An exactly solvable coarse-grained model for species diversity

Samir Suweis\textsuperscript{1}, Andrea Rinaldo\textsuperscript{2,3} and Amos Maritan\textsuperscript{1,4}

\textsuperscript{1} Dipartimento di Fisica G Galilei, INFN and CNISM Università di Padova, Via Marzolo 8, I-35151 Padova, Italy
\textsuperscript{2} Laboratory of Ecohydrology ECHO/ISTE/ENAC, Faculté ENAC, École Polytechnique Fédérale Lausanne (EPFL), Lausanne (CH), Switzerland
\textsuperscript{3} Dipartimento IMAGE, Università di Padova, I-35131 Padova, Italy
E-mail: suweis@pd.infn.it, andrea.rinaldo@epfl.ch and maritan@pd.infn.it

Received 13 March 2012
Accepted 25 June 2012
Published 19 July 2012

Online at stacks.iop.org/JSTAT/2012/P07017
doi:10.1088/1742-5468/2012/07/P07017

Abstract. We present novel analytical results concerning ecosystem species diversity that stem from a proposed coarse-grained neutral model based on birth–death processes. The relevance of the problem lies in the urgency for understanding and synthesizing both theoretical results from ecological neutral theory and empirical evidence on species diversity preservation. The neutral model of biodiversity deals with ecosystems at the same trophic level, where per capita vital rates are assumed to be species independent. Closed-form analytical solutions for the neutral theory are obtained within a coarse-grained model, where the only input is the species persistence time distribution. Our results pertain to: the probability distribution function of the number of species in the ecosystem, both in transient and in stationary states; the n-point connected time correlation function; and the survival probability, defined as the distribution of time spans to local extinction for a species randomly sampled from the community. Analytical predictions are also tested on empirical data from an estuarine fish ecosystem. We find that emerging properties of the ecosystem are very robust and do not depend on specific details of the model, with implications for biodiversity and conservation biology.

Keywords: correlation functions, coarse-graining (theory), stochastic processes (experiment), population dynamics (theory)

\textsuperscript{4} Author to whom any correspondence should be addressed.

© 2012 IOP Publishing Ltd and SISSA Medialab srl 1742-5468/12/P07017+18$33.00
1. Introduction

Statistical physics is decisively contributing to our understanding of ecological processes. In fact it is providing powerful theoretical tools and innovative steps towards the comprehension and the synthesis of broad empirical evidence on macroecological laws and emerging biodiversity patterns [1]–[5]. One field where statistical physics has been particularly successful is that of the neutral theory of biodiversity [6]–[13]. This theory is based on the assumption that, within the same trophic level, per capita vital rates are species independent. It offers a unified theoretical framework for ecosystem dynamics by invoking solely basic ecological processes such as birth, death, migration and dispersal limitation [6, 7]. Exact solutions have been found for several ecologically relevant quantities, such as: the relative species abundance distribution (RSA) [9, 10, 14]; species spatial patterns and clustering [12, 13]; patterns of $\beta$-diversity (i.e. intra-species and inter-species spatial correlation) [15, 16]; the species–area relationship (SAR) [4, 18]; and species' persistence time distributions [19]–[21]. These results enabled the theory to be tested by contrasting empirical data, and to assess the power of neutral models in predicting ecological patterns in many ecosystems.

Let us consider a given trophic level (e.g. plants) in a local ecosystem. ‘Local’ means a community immersed in a larger one, called the meta-community, that is considered as
An exactly solvable coarse-grained model for species diversity

infinite in size with respect to the local community (serving as a reservoir). We assume that each individual in the local community has a natural death rate $d$ and birth rate $b$ (linear birth and death rates). Moreover, due to speciation or diversification events (e.g. immigration from an outside region), new species (i.e. species not already present in the local community) enter the system with rate $\lambda$.

A general formulation of the neutral theory employs the birth–death master equation (ME) for the abundance dynamics of a species in the ecosystem [8, 11, 22]. Several exact results are known for birth–death models where the population, $N$, of the local community is strictly conserved and/or the corresponding $\lambda$ scales as $N$ (e.g., see the infinite alleles model with mutation [17] or the voter model [24]–[26]).

In particular, a very well known and well studied pattern in ecology is the RSA of the local community, defined as the fraction of species of a given abundance. It describes key elements of biodiversity, such as the frequency of rare species in the ecosystem [6, 7]. Another important quantity in conservation ecology is species richness: the number of different species in the community, independent of their abundance. Analytical solutions for both of these quantities can be found in the literature [8, 11, 17, 27].

Unfortunately, transient solutions for species richness and its $n$-point time correlations functions (i.e. the probability of having $S_1$ species at time $t_1$, $S_2$ species at time $t_2$, etc) are not easy to calculate using this approach. Moreover, greater complication comes into play if we want to generalize the dynamics with non-linear birth and death rates, i.e. birth and death rates which are not proportional to the population size [23]. For this reason, in section 2 we introduce a coarse-grained version of this model which will allow us to deduce in a simpler way novel analytical results concerning ecosystem biodiversity, where the only input is the species persistence time distribution. We note that, given the assumption of no interaction among species, no approximation is made to obtain these results—but they are limited to knowing whether a species is present/absent, as opposed to knowing how many individuals there are from that species at a given time. In particular we obtain closed-form analytical results for: the probability distribution function (pdf) of the number of species in the ecosystem in transient states; the $n$-point connected time correlation function; and the survival probability, $p_s(\tau)$, defined as the pdf of the time spans $\tau$ to local extinction for a species randomly sampled among the observed assemblages at a certain given time. A comparison of analytical predictions with empirical data from an estuarine fish ecosystem will be carried out. A set of conclusions then closes the paper.

2. The theoretical framework

2.1. The grand-canonical formulation of the neutral theory

In the natural/realistic case where the total numbers of species ($S$) and individuals ($N$) in the local community are not fixed, we can describe the abundance dynamics of the species in the system employing a ‘grand-canonical’ formulation of the neutral theory. In particular, if $\phi_j$ gives the number of species with $j$ individuals, and $\bar{\phi} = \{\phi_1, \phi_2, \ldots, \phi_\infty\}$, then the probability $P(\bar{\phi}, t)$—of having at time $t$, $\phi_1$ species with one individual, $\phi_2$ species with two individuals, and so on—is unequivocally described once the initial condition $P(\bar{\phi}, t = 0)$ is known and the rates of transition from $\bar{\phi}'$ to $\bar{\phi}$ due to birth/death or speciation events are given. In particular, for $k \geq 1$ with birth rate $b_k = b \cdot k \cdot \phi_k$ (birth events...
An exactly solvable coarse-grained model for species diversity

within a species of \( k \) individuals) we have the transitions \( \{ \phi_k \to \phi_k - 1; \phi_{k+1} \to \phi_{k+1} + 1 \} \), while with the death rate \( d_k = d \cdot k \cdot \phi_k \) (the death of an individual belonging to a species with \( k \) individuals) we have the transitions \( \{ \phi_k \to \phi_k - 1; \phi_{k-1} \to \phi_{k-1} + 1 \} \). Finally with rate \( \lambda \), the transition \( \{ \phi_1 \to \phi_1 + 1 \} \) occurs. We remark that in our framework, the total number of individuals \( N \) is not fixed and the rate at which a new species enters the system is independent of \( N \).

In our grand-canonical formulation, the stationary solution of the master equation corresponding to \( P(\phi, t) \) can be written as \( P_{\text{stat}}(\phi) = \lim_{t \to \infty} P(\phi, t) = (1/Z) \prod_k P(\phi_k) \), with

\[
P(\phi_k) = \left( \frac{\gamma x^k}{k!} \right) \frac{1}{\phi_k!}; \quad Z = e^{-\gamma \ln(1-x)},
\]

where \( \gamma = \lambda/b \) and \( x = b/d \). Note that \( x \) represents the ratio of effective per capita birth and death rates, and if \( \lambda \neq 0 \) it has to be less than 1 in order to avoid demographic explosion. On the other hand if \( \lambda = 0 \), then at equilibrium there are no individuals in the community because all species eventually go extinct [8, 9].

In our theoretical framework the RSA is deduced from the first moments of \( P_{\text{stat}} \), while species richness is described by the probability of having \( s \) different species in the community at stationarity, and turns out to be Poisson distributed with mean \( \langle S \rangle = -\gamma \ln(1-x) \):

\[
P(s) = \sum_{\phi_1, \phi_2, \ldots, \phi_\infty} P(\tilde{\phi}) \delta_K \left( \sum_j \phi_j - s \right) = \frac{(-\gamma \ln(1-x))^s}{s!} (1-x)^\gamma
\]

where \( \delta_K \) is the Kronecker delta, which is 1 when its argument is 0, and 0 otherwise.

Since to maintain a finite local community size, \( x \) must be less than 1, each species is eventually doomed to extinction. We thus can define the persistence time \( \tau \) of a species as the time between its emergence in the system (due to diversification or immigration) and its extinction [19]–[21]. In particular, persistence times for a species that undergoes the linear birth–death dynamics presented above are distributed according to the species persistence time (SPT) pdf [19]:

\[
p_{\text{SPT}}(t) = d \left[ \frac{1-x}{e^{d(1-x)t} - 1} \right]^2 e^{d(1-x)t}.
\]

From now on, without loss of generality, we set \( d = 1 \). From equation (3) we can express the mean persistence time, a key ecological quantity also known as the mean extinction time [19, 17, 27], as \( \langle \tau \rangle = \int_0^\infty \tau p_{\text{SPT}}(\tau) \, d\tau = -\ln(1-x)/x \). Finally, as expected, the mean number of species at stationarity is related to \( \langle \tau \rangle \), through equations (2) and (3), by the simple formula \( \langle S \rangle = \lambda \langle \tau \rangle \).

2.2. The coarse-grained model

A coarse-grained view of the grand-canonical formulation of the neutral theory can be described as follows. Each species \( i \) within the local community emerges at a random time \( \{ t_i \}_{i \geq 0} \) with a rate \( \lambda \), that is, the probability that a new species emerges in the infinitesimal time interval \( dt \) is \( \lambda \, dt \). Therefore the probability of having \( k \) new species in
An exactly solvable coarse-grained model for species diversity

Figure 1. Coarse-grained view of the ecosystem species dynamics. Species emerge in the ecosystem uniformly in time, and each persists for a random time $\tau$ drawn from the SPT pdf $p_{spt}$.

the ecosystem up to time $t$, $U_k(t)$, is given by (see the appendix)

$$U_k(t) = e^{-\lambda t} \sum_{s_0=1}^{k} (\lambda t)^{k-s_0} \frac{(k-s_0)!}{(k-s_0)!},$$

where $s_0$ is the number species at time $t = 0$, i.e., $U_k(t = 0) = \delta_k(k-s_0)$. Then the newcomer species $i$ persists for a random duration $\tau_i$ where the random variables $\tau_i$ are independent and identically distributed with a given pdf $p_{spt}(\tau)$. Therefore species arrive as a Poisson process (with rate $\lambda$), and then depart after some non-Poissonian waiting time $\tau \sim p_{spt}$.

This model is coarse grained in the sense that we do not take explicitly into account information on species abundance, implicitly contained in the functional form of $p_{spt}(\tau)$ (see figure 1). This approach enables us to go beyond equation (2), and to generalize these results also for general birth and death rates, by taking different functional shapes of the SPT pdf. We remark that the mapping is exact only with respect to the grand-canonical formulation of the neutral theory, while, as we will show, it is only a good approximation of the classic birth–death ME approach in the limit of large local community ($N$ fixed and $N \to \infty$).

Within our framework, the number of persistent species in the ecosystem at a given time $t$ (see figure 2) is given by

$$S(t) = k(t) \sum_{i=1}^{k(t)} \Theta(t_i + \tau_i - t),$$

where $k(t)$ is the number of new species that entered the system in the time interval $[0, t)$, i.e. $t_i < t$ for $i \leq k(t)$ and $t_i \geq t$ for $i > k(t)$. $\Theta(z)$ is the Heaviside step function, which is 1 when its argument is positive and 0 otherwise. $t_i$ is the time when the $i$th species enters the ecosystem and $t_i + \tau_i$ when it goes extinct. The probability of having $s$ species present in the ecosystem at time $t$ is thus

$$P(s, t) = \langle \delta_k(s - S(t)) \rangle,$$

i.e.

$$P(s, t) = \sum_{k=0}^{+\infty} U_k(t) \int_0^t \prod_{i=1}^{k} \frac{dt_i}{t_i} \int_0^\infty \prod_{j=1}^{k} d\tau_j p_{spt}(\tau_j) \delta_k \left( s - \sum_{i=1}^{k} \Theta(t_i + \tau_i - t) \right).$$

It is useful and customary to define the generating function of the process (the discrete Laplace transform) as $\hat{P}(z, t) = \sum_{s=0}^{+\infty} z^s P(s, t)$, and analogously for $U_k(t)$. Equations (4)
Figure 2. Schematic representation of survival times $\tau_s$, defined as the time to local extinction of a species randomly sampled among the observed assemblages at a certain time $T$. $\tau$, however, denotes the persistence time of a species, and it is defined as the time between its emergence in the system and its extinction. Equation (43) gives the relation between the two distributions, independently of the functional form of the birth and death rates.

and (6) lead to (see the appendix)

$$\hat{P}(z,t) = \hat{U} \left( 1 + \frac{z-1}{t} f(t), 0 \right) e^{\lambda f(t)(z-1)},$$

(7)

where

$$f(t) = \int_0^t d\tau P_>(\tau) = \int_0^{+\infty} p_{\text{pspt}}(\tau) \min[t, \tau] d\tau,$$

(8)

with $P_>(t) = \int_0^{+\infty} p_{\text{pspt}}(\tau)$. If, for example, we assume that $U_s(0) = \delta_K(s-1)$ (and therefore $\hat{U}(z,0) = z$), then from equation (7) we get

$$P(s,t) = e^{-\lambda f(t)} \frac{(\lambda f(t))^s}{s!} \left( 1 + \frac{s - \lambda f(t)}{\lambda t} \right).$$

(9)

One can see that $f(t)/t = \int_0^{+\infty} p_{\text{pspt}}(\tau) \min(1, \tau/t) d\tau \to 0$ in the $t \to \infty$ limit. This follows from $\min(1, \tau/t) \leq 1$, $\lim_{t \to \infty} \min(1, \tau/t) = 0$ for all $\tau$ and the Lebesgue dominated convergence theorem (see for example [28]). Since $\hat{U}(z,0)$ is continuous in the interval $-1 \leq z \leq 1$, we have $\hat{U}((1+(z-1)/t)f(t), 0) \to \hat{U}(1,0) = 1$ in the large time limit, implying that the initial condition is forgotten in this limit, as expected. If $\int_0^{+\infty} p_{\text{pspt}}(\tau) \tau d\tau = \langle \tau \rangle < \infty$, then $\lim_{t \to \infty} f(t) = \langle \tau \rangle$ and from equation (7) we get $\hat{P}(z) = \lim_{t \to \infty} \hat{P}(z,t) = e^{(1-z)\lambda \langle \tau \rangle}$, or equivalently,

$$P(s) = \lim_{t \to \infty} P(s,t) = \frac{(\lambda \langle \tau \rangle)^s}{s!} e^{-\lambda \langle \tau \rangle},$$

(10)

i.e. a Poisson pdf with mean given by the product of the species emergence rate $\lambda$ and the mean average persistence time. As expected, we found the same result as is given by equation (2).
Figure 3. (a) Comparison between the empirical SPT pdf (histogram) and theoretical $P(s)$ given by equation (10) (blue line). (b) Monthly time series of the analyzed estuarine fish ecosystem. The black dashed line represents the mean number of species $S_t$, while the green dot–dashed line represents the predicted average $\langle S \rangle$. (c) The analytical solution for $P(s)$ has been verified numerically.

From equation (7) we can easily calculate the average numbers of species at a generic time $t$. Indeed, if $\langle S \rangle_{t=0} = \sum_s S U_s(t=0) < \infty$, we have that $\hat{U}(z,0)$ admits a left derivative at $z=1$, and taking the left derivative at $z=1$ of equation (7) we obtain

$$\langle S \rangle_t = \frac{\partial}{\partial z} \hat{P}(z,t)|_{z=1} = \langle S \rangle_{t=0} \frac{f(t)}{t} + \lambda f(t) t \to \infty \lambda \langle \tau \rangle,$$

where again we see that the initial condition is forgotten in the large time limit. These results have been tested numerically (see figure 3).

3. The generating function approach

We are interested in the $n$-point connected time correlation function of our process (or the cumulant; see [29]) $\langle S(t^1)S(t^2)\ldots S(t^n) \rangle_C$ in stationary conditions ($\min_i t_i \to \infty$) and for $t^{i+1} - t^i$ fixed, for $i = 1, \ldots, n - 1$. The generating function is given by

$$Z_T(\{h\}) = \langle e^{-\int_0^T dt h(t)S(t)} \rangle = \sum_{k=0}^{+\infty} U_k(T) \int_0^T \prod_{i=1}^k \frac{dt_i}{T} \int_0^{+\infty} \prod_{j=1}^k d\tau_j p_{sp} (\tau_j)$$

$$\times \exp \left\{ - \sum_{i=1}^k \int_0^T dt' h(t') \Theta(t_i + \tau_i - t') \Theta(t' - t_i) \right\}$$

$$= \sum_{k=0}^{+\infty} U_k(T) z(T)^k = \hat{U}(z(T),0) e^{XT(z(T)-1)},$$

where

$$z(T) = \int_0^T \frac{dt}{T} \left\langle \exp \left\{ - \int_0^T dt' h(t') \Theta(t + \tau - t') \Theta(t' - t) \right\} \right\rangle_{\tau},$$

with $\langle \cdot \rangle_{\tau} = \int_0^\infty d\tau \cdot p_{sp} (\tau)$.

doi:10.1088/1742-5468/2012/07/P07017
The $n$-point correlation function is obtained by choosing

$$
h(t) = \sum_{i=1}^{n} \delta(t - t^i)h_i \quad \text{with } h_i \geq 0
$$

(14)

and

$$
\lim_{t^i \to \infty} \langle S(t^1)S(t^2)\cdots S(t^n) \rangle_C = \lim_{T \to \infty} \frac{\partial^n}{\partial h_1 \cdots \partial h_n} \ln[Z(\{h\})]_{|z=0},
$$

(15)

where we assume that $0 < t^1 \leq t^2 \leq \cdots \leq t^n < T$.

We will show that in stationary conditions (i) $\lim_{T \to \infty} z(T) = 1$ and (ii) if $\langle \tau \rangle_T < \infty$, then

$$
\langle S(t^1)S(t^2)\cdots S(t^n) \rangle_C = \lim_{T \to \infty} \lambda T \frac{\partial^n}{\partial h_1 \cdots \partial h_n} z(T)_{h_i=0}.
$$

(16)

Using equations (13) and (14) we obtain

$$
z(T) = \int_0^{T} \frac{dt}{T} \left\{ \exp \left\{ - \sum_{i=1}^{n} h_i \Theta(t + \tau - t^i)\Theta(t^i - t) \right\} \right\}_\tau
$$

$$
= \int_0^{T} \frac{dt}{T} \left\{ \prod_{i=1}^{n} \left[ (e^{-h_i} - 1)\Theta(t + \tau - t^i)\Theta(t^i - t) + 1 \right] \right\}
$$

$$
= 1 + \sum_{k=1}^{n} \sum_{i_1 < i_2 < \cdots < i_k} \langle \max\left\{ 0, \frac{t^{i_1} - \max\{0, t^{i_k} - \tau\} - T}{T} \right\} \rangle_\tau \prod_{j=1}^{k} (e^{-h_{ij}} - 1),
$$

(17)

(18)

where we have expanded the product on the rhs of equation (17) as $\prod_{i=1}^{n} (1 + v_i) = 1 + \sum_{j=1}^{n} x_i + \sum_{i_1 < i_2 < \cdots < i_k} x_{i_1}v_{i_2}v_{i_3} \cdots v_{i_k}$. Using the Lebesgue dominated convergence theorem, we have that the average in equation (18) tends to zero in the $T \to \infty$ limit and thus $\lim_{T \to \infty} z(T) = 1$, which proves (i). Let us consider now

$$
\lim_{T \to \infty} T(z(T) - 1) = \lim_{T \to \infty} \sum_{k=1}^{n} \sum_{i_1 < i_2 < \cdots < i_k} \langle \max\{0, t^{i_1} - \max\{0, t^{i_k} - \tau\} \} \rangle_\tau \prod_{j=1}^{k} (e^{-h_{ij}} - 1).
$$

(19)

Since $\max\{0, t^{i_1} - \max\{0, t^{i_k} - \tau\} \} < \tau$ and $\lim_{T \to \infty} \max\{0, t^{i_1} - \max\{0, t^{i_k} - \tau\} \} = \max\{0, \tau - (t^{i_k} - t^{i_1})\}$, again for the Lebesgue dominated convergence theorem, equation (19) leads to

$$
\lim_{T \to \infty} T(z(T) - 1) = \sum_{k=1}^{n} \sum_{i_1 < i_2 < \cdots < i_k} \prod_{j=1}^{k} (e^{-h_{ij}} - 1) \langle \max\{0, \tau - (t^{i_k} - t^{i_1})\} \rangle_\tau \equiv \mathcal{F}(\{h\}),
$$

(20)

and therefore we have that equation (12) becomes

$$
\lim_{T \to \infty} Z_T(\{h\}) = e^{\lambda \mathcal{F}(\{h\})}.
$$

(21)

doi:10.1088/1742-5468/2012/07/P07017
Using equations (15) and (21) we finally prove equation (16). In particular, we note that in the large time limit, the $n$-point connected time correlation function

$$\lim_{t^1 \to \infty} \langle S(t^1) S(t^2) \cdots S(t^n) \rangle_C = \lambda \langle \tau - (t^n - t^1) \rangle \Theta(\tau - (t_n - t_1)),$$

is independent of $t^2 \cdots t^{n-1} \in (t^1, t^n)$.

4. The survival times pdf

The survival time $\tau_s$ is defined as the time to local extinction of a species randomly sampled among the observed assemblage of species at a certain time $T$ (see figure 2).

We can express $\tau_s$ as a function of the random variables $t_0$ (the emergence time of that species) and $\tau$ for which the pdf is known:

$$\tau_s = t_0 + \tau - T \quad \text{if} \quad 0 < t_0 < T \quad \text{and} \quad t_0 + \tau \geq T. \quad (23)$$

We then express the survival time (ST) pdf conditional on a persistence $\tau$ as

$$p_s(t | \tau) = C \langle \delta(t - (t_0 + \tau - T)) \Theta(t_0 + \tau - T) \Theta(T - t_0) \Theta(t_0) \rangle, \quad (24)$$

where the constant $C$ ensures normalization. Solving for the ensemble average operators yields

$$p_s(t | \tau) = C \Theta(\tau - t) \Theta(t - \tau + T) \Theta(t),$$

and, by marginalizing over $\tau$, we obtain the ST pdf:

$$p_s(t) = \frac{1}{(\min(\tau, T))^{\tau_0}} \int_t^{t+T} p_{spt}(\tau) d\tau. \quad (24)$$

Without loss of generality, it can be assumed that $T$ is not affected by the boundary condition in $t = 0$, i.e. $T \to \infty$ and $p_s(t) = (1/(\tau)^{\tau_0}) \int_t^{\infty} p_{spt}(\tau) d\tau$. Particularizing now to the case of persistence distributions for linear birth–death processes (equation (3)), the survival pdf asymptotic behavior is

$$p_s(t) \propto \int_t^{\infty} \tau^{-2}((1 - x)\tau/(1 - e^{-(1-x)\tau}))^2 e^{-(1-x)\tau} d\tau \propto \frac{1}{e^{(1-x)t} - 1} \sim \begin{cases} t^{-1}, & \text{for } t \ll t^* \\ e^{-t/t^*}, & \text{for } t \gg t^* \end{cases} \quad (25)$$

where $t^* = (1 - x)^{-1}$.

5. Applications and comparison with empirical data

5.1. The Gamma species persistence time distribution

Recent results [20, 21] have shown that SPT distributions for several different kinds of ecosystems exhibit power-law behavior with an exponential-like cutoff:

$$p_{spt}(t) = A^{-1} t^{-\alpha} e^{-(1-x)t} \Theta(t - \tau_0), \quad (26)$$

where $A = (1 - x)^{\alpha - 1} \Gamma(1 - \alpha, (1 - x)\tau_0)$ is the normalization constant, and $\Gamma(a,b)$ is the incomplete Gamma function. The exponent $\alpha$ of the power law is suggested to depend on
the spatial structure of the embedding ecosystem. SPT distributions exhibit progressively smaller scaling exponents $\alpha$ for increasing constraints in the connectivity structure of the environmental matrix [20]. Specifically, numerical simulations [20] show that the exponents range from $\alpha = 2$, which corresponds to the case of global dispersal (mean field), through $\alpha = 1.91 \pm 0.01$ for a 3D lattice, $\alpha = 1.83 \pm 0.02$ in a savannah (2D lattice), and $\alpha = 1.64 \pm 0.02$ for river network topology, up to $\alpha = 3/2$ for 1D systems.

For the functional shape of the SPT distribution given by equation (26) with $\alpha \leq 2$, the following asymptotic limits are obtained:

$$P_>(t) = \int_t^\infty p_{spt}(t') \, dt' = \langle \tau \rangle_{\tau \geq t} p_{spt}(t)$$

$$= \begin{cases} \frac{1}{\Gamma(1-\alpha,(1-x)\tau_0)} e^{-(1-x)t'(1-x)t}^{-\alpha} \left( 1 - \frac{1}{t(1-x)} \right) + o \left( \left( \frac{1}{t(1-x)} \right)^2 \right), & \text{for } t \gg \tau^*, \\ (\tau_0/t)^{\alpha-1}, & \text{for } \tau_0 < t \ll \tau^* \end{cases}$$

while $P_>(t) \sim 1$ for $t \to 0$. The average persistence time follows from equation (8):

$$\langle \tau \rangle = \int_0^{+\infty} d\tau P_>(\tau) = f_\infty = \frac{\tau_0 E_{\alpha-1}((1-x)\tau_0)}{E_{\alpha}(1-x)\tau_0},$$

where $E_n(z) = \int_1^\infty e^{-zt/t^n} \, dt$ is the exponential integral function. Setting $\tau_0 = 1$ for simplicity in equation (26), the corresponding two-point correlation function derived from (22) is

$$\text{cov}_S(t) = \frac{\lambda}{(1-x)} \frac{\Gamma(2-\alpha,t(1-x)) - t(1-x)\Gamma(1-\alpha,\tau(1-x))}{\Gamma(1-\alpha,1-x)}, \quad \text{for } t > \tau_0.$$ (29)

Therefore, the covariance is a decreasing function of $t$, that for large times goes to zero, i.e., $\lim_{t \to +\infty} \text{cov}_S(t) \sim e^{-(1-x)t}t^{-\alpha} = 0$.

### 5.2. Scale-free SPT distributions

An interesting case is obtained when $x \to 1$, i.e. the system is in the scaling regime where it does not exhibit a characteristic time scale. This is typically reported when SPT of families or genera—as opposed to species—are considered, possibly measuring them from the fossil record [30]. Thus, longer time scales and long time series must be considered.

Under such assumptions, the normalized SPT pdf reads as

$$p_{spt}(t) = (\alpha - 1) \tau_0^\alpha t^{-\alpha} \theta(t - \tau_0); \quad \alpha > 1,$$ (30)

whereas the cumulative SPT distribution is $P_>(t) = (\tau_0/t)^{\alpha-1} \theta(t - \tau_0) + \theta(\tau_0 - t)$, and from equation (8) we obtain

$$f(t) = \begin{cases} \min[\tau_0, t] + \tau_0 \theta(t - \tau_0) \left( \frac{t}{\tau_0} \right)^{2-\alpha} - 1, & \alpha \neq 2; \\ \min[\tau_0, t] + \tau_0 \theta(t - \tau_0) \ln(t/\tau_0), & \alpha = 2. \end{cases}$$ (31)
For this case, and in the limit $T \gg t$, the ST pdf given by equation (24) becomes

$$p_s(t) = \begin{cases} 
\frac{2-\alpha}{T} \left( \frac{t}{T} \right)^{1-\alpha}, & \alpha \neq 2; \\
\frac{1}{t \ln (t/\tau_0)}, & \alpha = 2.
\end{cases} \quad (32)$$

We note that two different cases can be pointed out. (1) $P_{\text{stat}}(S)$ exists and depends on a microscopic time scale $\tau_0$. This implies that $\alpha > 2$, leading to $\lim_{t \to +\infty} f(t) = \tau_0((\alpha-1)/(\alpha-2)) \equiv f_\infty$. (2) For $\alpha \leq 2$ the stationary pdf $P_{\text{stat}}(S)$ does not exist. In fact in this case $f(t) \sim \tau_0(t/\tau_0)^{2-\alpha}$ for $\alpha < 2$ and $f(t) = \tau_0 \ln (\max [t, \tau_0]/(\tau_0/e))$ for $\alpha \leq 2$.

5.3. The Hinkley fish estuarine database

To test the results of the coarse-grained neutral null model, we employ a long-term monthly database for an estuarine fish ecosystem. Fish samples were collected from the cooling-water filter screens at Hinkley Point Power Station, located on the southern bank of the Bristol Channel in Somerset (UK). The data span the period from January 1981 to January 2010. A full description of the intake configuration and sampling methodology is given in [31]. A matrix $P$ is built using presence–absence records for each month during the 29 years. Each element $P_{st}$ of the matrix is equal to 1 if species $s$ is observed during month $t$; otherwise $P_{st} = 0$. The empirical persistence time is defined as the number of consecutive months in which the measurements reveal the presence of the species. The presence–absence time series thus form a vector of length $T$, where $T$ is the total number of months of observation.

Persistence time is defined as the length of a contiguous sequence of 1 s in the time series. From such time series we can thus reliably estimate the empirical average persistence time $\langle \tau \rangle$ and the species emergence rate $\lambda$. For the analyzed fish estuarine ecosystem we find $\langle \tau \rangle = 3.02$ (month) and $\lambda = 4.83$ (month$^{-1}$). Moreover, by summing over rows of the matrix $P_{st}$, we obtain the total number of observed species in month $t$, i.e., $S_t = \sum_s P_{st}$. We can thus calculate (assuming stationarity) the distribution of the number of persistent species in the system (figure 3(a)), its first moments, $\bar{S} = \sum_{t=1}^{T} S_t/T$ (figure 3(b)), and the empirical two-point correlation function (figure 4):

$$\rho(\Delta t) = \frac{\sum_{t=\Delta t+1}^{T} T(S_t - \bar{S})(S_{t-\Delta t} - \bar{S})}{(T - \Delta t)(\sigma_S^2)},$$ \quad (34)

where $\Delta t$ is the time lag and $\sigma_S^2 = (1/T)\sum_{t=1}^{T}(S_t - \bar{S})^2$, so $\rho(0) = 1$.  

doi:10.1088/1742-5468/2012/07/P07017
Figure 4. (a) Power spectrum ($\Xi[\omega_j]$) analysis of the time series $S_t$ and of the autocorrelation function $\rho(\Delta t)$ given by equation (34) (in the inset). We find a periodic behavior of the time series with period $t_p \approx 12$ month. (b) We test the analytical result for the $n$-point correlation function (red dashed line) given by equation (22) for $p_{opt}(t) = (1/\tau_0)e^{-t/\tau_0}$ via numerical simulations. The parameters used in the simulation are $T = 1000$, $\tau_0 = 2$ and $\lambda = 10$. Gray squares represents the two-point $\langle (S(t) - \bar{S})(S(t + \Delta t) - \bar{S}) \rangle$ function, while green and blue dots show the three-point correlation function $\langle (S(t) - \bar{S})(S(t + \alpha \Delta t) - \bar{S})(S(t + \Delta t) - \bar{S}) \rangle$, with $\alpha = 1/3, 2/3$, respectively. In the inset: comparison between the empirical autocorrelation function $\rho(\Delta t)$ (that does not take in account seasonal periodicity in $S_t$; black dots) and the analytical autocorrelation function predicted by the coarse-grained model using equation (29) with $(1 - x) = 0.0001$, $\alpha = 2$ and $\lambda = 4.83$ (gray solid line). The dashed gray horizontal lines represent the 5% confidence interval with respect to $\rho = 0$.

The average number of species predicted by the model can be obtained from equation (11), $\langle S \rangle = \lambda/\tau = 14.6$, whereas the standard deviation is $\sqrt{\lambda/\tau_S} = 3.8$ which is in good agreement with the observed ecosystem diversity $\bar{S} \pm \sigma_S = 14.7 \pm 3.7$ calculated from the presence–absence matrix $P_{st}$ (see figure 3(a)). The null hypothesis of a Poisson species distribution given by equation (10) is accepted by both the Kolmogorov–Smirnov test and the $\chi^2$ test within a 95% confidence interval ($P_{value} = 0.05$). This analysis suggests that species diversity data cannot be used by themselves to discriminate among different mechanisms of demographic growths, i.e., kinds of birth–death processes. Similar conclusions have been achieved using RSA data [23].
An exactly solvable coarse-grained model for species diversity

The autocorrelation reveals a relevant periodic behavior in the species times series (inset, figure 4(a)). To investigate such periodicity we study the power spectrum of the time series \( \tilde{S}_t = S_t - \bar{S}_t \), i.e. \( \Xi[\omega_j] = (1/2\pi T)|\chi(\omega_j)|^2 \), where \( \chi(\omega_j) = \sum_{t=1}^T S_t e^{i\omega_j t} \) is the Fourier transform of \( S_t \) and \( \omega_j = 2j\pi/T \) \( (j = 1, 2, \ldots, T) \) is the frequency. The spectrum exhibits a peak for \( j \approx 30 \), indicating that the period is \( t_p \approx 12 \) month (see figure 4(a)). Therefore the periodicity is a trivial effect due to seasonality of weather patterns and is not related to the fluctuations or intrinsic noise of the system dynamics [32]. To test equation (29) against empirical data, we then smooth out the periodicity due to seasonality. In order to do that, we calculate the empirical autocorrelation function only considering a specific month for every year and then averaging over all twelve months, i.e., \( \bar{\rho}^j(\Delta t) = \sum_{i=\Delta t+1}^{M_j} M_j (S_t^j - \bar{S}_t^j)(S_t^{\Delta t} - \bar{S}_t^{\Delta t})/((M_j - \Delta t)(\sigma_j^2)) \), and \( \bar{\rho} = \sum_j \bar{\rho}_j/12 \), where \( j = \{ \text{Jan, Feb, ..., Dec} \} \) and \( M_j \) is the total number of \( j \)-type months in the whole time series. As can be seen from the inset in figure 4(b), once we remove seasonality, equation (29) describes well the autocorrelation function \( \bar{\rho} \).

An analysis of the fish SPT distribution has also been carried out. We find that SPT distributions are well described by the Gamma SPT with parameters compatible with \( \alpha \approx 2 \) and \( x = 0.9999 \). This result holds with particular accuracy only when large SPT are considered. In fact, at monthly time scales, seasonality affects short SPT, increasing the slope of the first part of the SPT pdf. Due to the negligible effect of the cutoff at the monthly time scale, we find that also a power-law SPT with \( \alpha \approx 1.9 \) fits the empirical SPT pdf. Note that here we are only interested in the estimation of the input parameters of the coarse-grained neutral model, \( \langle \tau \rangle \) and \( \lambda \), as they can be calculated directly from the SPT measured time series without the need for fitting the entire SPT distribution.

6. Robustness of the results

6.1. The agreement with mean field voter model results

As we claimed in section 2, the proposed model is a coarse-grained view of the grand-canonical version of the neutral theory. Nevertheless we now show that our coarse-grained model gives a good approximation of results stemming from the mean field approximation of the voter model with speciation [16, 24, 25] or, equivalently, the infinite alleles model with mutation [17] (for details on the mapping between these two models we refer the reader to [26]). The scheme of these models is the following. Consider a local community of \( N \) individuals. At every time step a randomly selected individual in the ecosystem dies. With probability \( 1 - \nu \) the individual is replaced by a colonizing offspring of another individual, randomly selected within the ecosystem, whereas with probability \( \nu \) the offspring belongs to a new species. For \( N \) fixed and \( N \to \infty \) the above dynamics is described by a linear birth–death ME with \( b = 1 - \nu \) and \( d = 1 \) (and thus \( x = b/d = 1 - \nu \)):

\[
\frac{dp_n^{(i)}(t)}{dt} = x(n-1)p_{n-1}^{(i)}(t) + (n+1)p_{n+1}^{(i)}(t) - (xn+n)p_n^{(i)}(t) \quad \text{for } n \geq 1
\]

where \( i = 1, 2, \ldots, S \leq N \) and \( p_n^{(i)}(t) \) is the probability of the \( i \)th species having an abundance \( n \) at time \( t \). For the hypothesis of neutrality we have that \( p_n^{(i)}(t) \) can be expressed as

\[
p_n^{(i)}(t) = p_n^*(t - t_i),
\]

doi:10.1088/1742-5468/2012/07/P07017
where \( p_n^*(t) \) is the probability of a single species having abundance \( n \) at time \( t \) after its emergence (under the neutral assumption, \( p_n^*(t) \) is species invariant). The average population of the \( i \)th species, given that it is still present at time \( t \), is \( \langle n_i \rangle = \sum_{n \geq 1} np_n^*(t) / \sum_{n \geq 1} p_n^*(t) \). The mean population size at time \( t \) for the mean field voter model can be calculated as

\[
\langle n(t) \rangle = \left( \sum_{i=1}^{D(t)} \frac{\sum_{n=0}^{+\infty} p_n^*(t-t_i) n}{\sum_{i=1}^{D(t)} \sum_{n=1}^{+\infty} p_n^*(t-t_i)} \right) = \frac{\sum_{i=1}^{D(t)} \sum_{n=0}^{+\infty} p_n^*(t-t_i) n}{\sum_{i=1}^{D(t)} \sum_{n=1}^{+\infty} p_n^*(t-t_i)},
\]

where \( D(t) \) is the number of diversification events occurring up to time \( t \) and the ensemble average is over the random variables \( t_i \) and \( D(t) \). By using the fact that for all \( i, t_i \) is Poisson distributed with the same frequency \( \lambda \) and \( U_D(t) \) is the pdf of the variable \( D \) (see equation (4)), then equation (36) simplifies to

\[
\langle n(t) \rangle = \int_0^t dt' \sum_{n} n p_n^*(t-t') / \int_0^t dt' \sum_{n \geq 1} p_n^*(t').
\]

Thus, from the definition of \( p_n^*(t) \), it follows that \( \sum_{n} n p_n^*(t') = \langle n^* \rangle_t, \) or the average population of the species has more than one individual at time \( t \) from its emergence. Using equation (35), we have that it obeys the deterministic equation \( d \langle n^* \rangle_t / dt = -(1-x) \langle n^* \rangle_t \), whose solution is \( \langle n^* \rangle_t = \langle n^* \rangle_0 \exp(-(1-x)t) \) [19]. Thus we have that \( \int_0^t \langle n^* \rangle_0 \exp(-(1-x)t') dt' = \langle n^* \rangle_0 (1 - \exp(-(1-x)t))/ (1-x) \). We observe that \( \sum_{n \geq 1} p_n^*(t') \) is the probability that the species has more than one individual at time \( t \), which is the cumulative distribution of the SPT pdf \( P_\tau(t) = \sum_{n \geq 1} p_n^*(t) = \int_{+\infty}^t p_{SP}(\tau) d\tau, \) and therefore

\[
\int_0^t dt' \sum_{n \geq 1} p_n^*(t') = \int_0^t dt' \int_{+\infty}^{+\infty} p_{SP}(\tau) d\tau = \langle \min(t, \tau) \rangle \tau = f(t).
\]

Using the above relations and \( \langle n^* \rangle_0 = 1 \) (since by definition, \( p_n^*(t=0) = \delta_K(n-1) \)), we obtain

\[
\langle n(t) \rangle = \frac{1}{(1-x)f(t)} \frac{1}{t \rightarrow +\infty} \frac{1}{(1-x)(\tau)},
\]

from which it follows that \( \langle S \rangle = N/\langle n \rangle = N/(1-x) \langle \tau \rangle \). Therefore if we approximate \( N \) with the average number of individuals in the corresponding grand-canonical ensemble equation (2), i.e. \( \langle N \rangle = x \partial_x \ln Z = \lambda/(1-x) \), we find that equation (38) is the same result as we have obtained from the coarse-grained neutral model (compare with equation (11)).

Interestingly, it has also been found that, for linear birth and death processes, the survival probability function has the same asymptotic functional shape as is given by equation (25) [19]. In fact, by assuming an initial population distribution given by the Fisher log series \( P_{RSA}(n) = -x^n/(n \log(1-x)) \) [8], and defining \( p_{SP}(t|n_0) \) as the probability that a species starting with \( n_0 \) individuals is still present at time \( t \), it has been shown that [19]

\[
p_s(t) = \sum_{n_0 = 1}^{\infty} p_{SP}(t|n_0) P_{RSA}(n_0) = \begin{cases} t^{-1}, & \text{for } t \ll t^*; \\ e^{-(1-x)t}, & \text{for } t \gg t^*, \end{cases}
\]

that is, it has the same asymptotic behavior of the ST pdf as was obtained with equation (25) from the coarse-grained model.
6.2. The universal relation between persistence and survival distributions

The fact that the survival probability function has the same asymptotic behavior in two different neutral models suggests that, rather than occurring by chance, it possibly happens as a consequence of a deeper, and more general relationship between the SPT distribution and the ST probability function, valid regardless of the specific birth and death process assumptions or, in other words, independently of the functional shapes of \( b(n) \) and \( d(n) \).

To address this issue, we start by calculating the RSA at a stationary time \( T \) (with an absorbing boundary condition at \( n = 0 \)), assuming that each species has only one individual when it emerges. Regardless of the type of birth/death rate, such a relation can be written as

\[
P_{\text{RSA}}^a(n) = \lim_{T \to +\infty} \frac{1}{\sum_{n=1}^{\infty} \int_0^T du P(n, T - u | 1, 0)} \int_0^T du P(n, T - u | 1, 0)
\]

\[
= \lim_{T \to +\infty} \frac{1}{f(T)} \int_0^T du P(n, u | 1, 0).
\]  

(40)

where \( P(k, t|n, t_0) \) is the probability of finding \( k \) individuals at time \( t \) given that there are \( n \) at time \( t_0 < t \). The normalization has been calculated using equation (37).

The ST probability function is thus

\[
p_s(t) = \sum_{n=1}^{\infty} p_{\text{spt}}(t|n) P_{\text{RSA}}^a(n) = \lim_{T \to +\infty} \sum_{n=1}^{\infty} \frac{d}{dt} P(0, t|n, 0) \lambda \int_0^T du P(n, u | 1, 0).
\]

(41)

As this is at stationarity, the time-translational invariance \( P(n_2, t_2|n_1, t_1) = P(n_2, t_2 - t_1|n_1, 0) \) holds:

\[
p_s(t) = \lim_{T \to +\infty} \frac{1}{f(T)} \int_0^T dt \sum_{n=1}^{\infty} P(0, t|n, 0) P(n, 0|1, -u) du
\]

\[
= \lim_{T \to +\infty} \frac{1}{f(T)} \int_0^T dt \left[ P(0, t|1, -u) - P(0, t|0, 0) P(0, 0|1, -u) \right] du
\]

\[
= \lim_{T \to +\infty} \frac{1}{f(T)} \int_0^T dt P(0, t + u|1, 0)
\]

\[
= \lim_{T \to +\infty} \frac{1}{f(T)} \int_0^T P(0, t + u|1, 0)
\]

\[
= \lim_{T \to +\infty} \frac{1}{f(T)} [P_>(t|1, 0) - P_>(t + T|1, 0)],
\]

(42)

where \( P(0, t|1, 0) \equiv 1 - P_>(t|1, 0) \) and \( P_>(t|1, 0) = \sum_{n=1}^{\infty} P(n, t|1, 0) = \int_t^\infty p_{\text{spt}}(\tau) d\tau \). We thus obtain

\[
p_s(t) = \frac{1}{\langle \tau \rangle} \int_t^\infty p_{\text{spt}}(\tau) d\tau,
\]

(43)

and this relation is valid in general, independently of the specific master equations obeyed by \( P(n, t|1, 0) \).
7. Conclusions

In this paper we have proposed a neutral model for generic birth–death processes, where information on species abundance is subsumed by the SPT distribution, \( p_{\text{SPT}} \), associated with the ecosystem under study. The model can be seen as a coarse-grained version of the grand-canonical approach to neutral theory. This framework has two main advantages: (1) it allowed us to obtain analytical results such as the transient (and stationary) dynamics of ecosystem species richness, as well as the complete analytical description of the \( n \)-point correlation function of species diversity; (2) it provides a simple null model, incorporating all the main features of any neutral model based on birth–death processes, as a function of one \textit{a priori} distribution, namely the SPT distribution. This highlights the important role of \( p_{\text{SPT}} \), as a synthetic descriptor of the ecosystem dynamics. One of the main results obtained is given by the relation equation (41), valid for any birth and death process, between the persistence time distribution \( p_{\text{SPT}} \) and the survival probability function \( p_s \).

All the results presented are exact on the assumption that species are non-interacting. We have compared the analytical results from our model with empirical data on species diversity for an estuarine ecosystem. The analysis shows that, in spite of the minimalist assumptions of our model, complex emergent patterns in the ecosystem dynamics can be captured by the proposed coarse-grained neutral framework. This suggest that species diversity data cannot be used by themselves to discriminate among different kinds of birth–death processes. Further studies and different approaches are required to determine how species diversity patterns are related to different kinds of demographic dynamics. Two major simplifications of our analysis are the non-interacting ideal-gas-like assumption and ignoring the effects of the spatial distribution. Further research will probe what qualitative changes would arise on relaxing the mean-field-like assumption presented here, by accounting for dispersal limitation.

Appendix

In this appendix we provide some mathematical details for the results presented in the main text. We start by obtaining the probability of having \( s \) new species in the ecosystem during the time interval \([0, t]\), \( U_s(t) \), given by equation (4) in the main text. It is easy to write the ME for \( U_s(t) \), \( \dot{U}_s(t) = \lambda(U_{s-1}(t) - U_s(t)) \), i.e. new species enter the system at rate \( \lambda \). The corresponding differential equation for the generating function \( \hat{U}(z, t) = \sum_{s=1}^{\infty} z^s U_s(t) \) is \( \dot{\hat{U}}(z, t) = \lambda(z - 1)\hat{U}(z, t) \), which leads to \( \hat{U}(z, t) = e^{\lambda(z-1)\hat{U}(z, 0)} \).

Assuming the initial condition \( U_s(0) = \delta_K(s - s_0) \), then \( \hat{U}(z, 0) = z^{s_0} \) and

\[
\hat{U}(z, t) = e^{-\lambda t} z^{s_0} e^{\lambda t} = e^{-\lambda t} \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} z^{s_0 + k} = e^{-\lambda t} \sum_{s=s_0}^{\infty} \frac{(\lambda t)^{k-s_0}}{(k-s_0)!} z^k. \tag{A.1}
\]

From equation (A.1) it follows that

\[
U_k(t|s_0) = \begin{cases} 
\frac{(\lambda t)^{k-s_0}}{(k-s_0)!} \exp(-\lambda t), & \text{for } k \geq s_0; \\
0, & \text{for } k < s_0.
\end{cases} \tag{A.2}
\]
which in turn, on noting that $U_k(t) = \sum_{s_0=1}^{\infty} U_k(t|s_0) U_{s_0}(0)$, leads to equation (4). Let us derive now the probability distribution of the number of persistent species given by equation (9).

Using equations (6) and (A.2), the generating function of $P(s, t)$ reads as

$$\hat{P}(z, t) = \sum_{n=0}^{\infty} \sum_{s_0=1}^{\infty} \Theta(n - s_0) U_{s_0}(0) e^{-\lambda t} \frac{(\lambda t)^{n-s_0}}{(n-s_0)!}$$

$$\times \int_0^t \prod_{i=1}^n \frac{dt_i}{t_i} \int_0^\infty \prod_{j=1}^n \frac{d\tau_j}{\tau_j} p_{spt}(\tau_j) z^{\sum_{i=1}^{n} \Theta(t_i + \tau_i - t_i)}.$$  

(A.3)

Because of the independence of the random variables $t_i$ and $\tau_j$ we can write

$$\hat{P}(z, t) = \sum_{n=0}^{\infty} \sum_{s_0=1}^{\infty} \Theta(n - s_0) U_{s_0}(0) I(z, t)^{s_0} e^{-\lambda t} \frac{(\lambda t)^{n-s_0}}{(n-s_0)!} I(z, t)^{n-s_0},$$  

(A.4)

where we have set $t_i = t_0$ and $\tau_i = \tau$ for $i = 1, 2, \ldots, n$ and $I(z, t) \equiv \int_0^t (dt_0/t) \int_0^\infty d\tau p_{spt}(\tau) z^{\Theta(t_0 + \tau - t)}$. Finally, using $\sum_{n=0}^{\infty} x^n/n! = e^x$, and through the relation

$$z^{\Theta(t_0 + \tau - t)} = \int_0^t \Theta(t - t) z + \Theta(t - \tau) [z\Theta(t_0 - (t - \tau)) + \Theta(t - (t_0 + \tau))] dt_0$$

$$= z t \Theta(t - t) + (1 - \Theta(t - t))(z \tau + t - \tau)$$

(A.5)

we have

$$I(z, t) = \int_0^t \frac{dt_0}{t} \int_0^\infty d\tau p_{spt}(\tau) z t \Theta(t - t) + (1 - \Theta(t - t))(z \tau + t - \tau)$$

$$= (z - 1) \left( \int_0^t d\tau p_{spt}(\tau) + t(P_>(t) - 1) \right) = (z - 1)f(t) - t,$$  

(A.6)

where we have used $\int_0^t d\tau p_{spt}(\tau) = -\int_0^t d\tau \tau P_>(\tau) = -t P_>(t) + \int_0^t d\tau P_>(\tau)$ and

$$f(t) \equiv \int_0^t d\tau P_>(\tau) = \int_0^{+\infty} p_{spt}(\tau) \min[t, \tau] d\tau,$$  

(A.7)

which is equation (8) in the main text.

Substituting equations (A.8) and (A.6) in equation (A.4), together with observing that

$$\sum_{s_0=1}^{\infty} \Theta(n - s_0) U_{s_0}(0) I(z, t)^{s_0} = \hat{U} \left( 1 + \frac{z-1}{t} f(t), 0 \right),$$

(A.8)

leads to equation (7) presented in the main text.

References

[1] Chave J, Muller-Landau H C and Levin S A, Comparing classical community models: theoretical consequences for patterns of diversity, 2002 Am. Naturalist 159 1

[2] Holts D R, Emergent neutrality, 2006 Trends Ecol. Evolut. 21 531

[3] Dewar R C and Porté A, Statistical mechanics unifies different ecological patterns, 2008 J. Theor. Biol. 251 389

doi:10.1088/1742-5468/2012/07/P07017

J. Stat. Mech. (2012) P07017
An exactly solvable coarse-grained model for species diversity

[4] Zillio T, Banavar J R, Green J L, Harte J and Maritan A, *Incipient criticality in ecological communities*, 2008 Proc. Nat. Acad. Sci. USA 105 18714

[5] McGill B J, *Towards a unification of unified theories of biodiversity*, 2010 Ecol. Lett. 13 627

[6] Hubbell S, 2001 The Unified Theory of Biodiversity and Biogeography (Princeton, NJ: Princeton University Press)

[7] Bell G, *Neutral macroecology*, 2001 Science 293 2413

[8] Volkov I, Banavar J R, Hubbell S P and Maritan A, *Neutral theory and relative species abundance in ecology*, 2003 Nature 424 1035

[9] Azzele S, Pigolotti S, Banavar J R and Maritan A, *Dynamical evolution of ecosystems*, 2006 Nature 444 926

[10] Volkov I, Banavar J R, Hubbell S P and Maritan A, *Patterns of relative species abundance in rainforests and coral reefs*, 2007 Nature 450 45

[11] Vallade M and Houchmandzadeh B, *Analytical solution of a neutral model of biodiversity*, 2003 Phys. Rev. E 68 061902

[12] Houchmandzadeh B and Vallade M, *Clustering in neutral ecology*, 2003 Phys. Rev. E 68 061912

[13] Houchmandzadeh B, *Neutral clustering in a simple experimental ecological community*, 2008 Phys. Rev. Lett. 101 078103

[14] Mckane A J, Alonso D and Sole R, *Analytic solution of Hubbell model of local community dynamics*, 2004 Theor. Popul. Biol. 65 6773

[15] Condit R, *et al Beta-diversity in tropical forest trees*, 2002 Science 295 666

[16] Zillio T, Volkov I, Banavar J, Hubbell S and Maritan A, *Spatial scaling in model plant communities*, 2005 Phys. Rev. Lett. 95 098101

[17] Ewens W J, 2004 Mathematical Population Genetics (Berlin: Springer)

[18] He F and Hubbell S, *Species–area relationships always overestimate extinction rates from habitat loss*, 2011 Nature 473 308

[19] Pigolotti S, Flammini A, Marsili M and Maritan A, *Species lifetime distribution for simple models of ecologies*, 2005 Proc. Nat. Acad. Sci. USA 102 15747

[20] Bertuzzo E, Suweis S, Mari L, Maritan A, Rodriguez-Iturbe I and Rinaldo A, *Spatial effects on species persistence and implications for biodiversity*, 2011 Proc. Nat. Acad. Sci. USA 108 15836

[21] Suweis S, Bertuzzo E, Mari L, Maritan A, Rodriguez-Iturbe I and Rinaldo A, *On species persistence time distributions*, 2012 J. Theor. Biol. 303 15

[22] Gardiner C W, 2004 Handbook of Stochastic Methods: for Physics, Chemistry and the Natural Sciences (Berlin: Springer)

[23] Blyth R A and McKane A J, *Stochastic models of evolution in genetics, ecology and linguistics*, 2008 J. Stat. Mech. P07018

[24] Crow J F and Kimura M, 2009 An Introduction to Population Genetics Theory (Caldwell, NJ: The Blackburn Press)

[25] Bartle R G, 1995 *The Elements of Integration and Lebesgue Measure* (New York: Wiley-Interscience)

[26] Kubo R, 1962 J. Phys. Soc. Japan 17 1100

[27] Newman M E J and Sibani P, *Extinction diversity and survivorship of taxa in the fossil record*, 1999 Proc. R. Soc. London B 266 1593

[28] Henderson P A and Magurran A E, *Linking species abundance distributions in numerical abundance and biomass through simple assumptions about community structure*, 2010 Proc. R. Soc. B 277 1561