

Self-consistent approach for neutral community models with speciation

Bart Haegeman*

INRIA Research Team MERE, UMR Systems Analysis and Biometrics, 2 Place Pierre Viala, F-34060 Montpellier, France

Rampal S. Etienne

Community and Conservation Ecology Group, Centre for Ecological and Evolutionary Studies, University of Groningen, P.O. Box 14, 9750 AA Haren, The Netherlands

(Received 2 September 2009; published 11 March 2010)

Hubbell's neutral model provides a rich theoretical framework to study ecological communities. By incorporating both ecological and evolutionary time scales, it allows us to investigate how communities are shaped by speciation processes. The speciation model in the basic neutral model is particularly simple, describing speciation as a point-mutation event in a birth of a single individual. The stationary species abundance distribution of the basic model, which can be solved exactly, fits empirical data of distributions of species' abundances surprisingly well. More realistic speciation models have been proposed such as the random-fission model in which new species appear by splitting up existing species. However, no analytical solution is available for these models, impeding quantitative comparison with data. Here, we present a self-consistent approximation method for neutral community models with various speciation modes, including random fission. We derive explicit formulas for the stationary species abundance distribution, which agree very well with simulations. We expect that our approximation method will be useful to study other speciation processes in neutral community models as well.

DOI: [10.1103/PhysRevE.81.031911](https://doi.org/10.1103/PhysRevE.81.031911)

PACS number(s): 87.23.-n

I. INTRODUCTION

Neutral community theory assumes that all species in an ecological community have the same ecological function (the neutrality assumption) and explains community diversity as a stochastic balance between species origination and species extinction rather than the result of niche assembly where each species specializes to occupy a unique niche in resource utilization [1–4]. Species extinction is due the stochastic fluctuations of birth-death processes governing the abundance dynamics of the species. Species origination is due to formation of new species (speciation) or immigration. Note that speciation in the context of neutral community models has a rather restricted meaning, because new species still have the same function as existing species, even though they can be morphologically different. In Hubbell's most used neutral community model [3,4] these two mechanisms are treated on two different scales: speciation occurs at the large regional scale, while at the small local scale there is immigration from the regional community. Speciation and immigration processes thus counterbalance species extinctions in the regional and local communities, respectively.

In the first years after its launch in 2001 [2], neutral community theory was heavily criticized for its seemingly outrageous neutrality assumption (e.g., [5–10]) which stands in sharp contrast to classical niche theory in which all species are functionally different. Indeed, one might argue that a neutrality theory of biodiversity is a *contradictio in terminis*, because it assumes that there is no functional diversity. Neutral theory allows for morphological diversity and genetic diversity, however, and is usually restricted to communities

of a single functional group on the same trophic level, such as corals or tropical forests [11]; not even the strongest supporter of neutral theory would assume functional equivalence of predators and prey, or plants and pollinators. Seen in this light, functional groups could be defined by neutral theory: it is an assembly of species that matches predictions of neutral theory well. Neutral theory, then, is a potential solution to the paradox of the plankton [4,12] or, in this context, more appropriately called the paradox of the tropical forest: the amazingly high tree diversity on a small plot of tropical forest requires an unlikely large number of niches. Alternatively, neutrality could have arisen as an emergent consequence of community evolution [13–15]. Most neutralists, however, do not advocate this restricted form of neutrality: they only view neutrality as a first approximation to reality, similar to ideal gas theory in physics [3]: much of the observed patterns in nature can already be explained by simple rules that do not depend on differences between species. It is therefore a natural starting point, a null model, of community structure. Observed deviations from the predictions of neutral theory are not so much regarded as rejecting the theory, but as the interesting patterns that require an ecological explanation. To this end, neutral theory must mature first, particularly in the way that other elements of the theory, such as speciation and dispersal, are modeled; and more quantitative predictions need to be derived for the resulting models.

Here, we consider quantitative predictions at the regional scale, where speciation maintains diversity. Hubbell's basic model contains the point-mutation (PM) mode of speciation: each birth event has a small probability of producing a mutation that leads to an individual of a new species. Chromosome doubling is a simple example of this mode of speciation. Point-mutation speciation can be mimicked by assuming a constant inflow of new species as singletons [16]. Stationary species abundance distribution can be obtained

*bart.haegeman@inria.fr

exactly for the PM mode of speciation [17,18]. Hubbell also introduced an alternative model, the random-fission (RF) mode of speciation in which new species appear by randomly splitting up existing populations. This is a phenomenological model of allopatric speciation. No rigorous analytical expressions exist for the stationary species abundance distribution with RF speciation (see [19] for some approximate formulas based on simulations) or any other speciation model that is essentially different from the PM model. Here, we study a general class of neutral community models with speciation, including the PM and RF modes of speciation, and derive an excellent approximation for the stationary species abundance distribution.

The approximation is based on the fact that the two-scale model (implicitly) assumes a very large number of individuals in the regional community, even though the number of individuals in the local community can be relatively small [20]. Hence, we can utilize techniques from statistical mechanics to study the properties of regional-scale neutral communities; the thermodynamical limit corresponds to taking the regional community size to infinity. The correspondence with statistical mechanics is not straightforward, however, because the number of species increases as the number of individuals tends to infinity. Hence, the number of individuals in a given species can have a scaling behavior that is different from the total number of individuals in the community. We circumvent this complication by constructing an approximation scheme that is independent of such scaling relations.

II. HUBBELL'S NEUTRAL COMMUNITY MODEL

Neutral community theory uses stochastic models to describe the dynamics of the species composition of an ecological community. We first introduce the state space of these models and discuss two different descriptions of community composition. Next, we define Hubbell's regional-scale community model by specifying the transition rates on the state space. For the sake of completeness and comparison, we also give the transition rates for Hubbell's local-scale community model.

A. State space

There are basically two different ways to describe the species composition of a community. As we will use both throughout this paper, we introduce them here.

The first description considers species as labeled entities. Suppose we have a pool of S_T species (where S_T may be infinite; we take this limit later on). Each individual present in the community belongs to one of these S_T species (some species in the pool can be absent in the community). We specify for each of the S_T species its abundance. Hence, the labeled species description is given by a vector \vec{N} with S_T components,

$$\vec{N} = (N_1, N_2, \dots, N_{S_T}).$$

Component i gives the number of individuals N_i that belong to species i . The number of species S present in the community is given by

$$S = \sum_{i=1}^{S_T} \delta(N_i \geq 1),$$

where $\delta(N_i \geq 1) = 1$ if the condition $N_i \geq 1$ is satisfied and $\delta(N_i \geq 1) = 0$, otherwise. The number of individuals N in the community is given by

$$N = \sum_{i=1}^{S_T} N_i.$$

The second description considers species as unlabeled entities. Instead of specifying the abundance of each species individually, we specify the number of species with a given abundance. Hence, the unlabeled species description is given by a vector \vec{S} with N components,

$$\vec{S} = (S_1, S_2, \dots, S_N).$$

Component k gives the number of species S_k that have k individuals. The number of species S present in the community is given by

$$S = \sum_{k \geq 1} S_k = S_T - S_0,$$

where S_0 denotes the number of species in the pool that are absent from the community. The number of individuals N in the community is given by

$$N = \sum_{k \geq 1} k S_k.$$

Local-scale neutral community models can be formulated using the labeled species description. Indeed, the regional community fixes the pool of S_T species that are possibly present in the local community. Hence, the state of the local community is fully specified by the abundance of each of the S_T species. In Hubbell's neutral community model the number of individuals N is considered fixed (as we will discuss later, we can relax this assumption). The state space of the local community model is given by

$$\mathcal{N}_N = \left\{ \vec{N} \mid \vec{N} = (N_1, N_2, \dots, N_{S_T}), \sum_i N_i = N \right\}, \quad (1)$$

i.e., we consider all communities \vec{N} with a given number of individuals N . For example, if $S_T = 2$, then

$$\mathcal{N}_3 = \{(3, 0), (2, 1), (1, 2), (0, 3)\}.$$

In words, there are four labeled species states for three individuals and two distinct (i.e., labeled) species: the first species has three individuals, and the second species has no individuals [$\vec{N} = (3, 0)$]; the first species has two individuals, and the second species has one individual [$\vec{N} = (2, 1)$]; the first species has one individual, and the second species has two individuals [$\vec{N} = (1, 2)$]; or the first species has no individuals, and the second species has three individuals [$\vec{N} = (0, 3)$].

Regional-scale neutral community models are most conveniently described in the unlabeled species description. Indeed, the speciation process continually introduces new spe-

cies in the community, so that we cannot specify beforehand a pool of species that can be present in the community. In Hubbell's neutral community model the number of individuals N is constant over time. The state space \mathcal{S}_N is therefore

$$\mathcal{S}_N = \left\{ \vec{S} | \vec{S} = (S_1, S_2, \dots, S_N), \sum_k k S_k = N \right\}, \quad (2)$$

i.e., we consider all communities \vec{S} with a fixed number of individuals N . For example,

$$\mathcal{S}_3 = \{(3,0,0), (1,1,0), (0,0,1)\},$$

meaning that a community of three individuals is described by one of three unlabeled species states: all individuals belong to different species [$\vec{S}=(3,0,0)$]; two individuals belong to the same species, and the remaining individual belongs to a different species [$\vec{S}=(1,1,0)$]; or all individuals belong to the same species [$\vec{S}=(0,0,1)$]. Similarly, a community of four individuals is described by one of five unlabeled species states,

$$\mathcal{S}_4 = \{(4,0,0,0), (2,1,0,0), (0,2,0,0), (1,0,1,0), (0,0,0,1)\}.$$

B. Transition rates for regional community

Hubbell's regional-scale neutral community models can be considered as a continuous-time Markovian process on the state space \mathcal{S}_N . We denote the transition rate to go from state \vec{S}' to state \vec{S}'' by $R(\vec{S}', \vec{S}'')$. Specifying the matrix of transition rates $R(\vec{S}', \vec{S}'')$, $\vec{S}', \vec{S}'' \in \mathcal{S}_N$, fully defines the community model. Denoting by $P(\vec{S}(t)=\vec{S}')$ the probability that at time t the process' state is \vec{S}' , the master equation [21] is given by

$$\begin{aligned} \frac{d}{dt} P(\vec{S}(t) = \vec{S}') &= \sum_{\vec{S}'' \neq \vec{S}'} [P(\vec{S}(t) = \vec{S}'') R(\vec{S}'', \vec{S}') \\ &\quad - P(\vec{S}(t) = \vec{S}') R(\vec{S}', \vec{S}'')]. \end{aligned} \quad (3)$$

There are two types of events in the regional community: death-birth events, with transition rate $R^{\text{DB}}(\vec{S}', \vec{S}'')$, and speciation events, with transition rate $R^{\text{SP}}(\vec{S}', \vec{S}'')$. Hence, the total transition rate is

$$R(\vec{S}', \vec{S}'') = R^{\text{DB}}(\vec{S}', \vec{S}'') + R^{\text{SP}}(\vec{S}', \vec{S}'').$$

In a death-birth event, first an individual is selected to die (all N individuals have the same probability to be selected), and then another individual is selected to reproduce (all remaining $N-1$ individuals have the same probability to be selected). Note that such an event conserves the number of individuals N . Assume that the dying individual belongs to a species with abundance k and that the reproducing individual belongs to a species with abundance ℓ . Then, the resulting state \vec{S}'' of a transition $R^{\text{DB}}(\vec{S}', \vec{S}'')$ is

$$\vec{S}'' = \vec{S}' - \vec{e}_k + \vec{e}_{k-1} - \vec{e}_\ell + \vec{e}_{\ell+1},$$

with \vec{e}_k as the k th unit vector. The corresponding transition rate is

$$R^{\text{DB}}(\vec{S}, \vec{S} - \vec{e}_k + \vec{e}_{k-1} - \vec{e}_\ell + \vec{e}_{\ell+1}) = \mu \frac{k S_k}{N} \frac{\ell S_\ell}{N-1}, \quad (4)$$

with μ as the community-level death-birth rate. We will use the shorthand notation $R_{k,\ell}^{\text{DB}}(\vec{S})$ for the transition rate (4). Note that a death-birth event with $\ell=k-1$ has no net effect in the unlabeled species description (a species with abundance k loses an individual and a species with abundance $k-1$ gains an individual, so that S_k and S_{k-1} are constant).

In a speciation event, first an individual is selected (all N individuals have the same probability to be selected). The species the individual belongs to undergoes speciation. Assume that this species has abundance k . Then, after the speciation event, this species has abundance $k-\ell$ and a new species with ℓ individuals enters the community. To determine the abundance ℓ of the new species, we sample from a probability distribution $s^{(k)}$ on the set $\{1, 2, \dots, k-1\}$. Note that such a speciation event conserves the number of individuals N . Starting from a state \vec{S}' , the resulting state \vec{S}'' is

$$\vec{S}'' = \vec{S}' - \vec{e}_k + \vec{e}_\ell + \vec{e}_{k-\ell},$$

and the corresponding transition rate is

$$R^{\text{SP}}(\vec{S}, \vec{S} - \vec{e}_k + \vec{e}_\ell + \vec{e}_{k-\ell}) = \nu \frac{k S_k}{N} s^{(k)}(\ell), \quad (5)$$

with ν as the community-level speciation rate. We will use the shorthand notation $R_{k,\ell}^{\text{SP}}(\vec{S})$ for the transition rate (5). Note that a speciation event with $k=1$ has no net effect in the unlabeled species description (a singleton species is replaced with a new species with one individual, so that S_1 is constant). Also, in the unlabeled species description an event $R_{k,\ell}^{\text{SP}}(\vec{S})$ is equivalent to an event $R_{k,k-\ell}^{\text{SP}}(\vec{S})$.

We consider two explicit models of the speciation process. In the PM mode of speciation, a new species always consists of a single individual, i.e.,

$$s^{(k)}(\ell) = \delta_1(\ell) \quad \text{for point mutation.} \quad (6)$$

The transition rate of PM speciation events is given by

$$R^{\text{PM}}(\vec{S}, \vec{S} - \vec{e}_k + \vec{e}_{k-1} + \vec{e}_1) = \nu \frac{k S_k}{N}. \quad (7)$$

Alternatively, in the RF mode of speciation, the abundance of the new species is determined by randomly splitting the old species into two fragments. All fragment sizes have the same probability, i.e.,

$$s^{(k)}(\ell) = \frac{1}{k-1} \quad \text{for random fission.} \quad (8)$$

The transition rate of RF speciation events is given by

$$R^{\text{RF}}(\vec{S}, \vec{S} - \vec{e}_k + \vec{e}_\ell + \vec{e}_{k-\ell}) = \nu \frac{k S_k}{N} \frac{1}{k-1}, \quad (9)$$

for which we will use the shorthand notation $R_k^{\text{RF}}(\vec{S})$.

Hubbell's basic neutral model combines death-birth events (4) with PM speciation events (7). This defines an irreducible Markovian process on the finite state space \mathcal{S}_N , so that there exists a unique stationary distribution on \mathcal{S}_N [21].

Different techniques have been used to obtain this stationary distribution for any value of N . A straightforward derivation is based on the detailed-balance condition [22].

Hubbell's neutral model with RF speciation combines death-birth events (4) with speciation events (9). Again, we obtain an irreducible Markovian process with a unique stationary distribution on \mathcal{S}_N . However, the stationary distribution seems to be difficult to compute analytically; detailed balance is not satisfied. Some stationary properties of this model have been studied using numerical simulations [2,19]. In this paper we propose an approximation scheme that yields excellent results for large N .

C. Transition rates for local community

This paper focuses on Hubbell's regional community model. However, our analysis exploits the analogy with the local community model. Therefore, we give a brief description of Hubbell's local community model. More details can be found in [4,23].

As mentioned above, the local community model uses the labeled species description. Species can be labeled because the local community is assumed to be coupled to a very large regional community, the composition of which is constant on the local time scale (see [20] for a justification of this assumption). The regional community consists of S_T species; the relative abundance of species i is denoted by p_i . These are positive numbers summing up to 1,

$$\sum_{i=1}^{S_T} p_i = 1.$$

The local community composition is described by the abundance vector \vec{N} with S_T components, corresponding to the S_T species of the regional community. Moreover, the number of individuals N in the local community is constant over time.

There are two types of local events. Local death-birth events are analogous to regional death-birth events. First an individual is selected to die, and then another individual is selected to reproduce. Assuming that the current state is \vec{N}' , the dying individual belongs to species i , and the reproducing individual belongs to species $j \neq i$, the new state \vec{N}'' is

$$\vec{N}'' = \vec{N}' - \vec{e}_i + \vec{e}_j,$$

and the corresponding transition rate is

$$R^{\text{DB}}(\vec{N}, \vec{N}' - \vec{e}_i + \vec{e}_j) = \eta \frac{N_i}{N} \frac{N_j}{N-1},$$

with η as the total death-birth rate in the local community.

Local immigration events are comparable to regional speciation events. First an individual from the local community (suppose it belongs to species i) is selected to die, and then an individual from the regional community enters the local community. The probability that the immigrating individual belongs to the species j is equal to the relative abundance n_j . If the current state is \vec{N}' , then the new state \vec{N}'' is

$$\vec{N}'' = \vec{N}' - \vec{e}_i + \vec{e}_j,$$

and the corresponding transition rate is

$$R^{\text{IMM}}(\vec{N}, \vec{N}' - \vec{e}_i + \vec{e}_j) = \zeta \frac{N_i}{N} p_j,$$

with ζ as the total immigration rate from regional to local community.

The combination of local death-birth events with local immigration events defines an irreducible Markovian process on the finite state space \mathcal{N}_N . Hence, there exists a unique stationary distribution. Note that this local community distribution depends on the regional community composition (i.e., the relative abundances p_i). The local stationary distribution can be averaged over the regional stationary distribution. This yields a local community distribution that depends only on the rate parameters μ , ν , η , and ζ (more precisely, on the ratios $\frac{\nu}{\mu}$ and $\frac{\zeta}{\eta}$ of rate parameters). The latter distribution is the starting point for comparison between theory and observation; it can be used to infer the likelihood of the neutral community model given field data [24]. In this paper we deal with the stationary distribution of the regional community for a general class of speciation models and illustrate this for PM and RF; the derivation of the local community distribution and the comparison with field data for these specific models have been or will be reported elsewhere [24,25].

III. SELF-CONSISTENT APPROXIMATION

In the previous section we have introduced the neutral community model by specifying the transition rates (4) and (5). The neutrality structure is clearly present in these transition rates: the probability that a given species is selected to die, reproduce, or speciate is proportional to its abundance and independent of its identity. Hence, all individuals undergo the different processes with identical rates, independent of the species they belong to.

Consequently, when we focus on a particular species i , its dynamics are completely determined by its own abundance N_i and by the abundance of all other species taken together, equal to $N - N_i$. Indeed, the only coupling with the other species is due to the constant community size constraint (known as the zero-sum assumption in the ecological literature). However, this coupling is rather weak as one can expect that, for large community size N and for large number of species S , the community size constraint is satisfied (approximately) just by statistical averaging. This mechanism is analogous to the equivalence of ensembles in statistical mechanics. Recall that both N and S are large in the regional community models we are considering here.

Dropping the community constant size constraint corresponds to decoupling death and birth events. This means replacing the transitions $R_{k,\ell}^{\text{DB}}(\vec{S})$ [see Eq. (4)] with two types of events:

$$R^{\text{B}}(\vec{S}, \vec{S} - \vec{e}_k + \vec{e}_{k+1}) = \beta k S_k \quad \text{for birth events,}$$

$$R^{\text{D}}(\vec{S}, \vec{S} - \vec{e}_k + \vec{e}_{k-1}) = \delta k S_k \quad \text{for death events,}$$

with β and δ as the per capita death and birth rates, respectively. One must assume $\beta < \delta$ to avoid the community size N running away to infinity; it is rescued from extinction by constant immigration.

For local community models, the decoupling of death and birth events leads to the decoupling of the dynamics of different species. The resulting independent species model is particularly simple, because species can be considered separately. The stationary distribution can be computed explicitly and turns out to be identical to the stationary distribution of the constant community size model, when conditioned on total community size [23,26].

The situation is more delicate for the regional community model. Whereas the immigration process in the local community model is independent of the local community composition, the speciation process in the regional community model does depend on the composition of the regional community composition. Therefore, decoupling death and birth events does not lead to an independent species model. Moreover, with variable community size N , the stochastic community dynamics get trapped in the absorbing state $N=0$ (assuming that $\beta < \delta$). Indeed, whereas the immigration process in the local community model is still active in the state $N=0$, the speciation process in the regional community model gets halted in the state $N=0$: there are no individuals to speciate.

We propose the following approach to this problem. Our goal is to describe speciation (which is an internal process) as immigration (which is an external process) because this is much more tractable. We are motivated by simple models in which the results are identical [23]. Therefore, we first solve a community model in which new species enter the community as immigrants (Sec. IV). We compute the stationary distribution of this model, a computation similar to the local community model. Our model is slightly more general, because we allow for immigration events with several immigrants at once. It reduces to an independent species model, which is most easily solved in a labeled species description. We obtain a nontrivial stationary distribution, which we then convert back to the unlabeled species description.

Because speciation is modeled as immigration, the speciation rate, i.e., the rate at which new species enter the community, is constant over time. Similarly, the abundance of new species entering the community is taken from a fixed distribution, which is independent of community composition. However, as stated earlier, the speciation process (5) in the regional community model is controlled by the composition of the community. We approximate this using a self-consistent approach: we match the species inflow (both in terms of rate and abundance distribution) with the corresponding stationary community composition (Sec. V). In other words, we use an appropriately chosen externally controlled species inflow to construct the stationary distribution for internally controlled species inflow. We show that this approach yields accurate results for a range of speciation models.

IV. EXTERNALLY CONTROLLED SPECIES INFLOW

The neutral community model given by transition rates (4) and (5) describes a rather intricate speciation process, as species inflow depends on community composition and can therefore fluctuate in time. In this section we study a simpli-

fied neutral community model, in which species inflow is constant over time. The stationary distribution of this simplified model will be helpful to construct an approximate stationary distribution of the full community model, as we will show in the next section.

A. Immigration model

We consider a community model in which three types of events occur: birth, death, and immigration events. We work with labeled species, assume a fixed pool of S_T species, and allow the community size N to fluctuate, in contrast with Hubbell’s community model (Sec. II). The state space of the community model of this section is given by all vectors \vec{N} with S_T components, without restriction on community size.

The dynamics of different species are independent, so it suffices to consider a single species i . Denote the per capita birth rate by β , the per capita death rate by δ , and the species-level immigration rate by λ . We allow for immigration events with several individuals. When an immigration event occurs, the probability that k individuals immigrate at once is given by q_k . These are positive numbers summing up to 1,

$$\sum_{k \geq 1} q_k = 1. \tag{10}$$

We consider the case for which the immigration rate λ is small, i.e., $\lambda \ll \beta < \delta$. In fact, we will be interested in the limit $\lambda \rightarrow 0$ (see below).

We consider the continuous-time Markovian process for the abundance N_i of species i . Using the shorthand notation $p_k(t) = \mathbb{P}(N_i(t) = k)$, the master equation is given by

$$\begin{aligned} \frac{d}{dt} p_k(t) &= \beta(k-1)p_{k-1}(t) + \delta(k+1)p_{k+1}(t) - (\beta + \delta)kp_k(t) \\ &+ \lambda \sum_{\ell=1}^k q_\ell p_{k-\ell}(t) - \lambda p_k(t). \end{aligned} \tag{11}$$

We look for the stationary probabilities that species i has k individuals. Using the shorthand notation $p_k = \lim_{t \rightarrow \infty} \mathbb{P}(N_i(t) = k)$, the equations are

$$\begin{aligned} \delta p_1 &= \lambda \sum_{k \geq 1} q_k p_0, \\ 2\delta p_2 &= \beta p_1 + \lambda \sum_{k \geq 2} q_k p_0 + \lambda \sum_{k \geq 1} q_k p_1, \\ 3\delta p_3 &= 2\beta p_2 + \lambda \sum_{k \geq 3} q_k p_0 + \lambda \sum_{k \geq 2} q_k p_1 + \lambda \sum_{k \geq 1} q_k p_2, \\ &\dots \end{aligned}$$

We also have the normalization condition

$$\sum_{k \geq 0} p_k = 1.$$

Solving these equations in terms of p_0 , we find

$$\frac{p_1}{p_0} = \frac{\lambda}{\delta} \sum_{k \geq 1} q_k,$$

$$\frac{p_2}{p_0} = \frac{\beta\lambda}{2\delta^2} \sum_{k \geq 1} q_k + \frac{\lambda}{2\delta} \sum_{k \geq 2} q_k + \mathcal{O}(\lambda^2),$$

$$\frac{p_3}{p_0} = \frac{\beta^2\lambda}{3\delta^3} \sum_{k \geq 1} q_k + \frac{\beta\lambda}{3\delta^2} \sum_{k \geq 2} q_k + \frac{\lambda}{3\delta} \sum_{k \geq 3} q_k + \mathcal{O}(\lambda^2),$$

....

We rewrite this solution as

$$p_k = \begin{cases} \frac{\lambda}{\delta} \frac{Q_k}{k} + \mathcal{O}(\lambda^2) & \text{for } k \geq 1 \\ \frac{1}{Z_1} & \text{for } k = 0, \end{cases} \quad (12)$$

with

$$Q_k = \sum_{\ell=1}^k \rho^{k-\ell} \sum_{m \geq \ell} q_m, \quad \text{with } \rho = \frac{\beta}{\delta}, \quad (13)$$

$$Z_1 = 1 + \frac{\lambda}{\delta} \sum_{k \geq 1} \frac{Q_k}{k} + \mathcal{O}(\lambda^2). \quad (14)$$

By combining the one-species stationary distributions (12), we get the stationary distribution for the (labeled) community composition \vec{N} . Because species are independent, we have to take the product of the one-species distributions,

$$\begin{aligned} P(\vec{N}) &= \frac{1}{Z_1^{S_0}} \prod_{i: N_i \geq 1} \left(\frac{\lambda}{\delta} \frac{Q_{N_i}}{N_i} + \mathcal{O}(\lambda^2) \right) \\ &= \frac{1}{Z_1^{S_0}} \left(\frac{\lambda}{\delta} \right)^{S_T - S_0} \left(\prod_{i: N_i \geq 1} \frac{Q_{N_i}}{N_i} \right) + \mathcal{O}(\lambda^{S_T - S_0 + 1}), \end{aligned}$$

where S_0 is the number of species in the pool that are absent from the community composition \vec{N} . We are interested in the stationary distribution for unlabeled community composition \vec{S} . To obtain this we remove the species labels,

$$\begin{aligned} P(S_0, S_1, S_2, \dots) &= \frac{S_T!}{S_0! \prod_{k \geq 1} S_k!} \left\{ \frac{1}{Z_1^{S_0}} \left(\frac{\lambda}{\delta} \right)^{S_T - S_0} \left[\prod_{k \geq 1} \left(\frac{Q_k}{k} \right)^{S_k} \right] \right. \\ &\quad \left. + \mathcal{O}(\lambda^{S_T - S_0 + 1}) \right\} \\ &= \frac{S_T!}{S_0!} \left\{ \frac{1}{Z_1^{S_0}} \left(\frac{\lambda}{\delta} \right)^{S_T - S_0} \left[\prod_{k \geq 1} \frac{1}{S_k!} \left(\frac{Q_k}{k} \right)^{S_k} \right] \right. \\ &\quad \left. + \mathcal{O}(\lambda^{S_T - S_0 + 1}) \right\} \\ &= \frac{S_T!}{S_0!} \frac{1}{Z_1^{S_0}} \left(\frac{\lambda}{\delta} \right)^{S_T - S_0} \left[\prod_{k \geq 1} \frac{1}{S_k!} \left(\frac{Q_k}{k} \right)^{S_k} + \mathcal{O}(\lambda) \right]. \end{aligned}$$

B. Speciation as immigration

We have obtained the stationary distribution of the community immigration model (for small λ). Now we take into account that we are interested in the immigration process as a speciation model. Thus, every immigration event introduces a new species into the community. This situation corresponds to a species pool with very high diversity, i.e., we have to take the limit $S_T \rightarrow \infty$. Simultaneously, we keep the community-level immigration (or speciation) rate $\nu = S_T \lambda$ constant. This implies that for the species-level immigration rate λ we must take the limit $\lambda \rightarrow 0$. We get

$$\begin{aligned} P(\vec{S}) = P(S_1, S_2, \dots) &= \frac{1}{Z_2} \prod_{k \geq 1} \frac{1}{S_k!} \left(\frac{\nu Q_k}{\delta k} \right)^{S_k} \\ &= \frac{1}{Z_2} \prod_{k \geq 1} \frac{1}{S_k!} \left(\frac{\rho \theta Q_k}{k} \right)^{S_k}, \end{aligned} \quad (15)$$

with

$$\theta = \frac{\nu}{\beta}, \quad Z_2 = \exp \left(\rho \theta \sum_{k \geq 1} \frac{Q_k}{k} \right). \quad (16)$$

Equation (15) gives the stationary distribution for externally controlled species inflow. It has a particularly simple structure: the components of vector \vec{S} are mutually independent, and the component S_k is Poisson distributed with mean

$$\mathbb{E}S_k = \rho \theta \frac{Q_k}{k}. \quad (17)$$

Therefore, the total number of species S in the community is also Poisson distributed with mean

$$\mathbb{E}S = \sum_{k \geq 1} \mathbb{E}S_k = \rho \theta \sum_{k \geq 1} \frac{Q_k}{k},$$

so that $Z_2 = \exp(\mathbb{E}S)$ [see Eq. (16)]. The distribution for the number of individuals N satisfies

$$\mathbb{E}N = \sum_{k \geq 1} k \mathbb{E}S_k = \rho \theta \sum_{k \geq 1} Q_k,$$

$$\text{Var } N = \sum_{k \geq 1} k^2 \text{Var } S_k = \rho \theta \sum_{k \geq 1} k Q_k. \quad (18)$$

C. Community size constraint

The immigration model of this section has a fluctuating community size N . However, Hubbell's community model, given by transition rates (4) and (5), has a fixed community size N . In Hubbell's model the number of individuals N is a parameter, whereas in the immigration model the number of individuals N is a stochastic variable. Because we will use the stationary distribution of the immigration model as an approximation for the stationary distribution of Hubbell's model, we have to establish the link between both.

To do so, we choose the parameter ρ such that the expected community size $\mathbb{E}N$ in the immigration model equals the community size constraint N of Hubbell's model. Typi-

cally, this choice of ρ effectively imposes the community size constraint, because the stationary distribution for the community size N is often sharply peaked, i.e., $\sqrt{\text{Var } N} \ll \mathbb{E}N$. Note that this inequality also implies that the state $N=0$ has an extremely low stationary probability, so that community extinction is highly improbable.

Alternatively, we can restrict the stationary distribution (15) of the immigration model to those states \vec{S} with appropriate community size N , i.e., the community size N imposed by Hubbell's model. By conditioning Eq. (15) on N ,

$$P(\vec{S}|N) = \frac{1}{Z_3(N)} \prod_{k \geq 1} \frac{1}{S_k!} \left(\frac{\rho \theta Q_k}{k} \right)^{S_k}. \quad (19)$$

In the examples we work out in the next sections, the conditioned distribution $P(\vec{S}|N)$ is independent of ρ , so that conditioning can be used directly (i.e., without first determining ρ) to impose the community size constraint. Obviously, conditioning on N also avoids community extinction.

From Eq. (19) we find that

$$\mathbb{E}(S_k|N) = \frac{Z_3(N-k)}{Z_3(N)} \frac{\rho \theta Q_k}{k}, \quad (20)$$

and, using Eq. (15), that

$$P(N) = \frac{Z_3(N)}{Z_2}. \quad (21)$$

The normalization constants $Z_3(N)$ can be computed with a generating function. On one hand we have, using Eq. (21),

$$\mathbb{E}z^N = \sum_{N \geq 0} P(N)z^N = \frac{1}{Z_2} \sum_{N \geq 0} Z_3(N)z^N,$$

and on the other hand we have, using Eq. (15),

$$\mathbb{E}z^N = \sum_{\vec{S}} P(\vec{S})z^N = \frac{1}{Z_2} \exp\left(\rho \theta \sum_{k \geq 1} \frac{Q_k}{k} z^k\right) = \frac{1}{Z_2} \exp\left(\sum_{k \geq 1} \mathbb{E}S_k z^k\right).$$

Equating these expressions leads to

$$\sum_{N \geq 0} Z_3(N)z^N = \exp\left(\sum_{k \geq 1} \mathbb{E}S_k z^k\right). \quad (22)$$

To get the normalization constants $Z_3(N)$, we have to expand the right-hand side in powers of z^N . This expansion can be computed numerically (if not analytically) with the fast Fourier transform.

V. INTERNALLY CONTROLLED SPECIES INFLOW

In the previous section we computed the stationary distribution (15) of a neutral community model in which new species arrive with fixed abundance distribution. However, in the speciation process (5) the inflow of new species is controlled by the current community composition. In this section we study the feedback of community composition on the speciation process and propose a self-consistent approximation scheme to compute the stationary distribution of the neutral community model with this feedback, expressed in Eqs. (4) and (5).

Consider a community described by an abundance vector \vec{S} with $N = \sum_k S_k$ individuals. Suppose a speciation event occurs. The probability that the speciation happens in a species of abundance k and yields a new species of abundance ℓ is given by

$$\frac{k S_k}{N} s^{(k)}(\ell).$$

Summing this expression over the abundance k of the speciating species, we get the probability $q_\ell(\vec{S})$ that, if a new species appears in the community, it will have abundance ℓ ,

$$q_\ell(\vec{S}) = \sum_{k > \ell} \frac{k S_k}{N} s^{(k)}(\ell).$$

If the community is described by a probability distribution $P(\vec{S})$ or $P(\vec{S}|N)$, the probability vector $\vec{q}(\vec{S})$ is also stochastic. Its probability distribution has typically little dispersion, so that $q_\ell(\vec{S})$ can be replaced with the expectation $\mathbb{E}q_\ell$ [expectation with respect to $P(\vec{S})$ or $P(\vec{S}|N)$]. For a distribution $P(\vec{S}|N)$ with fixed community size, such as Eq. (19), we have

$$\mathbb{E}q_\ell = \sum_{k > \ell} \frac{k \mathbb{E}(S_k|N)}{N} s^{(k)}(\ell). \quad (23)$$

For a distribution $P(\vec{S})$ with variable community size, such as Eq. (15), we have

$$\mathbb{E}q_\ell = \sum_N P(N) \sum_{k > \ell} \frac{k \mathbb{E}(S_k|N)}{N} s^{(k)}(\ell) \approx \sum_{k > \ell} \frac{k \mathbb{E}S_k}{\mathbb{E}N} s^{(k)}(\ell). \quad (24)$$

The latter approximation assumes the distribution of the community size N to be sharply peaked, which is typically the case as we noted earlier. This assumption also implies that Eqs. (23) and (24) yield equivalent results. We prefer to work with Eq. (24) as it is often easier to compute in practical examples.

Note that Eq. (24) does not satisfy the normalization condition (10),

$$\begin{aligned} \sum_{\ell \geq 1} \mathbb{E}q_\ell &\approx \sum_{\ell \geq 1} \sum_{k > \ell} \frac{k \mathbb{E}S_k}{\mathbb{E}N} s^{(k)}(\ell) = \sum_{k \geq 2} \frac{k \mathbb{E}S_k}{\mathbb{E}N} \sum_{\ell=1}^{k-1} s^{(k)}(\ell) \\ &= \sum_{k \geq 2} \frac{k \mathbb{E}S_k}{\mathbb{E}N} = 1 - \frac{\mathbb{E}S_1}{\mathbb{E}N}. \end{aligned}$$

To keep the normalization, we simply add the missing term $\mathbb{E}S_1/\mathbb{E}N$ in component $\mathbb{E}q_1$. This corresponds to speciation in a species with a single individual. Note that such a speciation event has no net effect in terms of (unlabeled) species abundances, as the old species (with one individual) is entirely replaced with a new species (with one individual). Adding the missing term, we obtain the *computed* abundance distribution of immigrating species (denoted with the superscript cmp),

$$q_\ell^{\text{cmp}} = \frac{\mathbb{E}S_1}{\mathbb{E}N} \delta_1(\ell) + \sum_{k>\ell} \frac{k\mathbb{E}S_k}{\mathbb{E}N} s^{(k)}(\ell). \quad (25)$$

Now we have all ingredients to formulate the self-consistency problem. We first *assume* an abundance distribution of immigrating species (denoted with the superscript *ass*),

$$\vec{q}^{\text{ass}} = (q_1^{\text{ass}}, q_2^{\text{ass}}, \dots).$$

Using this distribution \vec{q}^{ass} , we compute the expected number of species $\mathbb{E}S_k$ with abundance k using Eqs. (13), (17), and (18). Equation (25) gives the computed abundance distribution of immigrating species,

$$\vec{q}^{\text{cmp}} = (q_1^{\text{cmp}}, q_2^{\text{cmp}}, \dots).$$

The vector \vec{q}^{cmp} has to be matched with the vector \vec{q}^{ass} we started with. The self-consistency problem consists of finding the abundance distribution \vec{q}^{ass} , so that the resulting abundance distribution $\vec{q}^{\text{cmp}} = \vec{q}^{\text{ass}}$.

The self-consistency equations can be written explicitly. Starting with a vector \vec{q}^{ass} , we have from Eq. (13),

$$Q_k = \sum_{\ell=1}^k \rho^{k-\ell} \sum_{m \geq \ell} q_m^{\text{ass}}.$$

Hence, from Eq. (17),

$$\mathbb{E}S_k = \rho \theta \frac{Q_k}{k} = \rho \theta \frac{1}{k} \sum_{\ell=1}^k \rho^{k-\ell} \sum_{m \geq \ell} q_m^{\text{ass}},$$

and from Eq. (18),

$$\mathbb{E}N = \rho \theta \sum_{k \geq 1} \sum_{\ell=1}^k \rho^{k-\ell} \sum_{m \geq \ell} q_m^{\text{ass}}.$$

Substituting the last two equations into Eq. (25),

$$q_n^{\text{cmp}} = \frac{\delta_1(n)}{\sum_{k \geq 1} \sum_{\ell=1}^k \rho^{k-\ell} \sum_{m \geq \ell} q_m^{\text{ass}}} + \frac{\sum_{k>n} \sum_{\ell=1}^k \rho^{k-\ell} \sum_{m \geq \ell} q_m^{\text{ass}} s^{(k)}(n)}{\sum_{k \geq 1} \sum_{\ell=1}^k \rho^{k-\ell} \sum_{m \geq \ell} q_m^{\text{ass}}}.$$

By interchanging the summations, we get

$$q_n^{\text{cmp}} = \delta_1(n) \frac{1-\rho}{\langle q^{\text{ass}} \rangle} + \sum_{m \geq 1} A_{nm} \frac{q_m^{\text{ass}}}{\langle q^{\text{ass}} \rangle},$$

with $\langle q^{\text{ass}} \rangle$ as the mean abundance of an immigrating species,

$$\langle q^{\text{ass}} \rangle = \sum_{m \geq 1} m q_m^{\text{ass}},$$

$$A_{nm} = \sum_{k>n} \rho^k (\rho^{-\min\{k,m\}} - 1) s^{(k)}(n).$$

Therefore, the self-consistency equations read

$$\langle q \rangle q_n - \sum_{m \geq 1} A_{nm} q_m = \delta_1(n) (1-\rho) \quad \text{for all } n \geq 1,$$

or in matrix notation,

$$(\langle q \rangle \mathbb{1} - A) \vec{q} = (1-\rho) \vec{e}_1, \quad (26)$$

with $\mathbb{1}$ as the identity matrix and \vec{e}_1 as the first unit vector. Equation (26) has to be solved for the vector $\vec{q} = (q_1, q_2, \dots)$. Note that Eq. (26) becomes a linear system of equations if $\langle q \rangle$ can be considered a constant. This property can be exploited to solve Eq. (26) numerically if an analytical solution is not possible.

VI. POINT-MUTATION MODEL

As a first, rather trivial, illustration of our self-consistent approach, we consider the point-mutation (PM) model. In this model a speciation event consists of splitting off one individual from an existing species [see Eq. (6)], so that new species arrive as single individuals,

$$q_\ell = \delta_1(\ell). \quad (27)$$

Actually, there is no self-consistency problem to solve, because the distribution \vec{q} is known beforehand, but the example is still illustrative of our approach. We substitute Eq. (27) in the solution for externally controlled species inflow. From Eq. (13) we find

$$Q_k = \rho^{k-1},$$

and from Eq. (15) we have

$$P(\vec{S}) = \frac{\rho^N}{Z_2} \prod_{k \geq 1} \frac{\theta^{S_k}}{S_k! k^{S_k}}.$$

The normalization constant is [see Eq. (16)]

$$\ln Z_2 = \theta \sum_{k \geq 1} \frac{\rho^k}{k} = -\theta \ln(1-\rho).$$

Hence,

$$P(\vec{S}) = (1-\rho)^\theta \rho^N \prod_{k \geq 1} \frac{\theta^{S_k}}{S_k! k^{S_k}}. \quad (28)$$

The distribution for the community size N can be obtained from Eq. (22),

$$Z_3(N) = \frac{\rho^N}{N!} (\theta)_N, \quad \text{with } (\theta)_N = \theta(\theta+1) \cdots (\theta+N-1),$$

and from Eq. (21),

$$P(N) = (1-\rho)^\theta \frac{(\theta)_N}{N!} \rho^N. \quad (29)$$

This is a negative binomial distribution.

The PM model we are approximating has fixed community size N . As explained above, we can impose this constraint by determining ρ such that $\mathbb{E}N$ equals the community size constraint N . From Eq. (29),

$$\mathbb{E}N = \frac{\rho \theta}{1-\rho},$$

so that

$$\rho = \frac{N}{N + \theta}. \tag{30}$$

Conditioning Eq. (28) on community size we find, using Eq. (19),

$$P(\vec{S}|N) = \frac{N!}{(\theta)_{N \geq 1}} \prod_{S_k} \frac{\theta^{S_k}}{S_k! k^{S_k}}. \tag{31}$$

This is the so-called Ewens' distribution [27], which is the exact stationary distribution of the regional community model with point mutation, i.e., with transition rates (4) and (7). Note that Eq. (31) does not depend on ρ , so that imposing Eq. (30) is superfluous.

In terms of the parameters μ , ν , and N of the PM model, the parameter θ appearing in the exact solution (31) is

$$\theta = \frac{\nu(N - 1)}{\mu}. \tag{32}$$

This equation allows us to establish the link with the parameters β , δ , and ν of the immigration model. For large community size N , we know from Eq. (30) that $\rho \approx 1$. Because $\rho = \frac{\beta}{\delta}$, $\theta = \frac{\nu}{\beta}$, and from Eq. (32), we get

$$\theta \approx \frac{\nu N}{\mu}, \quad \beta \approx \delta \approx \frac{\mu}{N} \quad \text{for large } N. \tag{33}$$

Recall that μ is the community-level death-birth rate, whereas β and δ are per capita birth and death rates, so that this correspondence is not surprising. In fact, relations (33) do not only hold for the PM model, but are generally valid.

Figure 1(a) shows the abundance distribution for the PM model. We plotted the expected number of species per logarithmic (base 2) abundance interval or, more precisely, we plotted $\rho(\log_2(k)) = \ln(2)k \mathbb{E}S_k$ vs $\log_2(k)$, so that the integral equals the expected total number of species,

$$\begin{aligned} \int \rho(\log_2(k)) d \log_2(k) &= \int \ln(2)k \mathbb{E}S_k \frac{dk}{\ln(2)k} \\ &= \int \mathbb{E}S_k dk \approx \mathbb{E}S. \end{aligned}$$

Here, we used a continuum approximation for the abundances k . Logarithmic abundance classes are commonly used in ecology to represent the community composition. The curves for the expected number of species without conditioning ($\mathbb{E}S_k$, blue circles) and with conditioning [$\mathbb{E}(S_k|N)$, red squares] almost coincide, indicating the equivalence between both. The agreement with the simulated curve (green triangles) is excellent, as expected because in this case our solution is exact.

VII. RANDOM-FISSION MODEL

The second application of our self-consistent approximation scheme deals with the random-fission (RF) model. In this model a speciation event consists of splitting a species into two fragments, so that all fragment sizes are equally probable [see Eq. (8)]. The abundance distribution \vec{q} of im-

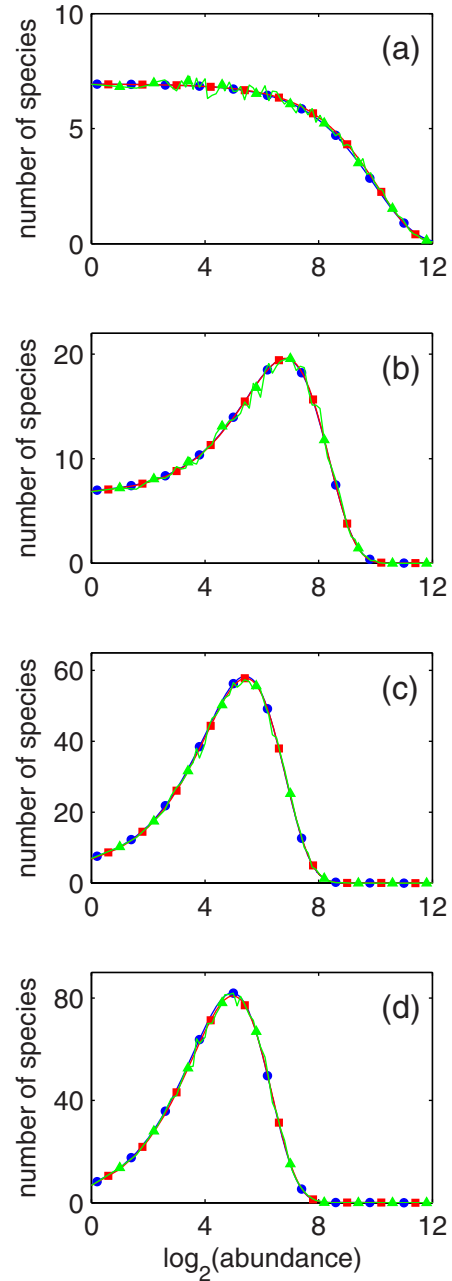


FIG. 1. (Color online) Species abundance distributions as predicted by self-consistent approach. We use Preston-like plots, i.e., the expected number of species per logarithmic (base 2) abundance interval. Blue circles: the expected number of species $\mathbb{E}S_k$ without conditioning on the number of individuals N [see Eq. (17)]. Red squares: the expected number of species $\mathbb{E}(S_k|N)$ with conditioning on the number of individuals [see Eq. (20)]. Green triangles: the expected number of species $\mathbb{E}(S_k)$ as obtained from a simulation of the full model [transition rates (4) and (5)]. First we simulated 10^4 events to reach the stationary regime, and then registered 1000 vectors \vec{S} with intervals of 10^4 events. We computed the mean of these 1000 vectors and regrouped the mean numbers \bar{S}_k in logarithmic (base 1.1) abundance k intervals. The self-consistent curves (blue circles and red squares) almost coincide and convincingly agree with the simulated curves (green triangles). Parameters: $N=10^4$, $\theta = 10$ for all panels, and α increases from top to bottom: (a) $\alpha=0$ or PM, (b) $\alpha=0.05$, (c) $\alpha=0.5$, and (d) $\alpha=1$ or RF.

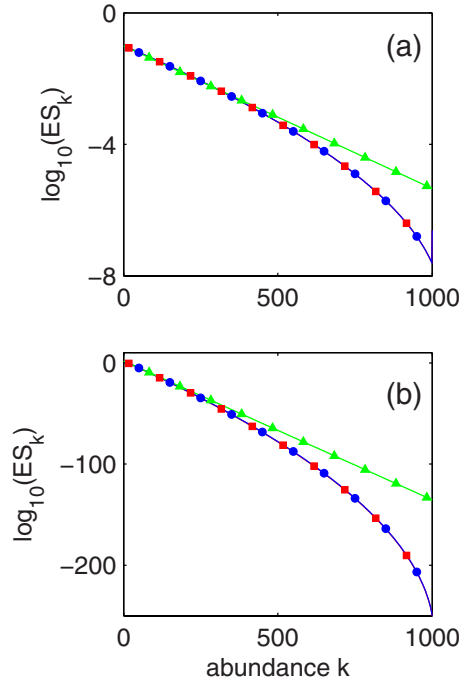


FIG. 2. (Color online) Abundance distributions for random-fission model and its self-consistency scheme. Blue circles: exact equilibrium values $\mathbb{E}(S_k|N)$ computed from Eq. (34). Green triangles: exponentially decreasing functions (35) used to initiate self-consistency scheme. Red squares: approximated values $\mathbb{E}(S_k|N)$ as obtained from self-consistency scheme, which coincide with the exact values (curves of blue circles and red squares coincide). Parameters: $N=1000$ and $\mu=1$. (a) $\theta=0.1$, $\nu=10^{-4}$; (b) $\theta=100$, $\nu=0.1$.

migrating species has to be found self-consistently.

To obtain an ansatz for the distribution \vec{q} we proceed as follows. The master equation (3) can be used to construct an equation for the expected number of species $\mathbb{E}S_\ell$ with abundance ℓ . This equation reads

$$\begin{aligned} \frac{d}{dt}\mathbb{E}S_\ell = & r_{\ell-1}\mathbb{E}S_{\ell-1} - (2r_\ell + s_\ell)\mathbb{E}S_\ell + r_{\ell+1}\mathbb{E}S_{\ell+1} \\ & + \sum_{k=\ell+1}^N s_k[s^{(k)}(\ell) + s^{(k)}(k-\ell)]\mathbb{E}S_k, \end{aligned} \quad (34)$$

with

$$r_\ell = \mu \frac{\ell(N-\ell)}{N(N-1)}, \quad s_\ell = \nu \frac{\ell}{N}.$$

Note that the system of differential equations (34) for $\ell=1, 2, \dots, N$ is autonomous. Indeed, the equations only depend on the set of first-order moments $\mathbb{E}S_k$, and not on higher-order moments such as $\mathbb{E}S_k S_\ell$. This is quite remarkable and seems to be a general property of neutral community models [17,28].

As a consequence, we can solve for the equilibrium solution of Eq. (34) without having to consider the full master equation (3). Figure 2 shows the equilibrium $\mathbb{E}S_k$ as a function of k (blue circles). The solutions are (to a good approxi-

mation) exponentially decreasing, except for large values of k (k comparable to N), for which $\mathbb{E}S_k$ is very small. To find an ansatz for the self-consistency problem, we are mainly interested in abundances $k \ll N$, because $\mathbb{E}S_k$ for large k will be modified by the community size constraint [which is taken into account in Eq. (34), but not in the self-consistency ansatz]. Thus, we fit an exponentially decreasing function to $k \mapsto \mathbb{E}S_k$ (Fig. 2, green triangles) and find

$$\mathbb{E}S_k \approx \frac{\nu}{\mu} N \exp\left(-\sqrt{\frac{\nu}{\mu}} k\right),$$

or in terms of the parameters of the self-consistency problem [see Eq. (33)],

$$\mathbb{E}S_k \approx \theta \exp\left(-\sqrt{\frac{\theta}{N}} k\right). \quad (35)$$

The following assumption for the immigrant abundance distribution \vec{q} also leads to an exponentially decreasing function $k \mapsto \mathbb{E}S_k$,

$$q_\ell^{\text{ass}} = (1-\rho)\rho^{\ell-1}. \quad (36)$$

Indeed,

$$\sum_{m \geq \ell} q_m^{\text{ass}} = \rho^{\ell-1},$$

and, from Eq. (13),

$$Q_k = \sum_{\ell=1}^k \rho^{k-\ell} \rho^{\ell-1} = k \rho^{k-1}, \quad (37)$$

and, using Eq. (17),

$$\mathbb{E}S_k = \rho \theta \frac{Q_k}{k} = \theta \rho^k. \quad (38)$$

The parameter ρ has to be determined so that the expected community size $\mathbb{E}N$, given by

$$\mathbb{E}N = \rho \theta \sum_{k \geq 1} Q_k = \theta \sum_{k \geq 1} k \rho^k = \frac{\rho \theta}{(1-\rho)^2}, \quad (39)$$

equals the community size constraint N . For large N , we have

$$\rho \approx 1 - \sqrt{\frac{\theta}{N}} \approx \exp\left(-\sqrt{\frac{\theta}{N}}\right), \quad (40)$$

so that Eqs. (35) and (38) are consistent.

A further consistency check can be performed by computing the abundance distribution $\mathbb{E}(S_k|N)$ with community size constraint. We determine the normalization constants $Z_3(N)$ from Eq. (22) using the fast Fourier transform and substitute these constants into expression (20) for $\mathbb{E}(S_k|N)$. The result (Fig. 2, red squares) coincides with the exact expression, i.e., the equilibrium solution of Eq. (34), despite the very small values of $\mathbb{E}S_k$ (Fig. 2; note the logarithmic scale on the y axis).

Next, we consider the self-consistency problem as such. Using Eqs. (38) and (39) for $\mathbb{E}S_k$ and $\mathbb{E}N$, the computed immigrant abundance distribution (25) is

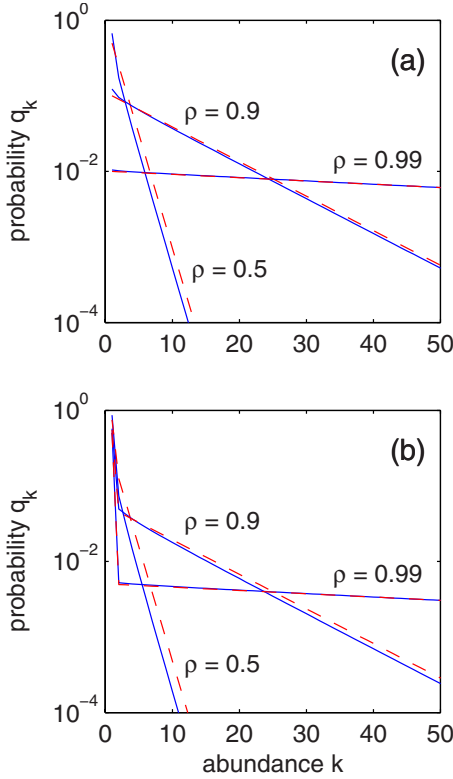


FIG. 3. (Color online) Comparison between assumed and computed immigrant abundance distributions. Dashed red line: assumed distribution \bar{q}^{ass} [see Eqs. (36) and (44)]. Solid blue line: computed distribution \bar{q}^{cmp} [see Eqs. (41) and (46)]. (a) RF model or $\alpha=1$; (b) interpolating model for $\alpha=0.5$. The distributions are plotted for three values of $\rho=0.5, 0.9, 0.99$ and approach each other for increasing values of ρ , i.e., increasing community size N . Assumed (dashed red line) and computed (solid blue line) distributions coincide for $\rho=0.99$.

$$q_\ell^{\text{cmp}} = (1-\rho)^2 \delta_1(\ell) + (1-\rho)^2 \sum_{k>\ell} \rho^{k-1} \frac{k}{k-1}. \quad (41)$$

Figure 3(a) compares this distribution \bar{q}^{cmp} with the ansatz distribution \bar{q}^{ass} [see Eq. (36)]. The agreement between both immigrant abundance distributions is excellent for $\rho \approx 1$, which is satisfied for large N . Based on this observation, together with previous consistency checks, we conclude that we have solved the self-consistency problem.

The corresponding species abundance distribution (15) reads

$$P(\vec{S}) = \frac{\rho^N}{Z_2} \prod_{k \geq 1} \frac{\theta^{S_k}}{S_k!},$$

with normalization constant [see Eq. (16)]

$$\ln Z_2 = \theta \sum_{k \geq 1} \rho^k = \frac{\theta \rho}{1 - \rho}.$$

Hence,

$$P(\vec{S}) = e^{-\theta \rho / (1 - \rho)} \rho^N \prod_{k \geq 1} \frac{\theta^{S_k}}{S_k!}. \quad (42)$$

This distribution can be written in terms of the parameters μ , ν , and N of the RF model using Eqs. (33) and (40),

$$\theta \approx \frac{\nu N}{\mu}, \quad \rho \approx \exp\left(-\sqrt{\frac{\nu}{\mu}}\right).$$

The RF model we are approximating has fixed community size N . This constraint is already implicitly present in distribution (42) for $\rho \approx 1$ or large N , because the corresponding distribution $P(N)$ is then sharply peaked at the community size constraint N . We can also explicitly condition on the community size constraint N [see Eq. (19)]

$$P(\vec{S}|N) = \frac{\rho^N}{Z_3(N)} \prod_{k \geq 1} \frac{\theta^{S_k}}{S_k!}, \quad (43)$$

with normalization constant, using Eq. (22),

$$Z_3(N) = \rho^N \sum_{S=1}^N \frac{\theta^S}{S!} \binom{N-1}{S-1}.$$

Note that $P(\vec{S}|N)$ does not depend on the parameter ρ .

It is interesting to note that not only the distribution $P(N)$ is sharply peaked, but also the distribution $P(S)$ [both are marginal distributions of Eq. (42)]. Hence, the unconditioned distribution (42) is equivalent to the distribution conditioned on the mean values for N and S ,

$$P(\vec{S}|N, S) = \frac{S!}{\binom{N-1}{S-1}} \prod_{k \geq 1} \frac{1}{S_k!}.$$

This distribution no longer depends on the parameters ρ and θ of the self-consistency problem. Its formula becomes particularly simple when written in terms of labeled species,

$$P(\vec{N}|N, S) = \frac{1}{\binom{N-1}{S-1}}.$$

Thus, the (approximated) stationary distribution of the RF model assigns the same probability to all labeled species states \vec{N} for a fixed number of species S .

Figure 1(d) shows the species abundance distribution for the RF model. Conditioning on N has no impact on the abundance distribution (blue circle curve and red square curve coincide), in agreement with our self-consistent approach. The approximate abundance distributions correspond very well with the simulated one (green triangles), as expected because the approximation coincides with the exact solution (see Fig. 2).

VIII. INTERPOLATING PM AND RF

Abundance distributions such as Eqs. (31) and (43) can be used to infer information about the speciation process active in the ecological community. The question whether field data

support PM or RF is most conveniently evaluated using a generalized model that incorporates both speciation modes with a parameter controlling the relative importance of the two speciation modes. In this section we introduce and solve such a generalized speciation model.

To do so, we consider a fragment size distribution $s^{(k)}$ that is a convex combination of PM and RF fragment size distributions,

$$s^{(k)}(\ell) = (1 - \alpha)\delta_1(\ell) + \alpha\frac{1}{k-1}.$$

The weight parameter α interpolates between PM ($\alpha=0$) and RF ($\alpha=1$). Comparing the immigrant abundance distributions \bar{q} , Eq. (27) for PM and Eq. (36) for RF, suggests the following ansatz for the combined model:

$$q_\ell^{\text{ass}} = (1 - \alpha)\delta_1(\ell) + \alpha(1 - \rho)\rho^{\ell-1}. \quad (44)$$

Because $\mathbb{E}S_k$ and $\mathbb{E}N$ are linearly dependent on \bar{q} , we get

$$\begin{aligned} \mathbb{E}S_k &= \theta(1 - \alpha + \alpha k)\frac{\rho^k}{k}, \\ \mathbb{E}N &= \rho\theta\frac{(1 - \alpha)(1 - \rho) + \alpha}{(1 - \rho)^2}. \end{aligned} \quad (45)$$

Hence, the computed immigrant abundance distribution (25) is

$$\begin{aligned} q_\ell^{\text{cmp}} &= \left((1 - \alpha) + \frac{\alpha(1 - \rho)^2}{(1 - \alpha)(1 - \rho) + \alpha} \right) \delta_1(\ell) \\ &+ \frac{\alpha(1 - \rho)^2}{(1 - \alpha)(1 - \rho) + \alpha} \sum_{k>\ell} \rho^{k-1} \frac{1 + \alpha(k-1)}{k-1}. \end{aligned} \quad (46)$$

Figure 3(b) compares the self-consistency distributions \bar{q}^{ass} and \bar{q}^{cmp} [see Eqs. (44) and (46)]. The agreement between both distributions is excellent for $\rho \approx 1$, which is equivalent to large N . Thus, an appropriate combination of the PM and RF solutions yields the solution of the interpolating model. Note that this is a nontrivial result, because the self-consistency equations (26) are nonlinear.

The self-consistent approximation (15) of the abundance distribution for the interpolating model reads

$$P(\vec{S}) = \frac{\rho^N}{Z_2} \prod_{k \geq 1} \frac{[\theta(1 - \alpha + \alpha k)]^{S_k}}{S_k! k^{S_k}},$$

with normalization constant [see Eq. (16)]

$$\ln Z_2 = \theta \sum_{k \geq 1} \frac{1 - \alpha + \alpha k}{k} \rho^k = -(1 - \alpha)\theta \ln(1 - \rho) + \alpha\theta \frac{\rho}{1 - \rho}.$$

Hence,

$$P(\vec{S}) = (1 - \rho)^{(1-\alpha)\theta} e^{-\alpha\theta\rho/(1-\rho)} \rho^N \prod_{k \geq 1} \frac{[\theta(1 - \alpha + \alpha k)]^{S_k}}{S_k! k^{S_k}}. \quad (47)$$

We still have to impose the fixed community size of the interpolating model. To do so, we determine ρ such that $\mathbb{E}N$

for distribution (47) equals the community size constraint N . We assume large N or $\rho \approx 1$, so that Eq. (45) becomes

$$\frac{N}{\theta} \approx \frac{1 - \alpha}{1 - \rho} + \frac{\alpha}{(1 - \rho)^2}.$$

This is a quadratic equation in ρ , which can be inverted to give the value of ρ . Together with Eq. (33) it allows us to express distribution (47) in terms of the parameters μ , ν , α , and N of the interpolating model. As in the PM and RF models, the community size constraint can also be imposed by conditioning on N . The conditional abundance distribution (19) can be computed using the fast Fourier transform.

Figures 1(b) and 1(c) illustrate how the species abundance distributions of the interpolating model lie in between the solution for the PM model [Fig. 1(a)] and the RF model [Fig. 1(b)]. Again, the abundance distributions are not noticeably modified by conditioning on the number of individuals N (blue circle curve and red square curve coincide) and agree nicely with the simulated distributions (green triangles).

IX. DISCUSSION

We have introduced a self-consistent approximation scheme to obtain stationary abundance distributions in neutral community theory with various speciation processes. Except for the most rudimentary speciation models, the inflow of species in the community is determined by the community composition. This feedback of the speciation process on community structure complicates the model. Our approximation is based on cutting the feedback loop and matching self-consistently the abundance distribution of immigrating species. This yields explicit expressions for the stationary species abundance distributions, which agree well with direct simulation results.

The study of the regional community model constitutes, first, a crucial step to compare model predictions with empirical data. However, more theoretical work is needed before data comparison can be carried out. Indeed, species abundance data are rarely available for the entire regional-scale community. Rather, data are usually available for one or more small and spatially localized samples of individuals taken from the regional community [24]. This sampling process can be modeled as one or more local communities receiving immigrants from the much larger regional community. The derivation of the species abundance distribution for the local communities constitutes, next, a nontrivial step. This derivation, together with a proper data comparison, has been carried out extensively for the PM model [24,29,30] and is currently performed for the RF model [25].

We have shown that the self-consistency problem can be formulated as a nonlinear self-consistency equation (26). This equation is rarely exactly solvable (except for very simple speciation models, such as the PM model). Numerical techniques could be used to solve this equation directly. Alternatively, one can solve the autonomous equations for the expected numbers $\mathbb{E}S_k$ and construct an ansatz for the immigrant abundance distribution \bar{q} . We have used this procedure to solve the RF model. One could also simulate the stochastic dynamics of the community and keep track of the specia-

tion events. Reconstructing the distribution of the immigrating species abundance could lead to a reasonable guess for \bar{q} . For the interpolating model, we were able to obtain an ansatz by combining the PM and RF solutions.

Our approximation scheme yields excellent results for different speciation models. It would be interesting to test the versatility of our approach on a wider range of problems. For example, the speciation models we studied in this paper have a constant speciation rate per individual. Other speciation models consider that the speciation rate is constant per spe-

cies [16]. In the latter models, as the number of species fluctuates, the community-level speciation rate varies over time, which might complicate the analysis. Also, we restricted our attention to approximating the stationary abundance distribution of nonspatial speciation models. It remains to be investigated whether our approach can be extended to study dynamical community properties [28], spatial speciation models [31], or other community structure characteristics such as phylogenetic relatedness between species [32].

-
- [1] G. Bell, *Science* **293**, 2413 (2001).
 [2] S. P. Hubbell, *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton University Press, Princeton, NJ, 2001).
 [3] D. Alonso, R. S. Etienne, and A. J. McKane, *Trends Ecol. Evol.* **21**, 451 (2006).
 [4] R. S. Etienne and D. Alonso, *J. Stat. Phys.* **128**, 485 (2007).
 [5] P. A. Abrams, *Nature (London)* **412**, 858 (2001).
 [6] J. H. Brown, *Evolution (Lawrence, Kans.)* **55**, 2137 (2001).
 [7] C. de Mazancourt, *Science* **293**, 1772 (2001).
 [8] J. Bengtsson, *Ecologic. Econ.* **42**, 497 (2002).
 [9] J. S. Clark and J. S. McLachlan, *Nature (London)* **423**, 635 (2003).
 [10] J. Fargione, C. S. Brown, and D. Tilman, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 8916 (2003).
 [11] I. Volkov *et al.*, *Nature (London)* **450**, 45 (2007).
 [12] G. E. Hutchinson, *Am. Nat.* **95**, 137 (1961).
 [13] R. D. Holt, *Trends Ecol. Evol.* **21**, 531 (2006).
 [14] S. P. Hubbell, *Ecology* **87**, 1387 (2006).
 [15] M. Scheffer and E. H. van Nes, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 6230 (2006).
 [16] R. S. Etienne *et al.*, *Oikos* **116**, 241 (2007).
 [17] M. Vallade and B. Houchmandzadeh, *Phys. Rev. E* **68**, 061902 (2003).
 [18] I. Volkov *et al.*, *Nature (London)* **424**, 1035 (2003).
 [19] R. E. Ricklefs, *Oikos* **100**, 185 (2003).
 [20] M. Vallade and B. Houchmandzadeh, *Phys. Rev. E* **74**, 051914 (2006).
 [21] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, Amsterdam, 1992).
 [22] B. Haegeman and R. S. Etienne, *Bull. Math. Biol.* **71**, 1507 (2009).
 [23] R. S. Etienne, D. Alonso, and A. J. McKane, *J. Theor. Biol.* **248**, 522 (2007).
 [24] R. S. Etienne, *Ecol. Lett.* **8**, 253 (2005).
 [25] R. S. Etienne and B. Haegeman, *Theor. Ecol.* (to be published).
 [26] B. Haegeman and R. S. Etienne, *J. Theor. Biol.* **252**, 288 (2008).
 [27] W. J. Ewens, *Theor. Popul. Biol.* **3**, 87 (1972).
 [28] D. Vanpeteghem, O. Zemb, and B. Haegeman, *Math. Biosci.* **212**, 88 (2008).
 [29] R. S. Etienne, *Ecol. Lett.* **10**, 608 (2007).
 [30] R. S. Etienne, *J. Theor. Biol.* **257**, 510 (2009).
 [31] J. Rosindell and S. J. Cornell, *Ecol. Lett.* **10**, 586 (2007).
 [32] F. Jabot and J. Chave, *Ecol. Lett.* **12**, 239 (2009).