

Dynamics of neutral biodiversity

Dimitri Vanpeteghem^a, Olivier Zemb^{a,b}, Bart Haegeman^{a,c,*}

^aINRA, UR50, Laboratoire de Biotechnologie de l'Environnement, Avenue des Etangs, F-11100 Narbonne, France

^bBiotechnology and Biomolecular Sciences, CMB, University of New South Wales, Sydney, New South Wales 2052, Australia

^cINRA-INRIA Research Team MERE, UMR Analyse des Systèmes et Biométrie, 2 Place Pierre Viala, F-34060 Montpellier, France

Received 25 April 2007; accepted 8 January 2008

Available online 17 January 2008

Abstract

Hubbell's neutral model has become a major paradigm in ecology. Whereas the steady-state structure is well understood, results about the dynamical aspects of the model are scarce. Here we derive dynamical equations for the Simpson diversity index. Both mean and variance of the diversity are proven to satisfy stable linear system dynamics. We show that in the stationary limit we indeed recover previous results, and we supplement this with numerical simulations to validate the dynamical part of our analytical computations. These findings are especially relevant for experiments in microbial ecology, where the Simpson diversity index can be accurately measured as a function of time.

© 2008 Elsevier Inc. All rights reserved.

Keywords: Neutral community model; Simpson diversity index; Biodiversity dynamics; Microbial ecology

1. Introduction

Neutral community theory as proposed by Hubbell [1] follows a stochastic approach to model ecological assemblages. It describes the species abundance dynamics of a local community in contact with a much larger regional community. When an individual in the local community dies, it is replaced by the offspring of another local individual, or by an immigrant of the regional community. The number of individuals in the local community remains therefore constant over time, which is called the zero-sum assumption. The neutrality assumption, on the other hand, states that all individuals, regardless of the species they belong to, behave identically under identical circumstances.

Although these assumptions, and neutrality in particular, are outrageous from a biological viewpoint, the model predicts stationary species abundance distributions remarkably close to those observed in nature [2–4]. The

neutral steady-state fits experimental data as accurately as most popular ecological distributions, e.g. the lognormal one [3,5]. However, neutral theory has the advantage to provide a dynamical framework, with a clear interpretation of the model parameters.

The dynamics of the neutral model have been considered in a limited number of studies. When the number of individuals in the local community is large, a continuous approximation is often justified [6]. This approach has been used to compute the species extinction-time distribution [7], or to study some dynamical aspects at or close to the stationary state [8]. In most of the papers dealing with neutral dynamics however, the model equations are used exclusively to study stationary properties.

Although biodiversity is a central notion in the neutral theory, its dynamics have not been investigated as such. It has been noted that the Simpson diversity index [9] enters the neutral theory in a natural way. Indeed, the average Simpson diversity in the regional community stationary state is directly related to the so-called fundamental biodiversity parameter [1]. Steady-state fluctuations have been computed [10,11] and also a dynamical equation for the average Simpson diversity in the regional community was reported [1,5].

* Corresponding author. Address: INRA, UR050, Laboratoire de Biotechnologie de l'Environnement, Avenue des Etangs, F-11100 Narbonne, France. Tel.: +33 (0) 468 425 161; fax: +33 (0) 468 425 160.

E-mail address: bart.haegeman@inria.fr (B. Haegeman).

In the present paper, we attempt a rigorous derivation of the biodiversity dynamics for the full neutral model, i.e. without a continuous approximation. We posit the dynamical description of Hubbell’s neutral model for the local community with a given species abundance distribution for the regional community. It is a master equation for the probability distribution on the abundances of all species. We then derive dynamical equations for the average Simpson diversity index and its fluctuations. Comparison with steady-state values and numerical simulations both validate our results. Finally, we argue how this work provides a link between neutral theory and microbial ecology.

2. Neutral model

Neutral community theory starts out by separating the local from the regional community. On the timescale of the local community, which is the one of interest to us, the regional community does not evolve. It consists of N species each with a fixed abundance. The regional community is assumed to be so large that only relative abundances are of importance. We denote the relative abundance of species k in the regional community by p_k , for $k = 1, 2, \dots, N$. We also introduce the relative abundance vector

$$\vec{p} = [p_1 \ p_2 \ \dots \ p_N].$$

Note that $\sum_k p_k = 1$.

The local community consists of X individuals, all belonging to one of the N species present in the regional community. We denote the absolute abundance of species k in the local community by X_k , an integer possibly zero. The absolute abundance vector

$$\vec{X} = [X_1 \ X_2 \ \dots \ X_N]$$

sums up to X , thus $\sum_k X_k = X$.

The dynamics in the local community is triggered by death events. The mortality rate is denoted by μ . Thus, the probability that in the interval $[t, t + \epsilon]$ one of the X individuals dies, is given by $\mu\epsilon + o(\epsilon)$ as $\epsilon \rightarrow 0$. In that case, the death is immediately compensated, either by the immigration of a new individual from the regional community (with probability m), or by the reproduction of some other individual in the local community (with probability $1 - m$). There are no other events in the local community apart from the two described: death followed by immigration and death followed by reproduction. As a consequence, the number of individuals in the local community remains constant, equal to X . This is the zero-sum assumption.

This formulation of Hubbell’s model leads to a continuous-time Markov process. The rate for a transition that decreases the abundance of species i by one, and increases the abundance of species $j \neq i$ by one. This rate is given by

$$R(X_i; p_j, X_j) = \mu \frac{X_i}{X} \left(m p_j + (1 - m) \frac{X_j}{X - 1} \right).$$

The probability that an individual of species i dies, is proportional to its abundance X_i . If the dead individual is re-

placed by an individual of the regional community, the probability that this new individual belongs to species j , equals the regional abundance p_j . Otherwise, if the dead individual is replaced by the offspring of an individual in the local community, the probability that this new individual belongs to species j , is proportional to its abundance X_j . As all these probabilities are simply proportional to the species abundance in the local or regional community, no differences are assumed between individuals of different species. This is the neutrality assumption.

By summing over all possible events,

$$\sum_{\substack{i,j \\ i \neq j}} R(X_i; p_j, X_j) + \sum_i R(X_i; p_i, X_i - 1) = \mu,$$

we retrieve the mortality rate μ . The second term in the left-hand side corresponds to events where the species of the replacing individual is the same as that of the dead individual. Such events do not change the abundance vector \vec{X} .

As our model is stochastic, we need some notation to deal with randomness. Bold symbols are used for random variables. For instance, the random variable corresponding to the abundance of species k is denoted by \mathbf{X}_k , and the vector containing the N random species abundances by $\vec{\mathbf{X}}$. To simplify notation, we do not distinguish consistently a random variable from its realisation. For instance, we denote the probability that $\vec{\mathbf{X}}$ takes the value \vec{X} by $\mathbb{P}(\vec{X})$. Marginal probability distributions are denoted by superscripts. For instance, $\mathbb{P}^i(X_i)$ stands for the probability distribution restricted to species i . Similarly, we use $\mathbb{P}^{ij}(X_i, X_j)$ and $\mathbb{P}^{ijk}(X_i, X_j, X_k)$ for the bivariate and trivariate distributions.

The master equation for \vec{X} reads

$$\begin{aligned} \frac{d}{dt} \mathbb{P}(\vec{X}) &= \sum_{\substack{i,j \\ i \neq j}} R(X_i + 1; p_j, X_j - 1) \mathbb{P}(\vec{X} + \vec{e}_i - \vec{e}_j) \\ &\quad - \sum_{\substack{i,j \\ i \neq j}} R(X_i; p_j, X_j) \mathbb{P}(\vec{X}), \end{aligned} \tag{1}$$

where we used basis vectors \vec{e}_i with components $(\vec{e}_i)_j = \delta_{ij}$, the Kronecker delta. The special structure of the neutral model implies the existence of an autonomous master equation for the abundance X_k ,

$$\begin{aligned} \frac{d}{dt} \mathbb{P}^k(X_k) &= R(X_k + 1; 1 - p_k, X - X_k - 1) \mathbb{P}^k(X_k + 1) \\ &\quad + R(X - X_k + 1; p_k, X_k - 1) \mathbb{P}^k(X_k - 1) \\ &\quad - R(X_k; 1 - p_k, X - X_k) \mathbb{P}^k(X_k) \\ &\quad - R(X - X_k; p_k, X_k) \mathbb{P}^k(X_k). \end{aligned} \tag{2}$$

See [Appendix A](#) for the derivation. Most papers take (2) as the starting point of their analysis. However, these equations do not form a complete description of the model. For most computations, e.g. the stationary distribution or the mean Simpson diversity, the marginals on one variable are sufficient. It will turn out that computing the var-

iance of the Simpson diversity requires also the bivariate distributions.

3. Simpson diversity: analytics

The most obvious way to quantify diversity is the number of species present in the community. More elaborate diversity notions have been proposed, taking species abundances into account. The Simpson concentration index [9] is defined as the probability that two individuals drawn randomly from the community belong to the same species. If the individuals are drawn with replacement, this leads to

$$Z_1 = \frac{1}{X^2} \sum_{k=1}^N X_k^2.$$

Without replacement, one obtains

$$Z_2 = \frac{1}{X(X-1)} \sum_{k=1}^N X_k(X_k - 1).$$

If the number of individuals X is large, as is usually the case, $Z_1 \approx Z_2$.

The larger the Simpson concentration, the smaller the diversity in the community. It is therefore customary to apply a decreasing transformation to the concentration index, to get diversity indices like $1 - Z_1$, $1/Z_1$ or $-\ln Z_1$. In the present context, the Simpson concentration is a random variable \mathbf{Z}_1 or \mathbf{Z}_2 . We derive the dynamics of its mean and variance.

3.1. Mean Simpson diversity

Under neutral dynamics, the equation for the expected value of some function $f(\mathbf{X}_k)$ of the abundance \mathbf{X}_k reads,

$$\begin{aligned} \frac{d}{dt} \mathbb{E}[f(\mathbf{X}_k)] &= \mathbb{E}[(f(\mathbf{X}_k - 1) - f(\mathbf{X}_k))R(\mathbf{X}_k; 1 - p_k, X - \mathbf{X}_k)] \\ &\quad + \mathbb{E}[(f(\mathbf{X}_k + 1) - f(\mathbf{X}_k))R(X - \mathbf{X}_k; p_k, \mathbf{X}_k)]. \end{aligned} \quad (3)$$

See Appendix B for the derivation. Note that this equation only depends on \mathbf{X}_k and not on the other abundances.

Eq. (3) allows to compute the dynamics of the momenta $\mathbb{E}X_k^a$, see Appendix C and Eqs. (14)–(17). Note that the dynamical equation for $\mathbb{E}X_k^a$ does not contain powers of order higher than a . This leads to autonomous systems of equations describing the dynamics of these momenta. For an alternative derivation using characteristic functions, see Appendix D.

The dynamics of the mean Simpson concentration can be obtained from (14) and (15). Indeed,

$$\mathbb{E}\mathbf{Z}_1 = \frac{1}{X^2} C_2^0 \quad \text{and} \quad \mathbb{E}\mathbf{Z}_2 = \frac{1}{X(X-1)} (C_2^0 - X), \quad (4)$$

where we introduced the notation

$$C_a^z = \sum_k p_k^z \mathbb{E}X_k^a.$$

Summing Eq. (15) over all k , we obtain

$$\begin{aligned} \frac{d}{dt} C_2^0 &= -2\mu \left(\frac{m}{X} + \frac{1-m}{X(X-1)} \right) C_2^0 + 2\mu \left(m - \frac{m}{X} \right) C_1^1 \\ &\quad + 2\mu \left(m + \frac{(1-m)X}{X-1} \right). \end{aligned} \quad (5)$$

This equation is not autonomous because C_1^1 appears in it. Multiplying (14) by p_k and summing over all k , we get

$$\frac{d}{dt} C_1^1 = -\mu \frac{m}{X} C_1^1 + \mu m C_0^2. \quad (6)$$

Eqs. (5) and (6) form an autonomous dynamical system. The dynamics are linear with eigenvalues

$$-\mu \frac{m}{X} \quad \text{and} \quad -2\mu \left(\frac{m}{X} + \frac{1-m}{X(X-1)} \right).$$

They are strictly negative, establishing exponential stability. One can go a step further and combine (5) and (6) to obtain a second order differential equation in C_2^0 alone.

3.2. Variance of Simpson diversity

The dynamics of the mean Simpson diversity has been studied previously [1,5]. It should not be overlooked however, that the mean of a stochastic variable alone does not have much meaning. For all we know, the variance might be so large as to make the mean virtually useless. Even worse, the variance might go to infinity as time increases. In this section we ascertain that the variance of the Simpson concentration \mathbf{Z}_1 or \mathbf{Z}_2 does not blow up.

To compute the variance, we need the dynamical equation for the expected value of $f(\mathbf{X}_k, \mathbf{X}_l)$,

$$\begin{aligned} \frac{d}{dt} \mathbb{E}[f(\mathbf{X}_k, \mathbf{X}_l)] &= \mathbb{E}[(f(\mathbf{X}_k - 1, \mathbf{X}_l) - f(\mathbf{X}_k, \mathbf{X}_l)) \\ &\quad \times R(\mathbf{X}_k; 1 - p_k - p_l, X - \mathbf{X}_k - \mathbf{X}_l)] \\ &\quad + \mathbb{E}[(f(\mathbf{X}_k + 1, \mathbf{X}_l) - f(\mathbf{X}_k, \mathbf{X}_l)) \\ &\quad \times R(X - \mathbf{X}_k - \mathbf{X}_l; p_k, \mathbf{X}_k)] \\ &\quad + \mathbb{E}[(f(\mathbf{X}_k, \mathbf{X}_l - 1) - f(\mathbf{X}_k, \mathbf{X}_l)) \\ &\quad \times R(\mathbf{X}_l; 1 - p_k - p_l, X - \mathbf{X}_k - \mathbf{X}_l)] \\ &\quad + \mathbb{E}[(f(\mathbf{X}_k, \mathbf{X}_l + 1) - f(\mathbf{X}_k, \mathbf{X}_l)) \\ &\quad \times R(X - \mathbf{X}_k - \mathbf{X}_l; p_l, \mathbf{X}_l)] \\ &\quad + \mathbb{E}[(f(\mathbf{X}_k - 1, \mathbf{X}_l + 1) - f(\mathbf{X}_k, \mathbf{X}_l))R(\mathbf{X}_k; p_l, \mathbf{X}_l)] \\ &\quad + \mathbb{E}[(f(\mathbf{X}_k + 1, \mathbf{X}_l - 1) - f(\mathbf{X}_k, \mathbf{X}_l))R(\mathbf{X}_l; p_k, \mathbf{X}_k)] \end{aligned} \quad (7)$$

The derivation of this formula goes along the lines of the derivation of formula (3).

Eq. (7) allows to derive the dynamical equations for the momenta $\mathbb{E}X_k^a X_l^b$, see Appendix C and Eqs. (18)–(20). Again, the dynamics of $\mathbb{E}X_k^a X_l^b$ are expressed in terms of lower order momenta, leading to autonomous systems of equations. Characteristic functions can be used to obtain the same equations, see Appendix D.

The variance of the Simpson concentration is

$$\text{Var } \mathbf{Z}_1 = \frac{C_4^0 + C_{22}^{00} - (C_2^0)^2}{X^4} \quad \text{and}$$

$$\text{Var } \mathbf{Z}_2 = \frac{C_4^0 + C_{22}^{00} - (C_2^0)^2}{X^2(X-1)^2}. \quad (8)$$

Here we used the notation

$$C_{ab}^{\alpha\beta} = \sum_{\substack{k,l \\ k \neq l}} p_k^\alpha p_l^\beta \mathbb{E} X_k^\alpha X_l^\beta.$$

Note that $C_{ab}^{\alpha\beta} = C_{ba}^{\beta\alpha}$.

The dynamics of the variance can thus be computed from the dynamics of C_2^0 , C_4^0 and C_{22}^{00} . Dynamical equations for these quantities can be derived by combining Eqs. (14)–(17) and Eqs. (18)–(20).

The dynamics of C_4^0 is a 9th order linear system. In terms of the vector

$$v_1 = [C_4^0 \quad C_3^1 \quad C_3^0 \quad C_2^2 \quad C_2^1 \quad C_2^0 \quad C_1^3 \quad C_1^2 \quad C_1^1],$$

it reads

$$\dot{v}_1 = A_1 v_1 + w_1, \quad (9)$$

where A_1 is the 9-dimensional matrix,

$$A_1 = \begin{bmatrix} 4a_3 & 4b_3 & 6c_2 & 0 & 6b_0 & 4a_{1/2} & 0 & 0 & 4b_{1/2} \\ 0 & 3a_2 & 0 & 3b_2 & 3c_2 & 0 & 0 & 3b_0 & a_0 \\ 0 & 0 & 3a_2 & 0 & 3b_2 & 3c_2 & 0 & 0 & 3b_0 \\ 0 & 0 & 0 & 2a_1 & 0 & 0 & 2b_1 & c_2 & 0 \\ 0 & 0 & 0 & 0 & 2a_1 & 0 & 0 & 2b_1 & c_2 \\ 0 & 0 & 0 & 0 & 0 & 2a_1 & 0 & 0 & 2b_1 \\ 0 & 0 & 0 & 0 & 0 & 0 & a_0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_0 \end{bmatrix},$$

and w_1 is the 9-dimensional vector,

$$w_1 = [2Xc_1 \quad \mu m C_0^2 \quad 0 \quad \mu m C_0^3 \quad \mu m C_0^2 \quad 2Xc_1 \quad \mu m C_0^4 \quad \mu m C_0^3 \quad \mu m C_0^2],$$

with

$$a_x = -\frac{\mu m}{X} - \alpha \frac{\mu(1-m)}{X(X-1)}$$

$$b_x = \mu m - \alpha \frac{\mu m}{X}$$

$$c_x = \frac{\mu m}{X} + \alpha \frac{\mu(1-m)}{X-1}.$$

The eigenvalues of the system dynamics are $4a_3$, $3a_2$ (with degeneracy 2), $2a_1$ (with degeneracy 3) and a_0 (with degeneracy 3). They are all strictly negative, implying exponential stability.

The dynamics of C_{22}^{00} is a 12th order linear system. In terms of the vector

$$v_2 = [C_{22}^{00} \quad C_{21}^{01} \quad C_{21}^{00} \quad C_{20}^{02} \quad C_{20}^{01} \quad C_{11}^{11} \quad C_{11}^{10} \quad C_{11}^{00} \quad C_{10}^{12} \quad C_{10}^{11} \quad C_{10}^{02} \quad C_{10}^{01}],$$

it reads

$$\dot{v}_2 = A_2 v_2 + w_2, \quad (10)$$

where A_2 is the 12-dimensional matrix,

$$A_2 = \begin{bmatrix} 4a_3 & 4b_3 & 2c_2 & 0 & 2b_2 & 0 & 4c_0 & 2d & 0 & 0 & 0 & 2c_0 \\ 0 & 3a_2 & 0 & b_2 & 0 & 2b_2 & c_2 & 0 & 0 & b_1 & c_0 & 0 \\ 0 & 0 & 3a_2 & 0 & b_2 & 0 & 2b_2 & c_2 & 0 & 0 & 0 & b_0 \\ 0 & 0 & 0 & 2a_1 & 0 & 0 & 0 & 0 & 2b_1 & 0 & c_2 & 0 \\ 0 & 0 & 0 & 0 & 2a_1 & 0 & 0 & 0 & 0 & 2b_1 & 0 & c_2 \\ 0 & 0 & 0 & 0 & 0 & 2a_1 & 0 & 0 & 2b_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2a_1 & 0 & 0 & b_1 & b_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2a_1 & 0 & 0 & 0 & 2b_1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_0 \end{bmatrix},$$

and w_2 the 12-dimensional vector,

$$w_2 = [0 \quad 0 \quad 0 \quad \mu m C_{00}^{21} \quad \mu m C_{00}^{11} \quad 0 \quad 0 \quad 0 \quad \mu m C_{00}^{22} \quad \mu m C_{00}^{21} \quad \mu m C_{00}^{21} \quad \mu m C_{00}^{11}],$$

with

$$d = \frac{\mu(1-m)}{X(X-1)}.$$

The eigenvalues of the system dynamics are $4a_3$, $3a_2$ (degeneracy 2), $2a_1$ (degeneracy 5) and a_0 (degeneracy 4). They are all strictly negative, proving exponential stability.

As a result, we have obtained dynamical equations for C_2^0 (Eqs. (5) and (6)), for C_4^0 (Eq. (9)) and for C_{22}^{00} (Eq. (10)). They all satisfy stable linear dynamics, implying that the variance for the Simpson diversity (8) will reach a finite stationary value.

4. Simpson diversity: numerics

The analytical results obtained in the previous section are validated in two different ways. First, we simulate the full stochastic model, and compare simulated trajectories with predicted mean and variance. Next, steady-state values of our dynamical equations are checked against previously published formulas.

4.1. Simulations

As long as the number of individuals X is not too large, the neutral model can be simulated directly. We generated trajectories for a system with $X = 10^4$ individuals and $N = 10^2$ species. The abundance p_k in the regional community were taken proportional to $1/k$. The immigration probability was $m = 5 \cdot 10^{-3}$, and the mortality rate, which fixes the timescale, $\mu = 1$. From the generated abundance vectors \vec{X} , the Simpson concentration Z_1 was computed. For the same parameters, we integrated the dynamical equations for C_2^0 , C_4^0 and C_{22}^{00} , see Eqs. (5), (6), (9) and

(10). The solutions were then combined in Eqs. (4) and (8) to obtain $\mathbb{E}Z_1$ and $\text{Var} Z_1$.

Fig. 1 compares the simulated Simpson diversity $1 - Z_1$ with our analytical computations. The left part compares one randomly generated trajectory with three reference curves:

$$\begin{aligned} \mathbb{E}[1 - Z_1] &= 1 - \mathbb{E}Z_1, \\ \mathbb{E}[1 - Z_1] \pm \sqrt{\text{Var}[1 - Z_1]} &= 1 - \mathbb{E}Z_1 \pm \sqrt{\text{Var} Z_1}. \end{aligned}$$

In the right part, the same three reference curves were estimated based on 100 simulated trajectories. The agreement between analytical and numerical computations is excellent.

4.2. Steady-state

The stationary composition of the local community under neutral dynamics is explicitly known [12]. Given the relative abundance vector \vec{p} for the regional community, the probability distribution for absolute abundance vector \vec{X} for the local community is

$$\lim_{t \rightarrow \infty} \mathbb{P}(\vec{X}) = \binom{X}{X_1 \dots X_N} \frac{\prod_{k=1}^N (I p_k)_{X_k}}{(I)_X}, \quad (11)$$

with

$$\begin{aligned} (a)_n &= a(a+1) \dots (a+n-1) \quad \text{and} \\ I &= \frac{m}{1-m} (X-1). \end{aligned}$$

This distribution allows to compute the stationary value of $\mathbb{E}X_k$,

$$\lim_{t \rightarrow \infty} \mathbb{E}X_k = X p_k, \quad (12)$$

which is identical to the equilibrium of (14). Similarly, the stationary value of $\mathbb{E}X_k^2$ from the distribution (11) is

$$\begin{aligned} \lim_{t \rightarrow \infty} \mathbb{E}X_k^2 &= X(X-1)p_k \frac{I p_k + 1}{I + 1} + X p_k \\ &= \frac{mX(X-1)^2}{mX+1-2m} p_k^2 + \frac{X(X-m)}{mX+1-2m} p_k, \end{aligned} \quad (13)$$

which is identical to the equilibrium of (15). This procedure can be continued to check the other stationary moments of Eqs. (16)–(20).

Combining (12) and (13),

$$\text{Var} X_k = \mathbb{E}X_k^2 - (\mathbb{E}X_k)^2 = p_k(1-p_k) \frac{X(X+I)}{I+1},$$

which shows that the fluctuations decrease monotonically with I and thus with m . Indeed, when the local community is strongly isolated from the regional community, the species abundance can fluctuate wildly. It is easy to check that

$$\text{Var} \frac{X_k}{X} = p_k(1-p_k) \quad \text{when } m \rightarrow 0,$$

i.e. the relative abundance in the local community behaves like a Bernoulli random variable. With strong immigration from the regional community, the fluctuations are

$$\text{Var} \frac{X_k}{X} = \frac{p_k(1-p_k)}{X} \quad \text{when } m \rightarrow 1,$$

i.e. the relative abundance in the local community becomes sharply peaked for large X .

Eq. (13) also allows to compute the stationary value for the Simpson concentration Z_1 ,

$$\lim_{t \rightarrow \infty} \mathbb{E}Z_1 = \frac{m(X-1)^2}{X(mX+1-2m)} \sum_k p_k^2 + \frac{X-m}{X(mX+1-2m)}.$$

This agrees with the known stationary value for the Simpson concentration Z_2 [13],

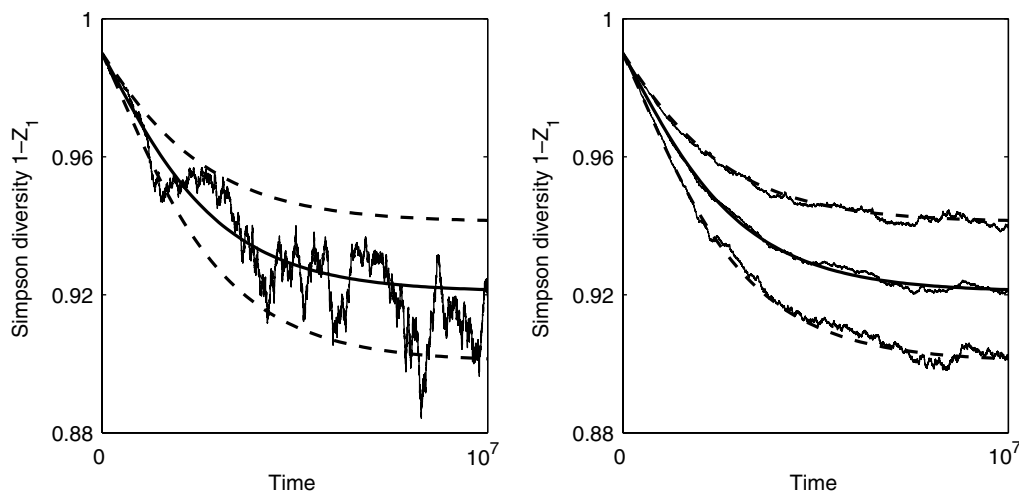


Fig. 1. Dynamics of Simpson diversity. Left: Simulated trajectory compared with analytical predictions of mean and standard deviation of Simpson diversity $1 - Z_1$. Right: Mean and standard deviation of Simpson diversity $1 - Z_1$ estimated from 100 simulated trajectories. The parameters used are $X = 10^4$, $N = 100$, $m = 5 \cdot 10^{-3}$, $\mu = 1$ and regional species abundances p_k are proportional to $1/k$. The initial condition was taken to be deterministic, with all species having equal abundance. Full line: expected values; dashed line: means \pm standard deviation.

$$\begin{aligned} \lim_{t \rightarrow \infty} \mathbb{E}Z_2 &= 1 - \frac{I}{I+1} \left(1 - \sum_k p_k^2 \right) \\ &= \frac{m(X-1)}{mX+1-2m} \sum_k p_k^2 + \frac{1-m}{mX+1-2m}. \end{aligned}$$

Fig. 2 compares the variability of species abundances and Simpson diversity. We took a regional community with $N = 10^6$ species with abundances p_k proportional to $1/k$. The local community consists of $X = 10^8$ individuals, which is too large to simulate the model directly. The mortality rate is $\mu = 1$, whereas immigration probabilities $m = 10^{-6}$, $m = 10^{-4}$ and $m = 10^{-2}$ were considered. The top row shows three species abundances curves as a function of species index:

$$\mathbb{E} \frac{X_k}{X}, \quad \mathbb{E} \frac{X_k}{X} \pm \sqrt{\text{Var} \frac{X_k}{X}}.$$

The bottom row shows three Simpson diversity curves as a function of time:

$$\mathbb{E}Z_1, \quad \mathbb{E}Z_1 \pm \sqrt{\text{Var} Z_1}.$$

Both sets of curves are shown on a logarithmic scale. As noted before, the variability decreases as the immigration probability increases. The variability for the Simpson diversity is systematically smaller than for the species abundances. Indeed, as all species contribute to the Simpson diversity, species abundance variabilities are averaged out.

5. Discussion

We have derived the dynamics of the biodiversity for a neutral local community. The biodiversity was quantified via the Simpson diversity index, which compared to other diversity measures is relatively easy to deal with analytically. Indeed, it has been noted previously [1,5,11] that the Simpson diversity is somehow compatible with the neutral theory. We also computed the variance on the expected dynamics of the Simpson diversity. In particular, for a large local community that has sufficient contact with the regional community (m not too small), the diversity fluctuations were found to be small.

One can expect that the larger the local community, the higher the ratio of dead individuals replaced by local offspring than by regional immigration, and thus the smaller the immigration probability m . For a given experimental system, delimiting an appropriate local community for theoretical analysis, and so fixing parameters X , m and μ , might be a delicate issue. To fix ideas, parameter values used in Fig. 1 could correspond to a community of macro-organisms, the tropical tree forest being the standard example in neutral theory. The parameters used in Fig. 2 rather suggest a microbial community.

To measure the diversity of a forest, one has to collect species data of individual trees, and use this to estimate, e.g. the Simpson diversity index. Linking experiment and theory proceeds therefore most easily via species abun-

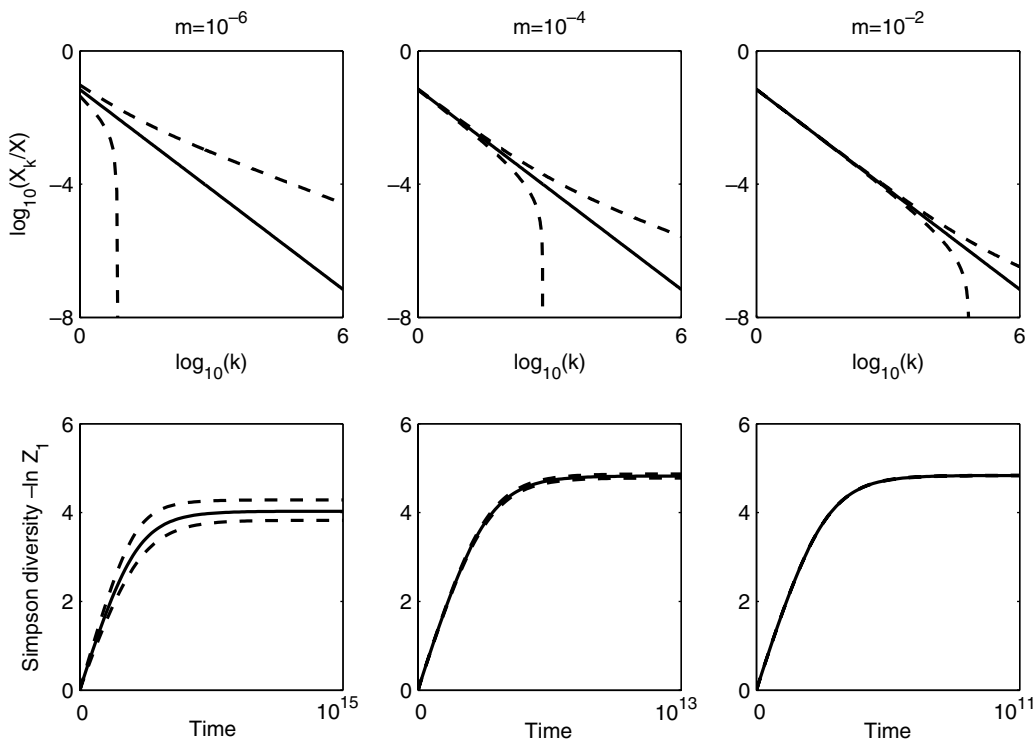


Fig. 2. Variability of species abundances and Simpson diversity. Top: Stationary species abundances in local community, together with their standard deviation, as a function of species index. Bottom: Mean and standard deviation of Simpson diversity $-\ln Z_1$ as a function of time. The parameters used are $X = 10^8$, $N = 10^6$, $\mu = 1$ and regional species abundances p_k are proportional to $1/k$. The initial condition was taken to be deterministic, with only species $k = 1$ present. Left: immigration probability $m = 10^{-6}$. Middle: $m = 10^{-4}$. Right: $m = 10^{-2}$. Full line: expected values; dashed line: means \pm standard deviation.

dance data, rather than making a detour via diversity. For microbial systems, however, this detour seems unavoidable. Indeed, due to the huge microbial diversity, the acquisition of accurate species abundance data is very difficult. Even a rough estimation is expensive and time consuming. Fortunately, cheap and fast DNA-based techniques exist that allow to assess diversity directly, without having to analyse individual microbes. Molecular fingerprinting techniques, for example, encode rather accurately the Simpson diversity [14].

More generally, it has been suggested that microbial communities could be more appropriate than traditional field studies to test ecological theories [15,16]. Indeed, microbial microcosms allow to perform ecological experiments during a few days, where other experimental systems would require several years. They occupy a limited space, but still contain billions of individuals and thousands of species, making them ideal for systematic studies. Moreover, molecular fingerprinting techniques allow to rapidly visualise the community, so that dynamics can be followed closely.

The combination of neutral community theory, a simple model with remarkable predictions, and microbial microcosms, a laboratory study of ecological communi-

ties, looks promising. Some work has been reported in this direction [17,18]. We believe that quantitative tests of microbial neutral dynamics will involve Simpson diversity. Our contribution could provide the theoretical framework for this kind of research.

Acknowledgement

It is a pleasure to thank N. Desassis, R.S. Etienne, J. Hamelin and C. Lobry for helpful suggestions. D.V. acknowledges a postdoctoral fellowship from the Environment and Agronomy Department (EA) of the French National Institute for Agricultural Research (INRA).

Appendix A

We show that the marginal distribution for the abundance of one species satisfies an autonomous master equation. This marginal is defined by

$$\mathbb{P}^k(Y_k) = \sum_{\substack{\vec{X} \\ X_k=Y_k}} \mathbb{P}(\vec{X}).$$

Using the master equation of the full model (1),

$$\begin{aligned} & \frac{d}{dt} \mathbb{P}^k(Y_k) \\ &= \sum_{\vec{X}} \frac{d}{dt} \mathbb{P}(\vec{X}) \\ &= \sum_{\substack{\vec{X} \\ X_k=Y_k}} \sum_{\substack{i,j \\ i \neq j}} R(X_i+1; p_j, X_j-1) \mathbb{P}(\vec{X} + \vec{e}_i - \vec{e}_j) - R(X_i; p_j, X_j) \mathbb{P}(\vec{X}) \\ &= \sum_{\substack{i,j \\ i \neq j \\ i \neq k \\ j \neq k}} \left[\sum_{X_i=0}^{X-Y_k-1} \sum_{X_j=1}^{X-X_i-Y_k} R(X_i+1; p_j, X_j-1) \mathbb{P}^{ijk}(X_i+1, X_j-1, Y_k) \right. \\ & \quad \left. - \sum_{X_i=1}^{X-Y_k} \sum_{X_j=0}^{X-X_i-Y_k} R(X_i; p_j, X_j) \mathbb{P}^{ijk}(X_i, X_j, Y_k) \right] \\ & \quad + \sum_{\substack{j \\ j \neq k}} \sum_{X_j=1}^{X-Y_k} R(Y_k+1; p_j, X_j-1) \mathbb{P}^{kj}(Y_k+1, X_j-1) \\ & \quad + \sum_{\substack{i \\ i \neq k}} \sum_{X_i=0}^{X-Y_k} R(X_i+1; p_k, Y_k-1) \mathbb{P}^{ik}(X_i+1, Y_k-1) \\ & \quad - \sum_{\substack{j \\ j \neq k}} \sum_{X_j=0}^{X-Y_k} R(Y_k; p_j, X_j) \mathbb{P}^{kj}(Y_k, X_j) \\ & \quad - \sum_{\substack{i \\ i \neq k}} \sum_{X_i=1}^{X-Y_k} R(X_i; p_k, Y_k) \mathbb{P}^{ik}(X_i, Y_k) \end{aligned}$$

The first and second line cancel. For the third line,

$$\begin{aligned} & \sum_{j \neq k} \sum_{X_j=1}^{X-Y_k} R(Y_k + 1; p_j, X_j - 1) \mathbb{P}^{kj}(Y_k + 1, X_j - 1) \\ &= \sum_{j \neq k} \sum_{X_j=0}^{X-Y_k-1} R(Y_k + 1; p_j, X_j) \mathbb{P}^{kj}(Y_k + 1, X_j) \\ &= \sum_{\vec{X}} \left[\sum_{j \neq k} R(Y_k + 1; p_j, X_j) \right] \mathbb{P}(\vec{X}) \\ &= \sum_{\vec{X}} R(Y_k + 1; 1 - p_k, X - Y_k - 1) \mathbb{P}(\vec{X}) \\ &= R(Y_k + 1; 1 - p_k, X - Y_k - 1) \mathbb{P}^k(Y_k + 1) \end{aligned}$$

The other lines can be computed similarly. As a result,

$$\begin{aligned} \frac{d}{dt} \mathbb{P}^k(Y_k) &= R(Y_k + 1; 1 - p_k, X - Y_k - 1) \mathbb{P}^k(Y_k + 1) \\ &\quad + R(X - Y_k + 1; p_k, Y_k - 1) \mathbb{P}^k(Y_k - 1) \\ &\quad - R(Y_k; 1 - p_k, X - Y_k) \mathbb{P}^k(Y_k) - R(X \\ &\quad - Y_k; p_k, Y_k) \mathbb{P}^k(Y_k) \end{aligned}$$

Appendix B

We prove the dynamical equation for the expected value of $f(\mathbf{X}_k)$. Using the master Eq. (1),

The first and second line cancel. For the other lines,

$$\begin{aligned} & \frac{d}{dt} \mathbb{E}[f(\mathbf{X}_k)] \\ &= \sum_{j \neq k} \sum_{X_k=1}^X \sum_{X_j=0}^{X-X_k} (f(X_k - 1) - f(X_k)) R(X_k; p_j, X_j) \mathbb{P}^{(k,j)}(\vec{X}) \\ &\quad + \sum_{i \neq k} \sum_{X_i=1}^X \sum_{X_k=0}^{X-X_i} (f(X_k + 1) - f(X_k)) R(X_i; p_k, X_k) \mathbb{P}^{(i,k)}(\vec{X}) \\ &= \sum_{j \neq k} \mathbb{E}[(f(\mathbf{X}_k - 1) - f(\mathbf{X}_k)) R(\mathbf{X}_k; p_j, \mathbf{X}_j)] \\ &\quad + \sum_{i \neq k} \mathbb{E}[(f(\mathbf{X}_k + 1) - f(\mathbf{X}_k)) R(\mathbf{X}_i; p_k, \mathbf{X}_k)] \\ &= \mathbb{E}[(f(\mathbf{X}_k - 1) - f(\mathbf{X}_k)) R(\mathbf{X}_k; 1 - p_k, X - \mathbf{X}_k)] \\ &\quad + \mathbb{E}[(f(\mathbf{X}_k + 1) - f(\mathbf{X}_k)) R(X - \mathbf{X}_k; p_k, \mathbf{X}_k)] \end{aligned}$$

This proves formula (3). One can proceed in a similar way to prove formula (7).

Appendix C

We compute the dynamical equations for the momenta of the probability distribution $\mathbb{P}(\vec{X})$. Applying formula (3) for $f(\mathbf{X}_k) = \mathbf{X}_k$, we get

$$\frac{1}{\mu} \frac{d}{dt} \mathbb{E} \mathbf{X}_k = -\frac{m}{X} \mathbb{E} \mathbf{X}_k + m p_k. \tag{14}$$

$$\begin{aligned} & \frac{d}{dt} \mathbb{E}[f(\mathbf{X}_k)] \\ &= \sum_{\vec{X}} f(\mathbf{X}_k) \frac{d}{dt} \mathbb{P}(\vec{X}) \\ &= \sum_{\substack{i,j \\ i \neq j}} \left[\sum_{\vec{X}} f(\mathbf{X}_k) R(X_i + 1; p_j, X_j - 1) \mathbb{P}(\vec{X} + \vec{e}_i - \vec{e}_j) - \sum_{\vec{X}} f(\mathbf{X}_k) R(X_i; p_j, X_j) \mathbb{P}(\vec{X}) \right] \\ &= \sum_{\substack{i,j \\ i \neq j \\ j \neq k}} \left[\sum_{X_i=0}^{X-1} \sum_{X_j=1}^{X-X_i} \sum_{X_k=0}^{X-X_i-X_j} f(\mathbf{X}_k) R(X_i + 1; p_j, X_j - 1) \mathbb{P}^{ijk}(X_i + 1, X_j - 1, X_k) \right. \\ &\quad - \sum_{X_j=1}^X \sum_{X_k=0}^{X-X_j} \sum_{X_i=0}^{X-X_j-X_k} f(\mathbf{X}_k) R(X_i; p_j, X_j) \mathbb{P}^{ijk}(X_i, X_j, X_k) \left. \right] \\ &\quad + \sum_{j \neq k} \left[\sum_{X_k=0}^{X-1} \sum_{X_j=1}^{X-X_k} f(\mathbf{X}_k) R(X_k + 1; p_j, X_j - 1) \mathbb{P}^{kj}(X_k + 1, X_j - 1) \right. \\ &\quad - \sum_{X_k=1}^X \sum_{X_j=0}^{X-X_k} f(\mathbf{X}_k) R(X_k; p_j, X_j) \mathbb{P}^{kj}(X_k, X_j) \left. \right] \\ &\quad + \sum_{i \neq k} \left[\sum_{X_i=0}^{X-1} \sum_{X_k=1}^{X-X_i} f(\mathbf{X}_k) R(X_i + 1; p_k, X_k - 1) \mathbb{P}^{ik}(X_i + 1, X_k - 1) \right. \\ &\quad - \sum_{X_i=1}^X \sum_{X_k=0}^{X-X_i} f(\mathbf{X}_k) R(X_i; p_k, X_k) \mathbb{P}^{ik}(X_i, X_k) \left. \right] \end{aligned}$$

For $f(\mathbf{X}_k) = \mathbf{X}_k^2$,

$$\begin{aligned} \frac{1}{\mu} \frac{d}{dt} \mathbb{E} \mathbf{X}_k^2 &= \left(-\frac{2m}{X} - \frac{2(1-m)}{X(X-1)} \right) \mathbb{E} \mathbf{X}_k^2 \\ &+ \left(2m - \frac{2m}{X} \right) p_k \mathbb{E} \mathbf{X}_k \\ &+ \left(\frac{m}{X} + \frac{2(1-m)}{X-1} \right) \mathbb{E} \mathbf{X}_k + mp_k. \end{aligned} \tag{15}$$

For $f(\mathbf{X}_k) = \mathbf{X}_k^3$,

$$\begin{aligned} \frac{1}{\mu} \frac{d}{dt} \mathbb{E} \mathbf{X}_k^3 &= \left(-\frac{3m}{X} - \frac{6(1-m)}{X(X-1)} \right) \mathbb{E} \mathbf{X}_k^3 \\ &+ \left(3m - \frac{6m}{X} \right) p_k \mathbb{E} \mathbf{X}_k^2 \\ &+ \left(\frac{3m}{X} + \frac{6(1-m)}{X-1} \right) \mathbb{E} \mathbf{X}_k^2 + 3mp_k \mathbb{E} \mathbf{X}_k \\ &- \frac{m}{X} \mathbb{E} \mathbf{X}_k + mp_k. \end{aligned} \tag{16}$$

For $f(\mathbf{X}_k) = \mathbf{X}_k^4$,

$$\begin{aligned} \frac{1}{\mu} \frac{d}{dt} \mathbb{E} \mathbf{X}_k^4 &= \left(-\frac{4m}{X} - \frac{12(1-m)}{X(X-1)} \right) \mathbb{E} \mathbf{X}_k^4 \\ &+ \left(4m - \frac{12m}{X} \right) p_k \mathbb{E} \mathbf{X}_k^3 \\ &+ \left(\frac{6m}{X} + \frac{12(1-m)}{X-1} \right) \mathbb{E} \mathbf{X}_k^3 + 6mp_k \mathbb{E} \mathbf{X}_k^2 \\ &+ \left(-\frac{4m}{X} - \frac{2(1-m)}{X(X-1)} \right) \mathbb{E} \mathbf{X}_k^2 \\ &+ \left(4m - \frac{2m}{X} \right) p_k \mathbb{E} \mathbf{X}_k \\ &+ \left(\frac{m}{X} + \frac{2(1-m)}{X-1} \right) \mathbb{E} \mathbf{X}_k + mp_k. \end{aligned} \tag{17}$$

Applying formula (7) to $f(\mathbf{X}_k, \mathbf{X}_l) = \mathbf{X}_k \mathbf{X}_l$,

$$\begin{aligned} \frac{1}{\mu} \frac{d}{dt} \mathbb{E} \mathbf{X}_k \mathbf{X}_l &= \left(-\frac{2m}{X} - \frac{2(1-m)}{X(X-1)} \right) \mathbb{E} \mathbf{X}_k \mathbf{X}_l \\ &+ \left(m - \frac{m}{X} \right) (p_k \mathbb{E} \mathbf{X}_l + p_l \mathbb{E} \mathbf{X}_k). \end{aligned} \tag{18}$$

For $f(\mathbf{X}_k, \mathbf{X}_l) = \mathbf{X}_k^2 \mathbf{X}_l$,

$$\begin{aligned} \frac{1}{\mu} \frac{d}{dt} \mathbb{E} \mathbf{X}_k^2 \mathbf{X}_l &= \left(-\frac{3m}{X} - \frac{6(1-m)}{X(X-1)} \right) \mathbb{E} \mathbf{X}_k^2 \mathbf{X}_l \\ &+ \left(m - \frac{2m}{X} \right) p_l \mathbb{E} \mathbf{X}_k^2 \\ &+ \left(2m - \frac{4m}{X} \right) p_k \mathbb{E} \mathbf{X}_k \mathbf{X}_l \\ &+ \left(\frac{m}{X} + \frac{2(1-m)}{X-1} \right) \mathbb{E} \mathbf{X}_k \mathbf{X}_l \\ &+ \left(m - \frac{m}{X} \right) p_k \mathbb{E} \mathbf{X}_l + \frac{m}{X} p_l \mathbb{E} \mathbf{X}_k. \end{aligned} \tag{19}$$

For $f(\mathbf{X}_k, \mathbf{X}_l) = \mathbf{X}_k^2 \mathbf{X}_l^2$,

$$\begin{aligned} \frac{1}{\mu} \frac{d}{dt} \mathbb{E} \mathbf{X}_k^2 \mathbf{X}_l^2 &= \left(-\frac{4m}{X} - \frac{12(1-m)}{X(X-1)} \right) \mathbb{E} \mathbf{X}_k^2 \mathbf{X}_l^2 \\ &+ \left(2m - \frac{6m}{X} \right) (p_k \mathbb{E} \mathbf{X}_k \mathbf{X}_l^2 + p_l \mathbb{E} \mathbf{X}_k^2 \mathbf{X}_l) \\ &+ \left(\frac{m}{X} + \frac{2(1-m)}{X-1} \right) (\mathbb{E} \mathbf{X}_k^2 \mathbf{X}_l + \mathbb{E} \mathbf{X}_k \mathbf{X}_l^2) \\ &+ \left(m - \frac{2m}{X} \right) (p_l \mathbb{E} \mathbf{X}_k^2 + p_k \mathbb{E} \mathbf{X}_l^2) \\ &+ \frac{2m}{X} (p_k \mathbb{E} \mathbf{X}_k \mathbf{X}_l + p_l \mathbb{E} \mathbf{X}_k \mathbf{X}_l) \\ &+ \frac{2(1-m)}{X(X-1)} \mathbb{E} \mathbf{X}_k \mathbf{X}_l + \frac{m}{X} (p_l \mathbb{E} \mathbf{X}_k + p_k \mathbb{E} \mathbf{X}_l). \end{aligned} \tag{20}$$

Appendix D

We describe an alternative method to derive dynamical equations for the momenta. It is based on the characteristic function

$$\Phi(\vec{z}) = \mathbb{E}[z_1^{X_1} \dots z_N^{X_N}] = \sum_{\vec{X}} z_1^{X_1} \dots z_N^{X_N} \mathbb{P}(\vec{X}).$$

A dynamical equation for Φ can be constructed from the master Eq. (1),

$$\begin{aligned} \frac{d}{dt} \Phi(\vec{z}) &= \sum_{\substack{i,j \\ i \neq j}} \left[\mu m p_j \frac{(z_j - z_i)}{X} \frac{\partial}{\partial z_i} \Phi(\vec{z}) \right. \\ &\left. + \mu(1-m) \frac{z_j(z_j - z_i)}{X(X-1)} \frac{\partial^2}{\partial z_i \partial z_j} \Phi(\vec{z}) \right], \end{aligned} \tag{21}$$

or formally,

$$\frac{d}{dt} \Phi(\vec{z}) = \sum_{\substack{i,j \\ i \neq j}} R \left((z_j - z_i) \frac{\partial}{\partial z_i}; p_j, z_j \frac{\partial}{\partial z_j} \right) \Phi(\vec{z}).$$

The equations for the momenta follow by taking derivatives of Φ . For instance,

$$\left. \frac{\partial^3 \Phi}{\partial z_k^2 \partial z_l} \right|_E = \mathbb{E}[\mathbf{X}_k(\mathbf{X}_k - 1)\mathbf{X}_l],$$

where evaluation in E stands for $z_1 = \dots = z_N = 1$.

For our purpose, this method is at least equally laborious as the procedure outlined previously. However, it shows why the dynamical equations for Z_2 are simpler than for Z_1 . By way of illustration, we compute $\mathbb{E} \mathbf{X}_k$ and $\mathbb{E}[\mathbf{X}_k(\mathbf{X}_k - 1)]$.

Taking the derivative with respect to z_k of (21), we obtain

$$\begin{aligned} \frac{d}{dt} \frac{\partial \Phi(\vec{z})}{\partial z_k} &= \sum_{\substack{i,j \\ i \neq j}} \left[\mu m p_j \frac{(\delta_{kj} - \delta_{ki})}{X} \frac{\partial \Phi(\vec{z})}{\partial z_i} \right. \\ &\left. + \mu m p_j \frac{(z_j - z_i)}{X} \frac{\partial^2 \Phi(\vec{z})}{\partial z_k \partial z_i} \right] \end{aligned}$$

$$\begin{aligned}
 & + \mu(1-m) \frac{\delta_{kj}(z_j - z_i)}{X(X-1)} \frac{\partial^2 \Phi(\vec{z})}{\partial z_i \partial z_j} & + 2\mu(1-m) \frac{\delta_{kj}(z_j - z_i)}{X(X-1)} \frac{\partial^3 \Phi(\vec{z})}{\partial z_k \partial z_i \partial z_j} \\
 & + \mu(1-m) \frac{z_j(\delta_{kj} - \delta_{ki})}{X(X-1)} \frac{\partial^2 \Phi(\vec{z})}{\partial z_i \partial z_j} & + 2\mu(1-m) \frac{z_j(\delta_{kj} - \delta_{ki})}{X(X-1)} \frac{\partial^3 \Phi(\vec{z})}{\partial z_k \partial z_i \partial z_j} \\
 & + \mu(1-m) \frac{z_j(z_j - z_i)}{X(X-1)} \frac{\partial^3 \Phi(\vec{z})}{\partial z_k \partial z_i \partial z_j} \Big]. & + \mu(1-m) \frac{z_j(z_j - z_i)}{X(X-1)} \frac{\partial^4 \Phi(\vec{z})}{\partial z_k^2 \partial z_i \partial z_j} \Big].
 \end{aligned} \tag{22}$$

When evaluating in E , we see that the second, third and fifth term drop. The fourth term is antisymmetric for the interchange $i \leftrightarrow j$, such that after summing it cancels as well. We are thus left with

Again, evaluating in E and using antisymmetry where applicable, we get

$$\begin{aligned}
 & \frac{d}{dt} \mathbb{E}[\mathbf{X}_k(\mathbf{X}_k - 1)] \\
 & = \sum_{\substack{i,j \\ i \neq j}} \left[2\mu p_j \frac{(\delta_{kj} - \delta_{ki})}{X} \frac{\partial^2 \Phi(\vec{z})}{\partial z_k \partial z_i} \Big|_E \right. \\
 & \quad \left. + 2\mu(1-m) \frac{\delta_{kj}(\delta_{kj} - \delta_{ki})}{X(X-1)} \frac{\partial^2 \Phi(\vec{z})}{\partial z_i \partial z_j} \Big|_E \right] \\
 & = \sum_{\substack{i \\ i \neq k}} \frac{2\mu p_k}{X} \frac{\partial^2 \Phi(\vec{z})}{\partial z_k \partial z_i} \Big|_E - \sum_{\substack{j \\ j \neq k}} \frac{2\mu p_j}{X} \frac{\partial^2 \Phi(\vec{z})}{\partial z_k^2} \Big|_E + \sum_{\substack{i \\ i \neq k}} \frac{2\mu(1-m)}{X(X-1)} \frac{\partial^2 \Phi(\vec{z})}{\partial z_i \partial z_k} \Big|_E \\
 & = \frac{2\mu p_k}{X} \mathbb{E}[(X - \mathbf{X}_k)\mathbf{X}_k] - \frac{2\mu m(1-p_k)}{X} \mathbb{E}[\mathbf{X}_k(\mathbf{X}_k - 1)] + \frac{2\mu(1-m)}{X(X-1)} \mathbb{E}[(X - \mathbf{X}_k)\mathbf{X}_k] \\
 & = -2\mu \left(\frac{m}{X} - \frac{1-m}{X(X-1)} \right) \mathbb{E}[\mathbf{X}_k(\mathbf{X}_k - 1)] + 2\mu \left(\frac{m(X-1)}{X} p_k + \frac{1-m}{X} \right) \mathbb{E}\mathbf{X}_k,
 \end{aligned}$$

$$\begin{aligned}
 & \frac{d}{dt} \mathbb{E}\mathbf{X}_k = \sum_{\substack{i,j \\ i \neq j}} \mu p_j \frac{(\delta_{kj} - \delta_{ki})}{X} \mathbb{E}\mathbf{X}_i \\
 & = \sum_{\substack{i,j \\ i \neq j}} \mu p_j \frac{\delta_{kj}}{X} \mathbb{E}\mathbf{X}_i - \sum_{\substack{i,j \\ i \neq j}} \mu p_j \frac{\delta_{ki}}{X} \mathbb{E}\mathbf{X}_i \\
 & = \sum_{\substack{i \\ i \neq k}} \frac{\mu p_k}{X} \mathbb{E}[\mathbf{X}_i] - \sum_{\substack{j \\ j \neq k}} \frac{\mu p_j}{X} \mathbb{E}[\mathbf{X}_k] \\
 & = \frac{\mu p_k}{X} \mathbb{E}[X - \mathbf{X}_k] - \frac{\mu m(1-p_k)}{X} \mathbb{E}\mathbf{X}_k \\
 & = \frac{\mu m}{X} (X p_k - \mathbb{E}\mathbf{X}_k),
 \end{aligned}$$

which follows also from (14) and (15).

which is identical to (14).

Taking the derivative of (22) with respect to z_k , we obtain

$$\begin{aligned}
 & \frac{d}{dt} \frac{\partial^2 \Phi(\vec{z})}{\partial z_k^2} = \sum_{\substack{i,j \\ i \neq j}} \left[2\mu p_j \frac{(\delta_{kj} - \delta_{ki})}{X} \frac{\partial^2 \Phi(\vec{z})}{\partial z_k \partial z_i} \right. \\
 & \quad + \mu p_j \frac{(z_j - z_i)}{X} \frac{\partial^3 \Phi(\vec{z})}{\partial z_k^2 \partial z_i} \\
 & \quad \left. + 2\mu(1-m) \frac{\delta_{kj}(\delta_{kj} - \delta_{ki})}{X(X-1)} \frac{\partial^2 \Phi(\vec{z})}{\partial z_i \partial z_j} \right]
 \end{aligned}$$

References

- [1] S.P. Hubbell, The Unified Neutral Theory of Biodiversity and Biogeography, Princeton University, Princeton, NJ, 2001.
- [2] M. Vallade, B. Houchmandzadeh, Analytical solution of a neutral model of biodiversity, Phys. Rev. E 68 (2003) 061902.
- [3] I. Volkov, J.R. Banavar, S.P. Hubbell, A. Maritan, Neutral theory and relative species abundance in ecology, Nature 424 (2003) 1035.
- [4] A.J. McKane, D. Alonso, R.V. Solé, Analytic solution of Hubbell's model of local community dynamics, Theor. Popul. Biol. 65 (2004) 67.
- [5] R.S. Etienne, H. Olf, A novel genealogical approach to neutral biodiversity theory, Ecol. Lett. 7 (2004) 170.
- [6] A.J. McKane, T.J. Newman, Stochastic models in population biology and their deterministic analogs, Phys. Rev. E 70 (2004) 041902.
- [7] P. Babak, Continuous probabilistic approach to species dynamics in Hubbell's zero-sum local community, Phys. Rev. E 74 (2006) 021902.
- [8] S. Azaele, S. Pigolotti, J.R. Banavar, A. Maritan, Dynamical evolution of ecosystems, Nature 444 (2006) 926.
- [9] E.H. Simpson, Measurement of diversity, Nature 163 (1949) 688.
- [10] F. Stewart, Variability in the amount of heterozygosity maintained by neutral mutations, Theor. Popul. Biol. 9 (1976) 188.
- [11] F. He, X.-S. Hu, Hubbell's fundamental biodiversity parameter and the Simpson diversity index, Ecol. Lett. 8 (2005) 386.
- [12] R.S. Etienne, D. Alonso, A dispersal-limited sampling theory for species and alleles, Ecol. Lett. 8 (2005) 1147.
- [13] R.S. Etienne, A new sampling formula for neutral biodiversity, Ecol. Lett. 8 (2005) 253.

- [14] B. Haegeman, B. Sen, J.-J. Godon, J. Hamelin, Accurate measurement of microbial Simpson diversity using fingerprints, 2007, manuscript in preparation.
- [15] C.M. Jessup, R. Kassen, S.E. Forde, B. Kerr, A. Buckling, P.B. Rainey, B.J.M. Bohannan, Big questions small worlds: microbial model systems in ecology, *Trends Ecol. Evol.* 19 (2004) 189.
- [16] J.I. Prosser, B.J.M. Bohannan, T.P. Curtis, R.J. Ellis, M.K. Firestone, R.P. Freckleton, J.L. Green, L.E. Green, K. Killham, J.J. Lennon, A.M. Osborn, M. Solan, C.J. van der Gast, J.P.W. Young, The role of ecological theory in microbial ecology, *Nature Rev. Microbiol.* 5 (2007) 384.
- [17] W.T. Sloan, M. Lunn, S. Woodcock, I.M. Head, S. Nee, T.P. Curtis, Quantifying the roles of immigration and chance in shaping prokaryote community structure, *Environ. Microbiol.* 8 (2005) 732.
- [18] M. Manefield, A. Whiteley, T. Curtis, K. Watanabe, Influence of sustainability and immigration in assembling bacterial populations of known size and function, *Microb. Ecol.* 53 (2007) 348.