# Theory of neutral clustering for growing populations

Bahram Houchmandzadeh
*Laboratoire de Spectrométrie Physique, CNRS & Grenoble Universités, Boîte Postale 87, 38402 St. Martin d'Hères Cedex, France*
(Received 30 July 2009; revised manuscript received 10 September 2009; published 24 November 2009)

The spatial distribution of most species in nature is nonuniform. We have shown recently [B. Houchmandzadeh, Phys. Rev. Lett. **101**, 078103 (2008)] on an experimental ecological community of amoeba that the most basic facts of life—birth and death—are enough to cause considerable aggregation which cannot be smoothened by random movements of the organisms. This clustering, termed neutral and always present, is independent of external causes and social interaction. We develop here the theoretical groundwork of this phenomenon by explicitly computing the pair-correlation function and the variance to mean ratio of the above neutral model and its comparison to numerical simulations.

DOI: 10.1103/PhysRevE.80.051920 PACS number(s): 87.23.Cc, 05.40.−a, 89.75.Fb

## I. INTRODUCTION

A major issue of population dynamics is to determine the causes governing the spatial distribution of organisms in nature. It has been known for a long time that individuals of a given species are not randomly distributed, but form patches and aggregates (Fig. 1). Taylor *et al.* [1], for example, in their comprehensive investigations, surveyed 4000 samples from 100 species across different kingdoms and concluded that nearly all the species under review had nonuniform distribution. Since then patchy spatial distribution have been reported in an extremely wide range of biological communities, from terrestrial or marine microbes [2,3] to trees in tropical forests [4] and even to biological systems which were not thought before as ecological communities such as biofilms [5] and cancerous cells [6].

Nonuniform distributions seems surprising because random movements will eventually give rise to a uniform one [7]. Therefore, when nonuniform distributions are observed, we look for causes capable of countering the homogenizing tendency of diffusion (migration). There are two obvious causes for spatial aggregation of species in nature. The first cause is the environment heterogeneity: the patchiness of species distribution can be a read out of the environment's "hostility map" toward the given species. This line of reasoning has seen a great development during the last 20 years. Such models first formulated by Levins [8] and now generally termed metapopulation ecology [9], consider the environment as formed of multiple "niches," with organisms capable of migrating between them and repopulating niches which have gone extinct. Population density will therefore be a complicated interplay between the niches topology, migration rates between them and the capability of each niche to sustain its population. The second cause of nonuniform distribution for some species can be social interactions such as search for mate or increase in food retrieval or security; herds, swarming, fruiting body formation in Dictyostelia or myxobacteria, etc. are provoked by social interactions. In dealing with these causes, the most common models use partial differential equations modeling: populations are characterized by their density $c(\mathbf{x})$, favorable niches by some local potential $V(\mathbf{x})$, which is coupled to birth/death rates, social interactions by two-particle interactions $V(\mathbf{x},\mathbf{x}')$, random motion and migrations by a diffusion term $\nabla^2 c$; differential equations are then used to link these quantities [10].

The patchy distribution of species however is so ubiquitous that one can wonder if there aren't more fundamental processes at work. The last decade has seen the rise of alternative/complementary theories, generally termed *neutral* [11], seeking the cause of patchiness in the very nature of life: birth and death phenomena. Living species have two very unique properties: the number of individuals can change only in integer units, and the apparition of a new individual (birth) always occurs close to a parent. These two simple facts can dramatically alter what we expect from continuous models and our intuition of migration and diffusive processes.

Consider for example the case of Brownian bugs with birth and death rates $\alpha$ and $\mu$ and mobility (migration rate) $D$. The naive differential equation governing their concentration $c(\mathbf{x})$ reads as

$$\partial_t c = D\nabla^2 c + (\alpha - \mu)c, \tag{1}$$

which is reduced to a simple diffusion equation $\partial_t c = D\nabla^2 c$ if the birth and death rates are equal; in this critical case, what we expect is the homogenization of the spatial distribution of the bugs. If instead of the differential equation, the population dynamics is computed by numerically simulating the bugs *individually*, then the exact contrary is observed: the distribution becomes extremely patchy as time flows [12]. The error in using continuous equations such as Eq. (1) comes from neglecting fluctuations. A concentration $c(\mathbf{x})$ is the number of organisms contained in a small area $d\mathbf{x}$ around the position $\mathbf{x}$, normalized by the size of the area. Birth and death events however are stochastic processes and $c(\mathbf{x})$ is indeed a random variable. The quantity $c$ referred to in Eq. (1) is in fact an ensemble *average*. Its use is only meaningful if the fluctuation $\Delta c$ of the random variable is small compared to the average $c$. In many physical systems, fluctuations are indeed small and partial differential equations are used with success. In living species however, fluctuations can be much wider and put in doubt the legality of using differential equations. We have shown by direct probabilistic modelization that for the above example of critical Brownian bugs or other similar living systems, the fluctuations become arbitrarily large compared to averages, giving rise to ex-

 

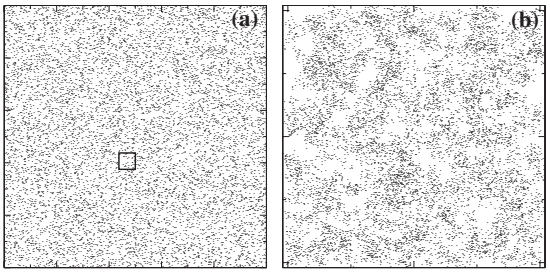


FIG. 1. Examples of (a) random uniform distribution; (b) patchy nonuniform distribution. A robust quantifier for patchiness is computed by dividing the space into contiguous squares (quadrats) of size $L$, counting the number $N_i$ of particles in each quadrat and forming the ratio of the variance to the mean ($V_m$) of these numbers. If the $V_m \approx 1$ the distribution is uniform or (Poisson) random (a). If $V_m > 1$, the distribution is nonuniform or clustered [(b), $V_m \approx 6$ computed over the squares shown in panel (a)]. $V_m < 1$ denotes an ordered distribution and is seldom encountered in natural environments.

tremely aggregated distributions [13,14]. We term this phenomena neutral clustering.

Clustered distribution in living organisms should therefore be in part caused by neutral phenomena. Consider for example the experimental model system of Fig. 2 when a small number of *D. Discoïdum* amoebae are spread uniformly in a petri dish and allowed to grow and move randomly on the petri surface [15]. Figure 2 shows the spatial distribution of the microorganisms after about eight generations. These amoebae are kept in a perfectly homogeneous environment; moreover, known chemical communications in these particular strains have been deactivated by mutations [16]. We can suspect that the extremely clustered distribution observed there is a neutral one, but this suspicion has to be confirmed by rigorous criteria as we cannot rule out *a priori* other unknown interactions. More generally, we need a mathematical tool to compare an observed spatial distribution to the expectation of a purely neutral one and deduce the importance of neutral causes. This is similar in principle to determining the bias of a coin by tossing it many times and comparing the number of heads and tails to the theoretical prediction of an unbiased one.

The object of the present article is to develop such a neutral framework for growing and randomly moving organisms, to which observations such as displayed in Fig. 2 can be compared. The most precise tool for such comparison is the pair-correlation function, i.e., the histogram of distances between all individuals. Another indicator of patchiness, the Variance to mean ratio ($V_m$) (see Fig. 1) can then be deduced from this function. If the pair-correlation function of the spatial distribution of observed species is equal to the one predicted by the neutral framework, we could then assert its neutrality ; this is indeed how we have demonstrated that the aggregation displayed in Fig. 2 is a neutral one [15]. If the pair-correlation function is markedly different from the neutral one, we can attribute the difference to other causes such as social interactions and/or environment. Moreover, the difference function is a readout of the interactions and can be used as a theoretical tool to investigate the nature of interactions or external causes.

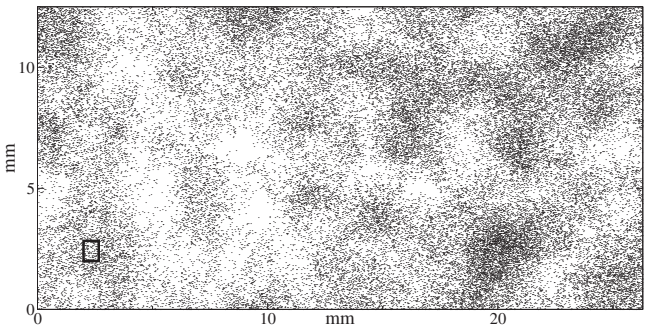


FIG. 2. Example of spatial distribution in an experimental community [15]: map of ≈55 000 pdsA-*D. Discoïdum* mutant after 72 h (7.5 generations) of growth in a petri dish. The map is obtained by scanning the bottom of the petri dish by an automated 10X objective and taking contiguous photograph; the rectangle denotes the relative size of one photograph. The position of each amoeba in each photograph is detected by an image analysis program. The result is assembled into the above map, where each dot represents one amoebae (see the Appendix). $V_m = 35$ for quadrats the size of the black rectangle. Identical results are obtained by using WT *D. Discoïdum.*

This paper is organized as follows. Section II develops the neutral framework at zero dimension, when no spatial dispersal can take place. This section stresses the importance of fluctuations due to birth and death phenomena. Spatial dispersal is introduced in Sec. III and the autocorrelation function is derived. We show there that the smoothening by random motion cannot overcome the creation of heterogeneity due to birth and death. Section IV is devoted to the derivation of the $V_m$, which is a more useful tool for experimental comparison. Concluding remarks are included in Sec. V.

## II. PROBABILISTIC FORMULATION AT ZERO DIMENSION

Before introducing the cumbersome spatial Focker-Planck equation, it can be insightful to consider a zero dimensional model where the origin of fluctuations can be simply pointed out. Consider the following thought experiment: a collection of microwells filled with nutrients; in each of the wells we deposit a precise number $n_0$ of microorganisms at time $t=0$ and then allow them randomly to duplicate with rate $\alpha$ and die with rate $\mu$. We suppose that all the microorganisms are similar. Let us call $n_i(t)$ the number of organisms in well $i$ at time $t$. Birth and death events are stochastic phenomena, therefore we cannot predict or compute $n_i(t)$ but only statistical properties like the average $\langle n(t) \rangle$ or the variance $V(t)$ over all wells. Denote $P(n_i,t)$ the probability of observing $n_i$ individuals at time $t$ in well $i$. The probability for a community of size $n$ to witness one birth (or death) and increase (or decrease) its number by one individual during a short time $dt$ is $W^+(n)dt = \alpha n dt$ [or $W^-(n)dt = \mu n dt$]. The master equation for the probability $P(n,t)$, counting per unit of time the number of ways the community can arrive at size $n$ or leave it is:

$$\frac{\partial P(n,t)}{\partial t} = [W^+(n-1)P(n-1,t) - W^+(n)P(n,t)] + [W^-(n+1)P(n+1,t) - W^-(n)P(n,t)]. \quad (2)$$

Various moments $\langle n^k \rangle$ can be extracted directly from the

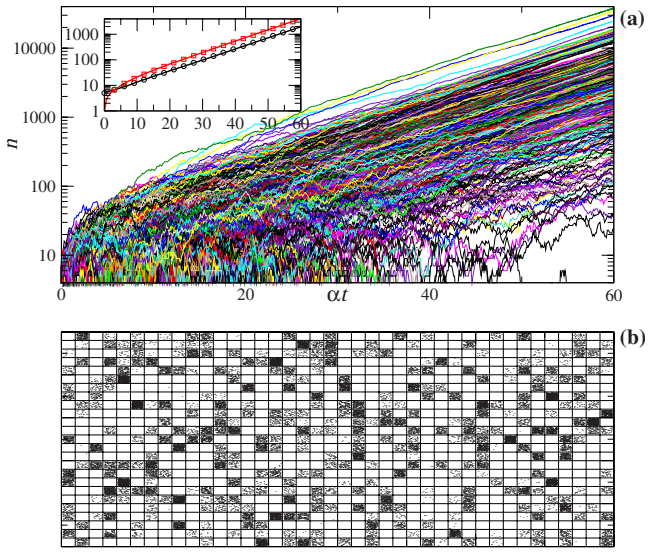


FIG. 3. (Color online) (a) A stochastic numerical simulation of the number of individuals $n_i(t)$ as of function of generation time ($\alpha t$) for 1000 wells. Each curves represents the population in one well. To generate the paths, at each time step, the number of birth and death are generated by a Poisson process. $\alpha=1$. $\mu=0.9$. [(a), inset] evolution of the average $\langle n(t)\rangle$ (black circles) and $\sqrt{V(t)}$ (red squares) and their comparison to theoretical expressions (3) and (4). (b) A visual map of the population number in each well after 60 generations. Wells are arranged into a $25\times 40$ array, each dot represents 50 individuals.

master equation (see the Appendix). The most useful ones are the average number of individuals $\langle n(t)\rangle$ and the variance, i.e., the centered second moment $V(t)=\langle n^2(t)\rangle-\langle n(t)\rangle^2$,

$$\frac{d\langle n\rangle}{dt}=(\alpha-\mu)\langle n\rangle,$$

$$\frac{dV}{dt}=2(\alpha-\mu)V+(\alpha+\mu)\langle n\rangle.$$

Note that in this community model, equations governing changes in moments of order $k$ over time involve only moments of order $\leq k$ and no moment closure approximation is needed. For the initial condition of precisely $n_0$ individuals per well $P(n,t=0)=\delta_{n,n_0}$, the solution reads as

$$\langle n(t)\rangle=n_0\exp((\alpha-\mu)t), \tag{3}$$

$$V(t)=\frac{1}{n_0}\frac{\alpha+\mu}{\alpha-\mu}\langle n\rangle(\langle n\rangle-n_0). \tag{4}$$

In the following, we will consider only growing populations $\alpha-\mu\geq 0$. In this case, the variance to mean ratio $V_m$ grows exponentially $\propto\exp((\alpha-\mu)t)$ and becomes arbitrarily large for long time. In the critical case of equality between birth and death rates, the average remains constant but the variance grows linearly: $V(t)=(\alpha+\mu)n_0t$. Note that the variance to mean ratio does not depend on $n_0$, but the coefficient of variation does: $\sqrt{V}/\langle n\rangle\propto 1/\sqrt{n_0}$.

Figure 3(a) shows a numerical simulation of the number of individuals $n_i(t)$ as a function of time in 1000 wells, where it is obvious that fluctuations dominate the behavior of the system and the number of individuals in the wells after some 60 generations can vary over 4 orders of magnitude. Figure 3(b) is a visual representation of the situation after 60 generations. We observe here on a simple example, that representing the population dynamics of a growing population by the average $\langle n(t)\rangle$ can be misleading.

## III. SPATIAL NEUTRAL CLUSTERING WITH MIGRATIONS

Figure 3(b) displays the extreme variability from well to well of a growing population. What would happen if there were no separation between wells and microorganisms were allowed to freely disperse? We know that dispersal will act as a smoothening force and the individuals will flow from high-concentration regions to low-concentration ones. However, we will show below by computing exactly the pair-correlation function, that this smoothening cannot overcome the patchiness induced by birth and death phenomena.

The pair-correlation function, i.e., the probability of finding an organism at a distance $\mathbf{r}$ of another one or the normalized histogram of all the distances between pairs in the ecosystem, is the main mathematical tool to investigate spatial clustering [17]. A closely related quantity used in ecological literature is the $\beta$-diversity [18], the rate at which similarities between quadrats decreases as a function of distance (see the Appendix). A uniform random (Poissonian) spatial distribution will have a flat pair-correlation function ; clustered distribution on the other hand are enriched by short pair distances and their pair-correlation function will show a peak at small scales.

For the sake of simplicity, we will derive the statistical properties of the growing population first in one dimension, the generalization to higher dimensions will be straightforward. Consider a collection of Brownian individuals moving randomly with diffusion coefficient (migration rate) $D/2$, duplicating with rate $\alpha$ and dying with rate $\mu$. By randomly moving Brownian individuals we mean that if the relative position $x(t)$ of any particular individual were followed during time, then $\langle x(t)\rangle=0$ and $\langle x^2(t)\rangle=Dt$. Let us divide the space into cells of size $\ell$ and call $n_i(t)$ the number of individuals in cell $i$ at time $t$. We will consider in this article infinitely wide space, neglecting finite-size corrections ; this approximation is valid when the physical size of the system is large compared to the migration length (see below). The probability for the change in the number of individuals in cell $i$ will be given as before by the birth and deaths which occur in this cells and the additional effect of migrations and exchange between neighboring cells.

To reduce the cumbersome notations, we will use the vector $\mathbf{n}=(\ldots,n_i,\ldots)$ to represent the state of the whole system; the vector $a_i\mathbf{n}=(\ldots,n_i-1,\ldots)$ to represent the same state as $\mathbf{n}$ except that one individual has been subtracted from cell $i$; $a_i^\dagger\mathbf{n}=(\ldots,n_i+1,\ldots)$ to represent the addition of one individual to cell $i$; the number of organisms $n_i$ in cell $i$ will be noted $\eta_i\mathbf{n}$. One birth in cell $i$ will transform $\mathbf{n}$ into $a_i^\dagger\mathbf{n}$ with the probability $W^+(\mathbf{n},i)dt=\alpha\eta_i\mathbf{n}dt$; one migration from cell

$i-1$ to cell $i$ transforms $\mathbf{n}$ into state $a_i^\dagger a_{i-1}\mathbf{n}$ with the probability $W^m(\mathbf{n},i-1,i)dt=\beta\eta_{i-1}\mathbf{n}$ where $\beta=D/(2\ell^2)$.

Summing up all the contributions, we can now write the master equation for the probability $P(\mathbf{n},t)$ of observing the system in state $\mathbf{n}$ at time $t$,

$$\begin{aligned}\frac{\partial P(\mathbf{n},t)}{\partial t} &= \sum_i W^+(a_i\mathbf{n},i)P(a_i\mathbf{n},t) - W^+(\mathbf{n},i)P(\mathbf{n},t)\\ &+ W^-(a_i^\dagger\mathbf{n},i)P(a_i^\dagger\mathbf{n},t) - W^-(\mathbf{n},i)P(\mathbf{n},t)\\ &+ W^m(a_{i-1}^\dagger a_i\mathbf{n},i-1,i)P(a_{i-1}^\dagger a_i\mathbf{n},t)\\ &- W^m(\mathbf{n},i,i+1)P(\mathbf{n},t) + W^m(a_{i+1}^\dagger a_i\mathbf{n},i+1,i)\\ &\times P(a_{i+1}^\dagger a_i\mathbf{n},t) - W^m(\mathbf{n},i,i-1)P(\mathbf{n},t). \qquad (5)\end{aligned}$$

Averages $\langle n_k\rangle$ and correlations $\langle n_k n_{k+m}\rangle$ can be extracted directly from the above equation as before (see the Appendix). We will assume uniform initial condition of exactly $\nu_0$ particles on all site at time $t=0$ ($P(\mathbf{n},0)=\Pi_i\delta_{n_i,\nu_0}$); the system is therefore symmetric upon translation and mirror reflection. The averages which do not depend on the position of the site $\langle n_k\rangle=\langle n\rangle$ read as

$$d\langle n\rangle/dt=(\alpha-\mu)\langle n\rangle. \qquad (6)$$

Correlations are only functions of the distance between sites $\langle n_k n_{k+m}\rangle=f(|m|)$. The equations governing the evolution of $u_m=\langle n_k n_{k+m}\rangle-\langle n\rangle^2-\langle n\rangle\delta_{m,0}$ (the centered correlations without self contribution) read as (Appendix)

$$\frac{du_m}{dt}=2(\alpha-\mu)u_m+2\beta(-2u_m+u_{m-1}+u_{m+1})+2\alpha\langle n\rangle\delta_{m,0}. \qquad (7)$$

The term $2\alpha\langle n\rangle\delta_{m,0}$ is the correlation creation due to births reflecting the fact that new organisms appear close to their parents. These equations can be directly solved in terms of combination of Bessel $I_m(t)$ and exponentials. It is however more fruitful to take the continuum limit of cell size $\ell\to 0$ and use $n_k=\ell c(x)$, where the cells are now enumerated with the continuum index $x=\ell k$, and $c(x,t)$ is the concentration at position $x$ and time $t$. Denoting $c(t)=\langle c(x,t)\rangle$, Eq. (6) transforms into

$$dc/dt=(\alpha-\mu)c$$

and therefore $c(t)=c_0\exp[(\alpha-\mu)t]$. Using the centered, *normalized* pair-correlation function

$$g(x,t)=(\langle c(y)c(y+x)\rangle-c^2-c\,\delta(x))/c^2, \qquad (8)$$

where $\delta(x)$ is the Dirac distribution, Eq. (7) transforms into

$$\frac{\partial g}{\partial t}=D\frac{\partial^2 g}{\partial x^2}+\frac{2\alpha}{c}\delta(x). \qquad (9)$$

There is no complication in generalizing this derivation to higher dimensions: the scalar position $x$ and the operator $\partial^2/\partial x^2$ have to be replaced by the vectorial position $\mathbf{r}$ and the Laplacian $\Delta$. It will be more fruitful to formulate the above equation in the natural scales of the problem: the generation time $\alpha^{-1}$ and the migration length $\lambda=\sqrt{2D/\alpha}$, i.e., the average distance one organism travel during one generation (or the mean-square distance of seed dispersion for plants communities). These units allow one to compare easily different experiments. Measured in these units, Eq. (9) reduces to the dimensionless equation

$$\frac{\partial g}{\partial t}=\frac{1}{2}\Delta g+\frac{2}{c}\delta(\mathbf{r}). \qquad (10)$$

The solution of Eq. (10), with the initial condition $g(\mathbf{r},0^-)=0$ (homogeneous initial distribution) reads as

$$g(\mathbf{r},t)=\int_0^t\frac{1}{(2\pi s)^{d/2}}\exp\left(-\frac{r^2}{2s}\right)\frac{2}{c(t-s)}ds \qquad (11)$$

$$=\frac{1}{c_0}\int_0^t\frac{2}{(2\pi s)^{d/2}}\exp\left(-\frac{r^2}{2s}\right)\exp((1-\mu/\alpha)(s-t))ds. \qquad (12)$$

For the critical case $\alpha=\mu$, the above expression has an exact expression,

$$g(\mathbf{r},t)=(1/c_0)\pi^{-d/2}r^{2-d}\Gamma(-1+d/2,r^2/2t)$$

(where $\Gamma$ is the incomplete gamma function) and displays logarithmic divergence at $d=2$ for large times [13].

For growing population ($\alpha>\mu$), Fig. 4 displays a spatial density map of the individual-based numerical simulation of a population and its pair-correlation function (see the Appendix) compared to the theoretical predictions ; it can be observed that they perfectly agree.

The above computation allows one to assess the importance of neutral clustering in natural or experimental ecosystems. Every population will display a certain degree of aggregation due to neutral phenomena and the observation of clustered population per se should not be surprising. It is only by comparison to neutral clustering than the importance of other phenomena (environment, social or interspecies interaction, etc.) can be measured. This is how we have been able to demonstrate that the clustering observed in an experimental ecosystem of Dictyostelia [15] was purely neutral and not due to such phenomena as chemical communications between microorganisms.

## IV. VARIANCE TO MEAN RATIO

The pair-correlation function can be accurately measured only when the data originate from numerical simulations or carefully designed experiments. In general, it is impossible on the field to record the position of *each* individuals. The general approach is to divide the space into squares (quadrats) of size $L$ and measure the *number* $N_i$ of individuals in each quadrat $i$. This measurement is done by various methods such as direct counting or optical methods (spectrometry methods, fluorescence, flow citometry on samples, etc.). Various statistics can then be performed to quantify the distribution heterogeneity [19]. The most robust measurement is the variance to mean ratio ($V_m$): A random (Poissonian) distribution of organism will have $V_m=1$; a significant deviation

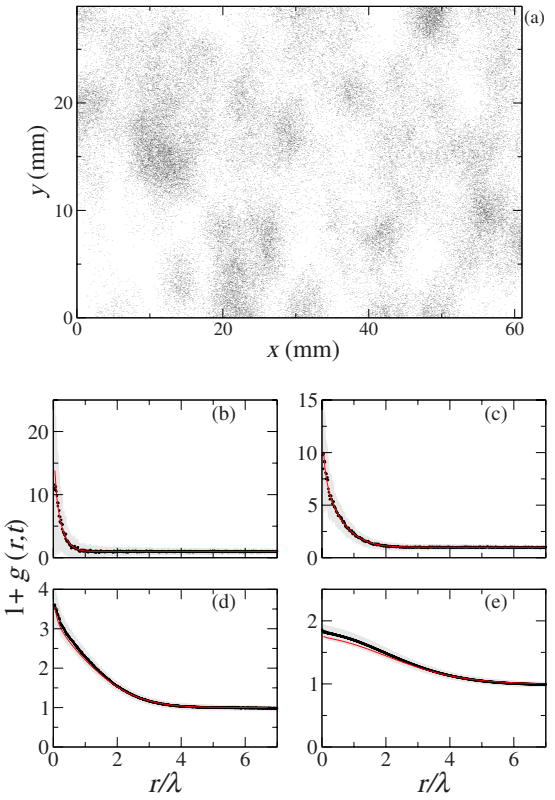


FIG. 4. (Color online) Numerical simulation of a growing population and its pair-correlation function. Data for numerical simulations are similar to the experimental ecosystems of [15]: $\alpha=1$, $\mu=0$, $D=1/2$, and $c_0=9.6\times10^{-2}$; time scales are in generation units, spatial scales in migration length units. (a) Spatial map of the organism at $t=6$; each dot represent one organism. (b)–(e) pair-correlation functions for four times: (b) $t=0.2$; (c) $t=1$; (d) $t=3$; and (e) $t=6$. Circles and shaded area: average and ± one standard deviation computed on 25 numerical simulations; (red online) solid lines: theoretical predictions. The small difference at short distances for large times (e) is due to the finite size of the ecosystem.

beyond this value signify clustering and the dependence of $V_m$ upon time and scale will inform us on the processes governing the ecosystem [20,21]. The $V_m$ of the spatial distribution of a growing population such as described above reaches extremely high values. For example, for the population displayed in Fig. 4(a), The $V_m$ computed over a quadrat $L=1$ (in migration length unit) is $\approx 30$.

In the case studied here, as we know the expected pair-correlation function, we can derive exactly the $V_m$ of the growing community. Consider in $d$ dimension the number of individual $N$ contained in a ($d$-dimensional) quadrat $\mathcal{A}$ of size $L^d$. This random variable reads as

$$N = \int_{\mathcal{A}} c(\mathbf{r},t)d\mathbf{r}$$

and its variance to mean ratio $V_m(L,t)$ is

$$\frac{\langle N^2\rangle - \langle N\rangle^2}{\langle N\rangle} = \frac{1}{L^d c(t)}\int_{\mathcal{A}\times\mathcal{A}} [\langle c(\mathbf{r},t)c(\mathbf{r}',t)\rangle - c(t)^2]d\mathbf{r}d\mathbf{r}'$$

$$= 1 + \frac{c(t)}{L^d}\int_{\mathcal{A}\times\mathcal{A}} g(\mathbf{r}-\mathbf{r}',t)d\mathbf{r}d\mathbf{r}'$$

$$= 1 + \frac{1}{L^d}\int_{\mathcal{A}\times\mathcal{A}}\int_0^t \frac{2}{(2\pi s)^{d/2}} \times \exp\left[-\frac{(\mathbf{r}-\mathbf{r}')^2}{2s}\right]\frac{c(t)}{c(t-s)}dsd\mathbf{r}d\mathbf{r}'.$$

Note that the factor $c(t)/c(t-s)=\exp[(1-\mu/\alpha)s]$ removes any dependency on the initial concentration $c_0$. By first computing the integral over spatial variables and denoting $h(u)$ the function (see the Appendix)

$$h(u) = \int_0^1 (1-x)\exp(-x^2/2u)dx \tag{13}$$

the $V_m$ reads as

$$V_m(L,t) = 1 + (2L)^d \int_0^t \frac{2\exp[(1-\mu/\alpha)s]}{(2\pi s)^{d/2}} h^d(s/L^2)ds.$$

For the critical case $\alpha=\mu$, $V_m(L,t)$ diverges at $d\leq 2$ for large times $t \gg L^2$, even though the average concentration $c(t)=c_0$ remains constant; the divergence is logarithmic at $d=2$.

For a growing population $\alpha>\mu$ the $V_m(L,t)$ grows as $L^d(1-\mu/\alpha)^{-1}\exp[(1-\mu/\alpha)t]/t^{d/2}$ for large times $t \gg L^2$ (see the Appendix and Fig. 6 inset). On the other hand, for large scale $L \gg \sqrt{t}$, the $V_m(L,t)$ saturates at

$$1 + 2(1-\mu/\alpha)^{-1}(e^{(1-\mu/\alpha)t} - 1). \tag{14}$$

Figure 5 displays the behavior of the $V_m$ both as a function of time and scale, and its comparison to individual-based numerical simulation of the population. As it can be seen, the $V_m$ reaches very high values for a growing population, much beyond the Poissonian distribution of $V_m=1$, indicating the clustered nature of population distribution.

The study of the dependence of the $V_m$ as a function of scale is similar in principle to some ecological measurements such as fractal index computation [22] or the presence-absence map [23]. It has however a more intuitive interpretation [24].

## V. CONCLUDING REMARKS

In the present work, we have shown how the very forces behind life, birth and death phenomena, give rise to organisms aggregation and strongly clustered spatial distributions ($V_m \gg 1$). Birth creates correlations at short distances and death removes correlations at all distances: organisms are born close to a parent and die everywhere. This clustering cannot be overcome and smoothened by random motions of the organisms and migrations, specially in dimensions one (communities along a shore line, a river bed, etc.) and two (the major forms of life, where the horizontal extension of the community is much larger than its vertical one). The

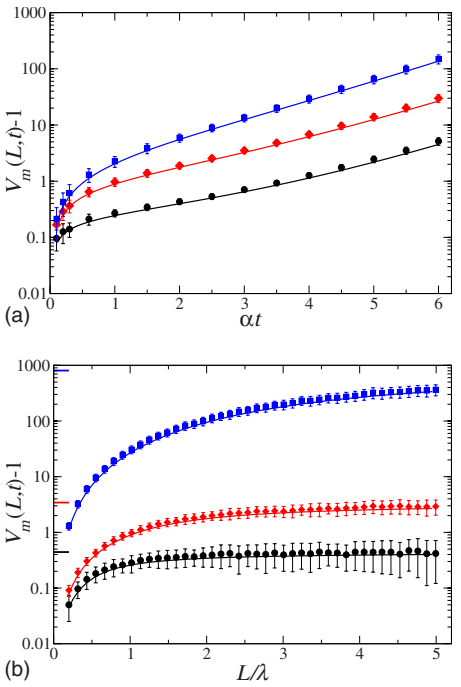


FIG. 5. (Color online) The theoretical $V_m$ (solid curves) and its comparison to the function computed on individual-based numerical simulations (symbols). Data represent averages over 25 simulated populations with the same parameters as in Fig. 4, and error bars are ± one standard deviation over the same samples. (a) $V_m$ as a function of time $t$ for three different values of $L$: $L$=0.4 (circles); $L$=1 (diamonds); and $L$=2.5 (squares). (b) $V_m$ as a function of scale $L$ for three different values of time $t$: $t$=0.2 (circles); $t$=1 (diamonds); and $t$=6 (squares). Thick solid bars on the left side represent saturation values given by expression (14).

neutral clustering of organisms was shown to be a relevant factor for ecological communities in a carefully designed experiment where all the parameters could be precisely measured [15] and the present work develops the theoretical ground of the neutral clustering phenomena by computing exactly the pair-correlation function and the variance to mean ratio. The observation of the widespread spatially aggregated distribution of organisms in ecological communities therefore should not be a surprise by itself. In observing the aggregation of organisms, one should first assess the amount of aggregation expected from neutral causes; when the aggregation deviates significantly from neutral ones one should invoke external causes such as environment heterogeneity. It would be an extremely naive view to claim that aggregation in ecological communities are mainly due to neutral causes; the aim of the present work is only to bring to the forefront neutral causes as one among the many factors relevant for the spatial distribution of species, a fact which is mostly ignored in the ecological literature.

The aggregation we have described here belongs to the class of *multiplicative noise* induced phenomena [25]; our work applies to growing populations such as plankton blooms, populations extending their geographical range due to environmental change [26] or more generally when there are successive or seasonal events of extinctions a recolonization. A natural extension of the present work would be to predator-prey communities where large stochastic fluctuations can lead to extinction and where the spatial extension is believed to play an important role in stabilizing the community [27,28].

## APPENDIX: MATHEMATICAL AND EXPERIMENTAL DETAILS

### 1. Extracting moments from the master equation

Consider the master Eq. (2). Let us multiply both sides of the equation by $f(n)$ and sum over all $n$. The left-hand side will give us $d\langle f(n)\rangle/dt$. On the right-hand side, we will have terms such as $\Sigma_n f(n)W^+(n-1)P(n-1,t)$ which, upon a change of index $n\to n+1$ transforms into $\Sigma_n f(n+1)W^+(n)P(n,t)$. Grouping all terms after such transformations, we get

$$\frac{d\langle f(n)\rangle}{dt} = \langle (f(n+1)-f(n))W^+(n)\rangle - \langle (f(n)-f(n-1))W^-(n)\rangle. \tag{A1}$$

Note that for $f(n)=n^k$, $f(n)-f(n-1)$ is a polynomial of order $k-1$; the jump rates are linear in $n$, therefore both sides of Eq. (A1) are polynomials of the same degree.

The spatial case seems more complicated because the master Eq. (5) involves a summation over all sites. The notations introduced in Sec. III however allow us to manipulate it by basically the same technique. These notations are a simpler version of those used in quantum field theory [29] because the creation $a_i^\dagger$ and annihilation operators $a_i$ as defined here commute: $a_i a_i^\dagger \mathbf{n} = a_i^\dagger a_i \mathbf{n}$. If we are interested in the average of $f(\eta_k \mathbf{n})$, i.e., a function of $n_k$ where $k$ is a given site, we multiply both sides of master Eq. (5) by $f(\eta_k \mathbf{n})$ and sum over all $\mathbf{n}$. The left-hand side will give, as before, $d\langle f(n_k)\rangle/dt$. For the right-hand side, consider for example the first part of the first line

$$\sum_{\mathbf{n}} \sum_i f(\eta_k \mathbf{n}) W^+(a_i \mathbf{n}, i) P(a_i \mathbf{n}, t)$$

which upon changing $\mathbf{n}\to a_i^\dagger \mathbf{n}$ and grouping with the second part of the first line (its conjugate expression), transforms into

$$\sum_{\mathbf{n}} \sum_i (f(\eta_k a_i^\dagger \mathbf{n}) - f(\eta_k \mathbf{n})) W^+(\mathbf{n}, i) P(\mathbf{n}, t). \tag{A2}$$

Because $\eta_k a_i^\dagger \mathbf{n} = \eta_k \mathbf{n} + \delta_{k,i}$, there is only one nonvanishing term in summation (A2) over all sites which then reads as

$$\sum_{\mathbf{n}} (f(\eta_k \mathbf{n}+1) - f(\eta_k \mathbf{n})) W^+(\mathbf{n}, k) P(\mathbf{n}, t) = \langle (f(\eta_k \mathbf{n}+1) - f(\eta_k \mathbf{n})) W^+(\mathbf{n}, k)\rangle.$$

This process has to be applied to all lines of master Eq. (5). Averages are computed by using the function $f(\eta_k \mathbf{n}) = \eta_k \mathbf{n}$; correlations are slightly more cumbersome; they are computed by using the function $f(\eta_k \mathbf{n}\, \eta_l \mathbf{n}) = \eta_k \mathbf{n}\, \eta_l \mathbf{n}$ where the cases $k=l$, $|k-l|=1$, and $|k-l|>1$ have to be treated separately.

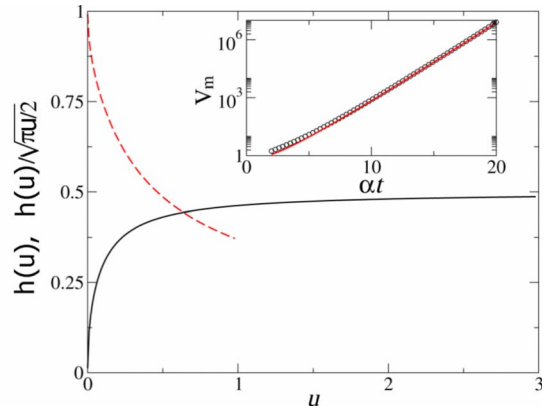


FIG. 6. (Color online) The function $h(u)$ (solid curve) and $h(u)/\sqrt{\pi u/2}$ (dotted curve). Inset: comparison of the function $V_m(1,t)$ (circles) to its long-time approximation $\exp(t)/\pi t$ (solid curve); $d=2$ and $\mu=0$.

#### 2. $V_m$ computation

The function $h(u)$ used in the $V_m$ has an explicit form,

$$h(u) = (e^{-1/2u} - 1)u + \sqrt{\pi u/2}erf(1/\sqrt{2u}),$$

$h(u)$ is a growing function of its argument which is $\simeq\sqrt{\pi u/2}$ for $u \ll 1$ and saturates at 1/2 for $u \gtrsim 1$ (Fig. 6). These approximations are used to obtain large scale and large time limits of the $V_m(L,t)$.

#### 3. Individual-based numerical simulations and correlation computation

Numerical simulation of discrete organisms present no particular difficulties: (i) select a discretization time $dt$ and generate an array of $n$ elements containing the position of each organism, distributed homogeneously at $t=0$ inside a given area; (ii) At each time step, generate two Poisson random numbers $N_1$ of parameter $\mu ndt$ and $N_2$ of parameter $\alpha ndt$; eliminate $N_1$ elements from and add $N_2$ elements to the array; each new element inherits its positions from a parent chosen at random; (iii) for each element in the array, generate two uniform random numbers $d_x$ and $d_y \in [-\sqrt{3Ddt}, \sqrt{3Ddt}]$ and increment its positions by these numbers; and (iv) repeat these operations until a specified amount of time has been spent.

#### 4. Pair correlation and $\beta$ diversity

The $\beta$-diversity measures the degree of similarity between two plots in a multicomponent ecosystem and is defined as [30]

$$F(\mathbf{r}) = \left\langle \sum_{i=1}^{S} n_i^A n_i^B \right\rangle,$$

where the average is taken over all plots $A$ and $B$ separated by $\mathbf{r}$, the sum is taken over all $S$ species present in the ecosystem, $n_i^A$ (or $B$) is the number of individuals of species $i$ in plot $A$ (or $B$). This is the probability that two individuals chosen at random in two plots belong to the same species. The pair-correlation function we use here [Eq. (7)] corresponds to the case $S=1$. The function $g(\mathbf{r},t)$ which is derived [Eq. (8)] is proportional to the probability of finding one individual at distance $\mathbf{r}$ of a given one.

#### 5. Experimental details

The experimental work this paper refers to has been already published [15]. For consistency, We give here a brief sketch of the measurements which leads to maps as displayed in Fig. 2. (a) Axenic *Dictyostelia* are spread and put to growth in a 50 mm petri dish under controlled conditions. At regular time intervals, the petri dish is scanned by a computer controlled microscope objective and contiguous photographs ($660\times 832$ $\mu$m) are taken. Paving the area of Fig. 2 requires 600 photographs. (b) A home made image analysis program detects the position of each *Dictyostelium* in each photograph. A global map such as Fig. 2 is reconstituted by combining *Dictyostelia* positions from all photographs in a sequence. By repeating the operation (4–12 h intervals, depending on the growth rate), a spatiotemporal map is constituted. One experiment can require up to 10 000 photos. (c) To measure the Brownian diffusivity (migration rate), the movements of Dictyostelia in a microscope frame are recorded and the cells trajectories are reconstituted. The mean-square linear displacement function $\langle x^2\rangle$ of all trajectories is computed for various time intervals $t$. The diffusion coefficient (migration rate) $D$ is recovered from the linear regression $\langle x^2\rangle = Dt$. (d) The growth rate $\alpha$ is measured by reporting the total number $N$ of *Dictyostelia* in each map versus time: $N=N_0\exp(\alpha t)$.

Axenic (Ax2) *Dictyostelia* were a gift from Pr. Franz Bruckert. *pdsA* mutants were obtained from the Dicty Stock Center (Columbia University, R. Kessin and J. Franke). During their growth, they are maintained at 22°C and stable humidity. The base liquid growth medium is HL5 (Division time $T_2=10$ h); for 1 L, the medium contains: bacteriological peptone (L35 Oxoïd) 14.30 g; yeast extract (L21 Oxoïd) 7.15 g; maltose 18 g; Na2HPO4 0.51 g; KH2PO4 0.48 g; dihydrostreptomycin 0.25. HL5 is an extremely rich medium were Dictyostelia have high growth rates. In order to vary the growth speed, the amino-acid source is limited either by reducing peptone (X/5, medium A, $T_2=12.8$ h) or totally removed (medium C, $T_2=23.1$ h). It was found that aged medium A is useful for very slow growth (medium D $T_2=28.9$ h). All growth media are 0.2 $\mu$m filtered to simplify image analysis. The medium is never limiting and the growth remains exponential during the entire experiment.

[1] L. R. Taylor, I. P. Woiwod, and J. N. Perry, J. Anim. Ecol. **47**, 383 (1978).
[2] A. P. Martin, M. V. Zubkov, P. H. Burkill, and R. J. Holland, Biol. Lett. **1**, 366 (2005).
[3] J. L. Green, A. J. Holmes, M. Westoby, I. Oliver, D. Briscoe, M. Dangerfield, M. Gillings, and A. J. Beattie, Nature (London) **432**, 747 (2004).
[4] R. Condit *et al.*, Science **295**, 666 (2002).
[5] T. J. Battin, W. T. Sloan, S. Kjelleberg, H. Daims, I. M. Head, T. P. Curtis, and L. Eberl, Nat. Rev. Microbiol. **5**, 76 (2007).
[6] I. González-García, R. V. Solé, and J. Costa, Proc. Natl. Acad. Sci. U.S.A. **99**, 13085 (2002).
[7] A. Okubo and S. A. Levin, *Diffusion and Ecological Problems: Modern Perspectives* (Springer, New York, 2001), Chaps. 1 and 2, pp. 1–27.
[8] R. Levins, American Entomol. **15**, 237 (1969).
[9] *Ecology, Genetics and Evolution of Metapopulations*, edited by I. Hanski and O. E. Gaggiotti (Academic, New York, 2004).
[10] A. Okubo and J. G. Mitchell, *Diffusion and Ecological Problems: Modern Perspectives* (Springer, New York, 2001), Chap. 9, p. 268.
[11] G. Bell, Science **293**, 2413 (2001).
[12] W. R. Young, A. J. Roberts, and G. Stuhne, Nature (London) **412**, 328 (2001).
[13] B. Houchmandzadeh, Phys. Rev. E **66**, 052902 (2002).
[14] B. Houchmandzadeh and M. Vallade, Phys. Rev. E **68**, 061912 (2003).
[15] B. Houchmandzadeh, Phys. Rev. Lett. **101**, 078103 (2008).
[16] R. Sucgang, C. J. Weijer, F. Siegert, J. Franke, and R. H. Kessin, Dev. Biol. **192**, 181 (1997).
[17] B. Bolker and S. W. Pacala, Theor. Popul. Biol. **52**, 179 (1997).
[18] J. C. Nekola and P. S. White, J. Biogeo. **26**, 867 (1999).
[19] M. G. Turner, Annu. Rev. Ecol. Evol. Syst. **36**, 319 (2005).
[20] L. Benedetti-Cecchi, Ecology **84**, 2335 (2003).
[21] A. Mahadevan and J. W. Campbell, *Handbook of Scaling Methods in Aquatic Ecology: Measurement, Analysis, Simulation* (CRC, Boca Raton, FL, 2003), Chap. 14, p. 215.
[22] T. W. Crawford, J. A. Commito, and A. M. Borowik, Landscape Ecol. **21**, 1033 (2006).
[23] J. P. M. Witte and P. J. J. F. Torfs, Ecography **26**, 60 (2003).
[24] J. M. Halley, S. Hartley, A. S. Kallimanis, W. E. Kunin, J. J. Lennon, and S. P. Sgardelis, Ecol. Lett. **7**, 254 (2004).
[25] C. Van den Broeck, J. M. R. Parrondo, R. Toral, and R. Kawai, Phys. Rev. E **55**, 4084 (1997).
[26] R. J. Wilson, C. D. Thomas, R. Fox, D. B. Roy, and W. E. Kunin, Nature (London) **432**, 393 (2004).
[27] M. J. Keeling, H. B. Wilson, and S. W. Pacala, Science **290**, 1758 (2000).
[28] S. P. Ellner, E. McCauley, B. E. Kendall, C. J. Briggs, P. R. Hosseini, S. N. Wood, A. Janssen, M. W. Sabelis, P. Turchin, R. M. Nisbet, and W. W. Murdoch, Nature (London) **412**, 538 (2001).
[29] D. C. Mattis, Rev. Mod. Phys. **70**, 979 (1998).
[30] J. Chave and G. Leigh, Theor. Popul. Biol. **62**, 153 (2002).