

LETTER

Confronting different models of community structure to species-abundance data: a Bayesian model comparison

Abstract

Rampal S. Etienne* and Han Olff
Community and Conservation
Ecology Group, University of
Groningen, PO Box 14, 9750 AA
Haren, The Netherlands
*Correspondence: E-mail:
r.s.etienne@rug.nl

Species abundances are undoubtedly the most widely available macroecological data, but can we use them to distinguish among several models of community structure? Here we present a Bayesian analysis of species-abundance data that yields a full joint probability distribution of each model's parameters plus a relatively parameter-independent criterion, the posterior Bayes factor, to compare these models. We illustrate our approach by comparing three classical distributions: the zero-sum multinomial (ZSM) distribution, based on Hubbell's neutral model, the multivariate Poisson lognormal distribution (MPLN), based on niche arguments, and the discrete broken stick (DBS) distribution, based on MacArthur's broken stick model. We give explicit formulas for the probability of observing a particular species-abundance data set in each model, and argue that conditioning on both sample size and species count is needed to allow comparisons between the two distributions. We apply our approach to two neotropical communities (trees, fish). We find that DBS is largely inferior to ZSM and MPLN for both communities. The tree data do not allow discrimination between ZSM and MPLN, but for the fish data ZSM (neutral model) overwhelmingly outperforms MPLN (niche model), suggesting that dispersal plays a previously underestimated role in structuring tropical freshwater fish communities. We advocate this approach for identifying the relative importance of dispersal and niche-partitioning in determining diversity of different ecological groups of species under different environmental conditions.

Keywords

Biodiversity, broken stick model, community, neutral model, niche-based model, Poisson lognormal distribution, zero-sum multinomial distribution.

Ecology Letters (2005) 8: 493–504

INTRODUCTION

As biodiversity is currently lost at an alarming rate with both profound ethical and socio-economical consequences (Costanza *et al.* 1997; Chapin *et al.* 2000), understanding the mechanisms underlying coexistence of species is crucial for a sustainable future of life on our planet. Large-scale observational biodiversity patterns (e.g. species-abundance distributions (SAD), species–area curves, body size-diversity distributions, etc.) are considered to reflect the underlying processes that structure ecological communities. After early studies by Motomura (1932), Fisher *et al.* (1943), Preston (1948, 1962) and MacArthur (1957, 1960), the study of these patterns was revived by Brown & Maurer (1989) who labelled this scientific field 'macroecology', which has since

then developed into an autonomous discipline (Brown 1995; Gaston & Blackburn 1999, 2000; Lawton 1999; Maurer 1999; Blackburn & Gaston 2003).

Macroecological patterns undeniably contain some information about the underlying processes, but it is not always clear how we can extract that information and to what extent this information will discriminate between alternative mechanistic hypotheses. As these patterns are often the only source of information we have on the processes that determine biodiversity (particularly for diverse tropical communities), it is worth improving our methods to extract from these patterns as much information as possible. Data on SAD, i.e. abundance counts of all species in an ecological community, play a central role in the current discussion on adaptive vs.

neutral explanations for the maintenance of biodiversity in ecological communities (Hubbell 2001; McGill 2003a; Volkov *et al.* 2003; Etienne & Olf 2004b). As the models corresponding to these explanations seem to fit at least some observed species-abundance patterns equally well (Harte 2003), it is even more urgent to develop powerful statistical tests that can detect any differences between these distributions.

Although the study of SAD was originally purely statistical (Motomura 1932; Fisher *et al.* 1943; Preston 1948, 1962) and statistical distributions without any mechanistic underpinning have continued to receive attention (Pielou 1975; Hengeveld & Stam 1978), surprisingly little attention has lately been paid to proper statistical analysis (McGill 2003b); sometimes distributions are even fitted by eye (Hubbell 2001). Yet, with computers becoming ever more powerful and mathematical and statistical tools being rapidly developed, it is now possible to carefully scrutinize the various models of community structure statistically, using the full information content in the data. Classical statistical tools can be used for this purpose (McGill 2003a; Volkov *et al.* 2003), but in this paper we take a Bayesian approach because of its great utility in parameter estimation and model comparison (Ellison 1996, 2004) and the fact that it produces a full probability distribution of parameter values rather than point estimates. To our knowledge, it has never been applied to species-abundance data before (except our preliminary exercises in Etienne & Olf 2004b).

First we describe the distributions to be compared in this Bayesian framework. We chose three distributions that are strongly rooted in the ecological literature. The first distribution is the discrete broken stick (DBS) distribution, which is based on broken stick model of MacArthur (1957, 1960). The second distribution is the lognormal distribution, or better (see below), the multivariate Poisson lognormal distribution (MPLN), that results from niche arguments (Bulmer 1974; May 1975), and has been used since the classical work by Preston (1948, 1962). The third distribution is the zero-sum multinomial (ZSM) distribution, which is derived from the neutral model of biodiversity (Hubbell 2001; McGill 2003a; Volkov *et al.* 2003; Etienne & Olf 2004b; Etienne 2005) of which a limiting distribution was introduced by Fisher *et al.* (1943). The lognormal and the ZSM have recently dominated the literature on determinants of biodiversity. We provide explicit formulas for the probability of observing a particular species-abundance data set in each model. For ZSM this formula has only recently been reported (Etienne 2005), whereas for DBS and MPLN these formulae are presented here for the first time. Then we briefly review the Bayesian statistical framework and the particular method that we employ, i.e. Markov Chain Monte Carlo (MCMC) simulation (Chen *et al.*

2000; for an ecological setting see Ter Braak & Etienne 2003; Etienne *et al.* 2004). We specify the ingredients of MCMC for each distribution (DBS, MPLN and ZSM). Finally, we apply our approach to a species-abundance data set of a neotropical forest (Condit *et al.* 1996, 2002), that has rapidly developed into a 'classical' data set for community model comparisons and a data set of a neotropical freshwater fish community (Winemiller 1990). To study one of our results in more detail, we also apply our approach to simulated data sets. We end the paper with a discussion of our approach and our results.

METHODS

Three models of community structure

MacArthur's broken-stick model: DBS

MacArthur (1957, 1960) introduced the broken-stick hypothesis in which the available resources for S species are likened to a stick of unit length which is randomly broken into S pieces, the length of each piece representing the relative abundance of a species. Sugihara (1980), extending work by Bulmer (1974), introduced a variation on MacArthur's broken-stick model. While in MacArthur's hypothesis the stick is *simultaneously* broken into S pieces, in Sugihara's version the stick is *sequentially* broken into fragments. These descriptions are mathematically equivalent (Tokeshi 1990, 1993), but may have different interpretations. Although MacArthur (1957, 1960) envisaged the S pieces as niches (one for each species), the random simultaneous breakage seems a rather arbitrary division of resources and may therefore also be viewed as a neutral model. However, the sequential breakage can be interpreted as a model of the ecological/evolutionary process by which niches arise, that is, an adaptive model.

MacArthur (1957, 1960) allowed breakage to occur anywhere on the stick, so his model gave a continuous description of (relative) abundances. In this paper, we use a discrete version of this model, in which the stick has integer length J , corresponding to a total number of J individuals, and breakage can only occur at integer positions $(1, \dots, J-1)$, such that abundances also have integer values. Multiple breakages at the same point are not allowed. The probability that one observes a specific SAD $D = (n_1, n_2, \dots, n_S)$ given the number of species S and the total number of individuals J is then described by a distribution that we will call the DBS distribution,

$$P_{\text{DBS}}[D|S, J] = \frac{\frac{J!}{\prod_{j=1}^S \Phi_j!}}{\binom{J-1}{S-1}} \quad (1)$$

where Φ_j is the number of species in the community that have abundance j . eqn 1 can be understood as

follows. Combinatorics teaches us that there are $\binom{J-1}{S-1} = \frac{(J-1)!}{(J-S)!(S-1)!}$ ways in which a stick of length J can be broken in S pieces of positive integer length. There are $\frac{S!}{\prod_{j=1}^S \Phi_j!}$ ways in which the stick can be broken that result in the same species-abundance data-set D . The probability that a given species-abundance data set is found is therefore given by the ratio of the latter to the former, so we get eqn 1. Note that eqn 1 does not contain any parameters and is conditioned on both S and J . The latter conditioning means that all resources are used at all times. MacArthur (1957, 1960) confronted his model with bird data and demonstrated that his model fits some bird species-abundance data fairly well.

May's niche-based model: MPLN

In the adaptive (niche-based) perspective, all species are considered to be functionally different, with different physiological, morphological and/or life-history traits. There are two different arguments as to why this leads to a lognormal SAD, i.e. a symmetric and unimodal distribution on a logarithmic abundance scale. The first argument comes from May (1975). The abundance of a species is governed by many more-or-less independent factors that interact multiplicatively rather than additively. Regardless of the distribution of these factors, the product of many such factors is asymptotically lognormally distributed, according to the Central Limit Theorem (CLT). The second argument is due to Bulmer (1974), but often attributed to Sugihara (1980). Consider the sequential breakage of the broken-stick model (an adaptive model). Now, the only requirement for a lognormal SAD to arise asymptotically (again invoking the CLT) is that the breakage occurs independently of fragment size. In MacArthur's model above this condition is *not* satisfied, because the probability of breakage is proportional to fragment size. For this reason MacArthur's broken-stick SAD has a more equitable SAD than the lognormal. The model where breakage occurs independently of fragment size has been dubbed the random fraction model by Tokeshi (1990). Both the random fraction model (Tokeshi 1990) and its limiting distribution, the lognormal model (Preston 1948, 1962), have been reported to fit SADs of various taxonomic groups well.

Strictly, the lognormal distribution only applies to the entire community, not to a sample from it. Preston (1948) already noted that there is a substantial sampling effect at low abundances: rare species are likely to be unobserved. He accounted for this effect by simply truncating the lognormal distribution with the so-called 'veil line'. Pielou (1969) and Bulmer (1974), and more recently Dewdney (1998) and Diserud & Engen (2000) refined this crude truncation. They remarked that if the abundance of each species i is a Poisson variate with mean λ_i and the λ_i s are independent observations

from a continuous probability distribution $f(\lambda)$ (for example the lognormal), then the probability that a species will have abundance n is given by

$$P_n = \int_0^\infty \frac{(\mu\lambda)^n e^{-\mu\lambda}}{n!} f(\lambda) d\lambda \tag{2}$$

where μ is the sampling intensity. Define $f(\lambda)$ as the lognormal distribution with mean M and variance V ,

$$f(\lambda) := \frac{1}{\lambda\sqrt{2\pi V}} e^{-\frac{(\ln\lambda-M)^2}{2V}} \tag{3}$$

where by $:=$ we denote a definition. Then eqn 2 becomes

$$\begin{aligned} P_n &= \frac{1}{n!\sqrt{2\pi V}} \int_0^\infty \frac{(\mu\lambda)^n e^{-\mu\lambda}}{\lambda} e^{-\frac{(\ln\lambda-M)^2}{2V}} d\lambda \\ &= \frac{1}{n!\sqrt{2\pi V}} \int_0^\infty \frac{x^n e^{-x}}{x} e^{-\frac{(\ln x-M')^2}{2V}} dx \end{aligned} \tag{4}$$

where $x := \mu\lambda$ and $M' := M + \ln \mu$. Equation 4 is known as the compound Poisson lognormal distribution (Bulmer 1974). We note that if $f(\lambda)$ were the Gamma distribution with shape parameter $k \rightarrow 0$, the compound distribution would be the negative binomial with shape parameter $k \rightarrow 0$, which is Fisher's logarithmic distribution (Bulmer 1974; Watterson 1974; Engen & Lande 1996b). For $k = 1$ we obtain the broken-stick model of MacArthur (Bulmer 1974; Engen & Lande 1996b). As zero abundances are not observed we must truncate the compound Poisson lognormal by dividing P_n by $1 - P_0$ (Bulmer 1974; Diserud & Engen 2000).

The probability that one observes a specific SAD $D = (n_1, n_2, \dots, n_S)$ given the number of species S and the model parameters $\Theta_{\text{MPLN}} = (M', V)$, i.e. the likelihood, is the MPLN. It is given by

$$P[D|\Theta_{\text{MPLN}}, S] = \frac{S!}{\prod_{j=1}^S \Phi_j!} \prod_{i=1}^S \frac{P_{n_i}}{1 - P_0} \tag{5}$$

In eqn 5 we have multiplied by $\frac{S!}{\prod_{j=1}^S \Phi_j!}$ for the same reason as in eqn 1 to take into account that the species labels $1, \dots, S$ are arbitrary, so we consider $(n_1, n_2, n_3, \dots, n_S)$ to be equivalent to $(n_2, n_1, n_3, \dots, n_S)$. Note also that the likelihood is conditioned on the number of species S and that the model parameters M' and V have no biologically interesting meaning (unlike the parameters of the neutral model, see below).

For comparison with DBS and ZSM it is necessary to condition both on the number of species S and on the total number of individuals, $J := \sum_{i=1}^S n_i$ (see below). This requires the probability distribution $P[J|\Theta_{\text{MPLN}}, S]$. This is the sum of the probabilities of all possible combinations of (n_1, \dots, n_S) that have $\sum_{i=1}^S n_i = J$. In this case,

$$P[J|\Theta_{\text{MPLN}}, S] = \sum_{\{n_1, \dots, n_S | \sum_{i=1}^S n_i = J\}} \prod_{i=1}^S \frac{P_{n_i}}{1 - P_0} \quad (6)$$

Combining eqns 5 and 6 leads to

$$P[D|\Theta_{\text{MPLN}}, S, J] = \frac{P[D|\Theta_{\text{MPLN}}, S]}{P[J|\Theta_{\text{MPLN}}, S]} = \frac{\prod_{j=1}^S \Phi_j! \prod_{i=1}^S \frac{P_{n_i}}{1 - P_0}}{\sum_{\{n_1, \dots, n_S | \sum_{i=1}^S n_i = J\}} \prod_{i=1}^S \frac{P_{n_i}}{1 - P_0}} \quad (7)$$

A technical detail: mathematically, the probability distribution of the sum of S identically and independently distributed variables is called the convolution of these S probability distributions, that is, the product of these probability distributions in Fourier space. This mathematical trick allows for a faster computation of $P[J|\Theta_{\text{MPLN}}, S]$, because instead of an enormous sum of products, only a Fourier transform, a product, and an inverse Fourier transform need to be computed.

Hubbell's neutral model: ZSM

In the neutral perspective, all individuals regardless of species are considered functionally equivalent. SAD solely result from the stochastic processes of birth, death and speciation, and in some models, immigration. In the neutral model of Hubbell (2001), when individuals in a local community die, they are immediately replaced by offspring of other local individuals or by immigrants from the regional species pool (the metacommunity in Hubbell's terminology), such that the total number of individuals remains constant (the zero-sum assumption). The replacement probability is purely proportional to each species abundance in the local community (when replaced by a local), or in the regional species pool (when replaced by an immigrant); this proportionality is the neutrality assumption. The species abundances in the regional species pool are the result of a balance between speciation and extinction. Immigration is characterized by the immigration probability m and the SAD in the regional species pool is fully governed by the fundamental biodiversity number $\theta := 2J_M v'$, where J_M is the number of individuals in the regional species pool (also assumed to be constant) and v' is the dimensionless speciation rate scaled to the birth rate. Vallade & Houchmandzadeh (2003) define $\theta := \frac{v(J_M - 1)}{1 - v}$ which is, apart from a factor of 2, equivalent to definition of Hubbell (2001) in the limit that Hubbell takes in his derivation ($J_M \gg 1$) because $v' = \frac{v}{1 - v}$. The factor of 2 arises from whether or not multiple speciation events in one time-step are allowed.

The neutral model has been carefully studied analytically (Vallade & Houchmandzadeh 2003; Volkov *et al.* 2003; Alonso & McKane 2004; Etienne & Olf 2004b; McKane *et al.* 2004; Etienne 2005). Specifically, Etienne & Olf (2004b)

found a full and exact analytical expression for the multivariate probability that one observes a specific SAD $D = (n_1, n_2, \dots, n_S)$ given the number of individuals J and the model parameters $\Theta_{\text{ZSM}} = (\theta, m)$, i.e. the likelihood $P[D|\Theta_{\text{ZSM}}, J]$. As this expression was intractable in practice, it was further simplified to tractable proportions (Etienne 2005), yielding a new sampling formula for neutral biodiversity:

$$P[D|\Theta_{\text{ZSM}}, J] = \frac{J!}{\prod_{i=1}^S n_i \prod_{j=1}^J \Phi_j!} \frac{\theta^S}{(I)_J} \sum_{A=S}^J \frac{K(D, A) I^A}{(\theta)_A} \quad (8)$$

where Φ_j is the number of species with abundance j and I is related to the immigration probability m by $I = \frac{m(J-1)}{1-m}$. Furthermore, the notation $(x)_j$ is known as the Pochhammer symbol defined as

$$(x)_j := \prod_{i=1}^j (x + i - 1) = x(x + 1) \cdots (x + j - 1) \quad (9)$$

and $K(D, A)$ is a coefficient for each A that depends on the data set D . Etienne (2005) gives more details on eqn 8 and provides a formula and a program to calculate the $K(D, A)$.

The sampling formula eqn 8 reduces to the well-known Ewens sampling formula (Ewens 1972; Karlin & McGregor 1972; Tavaré & Ewens 1997; Hubbell 2001) in the limit $m \rightarrow 1$. Like the Ewens sampling formula (Hubbell 2001), it has an associated algorithm to generate samples from the distribution (Etienne 2005).

The preliminary Bayesian analysis of Etienne & Olf (2004b) was not based on eqn 8, but on its much more complicated predecessor, which could be handled, albeit with tremendous computational power, by introducing many latent variables in the parameter estimation procedure. With eqn 8 this is no longer necessary.

Equation 8 is conditioned on the number of individuals J , but for comparison with DBS and MPLN, we must also condition on the number of species S . The required probability, $P[S|\Theta_{\text{ZSM}}, J]$ is given by (see Supplementary Material)

$$P[S|\Theta_{\text{ZSM}}, J] = \sum_{A=S}^J \bar{\pi}(J, A) \frac{I^A}{(I)_J} \bar{\pi}(A, S) \frac{\theta^S}{(\theta)_A} \quad (10)$$

where $\bar{\pi}(J, A)$ and $\bar{\pi}(A, S)$ are unsigned Stirling number of the first kind. Therefore we obtain

$$P[D|\Theta_{\text{ZSM}}, S, J] = \frac{P[D|\Theta_{\text{ZSM}}, J]}{P[S|\Theta_{\text{ZSM}}, J]} = \frac{\prod_{i=1}^S n_i! \prod_{j=1}^J \Phi_j! \frac{\theta^S}{(I)_J} \sum_{A=S}^J \frac{K(D, A) I^A}{(\theta)_A}}{\sum_{A=S}^J \bar{\pi}(J, A) \frac{I^A}{(I)_J} \bar{\pi}(A, S) \frac{\theta^S}{(\theta)_A}} \quad (11)$$

Like DBS and MPLN, ZSM has been reported to fit at least some species-abundance data well (Hubbell 2001).

Bayesian data analysis

In Bayesian data analysis (see e.g. Gelman *et al.* 2003; Ellison 2004) one uses probability distributions to express uncertainty about values of the model parameters (denoted here by Θ). Before data are collected, one usually has some idea of the range of possible parameter values, e.g. from a small study or expert knowledge. This prior knowledge can be translated into a probability distribution; this is the prior distribution $P[\Theta]$. If there is no prior knowledge, one can use so-called non-informative prior distributions (see Supplementary Material). If more information about the parameters becomes available, for example in the form of a new data set, this information can be used to update the probability distribution with Bayes' formula (see Supplementary Material). The resulting probability distribution is called the posterior distribution $P[\Theta|D]$. Thus, the Bayesian approach provides a natural framework to take into account both expert knowledge, preliminary data and extensive studies. Moreover, one does not merely obtain point estimates of parameter values, but a full probability distribution that is indispensable in uncertainty analysis of model predictions.

A very popular algorithm to obtain the posterior distribution $P[\Theta|D]$ is MCMC simulation (e.g. Chen *et al.* 2000). The idea of MCMC simulation is to let the parameters perform a random walk in parameter space according to a Markov chain that is set up in such a way that its stationary distribution is the posterior distribution. The Metropolis-Hastings (MH) algorithm is a particularly efficient implementation of MCMC. In each subsequent iteration u a new proposal Θ^* for the values of the model parameters is generated with a so-called jumping distribution. This proposal is then compared with the previous value Θ^{u-1} and accepted ($\Theta^u = \Theta^*$) or rejected ($\Theta^u = \Theta^{u-1}$) according to the acceptance ratio, which depends on the likelihoods $P[D|\Theta^*]$ and $P[D|\Theta^{u-1}]$, the prior distributions $P[\Theta^*]$ and $P[\Theta^{u-1}]$, and the jumping distributions $J_u[\Theta^*|\Theta^{u-1}]$ and $J_u[\Theta^{u-1}|\Theta^*]$. The list of accepted values constitutes the posterior distribution. See Supplementary Material for more details and references.

To compare our models we use the posterior Bayes factor (Aitkin 1991). It is defined as the ratio of posterior marginal likelihoods of two models; the posterior marginal likelihood is the likelihood averaged over the posterior distribution. See the Supplementary Material and references therein for more details.

Bayesian analysis of the three community models

May's lognormal model assumes S to be given and predicts a probability distribution for J given the model parameters.

Hubbell's neutral model assumes J to be given and predicts a probability for S given the parameters. MacArthur's broken-stick model assumes both J and S to be given. Comparing the three models thus seems like comparing apples and oranges. However, we can make a fair comparison of DBS, MPLN and ZSM by assuming both S and J to be given for all models, which we accomplish by additional conditioning on J for MPLN and on S for ZSM; the resulting conditional probabilities are given in eqns 7 and 11.

The posterior Bayes factor for ZSM and MPLN is thus (assuming that there is no *a priori* preference for any particular model; $P[M_{\text{DBS}}] = P[M_{\text{MPLN}}] = P[M_{\text{ZSM}}]$),

$$\begin{aligned} B_{\text{posterior,ZSM,MPLN}} &= \frac{P_{\text{posterior}}[D|S,J,M_{\text{ZSM}}]}{P_{\text{posterior}}[D|S,J,M_{\text{MPLN}}]} \\ &= \frac{\int P[D|\Theta_{\text{ZSM}},S,J,M_{\text{ZSM}}]P[\Theta_{\text{ZSM}}|D,S,J,M_{\text{ZSM}}]d\Theta_{\text{ZSM}}}{\int P[D|\Theta_{\text{MPLN}},S,J,M_{\text{MPLN}}]P[\Theta_{\text{MPLN}}|D,S,J,M_{\text{MPLN}}]d\Theta_{\text{MPLN}}} \\ &= \frac{\int P[D|\Theta_{\text{ZSM}},S,J,M_{\text{ZSM}}]P[\Theta_{\text{ZSM}}|D,J,M_{\text{ZSM}}]d\Theta_{\text{ZSM}}}{\int P[D|\Theta_{\text{MPLN}},S,J,M_{\text{MPLN}}]P[\Theta_{\text{MPLN}}|D,S,M_{\text{MPLN}}]d\Theta_{\text{MPLN}}} \\ &\approx \frac{\frac{1}{n_{\text{end,ZSM}} - n_{\text{burn-in,ZSM}}} \sum_{i=n_{\text{burn-in,ZSM}}}^{n_{\text{end,ZSM}}} P[D|\Theta_{\text{ZSM}}^i,S,J,M_{\text{ZSM}}]}{\frac{1}{n_{\text{end,MPLN}} - n_{\text{burn-in,MPLN}}} \sum_{i=n_{\text{burn-in,MPLN}}}^{n_{\text{end,MPLN}}} P[D|\Theta_{\text{MPLN}}^i,S,J,M_{\text{MPLN}}]} \end{aligned} \quad (12)$$

where in the third line we made use of the fact that both S and J are contained in the data, and in the fourth line we have made the connection with the MCMC output (with sample sizes $n_{\text{end},i} - n_{\text{burn-in},i}$ for each model i). The conditional likelihoods $P[D|\Theta_{\text{MPLN}}^i,S,J,M_{\text{MPLN}}]$ and $P[D|\Theta_{\text{ZSM}}^i,S,J,M_{\text{ZSM}}]$ are given by eqns 7 and 11 respectively, and $P[\Theta|D,S,M_{\text{MPLN}}]$ and $P[\Theta|D,J,M_{\text{ZSM}}]$ are the posteriors following from the MCMC simulations for each model separately. The Bayes factor of DBS to ZSM and MPLN is similar. As DBS has no parameters to average over, the posterior marginal likelihood of DBS is simply the likelihood given by eqn 1 and there is no MCMC simulation. To carry out the MH algorithm for MPLN and ZSM we must specify two more ingredients for each model additional to the likelihoods given by eqns 7 and 11: the jumping distribution and the prior distribution.

Jumping distribution

We chose to sample both parameters simultaneously for the MPLN as well as the ZSM. For the jumping distributions in the case of the MPLN, we chose the normal distribution for M' and the lognormal distribution for V , treating M' and V as independent. In the case of the ZSM we used the lognormal distribution for both θ and I , also treating these parameters as independent (no covariance). Covariance can be introduced to speed up convergence of the Markov Chain, but we chose not to do so, as this is a mere technical point, and we attempted to make this exposition as transparent and parsimonious with assumptions as possible.

Prior distribution

For the prior distributions we chose the non-informative Jeffreys (1961), with minor modifications. In the case of the MPLN, this does not lead to a closed-form expression. Therefore we used the Jeffreys prior for the lognormal distribution. Furthermore, we adopted the convention in scale-location parameters (Berger 1985) to use the product of the Jeffreys prior for each parameter separately instead of the joint prior that follows from eqn 8 in Appendix S2 in the Supplementary Material. This leads to

$$P[\Theta_{\text{MPLN}}|S] = \frac{1}{V} \tag{13}$$

where we conditioned on S for consistency (it has no influence on the prior).

For the ZSM calculating the Jeffreys prior is also cumbersome. Therefore we applied eqn 8 in the Supplementary Material where the expectation is taken with respect to the underlying species-ancestry distribution of Etienne & Olf (2004b). This results in (see Supplementary Material)

$$P[\Theta_{\text{ZSM}}|J] = \sqrt{\left(\frac{1}{\theta} \sum_{i=1}^J \sum_{k=i}^J \frac{\bar{s}(J, k) I^k}{(I)_J} \frac{i-1}{(\theta+i-1)^2} \right) \left(\frac{1}{I} \sum_{i=1}^J \frac{i-1}{(I+i-1)^2} \right)} \tag{14}$$

We stress again that because the posterior marginal likelihood is insensitive to the choice of priors, the exact choice of priors is not very important. We tried other non-informative priors and indeed found no substantial differences. In this paper we just report results using the priors given above.

Acceptance ratio

The acceptance ratio r (given by eqn 9 in the Supplementary Material) can now be computed. For the MPLN we have, using eqn 5,

$$\begin{aligned} r_{\Theta_{\text{MPLN}}} &= \frac{P[D|\Theta_{\text{MPLN}}^*, S]P[\Theta_{\text{MPLN}}^*|S]J_u[\Theta_{\text{MPLN}}^{u-1}|\Theta_{\text{MPLN}}^*]}{P[D|\Theta_{\text{MPLN}}^{u-1}, S]P[\Theta_{\text{MPLN}}^{u-1}|S]J_u[\Theta_{\text{MPLN}}^*|\Theta_{\text{MPLN}}^{u-1}]} \\ &= \frac{\prod_{i=1}^S \frac{P_{n_i}^*}{1-P_0^*} V^{u-1} V^*}{\prod_{i=1}^S \frac{P_{n_i}^{u-1}}{1-P_0^{u-1}} V^* V^{u-1}} \\ &= \left(\frac{1-P_0^{u-1}}{1-P_0^*} \right)^S \prod_{i=1}^S \frac{P_{n_i}^*}{P_{n_i}^{u-1}} \end{aligned} \tag{15}$$

For the ZSM we obtain, using eqn 8,

$$\begin{aligned} r_{\Theta_{\text{ZSM}}} &= \frac{P[D|\Theta_{\text{ZSM}}^*, J]P[\Theta_{\text{ZSM}}^*|J]J_u[\Theta_{\text{ZSM}}^{u-1}|\Theta_{\text{ZSM}}^*]}{P[D|\Theta_{\text{ZSM}}^{u-1}, J]P[\Theta_{\text{ZSM}}^{u-1}|J]J_u[\Theta_{\text{ZSM}}^*|\Theta_{\text{ZSM}}^{u-1}]} \\ &= \frac{\frac{J!}{\prod_{i=1}^S n_i! \prod_{j=1}^J \Phi_j!} \frac{\theta^{uS}}{(I^*)^J} \sum_{A=S}^J \frac{K(D, A) I^{uA}}{(\theta^*)^A}}{\frac{J!}{\prod_{i=1}^S n_i! \prod_{j=1}^J \Phi_j!} \frac{(\theta^{u-1})^S}{(I^{u-1})^J} \sum_{A=S}^J \frac{K(D, A) (I^{u-1})^A}{(\theta^{u-1})^A}} \\ &\quad \times \frac{P_{\text{Jeffreys}}[\theta^*, I^*|J]}{P_{\text{Jeffreys}}[\theta^{u-1}, I^{u-1}|J]} \frac{\theta^* I^*}{\theta^{u-1} I^{u-1}} \\ &= \left(\frac{\theta^*}{\theta^{u-1}} \right)^S \frac{(I^{u-1})^J}{(I^*)^J} \frac{\sum_{A=S}^J \frac{K(D, A) I^{uA}}{(\theta^*)^A}}{\sum_{A=S}^J \frac{K(D, A) (I^{u-1})^A}{(\theta^{u-1})^A}} \frac{P_{\text{Jeffreys}}[\theta^*, I^*|J]}{P_{\text{Jeffreys}}[\theta^{u-1}, I^{u-1}|J]} \end{aligned} \tag{16}$$

with P_{Jeffreys} given by eqn 14.

RESULTS: TWO DATA SETS

We applied our Bayesian approach to several censuses of a well-studied neotropical tree community on a 50 ha plot on Barro Colorado Island (BCI), Panama (Condit *et al.* 1996, 2002) and to a neotropical freshwater fish community in Caño Maraca (CM), a creek-floodplain ecosystem in Venezuela (Winemiller 1990). The BCI data set has been the subject of previous comparisons between lognormal and neutral models (Hubbell 2001; McGill 2003a; Volkov *et al.* 2003; Etienne & Olf 2004b). We analysed the species-abundance data of each of five censuses (1982, 1985, 1990, 1995 and 2000), as they appeared on <http://ctfs.si.edu/datasets/bci/abundance/bciN100.html> in September 2004. These data are different from those reported by Condit *et al.* (1996, 2002), because occasionally errors are found that require correction in old data sets, and because species identifications are sometimes changed (R. Condit, personal communication). The CM data set is a new data set in model comparisons, only used by Hubbell (2001) to obtain parameter estimates for his neutral model. The BCI 1995 census and the CM data are pictured in Fig. 1, while summary statistics of all data are given in Table 1. It is clear that rare species are uncommon in CM.

Sample size of the MCMC was around 100 000 (MPLN) and 2 000 000 (ZSM), and the burn-in was set at 1000 (MPLN) and 10 000 (ZSM). The lag (number of iterations between recorded iterations) was set at five to reduce autocorrelation and to save disk space. The posterior distributions of the model parameters for both the MPLN and the ZSM and for BCI (1995 census) and CM are shown in Fig. 2. Table 1 gives values for the *a posteriori* most probable combination of the ZSM parameters θ_{opt} and m_{opt} . These values are close to the maximum likelihood values (Etienne 2005) and hence demonstrate the weak role of the prior distribution.

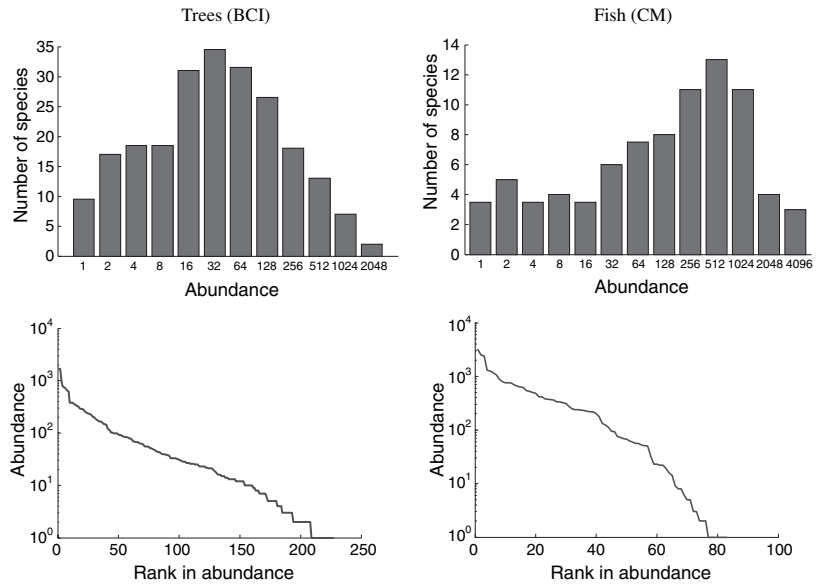


Figure 1 Species-abundance data for a 50 ha plot of neotropical forest on Barro Colorado Island, Panama (left column) and for a neotropical fish community in Caño Maraca, Venezuela (right column), displayed in two ways: species-abundance histogram with Preston binning (top row) and dominance-diversity plot also known as rank-abundance plot (bottom row).

The posterior marginal likelihoods and the posterior Bayes factors are also listed in Table 1. It is evident that the ZSM outperforms the MPLN for the BCI tree community, but by no means decisively, in line with previous assertions (Harte 2003; Volkov *et al.* 2003; Etienne & Olf 2004b). However, we find an overwhelmingly better fit of the ZSM relative to the MPLN for the CM fish community.

This result led us to hypothesize that the (dis)similarity of the MPLN and the ZSM is parameter dependent: for low immigration probability the ZSM diverges strongly from the MPLN as suggested by the CM data set, and perhaps also for high immigration probability. To investigate this idea we produced several artificial SAD with the neutral model with

parameters $\theta = 50$ and $m = 0.01$, $m = 0.1$ and $m = 1$, five for each value of m (see Etienne 2005 how such simulation data sets can be generated); sample size was set at $J = 20,000$. For all these simulated data sets and computed the posterior Bayes factor $B_{ZSM,MPLN}$ along with some other analyses. The results are listed in Table 2. First of all, the Bayes factor is always in favour of the ZSM, but this comes as no surprise, because one of the advantages of Bayesian model selection is that it is consistent, i.e. under mild conditions it selects the true model if this model is among the models to be compared and sufficient data are available (Berger 1985). But more importantly, the results support our hypothesis: for low and high immigration probabilities the fits of ZSM and MPLN differ substantially, whereas for

Table 1 Summary statistics of the example data sets

Data set		Descriptive statistics			Parameter estimates		$\ln P_{\text{posterior}}[D S,J,M]$			Model comparison	
Location	Census	J	S	T	θ_{opt}	m_{opt}	M_{ZSM}	M_{MPLN}	M_{DBS}	$B_{\text{ZSM,MPLN}}$	Conclusion
BCI (trees)	1982	20881	238	25	50.0	0.09	-304.498	-306.660	-418.540	8.7	ZSM, +
	1985	20719	237	29	49.7	0.10	-298.700	-301.104	-414.254	11.1	ZSM, +
	1990	21233	229	18	50.4	0.08	-299.108	-299.997	-401.577	2.4	ZSM, 0
	1995	21455	227	19	51.0	0.07	-307.025	-308.387	-408.985	3.9	ZSM, +
	2000	21205	226	14	48.1	0.08	-300.714	-300.272	-399.269	0.44 = $\frac{1}{1.56}$	MPLN, 0
CM (fish)	1984	28975	83	7	30.3	0.0024	-254.039	-262.402	-288.167	$4.3 \cdot 10^3$	ZSM, +++

J , number of individuals; S , number of species; T , number of singleton species; ZSM posterior mode ($\theta_{\text{opt}}, m_{\text{opt}}$), logarithm of the posterior marginal likelihoods of the models ZSM, MPLN and DBS, posterior Bayes factor $B_{\text{ZSM,MPLN}}$ and the conclusion on which model gives the best fit and the strength of the evidence for this conclusion indicated by 0, +, ++ or +++.

A Bayes factor between 1 and 3 implies that there is hardly any difference between the models, whereas larger values increasingly imply that the data provide evidence for the model in the numerator. See Table I in the Supplementary Material for more information on the interpretation of the Bayes factor.

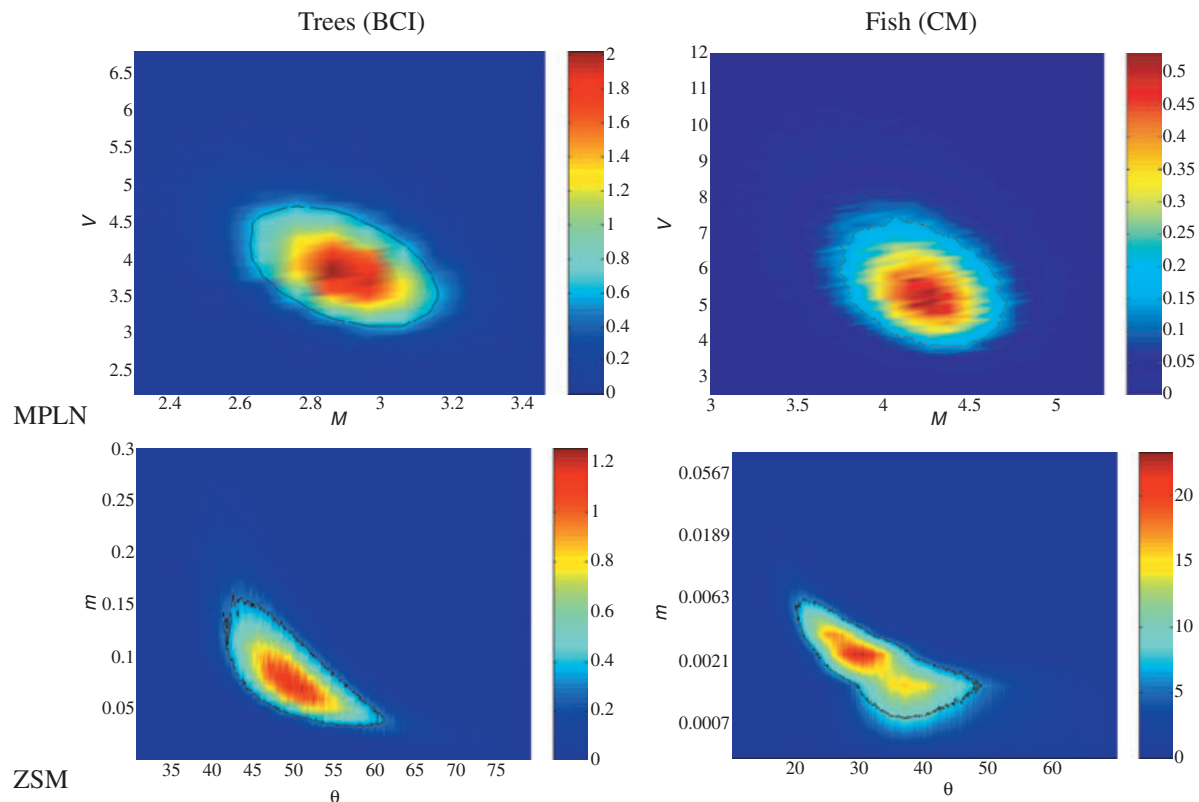


Figure 2 Posterior density plots of the model parameters corresponding to MPLN (top row) and to ZSM (bottom row) for the BCI tree data (left column) and the CM fish data (right column). The redder the colour, the higher the posterior density. Note the logarithmic axes in the ZSM plot of CM (bottom right panel). The black contour line girds the 68% credible region (i.e. the highest 68% of the posterior probability).

intermediate immigration probabilities, they are much more similar.

DISCUSSION

We have presented a Bayesian framework to compare different models of community structure on the basis of species-abundance data. We illustrated this by comparing the MPLN and ZSM distributions using neotropical tree and fish data. These distributions are considered as emergent properties of adaptive and neutral models of community structure respectively (Hubbell 2001; McGill 2003a; Volkov *et al.* 2003; Etienne & Olf 2004b). Our results suggest that the frequently used BCI tree community data do not point decisively in the direction of one of these models. This is in line with previous assertions (Harte 2003; Volkov *et al.* 2003; Etienne & Olf 2004b) which might suggest that the MPLN and ZSM are really not very different distributions. However, our example of the extremely dispersal limited CM fish community shows that species-abundance data are sometimes much better described by the ZSM. Our analysis of simulated data sets supports the idea that the resemblance

of the MPLN and the ZSM is parameter dependent: for low and high immigration probability the MPLN diverges strongly from the ZSM, whereas for intermediate immigration probability they are similar. This explains statements that species-abundance data cannot discriminate between models (e.g. Harte 2003), as these are mainly based on the mildly dispersal limited BCI tree community.

It is ironic that the adaptive, niche-based MPLN is so inferior to the ZSM for a fish community that is generally accepted to be controlled by strong adaptive mechanisms (Winemiller 1990). This result suggests that dispersal may be a key factor in further shaping community patterns. In contrast to metapopulation ecology that emphasizes the importance of dispersal, community ecology has traditionally focused on local interactions and mostly ignored dispersal. Exceptions include competition-colonization models (Pacala & Tilman 1994; Tilman 1994; Kinzig *et al.* 1999; Kneitel & Chase 2004) and metacommunity models (Loreau & Mouquet 1999; Mouquet & Loreau 2002, 2003; Marquet *et al.* 2003; review by Leibold *et al.* 2004). Our finding may also be due to the relatively poor mechanistic basis of the MPLN compared with the ZSM; hence, we

Table 2 Summary statistics of the simulation data sets

True parameters		Diversity		Parameter estimates				ln $P_{\text{posterior}}[D S,J,M]$			Model comparison	
θ_{true}	m_{true}	S	$E_{\text{ZSM}}[S]$	θ_{ML}	m_{ML}	$\theta_{\text{posterior}}$	$m_{\text{posterior}}$	M_{ZSM}	M_{MPLN}	M_{DBS}	$B_{\text{ZSM,MPLN}}$	Conclusion
50	0.01	138	149.2	50.8	0.0071	44.5	0.012	-281.805	-289.450	-312.276	$2.1 \cdot 10^3$	+++
50	0.01	158	149.2	42.6	0.021	48.8	0.012	-284.696	-289.938	-330.772	$1.9 \cdot 10^2$	+++
50	0.01	128	149.2	29.4	0.031	36.0	0.011	-263.676	-267.510	-311.333	$4.6 \cdot 10^1$	++
50	0.01	131	149.2	46.5	0.0072	41.5	0.011	-271.995	-279.402	-313.565	$1.6 \cdot 10^3$	+++
50	0.01	136	149.2	42.4	0.011	40.7	0.011	-284.136	-290.885	-320.598	$8.5 \cdot 10^2$	+++
50	0.10	228	232.4	48.1	0.11	51.7	0.062	-306.202	-311.477	-392.126	$2.0 \cdot 10^2$	+++
50	0.10	233	232.4	48.2	0.13	49.3	0.092	-292.000	-294.084	-395.645	$8.0 \cdot 10^0$	+
50	0.10	217	232.4	45.6	0.10	48.7	0.072	-292.797	-294.205	-376.229	$4.0 \cdot 10^0$	+
50	0.10	207	232.4	48.0	0.058	51.0	0.042	-295.001	-299.084	-372.036	$2.2 \cdot 10^1$	++
50	0.10	224	232.4	44.7	0.15	48.6	0.082	-296.799	-297.239	-394.789	$1.6 \cdot 10^0$	0
50	1.00	312	300.2	53.4	0.83	55.9	0.53	-304.888	-311.504	-493.329	$7.5 \cdot 10^2$	+++
50	1.00	320	300.2	54.0	1.00	56.3	0.87	-302.409	-308.082	-540.685	$2.9 \cdot 10^2$	+++
50	1.00	316	300.2	55.0	0.73	57.0	0.54	-314.797	-320.036	-509.684	$1.9 \cdot 10^2$	+++
50	1.00	322	300.2	55.9	0.77	58.1	0.61	-311.100	-321.959	-478.052	$5.2 \cdot 10^4$	+++
50	1.00	303	300.2	50.7	1.00	50.9	0.90	-300.620	-306.569	-519.149	$3.8 \cdot 10^2$	+++

True parameter values θ_{true} and m_{true} used to generate the data sets, observed number of species S , expected number of species $E_{\text{ZSM}}[S]$ for the true model parameters, ZSM maximum likelihood parameters ($\theta_{\text{ML}}, m_{\text{ML}}$), ZSM posterior mode ($\theta_{\text{opt}}, m_{\text{opt}}$), logarithm of the posterior marginal likelihoods of the models ZSM, MPLN and DBS, posterior Bayes factor $B_{\text{ZSM,MPLN}}$ and the strength of the evidence for the conclusion that the ZSM provides the better fit, indicated by 0, +, ++ or +++.

advocate further development of niche-based models (with or without dispersal) that yield explicit predictions for species abundances and have biologically meaningful parameters.

When comparing several models using likelihoods the conditioning on J and S is essential. Any comparison between adaptive and neutral models will need this. Given a sample area, the adaptive framework assumes that the sample area holds a fixed number of niches and hence species S , while the neutral framework assumes that the sample area provides resources for a fixed number of individuals J . The latter assumption constrains the SAD much more than the former. Without conditioning, the ZSM provides a much better fit than the MPLN for both BCI and CM, but then the information content of the assumption that J or S is fixed is not accounted for. After conditioning the better fit of the ZSM almost disappears for BCI, but still remains for CM. In the above-cited previous studies that compared the neutral model to the lognormal (they did not take Poisson sampling into account), no mention is made of this conditioning.

The MPLN and the ZSM are not the only distributions associated with niche and neutral models respectively. There are various distributions following from different niche-based sequential breakage models (MacArthur 1957, 1960; Sugihara 1980; Tokeshi 1990, 1993, 1996) and stochastic abundance models (Engen & Lande 1996a,b, Diserud &

Engen 2000). Different neutral models may lead to various distributions, including Fisher's logarithmic distribution (Kendall 1948; Watterson 1974; Caswell 1976; Dewdney 2000, see also the review by Chave 2004). The connection between neutral or niche and distribution is not completely exclusive. Watterson (1974) noted that a model from the adaptive perspective having an underlying Gamma distribution (see above) leads to the logarithmic distribution, which is usually associated with neutral models. In addition, as we noted above, MacArthur's (1957) broken-stick model is usually interpreted as adaptive (niche-based) when viewed as a sequential breakage model, but may just as well be regarded as neutral when viewed as stochastic simultaneous breakage model (Tokeshi 1990, 1993); this may be related to the fact that in many stick-breaking models there is conditioning on both J and S : the stick represents the zero-sum assumption and the number of breakages defines the number of species. Furthermore, in the derivation of the MPLN differences between species are not assumed; species are really symmetric and could thus be considered neutral (Hubbell & Lake 2003). This does not prevent us from comparing different distributions (such as MPLN and ZSM) on the basis of data, but it does imply a warning that we should be cautious in interpreting the results of our comparison in terms of underlying mechanisms: indeed pattern does not equal process (Cohen 1968; Clinchy *et al.* 2002). That is, we may be able to decide whether the ZSM

fits the data better than the MPLN, but this does not immediately mean that neutral models are better representations of community structure; perhaps, in the end, we may only be testing whether the zero-sum assumption holds. Yet, as we noted in the introduction, often we have no other sources of information available, so we argue that we had better explore these data extensively, but carefully, with the best possible methods. We believe the Bayesian approach provides a very good method, and the Bayes factor is an easy-to-interpret, more or less parameter independent (as it integrates over the posterior distribution) tool to compare different models.

Moreover, even if our approach is not used to compare different community models, it is still very useful in obtaining estimates of parameter values when one assumes a certain model to hold, based on other grounds (McGill 2003b; Nee 2003). For example, if the neutral model is supposed to be valid and one wants to assess the amount of dispersal limitation in a community, then our approach yields a full probability distribution for the immigration probability m . It tells one to what extent observed low local diversity is the result of high dispersal limitation or the result of low regional diversity. Parameter estimation can be combined with model comparison in unified models, i.e. models that contain elements of both niche and neutral theory; the estimated value and uncertainty of a parameter then gives an indication of the importance of the corresponding element. Such unified models are still rare, but some interesting suggestions have already been made (Marquet *et al.* 2003; Etienne & Olff 2004a).

Our Bayesian approach uses the likelihood of the full data set to obtain a measure of the goodness-of-fit (posterior marginal likelihood). In other words, it attempts to fit the dominance-diversity curve (bottom row of Fig. 1), whereas previous approaches (McGill 2003a; Volkov *et al.* 2003) have tried to fit the species-abundance histogram (top row of Fig. 1). In the latter case not all data are used, and the results may contain artefacts due to *ad hoc* choices of bin size (Lobo & Favila 1999). McGill (2003a) suggested to minimize these artefacts by using different bin sizes and averaging the results, but it is not as elegant and complete as fitting the dominance-diversity curve. Harte's (2003) statements that the fitted curves of lognormal (note that this is the lognormal, not the MPLN) and ZSM (note that this is the approximation of Volkov *et al.* 2003, not the exact multivariate distribution eqn 8) are practically indistinguishable and therefore macroecological study of this pattern is futile, are also based on fits of SADs (particularly of BCI), and should, for this reason alone, be interpreted with care. Our study shows that there is even more reason to hold on to species-abundance data, as they may sometimes contain valuable information about community structure. We therefore advocate that our approach

should be applied to a broad range of ecological groups of species under various environmental and geographical conditions.

ACKNOWLEDGEMENTS

We thank M.E.F. Apol, D. Alonso and two anonymous referees for their comments on an earlier version of this paper, R. Holtkamp and M.E.F. Apol for kindly allowing use of their computers, and D. Alonso, C. Bos, B. Carlin, M. Carlson, S. Chib, O. Diserud, P. Laud, A. McKane and I. Ntizoufras for discussions on their Bayesian and modelling approaches. Data of the neotropical fish assemblage in Caño Maraca, Venezuela, were kindly provided by K.O. Winemiller. See Winemiller (1990) for full acknowledgement. The Forest Dynamics Plot of BCI has been made possible through the generous support of the US National Science Foundation, The John D. and Catherine T. MacArthur Foundation, and the Smithsonian Tropical Research Institute and through the hard work of over 100 people from 10 countries over the past two decades. The BCI Forest Dynamics Plot is part the Center for Tropical Forest Science, a global network of large-scale demographic tree plots. Rampal Etienne was partly funded by the Priority Program 'Biodiversity in Disturbed Ecosystems' of the Netherlands Organization for Scientific Research. Part of the work for this article was carried out at the Tropical Nature Conservation and Vertebrate Ecology group, Wageningen University and Research Centre, Wageningen, The Netherlands, and at Environmental Science, Faculty of Geosciences, Utrecht University, Utrecht, The Netherlands.

SUPPLEMENTARY MATERIAL

The following material is available from <http://www.blackwellpublishing.com/products/journals/suppmat/ELE/ELE745/ELE745sm.htm>

Appendix S1 Bayesian model comparison.

Appendix S2 Derivation of eqns 10 and 14.

REFERENCES

- Aitkin, M. (1991). Posterior Bayes factors. *J. R. Stat. Soc. B*, 53, 111–142.
- Alonso, D. & McKane, A.J. (2004). Sampling Hubbell's neutral theory of biodiversity. *Ecol. Lett.*, 7, 901–910.
- Berger, J.O. (1985). *Statistical Decision Theory: Foundations, Concepts and Methods*. Springer, Berlin, Germany.
- Blackburn, T.M. & Gaston, K.J. (2003). *Macroecology: Concepts and Consequences*. the 43rd Annual Symposium of the British Ecological Society, held at the University of Birmingham, 10–17 April 2002. Blackwell, Maiden, MA.
- Brown, J.H. (1995). *Macroecology*. University of Chicago Press, Chicago, IL.

- Brown, J.H. & Maurer, B.A. (1989). Macroecology: the division of food and space among species on continents. *Science*, 243, 1145–1150.
- Bulmer, M.G. (1974). On fitting the Poisson lognormal distribution to species-abundance data. *Biometrics*, 30, 101–110.
- Caswell, H. (1976). Community structure: a neutral model analysis. *Ecol. Monogr.*, 46, 327–354.
- Chapin, F.S., Zavaleta, E.S., Eviner, V.T., Naylor, R.L., Vitousek, P.M., Reynolds, H.L. *et al.* (2000). Consequences of changing biodiversity. *Nature*, 405, 234–242.
- Chave, J. (2004). Neutral theory and community ecology. *Ecol. Lett.*, 7, 241–253.
- Chen, M.-H., Shao, Q.-M. & Ibrahim, J.G. (2000). *Monte Carlo Methods in Bayesian Computation*. Springer, New York, NY.
- Clinchy, M., Haydon, D.T. & Smith, A.T. (2002). Pattern does not equal process: what does patch occupancy really tell us about metapopulation dynamics? *Am. Nat.*, 159, 351–362.
- Cohen, J.E. (1968). Alternate derivations of a species-abundance relation. *Am. Nat.*, 102, 165–172.
- Condit, R., Hubbell, S.P. & Foster, R.B. (1996). Changes in tree species abundance in a neotropical forest: impact of climate change. *J. Trop. Ecol.*, 12, 231–256. Available at: <http://www.ctfs.si.edu/data/data.htm>.
- Condit, R., Pitman, N., Leigh, E.G., Chave, J., Terborgh, J., Foster, R.B. *et al.* (2002). Beta-diversity in tropical forest trees. *Science*, 295, 666–669.
- Costanza, R., d'Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon, B. *et al.* (1997). The value of the world's ecosystem services and natural capital. *Nature*, 387, 253–260.
- Dewdney, A.K. (1998). A general theory of the sampling process with applications to the "veil line". *Theor. Popul. Biol.*, 54, 294–302.
- Dewdney, A.K. (2000). A dynamical model of communities and a new species-abundance distribution. *Biol. Bull.*, 35, 152–165.
- Diserud, O.H. & Engen, S. (2000). A general and dynamic species abundance model, embracing the lognormal and the gamma models. *Am. Nat.*, 155, 497–511.
- Ellison, A.M. (1996). An introduction to Bayesian inference for ecological research and environmental decision-making. *Ecol. Appl.*, 6, 1036–1046.
- Ellison, A.M. (2004). Bayesian inference in ecology. *Ecol. Lett.*, 7, 509–520.
- Engen, S. & Lande, R. (1996a). Population dynamic models generating the lognormal species abundance distribution. *Math. Biosci.*, 132, 169–183.
- Engen, S. & Lande, R. (1996b). Population dynamic models generating the species abundance distributions of the Gamma type. *J. Theor. Biol.*, 178, 325–331.
- Etienne, R.S. (2005). A new sampling formula for neutral biodiversity. *Ecol. Lett.*, 8, 253–260.
- Etienne, R.S. & Olf, H. (2004a). How dispersal limitation shapes species – body size distributions in local communities. *Am. Nat.*, 163, 69–83.
- Etienne, R.S. & Olf, H. (2004b). A novel genealogical approach to neutral biodiversity theory. *Ecol. Lett.*, 7, 170–175.
- Etienne, R.S., ter Braak, C.J.F. & Vos, C.C. (2004). Application of stochastic patch occupancy models to real metapopulations. In: *Ecology, Genetics, and Evolution of Metapopulations* (eds Hanski, I. & Gaggiotti, O.E.). Elsevier Academic Press, London, UK, pp. 105–132.
- Ewens, W.J. (1972). The sampling theory of selectively neutral alleles. *Theor. Popul. Biol.*, 3, 87–112.
- Fisher, R.A., Corbet, A.S. & Williams, C.B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.*, 12, 42–58.
- Gaston, K.J. & Blackburn, T.M. (1999). Macroecology: a critique. *Oikos*, 84, 353–368.
- Gaston, K.J. & Blackburn, T.M. (2000). *Pattern and Process in Macroecology*. Blackwell, Oxford, UK.
- Gelman, A., Carlin, J.B., Stern, H.S. & Rubin, D.B. (2003). *Bayesian Data Analysis*. Chapman & Hall/CRC, London, UK.
- Harte, J. (2003). Tail of death and resurrection. *Nature*, 424, 1006–1007.
- Hengeveld, R. & Stam, A.J. (1978). Models explaining Fisher's log-series abundance curve. Proceedings of the Koninklijke Academie voor Wetenschappen, 81C, 415–427.
- Hubbell, S.P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton, NJ.
- Hubbell, S.P. & Lake, J. (2003). The neutral theory of biodiversity and biogeography, and beyond. In: *Macroecology: Patterns and Processes* (eds Blackburn, T. & Gaston, K.). Blackwell, Oxford, UK, pp. 45–63.
- Jeffreys, H. (1961). *Theory of Probability*. Oxford University Press, Oxford, U.K.
- Karlin, S. & McGregor, J. (1972). Addendum to a paper of W. Ewens. *Theor. Popul. Biol.*, 3, 113–116.
- Kendall, R.G. (1948). On some modes of population growth giving rise to R.A. Fisher's logarithmic series distribution. *Biometrika*, 35, 6–15.
- Kinzig, A.P., Levin, S.A., Dushoff, J. & Pacala, S. (1999). Limiting similarity, species packing, and system stability for hierarchical competition-colonization models. *Am. Nat.*, 153, 371–383.
- Kneitel, J.M. & Chase, J.M. (2004). Trade-offs in community ecology: linking spatial scales and species coexistence. *Ecol. Lett.*, 7, 69–80.
- Lawton, J. (1999). Are there general laws in ecology? *Oikos*, 84, 177–192.
- Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.F. *et al.* (2004). The metacommunity concept: a framework for multi-scale community ecology. *Ecol. Lett.*, 7, 601–613.
- Lobo, J.M. & Favila, M.E. (1999). Different ways of constructing octaves and their consequences on the prevalences of the bimodal species abundance distribution. *Oikos*, 87, 321–326.
- Loreau, M. & Mouquet, N. (1999). Immigration and the maintenance of local species diversity. *Am. Nat.*, 154, 427–440.
- MacArthur, R.H. (1957). On the relative abundance of bird species. *Proc. Natl Acad. Sci. USA*, 43, 293–295.
- MacArthur, R.H. (1960). On the relative abundance of species. *Am. Nat.*, 94, 25–36.
- Marquet, P.A., Keymer, J.E. & Cofré, H. (2003). Breaking the stick in space: of niche models, metacommunities and patterns in the relative abundance of species. In: *Macroecology: Patterns and Processes* (eds Blackburn, T. & Gaston, K.). Blackwell, Oxford, UK, pp. 64–84.
- Maurer, B.A. (1999). *Untangling Ecological Complexity: The Macroscopic Perspective*. University of Chicago Press, Chicago, IL.
- May, R.M. (1975). Patterns of species abundance and diversity. In: *Ecology and Evolution* (eds Cody, M.L. & Diamond, J.M.). Harvard University Press, Cambridge, MA, pp. 81–120.

- McGill, B.J. (2003a). A test of the unified neutral theory of biodiversity. *Nature*, 422, 881–885.
- McGill, B.J. (2003b). Strong and weak tests of macroecological theory. *Oikos*, 102, 679–685.
- McKane, A.J., Alonso, D. & Solé, R.V. (2004). Analytic solution of Hubbell's model of local community dynamics. *Theor. Popul. Biol.*, 65, 67–73.
- Motomura, I. (1932). A statistical treatment of communities. *Zoo. Mag. Tokyo*, 44, 379–383 (in Japanese).
- Mouquet, N & Loreau, M. (2002). Coexistence in metacommunities: the regional similarity hypothesis. *Am. Nat.*, 159, 420–426.
- Mouquet, N & Loreau, M. (2003). Community patterns in source-sink metacommunities. *Am. Nat.*, 162, 544–557.
- Nee, S. (2003). The unified phenomenological theory of biodiversity. In: *Macroecology: Patterns and Processes* (eds Blackburn, T. & Gaston, K.). Blackwell, Oxford, UK, pp. 31–41.
- Pacala, S.W. & Tilman, D. (1994). Limiting similarity in mechanistic and spatial models of plant competition in heterogeneous environments. *Am. Nat.*, 143, 222–257.
- Pielou, E.C. (1969). *An Introduction to Mathematical Ecology*. Wiley, New York, NY.
- Pielou, E.C. (1975). *Ecological Diversity*. Wiley, New York, NY.
- Preston, F.W. (1948). The commonness, and rarity, of species. *Ecology*, 29, 254–283.
- Preston, F.W. (1962). The canonical distribution of commonness and rarity. Parts I and II. *Ecology*, 43, 185–215, 410–432.
- Sugihara, G. (1980). Minimal community structure: an explanation of species-abundance patterns. *Am. Nat.*, 116, 770–787.
- Tavaré, S. & Ewens, W.J. (1997). Multivariate Ewens distribution. In: *Discrete Multivariate Distributions* (eds Johnson, N.L., Kotz, S. & Balakrishnan, N.). Wiley, New York, NY, pp. 232–246.
- Ter Braak, C.J.F. & Etienne, R.S. (2003). Improved Bayesian analysis of metapopulation data with an application to a tree frog metapopulation. *Ecology*, 84, 231–241.
- Tilman, D. (1994). Competition and biodiversity in spatially structured habitats. *Ecology*, 75, 2–16.
- Tokeshi, M. (1990). Niche apportionment or random assortment – species abundance patterns revisited. *J. Anim. Ecol.*, 59, 1129–1146.
- Tokeshi, M. (1993). Species abundance patterns and community structure. *Adv. Ecol. Res.*, 24, 111–186.
- Tokeshi, M. (1996). Power fraction: a new explanation of relative abundance patterns in species-rich assemblages. *Oikos*, 75, 543–550.
- Vallade, M. & Houchmandzadeh, B. (2003). Analytical solution of a neutral model of biodiversity. *Phys. Rev. E*, 68, 061902.
- Volkov, I., Banavar, J.R., Hubbell, S.P. & Maritan, A. (2003). Neutral theory and relative species abundance in ecology. *Nature*, 424, 1035–1037.
- Watterson, G.A. (1974). Models for the logarithmic species abundance distribution. *Theor. Popul. Biol.*, 6, 217–250.
- Winemiller, K.O. (1990). Spatial and temporal variation in tropical fish tropic networks. *Ecol. Monogr.*, 60, 331–367.

Editor, Marti Anderson

Manuscript received 8 October 2004

First decision made 16 November 2004

Manuscript accepted 26 January 2005