Analytical solution of a neutral model of biodiversity.

M. Vallade, B. Houchmandzadeh

Analytical solution of a neutral model of biodiversity

M. Vallade, B. Houchmandzadeh
Laboratoire de Spectrométrie Physique (CNRS), UMR 5588,
Université Joseph Fourier, BP 87,38402 Saint Martin d'Hères Cedex, France

The unified neutral model of biodiversity proposed by S. Hubbell is solved analytically: The distributions of species abundance in the metacommunity and in a local community are calculated exactly as a function of speciation and migration rates and of the size of the community. In the limit of large population sizes the densities of species of given relative abundance are found to be given by universal functions depending only on two parameters.
Pacs numbers: 87.23.-n, 87.10.+e, 02.50.Cw

One of the major goals of biodiversity theories is to explain the distribution of relative species abundance in ecological communities. Recently S. Hubbell has given an extensive description of a unified neutral theory of biodiversity[1] relevant for the description of taxons where trophically similar species such as trees in forests or benthic marine invertebrates communities are competing for resources. In this framework the population dynamics obeys a *zero-sum game*: large landscapes are essentially biotically saturated with individuals and every death is rapidly followed by the birth of an individual belonging to the same or to a different species. The time dependence of relative species abundance in a given community (an "island") is then the result of both internal competition and immigration from outside (the "metacommunity"). The second major Hubbell's hypothesis is that the game rules obey a principle of *neutrality*: all the species are equal competitors with the same per capita chances of dying and reproducing (the plausibility of such an assumption relies on the fact that bad competitors have been eliminated by Darwinian selection). For such a neutral theory, population dynamics is essentially a stochastic process and is expected to obey universal laws depending only on a few number of parameters independent of specific properties of the species involved. This is at variance with the so called " ecological niche theories" for which the particular fitness of a given species to a particular environment is the predominant feature. The neutral model provides a "null–hypothesis" to which actual data can be compared and the influence of other parameters can be assessed. In this respect, various aspects of neutral models have been reviewed by Bell[2]. Since its apparition, this model has generated considerable debate (see for example a critical review by Riklefs [3] Hubbell's reply[4], and [5] and references therein).

In addition to immigration, the species evolution depends also on speciation. Mutations, although being rare events, drive the long term drift of large communities. In Hubbell's theory they are considered as the fundamental process in determining the equilibrium distribution of species abundance in the *metacommunity*. He introduced a dimensionless parameter θ (the "fundamental biodiversity number") proportional to the product of the mutation rate ν with the total number of individuals in the metacommunity J_M (the time scale being fixed by the average lifetime of individuals).The relative species abundance in a local community depends essentially on its size J, the immigration rate from the metacommunity m and the mutation rate ν (through θ). The Hubbell's formulation of the problem can only be solved through numerical calculations. The purpose of the present paper is to report on an *exact analytical solution* of the model which gives directly the equilibrium distributions of species abundances both in the metacommunity and in the local community.

Let us first consider the metacommunity dynamics under the influence of mutations.We assume that the death of an individual in a metacommunity of size J_M is immediately followed by the birth of another one.The lifetimes and the mutation rates are assumed to be the same for all the species according to the neutral and the zero sum rules of the Hubbell's model. In the continuous time limit, the probability $P_k(t)$ that k individuals belonging to the same species are present at time t obeys a simple master equation:

$$dP_k/dt = A_{k-1}P_{k-1} + C_{k+1}P_{k+1} - (A_k + C_k)P_k \quad (1)$$

where the coefficients A_k and C_k are respectively the increase and decrease rates of a species of abundance k , expressed in units of individual lifetime. These coefficients are readily found to be given by [1]:

$$A_k = \frac{(J_M - k)k}{J_M(J_M - 1)}(1 - \nu) \quad (2)$$

$$C_k = \frac{k(J_M - k + (k-1)\nu)}{J_M(J_M - 1)} \quad (3)$$

The increase rate A_k for example is proportional to the probability of death of an individual which does not belong to the considered species $(J_M - k)/J_M$, times the probability of birth of an individual which belongs to the considered species $k/(J_M - 1)$, times the probability of no mutation $(1 - \nu)$.

Let $\langle \phi_k(t) \rangle_M$ designate the average number of species of abundance k. In the long time limit, the species which contribute to $\langle \phi_k(t) \rangle_M$ are those which have appeared by mutation at an earlier time $(t - u)$ and have reached the size k at time t . As all species are assumed to be equivalent, they have the same probability $p(u)du = \nu du$ to

appear between times u and $u+du$. The time evolution of $\langle\phi_k(t)\rangle_M$is then simply given by :

$$\langle\phi_k(t)\rangle_M = \int_0^t P_k(t-u)p(u)du = \nu\int_0^t P_k(u)du \quad (4)$$

where $P_k(t)$ is the solution of eq. 1 with initial condition $P_k(0)=\delta_{k,1}$. Using Laplace transformation, the equilibrium distribution of the number of species with a given abundance is then found to be (see Appendix):

$$\langle\phi_k\rangle_M = \frac{\theta\Gamma(J_M+1)\Gamma(J_M+\theta-k)}{k\Gamma(J_M+1-k)\Gamma(J_M+\theta)} \quad (5)$$

where θ is defined by:

$$\theta = \frac{(J_M-1)\nu}{1-\nu} \quad (6)$$

This definition of θ coincides (within a factor 2) with Hubbell's universal biodiversity number in the limit $J_M \gg 1$ and $\nu \ll 1$. One can easily check that $\langle\phi_k\rangle_M$ agrees exactly with the expression calculated at small J_M by the Hubbell's method. The total average number of species $\langle S\rangle_M = \sum_{k=1}^{J_M}\langle\phi_k\rangle_M$ can be put under the form :

$$\langle S\rangle_M = \sum_{i=0}^{J_M-1}\theta/(\theta+i) \quad (7)$$

in agreement with [6]. Typical distributions corresponding to Eq.5 are shown in fig. 1 using Preston plots (histogram of the number of species per octave of abundance).They are in perfect agreement with our numerical simulations of the problem.

For large J_M and finite θ, $g_M(\omega) = lim_{J_M\to\infty} J_M\langle\phi_k\rangle_M$, which represents the density of species of relative abundance $\omega = k/J_M$, reads:

$$g_M(\omega) = \frac{\theta(1-\omega)^{\theta-1}}{\omega} \quad (8)$$

This last result could have been obtained directly from the continuous reformulation of the evolution equation of $P_k(t)$ (Fokker–Planck equation), analog to the approach taken by Malécot[7].

Let us now consider the distribution of species in a local community of size J in contact with the above described metacommunity. If $J \ll J_M$ mutations within the community can be neglected and the equilibrium distribution in the metacommunity is negligibly modified by migrations from or towards the community. Let us now consider the probability $P_n(t;\omega)$ of finding a given species with abundance n in the community knowing that this species has relative abundance $\omega = k/J_M$ in the metacommunity. Because of the large size of the metacommunity compared to the community, ω can be considered constant over time scales relevant for the evolution of the community. Thus $P_n(t;\omega)$ obeys a differential equation similar to Eq 1 but with coefficients A_n and C_n given by:

$$A_n = \frac{(J-n)}{J}\left(\frac{n}{J-1}(1-m)+m\omega\right) \quad (9)$$

$$C_n = \frac{n}{J}\left(\frac{J-n}{J-1}(1-m)+m(1-\omega)\right) \quad (10)$$

The equilibrium solution $P_n(\omega)$ can be calculated (see Appendix) and it takes a concise form by introducing a re-scaled immigration parameter μ defined in a way similar to θ:

$$\mu = \frac{(J-1)m}{1-m} \quad (11)$$

$$P_n(\omega) = \binom{J}{n}\frac{(\mu\omega)_n(\mu(1-\omega))_{J-n}}{(\mu)_J} \quad (12)$$

In this expression $\binom{J}{n}$ is the binomial coefficient and $(a)_n = \Gamma(a+n)/\Gamma(a)$ is the Pochhammer coefficient[8] . Note that $P_n(\omega)$ is a polynomial of degree J in ω which can be conveniently written:

$$P_n(\omega) = \omega\sum_{k=1}^{J-1} f_{nk}(1-\omega)^k \quad (13)$$

where the coefficients f_{nk} depend on J and μ.

The important point is now that eq. 5 and eq. 12 can be combined to give the distribution of species $\langle\phi_n\rangle_C$ in the local community:

$$\langle\phi_n\rangle_C = \sum_{k=1}^{J_M} P_n(k/J_M)\langle\phi_k\rangle_M \quad (14)$$

This equation is the main result of the present paper. It expresses the way the internal population dynamics of an island is modified by immigration from the environment. In the extreme cases, if $m=1$, $\langle\phi_n\rangle_C = \langle\phi_k\rangle_M$ (complete "thermalisation") and if $m=0$, the isolated community undergoes total dominance by a single specie as will be discussed below.

The proof of this statement can be derived in the following way. Let $P_{jC}(n)$ be the probability that a given specie j has an abundance n in the community and $P_{jM}(k)$ the probability that this same specie has an abundance k in the metacommunity . Since $P_n(\omega)$ has the meaning of a conditional probability, one has:

$$P_{jC}(n) = \sum_{k=1}^{J_M} P_n(k/J_M)P_{jM}(k)$$

so that, if s is the total number of species:

$$\begin{aligned}\langle\phi_n\rangle_C &= <\sum_{j=1}^{s}\delta_{n,n_j}>_C = \sum_{j=1}^{s}P_{jC}(n)\\ &= \sum_{k=1}^{J_M}P_n(k/J_M)\sum_{j=1}^{s}P_{jM}(k) = \sum_{k=1}^{J_M}P_n(k/J_M)\langle\phi_k\rangle_M\end{aligned}$$

In the limit $J_M \to \infty$, $\langle \phi_n \rangle_C$ can be computed using eqs. 8 and 13:

$$\langle \phi_n \rangle_C = \int_0^1 P_n(\omega) g_M(\omega) d\omega = \sum_{k=1}^{J-1} f_{nk} \theta/(\theta+k) \quad (n \leq J-1) \tag{15}$$

Preston plots corresponding to this result are shown in Fig.2 . Our analytical results are again in very good agreement with numerical simulations. The existence of a maximum in these curves is a feature often observed in nature. It corresponds to an intermediate situation between full dominance of a single species in absence of immigration ($\mu = 0$) and full equilibrium with the metacommunity ($\mu = \infty$). Typical ecosystems for which the distribution of species abundance has been experimentally measured and can be well described by the present model are closed canopy forests (in Malaysia and Panama) [9] as cited in [1] .

The average total number of species in the local community $\langle S \rangle_C$ can be put in a form which generalizes eq. 7:

$$\langle S \rangle_C = \sum_{i=0}^{J-1} \frac{\theta}{(\theta+i)(\mu)_J} \sum_{j=i+1}^{J} \alpha_j \mu^j \tag{16}$$

where α_j are the coefficients of the polynomial expansion of the Pochhammer coefficient $(\mu)_J = \sum_{j=1}^{J} \alpha_j \mu^j$

For large J , the density of species $g_C(\omega) = lim_{J\to\infty} J \langle \phi_n \rangle_C$ in the continuum limit ($\omega = n/J$) is:

$$g_C(\omega) = \mu\theta \int_0^1 \left(\begin{array}{c} \mu \\ \mu u \end{array} \right) (1-\omega)^{\mu u - 1} \omega^{\mu(1-u)-1} u^{\theta} du \tag{17}$$

In this equation $\mu \simeq mJ$ is considered as finite when $J \to \infty$.

This equation describes the asymptotic behavior of large local communities, as a function of only two parameters θ and μ . Fig.3 shows $\omega g_C(\omega)$ for several values of μ. This function which is normalized to 1, represents the probability density of finding an individual belonging to a species of relative abundance ω . This figure clearly shows the crossover from dominance at small μ to diversity at large μ . When μ increases at constant θ, $\omega g_C(\omega)$ approaches continuously $\omega g_M(\omega)$. This is the full "thermalisation" of the community with the metacommunity when the immigration rate m approaches unity.

Eq.17 however must be used cautiously. Our numerical simulations show that species distribution depends critically on the ratio J/J_M when both J and J_M are large . This emphasizes the importance of discrete formula (eq. 14).

To conclude we have derived an analytical solution of the Hubbell's "unified" model of biodiversity: exact expressions for the equilibrium distributions of species abundance and for the total number of species have been found for a community and a metacommunity of finite sizes. In the limit of large population sizes we have shown that the equilibrium distributions of species are described by universal functions depending only on two scaling parameters, θ and μ, related respectively to the mutation and migration rates. The model discussed above is essentially "zero dimensional", which means that no spatial inhomogeneous distribution of the species has been considered. There is an obvious interest in extending the present results to 1 and 2 dimensional systems : important results have been obtained for the probability of sampling the same species at two different locations as a function of their separation, both theoretically[7, 10] and from field measurements[11]. The present work opens the way to further analytical results in that direction, specially for deriving the full species distribution similar to eq. 14.

Appendix

The set of master equations (1) can be written under a matrix form:

$$d\mathbf{P}/dt = \mathbf{H}.\mathbf{P}$$

with $\mathbf{P}$ a $(J+1)$ dimensional vector corresponding to $P_n(t)$ ($n = 0$ to J) and $\mathbf{H}$ a matrix defined by:

$$\mathbf{H} = \begin{bmatrix} -A_0 & C_1 & & & \\ A_0 & -(A_1+C_1) & C_2 & & \\ & A_1 & . & . & \\ & & . & . & C_J \\ & & & A_{J-1} & -C_J \end{bmatrix}$$

Using the Laplace transformation $\mathbf{Q}(E) = L[\mathbf{P}(t)]$ and defining the matrix $\mathbf{G}(E) = (E\mathbf{I} - \mathbf{H})^{-1}$ one has:

$$\mathbf{P}(t) = L^{-1}[\mathbf{G}(E)].\mathbf{P}(0)$$

The long time behavior of $\mathbf{P}(t)$ is obtained by considering the limit $E \to 0$ of $\mathbf{G}(E)$.

A1 -Metacommunity:

The probability $P_k(t)$ that a specie which has appeared at time $t = 0$ has an abundance k at time t is given by :

$$P_k(t) = L^{-1}[\mathbf{G}_{k1}(E)]$$

According to eq. 2, $A_0 = 0$. Using a recurrence method one finds that, when $E \to 0$:

$$\begin{aligned} \mathbf{G}_{01}(E) &= 1/E + \mathcal{O}(1) \\ \mathbf{G}_{11}(E) &= 1/C_1 + \mathcal{O}(E) \\ \mathbf{G}_{k1}(E) &= \frac{\prod_{j=1}^{k-1} A_j}{\prod_{j=1}^{k} C_j} + \mathcal{O}(E) \quad (1 < k \leq J_M) \end{aligned}$$

Therefore, when $t \to \infty$, $P_k(t) \to \delta_{k,0}$: every specie disappears by mutation. However, as new species appear continuously, the population distribution goes to a

well-defined stationary value which can be calculated by taking the Laplace transform of eq.4:

$$< \phi_k >_M = \nu L^{-1}[\mathbf{G}_{k1}(E)/E]_{E=0}$$

and one gets:

$$\begin{aligned} < \phi_1 >_M &= \nu/C_1 \\ < \phi_k >_M &= \nu \frac{\prod_{j=1}^{k-1} A_j}{\prod_{j=1}^{k} C_j} \quad (1 < k \leq J_M) \end{aligned}$$

Using the definition of θ (eq.6) , the coefficients C_j in eq.2 can be written:

$$C_j = \frac{j(J_M - j + \theta)}{J_M(J_M - 1)}(1 - \nu)$$

so that one gets:

$$< \phi_k >_M = \frac{\theta}{k} \prod_{j=1}^{k} \frac{(J_M - j + 1)}{(J_M - j + \theta)} = \frac{\theta}{k} \frac{\Gamma(J_M + 1)\Gamma(J_M + \theta - k)}{\Gamma(J_M + \theta)\Gamma(J_M + 1 - k)}$$

which is eq.5. One can easily check that this distribution satisfies correctly the sum rule:

$$\sum_{k=1}^{J_M} k < \phi_k >_M = J_M$$

Using this relation, the total average number of species in the metacommunity $< S(J_M) >_M = \sum_k < \phi_k >_M$ can be shown to obey the recurrence relation:

$$< S(J_M + 1) >_M = < S(J_M) >_M + \frac{\theta}{\theta + J_M}$$

Using $< S(1) >_M = 1$, one gets eq. 7

A2- Community:

In this case $A_0 \neq 0$. In the limit $E \to 0$ one then finds that:

$$\mathbf{G}_{nm}(E) = \frac{M_n}{E \sum_{n=0}^{J} M_n} + \mathcal{O}(1)$$

with:

$$\begin{aligned} M_n &= \prod_{j=0}^{n-1} A_j \prod_{j=n+1}^{J} C_j \quad (0 < n < J) \\ M_0 &= \prod_{j=1}^{J} C_j \\ M_J &= \prod_{j=0}^{J-1} A_j \end{aligned}$$

Introducing μ as defined in eq.11 into the coefficients A_n and C_n given in eq.9 and 10, one gets:

$$\begin{aligned} A_n &= \frac{m}{J\mu}(J - n)(n + \mu\omega) \\ C_n &= \frac{m}{J\mu} n(J - n + \mu(1 - \omega)) \end{aligned}$$

When $t \to \infty$ the probability $P_n(t;\omega)$ goes to a stationary value $P_n(\omega)$ given by:

$$P_n(\omega) = \frac{M_n}{\sum_{n=0}^{J} M_n}$$

The denominator of this expression is found to be to be independent of ω and the probability $P_n(\omega)$ can be written:

$$P_n(\omega) = \frac{\prod_{j=0}^{n-1}(J - j)(j + \mu\omega) \prod_{j=n+1}^{J} j(J - j + \mu(1 - \omega))}{\prod_{j=0}^{J-1}(J - j)(j + \mu)} \quad (0$$

which finally leads to eq.12. $P_n(\omega)$ is a polynomial of degree J which can be written:

$$P_n(\omega) = \omega \begin{pmatrix} J \\ n \end{pmatrix} \frac{\mu}{(\mu)_J} \prod_{j=1}^{n-1} (\mu + j - \mu(1-\omega)) \prod_{j=0}^{J-n-1} (\mu(1-\omega) + j)$$

which is of the form eq.13 and defines the coefficients f_{nk} in that equation.

[1] SP. Hubbell, *The unified neutral theory of biodiversity and biogeography,*Princeton University Press, Princeton (NJ) (2001).

[2] G. Bell, *Neutral Macroecology*, Science **293**,2413–2418 (2001).

[3] RE. Riklefs, *a comment on Hubbell's zero–sum ecological drift,* OIKOS, **100**, 185–192 (2003).

[4] SP. Hubbell, Modes of speciation and the lifespans of species under neutrality, OIKOS, **100**, 193–199 (2003).

[5] BJ. McGill, *A test of the unified neutral theory of biodiversity*, Nature, **422**,881–885 (2003).

[6] W. Ewens, *The sampling theory of selectively neutral alleles*, Th. Pop. Biol. **3**,87–112 (1972).

[7] G. Malécot, *The mathematics of heredity,* WH Freeman and Co., San Francisco (1969).Our eq.8 actually coincides with eq. 3.2.6 , page 57 of this book, in the special case : $v = t = w = 0$ and $4Nu = \theta$.

[8] M. Abramowitz and IA. Stegun, *Handbook of mathemat-*

ical functions, Dover publication, NY (1972).

[9] R. Condit *et al.*,*Species–area and species–individual relationships for tropical trees.* Journal of Ecology, **84**, 549–562 (1996).

[10] J. Chave and EG. Leigh, *A spatially explicit neutral model of β–diversity in tropical forests*, Th. Pop. Biol. **62**,153–68 (2002).

[11] R. Condit *et al.*, *β–diversity in tropical forest trees,* Science **295**,666–9 (2002).

Figure Captions.

Fig. 1 Preston plot (binned by octave) of metacommunity species distribution obtained by Monte–Carlo simulation : at each time step, an individual is removed from the community and replaced, with probability ν by a new mutant or, with probability $(1-\nu)$ by an existing individual replicate. $J_M = 1024$. Solid lines represent theoretical species distribution, as given by eq.8.

Fig. 2 Preston plot of local community species distribution obtained by Monte–Carlo simulation for $J_M = 1024$, $J = 128$ and 2 values of μ and θ. Dashed lines represent theoretical species distribution, as given by eq.14

Fig. 3 Probability density of finding an individual in a species of relative abundance ω in the local community for $J \to \infty$ (eq.17). $\theta = 5$, and various values of μ as indicated on each curve. $\mu = \infty$ corresponds to the metacommunity distribution.