Journal of Statistical Physics, Vol. 128, Nos. 1/2, July 2007 (© 2007)
DOI: 10.1007/s10955-006-9163-2

Neutral Community Theory: How Stochasticity and Dispersal-Limitation Can Explain Species Coexistence

Rampal S. Etienne[1] and David Alonso[2]

Received January 27, 2006; accepted June 30, 2006

Neutral community theory explains biodiversity, *i.e.* the coexistence of several species, as the result of a stochastic balance between immigration and extinction on a local level, and between speciation and extinction on a regional level. The most popular model, presented by Hubbell in 2001, has seen many analytical developments in recent years, which can be used in model analysis, model testing and model comparison. We review these developments here, and present alternative derivations and shine previously unnoticed lights on them.

KEY WORDS: biodiversity, community, neutral model, Ewens sampling formula.

1. INTRODUCTION

Physics has Pauli's exclusion principle, ecology has Gause's. It states that no two species can coexist indefinitely on the same resource: one will always outcompete the other(s). Gause(39) studied experimentally the coexistence of two closely related protozoa species (*Paramecium caudatum* and *Paramecium aurelia*). He found that both species experience sigmoidal growth when they are alone, but that *P. caudatum* is eliminated when the two species are put together. This has led to the theory that there is a limit to the similarity of coexistent species.(1) Only when they are sufficiently different, they can each occupy a different ecological niche (i.e. a unique way of utilizing available resources), enabling them to coexist. However, in nature, there are diverse communities that happen to accommodate a high

---

[1] Community and Conservation Ecology Group, Centre for Ecological and Evolutionary Studies, University of Groningen, P.O. Box 14, 9750 AA Haren, The Netherlands; e-mail: r.s.etienne@rug.nl.
[2] Ecology and Evolutionary Biology, University of Michigan, 830 North University Avenue, Ann Arbor, MI 48109-1048.

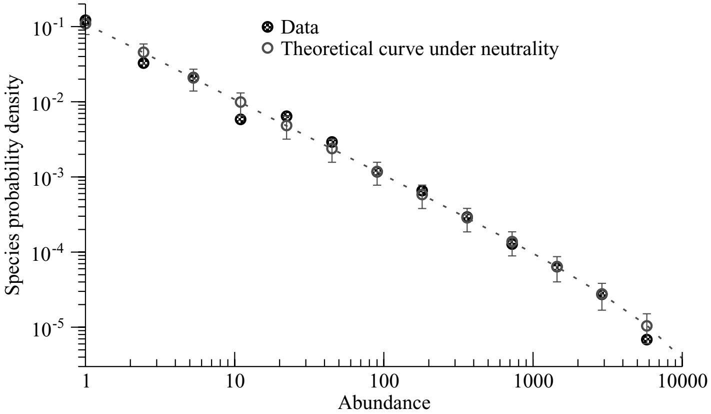

Fig. 1. The distribution of species abundances or abundance curve for a diverse group of marine algae, Mediterranean Diatomea (112,352 individuals, 107 species); data collected by Margalef(65) in the Mediterranean Sea and studied by Pueyo.(77) The species probability density is defined as the number of species in each abundance interval divided by the width of the interval and by the total number of species in the sample. The clear regularity in the species abundance distribution is one of the most striking predictions of neutral theory.

number of coexisting species occupying the same ecological niche. A paradigmatic example is marine plankton (see Fig. 1). There is an amazingly high diversity of microscopic algae in a few liters of sea water. Why should they coexist if they are all very similar in ecological requirements and all share the same homogeneous environment? This paradox became known as the paradox of the plankton.(55) A plausible solution for this paradox and a fierce challenge for traditional niche theory was recently provided by Hubbell(53) who explained coexistence of species, on a regional (continental) scale, as a stochastic equilibrium between origination of species (speciation) and disappearance of species (extinction), and on a local scale as a stochastic equilibrium between immigration and local extinction (due to either mortality or emigration). In this so-called neutral theory, all species at the same trophic level have the same demographic parameters that determine their birth-death dynamics, and all species have the same probability of colonizing empty sites. It is not so much species traits that determine their presence and abundance in a community, but stochastic demography, speciation and dispersal.

This recognition of the importance of chance and dispersal for community structure is not new. Grinnell(46) already argued that accidentals, i.e. species that have been observed only once, are just the result of dispersal. Gleason's(40,41) individualistic concept assumes that species are independent of one another and just happen to be in a locality due to chance and dispersal; this view challenged the then dominant view of a community as a higher level of organization, a superorganism.(16) The individualistic concept was largely ignored until Whittaker(94) started to endorse it again. Some years later, MacArthur(62,63) embraced randomness of species assembly in his broken-stick hypothesis: a stick (the community) is broken randomly into several pieces (niches) with the length of each piece representing the abundance of the species. Many similar hypotheses, the so-called sequential breakage models, were devised,(79,84–86) differing only in the rules governing the breakage process. Although the broken-stick theory was considered a niche assembly theory, it can be argued that it is a neutral theory,(33) because it treats species as equivalent, that is, the traits of the species do not matter in the amount of niche space that it receives in the stick-breaking procedure. However, as other useful models in community ecology,(17,95) broken-stick theory is a set of simple *algorithms* to generate abundance patterns resembling those actually observed in nature. By contrast, and this is an essential difference, neutral theory is based on realistic biological *processes*, such as birth, death, extinction, and dispersal. The first real break-through of neutral ideas in terms of such processes was established by MacArthur and Wilson(64) in their theory of island biogeography, which was a main source of inspiration for Hubbell.(53) This famous theory explains the diversity on islands as a balance between immigration and extinction of species. The main difference is that Hubbell(53) treats the individual as the basic unit, whereas MacArthur and Wilson(64) assigned this role to species. Individual-based neutral models such as Hubbell's were already developed by Caswell(10) who borrowed concepts from population genetics, as did Hubbell(53) (see below). Stochasticity and dispersal were also key elements in so-called lottery and voter models.(9,36) However, none of this work elicited such a strong debate as Hubbell's(53) book, not even Hubbell's own summary of his theory four years earlier.(52) The secret of this "success" probably lies in the fact that in his 2001 book Hubbell explained the theory in mostly non-technical terms, and thus popularized neutral ideas like Hutchinson had done for the ecological niche concept five decades earlier. Or it was just a perfect time for these ideas (see e.g. Ref. 7).

Hubbell's neutral theory is not the only solution to the paradox of the plankton. The ecological niche concept has also evolved considerably since it was coined by Grinnell(45)—yes, the same Grinnell as above—, endorsed by Elton(26) and formally defined by Hutchinson(54) as a hypervolume, a multi-dimensional space of resources (e.g. light, nutrients, structure) leading to standard niche theory that explains species coexistence as a result of resource partitioning along many niche axes (one niche, one species). This partitioning may be much more detailed than originally thought: for instance, plankton partition light intensity(71) and light frequencies/wave lengths.(83) Niche partitioning is not completely unrestricted, however, they are subject to trade-offs in life history traits.(59) Niche differentiation is often spatial: even if there is competitive exclusion locally, with resource heterogeneity there may be coexistence on larger spatial scales.(74,80) Niche differentiation along the time axis,(14) particularly in combination with non-equilibrium conditions,(15) is another important explanation of species coexistence. In these last examples, we see that stochasticity and dispersal have sneaked in in niche theory. Indeed, recent papers claiming to reconcile neutral ideas with niche theory incorporate these ingredients in niche models.(31,43,81)

One of the main merits of Hubbell's theory(4) is that it has introduced Ockham's razor to community ecology: while traditionally ecologists searched for a niche for each species in a community, neutral theory aims at finding the minimal model that can explain observed patterns. In other words, ecologists should not be trying to find differences between species, but determine to what extent differences between species matter. The neutral model can thus act as a null model (see also Ref. 42), to which any more complex model can be compared. To facilitate such comparisons, precise formulation of the predictions of the neutral model is crucial. One of the most successful and mostly studied predictions of neutral theory concerns the shape of the species abundance distribution. This distribution is usually presented in either of two ways: the number of species in different logarithmic abundance classes, or the logarithm of the abundance versus the rank of this abundance in the sample (but see Fig. 1 and Ref. 77 for a somewhat different presentation). With only two parameters the neutral model is able to reproduce the various observed shapes of this distribution for very different taxa. The species abundance distribution has intrigued ecologists for decades (see e.g. Refs. 38 and 76). Why this is so may be partly explained by the fact that these data are relatively easy to collect. But this distribution also lies at the heart of the niche vs. neutral debate: if the abundance of one species is larger than that of another, does this indicate intrinsic differences between the species, or is this just coincidence?

Considerable progress has recently been made on the precise predictions of the neutral model (mostly by physicists working in ecology), again particularly with respect to species abundance distributions, and in this paper we review the approaches that have been taken, but we provide novel derivations and mention previously unnoticed aspects. First we formulate the model. A schematic presentation of the model is given in Fig. 2.

2. THE MODEL

2.1. Regional Dynamics

Let the regional ecological community (metacommunity in Hubbell's terminology) consist of J_{M} individuals competing for the same resource (e.g. nutrients, light, space). It is easiest to envision a forest with sites acting as resource. The J_{M} individual trees are constrained by the so-called zero-sum assumption:(53) all of the available sites are being used all the time. This means that if a site becomes available, a new individual must emerge to colonize it. In Hubbell's model, the

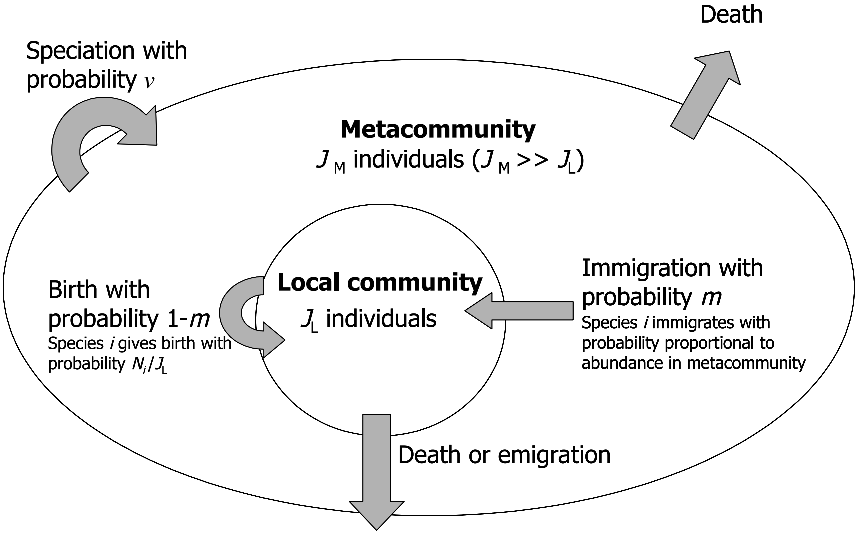

Fig. 2. A schematic representation of the most cited neutral model according to Hubbell.(53) The local community receives immigrants from the metacommunity that compete with local off-spring for available resources. The metacommunity is in a balance between speciation and extinction. Neutrality means that all individuals are equivalent, regardless of the species they belong to. This translates into probabilities of immigration, birth and death proportional to the abundance of species.

number of sites is assumed to be constant, so sites can only become available by the death of any of the individuals, which then must be replaced by a new individual. Such a new individual arises through reproduction; the assumption of neutrality requires that all reproducing individuals have an equal probability that their offspring colonizes the empty site. Note that this allows species to be different: the only requirement is that the colonization probability is equal for all species, but this may be realized by producing lots of offspring with low colonization potential or by producing few offspring with high colonization potential (e.g. if $c = 1 - (1 - p)^s$ where c is the overall colonization probability, p is the colonization probability of a single seed and s is the number of seeds produced per individual; the question remains, however, whether realistic trade-offs can be described in this simple way as seed production and seed survival are actually temporally separated events). For this reason, Hubbell(53) calls the species functionally equivalent; the species are identical as far as their function is concerned. The newly produced individual will evidently be of the same species as the parent, unless speciation occurs, with probability ν. In Hubbell's standard model, speciation is modelled in a very simple manner: a single mutation can cause a species' offspring to be a different species than its parent, and thus this mode of speciation is called the point mutation mode. Other modes of speciation are possible, and actually have been discussed by Hubbell,(52,53) but analytical approaches have not yet been applied to the corresponding neutral models. The neutral model with point mutation is an exact analog of the neutral model of molecular evolution(58) in population genetics, with individuals replacing genes, species replacing alleles and speciation replacing mutation. This opens up a box of mathematical and statistical tools developed in population genetics, for use in community ecology. Conversely, new mathematical results in community ecology will have their application in population genetics.

Hubbell(53) modeled the dynamics of the metacommunity in discrete time. At every time step, some individuals die. There are two extreme cases. The first extreme case is that, at every time step, all individuals die but, before doing so, reproduce (e.g. annual plants leaving seeds for next year) and the offspring compete with equal strengths for the empty site. The second extreme case is that, at every time step, only one individual dies, but reproduces before doing so, and this offspring competes with the offspring of the other individuals for the empty site. These models have exact parallels in population genetics, which has not been made sufficiently explicit in the recent ecological literature on the neutral model. The first extreme is known as the Wright–Fisher model,(37,96) in which there are no overlapping generations (since all individuals die simultaneously). The second extreme is known as the Moran(72,73) model, in which generations are overlapping. The version of Hubbell's model that has been subjected to careful analytical scrutiny, and will therefore be the main model discussed in this paper, is a variant of the Moran model. Where in the Moran model the individual that dies can still contribute to the pool of offspring competing for the site left vacant by the dying individual, in the analytically studied Hubbell model, the dying individual cannot contribute. We will come back to these three versions of the Hubbell model (which we will call the Wright–Fisher, Moran and Hubbell models) when we discuss the genealogical approach to the analysis of the neutral model.

Other neutral models, particularly models that allow metacommunity size to vary or that drop the zero-sum assumption, have also been studied recently.(48,89,91) The model of Volkov *et al.*(89) has the curious property that it drops the zero-sum assumption at the regional level, but makes use of this assumption at the local level. This property, which makes the model hard to interpret, has passed practically unnoticed. In this paper we focus on the standard Hubbell model with zero-sum assumption. Hubbell(53) argues that the zero-sum assumption has strong empirical support. It is also a simple way to incorporate interaction between species, which undoubtedly exist, instead of assuming them to be independent, as other neutral models have done.(22,27,28,48) For large, species-rich communities the zero-sum assumption and the independence assumption practically yield the same model predictions; for species-poor systems they may differ substantially.(3) However, for the equilibrium abundance distribution the models are equivalent as long as one conditions probabilities and expectations on the sample size (Ref. 94; Etienne *et al.* unpubl.).

2.2. Local Dynamics

In the local community there are J_L individuals, with $J_L \ll J_M$. Local dynamics are completely identical to regional dynamics, except that speciation is replaced by immigration from the region, with probability m. Any vacant site can thus be colonized by offspring of a local individual, or by offspring of a regional individual. In the local community, speciation is thus neglected, and therefore speciation plays only an indirect role by having an effect on the relative abundances of immigrating species. The boundaries of the local community are not well defined. In fact, the local community is usually set equal to a sample, as we now explain.

2.3. Sampling Theory

One of the great advantages of the neutral theory is that it is a sampling theory. Predictions are made for a sample of size J from the metacommunity or from the local community, and can thus readily be applied to data, which usually represent only a sample from the total (meta) community. The sampling property of the theory allows us to define the local community as a sample; hence it is usually assumed that $J = J_L$. The probability of immigration, m, thus depends on sample size, which complicates comparison of m-values for samples with different sample sizes. The sampling nature of neutral theory has led to the view of dispersal-limitation as a sampling effect. As we will see below, when a local community is sampled, the information we obtain from the potential regional diversity is filtered by the limited ability of species in the metacommunity to disperse and contribute to the actual local community being sampled. This filtering process can be precisely represented by a sampling distribution, the dispersal-limited multinomial.(30)

3. THREE APPROACHES

3.1. Simulations

Although Hubbell(53) translated some analytical results obtained in population genetics to community ecology for the regional community, he explored the model for the local community by simulations. One of the first contesters of the neutral theory, McGill,(67) also resorted to simulations. This can easily be done, because the model is a simple Markov model (Markov chain in discrete time). Yet, for accurate predictions of the species abundance distribution in or on the way to equilibrium, and hence reliable conclusions about the validity of the neutral model, simulations are not suitable, because the state space is enormously large for even small-sized communities requiring many time steps to reach convergence and many repetitions to obtain accurate estimates of mean and other moments. Therefore, there has been a quest for analytical solutions for the species abundance distribution in the local community, particularly for the equilibrium distribution, and this has resulted in two different, yet related approaches.

3.2. Master Equation Approach

The first approach sought for expressions of the expected number of species with a particular abundance n in a sample of size J, which we will denote by $E\left[S_n|\Theta, J\right]$ where Θ represents the model parameters. For the metacommunity, $E\left[S_n|\Theta, J\right]$ in equilibrium was already known in population genetics,(72,93) which was noted by Hubbell.(53) It is given by

$$E\left[S_n|\theta, J\right] = \frac{\theta}{n}\frac{(J+1-n)_n}{(J+\theta-n)_n} \tag{1}$$

where $(x)_N$ is the Pochhammer symbol defined as

$$(x)_N := \prod_{i=1}^{N}(x+i-1) = \frac{\Gamma(x+N)}{\Gamma(x)} = \sum_{j=1}^{N}\bar{s}(N,j)x^j \tag{2}$$

where $\Gamma(x)$ is the Gamma function and $\bar{s}(j,k)$ is the so-called unsigned Stirling number of the first kind. Furthermore, θ is the fundamental biodiversity parameter,(53) defined by

$$\theta := \frac{\nu}{1-\nu}(J_{\mathrm{M}}-1) \tag{3}$$

This θ is different from the formula, as defined by Hubbell,(53) which is $\theta := 2\nu J_{\mathrm{M}}$. The latter formula for θ is in fact the formula for the Wright–Fisher model, as we will see below.

For the local community, however, no such expressions were available in population genetics. Nevertheless, part of the solution was given in McKane *et al.*(69) and again in McKane *et al.*(70) where the local species abundance distribution was related to a given species abundance distribution of the metacommunity (which determines the species of the immigrants). The only missing step was the use of a metacommunity model to predict the species abundance distribution of the metacommunity, rather than the assumption of this distribution to be given. Volkov *et al.*(89) using the results by McKane *et al.*(69) took this step, but only partly so, because they used only an approximation of the metacommunity species abundance distribution. Vallade and Houchmandzadeh(87) gave the full solution for $E\left[S_n|\theta, m, J\right]$ in the local community, with the exact description of the species abundance distribution in the metacommunity, but their formula was erroneous.

This error was corrected by Etienne and Alonso$^{(30)}$ who, finally, gave the full solution:

$$\begin{aligned} E\left[S_n|\theta, m, J\right] &= \frac{\theta}{(I)_J}\binom{J}{n}\int_0^1 (Ix)_n\,(I\,(1-x))_{J-n}\,\frac{(1-x)^{\theta-1}}{x}dx \\ &= \sum_{j=1}^{J_\mathrm{M}}\binom{J}{n}\sum_{A=1}^{J}\sum_{a=1}^{n}\overline{s}\,(n,a)\,\overline{s}\,(J-n,A-a)\,\frac{I^A}{(I)_J}\frac{1}{\binom{A}{a}}\frac{\binom{j}{a}\binom{J_\mathrm{M}-j}{A-a}}{\binom{J_\mathrm{M}}{A}} \\ &\quad\times\frac{\theta}{j}\frac{(J_\mathrm{M}+1-j)_j}{(J_\mathrm{M}+\theta-j)_j} \end{aligned} \tag{4}$$

where

$$I := \frac{m}{1-m}\,(J_\mathrm{L}-1) \tag{5}$$

is the fundamental dispersal number.$^{(30)}$

McKane *et al.*$^{(69,70)}$ and Vallade and Houchmandzadeh$^{(87)}$ both used a master equation approach to derive their expressions that eventually led to (4). However, their equations were different because they had a different interpretation of speciation. McKane *et al.*$^{(69,70)}$ and Volkov *et al.*$^{(89,91)}$ used the following set of equations for the dynamics of the probability $P_n(t)$ that a species has abundance n in the metacommunity:

$$\frac{dP_0}{dt} = -sP_0 + r_1P_1 \tag{6a}$$

$$\frac{dP_1}{dt} = sP_0 + r_2P_2 - (g_1 + r_1)P_1 \tag{6b}$$

$$\frac{dP_n}{dt} = g_{n-1}P_{n-1} + r_{n+1}P_{n+1} - (r_n + g_n)P_n \quad \text{for } 1 < \text{n} < \text{c} \tag{6c}$$

$$\frac{dP_c}{dt} = g_{c-1}P_{c-1} - r_cP_c \tag{6d}$$

where c is the largest abundance that the species can have; c is usually equal to the total metacommunity size J_M, but it could also be a different ceiling common to all species. The parameters g_n and r_n denote the growth and decline functions for a species at abundance n. The parameter s is the rate of appearance of the species under consideration. If $s > 0$, each species has a non-zero probability of being "re-introduced" after extinction. McKane *et al.*$^{(69,70)}$ and Volkov *et al.*$^{(89,91)}$ regard speciation as a type of immigration process from the species pool with S species where the same species can reappear, and thus s is indeed non-zero in this interpretation.

Vallade and Houchmandzadeh$^{(87)}$ do not assume a species pool and take s to be a real speciation rate. Since it is then highly unlikely that exactly the same species reappears through speciation after going extinct, s should be zero which yields the following description:

$$\frac{dP_0}{dt} = r_1 P_1 \tag{7a}$$

$$\frac{dP_1}{dt} = r_2 P_2 - (g_1 + r_1) P_1 \tag{7b}$$

$$\frac{dP_n}{dt} = g_{n-1} P_{n-1} + r_{n+1} P_{n+1} - (r_n + g_n) P_n \quad \text{for} 1 < n < c \tag{7c}$$

$$\frac{dP_c}{dt} = g_{c-1} P_{c-1} - r_c P_c \tag{7d}$$

When time goes to infinity, every species will go extinct, resulting in $P_0(\infty) = 1$ and $P_n(\infty) = 0$ for $n > 0$. We are however interested in the distribution of species abundances conditional on non-extinction. As new species continuously enter the metacommunity through speciation, there is an equilibrium distribution of these abundances. To obtain this distribution Vallade and Houchmandzadeh(87) elegantly used Laplace transformation. The system (7) can be written in matrix form as

$$\frac{d\vec{P}(t)}{dt} = H\vec{P}(t) \tag{8}$$

where $\overrightarrow{P}(t) = (P_0(t), P_1(t), P_2(t), \ldots, P_{J_M}(t))$ and H is the transition matrix with elements (H_{00} is the element in the top left corner)

$$H_{ij} = \begin{cases} r_j & \text{for} \quad j = i + 1 \\ -(g_i + r_j) & \text{for} \quad j = i \\ g_j & \text{for} \quad j = i - 1 \\ 0 & \text{in all other cases} \end{cases} \tag{9}$$

The Laplace transform L of $\vec{P}(t)$ is defined by

$$L[\vec{P}(t)] := \int_0^\infty e^{-Et} \vec{P}(t)\, dt \tag{10}$$

When we apply Laplace transformation to (8), we obtain (using integration by parts),

$$EL[\vec{P}(t)] - \vec{P}(0) = HL[\vec{P}(t)] \tag{11}$$

where E is the argument of the Laplace transform, and $\vec{P}(0)$ is the initial condition. In this case we have $\vec{P}(0) = (0, 1, 0, \ldots, 0)$, i.e. $P_1(0) = 1$ and $P_n(0) = 1$ for $n \neq 1$). We can easily solve (11) for $L[\vec{P}(t)]$ and then take the inverse Laplace transform L^{-1} to arrive at the solution of (8) for $\vec{P}(t)$:

$$\vec{P}(t) = L^{-1}[G(E)\vec{P}(0)] \tag{12}$$

where we have defined

$$G(E) := (EI - H)^{-1} \tag{13}$$

with I the identity matrix. The expected number of species with abundance n at time t is the sum of all species that originated (through speciation at constant rate ν) at any previous time $\hat{t}$ between 0 and t and are still present with abundance n at time t:

$$\begin{aligned} E[S_n(t)] &= \nu \int_0^t P_n(t - \hat{t})d\hat{t} = \nu \int_0^t P_n(\hat{t})d\hat{t} \\ &= \nu L^{-1}\left[\left. \frac{(EI - H)^{-1}\vec{P}(0)}{E} \right|_n \right] = \nu L^{-1}\left[\frac{G_{n1}(E)}{E}\right] \end{aligned} \tag{14}$$

The second line follows from the first due to a property of the Laplace transform,

$$\int_0^t f(u)du = L^{-1}\left[\frac{L\,[f(t)]}{E}\right] \tag{15}$$

The long time behavior of $E[S_n(t)]$ can be studied by taking the limit $t \to \infty$. We can then use another property of the Laplace transform (the final value theorem),

$$\lim_{t\to\infty} f(t) = C \iff \lim_{E\to 0} EL[f(t)] = C \tag{16}$$

Taking $f(t) = E[S_n(t)]$ and using (14), we obtain

$$\lim_{E\to 0} EL[E[S_n(t)]] = \lim_{E\to 0} EL\left[\nu L^{-1}\left[\frac{G_{n1}(E)}{E}\right]\right] = \lim_{E\to 0} \nu \mathrm{G}_{n1}(E) \tag{17}$$

Vallade and Houchmandzadeh$^{(87)}$ used a recurrence method to expand the $G_{n1}(E)$ as power series in E for $n \geq 1$. Combining (16) and (17), we thus arrive at (for $n \geq 1$),

$$\lim_{t\longrightarrow\infty} E[S_n(t)] = \nu G_{n1,0} \tag{18}$$

where $G_{n1,0}$ is the zeroth order term of G_{n1}. Inserting the expressions obtained with their recurrence method, Vallade and Houchmandzadeh$^{(87)}$ found (1).

There is, however, a more direct way of deriving (4), which we will present here. It is based on Alonso$^{(2)}$ which was inspired by discussions with Alan McKane. Instead of considering the probability, one considers the dynamics of the expected number of species itself, $E\,[S_n|\theta, J_{\mathrm{M}}]$ for which we temporarily use the shorthand notation, S_n:

$$\frac{dS_1}{dt} = s + r_2 S_2 - (r_1 + g_1)S_1 \tag{19a}$$

$$\frac{dS_n}{dt} = g_{n-1}S_{n-1} + r_{n+1}S_{n+1} - (r_n + g_n)S_n \quad \text{for} \quad 1 < n < c \tag{19b}$$

$$\frac{dS_c}{dt} = g_{c-1}S_{c-1} - r_c S_c \tag{19c}$$

In this description there is no S_0 because there are no species with abundance 0. It is easily shown that the steady-state solution requires $g_{n-1}S_{n-1} = r_n S_n$ for all $n > 1$ and $s = r_1 S_1$ for $n = 1$ and is therefore given by

$$S_n = \frac{s}{r_1}\prod_{j=2}^{n}\frac{g_{j-1}}{r_j} \tag{20}$$

with the convention that the product returns 1 if $j > n$. Now for the metacommunity we have the dimensionless rates$^{(87)}$

$$g_n = \frac{(J_\mathrm{M} - n)\,n}{J_\mathrm{M}\,(J_\mathrm{M} - 1)}(1 - \nu) \quad \text{for} \quad n > 0 \tag{21a}$$

$$r_n = \frac{n}{J_\mathrm{M}}\left(\frac{J_\mathrm{M} - n}{J_\mathrm{M} - 1} + \frac{n-1}{J_\mathrm{M} - 1}\nu\right) \tag{21b}$$

$$s = \nu \tag{21c}$$

These rates can also be interpreted as probabilities in a discrete-time model as in Hubbell.$^{(53)}$ When we substitute these in (19), we obtain

$$\begin{aligned}
E\left[S_n|\theta, J_\mathrm{M}\right] = S_n &= \frac{s}{r_1}\prod_{j=2}^{n}\frac{g_{j-1}}{r_j} = \frac{\nu}{\frac{1}{J_\mathrm{M}}}\prod_{j=2}^{n}\frac{\frac{(J_\mathrm{M}-(j-1))(j-1)}{J_\mathrm{M}(J_\mathrm{M}-1)}(1-\nu)}{\frac{j}{J_\mathrm{M}}\left(\frac{J_\mathrm{M}-j}{J_\mathrm{M}-1} + \frac{j-1}{J_\mathrm{M}-1}\nu\right)} \\
&= \nu J_\mathrm{M}\prod_{j=2}^{n}\frac{(J_\mathrm{M} - (j-1))\,(j-1)\,(1-\nu)}{j\,(J_\mathrm{M} - 1 - (j-1) + (j-1)\,\nu)} \\
&= \frac{\nu J_\mathrm{M}}{n}\prod_{j=2}^{n}\frac{J_\mathrm{M} - (j-1)}{\frac{J_\mathrm{M}-1}{1-\nu} - (j-1)} = \frac{\nu J_\mathrm{M}}{n}\prod_{j=2}^{n}\frac{J_\mathrm{M} - (j-1)}{\theta + J_\mathrm{M} - 1 - (j-1)} \\
&= \frac{\nu J_\mathrm{M}}{n}\prod_{j=2}^{n}\frac{J_\mathrm{M} + 1 - j}{J_\mathrm{M} + \theta - j} = \frac{\nu}{n}(J_\mathrm{M} + \theta - 1)\prod_{j=1}^{n}\frac{J_\mathrm{M} + 1 - j}{J_\mathrm{M} + \theta - j} \\
&= \frac{\nu}{n}\left(\frac{J_\mathrm{M} - 1}{1 - \nu}\right)\prod_{j=1}^{n}\frac{J_\mathrm{M} + 1 - j}{J_\mathrm{M} + \theta - j} = \frac{\theta}{n}\prod_{j=1}^{n}\frac{J_\mathrm{M} + 1 - j}{J_\mathrm{M} + \theta - j} \\
&= \frac{\theta}{n}\prod_{j=1}^{n}\frac{J_\mathrm{M} + 1 - n + j - 1}{J_\mathrm{M} + \theta - n + j - 1} = \frac{\theta}{n}\frac{(J_\mathrm{M} + 1 - n)_n}{(J_\mathrm{M} + \theta - n)_n}
\end{aligned} \tag{22}$$

where we have used the identity $\theta + J_{\mathrm{M}} - 1 = \frac{J_{\mathrm{M}}-1}{1-\nu}$. It can be readily shown(30) that this formula is invariant under hypergeometric sampling, i.e. when a sample is taken without replacement, then the formula is still valid, but with J instead of J_{M}. Equation (22) also has an integral form(30):

$$E\left[S_n|\theta, J\right] = \int_0^1 \binom{J}{n} x^n \left(1-x\right)^{J-n} \Omega(x)dx \tag{23}$$

where $\Omega(x)dx$ is the number of species with relative abundance between x and $x + dx$ for a metacommunity of infinite size, which is given by Alonso and McKane(3) and Etienne and Alonso(30)

$$\Omega(x)dx := \theta \frac{(1-x)^{\theta-1}}{x} dx \tag{24}$$

Hypergeometric sampling from such an infinite metacommunity is equivalent to the much simpler binomial sampling.

For the local community we have, in the spirit of Vallade and Houchmandzadeh,(87)

$$g_n = \frac{J_{\mathrm{L}} - n}{J_{\mathrm{L}}} \left(\frac{n}{J_{\mathrm{L}} - 1}(1-m) + mx\right) \quad \text{for} \quad n > 0 \tag{25a}$$

$$r_n = \frac{n}{J_{\mathrm{L}}} \left(\frac{J_{\mathrm{L}} - n}{J_{\mathrm{L}} - 1}(1-m) + m\left(1-x\right)\right) \tag{25b}$$

$$s = mp(x, I, J_{\mathrm{L}}) \tag{25c}$$

where x is the relative abundance of the species in the metacommunity (which has infinite size in this derivation) and $p(x, I, J_{\mathrm{L}})$ is the probability that the immigrating individual has no conspecifics in the local community if it has relative abundance x in the metacommunity. We get

$$\begin{aligned}
E\left[S_n|m, x, J_{\mathrm{L}}\right] = S_n(x) &= \frac{s}{r_1} \prod_{j=2}^{n} \frac{g_{j-1}}{r_j} \\
&= \frac{mp(x, I, J_{\mathrm{L}})}{\frac{1}{J_{\mathrm{L}}}\left(1-mx\right)} \prod_{j=2}^{n} \frac{\frac{J_{\mathrm{L}}-(j-1)}{J_{\mathrm{L}}}\left(\frac{j-1}{J_{\mathrm{L}}-1}(1-m) + mx\right)}{\frac{j}{J_{\mathrm{L}}}\left(\frac{J_{\mathrm{L}}-j}{J_{\mathrm{L}}-1}(1-m) + m\left(1-x\right)\right)} \\
&= \frac{p(x, I, J_{\mathrm{L}})}{x} \prod_{j=1}^{n} \frac{J_{\mathrm{L}} - (j-1)}{j} \frac{j - 1 + Ix}{J_{\mathrm{L}} - j + I\left(1-x\right)} \\
&= \frac{p(x, I, J_{\mathrm{L}})}{x} \binom{J_{\mathrm{L}}}{n} \left(Ix\right)_n \prod_{j=1}^{n} \left(\frac{1}{I\left(1-x\right) + J_{\mathrm{L}} - j}\right)
\end{aligned}$$

$$\begin{aligned}
&\times \frac{\prod_{j=1}^{J_{\mathrm{L}}}\left(I\left(1-x\right)+J_{\mathrm{L}}-j\right)}{\prod_{j=1}^{J_{\mathrm{L}}}\left(I\left(1-x\right)+J_{\mathrm{L}}-j\right)}\\
&= \frac{p(x,I,J_{\mathrm{L}})}{x}\binom{J_{\mathrm{L}}}{n}\left(Ix\right)_{n}\frac{\prod_{j=n+1}^{J_{\mathrm{L}}}\left(I\left(1-x\right)+J_{\mathrm{L}}-j\right)}{\prod_{j=1}^{J_{\mathrm{L}}}\left(I\left(1-x\right)+J_{\mathrm{L}}-j\right)}\\
&= \frac{p(x,I,J_{\mathrm{L}})}{x}\binom{J_{\mathrm{L}}}{n}\left(Ix\right)_{n}\frac{\prod_{j=1}^{J_{\mathrm{L}}-n}\left(I\left(1-x\right)+j-1\right)}{\prod_{j=1}^{J_{\mathrm{L}}}\left(I\left(1-x\right)+j-1\right)}\\
&= \frac{p(x,I,J_{\mathrm{L}})}{x}\binom{J_{\mathrm{L}}}{n}\frac{\left(Ix\right)_{n}\left(I\left(1-x\right)\right)_{J_{\mathrm{L}}-n}}{\left(I\left(1-x\right)\right)_{J_{\mathrm{L}}}} \qquad (26)
\end{aligned}$$

This gives the abundance distribution assuming the relative abundance of all species is x. To account for the abundance distribution of the metacommunity, we must take the appropriate average over x:

$$\begin{aligned}
E\left[S_n|m,\theta,J_{\mathrm{L}}\right] &= \int_0^1 E\left[S_n|m,x,J_{\mathrm{L}}\right]x\Omega\left(x\right)dx\\
&= \int_0^1 \frac{p(x,I,J_{\mathrm{L}})}{x}\binom{J_{\mathrm{L}}}{n}\frac{\left(Ix\right)_{n}\left(I\left(1-x\right)\right)_{J_{\mathrm{L}}-n}}{\left(I\left(1-x\right)\right)_{J_{\mathrm{L}}}}x\theta\frac{\left(1-x\right)^{\theta-1}}{x}dx\\
&= \int_0^1 p(x,I,J_{\mathrm{L}})\binom{J_{\mathrm{L}}}{n}\frac{\left(Ix\right)_{n}\left(I\left(1-x\right)\right)_{J_{\mathrm{L}}-n}}{\left(I\left(1-x\right)\right)_{J_{\mathrm{L}}}}\theta\frac{\left(1-x\right)^{\theta-1}}{x}dx
\end{aligned} \qquad (27)$$

where $x\Omega\left(x\right)dx$ is the probability of an individual randomly taken from the metacommunity (which applies also to the immigrants) having abundance x. Summing this expression over $n=1,\ldots,J_{\mathrm{L}}$ and using Vandermonde's formula,(8)

$$\left(x+y\right)_N = \sum_{n=0}^{N}\binom{N}{n}\left(x\right)_n\left(y\right)_{N-n} \qquad (28)$$

we obtain

$$E\left[S|m,\theta,J_{\mathrm{L}}\right]=\int_{0}^{1}p(x,I,J_{\mathrm{L}})\frac{(I)_{J_{\mathrm{L}}}}{(I\,(1-x))_{J_{\mathrm{L}}}}\theta\frac{(1-x)^{\theta-1}}{x}dx$$

$$-\int_{0}^{1}p(x,I,J_{\mathrm{L}})\theta\frac{(1-x)^{\theta-1}}{x}dx=\int_{0}^{1}p(x,I,J_{\mathrm{L}})$$

$$\times\frac{(I)_{J_{\mathrm{L}}}}{(I\,(1-x))_{J_{\mathrm{L}}}}\Omega(x)\,dx-\int_{0}^{1}p(x,I,J_{\mathrm{L}})\Omega(x)\,dx \qquad (29)$$

Recalling the interpretations of $p(x, I, J_{\mathrm{L}})$ and $\Omega(x)$ above, the second term of this expression represents the expected number of species in the metacommunity that are not in the local community. Therefore the first term must represent the expected total number of species in the metacommunity. For an infinite metacommunity this is given by

$$\int_{0}^{1}\Omega(x)dx \qquad (30)$$

and we must therefore conclude that

$$p(x,I,J_{\mathrm{L}})=\frac{(I\,(1-x))_{J_{\mathrm{L}}}}{(I)_{J_{\mathrm{L}}}} \qquad (31)$$

and hence (27) becomes

$$E\left[S_n|m,\theta,J_{\mathrm{L}}\right]=\theta\binom{J_{\mathrm{L}}}{n}\int_{0}^{1}\frac{(Ix)_n\,(I\,(1-x))_{J_{\mathrm{L}}-n}}{(I)_{J_{\mathrm{L}}}}\frac{(1-x)^{\theta-1}}{x}dx \qquad (32)$$

Again, taking a hypergeometric sample from this distribution leaves the distribution invariant.(30)

We can also use the formalism provided in Etienne and Alonso(30) for the local community. This formalism states that the local community is not a random (hypergeometric or binomial), but a dispersal-limited sample from the metacommunity. For a random sample we have

$$E\left[S_n|\theta,J\right]=\sum_{j=n}^{J_{\mathrm{M}}}P_{\mathrm{hyp}}[n|j,J_{\mathrm{M}},J]E[S_j|\theta,J_{\mathrm{M}}] \qquad (33)$$

where

$$P_{\mathrm{hyp}}[n|j,J_{\mathrm{M}},J]:=\frac{\binom{j}{n}\binom{J_{\mathrm{M}}-j}{J-n}}{\binom{J_{\mathrm{M}}}{J}} \qquad (34)$$

is the hypergeometric distribution and $E[S_j|\theta, J_M]$ is given by (1). For a dispersal-limited sample we have a very similar expression:

$$E\left[S_n|\theta, m, J_M, J\right] = \sum_{j=1}^{J_M} P_{\mathrm{hyp}}^{\mathrm{DL}}\left[n|m, j, J_M, J\right] E[S_j|\theta, J_M] \tag{35}$$

where

$$\begin{aligned} P_{\mathrm{hyp}}^{\mathrm{DL}}\left[n|m, j, J_M, J\right] = \binom{J}{n} \sum_{A=1}^{J} \sum_{a=1}^{n} \overline{s}\,(n, a)\, \overline{s}\,(J-n, A-a)\, \frac{I^A}{(I)_J} \frac{1}{\binom{A}{a}} \\ \times P_{\mathrm{hyp}}[a|j, J_M, A] \end{aligned} \tag{36}$$

is the dispersal-limited hypergeometric distribution. One can show that (32), for a subsample of size J, and (35) are identical.

By writing the effect of dispersal limitation in the same way as the effect of random sampling, Etienne and Alonso(30) showed that dispersal limitation can be regarded as a sampling effect. Dispersal limitation is as ubiquitous as sampling effects, and likewise blurs our image of other possible processes that shape communities. By using dispersal-limited sampling, we are able to uncover the signature of such other underlying processes on the species abundance distribution. If only neutral process are at play, the structure we observe is a direct consequence of the dispersal-limited sampling of a neutral community. According to niche theory, there should still be some further structure left after accounting for all sampling effects.

The master equation approach has an important advantage that has hardly been explored. Vallade and Houchmandzadeh(87) provide a solution for the non-equilibrium case, that is for the time dependence of $P_n(t)$, the probability of a species having abundance n at time t, as we noted in (12). For our master equation in terms of S_n we can do something similar for the metacommunity case. The set of equations (19) can be put in matrix form,

$$\frac{d\overrightarrow{S}}{dt} = H\overrightarrow{S} + \overrightarrow{h} \tag{37}$$

where $\overrightarrow{S}$ is the vector $\overrightarrow{S} = (S_1, S_2, \ldots, S_{J_M})$, H is the transition matrix (9)—with element H_{11} in the top left corner—and $\overrightarrow{h}$ is a vector of length J_M with $h_1 = s$ and $h_j = 0$ for $j > 1$. This matrix equation is readily solved:

$$\overrightarrow{S} = e^{tH}(\overrightarrow{S}_0 + H^{-1}\overrightarrow{h}) - H^{-1}\overrightarrow{h} \tag{38}$$

where $\overrightarrow{S}_0$ is the initial abundance distribution.

3.3. Genealogical Approach

The second approach focused on the probability of observing a particular data set $\overrightarrow{D} = (n_1, n_2, \ldots, n_S)$ where n_i is the abundance of species i. For the metacommunity case, Hubbell(53) borrowed a result long known in population genetics as the Ewens sampling formula(35,57)

$$P[\overrightarrow{D}|\theta, J] = \frac{J!}{\prod_{i=1}^{S} n_i \prod_{j=1}^{J} \Phi_j!} \frac{\theta^S}{(\theta)_J} \tag{39}$$

where Φ_j is the number of species with abundance j in the data set. For the local community no such formula was available. Etienne and Olff(31,32) provided the first formula for this probability, and this was greatly simplified in Etienne.(29) We write it in a slightly different form here:

$$P[\overrightarrow{D}|\theta, m, J] = \frac{J!}{\prod_{i=1}^{S} n_i \prod_{j=1}^{J} \Phi_j!} \frac{\theta^S}{(I)_J} \sum_{a_1=1}^{n_1} \cdots \sum_{a_S=1}^{n_S} \frac{\overline{s}(n_i, a_i)\overline{s}(a_i, 1)}{\overline{s}(n_i, 1)} \frac{I^A}{(\theta)_A} \tag{40}$$

where

$$A = \sum_{k=1}^{S} a_k \tag{41}$$

Etienne and Alonso(30) showed that there is also an integral form of (40):

$$P[\overrightarrow{D}|\theta, m, J] = \frac{J!}{\prod_{i=1}^{S} n_i! \prod_{j=1}^{J} \Phi_j!} \frac{\theta^S}{(I)_J} \int_0^1 \cdots \int_0^1 \prod_{i=1}^{S} \left((I_i x_i)_{n_i} \frac{(1 - x_i)^{\theta-1}}{x_i} \right) \times dx_1, \ldots, dx_S \tag{42}$$

where

$$I_i = I \prod_{k=1}^{i-1} (1 - x_k) \tag{43}$$

Equation (42) provides a way to avoid Stirling numbers in computing the multivariate probability, e.g. by Monte Carlo integration. This will, however, be very computationally intensive for a large number of species S.

While the master equation approach is forwards in time, the genealogical approach follows the current, equilibrium, (meta)community backwards in time. The idea is described in detail in Etienne and Olff(31,32) and Etienne.(29) Here we illustrate the approach with the simplest possible data set $\overrightarrow{D}$, namely one consisting of two individuals. There are two possibilities: either the two individuals are of the same species, or they are of different species. In population genetics the probability that there are of the same species is indicated by F_2. We derive F_2 for the three models (Wright–Fisher, Moran, Hubbell) mentioned above. For completeness, we note that the original formulations of the Wright–Fisher and Moran models did not contain mutation (speciation). The versions of these models that we discuss here do.

In the Wright–Fisher model, all individuals die and leave off-spring which have equal chances of being in the next generation (so generations do not overlap). When two individuals are sampled at random, they can only be of the same species if no speciation occurred in the previous time step. This has probability $(1-\nu)^2$ where ν is the speciation probability per individual per time step. This probability is equal to the speciation probability per generation, u, in this model, as generations do not overlap. The individuals are conspecific, either if they have the same parent, which has probability $\frac{1}{J_\mathrm{M}}$, or if their parents were different individuals of the same species, which has probability $(1-\frac{1}{J_\mathrm{M}})F_2(t-1)$. Hence, we have the recursion

$$F_2(t)) = (1-\nu)^2\left(\frac{1}{J_\mathrm{M}} + \left(1-\frac{1}{J_\mathrm{M}}\right)F_2(t-1)\right) \tag{44}$$

In equilibrium $F_2(t) = F_2(t-1) =: F_2$, so

$$F_2 = \frac{1}{1+\frac{2\nu J_\mathrm{M}-\nu^2 J_\mathrm{M}}{(1-\nu)^2}} = \frac{1}{1+\frac{2uJ_\mathrm{M}-u^2 J_\mathrm{M}}{(1-u)^2}} \approx \frac{1}{1+2J_\mathrm{M}u} \tag{45}$$

where the approximation uses the fact that u is very small.

In the Moran model, only one individual dies and is replaced by offspring of itself or any of the other individuals with equal probability (so generations overlap). When two individuals are sampled at random, there is a $\frac{2}{J_\mathrm{M}}$ probability that either of these two is the new individual and there is a $\frac{1}{J_\mathrm{M}}$ probability that they then have the same parent. As speciation can only occur in offspring, they are then identical with probability $1-v$ where v is the speciation rate per time step. Because the average generation time of an individual is J_M, the per generation speciation rate is $u = \frac{v}{J_\mathrm{M}}$ and hence $v = J_\mathrm{M}u$. If neither of the sampled individuals is new, then they are identical if their parents are. We thus have the recursion

$$F_2 = \frac{2}{J_\mathrm{M}}(1-v)\frac{1}{J_\mathrm{M}} + \frac{2}{J_\mathrm{M}}(1-v)\frac{J_\mathrm{M}-1}{J_\mathrm{M}}F_2 + \left(1-\frac{2}{J_\mathrm{M}}\right)F_2 \tag{46}$$

from which

$$F_2 = \frac{1}{1+\frac{vJ_\mathrm{M}}{1-v}} = \frac{1}{1+\frac{J_\mathrm{M}^2 u}{1-v}} \approx \frac{1}{1+J_\mathrm{M}^2 u} \tag{47}$$

Hubbell's model is the same as Moran's, but the individual that dies does not contribute to the offspring, so we have $v = (J_\mathrm{M}-1)\,u$ and hence the recursion

$$F_2 = \frac{2}{J_\mathrm{M}}(1-v)\frac{1}{J_\mathrm{M}-1} + \frac{2}{J_\mathrm{M}}(1-v)\frac{J_\mathrm{M}-2}{J_\mathrm{M}-1}F_2 + \left(1-\frac{2}{J_\mathrm{M}}\right)F_2 \tag{48}$$

from which

$$F_2 = \frac{1}{1+\frac{v(J_M-1)}{1-v}} = \frac{1}{1+\frac{J_M(J_M-1)u}{1-v}} \approx \frac{1}{1+J_M(J_M-1)u} \tag{49}$$

For the three models described above F_2 is thus actually given (to a very good approximation) by the same formula

$$F_2 \approx \frac{1}{1+\theta} \tag{50}$$

where

$$\theta := 2J_e u \tag{51}$$

and J_e is the effective metacommunity size. In the three models we have

$$J_e = J_M, \text{ so } \theta = 2J_M u \text{ (Wright–Fisher)} \tag{52a}$$

$$J_e = \frac{J_M^2}{2}, \text{ so } \theta = J_M^2 u \; (Moran) \tag{52b}$$

$$J_e = \frac{J_M(J_M-1)}{2}, \text{ so } \theta = J_M(J_M-1)u \text{ (Hubbell)} \tag{52c}$$

which is in agreement with the results obtained in population genetics for the first two models.(92) Hubbell(53) used the Wright–Fisher version of the model to define θ, whereas his model description was interpreted as the modified Moran model, which we have referred to as the Hubbell model.

The form of F_2 given in (50) can be generalized to the full multivariate probability distribution (39). This means that all three models lead to the same sampling formula (39), provided that the speciation rate u is very small. Yet, because the effective metacommunity sizes in the three models depend differently on the real metacommunity size, the models do differ in their dynamics. Ecological drift (the analog of genetic drift, stochastic change in abundances) is more effective in the Wright–Fisher model than in the other two models, because the effective metacommunity size is smaller.

For the local community a similar reasoning can be followed. However, instead of speciation we have immigration and immigrants are not necessarily of a new species. Therefore, we find the same Ewens sampling formula, but this now represents the ancestry abundance distribution rather than the species abundance distribution and it has parameter I. The ancestry abundance distribution gives the distribution of the number of individuals that have the same immigrating ancestor, or in other words, the distribution of the number of descendants of immigrating ancestors. Because these ancestors have immigrated from the metacommunity, they form a random sample from this metacommunity. Thus the species abundance distribution of the local community is a kind of convolution of two Ewens distributions, one for the ancestors of the local community and the other for the species in the metacommunity. The result is given in (40).

Because the genealogical approach yields the full multivariate probability distribution, any statistic of the species abundances should be derivable from it. In particular, the expected number of species with abundance n can be found from the probability distributions (39) and (40). See Etienne and Alonso(30) for details. The multivariate probability distribution is useful for parameter estimation and model comparison as it represents the likelihood of the parameters given the full data. An example is given in Fig. 3. This figure also shows that sometimes dual likelihood optima are present.(34)

The genealogical approach cannot be used to determine the time evolution of the system. For this we need to write down the full transition matrix relating all possible states of the system to one another. As these become very large for even small (meta) community sizes, they are difficult if not impossible to study.

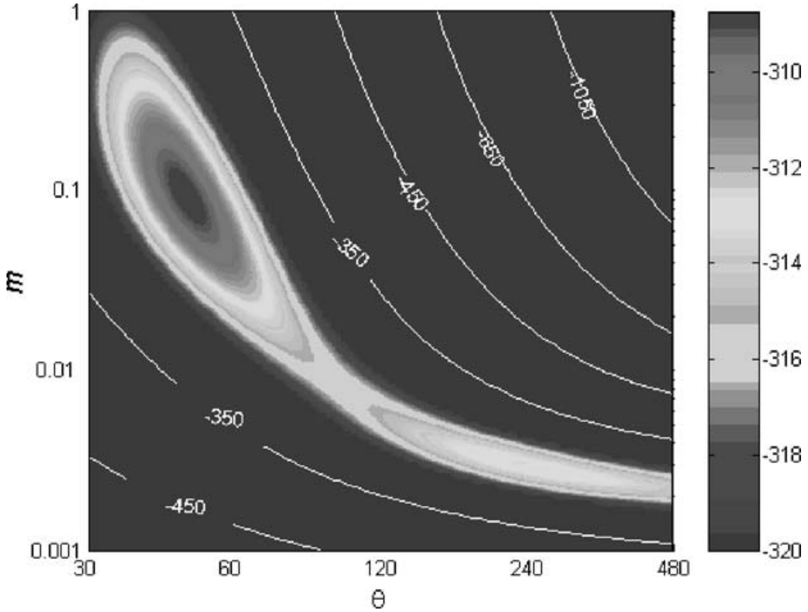

Fig. 3. Loglikelihood surface of the (θ, m)-parameter combination for the tree community on Barro Colorado Island, Panama (21457 individuals, 225 species, 1995 census, see Ref. 18). The colorbar shows loglikelihoods higher than −320, so the dark blue area represents loglikelihoods lower than −320, which are indicated by contour lines. Note the presence of a global and a local maximum.

4. DISCUSSION

Statistical physics aims at predicting the collective behavior of a system consisting of a large number of entities. Although these entities can have complex individual behavior by themselves, usually simple assumptions regarding the microscopic dynamics of these individual entities and their interactions are enough to recover most macroscopic properties of the system. The central question is precisely to figure out which assumptions are essential at the individual level to explain most macroscopic properties observed at the level of the whole system. This approach has been very successful in physics. During the last decade it has been applied to an increasing number of biological systems,(88) and, in particular, to community ecology.(53,61) Neutral community theory is a clear example of this "mechanico-statistical" approach. From extraordinarily simple assumptions about the birth-death dynamics of individuals a wide range of implications are derived for different properties of ecological communities from, for instance, the shape of phylogenetic trees to the geographic distribution of species in space and time.(53)

We have reviewed recent developments in neutral community theory. We have presented a more direct derivation of one of its key results, the expected number of species with a particular abundance. We have also provided a broader context for a second key result, the probability of observing a particular species abundance distribution. Thus, regarding the species abundance distribution in equilibrium, the main results have been found, and the neutral model seems to provide a good explanation of observed species abundance distributions, although there may be others that are equally good.(11,33,47) Particularly, recently a controversy has arisen over whether density-dependence can explain observed species abundance distributions as well as dispersal limitation can. Volkov *et al.*(91) claim this to be the case using a particular model of density-dependence, but Chave *et al.*(13) argue that this claim is not sufficiently substantiated, because Volkov *et al.*(91) do not provide the multivariate probability distribution. We add that even if the claim can be substantiated for the particular model studied by Volkov *et al.*,(91) it will not be true in general for more realistic models of density-dependence.

Yet, there are many more implications of neutral theory that do not seem to match observations so well. For example, it has been suggested that ecological drift cannot explain the observed rapid changes in abundance,(97) and that it yields species that are too old to be realistic.(75) However, this may only be true for the current version of the neutral model. Extended versions may be able to solve these problems. By adding dispersal as a key element to previous neutral theory in ecology,(10,93) current neutral biodiversity theory(53) has become essentially a spatial theory, yet space is still treated implicitly. Spatial ecological models are difficult to study analytically(6,82); results are usually found by simulation.(24,25,66) However, some analytical results have already been found in relation to how diversity changes under neutrality along the spatial dimension (i.e. beta diversity, Refs. 12, 99), spatial scaling(5) and clustering.(49,50) Nevertheless, this field is still largely untouched. Particularly, similarity between different communities belonging the same metacommunity has been claimed as yielding a new test for neutral theory,(23) but solid analytical tools are lacking. In sum, along with further development of the theory, we need to devise stronger tests(68) that can detect a non-neutral signal in the neutral noise.

To advance the field we believe that the parallel to population genetics and the neutral theory of molecular evolution(58) should be explored further, as many mathematical results have already been obtained in this field.(11,51) For example, spatial coalescent theory may help in incorporating phylogenetics in community ecology. The application of tools, concepts, and ideas from statistical physics to ecology also merits further study. Energy is a key concept both in physics and ecology. Interesting attempts have been made to introduce energetic constraints into the stochastic framework used to develop current neutral theory.(90) The ensemble concept is also starting to find its way to the ecological literature (Ref. 92, online material). And a general framework for the Ewens sampling formula as well as Maxwell–Boltzmann, Bose–Einstein and Fermi–Dirac statistics is provided by Costantini and Garibaldi.(19) In evolutionary biology, mathematical analogies with statistical physics has also opened the door for the powerful machinery of statistical fields to yield new insights in nearly-neutral evolution.(78)

Neutral theory is an ideal theory(4): neutral communities do not exist, in the same way that ideal gases do not exist. Nevertheless, neutral theory is and will remain very helpful in identifying non-neutral (niche) factors, and has put stochasticity, dispersal and speciation high on the agenda of community ecology. We hope that our paper will function as a basis and a stimulus for physicists to advance this elegant null model of biodiversity.

ACKNOWLEDGMENTS

We thank three anonymous reviewers for their helpful comments. We thank Alan McKane for inspiring discussions.

REFERENCES

1. P. A. Abrams, The theory of limiting similarity. *Ann. Rev. Ecol. Syst.* **14**:359–376 (1983).
2. D. Alonso, *The Stochastic Nature of Ecological Interactions. Communities, Metapopulations, and Epidemics.* Polytechnical University of Catalonia, Spain. PhD. thesis (2004).
3. D. Alonso and A. J. McKane, Sampling Hubbell's neutral theory of biodiversity. *Ecol. Lett.* **7**:901–910 (2004).
4. D. Alonso, R. S. Etienne and A. J. McKane, The merits of neutral theory. *Trends in Ecol. Evol.* **21**:451–547 (2006).
5. J. R. Banavar, J. L. Green, J. Harte and A. Maritan, Finite size scaling in ecology. *Phys. Rev. Lett.* **83**:4212 (1999).

6. J. Bascompte and R. V. Solé, eds., *Modeling Spatiotemporal Dynamics in Ecology*, Berlin, Germany: Springer and Landes Bioscience (1997).
7. G. Bell, Neutral macroecology. *Science* **293**:2413–2418 (2001).
8. G. Boros and V. Moll, *Irresistible Integrals: Symbolics, Analysis and Experiments in the Evaluation of Integrals*. Cambridge, U. K.: Cambridge University Press (2004).
9. M. Bramson, J. T. Cox and R. Durrett, A spatial model for the abundance of species. *Ann. Probability* **26**:658–709 (1997).
10. H. Caswell, Community structure: A neutral model analysis. *Ecol. Monogr.* **46**:327–354 (1976).
11. J. Chave, Neutral theory and community ecology. *Ecol. Lett.* **7**:241–253 (2004).
12. J. Chave and E. G. Leigh, A spatially explicit neutral model of β-diversity in tropical forests. *Theor. Popul. Biol.* **62**:153–168 (2002).
13. J. Chave, D. Alonso and R. S. Etienne, Comparing models of species abundance. *Nature* **441**:E1 (2006).
14. P. L. Chesson, Mechanisms of maintenance of species diversity. *Annu. Rev. Ecol. Systematics* **31**:343–366 (2000).
15. P. L. Chesson and R. R. Warner, Environmental variability promotes coexistence in lottery competitive systems. *Am. Nat.* **125**:923–943 (1981).
16. F. E. Clements, *Plant Succession: An Analysis of the Development of Vegetation*. Washington, DC: Carnegie Institute of Washington Publication, p. 242, (1916).
17. J. E. Cohen and C. M. Newman, A stochastic theory of community food webs: I. Models and aggregated data. *Proc. R. Soc. Lond. B* **224**:421–448 (1985).
18. R. Condit, N. Pitman, E. G. Leigh, J. Chave, J. Terborgh, R. B. Foster, P. Nuñez, S. Aguilar, R. Valencia, G. Villa, H. C. Muller-Landau, E. Losos and S. P. Hubbell, Beta-diversity in tropical forest trees. *Science* **295**:666–669 (2002).
19. D. Costantini and U. Garibaldi, The Ehrenfest fleas: from model to theory. *Synthese* **139**:107–142 (2004).
20. A. K. Dewdney, A general theory of the sampling process with applications to the “veil line.” *Theor. Popul. Biol.* **54**:294–302 (1998).
21. A. K. Dewdney, A dynamical model of communities and a new species-abundance distribution. *Biol. Bull.* **35**:152–165 (2000).
22. O. H. Diserud and S. Engen, A general and dynamic species abundance model, embracing the lognormal and the gamma models. *Am. Nat.* **155**:497–511 (2000).
23. M. Dornelas, S. R. Connolly and T. P. Hughes, Coral reef diversity refutes the neutral theory of biodiversity. *Nature* **440**:80–82 (2006).
24. R. Durrett, Stochastic spatial models. *SIAM Rev.* **41**:677–718 (1999).
25. R. Durrett and S. Levin, Spatial models for species-area curves. *J. Theor. Biol.* **179**:119–127 (1996).
26. C. Elton, *Animal Ecology*. London, U. K.: Sidgewick & Jackson (1927).
27. S. Engen and R. Lande, Population dynamic models generating the lognormal species abundance distribution. *Math. Biosci.* **132**:169–183 (1996).
28. S. Engen and R. Lande, Population dynamic models generating the species abundance distributions of the Gamma type. *J. Theor. Biol.* **178**:325–331 (1996).
29. R. S. Etienne, A new sampling formula for neutral biodiversity. *Ecol. Lett.* **8**:253–260 (2005).
30. R. S. Etienne and D. Alonso, A dispersal-limited sampling theory for species and alleles. *Ecol. Lett.* **8**:1147–1156 (2005). Erratum in *Ecol. Lett.* **9**:500.
31. R. S. Etienne and H. Olff, How dispersal limitation shapes species—body size distributions in local communities. *Am. Nat.* **163**:69–83 (2004).
32. R. S. Etienne and H. Olff, A novel genealogical approach to neutral biodiversity theory. *Ecol. Lett.* **7**:170–175 (2004).

33. R. S. Etienne and H. Olff, Confronting different model of community structure to species-abundance data: a Bayesian model comparison. *Ecol. Lett.* **8**:493–504 (2005).
34. R. S. Etienne, A. M. Latimer, J. A. Silander and R. M. Cowling, Comment on "Neutral ecological theory reveals isolation and rapid speciation in a biodiversity hot spot." *Science* **311**:610b (2006).
35. W. J. Ewens, The sampling theory of selectively neutral alleles. *Theor. Popul. Biol.* **3**:87–112 (1972).
36. T. Fagerström, Lotteries in communities of sessile organisms. *Trends in Ecol. Evol.* **3**:303–306 (1988).
37. R. A. Fisher, *The Genetical Theory of Natural Selection*. Oxford, U. K.: Clarendon Press (1930).
38. R. A. Fisher, A. S. Corbet and C. B. Williams, 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 (1943).
39. G. F. Gause, *The Struggle for Existence*. Baltimore, MD: Williams and Wilkins (1934).
40. H. A. Gleason, The individualistic concept of the plant association. *Bull. Torrey Botanical Club* **53**:7–26 (1926).
41. H. A. Gleason, The individualistic concept of the plant association. *Am. Midland Nat.* **21**:92–110 (1939).
42. N. J. Gotelli, Null model analysis of species co-occurrence patterns. *Ecology* **81**:2606–2621 (2000).
43. D. Gravel, C. D. Canham, M. Beaudet and C. Messier, Reconciling niche and neutrality: the continuum hypothesis. *Ecol. Lett.* **9**:399–409 (2006).
44. R. C. Griffiths and S. Lessard, Ewens' sampling formula and related formulae: Combinatorial proofs, extensions to variable population size and applications to ages of alleles. *Theor. Popul. Biol.* **68**:167–177 (2005).
45. J. Grinnell, The niche-relationships of the California Thrasher. *Auk* **34**:427–433 (1917).
46. J. Grinnell, On the role of the accidental. *Auk* **39**:373–380 (1922).
47. J. Harte, Tail of death and resurrection. *Nature* **424**:1006–1007 (2003).
48. F. L. He, Deriving a neutral model of species abundance from fundamental mechanisms of population dynamics. *Funct. Ecol.* **19**:187–193 (2005).
49. F. He and S. P. Hubbell, Percolation theory for the distribution and abundance of species. *Phys. Rev. Lett.* **91**:198103 (2003).
50. B. Houchmandzadeh and M. Vallade, Clustering in neutral ecology. *Phys. Rev. E* **68**:061912 (2003).
51. X.-S. Hu, F. He, and S. P. Hubbell, Neutral theory in macroecology and population genetics. *Oikos* **113**:548–556 (2006).
52. S. P. Hubbell, A unified theory of biogeography and relative species abundance and its application to tropical rain forests and coral reefs. *Coral Reefs* **16**:S9–S21 (1997).
53. S. P. Hubbell, *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton, NJ: Princeton University Press (2001).
54. G. E. Hutchinson, Concluding remarks. *Cold Spring Harbor Symp. Quant. Biol.* **22**:415–427 (1957).
55. G. E. Hutchinson, The paradox of the plankton. *Am. Nat.* **95**:137–145 (1961).
56. N. L. Johnson, S. Kotz and N. Balakrishnan, *Discrete Multivariate Distributions*. New York, NY: Wiley (1997).
57. S. Karlin and J. McGregor, Addendum to a paper of W. Ewens. *Theor. Popul. Biol.* **3**:113–116 (1972).
58. M. Kimura, *The Neutral Theory of Molecular Evolution*. Cambridge, U. K.: Cambridge University Press (1983).
59. J. M. Kneitel and J. M. Chase, Trade-offs in community ecology: linking spatial scales and species coexistence. *Ecol. Lett.* **7**:69–80 (2004).
60. R. Lande, S. Engen and B.-E. Saether, *Stochastic Population Dynamics in Ecology and Conservation*. Oxford Series in Ecology and Evolution. Oxford, U. K.: Oxford University Press (2003).

61. S. A. Levin, Complex adaptive systems: exploring the known, the unknown, and the unknowable. *Bull. Am. Math. Soc.* **40**:3–19 (2003).
62. R. H. MacArthur, On the relative abundance of bird species. *Proc. Natl. Acad. Sci., USA* **43**:293–295 (1957).
63. R. H. MacArthur, On the relative abundance of species. *Am. Nat.* **94**:25–36 (1960).
64. R. H. MacArthur and E. O. Wilson, *Island Biogeography*. Princeton, NJ: Princeton University Press (1967).
65. R. Margalef, Through the looking glass: how marine phytoplankton appears through the microscope when graded by size and taxonomically sorted. *Scientia Marina* **58**:87–101 (1994).
66. P. A. Marquet, J. E. Keymer and H. Cofré, Breaking the stick in space: of niche models, metacommunities and patters of relative abundance of species. in T. M. Blackburn & K. J. Gaston, eds. *Macroecology. Concepts and consequences*. Oxford, U. K.: Blackwell, pp. 64–84 (2003).
67. B. J. McGill, A test of the unified neutral theory of biodiversity. *Nature* **422**:881–885 (2003).
68. B. J. McGill, Strong and weak tests of macroecological theory. *Oikos* **102**:679–685 (2003).
69. A. J. McKane, D. Alonso and R. V. Solé, A mean field stochastic theory for species rich assembled communities. *Phys. Rev. E* **62**:8466–8484 (2000).
70. A. J. McKane, D. Alonso and R. V. Solé, Analytic solution of Hubbell's model of local community dynamics. *Theor. Popul. Biol.* **65**:67–73 (2004).
71. L. R. Moore, G. Rocap and S. W. Chisholm, Physiology and molecular phylogeny of coexisting prochlorococcus ecotypes. *Nature* **393**:464–467 (1998).
72. P. A. P. Moran, Random processes in genetics. *Proc. Cambridge Philos. Soc.* **54**:60–71 (1958).
73. P. A. P. Moran, *Statistical Processes of Evolutionary Theory.* Oxford, U. K.: Clarendon Press (1962).
74. N. Mouquet and M. Loreau, Coexistence in metacommunities: the regional similarity hypothesis. *Am. Nat.* **159**:420–426 (2002).
75. S. Nee, The neutral theory of biodiversity: do the numbers add up? *Funct. Ecol.* **19**:173–176 (2005).
76. F. W. Preston, The commonness and rarity of species. *Ecology* **29**:254–283 (1948).
77. S. Pueyo, Diversity: between neutrality and structure. *Oikos* **112**:392–405 (2006).
78. G. Sella and A. E. Hirsh, The application of statistical physics to evolutionary biology. *Proc. Natl. Acad. Sci., USA* **102**:9541–9546 (2005).
79. G. Sugihara, Minimal community structure: an explanation of species-abundance patterns. *Am. Nat.* **116**:770–787 (1980).
80. D. Tilman and S. Pacala, The maintenance of species richness in plant communities. in: *Species Diversity in Ecological Communities; Historical and Geographic Perspectives*, R. E. Ricklefs and D. Schluter, eds. Chicago, IL: University of Chicago Press, pp. 13–25 (1993).
81. R. Thompson and C. Townsend, A truce with neutral theory: local deterministic factors, species traits and dispersal limitation together determine patterns of diversity in stream invertebrates. *J. Anim. Ecol.* **75**:476–484 (2006).
82. D. Tilman, Diversity by default. *Science* **283**:495–496 (1999).
83. C. S. Ting, G. Rocap, J. King and S. W. Chisholm, Cyanobacterial photosynthesis in the oceans: the origins and significance of divergent light-harvesting strategies. *Trends in Microbiol.* **10**:134–142 (2002).
84. M. Tokeshi, Niche apportionment or random assortment—species abundance patterns revisited. *J. Anim. Ecol.* **59**:1129–1146 (1990).
85. M. Tokeshi, Species abundance patterns and community structure. *Adv. Ecol. Res.* **24**:111–186 (1993).
86. M. Tokeshi, Power fraction: a new explanation of relative abundance patterns in species-rich assemblages. *Oikos* **75**:543–550 (1996).

87. M. Vallade and B. Houchmandzadeh, Analytical solution of a neutral model of biodiversity. *Phys. Rev. E* **68**:061902 (2003).
88. T. Vicsek, *Fluctuations and Scaling in Biology*. Oxford, U. K.: Oxford University Press (2001).
89. I. Volkov, J. R. Banavar, S. P. Hubbell and A. Maritan, Neutral theory and relative species abundance in ecology. *Nature* **424**:1035–1037 (2003).
90. I. Volkov, J. R. Banavar, S. P. Hubbell and A. Maritan, Organization of ecosystems in the vicinity of a novel phase transition. *Phys. Rev. Lett.* **92**:218703 (2004).
91. I. Volkov, J. R. Banavar, F. He, S. P. Hubbell and A. Maritan, Density dependence explains tree species abundance and diversity in tropical forests. *Nature* **438**:658–661 (2005).
92. J. Wakeley, *Coalescent Theory. An Introduction*. Greenwood Village, CO: Roberts & Co (2004).
93. G. A. Watterson, Models for the logarithmic species abundance distribution. *Theor. Popul. Biol.* **6**:217–250 (1974).
94. R. H. Whittaker, A consideration of climax theory: the climax as a population and pattern. *Ecol. Monogr.* **23**:41–78 (1953).
95. R. J. Williams and N. D. Martinez, Simple rules yield complex food webs. *Nature* **404**:180–183 (2000).
96. S. Wright, Evolution in Mendelian populations. *Genetics* **16**:97–159 (1931).
97. D. W. Yu, J. W. Terborgh and M. D. Potts, Can high tree species richness be explained by Hubbell's null model? *Ecol. Lett.* **1**:193–199 (1998).
98. T. Zillio, I. Volkov, J. R. Banavar, S. P. Hubbell and A. Maritan, Spatial scaling in model plant communities. *Phys. Rev. Lett.* **95**:098101 (2005).