

# Spatial patterns emerging from a stochastic process near criticality

Fabio Peruzzo, Mauro Mobilia, and Sandro Azaele\*  
*Department of Applied Mathematics, University of Leeds*

There is mounting empirical evidence that many communities of living organisms display key features which closely resemble those of physical systems at criticality. We here introduce a minimal model framework for the dynamics of a community of individuals which undergoes local birth-death, immigration and local jumps on a regular lattice. We study its properties when the system is close to its critical point. Even if this model violates detailed balance, within a physically relevant regime dominated by fluctuations, it is possible to calculate analytically the probability density function of the number of individuals living in a given volume, which captures the close-to-critical behavior of the community across spatial scales. We find that the resulting distribution satisfies an equation where spatial effects are encoded in appropriate functions of space, which we calculate explicitly. The validity of the analytical formulæ is confirmed by simulations in the expected regimes. We finally discuss how this model in the critical-like regime is in agreement with several biodiversity patterns observed in tropical rain forests.

## I. INTRODUCTION

Several authors have showed that the parameters of models which describe biological systems are not located at random in their parameter space, but are preferably poised in the vicinity of a point or surface which separates regimes of qualitatively different behaviors [1]. In this sense, stationary states of living systems are not only far from equilibrium, but bring the hallmark of criticality. Although the connection between underpinning dynamics and measurable quantities is sometimes tenuous and hence conclusions about criticality loose, empirical examples span a wide range of biological organization, from gene expression in macrophage dynamics [2], to cell growth [3], relatively small networks of neurons [4], flocks of birds [5] and, possibly, tree populations in tropical forests [6].

In this article, we mainly focus on the spatial patterns emerging from a minimal model of population dynamics close to its critical point. This latter is identified as a singularity in the population size of the system, similarly to what happens in the theory of branching processes in the sub-critical regime [7], when the fluctuations play a crucial role. Therefore, in our model the critical point does not mark a transition between ordered and disordered phases *sensu* equilibrium statistical mechanics [8], although connections in a broader context may certainly exist. The emergent patterns are not calculated by using classical size-expansion methods, but introducing a parameter expansion which appropriately identifies criticality in the parameter space of the model.

The calculation of the probability distribution of large scale configurations emerging from the microscopic dynamics is challenging, even at stationarity [9, 10]. When stochastic processes violate detailed balance, they have a generator which is not self-adjoint [11] and different states are coupled by probability currents at mi-

croscopic level [12]. These flows of probability among microstates break detailed balance, time symmetry and produce macroscopic non-equilibrium behavior. A common way to overcome these hurdles is to formulate some kind of effective Langevin equation which describes the dynamics of the mesoscopic variables of interest, losing track, however, of the underlying microscopic dynamics [13–15]. Nonetheless, in this paper we study a model which, despite violating microscopic detailed balance [16, 17], allows one to study analytically (stationary) out-of-equilibrium properties of spatial patterns. These latter emerge mainly because of the large intrinsic fluctuations of the local population sizes. Also, the model's mathematical amenability allows us to analyze in detail those spatial ecological patterns and to compare them with observation data for ecosystems with large species richness. The agreement between model predictions and empirical data highlights the usefulness of the approach and strengthens the connections between physics and theoretical ecology.

In this spatial metacommunity model, local communities (or, equivalently, sites or voxels) are located on a  $d$ -dimensional regular graph (or lattice) where individuals are treated as well-mixed particles which undergo a birth and death process with local diffusion and constant colonization. We thoroughly investigate the spatial stochastic dynamics close to criticality and deduce the equation governing the evolution of the conditional distribution  $p(N|V; t)$ , the probability to find  $N$  individuals in a volume  $V$  at time  $t$  (in dimension  $d$ ). Within this regime we map the equation of  $p(N|V)$  of the out-of-equilibrium spatial model into an equation of a suitable stochastic process, which instead satisfies detailed balance. This model is described by functions of space, which we are able to calculate exactly. The exact stochastic simulations are always matched by our analytical formulæ in the expected regimes.

The rest of the paper is organized as follows: in section II we introduce the master equation of the model; in section III we calculate the mean and pairwise correlation; in section IV we study the generating function of the

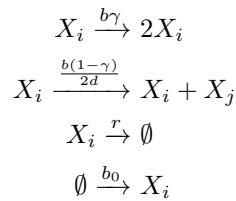
---

\* s.azaele@leeds.ac.uk

conditional probability density function of  $N(V)$ ; in section V and VI we calculate the population variance and the conditional pdf of  $N(V)$ , respectively, along with a comparison between simulations and analytical formulæ; in section VII we present an ecological application of the model; finally, section VIII includes some discussions and perspectives about the results.

## II. MASTER EQUATION OF THE MODEL

This is a metapopulation model in which individuals live in local communities (or sites) located on a  $d$ -dimensional lattice,  $\mathbb{L}$ , whose linear side is  $a$ . If  $X_i$ ,  $i \in \mathbb{L}$ , indicates an individual living in site  $i$ , the reactions defining the model's dynamics can be cast into the form



where  $j$  indicates a site which is a nearest neighbor of the site  $i$ . In this model, individuals within local communities (or, equivalently, sites or voxels) are treated as diluted, well-mixed point-like particles which undergo a minimal stochastic demographic dynamics: each individual may die at a constant death rate  $r$  and give birth to an offspring at a constant rate  $b$ . The newborn individual remains in the same community with probability  $\gamma$ , whereas it may hop onto one of the  $2d$  nearest neighbours with probability  $1 - \gamma$ . Also, all communities are colonized by external individuals at a constant immigration rate  $b_0$ , which prevents the system to end up into an absorbing state without individuals [18, 19].

Notice that (for  $0 \leq \gamma < 1$ ) spatial movement is always coupled to birth, so that only newborn individuals can move. This is because we have in mind an application to spatial ecology, where this model mimics the population dynamics in species-rich communities of trees, in which only seeds can move. However, it can be easily modified to include random walk behaviour or different dispersals – like those for bacteria or humans – in which the local birth and the hopping probability are in general decoupled. We have indeed verified that the generality of our final results does not depend on that coupling.

In the language of chemical reaction kinetics, the first reaction represents an autocatalytic production; this and the hopping move are responsible for the break of detailed balance as shown in Appendix A. Therefore, stationary states of this process are non-equilibrium steady states, albeit the model is defined by linear birth and death rates.

Indeed, it is worth emphasizing that each “patch” has not a maximum number of individuals which it can accommodate, but any population size is allowed, albeit

large sizes have an exponentially small probability to occur. This is because the model has no intrinsic carrying capacity which leads the population to saturation. A carrying capacity has the advantage to bring in more realistic features, but it also makes the model mathematically more complicated because of non-linear terms. Here we show that linear rates of a stochastic model *per se* can produce a relevant phenomenology within a stationary out-of-equilibrium model. Thus, since we wanted to focus on the regime near criticality, we have preferred to delve into a relatively simpler system, in which non-linearities are neglected in a first approximation. More complicated dynamics are definitely important and will be studied in future works.

Finally, for the system to avert demographic explosion we have to assume that  $b_0 > 0$  and  $0 < b < r$ , but it turns out that the most interesting features emerge when  $b \simeq r$ , i.e., close to its critical point. Indeed, as we will show in Section VII, for comparable birth and death rates the model is able to describe several spatial patterns of tree species in tropical forests [20].

Let us now indicate with  $n_i$  the number of individuals in site  $i$ . Assuming that within every site the spatial structure can be neglected and that we have perfect mixing, when the configuration of the system is  $\{n\} = \{n_i : i \in \mathbb{L}\}$ , the linear birth and death rates in site  $i$ , i.e.,  $\mathcal{W}_i^+(\{n\})$  and  $\mathcal{W}_i^-(\{n\})$ , read respectively

$$\begin{aligned} \mathcal{W}_i^+(\{n\}) &= \frac{b(1-\gamma)}{2d} \sum_{j:|j-i|=1} n_j + b\gamma n_i + b_0 \\ \mathcal{W}_i^-(\{n\}) &= r n_i \end{aligned} \quad (1)$$

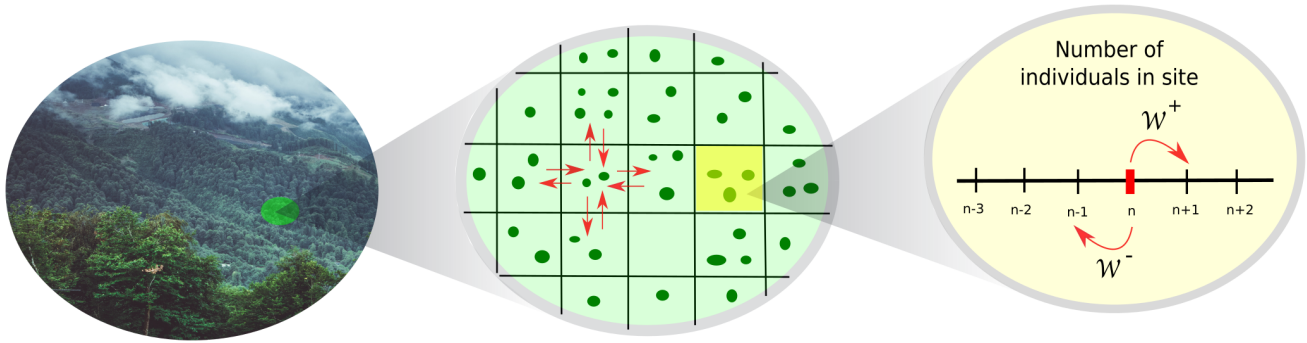
Let  $P(\{n\}, t)$  be the probability to find the system in the configuration  $\{n\}$  at time  $t$ . Then the master equation for  $P(\{n\}, t)$  reads

$$\begin{aligned} \frac{\partial}{\partial t} P(\{n\}, t) &= \sum_{i \in \mathbb{L}} \left\{ \right. \\ &\mathcal{W}_i^+(\{\dots, n_i - 1, \dots\}) P(\{\dots, n_i - 1, \dots\}, t) + \\ &-\mathcal{W}_i^+(\{n\}) P(\{n\}, t) + \\ &+\mathcal{W}_i^-(\{\dots, n_i + 1, \dots\}) P(\{\dots, n_i + 1, \dots\}, t) + \\ &\left. -\mathcal{W}_i^-(\{n\}) P(\{n\}, t) \right\} \end{aligned} \quad (2)$$

where the dots represent that all other occupation numbers remain as in  $\{n\}$  and it is intended that  $P(\cdot) = 0$  whenever any of the entrances is negative. The spatial generating function of the process is defined as

$$\begin{aligned} \zeta(\{H\}, t) &= \langle e^{\sum_{k \in \mathbb{L}} n_k H_k} \rangle = \\ &= \sum_{\{n\}} e^{\sum_{k \in \mathbb{L}} n_k H_k} p(\{n\}, t) \end{aligned}$$

where  $H_i \leq 0$  for every  $i \in \mathbb{L}$ . Multiplying both sides of eq.(2) by  $e^{\sum_{k' \in \mathbb{L}} n_{k'} H_{k'}}$  and summing over all configurations of the system, we find that  $\zeta(\{H\}, t)$  satisfies the



**FIG. 1: Illustration of the model.** Individual trees are represented by dark green circles within local communities which are located on a regular graph (lattice). Each individual may die or give birth to an offspring with constant per capita rates. New individuals may either remain in the community where they were born with probability  $\gamma$ , or hop onto one of the  $2d$  nearest neighbours with probability  $1 - \gamma$ . Finally, all communities are colonized by external individuals at a constant immigration rate  $b_0$ . The dynamics of the model is therefore defined by the jump rates defined in eq.(1).

equation

$$\frac{\partial}{\partial t} \zeta(\{H\}, t) = \sum_{i \in \mathbb{L}} \left\{ (e^{H_i} - 1) \left[ \frac{b(1-\gamma)}{2d} \sum_{j:|i-j|=1} \frac{\partial \zeta}{\partial H_j} \right. \right. \quad (3)$$

$$\left. \left. + b\gamma \frac{\partial \zeta}{\partial H_i} + b_0 \zeta \right] + r(e^{-H_i} - 1) \frac{\partial \zeta}{\partial H_i} \right\}.$$

This is the main equation of the model from which we will calculate the most important results. We were not able to find the full solution of this equation. However, one can gain a lot of information about the general properties of the process by looking into the probability distribution of the random variable  $N(V, t) = \sum_{i \in \mathcal{V}} n_i(t)$ , where  $\mathcal{V}$  is the set of sites in a  $d$ -dim volume. Before studying such a distribution, it is useful to calculate the mean number of individuals and the spatial correlation between any pair of sites.

### III. MEAN AND PAIR CORRELATION

The equation for the mean number of individuals in the site  $k$  can be obtained by taking the partial derivative of both sides of eq.(3) with respect to  $H_k$  and then setting  $\{H\} = 0$ :

$$\frac{\partial \langle n_k \rangle}{\partial t} = \frac{b(1-\gamma)}{2d} \Delta_k \langle n_k \rangle - \mu \langle n_k \rangle + b_0$$

where  $\mu := r - b$  and  $\Delta_k$  is the discrete Laplace operator, which is defined as

$$\Delta_k f(k) = \sum_{j:|k-j|=1} (f(j) - f(k)) .$$

This finite difference equation can be solved in full generality and at stationarity, i.e. for  $t \rightarrow \infty$ , we get simply  $\langle n \rangle = \frac{b_0}{\mu}$ , regardless of any spatial location.

The pairwise spatial correlation between sites  $k$  and  $l$ , i.e.,  $\langle n_k n_l \rangle$ , can also be obtained by taking the partial derivatives of both sides of eq.(3) with respect to  $H_k$  and  $H_l$ , and then setting  $\{H\} = 0$  (see Appendix B for details):

$$\frac{\partial}{\partial t} \langle n_k n_l \rangle = D \left( \Delta_k \langle n_k n_l \rangle + \Delta_l \langle n_k n_l \rangle \right) + \quad (4)$$

$$- 2\mu \langle n_k n_l \rangle + 2b_0 \langle n \rangle +$$

$$+ \delta_{k,l} \left( 2\sigma^2 \langle n \rangle + b_0 + D \Delta_k \langle n_k \rangle \right)$$

where we have used the notation

$$D := \frac{b(1-\gamma)}{2d} \quad \text{and} \quad \sigma^2 := \frac{r+b}{2} .$$

We also introduce

$$\lambda := \sqrt{\frac{D}{\mu}} \quad \text{and} \quad \rho := \sqrt{\frac{\sigma^2}{b_0}} ,$$

which are dimensionless parameters and provide important information about how spatial diffusion intermingles with demographic dynamics.

In order to solve eq.(4), we introduce a  $d$ -dim system of Cartesian coordinates where the coordinates of each site are given as a multiple of the lattice side  $a$ . Thus, a vector  $\mathbf{k}$  indicates the corresponding position of a site. In this way, we can calculate the stationary solution of eq.(4) by writing  $\langle n_{\mathbf{k}} n_{\mathbf{l}} \rangle$  as a Fourier series expansion. Exploiting translation invariance the final expression of the solution reads (see Appendix B)

$$\langle n_{\mathbf{k}} n_{\mathbf{l}} \rangle = \langle n \rangle^2 + \langle n \rangle^2 \rho^2 \left( 1 + \frac{\mu}{2\sigma^2} \right) \times$$

$$\times \left( \frac{a}{2\pi} \right)^d \int_{\mathcal{C}} d\mathbf{p} \frac{e^{i\mathbf{p} \cdot (\mathbf{k}-\mathbf{l})}}{1 + 2\lambda^2 \sum_{i=1}^d (1 - \cos(p_i a))} \quad (5)$$

where  $p_i$  is the  $i$ -th Cartesian component of the  $d$ -dim vector  $\mathbf{p}$  and  $\mathcal{C}$  is the hypercubic primitive unit cell of size  $2\pi/a$ . Interestingly, pairs of sites de-correlate for  $\gamma = 1$  or  $b = 0$ , when at stationarity we obtain  $\langle n_{\mathbf{k}} n_{\mathbf{l}} \rangle - \langle n \rangle^2 = c \delta_{\mathbf{k}, \mathbf{l}}$ , where  $c$  is a constant depending only on the demographic parameters and  $\delta_{\mathbf{k}, \mathbf{l}}$  is a Kronecker delta. However,  $c \neq \langle n \rangle$ , pointing out that local fluctuations are non-Poissonian.

Eq.(5) is amenable to a continuous spatial limit, obtained as  $a \rightarrow 0$ , and provides a closed analytic form for the pair correlation. Indicating now with  $n(x)$  and  $n(y)$  the density of individuals on the sites located at  $x$  and  $y$ , respectively, in continuous space (and rescaling parameters accordingly), we find (see Appendix B)

$$\frac{\langle n(x)n(y) \rangle}{\langle n \rangle^2} = 1 + \frac{\bar{\rho}^2}{(2\pi\bar{\lambda}^2)^{d/2}} \left(1 + \frac{\mu}{2\sigma^2}\right) \times \left(\frac{|x-y|}{\bar{\lambda}}\right)^{\frac{2-d}{2}} K_{\frac{2-d}{2}}\left(\frac{|x-y|}{\bar{\lambda}}\right) \quad (6)$$

where  $|x-y|$  is the distance between the sites located at  $x$  and  $y$ ,  $K_\nu$  is the modified Bessel function of the second kind of order  $\nu$  [21], and we have defined

$$\bar{\lambda} := \sqrt{\frac{\bar{D}}{\mu}} \quad \bar{\rho} := \sqrt{\frac{\sigma^2}{b_0}} \quad ,$$

where  $\bar{D} = Da^2$  and  $\bar{b}_0 = b_0/a^d$  are finite as  $a \rightarrow 0$ . As the asymptotic behavior of  $K_\nu$  as  $z \rightarrow \infty$  is  $K_\nu(z) \sim e^{-z} \sqrt{\frac{\pi}{2z}}$ ,  $\bar{\lambda}$  is the correlation length of the system;  $\bar{\rho}^2$  has the dimensions of a  $d$ -dim volume and gives the local intensity of the correlations. Because eq.(6) is the continuum limit of eq.(5), this expression of the pair correlation function is a good approximation of the one in eq.(5) only when  $|x-y| \gg a$  and  $\bar{\lambda} \gg a$ .

#### IV. GENERATING FUNCTION CLOSE TO THE CRITICAL POINT

In this section we introduce the parameters which allow us to identify a region close to the critical point of the model. This suggests an expansion which will lead to simplified equations which, nonetheless, carry a lot of information about the model.

A simple way to make progress with the master equation in eq.(2) is the use of a formal Kramers-Moyal expansion [22]. It is well-known that there are limitations to this procedure and it has been criticized, because one cannot always pinpoint a small parameter for the correct expansion [22, 23]. The system-size expansion solves these difficulties, but it has to be applied when the size of the system becomes large [23]. Here, however, it is not entirely evident what parameter should identify the size (population size or volume) of the system in the critical regime. Indeed, the model has no carrying capacity or maximum population size, and the total volume of the

system could only provide us with the macroscopic equation, which has no interest in the present case.

In order to make analytical progress, we have therefore introduced two dimensionless parameters,  $\varepsilon$  and  $\eta$ , which identify a non-trivial region when  $0 < b < r$ , but  $b \rightarrow r$ . We define  $\varepsilon := \frac{2(r-b)}{r+b}$  with the condition that  $\frac{b_0}{\mu}\varepsilon = \mathcal{O}(1)$  as  $\varepsilon \rightarrow 0^+$ ; in this way we obtain a constant  $\rho^2 = \mu/b_0\varepsilon$ , which in real systems is large because usually  $r/b_0 \gg 1$ . The parameter  $\varepsilon$  indicates how close the system is to the critical point, regardless of spatial diffusion. With the independent parameter  $\eta := \frac{D}{\sigma^2}$ , instead, we compare the importance of spatial diffusion with respect to demographic fluctuations. We will assume that  $\eta = \mathcal{O}(\varepsilon)$  as  $\varepsilon \rightarrow 0^+$  and hence  $\eta/\varepsilon = \lambda^2$ , a new independent constant.

When  $b$  and  $r$  are close to each other, we expect that population sizes can be well approximated with continuous random variables in each site. In order to understand when this is possible in relation to the parameters  $\varepsilon$  and  $\eta$ , we assumed that the generating function  $\zeta(\{H\}, t)$  is analytic at  $H_i = 0$  for any  $i$  and the most important contribution to the equation of  $\zeta(\{H\}, t)$  comes from a negative neighborhood of the origin with thickness  $\mathcal{O}(\varepsilon)$ . This is tantamount to introducing the change of variable  $H_i = \varepsilon S_i$  into eq.(3) and to expanding in powers of  $\varepsilon$ , assuming  $S_i = \mathcal{O}(1)$  and  $S_i \leq 0$ . Up to linear order in  $\eta$  and  $\varepsilon$  we obtain

$$\frac{\partial}{\partial t} \zeta(\{S\}, t) = \sum_{i \in \mathbb{L}} \sigma^2 S_i \left\{ \eta \Delta_i \frac{\partial \zeta}{\partial S_i} - \varepsilon \frac{\partial \zeta}{\partial S_i} + \frac{\varepsilon}{\rho^2} \zeta + \varepsilon S_i \frac{\partial \zeta}{\partial S_i} \right\} \quad (7)$$

or

$$\frac{\partial}{\partial T} \zeta(\{S\}, t) = \sum_{i \in \mathbb{L}} S_i \left\{ \lambda^2 \Delta_i \frac{\partial \zeta}{\partial S_i} - \frac{\partial \zeta}{\partial S_i} + \frac{1}{\rho^2} \zeta + S_i \frac{\partial \zeta}{\partial S_i} \right\} \quad , \quad (8)$$

where we have introduced the dimensionless time  $T := \mu t$  and now  $\zeta$ , with a slight abuse of notation, indicates the generating function corresponding to continuous (and dimensionless) random variables. Therefore, the evolution equation for the generating function of the population sizes becomes

$$\frac{\partial \zeta}{\partial T} = \sum_{i \in \mathbb{L}} H_i \left\{ \lambda^2 \Delta_i \frac{\partial \zeta}{\partial H_i} - \frac{\partial \zeta}{\partial H_i} + \langle n \rangle \zeta + \frac{\sigma^2}{\mu} H_i \frac{\partial \zeta}{\partial H_i} \right\} \quad , \quad (9)$$

where  $H_i$  is the variable conjugated to the continuous random variable  $n_i$ . The parameters also correspond to this continuum limit and now the population sizes  $n_i$  have an exponential cut-off with a (large) characteristic scale given by  $\sigma^2/\mu$ . Eq.(9) leads to the following Fokker-

Planck equation:

$$\begin{aligned} \frac{\partial}{\partial T} P(\{n\}, T) = & \\ & \sum_{i \in \mathbb{L}} \left\{ -\frac{\partial}{\partial n_i} \left[ \left( \lambda^2 \Delta_i n_i - n_i + \langle n \rangle \right) P(\{n\}, T) \right] + \right. \\ & \left. + \frac{\sigma^2}{\mu} \frac{\partial^2}{\partial n_i^2} \left[ n_i P(\{n\}, T) \right] \right\} \quad , \quad (10) \end{aligned}$$

where now  $\{n\}$  are continuous random variables and  $P(\{n\}, T)$  is the corresponding probability density function. It is interesting to note that this is not exactly equivalent to the naïve Kramers-Moyal expansion of eq.(2), which would entail additional terms in the diffusive part. Nevertheless it is a diffusive approximation of the process in the regime identified by the parameters  $\eta$  and  $\varepsilon$ .

### A. Conditional generating function

Eq.(10) cannot be solved in full generality, but we can better understand the underlying dynamics by studying the distribution of the population sizes in arbitrary volumes of space. Let us indicate with  $\mathcal{V}$  the set of sites belonging to a  $d$ -dim volume  $|\mathcal{V}| = V$ , and let us introduce the random variable  $N(V, t) = \sum_{i \in \mathcal{V}} n_i(t)$ , i.e. the total number of individuals in  $\mathcal{V}$  at time  $t$ . By indicating with  $P(N|V, t)$  the corresponding probability density function of  $N$ , we define the conditional generating function

$$Z(h|V, t) = \langle e^{hN(V,t)} \rangle = \int_0^\infty dN e^{hN} P(N|V, t)$$

where  $h \leq 0$ . We obtain the corresponding equation for  $Z$  by specifying  $H_i$ , i.e.

$$H_i = \begin{cases} h & \text{if } i \in \mathcal{V} \\ 0 & \text{if } i \notin \mathcal{V} \end{cases} \quad (11)$$

and substituting this into eq.(9). Thus,

$$\begin{aligned} \frac{\partial}{\partial T} Z(h|V, T) = & h \left[ \lambda^2 \sum_{i \in \mathcal{V}} \Delta_i f(i, h, V, T) - \frac{\partial Z}{\partial h} + \right. \\ & \left. + \langle n \rangle V Z \right] + \frac{\sigma^2}{\mu} h^2 \frac{\partial Z}{\partial h} \quad (12) \end{aligned}$$

where  $f(i, h, V, T) = \langle n_i(T) e^{hN(V,T)} \rangle$  and we have used the identity

$$\sum_{i \in \mathcal{V}} \langle n_i e^{hN} \rangle = \frac{\partial Z}{\partial h}(h|V, T) \quad . \quad (13)$$

This equation depends on  $f(i, h, V, T)$  which in general is unknown. An equation for  $f$  can be derived by

differentiating both sides of eq.(9) with respect to  $H_k$ . Assuming as before that  $k \in \mathcal{V}$ , we finally obtain the following equation for  $f(k, h, V, T)$  (abbreviated  $f(k, h)$ )

$$\begin{aligned} \frac{\partial}{\partial T} f(k, h) = & \lambda^2 \Delta_k f(k, h) - f(k, h) + \\ & + \frac{b_0}{\mu} Z + h \left[ \lambda^2 \sum_{i \in \mathcal{V}} \Delta_i g(i, k, h) + \right. \\ & \left. - \frac{\partial}{\partial h} f(k, h) + \langle n \rangle V f(k, h) \right] + \\ & + 2 \frac{\sigma^2}{\mu} h f(k, h) + \frac{\sigma^2}{\mu} h^2 \frac{\partial}{\partial h} f(k, h) \quad (14) \end{aligned}$$

where  $g(i, k, h, V, T) := \langle n_i(T) n_k(T) e^{hN(V,T)} \rangle$  and where we have used the identity

$$\sum_{i \in \mathcal{V}} \langle n_i n_k e^{hN} \rangle = \frac{\partial}{\partial h} f(k, h, V, T) \quad . \quad (15)$$

In eq.(14) the function  $g$  is still unknown, but it is possible to show that  $\Delta_i g(i, k, h, T) = \Delta_k g(i, k, h, T)$  to leading order as  $a \rightarrow 0$  (see Appendix C). This allows us to obtain a closed equation for  $f$ . Indeed, we have

$$\sum_{i \in \mathcal{V}} \Delta_i g(i, k, h, V, T) = \frac{\partial}{\partial h} \Delta_k f(k, h, V, T)$$

after using the identity (15).

The spatial continuous limit of eq.(12) as  $a \rightarrow 0$  reads

$$\begin{aligned} \frac{\partial}{\partial T} Z(h|V, T) = & h \left[ \bar{\lambda}^2 \int_{\mathcal{V}} dx \nabla_x^2 f(x, h, T) - \frac{\partial Z}{\partial h} + \right. \\ & \left. + \langle n \rangle V Z \right] + \frac{\sigma^2}{\mu} h^2 \frac{\partial Z}{\partial h} \quad , \quad (16) \end{aligned}$$

where in  $\langle n \rangle$  we have substituted  $b_0$  with  $\bar{b}_0$ . This equation is of crucial importance in what follows. Similarly, eq.(14) becomes

$$\begin{aligned} \frac{\partial}{\partial T} f(y, h, V, T) = & \bar{\lambda}^2 \nabla_y^2 f(y, h) - f(y, h) + \langle n \rangle Z + \\ & + h \left[ \bar{\lambda}^2 \frac{\partial}{\partial h} \nabla_y^2 f(y, h) - \frac{\partial}{\partial h} f(y, h) + \right. \\ & \left. + \langle n \rangle V f(y, h) \right] + \frac{\sigma^2}{\mu} \frac{\partial}{\partial h} \left[ h^2 f(y, h) \right] \quad , \quad (17) \end{aligned}$$

where with the continuous coordinates ( $n_i \rightarrow n(x)$  and  $n_k \rightarrow n(y)$ ), we have also  $f(x, h, V, T) = \langle n(x) e^{hN} \rangle$  and  $N(V, t) = \int_{\mathcal{V}} dx n(x, t)$ . When there are no spatial effects, i.e.,  $\bar{\lambda} = 0$ , eq.(16) reads

$$\frac{\partial}{\partial T} Z(h|V, T) = h \left[ - \left( 1 - \frac{\sigma^2}{\mu} h \right) \frac{\partial Z}{\partial h} + \langle n \rangle V Z \right] \quad , \quad (18)$$

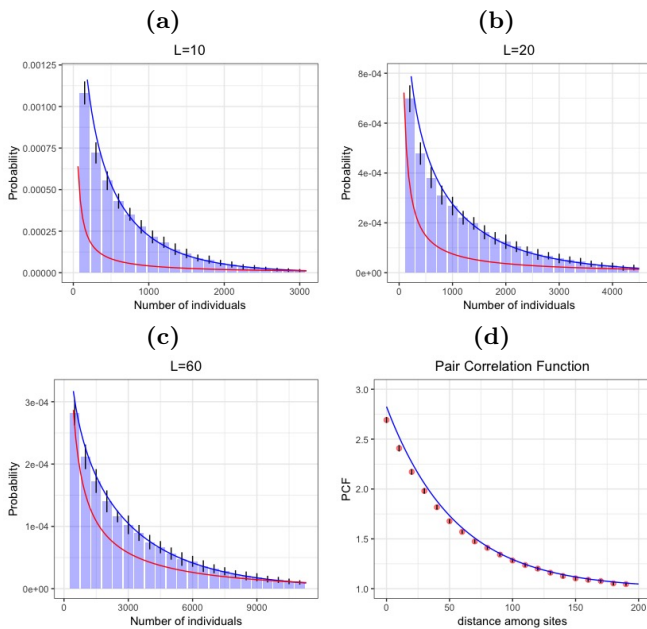
which has the following solution at stationarity

$$Z(h|V) = \left( 1 - \frac{\sigma^2}{\mu} h \right)^{-\frac{\mu \langle n \rangle V}{\sigma^2}} \quad , \quad (19)$$

where  $h \leq 0$  and  $\mu \langle n \rangle V / \sigma^2 = V / \bar{\rho}^2$ . This is the generating function of a gamma distribution [23] with mean  $\langle n \rangle V$  and variance  $\langle n \rangle^2 V \bar{\rho}^2$ . Also, it is not difficult to verify that when  $\bar{\lambda} = 0$ , the solution of eq.(17) is given by

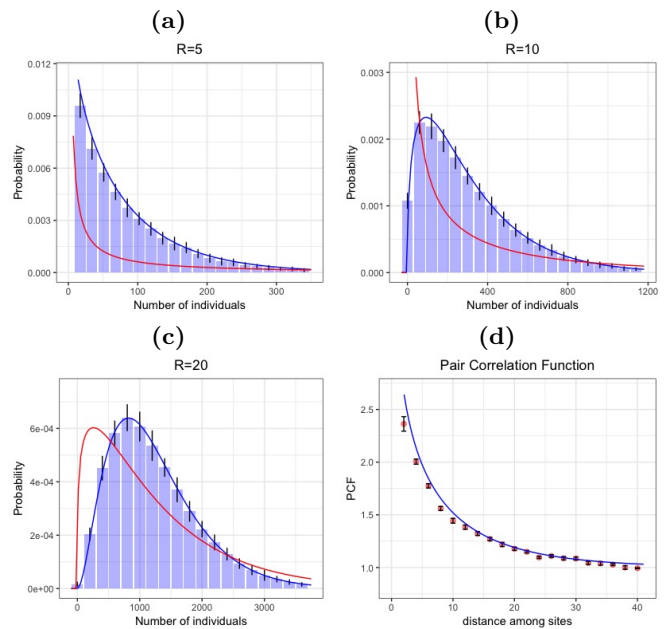
$$f(y, h, V, T) = \frac{1}{V} \frac{\partial Z}{\partial h}(h|V, T) \quad . \quad (20)$$

In the next section we calculate the stationary variance of the population sizes in a volume  $V$ , i.e., of the random variable  $N(V)$ .



**FIG. 2: Stationary distribution as obtained from the phenomenological algorithm ( $d = 1$ ).** Panels

(a), (b) and (c) present a comparison between the simulated model as obtained from the phenomenological algorithm (histograms) outlined in sec.VI and the stationary solution as calculated with eq.(35) (blue solid line). The lattice comprises 500 sites in total and we carried out 50,000 independent realizations in  $d = 1$  with periodic boundary conditions, where parameters are  $D = 30$ ,  $b_0 = 0.5$ ,  $\mu = 0.01$  and  $\sigma = 10$ , and hence  $\bar{\lambda} \approx 55$ . Panels show results for segments of different lengths which include 10, 20 and 60 adjacent sites, respectively. The size of error bars (black lines) are twice as much the standard deviation, while the red solid line represents the mean field solution of the system (i.e. eq.(35) where  $\Sigma = \sigma^2 / \mu$ ). Panel (d) presents the comparison between the simulated and the analytic pair correlation function at stationarity. Red dots are from simulations, while the blue solid line is the analytic solution given by eq.(6).



**FIG. 3: Stationary distribution as obtained from the phenomenological algorithm ( $d = 2$ ).** Panels

(a), (b) and (c) present a comparison between the simulated model as obtained from the phenomenological algorithm (histograms) outlined in sec.VI and the stationary solution as calculated with eq.(35) (blue solid line). The square lattice comprises  $200 \times 200$  sites. We carried out 50,000 independent realizations in  $d = 2$  with periodic boundary conditions, with parameters  $D = 20$ ,  $b_0 = 0.1$ ,  $\mu = 0.1$  and  $\sigma = 10$ , and hence  $\bar{\lambda} \approx 15$ . Panels (a-c) show results for different areas with radii of length 5, 10 and 20, respectively. The size of error bars (black lines) are twice as much the standard deviation, while the red solid line represents the mean field solution of the system (i.e., eq.(35) where  $\Sigma = \sigma^2 / \mu$ ). Panel (d) presents the comparison between the simulated and the analytic pair correlation function at stationarity. Red dots are from simulations, while the blue solid line is the analytic solution given by eq.(6).

## V. POPULATION VARIANCE

In this section we calculate the stationary variance of the population size, i.e.,  $N(V) = \int_V dx n(x)$ , where  $V$  is a finite volume. This quantity is of pivotal importance in the approximation that we are going to develop in the following sections and it is key when introducing spatial information in the equations. Thus we outline here the main calculations, leaving to Appendix D further details.

We start from the spatial continuous limit of eq.(4) as  $a \rightarrow 0$ , which at stationarity reads

$$\bar{\lambda}^2 \left( \nabla_x^2 \langle n(x)n(y) \rangle + \nabla_y^2 \langle n(x)n(y) \rangle \right) - 2 \langle n(x)n(y) \rangle + 2 \langle n \rangle^2 + 2 \frac{\sigma^2}{\mu} \langle n \rangle \delta(x-y) = 0 \quad (21)$$

where  $\delta(x-y)$  is a Dirac delta and we have included only the leading terms as  $\varepsilon \rightarrow 0$ . By integrating both sides of eq.(21) with respect to  $y \in \mathcal{V}$ , we obtain an equation for  $\langle n(x)N(V) \rangle$ . We plan to find an explicit expression for this quantity, because it plays a key role in the following when we will calculate  $Z(h|V)$ . Henceforth, we will take  $\mathcal{V}$  to be a  $d$ -dim ball of radius  $R$  and we will assume that the origin of the Cartesian coordinates is at its center. Thus we indicate with  $|x|$  the distance from the origin of the site located at  $x$  in this coordinate system. With this notation and using the symmetry of  $\langle n(x)n(y) \rangle$  with respect to  $x$  and  $y$ , we obtain at stationarity

$$\bar{\lambda}^2 \nabla_x^2 \langle n(x)N(R) \rangle - \langle n(x)N(R) \rangle + \langle n \rangle^2 V + \frac{\sigma^2}{\mu} \langle n \rangle \Theta(R - |x|) = 0 \quad (22)$$

This linear ODE has to be solved separately for  $|x| > R$  and  $|x| < R$ . The continuity of  $\langle n(x)N(R) \rangle$  and its first derivative at the boundary  $|x| = R$  provide the solution. The final results is (see Appendix D)

$$\langle n(x)N(R) \rangle = \langle n \rangle^2 V + \frac{\sigma^2}{\mu} \langle n \rangle \Psi\left(\frac{|x|}{\bar{\lambda}}, \frac{R}{\bar{\lambda}}\right) \quad (23)$$

where the function  $\Psi$  takes the following form for  $|x| \leq R$

$$\Psi\left(\frac{|x|}{\bar{\lambda}}, \frac{R}{\bar{\lambda}}\right) = 1 - \frac{\left(\frac{|x|}{R}\right)^{1-\frac{d}{2}} K_{\frac{d}{2}}\left(\frac{R}{\bar{\lambda}}\right) I_{\frac{d}{2}-1}\left(\frac{|x|}{\bar{\lambda}}\right)}{I_{\frac{d}{2}-1}\left(\frac{R}{\bar{\lambda}}\right) K_{\frac{d}{2}}\left(\frac{R}{\bar{\lambda}}\right) + I_{\frac{d}{2}}\left(\frac{R}{\bar{\lambda}}\right) K_{\frac{d}{2}-1}\left(\frac{R}{\bar{\lambda}}\right)} \quad (24)$$

where  $I_\nu(z)$  and  $K_\nu(z)$  are the modified Bessel functions of the first and second kind, respectively [21]. Integrating both sides of eq.(23) with respect to  $x \in \mathcal{V}$ , we obtain the equation for the second moment of  $N(R)$ , i.e.

$$\langle N(R)^2 \rangle = \langle n \rangle^2 V^2 + \frac{\sigma^2}{\mu} \langle n \rangle V \psi\left(\frac{R}{\bar{\lambda}}\right) \quad (25)$$

where  $\psi(R/\bar{\lambda})$  takes the following explicit form in dimension  $d$  (see Appendix D for its behavior):

$$\psi\left(\frac{R}{\bar{\lambda}}\right) = 1 - \frac{\frac{d\bar{\lambda}}{R} K_{\frac{d}{2}}\left(\frac{R}{\bar{\lambda}}\right) I_{\frac{d}{2}}\left(\frac{R}{\bar{\lambda}}\right)}{I_{\frac{d}{2}-1}\left(\frac{R}{\bar{\lambda}}\right) K_{\frac{d}{2}}\left(\frac{R}{\bar{\lambda}}\right) + I_{\frac{d}{2}}\left(\frac{R}{\bar{\lambda}}\right) K_{\frac{d}{2}-1}\left(\frac{R}{\bar{\lambda}}\right)} \quad (26)$$

These two functions, namely  $\langle n(x)N(R) \rangle$  in eq.(23) and  $\langle N(R) \rangle^2$  in eq.(25) will be used in the next sections to

calculate a first order approximation of the spatially explicit probability density function of  $N(V)$  in the vicinity of the critical point.

Finally, these solutions allow us to write down the analytic form of the variance of  $N(R)$  in dimension  $d$ , i.e.

$$\text{Var}[N(R)] = \langle N(R) \rangle \Sigma(R) \quad (27)$$

where  $\langle N(R) \rangle = \langle n \rangle V = \bar{b}_0 V / \mu$  and  $\Sigma(R) := \sigma^2 \psi(R/\bar{\lambda}) / \mu$ . The function  $\Sigma(R)$  is the spatial Fano factor and quantifies the deviations of the fluctuations from a Poisson process. Since  $\sigma^2 / \mu = \mathcal{O}((r-b)^{-1})$ , when the system is close to the critical point – for fixed  $R$  and  $\bar{\lambda}$  – the system has large fluctuations on all scales larger than the correlation length  $\bar{\lambda}$ . Also, in the regime  $R/\bar{\lambda} \rightarrow +\infty$  we obtain  $\mu \Sigma(R) / \sigma^2 = 1 + \mathcal{O}[(R/\bar{\lambda})^{-1}]$ , thus recovering the mean-field fluctuations as predicted by eq.(19).

### A. The spatial Taylor's law

Taylor's law was first observed in ecological communities [24, 25], where natural populations show some degree of spatial aggregation. This was phenomenologically captured by assuming a scaling relationship between the variance and mean of population sizes in different areas. More generally, and recently, Taylor's law denotes any power relation between the variance and the mean of random variables in complex systems [26, 27]. The law postulates a relation of the following form

$$\text{Var}[N(R)] = C \langle N(R) \rangle^\alpha \quad (28)$$

where  $C$  is a positive constant and  $\alpha$  typically assumes values between one and two [25, 27]. The spatial model which we have introduced can predict the behavior of this relation across scales without making specific assumptions. If we focus on the two dimensional case and fix  $\bar{\lambda}$ , we obtain  $\text{Var}[N(R)] = C_1 \langle N(R) \rangle^2 \log(\langle N(R) \rangle)$  for  $R \ll \bar{\lambda}$ , while  $\text{Var}[N(R)] = C_2 \langle N(R) \rangle$  for  $R \gg \bar{\lambda}$ . This latter situation corresponds to the mean field case in which  $C_2 = \sigma^2 / \mu$ . So for areas of radius much smaller than the correlation length the model is characterized by  $\alpha = 2$  with logarithmic corrections, while in the case of radii much larger than  $\bar{\lambda}$  we obtain  $\alpha = 1$ . This is in agreement with previous studies [24–26]. It has been proved that Taylor's law can emerge in a much more general class of stochastic processes, for example when the dynamical rates of the model are affected by environmental variability [26, 27]. In the case of this birth-death spatial model, the cross-over in the scaling exponent  $\alpha$  results from the different roles of dispersal played at different spatial scales.

### VI. SOLUTION OF THE CONDITIONAL PDF

The goal of this section is to derive the main result of the paper. With the spatial population variance ob-

tained in the previous section along with an appropriate approximation, we are now able to close and solve eq.(16) for the conditional generating function  $Z(h|V)$ . We will then invert this latter for the conditional probability density function  $P(N|V)$ , from which several characteristics of the spatial patterns can be deduced.

In order to calculate the explicit form of  $Z(h|V, T)$  from eq.(16), we first need to calculate the solution of eq.(17), which in turn has to satisfy the identity in eq.(13). We will first make use of this latter in the form

$$\int_{\mathcal{V}} f(x, h, V, T) dx = \frac{\partial Z}{\partial h}(h|V, T) \quad . \quad (28)$$

It is not difficult to verify that  $f$  can be expressed as

$$f(x, h, V, T) = \frac{1}{V} \frac{\partial Z}{\partial h} \left[ 1 + \sum_{i=1}^{\infty} h^i A_i(x, V, T) \right] \quad , \quad (29)$$

where the functions  $A_i$  are such that

$$\int_{\mathcal{V}} A_i(x, V, T) dx = 0 \quad (30)$$

for any  $i$ . One could calculate the explicit form of  $A_i$ , which depends on the spatial position, by substituting the expression in eq.(29) into eq.(17). However, the meaning of these functions provides a more efficient way for the calculation. Because of the definition of  $f$ , we readily obtain  $f(x, h=0) = \langle n \rangle$  and

$$\left. \frac{\partial^m f}{\partial h^m}(x, h) \right|_{h=0} = \langle n(x) N^m \rangle \quad (31)$$

for  $m = 1, 2, \dots$ , from which we can make explicit the expression of  $A_i$  by using eq.(29). For instance, it is not difficult to show (see Appendix E) that at stationarity

$$\begin{aligned} A_1(x) &= \frac{1}{\langle n \rangle} \left( \langle n(x) N \rangle - \frac{1}{V} \langle N^2 \rangle \right) = \\ &= \frac{\sigma^2}{\mu} \left[ \Psi \left( \frac{|x|}{\lambda}, \frac{R}{\lambda} \right) - \psi \left( \frac{R}{\lambda} \right) \right] \end{aligned}$$

where  $\Psi$  and  $\psi$  have been defined in eqs.(24) and (26), respectively. Similar relations, though more complicated, hold for  $i = 2, 3, \dots$ . Actually,  $\langle n(x) N^i \rangle$  can be calculated by integrating  $i$  times over the volume  $\mathcal{V}$  the  $(i+1)$ -th spatial correlation function.

Since we are interested in relatively large population sizes, we retain only the first two terms in the bracket of eq.(29). These latter provide an accurate approximation of  $f$  also in other important parameter regimes, namely those of large and small correlation length ( $\bar{\lambda} \rightarrow 0, \infty$ ). Thus,  $f$  at stationarity turns into

$$f(x, h, V) = \frac{1}{V} \frac{\partial Z}{\partial h} \left[ 1 + h A_1(x) \right] \quad , \quad (32)$$

where  $A_1$  is the one we have obtained before. It is remarkable that, when substituting eq.(32) into eq.(16), at stationarity one obtains (see Appendix E)

$$\left( 1 - h \Sigma(R) \right) \frac{\partial Z}{\partial h} = \langle n \rangle V Z \quad , \quad (33)$$

where  $\Sigma(R) = \sigma^2 \psi(R/\bar{\lambda})/\mu$  is the spatial Fano factor defined in the previous section. Therefore, eq.(16) can be closed when considering terms up to  $\mathcal{O}(h)$ . The form of eq.(33) and that of eq.(18) at stationarity are the same, provided we replace  $\sigma^2/\mu$  with  $\Sigma(R)$ . Therefore the solution of eq.(33) is

$$Z(h|R) = \left( 1 - \Sigma(R) h \right)^{-\frac{\langle n \rangle V(R)}{\Sigma(R)}} \quad , \quad (34)$$

which, when inverted for the probability density function, gives a gamma distribution of the form

$$P(N|R) = \left( \frac{1}{\Sigma(R)} \right)^{\frac{\langle n \rangle V(R)}{\Sigma(R)}} \frac{N^{\frac{\langle n \rangle V(R)}{\Sigma(R)} - 1} e^{-\frac{N}{\Sigma(R)}}}{\Gamma \left( \frac{\langle n \rangle V(R)}{\Sigma(R)} \right)} \quad . \quad (35)$$

Thus, while without spatial effects the characteristic scale of the population size is  $\epsilon^{-1} \simeq \sigma^2/\mu \gg 1$ , in this regime space introduces a space-dependent scale for fluctuations which is quantified by  $\Sigma(R) = \epsilon^{-1} \psi(R/\bar{\lambda})$ , where  $\psi$  is the function defined in eq.(26).

As a further insight, if one defines the new process for the random variable  $N(R, T)$  in the volume  $V$  of fixed radius  $R$  by the stochastic differential equation

$$\dot{N}(R) = \bar{b}_0 V(R) - \mu N(R) + \sigma \sqrt{\psi(R/\bar{\lambda})} N(R) \xi(t) \quad , \quad (36)$$

where  $\xi(t)$  is a zero-mean Gaussian white noise and  $\langle \xi(t) \xi(t') \rangle = 2\delta(t-t')$ , then the stationary pdf of  $N(V)$  is exactly eq.(35). Notice that space is taken into account only implicitly through the functions  $V(R)$  and  $\psi(R/\bar{\lambda})$ . Eq.(36) can also be obtained as an  $\epsilon$ -limit of a spatially-implicit master equation along the lines we have showed in Section IV. This process – unlike the spatial one – satisfies the detailed balance condition at stationarity as the flux at  $N = 0$  is set to zero. This result suggests that there are some families of spatially-explicit processes which, when restricted to a finite volume, can be well approximated by spatially-implicit processes. Whilst the former breaks detailed balance, the latter turns out to be simpler and satisfies the detailed balance condition. In this model the region of this approximation is close to the critical point of the process.

It remains to understand when the  $f$  in the form of eq.(32) solves eq.(17), i.e. the original equation for  $f$ . In



Appendix E we show that, for a fixed radius  $R$ , at stationarity  $f$  is a solution when  $h$  is small. Also, eq.(32) is a good approximation when  $\bar{\lambda}$  is finite, either very large or very small compared to  $R$ , regardless of the spatial dimension  $d$ . The limits  $\bar{\lambda} \rightarrow 0$  and  $\bar{\lambda} \rightarrow +\infty$  of eq.(35) lead to the mean-field expressions, respectively eq.(18) and  $P(N|R) = \delta(N - \langle n \rangle V(R))$  at leading order. Thus, eq.(35) captures the leading behavior of the distribution of the random variable  $N(R)$  in the large population regime and in the vicinity of the critical point. The simulations indeed confirm this with very good accuracy as shown in Figs.(2) and (3).

Because we assumed that  $A_1(x)$  is at stationarity, eq.(32) does not give all the correct terms for a time evolution of the process  $N(R, T)$ . However, relatively close to stationarity, even the temporal dynamics is accurately described by eq.(36). We have checked this idea and compared the exact simulations of the process – as provided by the Doob-Gillespie algorithm (see also the following section) – with the corresponding analytic temporal evolution as obtained from solving eq.(36). Assuming that initially in the volume  $V$  there are  $N_0$  individuals, we find the following time-dependent solution (see [28] for the details of the derivation):

$$P(N, t|N_0, 0) = \left( \frac{1}{\Sigma(R)} \right)^{\frac{\bar{v}_0 V}{\mu \Sigma(R)}} N^{\frac{\bar{v}_0 V}{\mu \Sigma(R)} - 1} e^{-\frac{N}{\Sigma(L)}} \frac{\left[ \left( \frac{1}{\Sigma(R)} \right)^2 N_0 N e^{-\mu t} \right]^{\frac{1}{2} - \frac{\bar{v}_0 V}{2\mu \Sigma(L)}}}{1 - e^{-\mu t}} \exp \left[ -\frac{\frac{1}{\Sigma(R)}(N + N_0)e^{-\mu t}}{1 - e^{-\mu t}} \right] I_{\frac{\bar{v}_0 V}{\mu \Sigma(R)} - 1} \left[ \frac{\frac{2}{\Sigma(R)} \sqrt{N_0 N e^{-\mu t}}}{1 - e^{-\mu t}} \right], \quad (37)$$

where we used reflecting boundary conditions at  $N = 0$  at any  $t > 0$ . This pdf indeed tends to the stationary solution in eq.(35) as  $t \rightarrow \infty$ . The agreement between simulations and eq.(37) is showed in Fig.(7) and (8).

### Simulations with an efficient algorithm: the phenomenological algorithm

We have compared our analytic predictions to simulations coming from a range of different parameter sets and from two different simulation schemes.

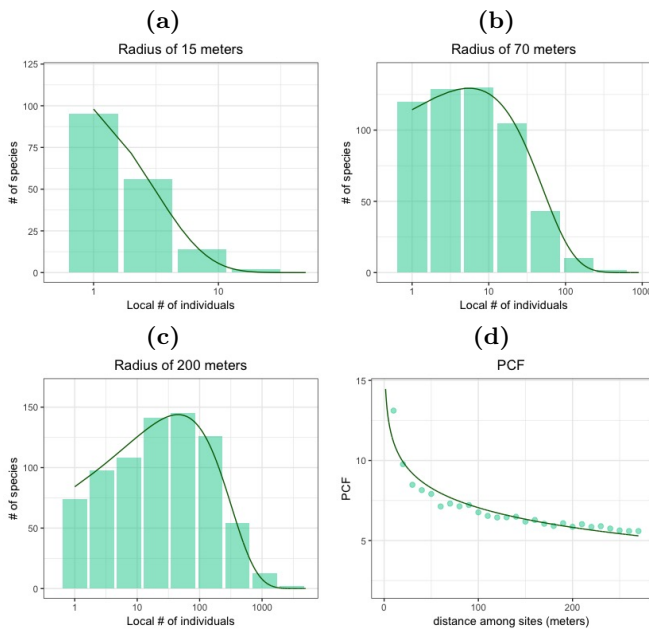
The first is the Doob-Gillespie's algorithm [29] for producing exact trajectories of Markovian processes. We have used periodic boundary conditions in 1- $d$  lattices of various sizes. Parameters were chosen so that the correlation length of the system was much smaller than the total size of the lattice. We have analyzed different sets of parameters and for each one we have run 50,000 independent realizations: error bars were calculated by grouping the results into 50 sets made of 1000 realizations each. In principle this simulation scheme allows us to obtain the exact trajectories of the system at any time, from an initial configuration up to stationarity. However, it is computationally very expensive, and therefore we have been forced to choose relatively small lattice sizes to investigate significant changes from the initial configuration. The results of this are shown in Figs.(7) and (8).

In order to analyze a wider set of parameters and larger lattices, we have simulated the process by using a new algorithm which generates the stationary random field obtained from the stochastic partial differential equation defined on the lattice. This was introduced in [30], and

modifies a previous scheme that was used for simulating models of directed percolation [31]. Here we briefly summarize the main steps of the pseudo-code. From the definition of the discrete Laplace operator, we split the term  $\lambda^2 \Delta_i n_i$  in eq.(10) into one part depending only on  $n_i$  ( $2^d \lambda^2 n_i$ ) and another one depending on the densities in the nearest neighboring sites ( $\lambda^2 \sum_{j:|i-j|=1} n_j$ ). Conditional on the values of  $n_j$  for  $j \neq i$ , the second term is constant and thus  $P(n_i|n_j)$  can be obtained as a gamma distribution at stationarity [30]. Starting from a random initial configuration, at each iteration  $m$ , we randomly select a site  $i$  and update the value of  $n_i^m$  in the next step by sampling from  $n_i^{m+1} \sim P(n_i|n_j^m)$ , conditional on the values of  $n_j^m$  (i.e.  $n_j^{m+1} = n_j^m$ ). These steps are repeated until lattice configurations become independent of initial conditions. While standard stochastic integration schemes fail to preserve the positivity of  $\{n\}$  at any step, this algorithm produces non-negative populations by construction. We have verified that the simulated distribution thus obtained for  $N(R)$  matches the exact simulations of the Doob-Gillespie's algorithm in  $d = 1$  and for different lengths, when the comparison was feasible (see Fig.(9)). By means of this algorithm we were able to study much larger lattice sizes at stationarity, and compare simulations against the predictions of the analytic solutions. The agreement was excellent in all the expected regimes (see Figs.(2) and (3)).

## VII. AN ECOLOGICAL APPLICATION

A simple, but far from trivial, application of the mathematical model we have previously described is the modelling of spatial patterns in ecosystems with a large num-



**FIG. 4: Species abundance distribution from a lowland tropical forest and prediction from the model.** Panels (a), (b) and (c) present the comparison

between the empirical data of the distribution of species' abundances from the lowland tropical forest inventory of the Pasoh natural reserve (Malaysia) and the prediction based on the model. Histograms represent the empirical data while the green lines are the predictions obtained from the solution of the model, i.e. from eq.(35). We highlight that these are not best fits to the empirical data, but genuine predictions that are formulated from the empirical measures of the pair correlation function, as described in the main text (see eq.(6)). The radii of the areas are 15, 70 and 200 meters as reported on the corresponding panels. In panel (d) we compare the empirical pair correlation function (green dots) with the best fit of eq.(6). The correlation length that is thus calculated is  $\bar{\lambda} \approx 2.5 \times 10^3$  meters, with  $\bar{\rho} \approx 8.9 \times 10^3$  and  $\langle n \rangle \approx 6.1 \times 10^{-4}$  trees per square meter for each species. The total number of species in the whole 50Ha forest stand is  $\approx 900$ .

ber of species. Examples include the species richness and abundance distribution of vascular plant species in tropical forest inventories. This model can be used to predict the abundance distribution of tree-species in a forest stand from measures of the two-point correlation function (PCF) and mean abundance per species. Because these latter descriptors are relatively easy to calculate, it turns out that we can obtain an estimate of how many rare (or abundant) species live in a region without surveying the entire system. Therefore, as well as being theoretically interesting, this approach has also an important practical advantage, because it ultimately allows one to infer the total number of species within a very large spatial region by utilizing only scattered and small-scale samples of the

region itself. This is a long-lasting problem which has received a lot of attention recently [6, 15, 20].

Our goal here is not to explain the upscaling method, but only show to what degree the spatial model is in agreement with ecological empirical data. The two-point correlation function (PCF) specifies how similar individuals are as a function of the geographic distance. Usually, it is more likely that close-by individuals belong to the same species than individuals that live farther apart. This translates into a PCF that is always positive, but decays with distance [32, 33]. A decline in similarity with increasing geographic distance indicates that the individuals of a community are spatially aggregated. Therefore, a simple random placement of individuals in space is not a good approximation of the configurations of the community as thought in the past [34]. On the contrary, the stationarity PCF of this model decays with distance, is always positive (see eq.(6)) and has two free parameters ( $\bar{\lambda}$  and  $\bar{\rho}$ ). When we calculate these latter from a best fit of the empirical data, we obtain a good agreement as shown in Fig.(4d).

The mean abundance per species is readily available, because the total number of species and individuals are known in this forest plot. Since at stationarity the model is fully specified by these three parameters ( $\langle n \rangle$ ,  $\bar{\lambda}$  and  $\bar{\rho}$ ), we are henceforth able to predict all the stationary patterns that we like to compare with those of the empirical ecosystem. One of them is the probability that a species has a given number of individuals within a specific region. In the ecological literature it is often referred to as species abundance distribution (SAD) [35]. In our model it is given by eq.(35) and the histograms of this pattern are reported in Fig.(4a-c) for different radii. The SAD represents one of the most commonly used static measures for summarizing information on ecosystem's diversity. Interestingly, the shape of the SAD in tropical forests has been observed to maintain similar features, regardless of the geographical location or the details of species interactions. Indeed, it often displays a unimodal shape at larger scales and a peak at small abundances at relatively small spatial scales. Numerous papers have focused on evaluating the processes that generate and maintain such observed characteristics [20, 36], but only few of them considered spatial effects in an explicit framework [19, 37].

The panels in Fig.4 show a comparison between the empirical data from the forest inventory of Pasoh Natural reserve in Malaysia (year 2005) and the predicted abundance distribution of species obtained from the model we have described in the previous sections. It is remarkable that the predicted curves in the first three panels are genuine inferences obtained from the mean abundance per species (i.e.,  $\langle n \rangle$ ) and the PCF (i.e.,  $\bar{\lambda}$  and  $\bar{\rho}$ ), and not best fitted curves to empirical SAD.

Such an agreement confirms that abundances of species are indeed characterized by very large fluctuations and that local populations appear correlated over very large spatial scales. This entails that complex ecosystems may

comprise a large number of rare species, whereas only a few have large abundances (hyperdominant species). Because this model does not explicitly include interactions nor environmental forcing, we cannot single out which ecological processes bring about this finding. Nevertheless, the model is able to explain this separation of population size scales by poisoning a system close to criticality.

The simplicity and generality of the model make it suitable for the description of patterns in other biological systems. Indeed, numerous recent studies have showed that spatial bio-geographical patterns emerge in marine ecosystems [38, 39], microorganisms [40], including bacteria [41, 42], archaea, viruses, fungi [43] and eukaryotes [44, 45]. Future work will include the application of the current framework to those biological communities.

### VIII. CONCLUSIONS

In this paper we have studied a spatial stochastic model which can be fruitfully used to describe the main large scale characteristics of species-rich ecosystems. We have shown how to calculate analytically some of the most important spatial patterns when the system is close to criticality, which is the regime where the most important features emerge.

The model encapsulates birth, death, immigration and local hopping of individuals. It describes the dynamics of point-like and well-mixed individuals living in a metacommunity defined on a  $d$ -dimensional regular graph. The model is also minimal, meaning that, without one of its components (i.e., birth, death, nearest-neighbor hopping and external immigration), it yields either trivial or well-known results. Despite its simplicity, however, it violates detailed balance and generates non-trivial spatial patterns when close to its critical point. These patterns are substantially due to stochastic effects and entail large fluctuations correlated on large spatial and temporal scales.

The linearity of the rates is not sufficient to derive a full solution (in a weak sense) of the model. However, in applications one is usually interested in the analytical properties of processes that are much less general than the spatial random field. Thus we restricted our analysis to the conditional distribution,  $p(N|V)$ , that  $N$  individuals are found in a volume  $V$ . This quantity is sufficiently general to describe a wealth of patterns in several systems. We have found that in the close-to-critical regime,  $p(N|V)$  satisfies an aptly derived equation which has the form of the corresponding mean-field equation of the process (i.e., without space). This equation includes functions of  $V$ , which we have exactly calculated. Such spatial redefinition of the parameters introduces strong deviations from the corresponding mean-field solutions, as confirmed by the exact stochastic simulations. Also, this shows that the process that governs the random variable  $N$  satisfies detailed balance in a first approximation and close to the critical point, thus being considerably

simpler than the distribution of the random field.

This result suggests the tantalizing hypothesis that the conditional distribution,  $p(N|V)$ , provided by some families of spatial stochastic processes, is described by much simpler processes within a specific region of the parameter space. In our model the region of this approximation is close to the critical point of the process and the distribution of the simpler process holds the same functional shape across all spatial scales. Of course, the hypothesis requires much more scrutiny, especially when spatial models include nonlinear terms which can jeopardize the methods developed here. On the other side, our approach allows for the analysis of several generalizations, including non-local dispersal kernels and different sources of noise (e.g., environmental noise).

On a more applied side, we have shown that when applying the model to large ecosystems, species are predicted to display a broad range of abundances as a consequence of the critical regime. Therefore, many of them are rare and only a few are very abundant. These large demographic fluctuations are correlated across large spatial scales as well as over long times, in agreement with several empirical datasets collected in well-known tropical forest inventories [6, 30, 46].

Of course, these findings do not prove that these biological systems are close to criticality, but suggest that it is worth pursuing that route further. In this article we have not investigated the reasons why those systems look nearly critical, nor how they can operate within such tiny regions of their parameter space. For this, one needs to look into how inter- and intra-interactions dynamically lead living systems towards the correct region, in which states are biologically meaningful – however, see [47–50]. Nonetheless, our results could help understand what key factors drive such dynamics, and possibly shed light on the importance and effects of non-linearities among interacting agents.

### IX. ACKNOWLEDGMENTS

FP thanks the NERC SPHERES DTP (NE/L002574/1) for funding his studentship. We thank Amos Maritan for insightful discussions.

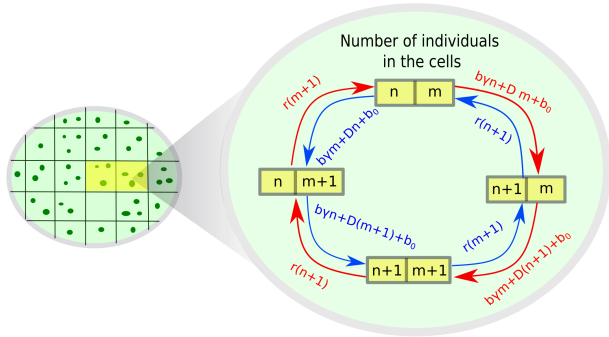
### Appendix A BROKEN DETAILED BALANCE

We here recall the main general properties of detailed balance. Let us denote with  $\mathbf{c}$  the configuration of a generic stochastic process and indicate with  $p(\mathbf{c}, t|\mathbf{c}_0, t_0)$  the probability that the configuration  $\mathbf{c}$  is seen at time  $t$ , given that the configuration at time  $t_0$  was  $\mathbf{c}_0$  (abbreviated  $p(\mathbf{c}, t)$ ). We introduce  $\mathcal{W}(\mathbf{c}'|\mathbf{c})$ , which is the (time-independent) rate to transit from state  $\mathbf{c}$  to  $\mathbf{c}'$ . If we consider Markovian dynamics, the evolution of  $p(\mathbf{c}, t)$

is given by the following master equation (ME) [22, 23]

$$\frac{\partial p(\mathbf{c}, t)}{\partial t} = \sum_{\mathbf{c}'} \left[ \mathcal{W}(\mathbf{c}|\mathbf{c}')p(\mathbf{c}', t) - \mathcal{W}(\mathbf{c}'|\mathbf{c})p(\mathbf{c}, t) \right] . \quad (38)$$

If it happens that  $\mathcal{W}(\mathbf{c}|\mathbf{c}')P(\mathbf{c}') - \mathcal{W}(\mathbf{c}'|\mathbf{c})P(\mathbf{c}) = 0$  for all configurations (this is the detailed balance (DB) condition), then the probability distribution  $P(\mathbf{c})$  is also a stationary solution of the ME, as we see from eq.(38). On the other hand, it is clear that not all stationary distributions satisfy the detailed balance condition.



**FIG. 5: Broken Detailed Balance.** The picture represents a closed path in the space of configurations of the process which has different probabilities depending on the direction of the path (see also [17]). We take two neighboring sites, which initially have  $n$  and  $m$  individuals. The rates of jumping into the next configuration of the path are reported in the image and the arrows indicate the direction. In this simple case  $D = b(1 - \gamma)$ . In the clockwise direction (red arrows) the total rate is  $[b\gamma n + Dm + b_0][b\gamma m + D(n+1) + b_0][r(n+1)][r(m+1)]$ . In the anti-clockwise direction (blue arrows) the total rate is  $[b\gamma m + Dn + b_0][b\gamma n + D(m+1) + b_0][r(m+1)][r(n+1)]$ . These two total rates must be equal for the detailed balance to hold. However, for any arbitrary configuration ( $m \neq n$ ) this is true only when  $D(D - b\gamma)r = 0$ , i.e. for  $b = 0$ ,  $r = 0$ ,  $\gamma = 1, 1/2$ .

It is possible to show [12] that a condition for the validity of DB is that the probability of following a closed path in the space of configurations does not depend on the orientation of the path. More precisely, DB is satisfied if and only if for any choice of a closed path  $\{\mathbf{c}_1, \dots, \mathbf{c}_m\}$ , with  $m$  an arbitrary number, the following holds

$$\begin{aligned} \mathcal{W}(\mathbf{c}_1|\mathbf{c}_2)\mathcal{W}(\mathbf{c}_2|\mathbf{c}_3) \cdots \mathcal{W}(\mathbf{c}_m|\mathbf{c}_1) &= \quad (39) \\ = \mathcal{W}(\mathbf{c}_1|\mathbf{c}_m)\mathcal{W}(\mathbf{c}_m|\mathbf{c}_{m-1}) \cdots \mathcal{W}(\mathbf{c}_2|\mathbf{c}_1) . \end{aligned}$$

This equation corresponds to microscopic reversibility and, when it is violated, the system can be found in non-equilibrium steady states.

Violation of DB is key to living systems and is the subject of intense recent interest [51, 52]. In fig. (5) we show

with a simple example that for this model the condition in eq.(39) is not satisfied, hence the detailed balance condition does not hold: it is possible to find a path in the space of configurations where the total rate does depend on the orientation of the closed path. Fig. (5) shows one of such paths. Notice that such counterexample does not hold when spatial dispersal is switched off (i.e.  $D = 0$  or  $\gamma = 1$ ), or when autocatalytic production and spatial dispersal rates are equal ( $\gamma = 1/2$ ).

## Appendix B SPATIAL CORRELATION OF THE BIRTH-DEATH MARKOV PROCESS

In the main text we have defined the spatial generating function of the model

$$\begin{aligned} \zeta(\{H\}, t) &= \langle e^{\sum_{k \in \mathcal{L}} n_k H_k} \rangle = \\ &= \sum_{\{n\}} e^{\sum_{k \in \mathcal{L}} n_k H_k} p(\{n\}, t) . \end{aligned}$$

From eq.(2), multiplying both sides through by  $e^{\sum_{s \in \mathcal{L}} n_s H_s}$  and summing over all states, we find eq.(3) of the main text. If we differentiate by  $H_k$  and impose  $\{H\} = 0$ , we find the equation for the mean number of individuals  $\langle n_k \rangle$ , reported in the main text. Taking another derivative with respect to  $H_l$  and setting  $\{H\} = 0$ , we obtain the equation for the spatial two-point correlation function (PCF) among the sites  $k$  and  $l$

$$\begin{aligned} \frac{\partial}{\partial t} \langle n_k n_l \rangle &= D \left( \Delta_l \langle n_k n_l \rangle + \Delta_k \langle n_k n_l \rangle \right) - 2\mu \langle n_k n_l \rangle + \\ &+ 2b_0 \langle n \rangle + \delta_{k,l} \left( 2\sigma^2 \langle n \rangle + b_0 + D\Delta_k \langle n_k \rangle \right) , \end{aligned}$$

where  $\sigma^2 := \frac{b+r}{2}$ ,  $D := \frac{b(1-\gamma)}{2d}$ ,  $\delta_{k,l}$  is a Kronecker delta, and  $\Delta$  is the discrete Laplace operator as defined in the main text.

Considering stationary patterns, because of homogeneity we have  $\Delta_k \langle n_k \rangle = 0$  and, introducing  $G_{k,l} = \langle n_k n_l \rangle - \langle n \rangle^2$ , we obtain

$$\begin{aligned} D(\Delta_k G_{k,l} + \Delta_l G_{k,l}) - 2\mu G_{k,l} + \\ + \left( 2\sigma^2 \langle n \rangle + b_0 \right) \delta_{k,l} = 0 . \end{aligned}$$

We will now consider the Fourier series expansion of  $G_{\mathbf{k},\mathbf{l}}$ , which we will write as

$$G_{\mathbf{k},\mathbf{l}} = \left( \frac{a}{2\pi} \right)^d \int_{\mathcal{C}} d\mathbf{p} \hat{G}(\mathbf{p}) e^{i\mathbf{p} \cdot (\mathbf{k} - \mathbf{l})} ,$$

where  $\mathbf{k}, \mathbf{l}$  are the Cartesian coordinates of the locations of sites on the lattice (in  $a$  units),  $\mathbf{p}$  is a vector with  $d$  components which belongs to  $\mathcal{C}$ , the hypercubic  $d$ -dimensional primitive unit cell of size  $2\pi/a$ . Upon substituting  $G_{\mathbf{k},\mathbf{l}}$  in the stationary equation, we get an expression for  $\hat{G}(\mathbf{p})$ , which is

$$\hat{G}(\mathbf{p}) = \left( \frac{\sigma^2 \langle n \rangle + b_0}{\mu} \right) \frac{1}{1 + \frac{2D}{\mu} \sum_{i=1}^d (1 - \cos(p_i a))} ,$$

where  $p_i$  is the  $i$ -th component of  $\mathbf{p}$ . Therefore

$$G_{\mathbf{k},\mathbf{l}} = \left(\frac{a}{2\pi}\right)^d \int_{\mathcal{C}} d\mathbf{p} \frac{\left(\frac{\sigma^2}{\mu}\langle n \rangle + \frac{b_0}{2\mu}\right) e^{i\mathbf{p}\cdot(\mathbf{k}-\mathbf{l})}}{1 + \frac{2D}{\mu} \sum_{i=1}^d (1 - \cos(p_i a))} .$$

We can obtain a good deal of simplification by taking a continuous spatial limit ( $a \rightarrow 0$ ). Renaming  $\mathbf{p} \rightarrow p$ ,  $\mathbf{k} \rightarrow x$  and  $\mathbf{l} \rightarrow y$  (now continuous variables in  $\mathbb{R}^d$ ), and appropriately rescaling the constants as explained in the main text, we arrive at the expression for the pairwise spatial correlation in dim  $d$ , i.e.

$$G(x, y) = \frac{\bar{\rho}^2 \langle n \rangle^2}{(2\pi)^d} \left(1 + \frac{\mu}{2\sigma^2}\right) \int dp \frac{e^{ip\cdot(x-y)}}{1 + \bar{\lambda}^2 p^2} ,$$

where we have replaced  $b_0$  with  $\bar{b}_0$  in  $\langle n \rangle$  and used the definitions of  $\bar{\rho}$ ,  $\bar{\lambda}$  given in the main text. Actually, the  $d$ -dim integral in the previous expression can be calculated

$$\begin{aligned} \int dp \frac{e^{ip\cdot(x-y)}}{1 + \bar{\lambda}^2 p^2} &= \int_0^\infty ds \int dp e^{ip\cdot(x-y) - (1 + \bar{\lambda}^2 p^2)s} \\ &= \int_0^\infty ds \left(\frac{\pi}{s}\right)^{\frac{d}{2}} e^{-\frac{|x-y|^2}{4\bar{\lambda}^2 s} - s} \\ &= \frac{(2\pi)^{d/2}}{\bar{\lambda}^d} \left(\frac{|x-y|}{\bar{\lambda}}\right)^{\frac{2-d}{2}} K_{\frac{2-d}{2}}\left(\frac{|x-y|}{\bar{\lambda}}\right) \end{aligned}$$

where  $K_a(z)$  is the modified Bessel function of the second kind and in the last step we used 9.6.24 from [53]. Eventually, the PCF is

$$G(x, y) = \frac{\bar{\rho}^2 \langle n \rangle^2}{(2\pi \bar{\lambda}^2)^{d/2}} \left(1 + \frac{\mu}{2\sigma^2}\right) \times \left(\frac{|x-y|}{\bar{\lambda}}\right)^{\frac{2-d}{2}} K_{\frac{2-d}{2}}\left(\frac{|x-y|}{\bar{\lambda}}\right) . \quad (40)$$

Because  $K_a(z)$  decays exponentially for large  $z$  regardless of  $a$ ,  $\bar{\lambda}$  plays the role of a correlation length of the spatial system. Instead,  $\bar{\rho}^2$  has dimensions of a  $d$ -dim volume and provides a characteristic volume of local demographic fluctuations.

### Appendix C APPROXIMATING $g(i, j, h, t)$ IN THE CONTINUOUS SPATIAL LIMIT

Equation (14) shows that the evolution of  $f(k, h, t)$  is not closed, being coupled to the generating function  $Z$  and to the function  $g(i, k, h, t) = \langle n_i n_k e^{hN(V)} \rangle$ . In this section we want to show that, at leading order in the limit  $a \rightarrow 0$ , the value of  $\Delta_i g(i, k, h, t)$  approaches  $\Delta_k g(i, k, h, t)$ . This makes the equation for  $f(k, h, t)$  much simpler, basically decoupling it with the other quantities. For simplicity we will only consider the one dimensional case, but this claim holds true in higher dimensions as well. Thus  $\mathcal{V}$  will be an interval of length  $2R$

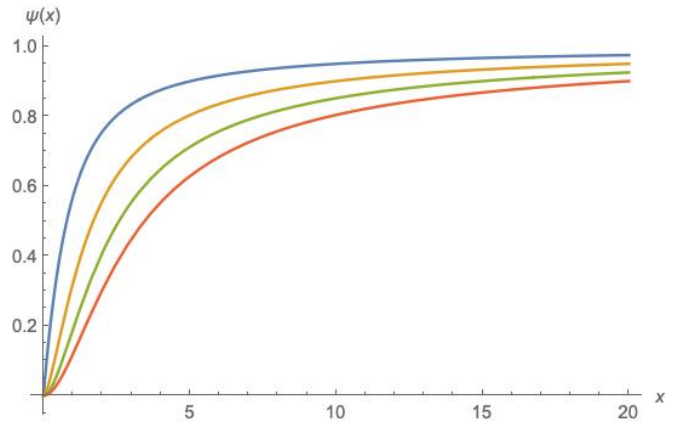


FIG. 6: **Behavior of the function  $\psi(x)$ .** The figure shows the behavior of  $\psi(x)$  as defined in eq.(26) in dim  $d = 1$  (blue solid curve),  $d = 2$  (yellow solid curve),  $d = 3$  (green solid curve) and  $d = 4$  (red solid curve).

and the origin of the coordinate system will be located at its center. By changing slightly the notation, we will now indicate  $N(V) = N(-R, R)$ , thus making explicit that  $V$  extends from site  $-R$  to  $R$ . Since the system is spatially homogeneous, we can write

$$\begin{aligned} \langle n_{i-a} n_k e^{hN(-R,R)} \rangle &= \\ &= \langle n_i n_{k+a} e^{hN(-R+a, R+a)} \rangle \end{aligned}$$

As  $a \rightarrow 0$  and at leading order, we can neglect  $a$  in the argument of  $N(-R, R)$ , thus obtaining

$$\begin{aligned} g(i-a, k, h, t) &= \langle n_{i-a} n_k e^{hN(-R,R)} \rangle = \\ &= \langle n_i n_{k+a} e^{hN(-R,R)} \rangle = \\ &= g(i, k+a, h, t) \end{aligned}$$

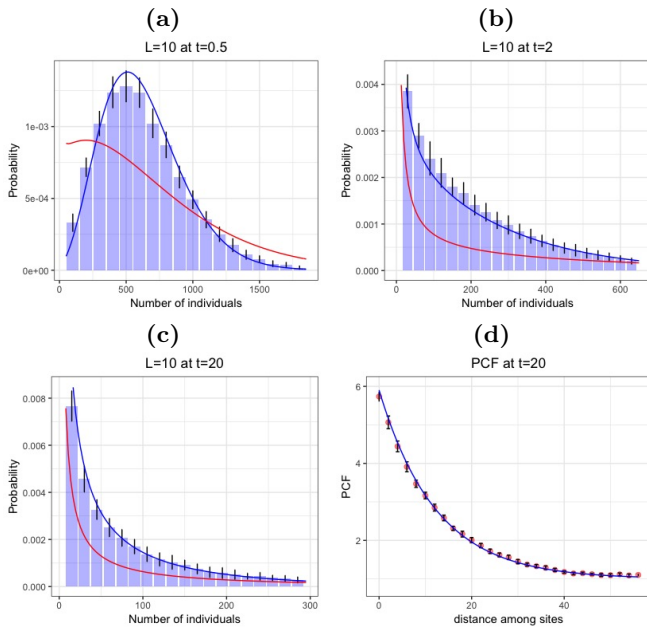
Similarly  $\langle n_{i+a} n_k e^{hN(-R,R)} \rangle = \langle n_i n_{k-a} e^{hN(-R,R)} \rangle$ . Thus, as  $a \rightarrow 0$  at leading order we can write

$$\begin{aligned} \Delta_i g(i, k) &= g(i+a, k) + g(i-a, k) - 2g(i, k) = \\ &= g(i, k-a) + g(i, k+a) - 2g(i, k) \\ &= \Delta_k g(i, k) . \end{aligned}$$

This approximation makes it possible to obtain eq.(17) in the continuous limit.

### Appendix D MULTIDIMENSIONAL VARIANCE

In the main text we have outlined how to calculate the second moment of the random variable  $N(V)$ . Here we provide some more details. We will take  $\mathcal{V}$  to be a  $d$ -dimensional ball with the origin of the Cartesian coordinates in its center. By integrating  $\langle n(x)n(y) \rangle$  over  $y \in \mathcal{V}$  in eq.(21) and using the symmetry of  $\langle n(x)n(y) \rangle$  in  $x$  and  $y$ , we obtain an equation for  $\langle n_x N(V) \rangle$ , which at stationarity reads



**FIG. 7: Time evolution with the Doob-Gillespie algorithm.** Panels (a), (b) and (c) present a comparison between the simulated model as obtained from the Doob-Gillespie algorithm (histograms), the analytic prediction as defined in eq.(37) (blue solid line), and the mean field solution (red line). Simulations were carried out in  $d = 1$  with periodic boundary conditions, where each site initially contained exactly 100 individuals, with a total number of 200 lattice sites;  $L$  comprises 10 adjacent sites and time  $t$  is expressed in units of  $\mu^{-1}$ . The size of error bars are twice as much the standard deviation. Results in panel (d) are from the same set of simulated data (red dots, with error bars) and compares simulations to the analytic curve (blue line) of the pair correlation function at large times ( $t = 20$ ). Parameters in these simulations are  $b = 600$ ,  $r = 601$ , with  $\gamma = 0.5$  and  $b_0 = 5$  ( $\lambda \approx 12$ ).

$$\frac{\bar{D}}{\mu} \nabla_x^2 \langle n(x)N(V) \rangle - \langle n(x)N(V) \rangle + \frac{\bar{b}_0}{\mu} \frac{\bar{b}_0 V}{\mu} + \frac{\sigma^2 \bar{b}_0}{\mu} \Theta(R - |x|) = 0 \quad , \quad (41)$$

where  $\Theta(z)$  is the Heaviside step function and  $V$  is the  $d$ -dim volume

$$V = \frac{\pi^{d/2}}{\Gamma\left(\frac{d}{2} + 1\right)} R^d$$

with radius  $R$ . Eq.(41) must be solved for  $R < |x|$  and  $R > |x|$  separately. Boundary conditions for  $\langle n(x)N(V) \rangle$  and its first derivative will fix the values of the integrating constants. For  $|x| < R$  we obtain

$$\langle n(x)N(V) \rangle = \left(\frac{\bar{b}_0}{\mu}\right) \frac{\bar{b}_0 V}{\mu} + \frac{\sigma^2 \bar{b}_0}{\mu} + A \left(\frac{|x|}{\lambda}\right)^{1-\frac{d}{2}} I_{\frac{d}{2}-1}\left(\frac{|x|}{\lambda}\right)$$

while for  $|x| > R$  instead

$$\langle n(x)N(V) \rangle = B \left(\frac{|x|}{\lambda}\right)^{1-\frac{d}{2}} K_{\frac{d}{2}-1}\left(\frac{|x|}{\lambda}\right) + \left(\frac{\bar{b}_0}{\mu}\right) \frac{\bar{b}_0 V}{\mu} \quad .$$

The constants  $A$  and  $B$  will be fixed using the aforementioned continuity conditions. Upon explicit calculation, for  $|x| \leq R$  we obtain

$$\begin{aligned} \langle n(x)N(V) \rangle &= \left(\frac{\bar{b}_0}{\mu}\right)^2 V + \frac{\sigma^2 \bar{b}_0}{\mu} \frac{\bar{b}_0 V}{\mu} \left[ 1 - \left(\frac{|x|}{R}\right)^{1-\frac{d}{2}} \frac{K_{\frac{d}{2}}\left(\frac{R}{\lambda}\right) I_{\frac{d}{2}-1}\left(\frac{|x|}{\lambda}\right)}{I_{\frac{d}{2}-1}\left(\frac{R}{\lambda}\right) K_{\frac{d}{2}}\left(\frac{R}{\lambda}\right) + I_{\frac{d}{2}}\left(\frac{R}{\lambda}\right) K_{\frac{d}{2}-1}\left(\frac{R}{\lambda}\right)} \right] \\ &= \left(\frac{\bar{b}_0}{\mu}\right)^2 V + \frac{\sigma^2 \bar{b}_0}{\mu} \frac{\bar{b}_0 V}{\mu} \Psi\left(\frac{|x|}{\lambda}, \frac{R}{\lambda}\right) \quad , \end{aligned} \quad (42)$$

where  $\Psi(a, b)$  was defined in eq.(24) of the main text and  $I_\nu(z)$  and  $K_\nu(z)$  are the modified Bessel functions of the first and second kind, respectively [21]. Integrating with respect to  $x \in \mathcal{V}$  and using the properties of  $I_\nu(z)$ , we can readily obtain the explicit form of the second moment of  $N(V)$ , i.e.

$$\langle N(V)^2 \rangle = \left(\frac{\bar{b}_0 V}{\mu}\right)^2 + \frac{\sigma^2 \bar{b}_0 V}{\mu} \frac{\bar{b}_0 V}{\mu} \left[ 1 - \frac{d \bar{\lambda}}{R} \frac{K_{\frac{d}{2}}\left(\frac{R}{\lambda}\right) I_{\frac{d}{2}}\left(\frac{R}{\lambda}\right)}{I_{\frac{d}{2}-1}\left(\frac{R}{\lambda}\right) K_{\frac{d}{2}}\left(\frac{R}{\lambda}\right) + I_{\frac{d}{2}}\left(\frac{R}{\lambda}\right) K_{\frac{d}{2}-1}\left(\frac{R}{\lambda}\right)} \right] \quad , \quad (43)$$

where we can read off the explicit expression for  $\psi(R/\lambda)$ , also reported in eq.(26). Finally, since  $\text{Var}(N(V)) = \langle N(V)^2 \rangle - \langle N(V) \rangle^2$  we can explicitly write down the variance of  $N(V)$ , which takes the form

$$\text{Var}(N(V)) = \frac{\bar{b}_0 V \sigma^2}{\mu} \left[ 1 - \frac{d \bar{\lambda}}{R} \frac{K_{\frac{d}{2}}\left(\frac{R}{\bar{\lambda}}\right) I_{\frac{d}{2}}\left(\frac{R}{\bar{\lambda}}\right)}{I_{\frac{d}{2}-1}\left(\frac{R}{\bar{\lambda}}\right) K_{\frac{d}{2}}\left(\frac{R}{\bar{\lambda}}\right) + I_{\frac{d}{2}}\left(\frac{R}{\bar{\lambda}}\right) K_{\frac{d}{2}-1}\left(\frac{R}{\bar{\lambda}}\right)} \right] = \langle N(V) \rangle \Sigma(R) \quad ,$$

where  $\Sigma(R/\bar{\lambda}) = \sigma^2 \psi(R/\bar{\lambda})/\mu$ .  $\Sigma(R/\bar{\lambda})$  is the spatial Fano factor and quantifies the relative importance of fluctuations in the system. Finally, Fig.(6) shows the behavior of  $\psi(x)$ .

## Appendix E EVALUATION OF THE REGIMES OF ACCURACY OF THE METHOD

In this section we show that the truncation of  $f(x, V, h)$  as in eq.(32) yields an accurate approximation for the conditional probability distribution of the model. By retaining the first two terms in the square brackets of eq.(29), at stationarity we are left with the following

$$f(x, V, h) = \frac{1}{V} \frac{\partial Z}{\partial h} \left[ 1 + h A_1(x, R) \right] \quad , \quad (44)$$

where  $Z(h)$  is the conditional generating function at stationarity. By taking the derivative with respect to  $h$  of both sides of eq.(44) and setting  $h = 0$ , we obtain

$$\left. \frac{\partial f}{\partial h} \right|_{h=0} = \langle n_x N(V) \rangle = \frac{1}{V} \langle N(V)^2 \rangle + \langle n \rangle A_1(x, R)$$

which gives

$$A_1(x, V) = \frac{1}{\langle n \rangle} \left( \langle n(x) N(V) \rangle - \frac{1}{V} \langle N(V)^2 \rangle \right) \quad .$$

Because in Appendix D we have already calculated  $\langle n(x) N(V) \rangle$  and  $\langle N(V)^2 \rangle$  (see eqs.(42) and (43)),  $A_1(x, V)$  is known explicitly. Substituting  $f(x, V, h)$  in eq.(44) with  $A_1(x, V)$  obtained before into eq.(16) at stationarity, we get

$$\begin{aligned} \frac{h}{\langle n \rangle V} \frac{\partial Z}{\partial h} \bar{\lambda}^2 \int_{\mathcal{V}} dx \nabla_x^2 \langle n(x) N(V) \rangle - \frac{\partial Z}{\partial h} + \\ + \langle n \rangle V Z + \frac{\sigma^2}{\mu} h \frac{\partial Z}{\partial h} = 0 \quad . \end{aligned}$$

Now we can readily simplify the term  $\bar{\lambda}^2 \nabla_x^2 \langle n(x) N(V) \rangle$  by making use of eq.(41). Integrating this latter with respect to  $x$  in  $\mathcal{V}$ , since  $\langle N(R) \rangle = \langle n \rangle V$  and  $\Sigma(R) = (\langle N(R)^2 \rangle - \langle N(R) \rangle^2)/\langle N(R) \rangle$  (see Appendix D), we are therefore left with the following equation

$$\left( 1 - h \Sigma(R) \right) \frac{\partial Z}{\partial h} = \langle n \rangle V Z \quad , \quad (45)$$

which therefore provides the equation for the generating function of  $N(R)$  up to terms  $\mathcal{O}(h)$ .

Similarly to what we have done so far, we can get further insight into the evolution of  $f(x, V, h)$ . We substitute  $f(x, V, h)$  from eq.(44) into eq.(17) and use eq.(41) to

obtain  $\bar{\lambda}^2 \nabla_x^2 \langle n_x N(V) \rangle$ . Eventually, at stationarity this yields

$$\begin{aligned} \frac{\partial}{\partial h} \left\{ h \left[ - \left( 1 - h \Sigma(R) \right) \frac{\partial Z}{\partial h} + \langle n \rangle L Z \right] \right\} + \\ + A_1(x, R) \left\{ \frac{\partial}{\partial h} \left[ \frac{\sigma^2}{\mu} h^3 \frac{\partial Z}{\partial h} \right] + h^2 \langle n \rangle V \frac{\partial Z}{\partial h} \right\} = 0 \quad . \end{aligned} \quad (46)$$

The first addend of eq.(46) is zero because of eq.(45), and the remaining terms are  $\mathcal{O}(h^2)$ , so  $f$  solves eq.(17) up to terms  $\mathcal{O}(h)$ .

On the other hand eq.(46) shows that  $f(x, V, h)$  is a good approximation of the solution of eq.(17) when  $A_1(x, R)$  approaches zero. Fixing the values of  $|x|$ ,  $R$  and  $\sigma^2/\mu$ , we can calculate explicitly the regimes of  $\bar{\lambda}$  where this condition is verified. From eqs.(24) and (26) in the main text it is easy to rewrite  $A_1(x, R)$

$$A_1(x, R) = \frac{\sigma^2}{\mu} \left[ \Psi \left( \frac{|x|}{\bar{\lambda}}, \frac{R}{\bar{\lambda}} \right) - \psi \left( \frac{R}{\bar{\lambda}} \right) \right]$$

The asymptotic expansion of the modified Bessel functions is [21]

$$\begin{aligned} I_\nu(z) &= \frac{e^z}{\sqrt{2\pi z}} \left( 1 + \mathcal{O}[1/z] \right) \\ K_\nu(z) &= e^{-z} \sqrt{\frac{\pi}{2z}} \left( 1 + \mathcal{O}[1/z] \right) \end{aligned}$$

when  $z \rightarrow \infty$ . From this it is not difficult to verify that as  $\bar{\lambda} \rightarrow 0$  we have

$$\begin{aligned} \Psi \left( \frac{|x|}{\bar{\lambda}}, \frac{R}{\bar{\lambda}} \right) &= 1 + \mathcal{O} \left[ e^{-\frac{R-|x|}{\bar{\lambda}}} \left( \frac{|x|}{R} \right)^{1-d} \right] \\ \psi \left( \frac{R}{\bar{\lambda}} \right) &= 1 + \mathcal{O} \left( \frac{\bar{\lambda}}{R} \right) \end{aligned}$$

and so for  $|x| < R$  indeed  $A_1(x, R) \rightarrow 0$ .

The case of  $\bar{\lambda} \rightarrow \infty$  is more elaborate. Let's call  $z = R/\bar{\lambda}$ . After some lengthy but otherwise straightforward calculations we can verify that as  $z \rightarrow 0$  we can write

$$\psi(z) = \begin{cases} z + \mathcal{O}(z^2) & \text{for } d = 1 \\ -\frac{z^2}{2} \log(z) + \mathcal{O}(z^2) & \text{for } d = 2 \\ \frac{2z^2}{5} + \mathcal{O}(z^3) & \text{for } d = 3 \end{cases}$$

and at leading order  $\psi(z)$  is proportional to  $z^2$  for  $d \geq 3$ . For the case of  $\Psi$  let us write  $z_x = |x|/\bar{\lambda}$ . We can verify

that

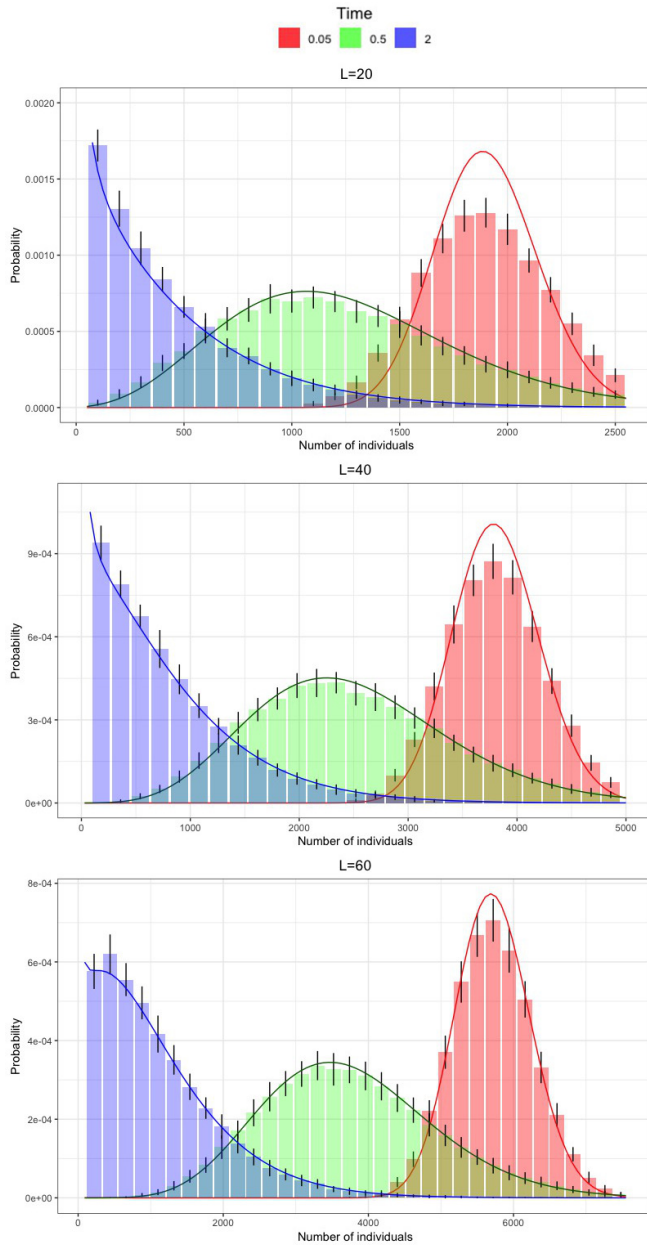
$$\Psi(z_x, z) = \begin{cases} z + \mathcal{O}(z^2, z_x^2) & \text{for } d = 1 \\ -\frac{z^2}{2} \log(z) + \mathcal{O}(z^3, z_x^2) & \text{for } d = 2 \\ \frac{z^2}{3} + \mathcal{O}(z^3, z_x^2) & \text{for } d = 3 \end{cases}$$

and at leading order  $\Psi$  is proportional to  $z^2$  for  $d \geq 3$ . Thus again  $A_1(x, R) \rightarrow 0$ . These findings follow from that  $f$  goes into mean-field regimes as  $\lambda \rightarrow 0, \infty$  and hence all spatial terms go to zero.

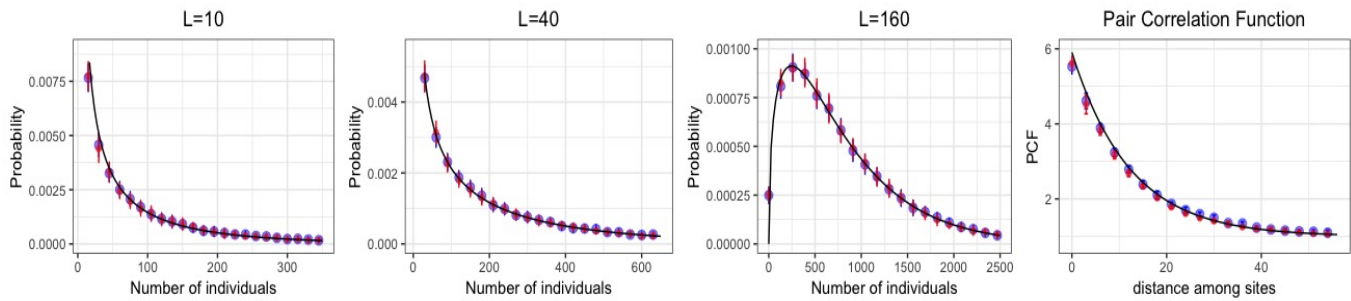
- 
- [1] Thierry Mora and William Bialek. Are biological systems poised at criticality? *Journal of Statistical Physics*, 144(2):268–302, 2011.
- [2] Matti Nykter, Nathan D Price, Maximino Aldana, Stephen A Ramsey, Stuart A Kauffman, Leroy E Hood, Olli Yli-Harja, and Ilya Shmulevich. Gene expression dynamics in the macrophage exhibit criticality. *Proceedings of the National Academy of Sciences*, 105(6):1897–1900, 2008.
- [3] Chikara Furusawa and Kunihiko Kaneko. Adaptation to optimal cell growth through self-organized criticality. *Physical review letters*, 108(20):208103, 2012.
- [4] Elad Schneidman, Michael J Berry II, Ronen Segev, and William Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007, 2006.
- [5] Andrea Cavagna, Alessio Cimarrelli, Irene Giardina, Giorgio Parisi, Raffaele Santagati, Fabio Stefanini, and Massimiliano Viale. Scale-free correlations in starling flocks. *Proceedings of the National Academy of Sciences*, 107(26):11865–11870, 2010.
- [6] Anna Tovo, Samir Suweis, Marco Formentin, Marco Favretti, Igor Volkov, Jayanth R Banavar, Sandro Azaele, and Amos Maritan. Upscaling species richness and abundances in tropical forests. *Science Advances*, 3(10):e1701438, 2017.
- [7] Krishna B Athreya, Peter E Ney, and PE Ney. *Branching processes*. Courier Corporation, 2004.
- [8] John Cardy. *Scaling and renormalization in statistical physics*, volume 5. Cambridge university press, 1996.
- [9] Tobias Jahnke and Wilhelm Huisinga. Solving the chemical master equation for monomolecular reaction systems analytically. *Journal of mathematical biology*, 54(1):1–26, 2007.
- [10] Andrew Mellor, Mauro Mobilia, and RKP Zia. Characterization of the nonequilibrium steady state of a heterogeneous nonlinear q-voter model with zealotry. *EPL (Europhysics Letters)*, 113(4):48001, 2016.
- [11] Grigorios A Pavliotis. *Stochastic processes and applications: diffusion processes, the Fokker-Planck and Langevin equations*, volume 60. Springer, 2014.
- [12] RKP Zia and B Schmittmann. Probability currents as principal characteristics in the statistical mechanics of non-equilibrium steady states. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(07):P07012, 2007.
- [13] Malte Henkel, Haye Hinrichsen, and Sven Lübeck. *Non-Equilibrium Phase Transitions: Volume 1: Absorbing Phase Transitions*. Springer Science & Business Media, 2008.
- [14] Jordi García-Ojalvo and José Sancho. *Noise in spatially extended systems*. Springer Science & Business Media, 2012.
- [15] Yahav Shem-Tov, Matan Danino, and Nadav M Shnerb. Solution of the spatial neutral model yields new bounds on the amazonian species richness. *Scientific reports*, 7:42415, 2017.
- [16] Pavel L Krapivsky, Sidney Redner, and Eli Ben-Naim. *A kinetic view of statistical physics*. Cambridge University Press, 2010.
- [17] Jacopo Grilli, Sandro Azaele, Jayanth R Banavar, and Amos Maritan. Absence of detailed balance in ecology. *EPL (Europhysics Letters)*, 100(3):38002, 2012.
- [18] James P O’Dwyer and Jessica L Green. Field theory for biogeography: a spatially explicit model for predicting patterns of biodiversity. *Ecology letters*, 13(1):87–95, 2010.
- [19] James P ODwyer and Stephen J Cornell. Cross-scale neutral ecology and the maintenance of biodiversity. *Scientific reports*, 8(1):10200, 2018.
- [20] Sandro Azaele, Samir Suweis, Jacopo Grilli, Igor Volkov, Jayanth R Banavar, and Amos Maritan. Statistical mechanics of ecological systems: Neutral theory and beyond. *Reviews of Modern Physics*, 88(3):035003, 2016.
- [21] N.N. Lebedev and R.A. Silverman. *Special Functions & Their Applications*. Dover Books on Mathematics. Dover Publications, 2012.
- [22] C.W. Gardiner. *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences*. Springer complexity. Springer, 2004.
- [23] N. G. Van Kampen. *Stochastic processes in physics and chemistry*, volume 1. Elsevier, 1992.
- [24] LR Taylor. Aggregation, variance and the mean. *Nature*, 189(4766):732, 1961.
- [25] LR Taylor and RAJ Taylor. Aggregation, migration and population mechanics. *Nature*, 265(5593):415, 1977.
- [26] Zoltán Eisler, Imre Bartos, and János Kertész. Fluctuation scaling in complex systems: Taylor’s law and beyond. *Advances in Physics*, 57(1):89–142, 2008.
- [27] Charlotte James, Sandro Azaele, Amos Maritan, and Filippo Simini. Zipf’s and taylor’s laws. *Physical Review E*, 98(3):032408, 2018.
- [28] Sandro Azaele, Simone Pigolotti, Jayanth R Banavar, and Amos Maritan. Dynamical evolution of ecosystems. *Nature*, 444(7121):926, 2006.
- [29] Daniel T Gillespie. Exact stochastic simulation of coupled chemical reactions. *The journal of physical chem-*



- istry*, 81(25):2340–2361, 1977.
- [30] Fabio Peruzzo and Sandro Azaele. A phenomenological spatial model for macro-ecological patterns in species-rich ecosystems. In *Stochastic Processes, Multiscale Modeling, and Numerical Methods for Computational Cellular Biology*, pages 349–368. Springer, 2017.
- [31] Ivan Dornic, Hugues Chaté, and Miguel A Munoz. Integration of langevin equations with multiplicative noise and the viability of field theories for absorbing phase transitions. *Physical review letters*, 94(10):100601, 2005.
- [32] Richard Condit, Nigel Pitman, Egbert G Leigh, Jérôme Chave, John Terborgh, Robin B Foster, Percy Núñez, Salomón Aguilar, Renato Valencia, Gorky Villa, et al. Beta-diversity in tropical forest trees. *Science*, 295(5555):666–669, 2002.
- [33] Jérôme Chave and Egbert G Leigh Jr. A spatially explicit neutral model of  $\beta$ -diversity in tropical forests. *Theoretical population biology*, 62(2):153–168, 2002.
- [34] Bernard D Coleman. On random placement and species-area relations. *Mathematical Biosciences*, 54(3-4):191–215, 1981.
- [35] Brian J McGill, Rampal S Etienne, John S Gray, David Alonso, Marti J Anderson, Habtamu Kassa Benecha, Maria Dornelas, Brian J Enquist, Jessica L Green, Fangliang He, et al. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology letters*, 10(10):995–1015, 2007.
- [36] Andrew J Black and Alan J McKane. Stochastic formulation of ecological models and their applications. *Trends in ecology & evolution*, 27(6):337–345, 2012.
- [37] Simone Pigolotti, Massimo Cencini, Daniel Molina, and Miguel A Muñoz. Stochastic spatial models in ecology: a statistical physics approach. *Journal of Statistical Physics*, 172(1):44–73, 2018.
- [38] Andrea Rinaldo, Amos Maritan, Kent K Cavender-Bares, and Sallie W Chisholm. Cross-scale ecological dynamics and microbial size spectra in marine ecosystems. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 269(1504):2051–2059, 2002.
- [39] Enrico Ser-Giacomi, Lucie Zinger, Shruti Malviya, Colomban De Vargas, Eric Karsenti, Chris Bowler, and Silvia De Monte. Ubiquitous abundance distribution of non-dominant plankton across the global ocean. *Nature ecology & evolution*, 2(8):1243, 2018.
- [40] Stephen Woodcock, Christopher J Van Der Gast, Thomas Bell, Mary Lunn, Thomas P Curtis, Ian M Head, and William T Sloan. Neutral assembly of bacterial communities. *FEMS microbiology ecology*, 62(2):171–180, 2007.
- [41] Andrea Giometto, Andrea Rinaldo, Francesco Carrara, and Florian Altermatt. Emerging predictable features of replicated biological invasion fronts. *Proceedings of the National Academy of Sciences*, 111(1):297–301, 2014.
- [42] Andrea Giometto, Marco Formentin, Andrea Rinaldo, Joel E Cohen, and Amos Maritan. Sample and population exponents of generalized taylor’s law. *Proceedings of the National Academy of Sciences*, 112(25):7755–7760, 2015.
- [43] China A Hanson, Jed A Fuhrman, M Claire Horner-Devine, and Jennifer BH Martiny. Beyond biogeographic patterns: processes shaping the microbial landscape. *Nature Reviews Microbiology*, 10(7):497, 2012.
- [44] Florian Altermatt, Emanuel A Fronhofer, Aurelie Garnier, Andrea Giometto, Frederik Hammes, Jan Klecka, Delphine Legrand, Elvira Mächler, Thomas M Massie, Frank Pennekamp, et al. Big answers from small worlds: a user’s guide for protist microcosms as a model system in ecology and evolution. *Methods in Ecology and Evolution*, 6(2):218–231, 2015.
- [45] Michael DJ Lynch and Josh D Neufeld. Ecology and exploration of the rare biosphere. *Nature Reviews Microbiology*, 13(4):217, 2015.
- [46] Sandro Azaele, Amos Maritan, Stephen J Cornell, Samir Suweis, Jayanth R Banavar, Doreen Gabriel, and William E Kunin. Towards a unified descriptive theory for spatial ecology: predicting biodiversity patterns across spatial scales. *Methods in Ecology and Evolution*, 6(3):324–332, 2015.
- [47] Iacopo Mastromatteo and Matteo Marsili. On the criticality of inferred models. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(10):P10012, 2011.
- [48] Alireza Goudarzi, Christof Teuscher, Natali Gulbahce, and Thimo Rohlf. Emergent criticality through adaptive information processing in boolean networks. *Physical review letters*, 108(12):128702, 2012.
- [49] Adam Z Stieg, Audrius V Avizienis, Henry O Sillin, Cristina Martin-Olmos, Masakazu Aono, and James K Gimzewski. Emergent criticality in complex turing b-type atomic switch networks. *Advanced Materials*, 24(2):286–293, 2012.
- [50] Jorge Hidalgo, Jacopo Grilli, Samir Suweis, Miguel A Muñoz, Jayanth R Banavar, and Amos Maritan. Information-based fitness and the emergence of criticality in living systems. *Proceedings of the National Academy of Sciences*, 111(28):10095–10100, 2014.
- [51] FS Gnesotto, Federica Mura, Jannes Gladrow, and CP Broedersz. Broken detailed balance and non-equilibrium dynamics in living systems: a review. *Reports on Progress in Physics*, 81(6):066601, 2018.
- [52] Christopher Battle, Chase P Broedersz, Nikta Fakhri, Veikko F Geyer, Jonathon Howard, Christoph F Schmidt, and Fred C MacKintosh. Broken detailed balance at mesoscopic scales in active biological systems. *Science*, 352(6285):604–607, 2016.
- [53] Milton Abramowitz and Irene A Stegun. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, volume 55. Courier Corporation, 1965.



**FIG. 8: Time comparison for different areas and different times in one dimension.** Simulations were carried out in  $d = 1$  with periodic boundary conditions, where each site initially contained exactly 100 individuals, with a total number of 200 lattice sites; parameters are the same as in Fig.(7). The three panels show comparisons of simulated data from the Doob-Gillespie algorithm and the analytic formula presented in eq.(37). The three panels show different segment lengths (20, 40 and 60 sites, respectively), and each presents three plots referring to different times  $T$  ( $T = 0.05$  for red histograms,  $T = 0.5$  for green histograms and  $T = 2$  for blue histograms, units of  $\mu^{-1}$ ). Histograms represent data from simulations, solid lines are the analytic predictions and error bars have length twice as much the standard deviation. Initially predictions and simulated data differ, as expected. However, as we approach stationarity regimes they improve significantly, and already at  $T = 0.5$  the analytical and simulated distribution match very well.



**FIG. 9: Comparison between Gillespie and Phenomenological simulation scheme.** The four plots show the conditional pdf,  $P(N|L)$ , simulated with the same set of parameters and on lattices of the same size, but using different algorithms at stationarity and in one dimension. Blue dots represent data from the phenomenological scheme (and blue lines are the respective error bars), while red dots are from Doob-Gillespie algorithm (with respective error bars). Black solid lines are the analytic predictions from eq.(35). The parameters are  $b = 600$ ,  $d = 601$ ,  $\gamma = 0.5$ ,  $b_0 = 5$  ( $D = 150$ ) and hence  $\lambda \approx 12$ . The lattices have 200 total sites and periodic boundary conditions have been used.