



HAL
open science

Theory of neutral clustering for growing populations.

Bahram Houchmandzadeh, Marcel Vallade

► **To cite this version:**

Bahram Houchmandzadeh, Marcel Vallade. Theory of neutral clustering for growing populations.. Physical Review E: Statistical, Nonlinear, and Soft Matter Physics, 2009, 80 (5), pp.051920. 10.1103/PhysRevE.80.051920 . hal-00633709

HAL Id: hal-00633709

<https://hal.science/hal-00633709>

Submitted on 19 Oct 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Theory of Neutral Clustering for growing populations.

Bahram Houchmandzadeh

CNRS & Grenoble Universités, Lab. Spectrométrie Physique, BP87, 38402 St-Martin d'Hères Cedex, France.

The spatial distribution of most species in nature is nonuniform. We have shown recently (PRL 101:078103) 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 smothered 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.

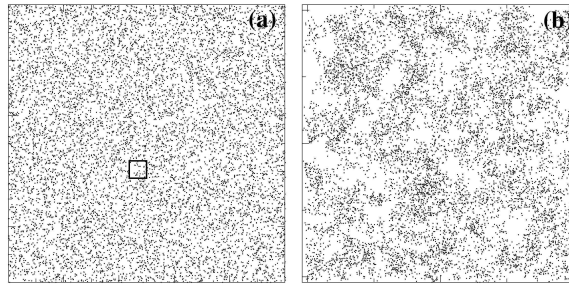


Figure 1. Examples of (a) random, uniform distribution ; (b) patchy, non-uniform distribution. A robust quantifier for patchiness is computed by dividing the space into contiguous squares (quadrats) of size L , counting the number N_i of particles in each quadrats and forming the ratio of the variance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

I. INTRODUCTION.

A major issue of ecology and 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 (Figure 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 non-uniform 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].

Non-uniform distributions seems surprising because random movements will eventually give rise to a uniform one. Therefore, when non-uniform distributions are observed, we look for causes capable of countering the homogenizing tendency of diffusion. 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 Levin[7] and now generally termed metapopulation biology[8], 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 community. The second cause of non-uniform 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 the same methodologies as in physics : 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 by a diffusion term $\nabla^2 c$, ...; differential equations are then used to link these quantities[9].

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*[10], 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 diffusive processes.

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

$$\partial_t c = D \nabla^2 c + (\alpha - \mu) c \quad (1)$$

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 homogenisation 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[11]. The error in using continuous equations such as (1) comes from neglecting fluctuations. A concentration $c(\mathbf{x})$ is the number of organisms contained in a small

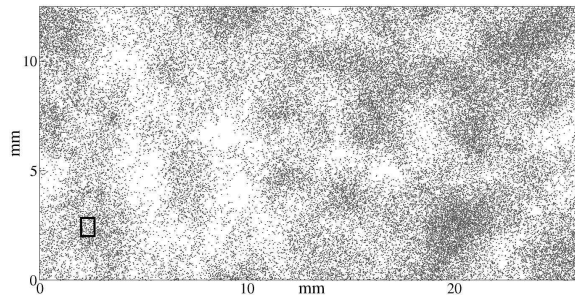


Figure 2. Example of spatial distribution in an experimental community[14] : map of ≈ 55000 pdsA- *D. Discoïdum* mutant after 72 hr (7.5 generations) of growth in a petri dish. The map is obtained by scanning the bottom of the petri dish by an automated 10X objective and taking contiguous photograph; the rectangle denotes the relative size of one photograph. The position of each amoebae in each photograph is detected by an image analysis program. The result is assembled into the above map, where each dot represents one amoebae. VMR = 35 for quadrats the size of the black rectangle.

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 equation (1) is in fact an ensemble *average*. Its use is only meaningful if the fluctuation Δc 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 extremely aggregated distributions[12, 13]. 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 ecosystem of figure 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[14]. Figure 2 shows the spatial distribution of the microorganisms after about 8 generations. These amoebae are kept in a perfectly homogeneous environment; moreover, known chemical communications in this particular strains have been deactivated by mutations[15]. 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 VMR (see Figure 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 figure 2 is a neutral one[14]. 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.

This article is organized as follow : the section II develops the neutral framework at zero dimension, when no spatial diffusion can take place. This section stresses the importance of fluctuations due to birth and death phenomena. Spatial diffusion is introduced at section 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 VMR, which is a more useful tool for experimental comparison. Concluding remarks are included in section V.

II. THE PROBABILISTIC FORMULATION AT ZERO DIMENSION.

Before introducing the cumbersome spatial Focker-Plank 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 micro-wells filled with nutriments ; 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 α and die with rate μ . 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 $W^+(n)dt = \alpha ndt$ (or $W^-(n)dt = \mu ndt$). The master equation for the probability $P(n, t)$, counting per unit of time the number of way the community can arrive at size n or leave it is:

$$\begin{aligned} \frac{\partial P(n, t)}{\partial t} &= (W^+(n-1)P(n-1, t) - W^+(n)P(n, t)) \\ &+ (W^-(n+1)P(n+1, t) - W^-(n)P(n, t)) \end{aligned} \quad (2)$$

Various moments $\langle n^k \rangle$ can be extracted directly from the master equation (see appendix A 1). The most useful ones are the average number of individuals $\langle n(t) \rangle$ and the variance, *i.e.* the centered second moment $V(t) = \langle n^2(t) \rangle - \langle n(t) \rangle^2$:

$$\begin{aligned} \frac{d\langle n \rangle}{dt} &= (\alpha - \mu) \langle n \rangle \\ \frac{dV}{dt} &= 2(\alpha - \mu)V + (\alpha + \mu) \langle n \rangle \end{aligned}$$

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 :

$$\langle n(t) \rangle = n_0 \exp((\alpha - \mu)t) \quad (3)$$

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

In the following, we will consider only growing populations $\alpha - \mu \geq 0$. In this case, the variance to mean ratio VMR 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 3a 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 behaviour of the system and the number of individuals in the wells after some 60 generations can vary over 4 orders of magnitude. Figure 3b 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.

Figure 3b 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 diffuse between them ? We know that diffusion 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 he main mathematical tool to investigate spatial clustering [16]. A closely related quantity used in ecological literature is the β -diversity[17], the rate at which similarities between quadrats decreases as a function of distance. 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 $D/2$, duplicating with rate α and dying with rate μ . 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 ℓ 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. 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 neighbouring cells.

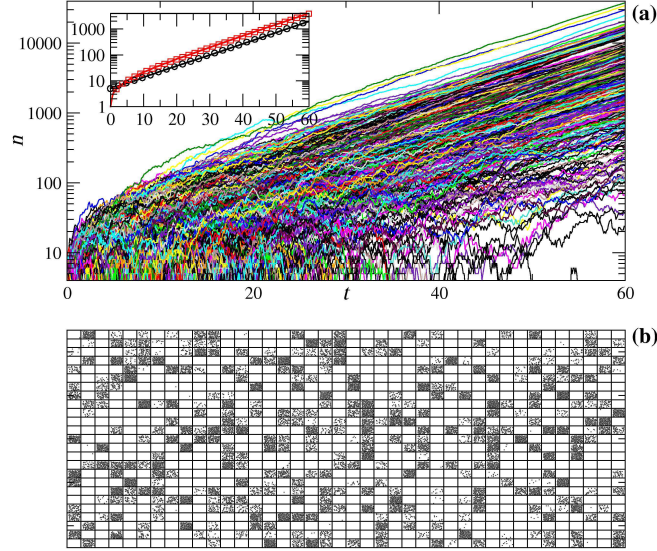


Figure 3. (a) A stochastic numerical simulation of the number of individuals $n_i(t)$ as of function of time for 1000 wells. Each curves represents the population in one well. To generate the paths, at each time step, the number of birth and death are generated by a Poisson process. Unit of time is the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

To reduce the cumbersome notations, we will use the vector $\mathbf{n} = (\dots, n_i, \dots)$ to represent the state of the whole system; the vector $a_i \mathbf{n} = (\dots, n_i - 1, \dots)$ to represent the same state as \mathbf{n} except that one individual has been subtracted from cell i ; $a_i^\dagger \mathbf{n} = (\dots, n_i + 1, \dots)$ 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 :

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

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 appendix A 1). We will assume uniform initial condition of exactly ν_0 particles on all site at time $t = 0$ ($P(\mathbf{n}, 0) = \prod_i \delta_{n_i, \nu_0}$); 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

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

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 (appendix A 1):

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

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

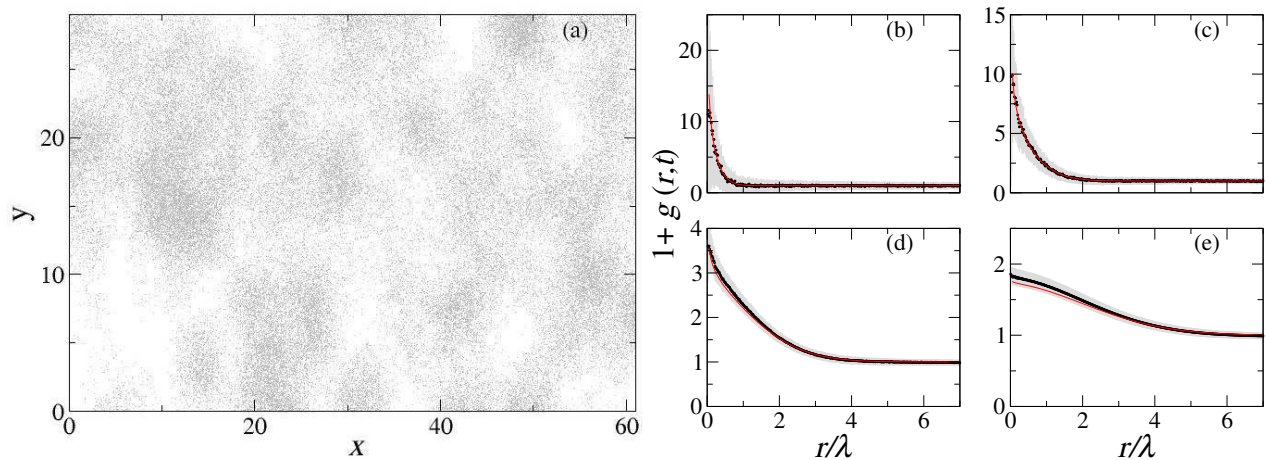


Figure 4. Numerical Simulation of a growing population and its pair correlation function. Data for numerical simulations are similar to the experimental ecosystems of [14] : $\alpha = 1$, $\mu = 0$, $D = 1/2$, $c_0 = 9.6 \times 10^{-2}$; time scales are in generation units, spatial scales in diffusion length units. (a) Spatial map of the organism at $t = 6$; each dot represent one organism. (b-e) pair correlation functions for 4 times : (b) $t = 0.2$; (c) $t = 1$; (d) $t = 3$; (e) $t = 6$. Circles and shaded area : average and \pm one standard deviation computed on 25 numerical simulations ; (red on line) solid lines : theoretical predictions. The small difference at short distances for large times (e) is due to the finite size of the ecosystem.

is however more fruitful to take the continuum limit of cell size $\ell \rightarrow 0$ 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 $c(t) = \langle c(x, t) \rangle$, equation (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

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

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

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

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 Δ . It will be more fruitful to formulate the above equation in the natural scales of the problem : the generation time α^{-1} and the diffusion 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 allows one to compare easily different experiments. Measured in these units, eq.(9) reduces to the dimensionless equation

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

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

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

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

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 Γ is the incomplete gamma function) and displays logarithmic divergence at $d = 2$ for large times[12].

For growing population ($\alpha > \mu$), figure 4 displays a spatial density map of the individual-based numerical simulation of a population and its pair correlation function (see Appendix A 3) 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 inter-species interaction, ...) can be measured. This is how we have been able to demonstrate that the clustering observed in an experimental ecosystem of *Dictyostelia* [14] was purely neutral and not due to such phenomena as chemical communications between microorganisms.

IV. THE 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 cytometry on samples, ...). Various statistics can then be performed to quantify the distribution heterogeneity[18]. The most robust measurement is the variance to mean ratio: A random (Poissonian) distribution of organism will have $\text{VMR} = 1$; a significant deviation beyond this value signify clustering and the dependence of VMR upon time and scale will inform us on the processes governing the ecosystem [19, 20]. The VMR of the spatial distribution of a growing population such as described above reaches extremely high values. For example, for the population displayed in 4a, The VMR computed over a quadrat $L = 1$ (in diffusion length unit) is ≈ 30 .

In the case studied here, as we know the expected pair correlation function, we can derive exactly the VMR 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 :

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

and its variance to mean ratio $\text{VMR}(L, t)$ is

$$\begin{aligned} \frac{\langle N^2 \rangle - \langle N \rangle^2}{\langle N \rangle} &= \frac{1}{L^d c(t)} \int_{\mathcal{A} \times \mathcal{A}} (\langle c(\mathbf{r}, t) c(\mathbf{r}', t) \rangle - c(t)^2) d\mathbf{r} d\mathbf{r}' \\ &= 1 + \frac{c(t)}{L^d} \int_{\mathcal{A} \times \mathcal{A}} g(\mathbf{r} - \mathbf{r}', t) d\mathbf{r} d\mathbf{r}' \\ &= 1 + \frac{1}{L^d} \int_{\mathcal{A} \times \mathcal{A}} \int_0^t \frac{2}{(2\pi s)^{d/2}} \exp\left(-\frac{(\mathbf{r} - \mathbf{r}')^2}{2s}\right) \frac{c(t)}{c(t-s)} ds d\mathbf{r} d\mathbf{r}' \end{aligned}$$

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 Appendix A 2)

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

the VMR reads

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

For the critical case $\alpha = \mu$, $\text{VMR}(L, t)$ 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 $\text{VMR}(L, t)$ grows as $L^d (1-\mu/\alpha)^{-1} \exp((1-\mu/\alpha)t) / t^{d/2}$ for large times $t \gg L^2$ (See Appendix A 2 and Fig. 6inset). On the other hand, for large scale $L \gg \sqrt{t}$, the $\text{VMR}(L, t)$ saturates at

$$1 + 2(1-\mu/\alpha)^{-1} \left(e^{(1-\mu/\alpha)t} - 1 \right) \quad (14)$$

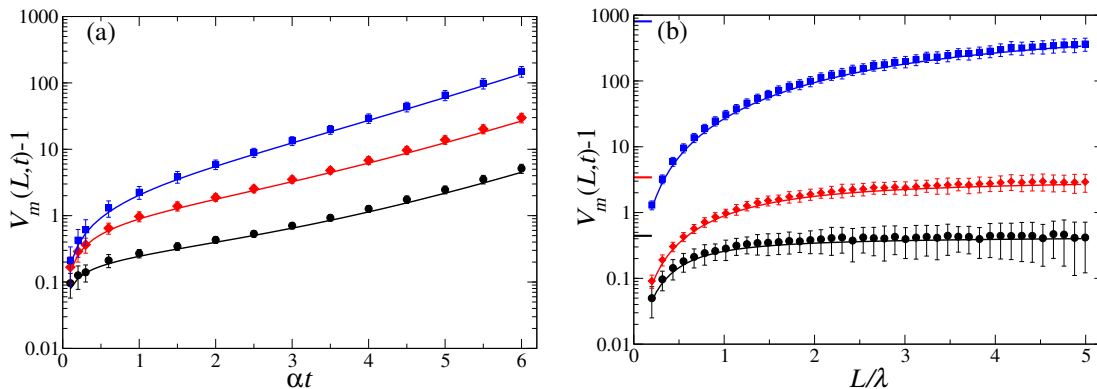


Figure 5. The theoretical VMR (solid curves) and its comparison to the function computed on individual-based numerical simulations (symbols). Data represent averages over 25 simulated populations with the same parameters as in figure 4, and error bars are \pm one standard deviation over the same samples.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

Figure 5 displays the behaviour of the VMR 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 VMR reaches very high values for a growing population, much beyond the Poissonian distribution of $\text{VMR} = 1$, indicating the clustered nature of population distribution.

The study of the dependence of the VMR as a function of scale is similar in principle to some ecological measurements such as fractal index computation [21] or the presence-absence map [22]. It has however a more intuitive interpretation [23].

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 ($\text{VMR} \gg 1$). 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 smoothed by random motions of the organisms and diffusion, specially in dimensions one (communities along a shore line, a river bed, ...) and two (the major forms of life, where the horizontal extension of the community is much larger than its vertical one). The 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 [14] 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 shouldn't 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 applies to growing populations such as plankton blooms, populations extending their geographical range due to environmental change [24] 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 [25, 26].

Appendix A: Mathematical Details.

1. Extracting moments from the master equation.

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

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

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 (A1) are polynomials of the same degree.

The spatial case seems more complicated because the master equation (5) involves a summation over all sites. The notations introduced in section 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[27] 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 the master equation (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 $\mathbf{n} \rightarrow a_i^\dagger \mathbf{n}$ and grouping with the second part of the first line (its conjugate expression), transforms into

$$\sum_{\mathbf{n}} \sum_i \left(f(\eta_k a_i^\dagger \mathbf{n}) - f(\eta_k \mathbf{n}) \right) W^+(\mathbf{n}, i) P(\mathbf{n}, t) \quad (\text{A2})$$

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

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

This process has to be applied to all lines of the master equation (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.

2. The VMR computation.

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

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

$h(u)$ 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 VMR(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 μndt and N_2 of parameter αndt ; eliminate N_1 elements from and add N_2 elements to the array; each new element inherits its positions from

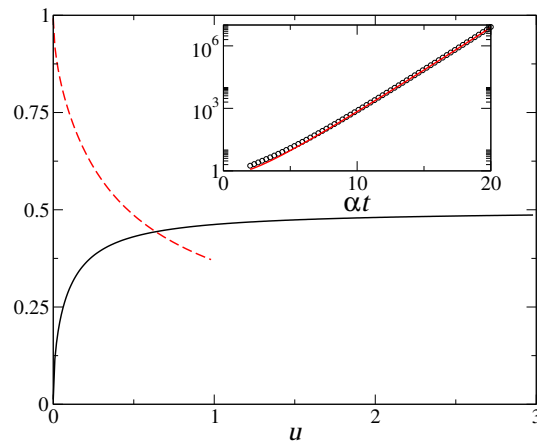


Figure 6. The function $h(u)$ (solid curve) and $h(u)/\sqrt{\pi u/2}$ (dotted curve). Inset : comparison of the function $\text{VMR}(1, t)$ (circles) to its long time approximation $\exp(t)/\pi t$ (solid curve) ; $d = 2$, $\mu = 0$.

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 ; (iv) repeat these operations until a specified amount of time has been spent.

Autocorrelation function is a (normalized) histogram of distances: if there are n_i organisms in cell \mathbf{i} and $n_{\mathbf{i}+\mathbf{j}}$ organisms in cell $\mathbf{i} + \mathbf{j}$, then there are $n_i n_{\mathbf{i}+\mathbf{j}}$ distances \mathbf{j} . The isotropy of the system behooves us to compute the number of distances between r and $r + dr$ (where dr is the bin size) and divide this number (after appropriate normalization by c^2 and the surface of the ecosystem) by $2\pi r$ to retrieve the autocorrelation function (8). Both real data from experiments and numerical simulations are files containing the *position* of each organism in the ecosystem. In principle, the correlation computation program will select an individual i , then run through all individuals j ($j \neq i$), compute the distances r_{ij} and increment the bin (r_{ij}/dr) by $1/2\pi r_{ij}$. This operation will be performed for all individual i . However, both real and numerical experiments take place in a finite world and the boundaries will affect the computation for large distances. The bias comes from the fact that the circle of radius r_{ij} around the organism i may not be contained entirely in the (real or numerical) ecosystem and cross the boundaries. To correct for this effect, one has to find the effective arc length a of radius r_{ij} around organism i which is contained *inside* the boundaries of the ecosystem and increment the bin (r_{ij}/dr) by $1/a$.

-
- [1] L. R. Taylor, I. P. Woiwod, and J. N. Perry. Density-dependence of spatial behavior and rarity of randomness. *Journal Of Animal Ecology*, 47:383–406, 1978.
 - [2] A.P. Martin, M.V. Zubkov, P.H. Burkill, and R.J. Holland. Extreme spatial variability in marine picoplankton and its consequences for interpreting eulerian time-series. *Biol Lett*, 1(3):366–369, Sep 2005.
 - [3] J. L. Green, A. J. Holmes, M. Westoby, I. Oliver, D. Briscoe, M. Dangerfield, M. Gillings, and A. J. Beattie. Spatial scaling of microbial eukaryote diversity. *Nature*, 432:747–750, 2004.
 - [4] R. Condit and et al. Beta-diversity in tropical forest trees. *Science*, 295:666–669, 2002.
 - [5] T. J. Battin, W. T. Sloan, S. Kjelleberg, H. Daims, I. M. Head, T. P. Curtis, and L. Eberl. Microbial landscapes: new paths to biofilm research. *Nat Rev Microbiol*, 5:76–81, 2007.
 - [6] I. González-García, R. V. Solé, and J. Costa. Metapopulation dynamics and spatial heterogeneity in cancer. *Proc Natl Acad Sci U S A*, 99:13085–13089, 2002.
 - [7] R. Levin. Some demographic and genetic consequences of environmental heterogeneity for biological controls. *Bulletin of entomological society of America*, 15:237–240, 1969.
 - [8] I. Hanski and O.E. Gaggiotti, editors. *Ecology, Genetics and Evolution of Metapopulations*. Academic Press, 2004.
 - [9] A. Okubo and J.G. Mitchell. *Diffusion and ecological problems : modern perspectives*, chapter 9., page 268. Springer, 2001.
 - [10] G. Bell. Neutral macroecology. *Science*, 293:2413–2418, 2001.
 - [11] W. R. Young, A. J. Roberts, and G. Stuhne. Reproductive pair correlations and the clustering of organisms. *Nature*, 412:328–331, 2001.
 - [12] B. Houchmandzadeh. Clustering of diffusing organisms. *Phys Rev E*, 66:052902, 2002.
 - [13] B. Houchmandzadeh and M. Vallade. Clustering in neutral ecology. *Phys Rev E*, 68:061912, 2003.
 - [14] B. Houchmandzadeh. Neutral clustering in a simple experimental ecological community. *Phys Rev Lett*, 101(7):078103, Aug 2008.

- [15] R. Sugang, C. J. Weijer, F. Siegert, J. Franke, and R. H. Kessin. Null mutations of the dictyostelium cyclic nucleotide phosphodiesterase gene block chemotactic cell movement in developing aggregates. *Dev Biol*, 192:181–192, 1997.
- [16] B. Bolker and S. W. Pacala. Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. *Theoretical Population Biology*, 52:179–197, 1997.
- [17] J. C. Nekola and P. S. White. The distance decay of similarity in biogeography and ecology. *Journal Of Biogeography*, 26:867–878, 1999.
- [18] M. G. Turner. Landscape ecology: What is the state of the science? *Annual Review Of Ecology Evolution And Systematics*, 36:319–344, 2005.
- [19] L. Benedetti-Cecchi. The importance of the variance around the mean effect size of ecological processes. *Ecology*, 84:2335–2346, 2003.
- [20] A. Mahadevan and J.W. Campbell. *Handbook of scaling methods in aquatic ecology : measurement, analysis, simulation.*, chapter 14 Biogeochemical Variability at the sea surface., page 215. CRC, 2003.
- [21] T. W. Crawford, J. A. Commito, and A. M. Borowik. Fractal characterization of mytilus edulis l. spatial structure in intertidal landscapes using gis methods. *Landscape Ecology*, 21(7):1033–1044, October 2006.
- [22] J. P. M. Witte and P. J. J. F. Torfs. Scale dependency and fractal dimension of rarity. *Ecography*, 26(1):60–68, February 2003.
- [23] J.M. Halley, S. Hartley, A.S. Kallimanis, W.E. Kunin, J.J. Lennon, and S.P. Sgardelis. Uses and abuses of fractal methodology in ecology. *Ecology Letters*, 7:254–271, 2004.
- [24] R. J. Wilson, C. D. Thomas, R. Fox, D. B. Roy, and W. E. Kunin. Spatial patterns in species distributions reveal biodiversity change. *Nature*, 432:393–396, 2004.
- [25] M. J. Keeling, H. B. Wilson, and S. W. Pacala. Reinterpreting space, time lags, and functional responses in ecological models. *Science*, 290:1758–1761, 2000.
- [26] 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. Habitat structure and population persistence in an experimental community. *Nature*, 412(6846):538–543, Aug 2001.
- [27] D.C. Mattis. The uses of quantum field theory in diffusion-limited reactions. *Rev. Mod. Phys.*, 70:979, 1998.