meteorological data were obtained from MeteoFrance and included monthly mean temperatures (abbreviated as T in the following equations) and precipitation (ppt) for the 2005–2010 period. In addition to these climatic data, global solar radiation (solrad) was calculated

Table 1
Summary of the cover data of the seventeen most abundant species of the study area, including mean cover (calculated from class data, assigning to each class a cover value equal to the mean value of the class) ± SD, and empirical probability of each Braun Blanquet cover class from 0 to 5, corresponding respectively to absence, less than 5%, between 5 and 25%, between 25 and 50%, between 50 and 75%, and more than 75% cover.

Species name	Species code	Mean cover (%)	P(Y = 0)	P(Y = 1)	P(Y = 2)	P(Y = 3)	P(Y = 4)	P(Y = 5)
Abies alba Mill.	abal	33.06 ± 30.47	0.18	0.17	0.15	0.13	0.26	0.10
Acer pseudoplatanus L.	acps	4.64 ± 6.64	0.32	0.50	0.17	0.01	0.00	0.00
Ajuga reptans L.	ajre	3.69 ± 6.38	0.48	0.37	0.14	0.01	0.00	0.00
Carex sylvatica Huds.	casy	4.45 ± 7.86	0.52	0.30	0.16	0.02	0.00	0.00
Corylus avellana L.	coav	8.44 ± 11.99	0.38	0.28	0.27	0.06	0.01	0.00
Dryopteris filix-mas (L.) Schott	drfi	4.35 ± 6.94	0.41	0.42	0.16	0.01	0.00	0.00
Fagus sylvatica L.	fasy	12.15 ± 15.07	0.29	0.25	0.31	0.11	0.03	0.00
Fraxinus excelsior L.	frex	4.57 ± 8.9	0.50	0.33	0.14	0.02	0.01	0.00
Fragaria vesca L.	frve	5.47 ± 8.4	0.36	0.42	0.19	0.02	0.00	0.00
Galium odoratum (L.) Scop	gaod	6.27 ± 10.15	0.50	0.23	0.22	0.04	0.00	0.00
Oxalis acetosella L.	oxac	9.05 ± 13.64	0.46	0.19	0.27	0.07	0.02	0.00
Picea abies (L.) Karst	piab	41.90 ± 28.46	0.10	0.07	0.20	0.20	0.30	0.14
Rubus fruticosus L.	rufr	9.37 ± 17.1	0.45	0.26	0.18	0.05	0.03	0.01
Rubus idaeus L.	ruid	6.25 ± 9.54	0.37	0.38	0.20	0.04	0.00	0.00
Sorbus aria (L.) Crantz	soar	2.95 ± 5.91	0.54	0.35	0.09	0.01	0.00	0.00
Sorbus aucuparia L.	soau	6.15 ± 8.74	0.33	0.43	0.22	0.03	0.00	0.00
Vaccinium myrtillus L.	vamy	10.32 ± 17.27	0.54	0.12	0.20	0.11	0.02	0.01