



Traditional occupancy–abundance models are inadequate for zero-inflated ecological count data

Gudeta Sileshi^{a,*}, Girma Hailu^b, Gerson I. Nyadzi^c

^a World Agroforestry Centre (ICRAF), SADC-ICRAF Agroforestry Programme, Chitedze Agricultural Research Station, P.O. Box 30798, Lilongwe, Malawi

^b 2193 Espirt Dr, K4A, 4Z1, Orleans, Ontario, Canada

^c UN Millennium Villages Project, P.O. Box 1561, Tabora, Tanzania

ARTICLE INFO

Article history:

Received 10 December 2007

Received in revised form 20 February 2009

Accepted 17 March 2009

Available online 21 May 2009

Keywords:

Model selection

Parsimony

Rare species

Repeated measures

ABSTRACT

Traditional occupancy–abundance and abundance–variance–occupancy models do not take into account zero-inflation, which occurs when sampling rare species or in correlated counts arising from repeated measures. In this paper we propose a novel approach extending occupancy–abundance relationships to zero-inflated count data. This approach involves three steps: (1) selecting distributional assumptions and parsimonious models for the count data, (2) estimating abundance, occupancy and variance parameters as functions of site- and/or time-specific covariates, and (3) modelling the occupancy–abundance relationship using the parameters estimated in step 2. Five count datasets were used for comparing standard Poisson and negative binomial distribution (NBD) occupancy–abundance models. Zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) occupancy–abundance models were introduced for the first time, and these were compared with the Poisson, NBD, He and Gaston's and Wilson and Room's abundance–variance–occupancy models. The percentage of zero counts ranged from 45 to 80% in the datasets analysed. For most of the datasets, the ZINB occupancy–abundance model performed better than the traditional Poisson, NBD and Wilson and Room's model. He and Gaston's model performed better than the ZINB in two out of the five datasets. However, the occupancy predicted by all models increased faster than the observed as density increased resulting in significant mismatch at the highest densities. Limitations of the various models are discussed, and the need for careful choice of count distributions and predictors in estimating abundance and occupancy parameter are indicated.

© 2009 Published by Elsevier B.V.

1. Introduction

A positive occupancy–abundance and abundance–variance relationship has been widely documented, both intra- and inter-specifically, at a range of spatial scales for a diverse array of animal and plant species (Brown, 1984; Gaston et al., 2000, 2006; He et al., 2002; Taylor, 1961). Since the first comprehensive treatment of the occupancy–abundance relationship (Brown, 1984), it has become a general mathematical expectation that the occupancy–abundance relationship will always be positive, although occasional zero and negative correlations have been reported (Gaston et al., 2000; Wilson, 2008). This relationship has received particular attention in the context of meta-population dynamics, conservation biology, agricultural entomology and epidemiology (Anderson and May, 1985; Gaston et al., 2006; Wilson and Room, 1983).

A suite of empirical and theoretical models has been widely employed to describe the occupancy–abundance rela-

tionship in various fields (He and Gaston, 2003). However, most are special forms of the negative binomial distribution (NBD) occupancy–abundance model (He and Gaston, 2003). Recently, He and Gaston (2003) derived a general abundance–variance–occupancy model by combining the abundance–variance relationship described by Taylor's power law (TPL) (Taylor, 1961) and the NBD occupancy–abundance model. The abundance–variance–occupancy model arguably has much wider ecological significance from the perspective of pattern unification and, as such, it may help in fundamental understanding of spatial variation in abundance (He and Gaston, 2003). However, this model assumes perfect detection of species, and occupancy and abundance to be temporally and spatially invariant (He and Gaston, 2003). In their current form, all the occupancy–abundance models also do not take into account zero-inflation and its impacts on estimates of abundance, variance and occupancy from count data. Therefore, there is a need to develop more robust models that account for zero-inflation, which may arise from various sources.

A wide range of ecological count data exhibit zero-inflation (Cunningham and Lindenmayer, 2005; Gray, 2005; Martin et al., 2005; Sileshi, 2006, 2008; Warton, 2005), and such data do not readily fit standard distributions such as the NBD (Hall, 2000). Two

* Corresponding author. Tel.: +265 1707329.

E-mail addresses: sgwelde@yahoo.com (G. Sileshi), girmahailuus@yahoo.com (G. Hailu), gerson.nyadzi@undp.org (G.I. Nyadzi).

types of zeros are often encountered in count data: structural zeros which are inevitable, and sampling zeros which occur by chance. Structural zeros consist of a large number of true zeros which arise when presence is not tenable (Cunningham and Lindenmayer, 2005). These are caused by the real ecological effects of interest (Martin et al., 2005). For example, the study of rare organisms will often lead to the collection of data with a high frequency of zeros (Welsh et al., 1996). Within almost all communities the vast majority of species are rare. Yet such species will frequently be of ecological, conservation or management interest in part because they may be among the extinction-prone taxa in an assemblage (Cunningham and Lindenmayer, 2005). Sampling zeros are random, and arise due to sampling where conditions are potentially suitable but absence is observed. False zeros (MacKenzie et al., 2002) occur when the species under study is present at the time of sampling, but the observer does not detect it because of its cryptic or secretive nature. Therefore, for rare species with low detection probability, excess zeroes could be substantial.

In this paper we illustrate a novel method for modelling the occupancy–abundance relationship for species with patchy distributions and, therefore, zero-inflated count datasets. We also propose two new occupancy–abundance models based on the zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) distributions. We then compare these models with the traditional occupancy–abundance models derived from the Poisson and negative binomial distribution (NBD), and with two abundance–variance–occupancy models derived from Taylor's power law. We illustrate the use of information criteria for model selection, and discuss the limitations of the various models.

2. Methods

2.1. The data

Five count datasets with varying levels of zero-inflation were used in this analysis. These datasets are by no means the most representative of zero-inflated counts. They were only used to illustrate the analytical methods proposed in Section 2.2. The first dataset consisted of counts of adults of the chrysomelid beetle *Mesoplatys ochroptera* Stål in western Kenya. This species was monitored in two experiments established during 1999–2000 and 2000–2001, each consisting of three agroforestry treatments consisting of pure *Sesbania sesban* (L.) Merrill, a mixture of *S. sesban* and *Tephrosia vogelii* Hook and *S. sesban* and *Crotalaria grahamiana* Whight & Arn. The sites were Dudi and Khumusalaba in Butere district, Mutumbu in Siaya district, and Lela in Kisumu district of western Kenya (Sileshi et al., 2006). In each treatment, the abundance of *M. ochroptera* was monitored on 15 randomly selected trees that were tagged using coloured plastic strings. The numbers of adults were recorded monthly from July to December in 1999 and 2000 on each tree. This constituted two years of data collected on six dates of sampling for each site and treatment. On each date, samples were taken from the same tree, and this constituted a repeated measures dataset. The total sample size was 2546 *S. sesban* trees, of which 2035 (79.9%) had zero counts of *M. ochroptera*. Years, dates, sites and treatments were used as covariates to estimate abundance parameters in Section 2.2.1.

The second dataset consisted of counts of the invasive species *Heteropsylla cubana* Craw per shoot of eight provenances of the fodder tree *Leucaena leucocephala* (Lam) de Wit in Tanzania. To assess psyllid abundance, three terminal shoots with the next three open leaves were randomly cut from three randomly selected trees per plot. Samples were bagged in polythene bags and taken to the laboratory where the shoots were examined under a dissecting microscope and the number of psyllid nymphs per shoot was

recorded. Data were recorded on six sampling dates. The total sample size was 861 shoots, of which 403 (46.8%) had zero counts. Sampling dates and provenances were used as covariates to estimate abundance parameters in Section 2.2.1.

The third dataset consisted of counts of the curculionid beetle *Diaecoderus* sp. per maize plant in eastern Zambia. Beetles were counted on 10 randomly selected maize plants in 13 agroforestry treatments in February 2002 and 2003. The treatments were replicated four times and arranged in a randomized complete blocks design. The total sample size was 990 plants, of which 444 (44.9%) were zero counts. Years and treatments were used as covariates for estimation of abundance parameters in Section 2.2.1.

The fourth dataset consisted of counts of the tenebrionid *Gonocephalum simplex* (F.) in soil monoliths from agroforestry practices in eastern Zambia. The study areas, treatment, experimental design and management of the experiments have been described in detail by Sileshi and Mafongoya (2007). Sampling was conducted three times between December 2003 and July 2004. Soil samples were collected using a soil monolith (25 cm × 25 cm and 25 cm depth) placed over a randomly selected spot, and driven into the soil to ground level using a metallic mallet. Adults were hand-sorted from the soil and counts recorded per soil monolith. The total sample size was 542 monoliths, of which 414 (76.4%) had zero counts. Sites and treatments were used as covariates to estimate the abundance parameters in Section 2.2.1.

The fifth dataset consisted of counts of the leaf beetle *Ootheca bennigseni* Weis in eastern Zambia. Beetles were monitored on bean and cowpea crops in experimental fields and two nearby farmers' fields at Msekera in February 2003. Each farm was divided into homogenous (2 m × 2 m) plots and beetle counts were recorded on 15 and 30 plants of each of bean and cowpea plants per plot in farmers' field and the experimental fields, respectively. The total sample size was 420 plants, of which 240 (68.6%) had zero counts. Fields and crops were used as covariates to estimate parameters of abundance in Section 2.2.1.

2.2. The modelling approach

The shape and interpretation of occupancy–abundance and abundance–variance–occupancy relationships are subject to the sampling scale (He et al., 2002). In practice, these relationships are established at some sampling scale using a range of sample mean abundance (m), variance (s^2) and occupancy (p_o). If the sample size is sufficiently large, m , s^2 and p_o are assumed to approach the true abundance (μ), variance (σ^2) and occupancy (p_p), respectively. For clarity, sample abundance is defined as the mean density of individuals in the sampling units (habitat patches) in which a species was recorded, and the observed occupancy as the proportion of occupied patches. When the sampling scale changes, values of m , s^2 and p_o will change, and this is likely to change the model that best fits the observed data. The computation of m , p_o and s^2 is sometimes done without due consideration for predictors (covariates) of μ , σ^2 and p_p . If not done according to covariates that significantly explain these parameters, they may be biased resulting in distortion of the occupancy–abundance relationship. In this paper we propose a three-step modelling approach that will account for zero-inflation and improve accuracy in parameter estimation. The steps include (1) selecting distributional assumptions and parsimonious models for the count data, (2) estimating abundance, occupancy and variance parameters as functions of site- and/or time-specific covariates, and (3) modelling the occupancy–abundance relationship using the parameters estimated in step 2.

2.2.1. Modelling abundance

The first two steps involved analysis of the datasets described above assuming Poisson, NBD, ZIP and ZINB, and jointly estimating

Download English Version:

<https://daneshyari.com/en/article/4377952>

Download Persian Version:

<https://daneshyari.com/article/4377952>

[Daneshyari.com](https://daneshyari.com)