

# On species occupancy-abundance models<sup>1</sup>

Fangliang HE<sup>2</sup>, Canadian Forest Service, Pacific Forestry Centre, 506 West Burnside Road, Victoria,  
British Columbia V8Z 1M5, Canada, e-mail: fhe@pfc.forestry.ca

Kevin J. GASTON, Biodiversity and Macroecology Group, Department of Animal and Plant  
Sciences, University of Sheffield, Sheffield S10 2TN, United Kingdom.

Jianguo WU, Department of Plant Biology, Arizona State University, P.O. Box 871601,  
Tempe, Arizona 85287-1601, U.S.A.

**Abstract:** One of the most pronounced generalizations in macroecology is the positive interspecific relationship between the distribution and abundance of species. Understanding of this relationship comes in two forms: statistical descriptions and insight into the biological mechanisms. Several statistical and mathematical models have been used to describe, to summarize, to interpret, and to predict the positive relationship, *e.g.*, the Poisson, negative binomial, Nachman's, power, and logistic models. In this study we derive a three-parameter model of which others are special cases. The unified model describes how occupancy depends on the density and the spatial distribution of species. It is shown that the unified model can be generalized from the balance between the colonization and extinction processes of a metapopulation, which provides a mechanistic basis for the model. A maximum likelihood procedure is developed to estimate the parameters, and several occupancy-abundance models are applied to data on bird and tree species distributions. It is concluded that the general model gives the best fits to the two data sets although Nachman's model is recommended for the bird species and the logistic model for the tree species because these models have only two parameters. We further discuss how to quantify occupancy-abundance patterns and to define species-area curves in terms of occurrence.

**Keywords:** abundance, distribution, incident probability, macroecology, maximum likelihood, occupancy, occurrence, species-area curve, *z*-value.

**Résumé :** En macroécologie, une des idées les plus répandues est la relation positive que l'on observe entre la répartition et l'abondance des espèces. On peut comprendre cette relation grâce à des descriptions statistiques et une analyse des mécanismes biologiques. Plusieurs modèles statistiques et mathématiques ont été utilisés pour décrire, résumer, interpréter et prédire cette relation positive, comme les modèles de Poisson, binomial négatif, de Nachman, de puissance et le modèle logistique. Dans cette étude, nous proposons un modèle utilisant trois paramètres. Le modèle décrit comment l'occupation d'un lieu dépend de la densité et de la répartition spatiale des espèces. Le modèle peut être généralisé en prenant en considération l'équilibre entre les processus de colonisation et d'extinction d'une métapopulation, ce qui fournit la base mécanique du modèle. Une procédure de vraisemblance maximale a été développée pour estimer les paramètres. Plusieurs modèles de présence-absence et d'abondance ont été testés grâce à des données de répartition d'espèces d'oiseaux et d'arbres. Le modèle général donne les meilleurs résultats avec les deux ensembles de données, bien que les modèles de Nachman (pour les oiseaux) et logistique (pour les arbres) soient recommandés car ils ne font appel qu'à deux paramètres. Enfin, nous discutons des façons de quantifier les patrons de présence-absence et d'abondance et de déterminer les courbes du nombre d'espèces en fonction de la superficie selon leur présence.

**Mots-clés :** abondance, distribution, probabilité incidente, macroécologie, vraisemblance maximale, occupation, présence, courbe aire-espèce, valeur normale centrée réduite.

## Introduction

The distribution and abundance of species have long been central issues in ecology and biogeography (Willis, 1922; Andrewartha & Birch, 1954; Williams, 1964; McNaughton & Wolf, 1970; Brown, 1984; Gaston & Lawton, 1988a; Hanski, Kouki & Halkka, 1993; Gaston & Blackburn, 2000). Recently, much attention has been paid to the interrelationship between these two quantities in the context of macroecology, which is a new research program focusing on the documentation and interpretation of patterns in a range of species attributes (*e.g.*, distribution, abundance, body size, life history traits) at broad (geographical) spatial scales (Brown & Maurer, 1989; Gaston, 1994; Brown, 1995; Gaston & Blackburn, 2000). Although the univariate, bivariate, and multivariate characteristics of these attributes may all be important in understanding the

structure of species assemblages in space and time, the bivariate correlation between distribution and abundance has become the most widely documented and extensively studied (Brown, 1984, 1995; Gotelli & Simberloff, 1987; Gaston & Lawton, 1988b; Kolasa, 1989; Collins & Glenn, 1990; Gaston, 1994, 1996; Durrer & Schmid-Hempel, 1995; Boecken & Shachak, 1998; Johnson, 1998).

Previous studies have convincingly demonstrated that there is a general positive interspecific relationship between occupancy and abundance, *i.e.*, locally abundant species tend to be widespread in space whereas rare species tend to be narrowly distributed. Attempts to determine why this pattern occurs have met with only limited success (Kolasa, 1989; Hanski, Kouki & Halkka, 1993; Gaston, 1994; Brown, 1995; Gaston, Blackburn & Lawton, 1997; Kolasa & Drake, 1998; Gaston & Blackburn, 2000). Doubtless, this has been principally because of difficulties in disentangling the possible effects of different mechanisms and the diffi-

<sup>1</sup>Rec. 2001-04-11; acc. 2001-09-25.

<sup>2</sup>Author for correspondence.

culty of conducting manipulative experiments at relevant spatial scales. However, the lack of any widely accepted general descriptive model of the occupancy-abundance relationship certainly has not helped and has also hampered progress in exploring the implications of the relationship for other facets of community structure (Leitner & Rosenzweig, 1997). This is despite the fact that a number of statistical models of occupancy-abundance relationships exist in the literature although they were developed for a variety of purposes (Nachman, 1981; Wright, 1991; Gaston, 1994; Hanski & Gyllenberg, 1997; Leitner & Rosenzweig, 1997).

To provide a more comprehensive and cohesive perspective on the occupancy-abundance relationship, here we derive a general model that unifies the existing and seemingly independent models (throughout, occupancy is defined as the area occupied by a species in a study area, and abundance is the number of individuals of the species per unit area or in a given area, which is equivalent to density). We then give an interpretation of this general model in terms of the dynamics of metapopulations although the general model by derivation is statistical and empirical. We further develop a maximum likelihood procedure for parameter estimation. Finally, we discuss how these models can be used to quantify occupancy-abundance patterns and to define related species-area curves.

### Models and mechanisms

The simplest occupancy-abundance model is that derived from the Poisson distribution by assuming that the individuals of a species are randomly and independently distributed in space. It has the form (Wright, 1991)

$$p = 1 - e^{-\mu} \quad [1]$$

where  $p$  is the probability of occurrence (or incidence) of a species in a sample unit (*e.g.*, an areal sample), and  $\mu$  is the mean local density of the species (*i.e.*, the mean density per sample). As an example, let's assume that a total of  $N$  individuals of a species are randomly distributed among a given number of sites  $M$  of which  $m$  sites are occupied;  $p$  is then the proportion of sites occupied ( $= m/M$ ) and  $\mu$  is the mean number of organisms per site ( $= N/M$ ). In reporting occupancy-abundance relationships, abundances are sometimes averaged only across occupied sites (Gaston, 1996). Accordingly, model [1] can easily be expressed in terms of the density averaged across occupied sites ( $\mu' = \mu M/m$ ):

$$p = 1 - e^{-\frac{M}{m}\mu} = 1 - e^{-p\mu'}$$

As He and Gaston (2000a) point out, the change in the definition of density will make no qualitative difference to the occupancy-abundance relationship. This will also be true for the other occupancy-abundance models that follow.

In practice, the individuals of few species are randomly distributed in space; instead, most are aggregated. The model most commonly used to describe such aggregation is the negative binomial distribution (NBD) (Evans, 1953; Boswell & Patil, 1970). From the NBD an occupancy-abundance model can be derived as (Wright, 1991)

$$p = 1 - \left(1 + \frac{\mu}{k}\right)^{-k} \quad [2]$$

where  $k$  is a positive aggregation parameter, with small  $k$  representing strong aggregation and large  $k$  a random distribution. Although  $k$  of the NBD is defined to be strictly positive, He and Gaston (2000a) show that  $k$  in the occupancy-abundance model [2] could take negative values. When  $k$  is negative, model [2] is in fact derived from a positive binomial distribution that describes a regular distribution of organisms. Therefore, the effective domain of  $k$  in model [2] is  $(-\infty, -\mu)$  and  $(0, +\infty)$ , *i.e.*, the left-hand domain describes a regular distribution of organisms whereas the right-hand domain constitutes aggregation (He & Gaston, 2000a).

The third occupancy-abundance model is initially given by Nachman (1981). It has the form

$$p = 1 - e^{-\alpha\mu^\beta} \quad [3]$$

where  $\alpha$  is a positive parameter, and  $\beta$  is a positive scale parameter that determines the shape and curvature of the  $p$  versus  $\mu$  curve. Model [3] has traditionally been used in agricultural entomology for estimating pest density  $\mu$  from  $p$  because of the convenience and economy in recording pest presence/absence on crop plants (Nachman, 1984; Ekbom, 1987; Perry, 1987; Kuno, 1991; Hepworth & MacFarlane, 1992). The model is now used to describe occupancy-abundance relationships (Gaston, 1994).

Recently there has been increasing interest in modeling species-area curves in terms of the range size or occupancy of species (Leitner & Rosenzweig, 1997; Hanski & Gyllenberg, 1997; Ney-Nifle & Mangel, 1999). In this regard, Leitner and Rosenzweig (1997) give a power occupancy-abundance model (also see Gaston, 1994):

$$p = \alpha\mu^\beta \quad [4]$$

where  $\alpha$  is positive, and  $\beta$  is a scale parameter.

In the same context, Hanski and Gyllenberg (1997) use a logistic model to describe occupancy-abundance relationships:

$$p = \frac{1}{1 + e^{-a-b\ln(\mu)}}$$

which can equivalently be written as

$$p = \frac{\alpha\mu^\beta}{1 + \alpha\mu^\beta} \quad [5]$$

where  $\alpha = e^a$  and  $\beta = b$  are two positive parameters.

Although these five models have been derived by different authors and applied in different kinds of situations, they are not independent. He and Gaston (2000a) have outlined possible links of the Poisson, the Nachman, and the logistic models with the NBD model [2]. In fact, a more general model can be formulated, of which the other five models are special cases. The general model has three parameters:

$$p = 1 - \left(1 + \frac{\alpha\mu^\beta}{k}\right)^{-k} \quad [6]$$

where  $\alpha$  is a positive parameter,  $\beta$  is a scale parameter, and  $k$  is a negative or positive binomial distribution parameter defined in the domain of  $(-\infty, -\alpha\mu^\beta)$  or  $(0, +\infty)$ . It is easy to show that:

- a) when  $\alpha = \beta = 1$ , [6] is the NBD model [2];
- b) when  $k = -1$ , [6] is the power model [4];
- c) when  $k = 1$ , [6] is the logistic model [5];
- d) when  $k \rightarrow \pm\infty$ , [6] is the Nachman model [3];
- e) when  $k \rightarrow \pm\infty$  and  $\alpha = \beta = 1$ , [6] is the Poisson model [1].

The general model [6] describes the direct dependence of occupancy on the density and spatial distribution of species, and the effects of density could take various forms (determined by the term  $\alpha\mu^\beta$  in equation [6]). Although the Poisson model [1] and the NBD model [2] can be derived from the assumption that species abundances in a given area follow the Poisson and the negative binomial distributions, others are largely empirical models lacking a biological or statistical basis. However, one plausible interpretation of these models can be found from metapopulation dynamics. Following Hanski (1994; 1997), when a metapopulation is in steady state, the occupancy probability (called incidence function in some literature) can be expressed as

$$p = \frac{C}{C + E} \tag{7}$$

where  $C$  is the colonization rate of empty sites, and  $E$  is the extinction rate of extant populations. Studies have shown that  $C$  is an increasing function, and  $E$  is a decreasing function, of population density  $\mu$  (Gilpin & Diamond, 1976; Hanski & Gyllenberg, 1997). Although the relationships between the colonization and extinction rates and population density may be defined on biological grounds, the ones typically used in the literature are conveniently given as  $C = a\mu^b$  and  $E = c\mu^{-d}$  (see Hanski, 1994), where  $a$ ,  $b$ ,  $c$ , and  $d$  are constants. These relationships capture the main features of increasing colonization and decreasing extinction along the axis of density (Figure 1a), and they make a great deal of biological sense in that a high colonization rate is associated with high population density whereas a high extinction rate is associated with a low density. As a result, it is easy to show that the incidence function [7] becomes the occupancy-abundance model [5] in the form

$$p = 1 - \left(1 + \frac{\alpha\mu^\beta}{1}\right)^{-1} \tag{8}$$

where  $\alpha = a/c$  and  $\beta = b + d$ . It is obvious that model [6] is a generalization of model [8]. Based on this generalization, we can further equate models [6] and [7] to obtain

$$\frac{C}{E} = \left(1 + \frac{\alpha\mu^\beta}{k}\right)^k - 1 \tag{9}$$

This model relates the spatial aggregation (represented by  $k$ ) of a species to the colonization and extinction processes of a metapopulation. It suggests that for a given popula-

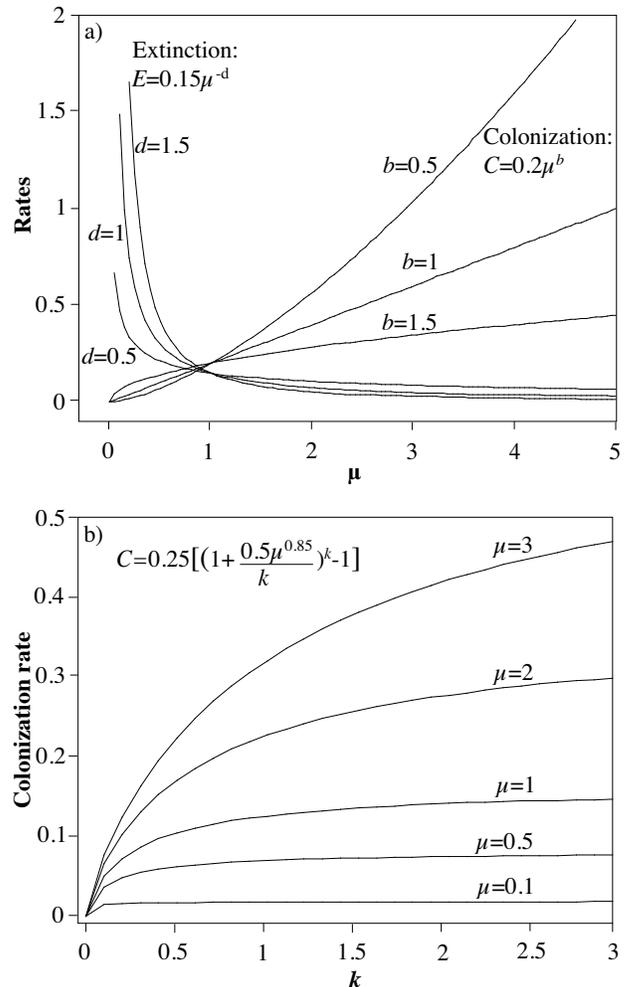


FIGURE 1. (a) Numerical results illustrating the relationships between colonization rate and population density, and extinction rate and population density; (b) the relationship between colonization rate  $C$  and spatial distribution of species (model [9]) for a given extinction rate ( $E = 0.2$ ) and a population level ( $\mu$  from 0.1 to 3).

tion abundance and extinction rate (*i.e.*,  $\alpha\mu^\beta$  and  $E$  are fixed), the colonization rate  $C$  is a function of spatial aggregation ( $k$ ) in such a way that if the species is highly aggregated (*i.e.*,  $k \rightarrow 0^+$ ),  $C$  is small, and if the species is widely distributed (*i.e.*,  $k$  is large),  $C$  is large (Figure 1b). In the extreme case where a species is so aggregated that all individuals inhabit a single location, it is natural to think that the species has a low colonization ability. These results seem to be rather reasonable and realistic and are consistent with evidence showing that species aggregation is closely associated with poor colonization ability (Tilman, Lehman & Kareiva, 1997).

PARAMETER ESTIMATION

Occupancy-abundance data are usually collected by observing the distribution of several species in many sample units as shown in Table I. Although the above occupancy-abundance models may be fitted by nonlinear least squares regression, the method is probably not appropriate because the response variable  $p$  (Table I) is bounded in  $[0, 1]$ , and its

distribution is usually highly skewed. One alternative solution is data transformation to normality (Williamson & Gaston, 1999). Another, which we give here, is the maximum likelihood estimation.

The occurrence of a given species in a particular sample unit is a binary random variable, hence the number of occupied units,  $y$ , is a binomial variable:

$$f(y) = \binom{n}{y} p^y (1-p)^{n-y}, y = 0, 1, \dots, n \quad [10]$$

Each of the  $s$  species in table I has a binomial distribution as defined in [10]. Therefore, the joint log-likelihood function for the  $s$  species is

$$l(y_1, \dots, y_s) = \sum_{i=1}^s \left[ y_i \ln \left( \frac{p_i}{1-p_i} \right) + n_i \ln(1-p_i) + \ln \binom{n_i}{y_i} \right] \quad [11]$$

The probability  $p_i$  varies from species to species because of the difference in density  $\mu_i$ . The parameters in occupancy-abundance models [1] to [6] can be estimated by substituting a corresponding model into [11] and then maximizing the log-likelihood function with respect to the parameters. For the logistic model [5], it is a typical generalized linear model with logit link (McCullagh & Nelder, 1989):

$$\ln \left( \frac{p}{1-p} \right) = \ln \alpha + \beta \ln \mu$$

Here we apply all six occupancy-abundance models (the five existing ones, plus the general model) to two data sets. The first set is derived from the most recent atlas of the distribution of breeding birds of Bedfordshire, United Kingdom (Dazley & Trodd, 1994). The entire county is divided into 378 2-km  $\times$  2-km grid cells. The occurrence of each of the 61 passerine species in each cell was recorded, and the abundance of each species in the entire county was also approximately recorded. The final data set has the form of table I with  $s = 61$ . The second data set is that of 824 tree/shrub species in a 500-m  $\times$  1000-m plot from a tropical rain forest in Pasoh, Malaysia (He, Legendre & LaFrankie, 1997; Manokaran *et al.*, 1999). The plot is divided into 800 25-m  $\times$  25-m grid cells. These data also have the form of table I with  $s = 824$ . Further detailed description of the two data sets can be found in He and Gaston (2000b).

The goodness-of-fit of each model is assessed by comparing the deviance with a  $\chi^2$  distribution (McCullagh & Nelder, 1989). The deviance is twice the difference between the maximum log-likelihood achievable in [11] and the log-likelihood for the model of interest. The maximum log-likelihood achievable (*i.e.*, the full model) is that obtained

TABLE I. Typical occupancy-abundance data. In some studies, the total number of sample units,  $n_i$ , may be equal for the different species.

Species	1	2	.....	$s$
No. of occupied units	$y_1$	$y_2$	.....	$y_s$
No. of empty units	$n_1 - y_1$	$n_2 - y_2$	.....	$n_s - y_s$
Total no. of units	$n_1$	$n_2$	.....	$n_s$
Proportion occupied	$p_1 = y_1/n_1$	$p_2 = y_2/n_2$	.....	$p_s = y_s/n_s$
Density	$\mu_1$	$\mu_2$	.....	$\mu_s$

by substituting the observed  $p_i$  into equation [11]. If the model of interest is adequate, the deviance should be small compared to the  $\chi^2_{s-l}$ , where  $s$  is the total number of species and  $l$  is the number of parameters in the model under study (roughly, the model is considered adequate if the deviance is smaller than the degrees of freedom  $s-l$ ).

By the  $\chi^2$  test, the results in table II show that none of the models significantly fits the data. However, the test statistic must be interpreted cautiously because the deviances are much larger than the degrees of freedom, suggesting that the models are over-dispersed, and therefore the  $\chi^2$  test for goodness-of-fit is not reliable. Readers are referred to McCullagh and Nelder (1989) and Dean (1992) for further details about testing and modeling over-dispersion problems. Furthermore, the asymptotic  $\chi^2$  distribution for the deviance may not be particularly relevant because the assumption of independence of the  $s$  species for the joint log-likelihood function [11] may not hold. Here we are more interested in the relative performances of the models than a rigorous statistical test and, hence, use the sum of absolute differences between the observed proportions of occurrence ( $p_i$ ) and the fitted probabilities ( $\hat{p}_i$ ) as a goodness-of-fit:

$$\sum_{i=1}^s |p_i - \hat{p}_i|$$

For the Bedfordshire bird data, the algorithm for the power model [4] did not converge (therefore, it is excluded from Table IIa). The three-parameter general model [6] and the Nachman model give identical predictions (Table IIa). This is because the estimated  $k$  for the general model is so large ( $\approx 32,4000$ ) that it actually becomes the Nachman model. As a result, both models have the same  $\alpha$  and  $\beta$  values, 0.172 (standard error, SE = 0.0054) and 0.674 (SE = 0.0097), respectively. The deviance and the absolute difference for the general model estimated using the nonlinear least squares method are also included for comparison (Table II). It is clear from Figure 2a that there is little difference between the maximum likelihood method and the nonlinear least squares method for the general model.

For the Pasoh tree data, the general model gives the best fit, but its marginal superiority over the logistic model is offset by an extra parameter (Table IIb). The estimated logistic model has parameters  $\alpha = 0.931$  (SE = 0.0039) and  $\beta = 1.036$  (SE = 0.0031). Again, the maximum likelihood method and the nonlinear least squares method give nearly identical results (Table IIb). However, for other models, the least squares regression may not be appropriate at all (*e.g.*, it cannot guarantee the power model to be within [0, 1] for the given density range of a study). All six models estimated using the maximum likelihood method are plotted in Figure 2b.

## Discussion

### MODELS AND PATTERNS OF OCCUPANCY-ABUNDANCE RELATIONSHIPS

The significance of the general model [6] is not only that it mathematically unifies other models, but that it also

TABLE II. Deviances and the absolute differences between the model predictions and the observed proportions of occurrence for comparing the six occupancy-abundance models. MLE: the models were fitted using the maximum likelihood method; NLS: the nonlinear least squares method.

	MLE					NLS	
	General	Poisson	Logistic	Nachman	NBD	Power	General
a. BEDFORDSHIRE BIRD DATA							
Deviance	1994.2	68140.0	2569.2	1994.2	4330.14		1994.2
(df. of $\chi^2$ )	(58)	(61)	(59)	(59)	(60)		(58)
Absolute difference	4.465	12.977	5.020	4.465	8.169		4.598
b. PASOH TREE DATA							
Deviance	14678.2	57205.9	14705.0	17130.3	15436.1	46296.3	14727.3
(df. of $\chi^2$ )	(821)	(824)	(822)	(822)	(823)	(822)	(821)
Absolute difference	26.264	47.399	26.324	29.709	26.524	64.265	26.307

points out the way that occupancy depends on density and distribution. The positive interspecific relationship between occupancy and abundance of species is captured by all the models described above although their forms vary. This variation reflects the flexibility of the general model [6] and the different aspects that the specific models emphasize. For instance, the Poisson model [1] represents a null hypothesis of a homogeneously random distribution of the individuals of a species, the NBD model [2] describes the effect of aggregation, and variation from the NBD is described by other models. Overall, the differences among these models can be interpreted in terms of the colonization and extinction processes experienced by a spatially structured set of populations as described by equation [9] and figure 1. However, except for the Poisson and the NBD models, precise mechanistic assessments for other models are difficult to make. According to equation [9], the Poisson and Nachman models may be more appropriate for species of higher colonization ability than other models because they are derived from the general model [6] at  $k \rightarrow \pm\infty$ . By the same argument, the logistic model may be more appropriate for species of lower colonization ability because the model is a special case of the general model [6] at small  $k$  ( $= 1$ ), which is associated with low colonization, according to equation [9], given a fixed extinction rate. However, it is worth noting that colonization is not the sole factor that determines the distribution of species; further information (concerning, *e.g.*, reproductive behavior, physiological tolerance and capacity, and landscape heterogeneity) is required to understand the limit of the distributions of species (Spicer & Gaston, 1999; Gaston & Blackburn, 2000). Nevertheless, as a very general rule, we suggest that the Nachman model is more appropriate for species of relatively high colonization ability (*e.g.*, birds) while the logistic model is better for species of relatively low colonization ability (*e.g.*, plants).

By comparing an observed pattern to the Poisson or NBD model, we may be able to make inferences about how spatial aggregation would affect occupancy-abundance relationships. As shown in table II, the NBD model fits better than the Poisson to the two data sets. This is not surprising given the wide variation in spatial distribution of species. The Nachman model is the best candidate for the Bedfordshire data while the logistic is the best for the Pasoh data. This seems to reflect the difference in colonization ability of the birds and plants as discussed above.

Gaston (1994) hypothesizes an overall triangular pattern for interspecific occupancy-abundance relationships (particularly at broad spatial scales and taxonomic diversity). The triangular pattern is obvious if the data for the Bedfordshire bird and the Pasoh tree species in figure 2 are plotted on an untransformed (rather than logarithmic) scale. Figure 3 shows the triangle for the Pasoh species, which suggests that it is broadly defined by the Poisson model [1] and the power model [4], with  $\alpha = 1/\mu_{\max}$  and  $\beta = 1$ , *i.e.*, the straight lines connecting  $(0, 0)$  and  $(\mu_{\max}, 1)$ . The upper limit of the triangle is defined by a random distribution of individuals whereas the lower bound is the straight line indicating that the proportion of occupancy by a species is linearly proportional to density. This triangle also holds for the Bedfordshire data. Obviously, the triangle rule is not exclusive. Strongly aggregated species can lie outside the triangle as shown in figure 3 for three species that are extremely aggregated in the Pasoh plot (He, Legendre & LaFrankie, 1997).

The interpretation of occupancy-abundance data and their models is subject to the sampling scale used. It is apparent that when scale changes, both the occupancy ( $p$ ) and abundance ( $\mu$ ) of a species will change. It is likely that the change in scale will also change the model that “best” fits the observed data. For example, if the sampling scale is so fine that each cell (site) contains only one organism, then the Poisson model [1] will just be adequate in describing the occupancy data even though at coarser scales other models may better describe the data. On the other hand, if sampling scale is so coarse that it is as large as the study area, then all species will have the same occupancy, with  $p = 1$ , regardless of the variation in abundance among species. In this trivial case, no occupancy-abundance model will be able satisfactorily to describe the data.

#### OCCUPANCY-ABUNDANCE AND SPECIES-AREA MODELS

There has been interest in modeling species-area curves in terms of species occurrence (Hanski & Gyllenberg, 1997; Leitner & Rosenzweig, 1997; Ney-Nifle & Mangel, 1999). Since an occupancy-abundance model represents the probability of species occurrence in a sample area, the expected number of species in the area is simply the sum of the probabilities across all species:

$$s_a = \sum_{i=1}^s p_i \quad [12]$$

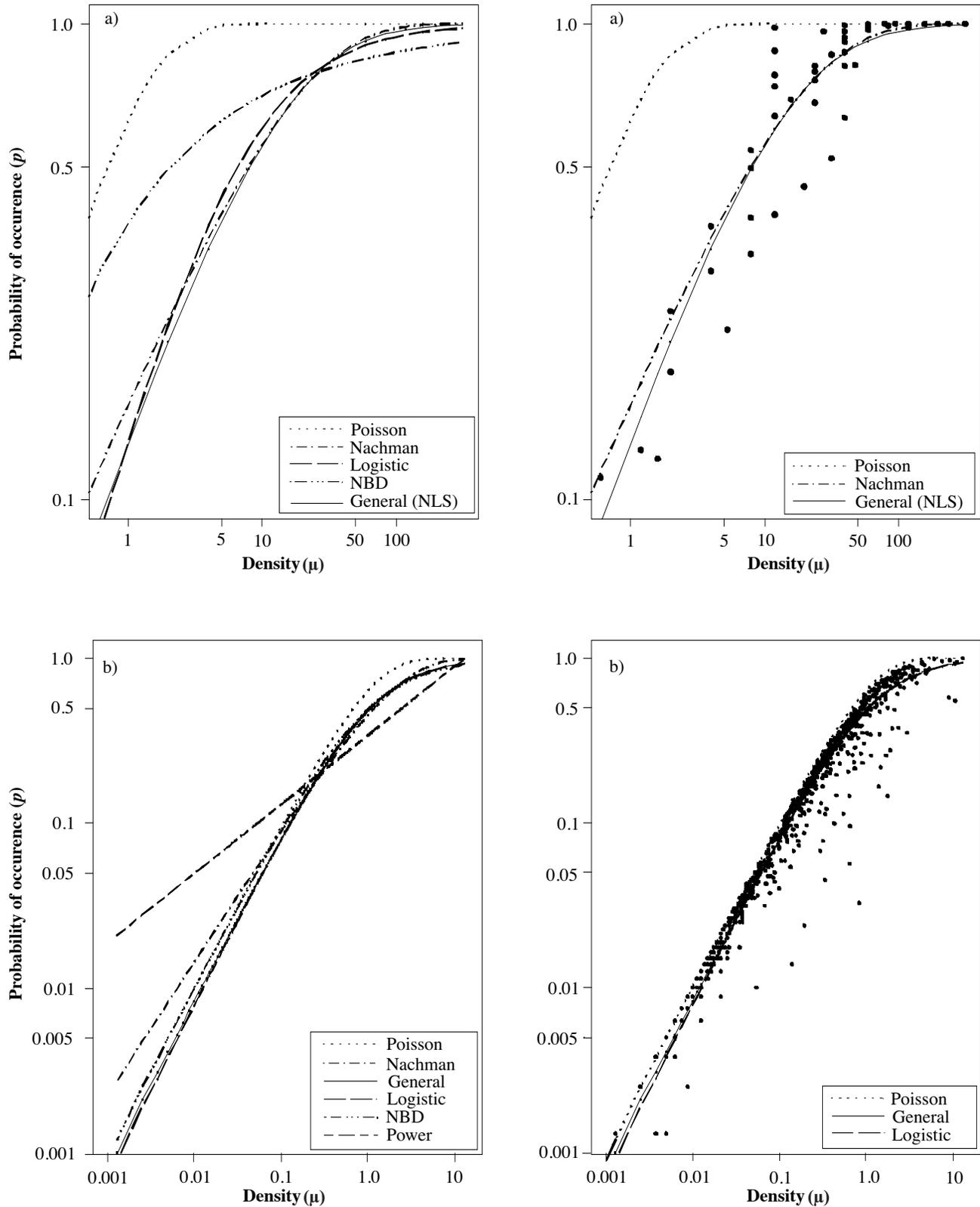


FIGURE 2. (a) Fitting the Bedfordshire bird data at a 2-km  $\times$  2-km scale using five models: [1], [2], [3], [5] and [6]. The left panel shows the five fitted models. The Nachman model is identical to the general model fitted by MLE. The right panel shows three fitted curves (Poisson, Nachman, and general [NLS]) plus the observed data. The general model was fitted using the nonlinear least squares method (NLS). (b) Fitting the Pasoh tree data at a 25-m  $\times$  25-m scale using all six models: [1] to [6]. The left panel shows the six fitted models. The right panel shows three fitted models (Poisson, general, and logistic) plus the observed data. The logistic model and the general model are the best fitting and are nearly identical. Dots in a and b are the observed data.

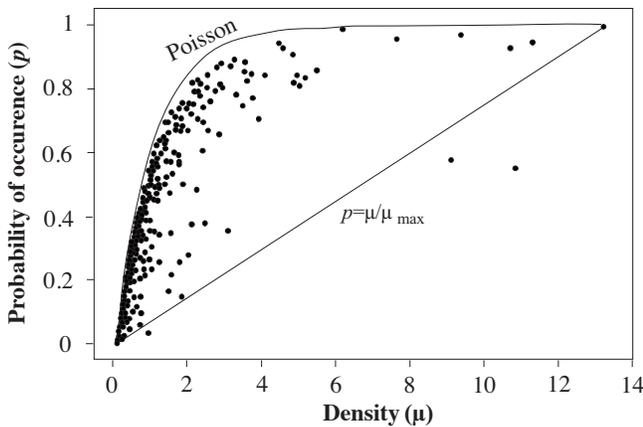


FIGURE 3. The occupancy-abundance pattern is broadly defined by a triangle with the upper bound of the Poisson model [1] and the lower bound of the model  $p = \mu/\mu_{\max}$ .

where  $s_a$  is the number of species in sample  $a$ , and  $s$  is the total number of species in the study area. For instance, substituting the Poisson model [1] into [12], the species-area curve is of the form that is a limit of the random placement model of Coleman (1981):

$$s_a = s - \sum_{i=1}^s e^{-\frac{a}{A}N_i} = s - \sum_{i=1}^s e^{-\mu_i} \quad [13]$$

where  $N_i$  is the total abundance for species  $i$  in the study area.

The species-area model [12] links occupancy-abundance relationships to species-area curves. Based on this model, we can analyze the effect of distribution ranges of species on the slope of a species-area curve:

$$z = \frac{d \ln s_a}{d \ln a} = \frac{a}{s_a} \frac{ds_a}{da} \quad [14]$$

The species-area curve corresponding to the general occupancy-abundance model [6] can be derived by substituting [6] into [12] and replacing  $\mu_i$  by  $aN_i/A$ . The  $z$ -value is then found by a derivative with respect to  $a$  following equation [14], resulting in

$$z = k\beta \frac{\sum_{i=1}^s [1 - p_i - (1 - p_i)^{1+1/k}]}{\sum_{i=1}^s p_i} \quad [15]$$

The  $z$ -values for the other occupancy-abundance models can accordingly be derived. It is apparent that the  $z$ -value is not a constant but depends on the occupancy-abundance model used. Leitner and Rosenzweig (1997) use the power model [4] approximately to derive a species-area curve. Based on that, they find that the  $z$ -value and the scale parameter  $\beta$  (their notation  $c$ ) form a near linear 1-1 line (their Figure 7). Based on equation [15], it is easy to show that this 1-1 line is exact. The power model corresponds to  $k = -1$  in equation [15], which results in  $z = \beta$ .

Hanski and Gyllenberg (1997) use the logistic model [5] to show the relatedness between species-area curves and occupancy-abundance models using species densities,

island areas, and the variations of these two quantities as a common currency. They conclude that the species-abundance distribution itself is not important in determining  $z$ -values. From equation [15], the  $z$ -value for the logistic model [5] when  $k = 1$  is

$$z = \beta \frac{\sum_{i=1}^s p_i(1 - p_i)}{\sum_{i=1}^s p_i} \quad [16]$$

Here we are interested in examining how different occupancy-abundance models affect the  $z$ -values. The results for the Bedfordshire bird and the Pasoh tree data using the six occupancy-abundance models are shown in table III. Differences in  $z$ -values for the Bedfordshire bird data are discernible whereas the  $z$ -values for the Pasoh tree data change little among the models, except for the power model. Nevertheless, overall, for a given data set the  $z$ -value is relatively insensitive to occupancy-abundance models. This is probably due to the inadequacy of the occupancy-abundance models in describing the spatial structure of occupied sites for a species. In comparing the  $z$ -values, we do not suggest that the species-area curve defined by [12] is adequately fitted by the  $\log(s_a)$ - $\log(a)$  species-area model. In fact, much evidence, particularly that from species-rich tropical forests, has shown that the log-log species-area curve is a poor model (Condit *et al.*, 1996; Plotkin *et al.*, 2000). The existence of a slope does not mean that the data points actually fall along a line. Therefore, for the different occupancy-abundance models, the actual species-area curves can still be very different.

TABLE III. The slopes ( $z$ -values) of the  $\log(s_a)$ - $\log(a)$  species-area curves defined by the various occupancy-abundance models for the Bedfordshire bird and the Pasoh tree data.

Data	Occupancy-abundance models					
	General	Poisson	Logistic	Nachman	NBD	Power
Bedfordshire	0.216	0.047	0.195	0.216	0.086	
Pasoh	0.568	0.567	0.567	0.566	0.539	1

### Acknowledgements

The large-scale forest plot at the Pasoh Forest Reserve is an ongoing project of the Malaysian Government, directed by the Forest Research Institute of Malaysia and the Smithsonian Tropical Research Institute, under the leadership of N. Manokaran, P. S. Ashton, and S. P. Hubbell. F. He is supported by the Biodiversity and Ecosystem Processes Networks of the Canadian Forest Service. K. J. Gaston is a Royal Society University Research Fellow. J. Wu gratefully acknowledges funding from the U.S. Environmental Protection Agency (EPA R827676-01-0).

### Literature cited

- Andrewartha, H. G. & L. C. Birch, 1954. The Distribution and Abundance of Animals. University of Chicago Press, Chicago, Illinois.
- Boecklen, B. & M. Shachak, 1998. The dynamics of abundance and incidence of annual plant species richness during colonization in a desert. *Ecography*, 21: 63-73.

- Boswell, M. T. & G. P. Patil, 1970. Chance mechanisms generating the negative binomial distributions. Pages 3-22 in G. P. Patil (ed.). *Random Counts in Models and Structures*. Pennsylvania State University Press, University Park, Pennsylvania.
- Brown, J. H., 1984. On the relationship between abundance and distribution of species. *American Naturalist*, 124: 255-279.
- Brown, J. H., 1995. *Macroecology*. University of Chicago Press, Chicago, Illinois.
- Brown, J. H. & B. A. Maurer, 1989. Macroecology: The division of food and space among species on continents. *Science*, 243: 1145-1150.
- Coleman, D. B., 1981. On random placement and species-area relations. *Mathematical Biosciences*, 54: 191-215.
- Collins, S. L. & S. M. Glenn, 1990. A hierarchical analysis of species' abundance patterns in grassland vegetation. *American Naturalist*, 135: 633-648.
- Condit, R., S. P. Hubbell, J. V. LaFrankie, R. Sukumar, N. Manokaran, R. B. Foster & P. S. Ashton, 1996. Species-area and species-individual relationships for tropical trees: A comparison of three 50-ha plots. *Journal of Ecology*, 84: 549-562.
- Dazley, R. A. & P. Trodd, 1994. *An atlas of the breeding birds of Bedfordshire: 1988-92*. Bedfordshire Natural History Society, Bedford.
- Dean, C. B., 1992. Testing for overdispersion in Poisson and binomial regression models. *Journal of the American Statistical Association*, 87: 451-457.
- Durrer, S. & P. Schmid-Hempel, 1995. Parasites and the regional distribution of bumblebee species. *Ecography*, 18: 114-122.
- Ekbom, B. S., 1987. Incidence counts for estimating densities of *Rhopalosiphum padi* (Homoptera: Aphididae). *Journal of Economic Entomology*, 80: 933-935.
- Evans, F. N., 1953. Experimental evidence concerning contagious distributions in ecology. *Biometrika*, 40: 186-211.
- Gaston, K. J., 1994. *Rarity*. Chapman & Hall, London.
- Gaston, K. J., 1996. The multiple forms of the interspecific abundance-distribution relationship. *Oikos*, 76: 211-220.
- Gaston, K. J. & T. M. Blackburn, 2000. *Pattern and Process in Macroecology*. Blackwell Science, Oxford.
- Gaston, K. J., T. M. Blackburn & J. H. Lawton, 1997. Interspecific abundance-range size relationships: An appraisal of mechanisms. *Journal Animal Ecology*, 66: 579-601.
- Gaston, K. J. & J. H. Lawton, 1988a. Patterns in the abundance and distribution of insect populations. *Nature*, 331: 709-712.
- Gaston, K. J. & J. H. Lawton, 1988b. Patterns in body size, population dynamics and regional distribution of bracken herbivores. *American Naturalist*, 132: 662-680.
- Gilpin, M. E. & J. M. Diamond, 1976. Calculation of immigration and extinction curves from the species-area-distance relation. *Proceedings of the National Academy of Sciences, U.S.A.*, 73: 4130-4134.
- Gotelli, N. J. & D. Simberloff, 1987. The distribution and abundance of tallgrass prairie plants: A test of the core-satellite hypothesis. *American Naturalist*, 130: 18-35.
- Hanski, I., 1994. A practical model of metapopulation dynamics. *Journal of Animal Ecology*, 63: 151-162.
- Hanski, I., 1997. Metapopulation dynamics, from concepts and observations to predictive models. Pages 69-91 in I. Hanski & M. E. Gilpin (ed.). *Metapopulation Biology*. Academic Press, San Diego, California.
- Hanski, I. & M. Gyllenberg, 1997. Uniting two general patterns in the distribution of species. *Science*, 275: 397-400.
- Hanski, I., J. Kouki & A. Halkka, 1993. Three explanations of the positive relationship between distribution and abundance of species. Pages 108-116 in R. E. Ricklefs & D. Schluter (ed.). *Species Diversity in Ecological Communities*. University of Chicago Press, Chicago, Illinois.
- He, F. & K. J. Gaston, 2000a. Estimating species abundance from occurrence. *American Naturalist*, 156: 553-559.
- He, F. & K. J. Gaston, 2000b. Occupancy-abundance relationships and sampling scales. *Ecography*, 23: 503-511.
- He, F., P. Legendre & J. V. LaFrankie, 1997. Distribution patterns of tree species in a Malaysian tropical rain forest. *Journal of Vegetation Science*, 8: 105-114.
- Hepworth, G. & J. R. MacFarlane, 1992. Systematic presence-absence sampling method applied to twospotted spider mite (Acari: Tetranychidae) on strawberries in Victoria, Australia. *Journal of Economic Entomology*, 85: 2234-2239.
- Johnson, C. N., 1998. Rarity in the tropics: Latitudinal gradients in distribution and abundance in Australian mammals. *Journal of Animal Ecology*, 67: 689-698.
- Kolasa, J., 1989. Ecological systems in hierarchical perspective: Breaks in community structure and other consequences. *Ecology*, 70: 36-47.
- Kolasa, J. & J. A. Drake, 1998. Abundance and range relationship in a fragmented landscape: Connections and contrasts between competing models. *Coenoses*, 13: 79-88.
- Kuno, E., 1991. Sampling and analysis of insect populations. *Annual Review of Entomology*, 36: 285-304.
- Leitner, W. A. & M. L. Rosenzweig, 1997. Nested species-area curves and stochastic sampling: a new theory. *Oikos*, 79: 503-512.
- Manokaran, N, J. V. LaFrankie, K. M. Kochummen, E. S. Quah, J. E. Klahn, P. S. Ashton & S. P. Hubbell, 1999. The Pasoh 50-ha forest dynamics plot: 1999 CD-ROM version. Forest Research Institute of Malaysia, Kepong.
- McCullagh, P. & J. A. Nelder, 1989. *Generalized Linear Models*. Chapman & Hall, London.
- McNaughton, S. J. & L. L. Wolf, 1970. Dominance and the niche in ecological systems. *Science*, 167: 131-139.
- Nachman, G., 1981. A mathematical model of the functional relationship between density and spatial distribution of a population. *Journal of Animal Ecology*, 50: 453-460.
- Nachman, G., 1984. Estimates of mean population density and spatial distribution of *Tetranychus urticae* (Acarina: Tetranychidae) and *Phytoseiulus persimilis* (Acarina: Phytoseiidae) based upon the proportion of empty sampling units. *Journal of Applied Ecology*, 21: 903-913.
- Ney-Nifle, M. & M. Mangel, 1999. Species-area curves based on geographical range and occupancy. *Journal of Theoretical Biology*, 196: 327-342.
- Perry, J. N., 1987. Host-parasitoid models of intermediate complexity. *American Naturalist*, 130: 955-957.
- Plotkin, J. B., M. D. Potts, D. W. Yu, S. Bunyavejchewin, R. Condit, R. Foster, S. Hubbell, J. LaFrankie, N. Manokaran, L. H. Seng, R. Sukumar, M. A. Nowak & P. Ashton, 2000. Predicting species diversity in tropical forests. *Proceedings of the National Academy of Sciences, U.S.A.*, 97: 10850-10854.
- Spicer, J. I. & K. J. Gaston, 1999. *Physiological Diversity and Its Ecological Implications*. Blackwell Science, Oxford.
- Tilman, D., C. L. Lehman & P. Kareiva, 1997. Population dynamics in spatial habitats. Pages 3-20 in D. Tilman & P. Kareiva (ed.). *Spatial Ecology*. Princeton University Press, Princeton, New Jersey.
- Williams, C. B., 1964. *Patterns in the Balance of Nature*. Academic Press, London.
- Williamson, M. & K. J. Gaston, 1999. A simple transformation for sets of range sizes. *Ecography*, 22: 674-680.
- Willis, J. C., 1922. *Age and Area*. Cambridge University Press, Cambridge.
- Wright, D. H., 1991. Correlations between incidence and abundance are expected by chance. *Journal of Biogeography*, 18: 463-466.