Quantifying evenly distributed states in exclusion and nonexclusion processes

Benjamin J. Binder*
School of Mathematical Sciences, University of Adelaide, South Australia 5005, Australia

Kerry A. Landman
Department of Mathematics and Statistics, University of Melbourne, Victoria 3010, Australia

(Received 23 November 2010; revised manuscript received 10 February 2011; published 20 April 2011)

Spatial-point data sets, generated from a wide range of physical systems and mathematical models, can be analyzed by counting the number of objects in equally sized bins. We find that the bin counts are related to the Pólya distribution. New measures are developed which indicate whether or not a spatial data set, generated from an exclusion process, is at its most evenly distributed state, the complete spatial randomness (CSR) state. To this end, we define an index in terms of the variance between the bin counts. Limiting values of the index are determined when objects have access to the entire domain and when there are subregions of the domain that are inaccessible to objects. Using three case studies (Lagrangian fluid particles in chaotic laminar flows, cellular automata agents in discrete models, and biological cells within colonies), we calculate the indexes and verify that our theoretical CSR limit accurately predicts the state of the system. These measures should prove useful in many biological applications.

DOI: 10.1103/PhysRevE.83.041914 PACS number(s): 87.17.−d, 87.18.−h, 87.85.−d

I. INTRODUCTION

Spatial-point data sets can be generated from a wide range of physical systems and mathematical models. Examples include the spatial positioning of Lagrangian fluid particles in chaotic laminar flows [1–8], cellular automata (CA) agents in discrete models [9–15], and cells within colonies in biological experiments [16–19]. The spatial distribution of a group of such objects can then be analyzed in a variety of ways. These include quadrant analysis of grid data sets, radial distribution functions, and correlation coefficients [20–22], or alternatively counting the number of objects in equally sized bins covering the spatial domain [3,6]. In this paper we focus on using bin counts: we derive new indexes to quantify spatial distributions for both exclusion and nonexclusion processes, based on the variance between the bin counts.

When each bin contains an equal (the average) number of objects, naturally the variance is zero. In this case, the state is evenly distributed. The most segregated state occurs when all objects reside in a single bin. In practical terms, an even distribution is an idealized state that is not often realized. Instead, the more realistic scenario occurs when each object is equally likely to lie in any bin. Such a state is termed *complete spatial randomness* (CSR) [3,6,21]. Once the CSR state is achieved the distribution of objects cannot be made any more even.

Phelps and Tucker [6] derived several indexes and their corresponding limiting values to ascertain when a distribution of tracked fluid particles in a chaotic flow reached the CSR state. This is an example of a nonexclusion process. The fluid particle positions are idealized points within the flow, they have no size, and there are no nearest-neighbor (exclusion) interactions. In contrast, for a simple exclusion process [23], an object's attempted movement within the domain is restricted and depends on whether the target space is already occupied or not (as for CA agents [10,12]). We are interested in generalizing the Phelps and Tucker indexes and their limiting values [6] to exclusion processes. This will allow us to determine when a distribution of objects generated from an exclusion process has attained the CSR state. The key feature in our derivation for an exclusion process is a Pólya-Eggenberger (Pólya [24]) probability mass function (PMF) of the bin counts. This differs from a nonexclusion process where there is a binomial PMF of the bin counts [3,6,21].

We derive the index and CSR limit for a nonexclusion and an exclusion process. Two simple discrete models are presented to illustrate the differences between these two types of processes. Using three case studies, we calculate the indexes and verify that our theoretical CSR limit accurately predicts the state of the system.

II. THEORY

We discuss two-dimensional data sets with two-dimensional bins. This work generalizes in a natural way to three dimensions.

A. Variance

Consider a $X \times Y$ rectangular domain which is populated with a total of n objects each of area s using either a nonexclusion or exclusion process. Following Phelps and Tucker [6], the domain is divided into M equal-sized bins each of area S. If b_j is the number of objects in bin j, $(j = 1, \ldots, M)$, then the average bin count is n/M and the variance is

$$\sigma^2 = \frac{1}{M} \sum_{j=1}^{M} \left(b_j - \frac{n}{M} \right)^2. \tag{1}$$

There are two limiting cases.

*benjamin.binder@adelaide.edu.au

(1) The state is an even distribution where each bin contains the average number of objects. Then $\sigma^2 = 0$.

(2) The state is completely segregated where all objects reside in a single bin. Without any loss of generality, we can write this as

$$b_j = \begin{cases} n, & j = 1, \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

Then Eq. (1) becomes

$$\sigma_0^2 = n^2 \left(\frac{M-1}{M^2} \right). \tag{3}$$

It is convenient to define a scaled variance, called an index [6], given by

$$I = \frac{\sigma^2}{\sigma_0^2}, \tag{4}$$

which ranges from unity (maximum segregation) to zero (even distribution). Note that for an exclusion process, a completely segregated state is only possible if n is sufficiently small since it is limited by the size of the region. Therefore, if n is sufficiently high, in practice, it is not possible to have a completely segregated state in an exclusion process. However, σ_0^2 can still be used as a scaling in the definition of the index.

B. CSR limits

Randomly distributed objects do not necessarily result in an even distribution, but rather the CSR state, where each object is equally likely to lie in any bin. We derive the CSR limits for both nonexclusion and exclusion processes, thereby extending the work of Phelps and Tucker [6].

We find that the bin counts are related to the Pólya distribution. If n objects are randomly distributed throughout the domain, and K represents the bin count b_j in any one of the M bins, then the distribution of K is the Pólya distribution with parameters n, S, $XY - S$, and $-s$, namely

$$\Pr(K = k,n) = \frac{\binom{n}{k} \prod_{i=0}^{k-1} (P + i\alpha) \prod_{i=0}^{n-k-1} (Q + i\alpha)}{\prod_{i=0}^{n-1} (1 + i\alpha)}, \tag{5}$$

where

$$P = \frac{S}{XY} = \frac{1}{M}, \quad Q = 1 - \frac{1}{M}, \quad \alpha = -\frac{s}{XY}. \tag{6}$$

The variance is

$$\sigma^2 = \frac{n}{M} \left(1 - \frac{1}{M} \right) \frac{1 + n\alpha}{1 + \alpha}. \tag{7}$$

For a nonexclusion process the objects have no size or area, so $s = 0$. The result is a binomial PMF of the bin counts [3,6,21], with variance

$$\sigma_{\text{csr}}^2 = \frac{n}{M} \left(1 - \frac{1}{M} \right). \tag{8}$$

Using Eqs. (3)–(8), the CSR limiting value is

$$I_{\text{csr}} = \frac{1}{n}, \tag{9}$$

for a nonexclusion process.

It is worthwhile mentioning that the limiting value in Eq. (9) is the same as the one reported in Phelps and Tucker [6], but we have derived ours in a slightly different way. The Phelps and Tucker derivation assumes that both $M \gg 1$ and $n \gg 1$. Under these assumptions the binomial PMF asymptotes to a Poisson PMF [25] with variance $\sigma_{\text{csr}}^2 = n/M$. Combining this with their approximation of $\sigma_0^2 = n^2/M$ (for $M \gg 1$) gives the same I_{csr} value as ours. The subtle difference in the way we choose to calculate the index in Eq. (4), using Eq. (3), is important for moderate values of M and n, even though the CSR limiting value is the same when both M and n are large.

For an exclusion process $s > 0$, with variance

$$\sigma_{\text{xcsr}}^2 = \frac{\frac{n}{M}\left(1 - \frac{1}{M}\right)\left(1 - \frac{ns}{XY}\right)}{1 - \frac{s}{XY}} \approx \frac{n}{M} \left(1 - \frac{1}{M} \right) \left(1 - \frac{ns}{XY} \right), \quad (XY \gg s). \tag{10}$$

Note, when $s = 1$ the result is a hypergeometric PMF of the bin counts. The CSR limiting value is

$$I_{\text{xcsr}} = \frac{1 - \frac{ns}{XY}}{n\left(1 - \frac{s}{XY}\right)} \approx \frac{1}{n} \left(1 - \frac{ns}{XY} \right), \quad (XY \gg s), \tag{11}$$

where the subscript x denotes an exclusion process. Here, it is convenient to define the average density or occupied area fraction of the domain to be

$$d = \frac{ns}{XY}. \tag{12}$$

This parameter can also be viewed as an average probability of occupancy of each bin. Equation (11) then becomes

$$I_{\text{xcsr}} = \frac{1 - d}{n}. \tag{13}$$

C. Example

We consider two simple discrete models to help illustrate the differences between the nonexclusion ($s = 0$) and exclusion ($s = 1$) process CSR limits as the number of objects n increases, for two values of bin number M. We will plot the index I for different values of d. It should be noted that for a nonexclusion process d cannot be interpreted as a density; instead it is a rescaling of the number of objects within area XY and, as such, there is no limiting value of d. However, for an exclusion process, it is a scaled density or average occupancy, and so $0 \leqslant d \leqslant 1$.

We use a two-dimensional rectangular lattice with unit spacing. Each site is indexed (i, j) where $i, j \in \mathbb{Z}$. For both models we randomly select a lattice site within the domain that may already be populated with objects. If the selected lattice site is unoccupied, place the object on the site. If the selected lattice is already occupied, then (i) for the nonexclusion model, an object is added to the total number of objects already residing on the lattice site and (ii) for the exclusion model, the selected lattice site is rejected, and another one is randomly chosen. These processes are repeated (for both models) until the total number of objects in the lattice reaches a predetermined value n.

Figure 1(a) illustrates a single realization of the exclusion model, where occupied sites are blue and empty sites are yellow; a single realization of a nonexclusion process is not

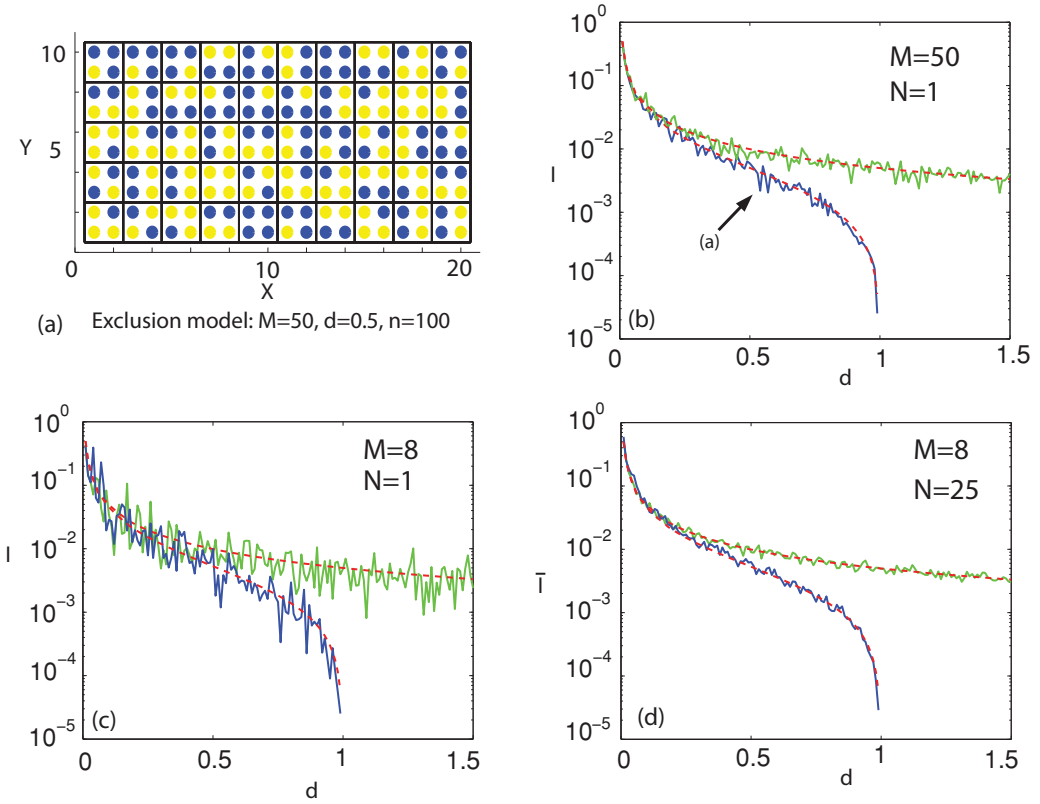

FIG. 1. (Color online) Nonexclusion and exclusion models on a 20×10 lattice ($X = 20$, $Y = 10$) with M bins. (a) Single realization of the exclusion model with $n = 100$, $M = 50$, and $d = 0.5$, showing occupied sites [blue (dark gray)] and empty sites [yellow (light gray)]. (b) Calculated values of the index I (from single realizations) for a nonexclusion [green (light gray)] and exclusion [blue (dark gray)] process versus d. The dashed red curves are the limiting CSR values $I_{\rm csr}$ [Eq. (9)] (upper) and $I_{\rm xcsr}$ [Eq. (11)] (lower) with $M = 50$. The arrow in (b) on the blue curve corresponds to the realization in (a). (c) As in (b) with $M = 8$. (d) Average index $\bar{I}$ (from $N = 25$ realizations) versus d for $M = 8$. The dashed red curves are $I_{\rm csr}$ (upper) and $I_{\rm xcsr}$ (lower).

shown. Values of the index I are calculated using Eq. (4) for a single realization of the nonexclusion and exclusion models as a function of d, shown in Fig. 1(b). Even for a single realization, the I curves compare well with the corresponding limiting CSR values $I_{\rm csr}$ and $I_{\rm xcsr}$, respectively. This is no surprise, as by construction (randomly choosing lattice sites to be populated with objects) we expect the distribution of objects to be at the CSR state. However, the simple discrete models provide a check that our analysis is correct.

When the density of the lattice is low, nearest-neighbor interactions in exclusion processes are less significant, so that we expect $I_{\rm xcsr}$ to be close to $I_{\rm csr}$ [Figs. 1(b) and 1(c)]. However, the two CSR limiting values deviate sharply as d increases. This can be explained by considering their values when $n = XY$. In this case, the nonexclusion process limiting value is $I_{\rm csr} = 1/XY$, but for the exclusion process, the domain is fully occupied (or at carrying capacity) with $d = 1$ and $I_{\rm xcsr} = 0$ – this is the evenly distributed state. Note that for a nonexclusion process the evenly distributed state is approached more slowly, only as $n \to \infty$ does $I_{\rm csr} \to 0$.

For a single realization, it would be expected that reducing the number of bins M (or number of samples) in calculating the index I would increase the fluctuations about the CSR limits; this is demonstrated in Figs. 1(b) and 1(c). To decrease the fluctuations, we sum over N realizations and define the average index as

$$\bar{I} = \frac{1}{N}\sum_{i=1}^{N} I_i, \tag{14}$$

where I_i is the ith realization of the index for a specified value of d. As expected, comparison between a single realization [Fig. 1(c)] and the average over 25 realizations [Fig. 1(d)] reduces the fluctuations about the CSR limits. The average index compares very well with both CSR limits. Similar results hold for other lattice dimensions with various values of M, when the values of MN are the same.

Although the number of bins M does not appear in either formulas of the CSR limits (9) and (13), some care is needed when interpreting results which may depend on M. For example, consider a situation in which the number of bins M equals the size of the domain XY. For an exclusion process each bin count b_j either has only one or no objects within it. Equation (1) can then be rewritten as

$$\sigma^2 = \frac{1}{XY}\left(\sum_{j=1}^{n}\left(1-\frac{n}{XY}\right)^2 + \sum_{j=1}^{XY-n}\left(\frac{n}{XY}\right)^2\right) = \frac{n}{XY}\left(1-\frac{n}{XY}\right). \tag{15}$$

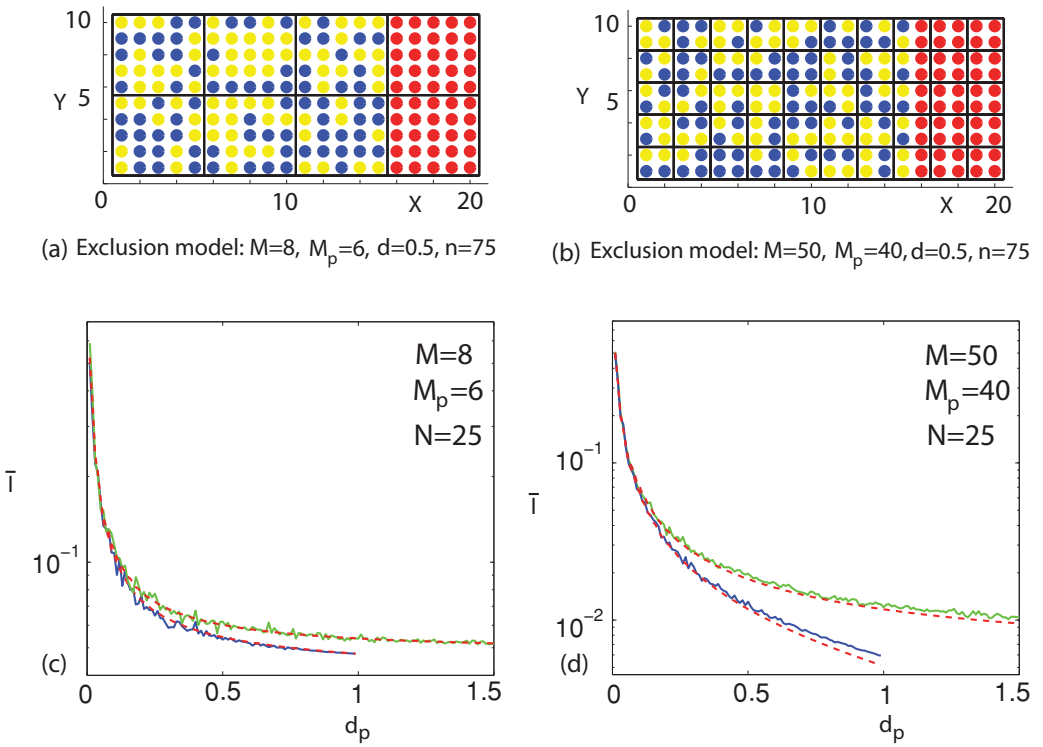

FIG. 2. (Color online) Nonexclusion and exclusion models on a 20×10 lattice ($X = 20$, $Y = 10$) with M bins. (a) Single realization of the exclusion model with $n = 75$, $M = 8$, $M_p = 6$, and $d = 0.5$, showing occupied sites [blue (dark gray)] and empty sites [yellow (light gray)]. The red (midgray on right) sites are inaccessible to (and therefore unoccupied by) blue (dark gray) objects. (b) As in (a) with $M = 50$ and $M_p = 40$. (c) Average index $\bar{I}$ (from $N = 25$ realizations) for both nonexclusion [green (light gray)] and exclusion [blue (dark gray)] processes versus d_p, with $M = 8$ and $M_p = 6$. The dashed red curves are I_{isl} (upper) and I_{xisl} (lower). (d) As in (c) with $M = 50$ and $M_p = 40$.

The corresponding value of the index (4) is

$$I = \frac{1-d}{n}, \qquad (XY \gg 1), \tag{16}$$

which is the same as the CSR limiting value (13). According to the calculation the distribution of objects is at the CSR state, regardless of how they are spatially organized within the domain, which is clearly not always the case. Typically we find the index (4) and average index (14) to be a reliable indicator of when the CSR state is reached if $n \gg M$. When this is not the case we check that our calculations are independent of M.

In the previous example, the whole region $X \times Y$ is accessible to distributed objects. Now consider the case when there is a subregion of the whole region which is inaccessible to the objects. This can occur, for example, when tracked fluid particles are distributed in a chaotic region of a laminar fluid flow possessing periodic islands in which no fluid particles are being tracked [6]. Hence there will be a number of occupied or populated bins, denoted M_p, and a number of unoccupied or unpopulated bins $M - M_p$. The variance of object counts among the occupied bins is

$$\sigma_p^2 = \frac{1}{M_p} \sum_{j=1}^{M_p} \left(b_j - \frac{n}{M_p} \right)^2. \tag{17}$$

It is simple to confirm that the variance over the M bins [Eq. (1)] can be expressed in terms of σ_p^2 as

$$\sigma^2 = \frac{n^2}{M} \left(\frac{1}{M_p} - \frac{1}{M} \right) + \frac{M_p}{M} \sigma_p^2. \tag{18}$$

Assuming the populated bins are at the CSR state, new limiting indexes can be obtained with a little algebra. For the nonexclusion process replace σ_p^2 with $\frac{n}{M_p}(1 - \frac{1}{M_p})$ from Eq. (8), then Eq. (18) together with Eqs. (3) through (4) defines a limiting index for a nonexclusion process with *island* regions [6], as

$$I_{\text{isl}} = \frac{M}{M-1} \left(\frac{1}{M_p} - \frac{1}{M} + \frac{1}{n} \frac{M_p - 1}{M_p} \right) \approx \left(\frac{1}{M_p} - \frac{1}{M} + \frac{1}{n} \right), \quad \text{when} \quad M_p \gg 1. \tag{19}$$

Similarly, for the exclusion process we replace σ_p^2 with $\frac{n}{M_p}(1 - \frac{1}{M_p})(1 - d_p)$, where $d_p = dM/M_p$ is the occupancy within the fraction of bins occupied ($0 \leqslant d_p \leqslant 1$). Using Eq. (10), a limiting index for an exclusion process with *island* regions is

$$I_{\text{xisl}} = \frac{M}{M-1} \left(\frac{1}{M_p} - \frac{1}{M} + \frac{1-d_p}{n} \frac{M_p - 1}{M_p} \right) \approx \left(\frac{1}{M_p} - \frac{1}{M} + \frac{1-d_p}{n} \right), \quad \text{when} \quad M_p \gg 1. \tag{20}$$

Since $(1/M_p - 1/M) > 0$, these limiting values are larger than their corresponding limiting CSR values [Eqs. (9) and (11), respectively].

Figure 2(a) is a realization of the exclusion model, where the part of the domain shown in red is inaccessible to the placement of objects. With the illustrated setup of bins and restriction on the domain, the average index $\bar{I}$ is determined for different object number n for both a nonexclusion and exclusion process. Figure 2(c) shows that the computed values of $\bar{I}$ as a function of d_p compares well with the limiting values I_{isl} and I_{xisl}, respectively. Similar to the two CSR limits, we see that I_{isl} and I_{xisl} are close for small d_p values. However, as d_p increases, the deviation between the two limits is much smaller than for the CSR limits. This occurs because I_{isl} and I_{xisl} are primarily controlled by the (same) area fraction of the domain which is unoccupied, rather than when $d_p = 1$ or $n \to \infty$ for the exclusion and nonexclusion processes, respectively.

Although this example enabled us to check our analysis, it is a rather artificial situation. It is far more likely, in practice, that the inaccessible part of the domain will overlap a number of bins in the calculation, as illustrated with the simple configuration in Fig. 2(b). The partially populated bins are now no longer at the CSR state. Hence, the limiting values Eqs. (19) and (20) become lower bounds for a calculation of the index, as demonstrated in Fig. 2(d).

III. CASE STUDIES

We consider three case studies drawn from different areas: fluid mechanics, agent-based models and cell colonies.

A. Nonexclusion process: Lagragian fluid particles

Laminar fluid flows occur in fluid mixing applications, for example, in a batch mixing process [5,7,8], where a vat of fluid is stirred by a number of rods. The stirring motion of the rods is usually executed periodically; it is called a stirring protocol. The mixing is known to be particularly effective provided that the Lagrangian particle paths are chaotic [1]. Nevertheless, even in a chaotic flow, there may be small [6] or large periodic islands [7,8], prohibiting an even spatial distribution of a passive tracer (tracked Lagrangian fluid particles) within the entire fluid flow domain.

Often the goal is to assess the quality of the fluid mixing for a given device, such as a stirring protocol for a batch mixer. A variety of mixing measures can be used [5]. Instead of these, we choose to use the index I and the limiting values I_{csr} and I_{isl}. This extends the work of Phelps and Tucker [6], where fluid particles were tracked in a time-periodic sine flow [26], on a two-dimensional rectangular domain, where the velocity field can be integrated analytically.

Here we consider a more complicated situation, where the geometry of the flow domain changes with time as the stirring rods move through a very viscous fluid in a circular vat (or domain). A slow viscous flow (Stokes flow) is simulated using the solution of Finn *et al.* [4] to determine the velocity field numerically. We choose the origin at the center of the circular vat and scale the lengths so that the vat has unit radius. Furthermore, we assume that the fluid particles, in addition to being advected with the flow, also diffuse, with diffusivity D.

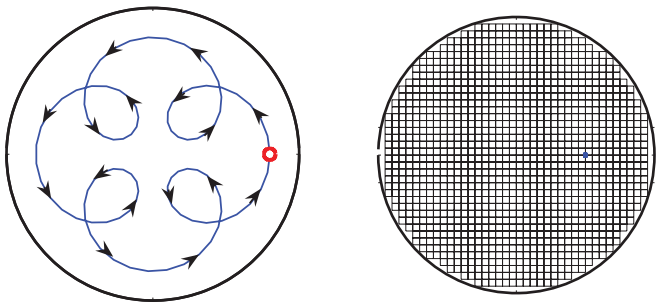

FIG. 3. (Color online) Stirring protocol and initial condition for fluid particle simulations in a vat of (scaled) unit radius. (a) Time-periodic stirring protocol, the stirring rod (red, small open circle) has radius $r = 0.02$. (b) The vat is partially covered with a total of $M = 1168$ square bins, each with equal area $A = 0.05^2$. A total of $n_t = 10^4$ particles are chosen to be tracked, initially evenly distributed among $M_p = 4$ bins, indicated by a blue square.

Following Aref *et al.* [2] and Finn *et al.* [5], we replace the usual advection equations with the stochastic differential equations

$$dx = u(x,y,t)dt + dN_x, \tag{21}$$

$$dy = v(x,y,t)dt + dN_y, \tag{22}$$

where (u,v) is the fluid velocity, while dN_x and dN_y are independent, identically distributed, Gaussian random variables with mean zero and variance $2Ddt$.

We examine the stirring protocol illustrated in Fig. 3(a), where a single stirring rod (red) moves through the fluid along a prescribed path. The time-periodic stirring protocol is normalized to have unit period. This stirring protocol is known to have visible periodic islands within the flow when there is no diffusion [8], making it an excellent candidate for us to analyze here.

Some further analysis is required to overcome the problem of dealing with a circular domain (the vat) in the calculation of the index and the CSR limit. Therefore, we partially grid most of the vat into M square bins each with equal area A, shown in Fig. 3(b). The gridded region (with area MA) used in the calculations is now a patchwork of square bins, rather than a rectangular domain. Let n_t denote the total number of fluid particles tracked in the vat. Of course, they may not all lie within the gridded region, and furthermore, the number within the gridded region, denoted by n, may change with time, affecting the calculation of the index in Eq. (4).

This also means that the CSR limit in Eq. (9) varies since it also depends on n. Of course, this is not an issue if we are content with calculating the CSR limit for each calculation of the index. However, we choose not to proceed this way and instead make an approximation to the CSR limit. Assuming that the entire fluid flow domain is at the CSR state, the number particles within the gridded region can be approximated by a constant, determined by the area fraction which is gridded, namely, $n \approx (MAn_t)/\pi$. Hence, the CSR limit in Eq. (9) becomes

$$I_{\text{csr}} = \frac{\pi}{MAn_t}. \tag{23}$$

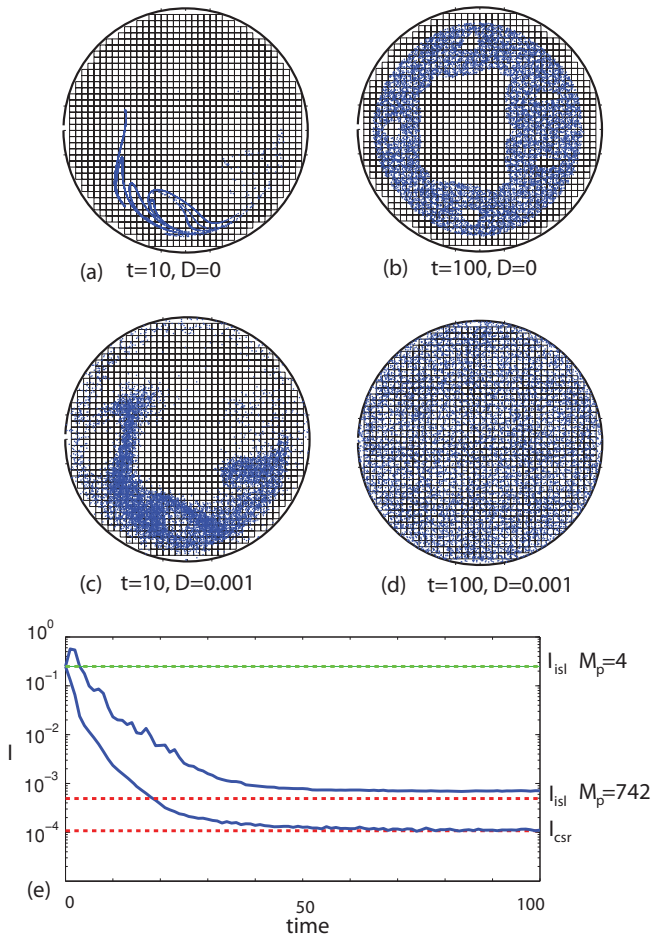

FIG. 4. (Color online) Particle simulations and evolution of the index I. (a–d) Simulations of the initial condition shown in Fig. 3(b). (e) Values of the index plotted as a function of time are given by the blue (dark gray) curves: $D = 0$ (upper) and $D = 0.001$ (lower). The dashed lines are limiting values of I_{isl} [Eq. (24)] for $M_p = 4$ [green, top (light gray) line] and $M_p = 742$ (red, middle line). The lowest dashed red line is I_{csr} from Eq. (23).

Typically in our calculations $n_t \geqslant n \gg M > M_p \gg 1$, and so the limiting value in Eq. (19) just simplifies to

$$I_{\text{isl}} = \frac{1}{M_p} - \frac{1}{M}. \tag{24}$$

The initial condition for our simulations has 10^4 fluid particles evenly distributed among four bins as in Fig. 3(b). This makes $n_t = 10^4$ and $M_p = 4$. The initial position of the particles is in a region of the fluid flow which is known to be chaotic (see the iterated mapping plots in Binder [8]). Using Eqs. (21) and (22), we track the particles as they move within the fluid for two values of the diffusivity, $D = 0$ and $D = 0.001$. We illustrate the resulting fluid particle positions at two times in Fig. 4(a–d).

When there is no diffusion, the particles eventually (after about 50 periods) appear randomly distributed throughout a chaotic region of the fluid flow, and the number of populated bins levels to a constant value of 742, giving $M_p = 742$. There are six visible periodic islands and an outer annular region in which there are no particles [Fig. 4(b)]. The upper blue curve in Fig. 4(e) shows how the index decays with time. It asympototes to a value which is slightly greater than the limiting value, I_{isl}, obtained from Eq. (24) with $M_p = 742$. Interestingly, for early time, the index increases above the other limiting value of $I_{\text{isl}} \approx 1/4$ [Eq. (24)] with $M_p = 4$, illustrated by the dashed green curve in Fig. 4(e). This means, at early times, the number of occupied bins will decrease from four so that the mixing becomes more segregated.

In contrast, when there is diffusion, the particle positions (after about 70 periods) appear to be randomly distributed throughout the entire domain, as seen in Fig. 4(d). The time evolution of the index (lower blue curve) tends to the CSR limit given by Eq. (23). This indicates when the fluid mixture has reached an homogenized state.

B. Exclusion process: Cellular automata agents

Discrete CA models can assist in the understanding of underlying mechanisms in biological systems; for example, neural crest cell (NCC) invasion in the developing enteric nervous system [9,11,12], and designing tissue engineered scaffolds [15,27]. The agents follow simple rules governed by probabilities. The resulting agent distributions are stochastic and illustrative of real biological observations and experiments. This means that information regarding the distribution of CA agents is a crucial part of the analysis.

In some circumstances, the average occupancy of the CA model can be intimately linked to a continuum model [9–11,28], providing an understanding of the variety of scales of interest in a biological process. A continuum model's predictions can also be compared to the average index $\bar{I}$ for a simple exclusion process. Therefore, we choose to investigate the motility mechanism in the CA model for the NCC invasion process [9,10], which, on average, satisfies the linear one-dimensional diffusion equation.

Consider a domain ($X \times Y$ rectangular lattice of integers) which contains n agents at any time t. During a single time step of CA model from t to $t + 1$, the agents are selected randomly and given the opportunity to move. An agent at (x,y) that is chosen to be motile attempts to move to one of the four nearest neighbours $(x \pm 1, y)$, $(x, y \pm 1)$ each with probability $1/4$. If the target site is occupied for any attempted move, it is aborted, as required in an exclusion process. With the initial condition given in Fig. 5(a), Figures 5(b) and 5(c) illustrate a single realization at two times ($t = 10$ and $t = 150$).

To compare the CA model's results with the continuum model, we first define $c_i(x,t)$ to be the total number of agents in column x of the lattice after t steps of the ith realization. Then for N realizations with the same initial condition, the average CA agent concentration (or average agent column density) is

$$C(x,t) = \frac{1}{NY} \sum_{i=1}^{N} c_i(x,t). \tag{25}$$

The corresponding continuum model [9,10] is the linear one-dimensional diffusion equation

$$\frac{\partial C(x,t)}{\partial t} = D \frac{\partial^2 C(x,t)}{\partial x^2}, \tag{26}$$

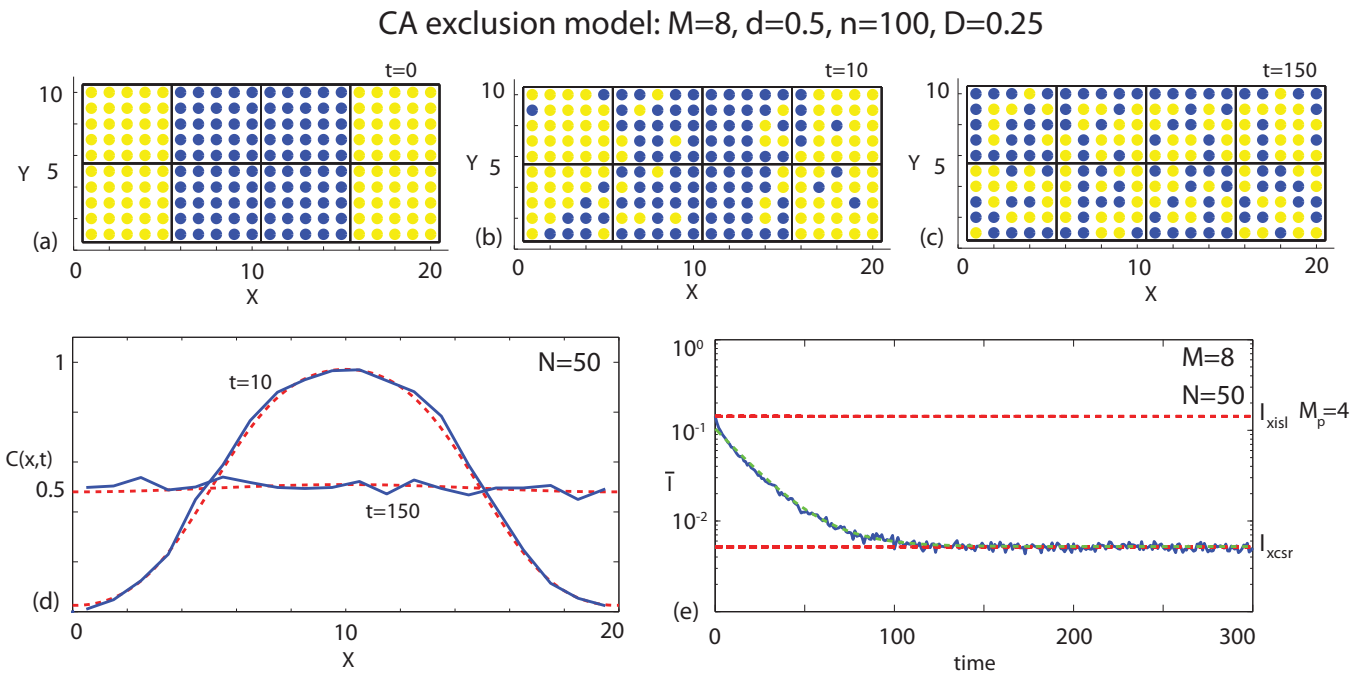

FIG. 5. (Color online) CA exclusion model results. (a) The initial condition with $n = 100, M = 8, d = 0.5, M_p = 4, d_p = 1$. (b–c) Typical realizations of the CA model at time $t = 10$ and $t = 150$. (d) Plots of the (average) CA concentration [blue (dark gray) line] compared with continuum solution to Eq. (26) with $D = 1/4$ (dashed red line). (e) Evolution of the average index $\bar{I}$ [blue (dark gray) line] with time over 50 realizations. The upper dashed red line is the limit I_{xisl} (at $t = 0$ with $M_p = 4$). The lower dashed red line is I_{xcsr} in Eq. (10). The middle dashed green curve is the time evolution of the index computed with the asymptotic form given by Eq. (27).

with diffusivity $D = 1/4$ (when the agent movement probability is unity, as chosen here). With the initial condition given in Fig. 5(a), a comparison between the average CA concentration and the solution of Eq. (26) is illustrated in Fig. 5(d). We find that the CA concentration $C(x,t) \approx 0.5 = d$ for all values of x when $t > 150$. Therefore, we expect the distribution of agents to be at the CSR state at this time. Furthermore, standard separation of variable techniques can be used to solve Eq. (26). We expect the long-time solution to be asymptotic to the first term of this series expansion. The concentration can be converted into a number b_j in each bin, by multiplying by bin height and integrating the concentration over the width of each bin. For the initial data and geometry in Fig. 5(a), it is simple to verify that

$$I(t) \sim I_{\text{xcsr}} + \frac{1}{\sigma_0^2}\left(L_x \frac{A_k}{\omega_k} e^{-\omega_k{}^2 Dt}\right)^2,$$
$$A_k = \frac{2}{k\pi}\left(\sin\frac{3k\pi}{4} - \sin\frac{k\pi}{4}\right), \quad \omega_k = \frac{k\pi}{X}. \qquad (27)$$

Here $X = 20$, $L_x = 5$ is the height of each bin, $D = 1/4$ and $k = 2$ corresponds to the leading term in the series expansion for $C(x,t)$. The first term in Eq. (27) accounts for random fluctuations in the discrete approach versus the continuum approach.

Figure 5(e) shows the evolution with time (blue curve) of the average index $\bar{I}$. The upper dashed curve is the limiting value I_{xisl} from Eq. (20) with $M_p = 4$, $d_p = 1$, similar to what we have seen in the fluid mixing study. The lower dashed red curve is the limiting value I_{xcsr} from Eq. (11), which indicates the distribution of agents is at the CSR state when $t > 150$. The dashed green curve is the time evolution of the index computed with the asymptotic form of the continuum solution to Eq. (26), given by Eq. (27). Just the first term (with $k = 2$) in the series expansion gives an excellent fit to the evolving index, again verifying that the continuum model provides the appropriate approximation to the average behavior of randomly moving objects in an exclusion process.

Although only the motility mechanism has been investigated here, we have demonstrated that the index and limiting values can, in principle, be used to assess the spatial distribution of agents in more sophisticated CA models.

C. Exclusion process: Colony of Escherichia coli cells

The spatial distribution and organization of many cell colonies have been analyzed [16–19]. Here we calculate the index for a colony of *Escherichia coli* (*E. coli*) cells growing in

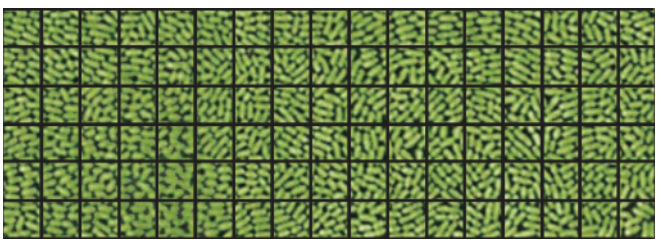

FIG. 6. (Color online) Typical image of a colony of *E. coli* cells [reprinted from Volfson *et al.* [16] Fig. 1(c); with permission from NAS]. On top of this image, a grid of equal sized bins has been placed, here $M = 102$ and $n = 1020$.

a microfluidic chemostat [16], such as the one shown in Fig. 6, where equal-sized bins have been placed over the image. We counted the number of cells in each bin. When an individual cell overlapped with a number of bins, it was included in the bin count which contained the majority of the cell. Notice also that individual cells vary in size. We calculate the index I for the cells in Fig. 6. Assuming that the distribution of cells is at the exclusion CSR state, this provides us with an estimate of the area fraction or density of the cells since $d = 1 - nI_{\text{xcsr}} = 0.85$, from Eq. (13). This compares well with value of the area fraction $d \approx 0.84$ obtained by Volfson *et al.* ([16], Fig. 1G).

This demonstrates that our index I can be used in a simple straightforward way to approximate the density of a population of cells, provided it is at the CSR state. Of course, we recognize that our formulations do not explicitly account for the variation in cell size or cells overlapping with a number of bins. Heuristically though, these effects cancel each other out, on average, as shown next.

First we investigate a case when an object may overlap with a number of bins. We consider the following exclusion model which corresponds to $s = 3$, in contrast with our previous exclusion models with $s = 1$. We randomly select a lattice site (x, y) within the domain. If the selected lattice site [Fig. 7(a), blue markers] and two neighboring sites $(x \pm 1, y)$ [Fig. 7(a), green markers] are unoccupied they are populated with the three markers (one blue and two green markers) denoting the object. If any of the selected sites is already occupied, then the process is rejected. This process is repeated until the total number of objects in the lattice reaches a predetermined value n. A typical simulation is shown in Fig. 7(a) with $n = 24$. Values of the average index $\bar{I}$ (blue line) are calculated as a function of d, shown in Fig. 7(b). Any given object is included in the bin count in which its blue marker lies. The results compare well with the I_{xcsr} limiting value. Notice that for this particular exclusion process (with $s = 3$), we are only able to simulate results when $d < 0.72$.

Finally, we consider an exclusion process where there are multiple species of different size objects. By way of example, consider objects with $s = 3$ and others with $s = 1$ [Fig. 7(c)]. Each subspecies can be considered individually, and their computed index matches their I_{xcsr} limiting values [Fig. 7(e)]. Treating all the objects together [Fig. 7(d)], we find that an average $\bar{d} = d_1 + d_2$ gives the correct I_{xcsr} limiting value [Fig. 7(e)]. In general, K species can be considered, and the total density is just $\bar{d} = \sum_{i=1}^{K} d_i = \sum_{i=1}^{K} n_i s_i/(XY) = n\bar{s}/(XY)$, where n_i is the number of objects with area s_i, $d_i = n_i s_i/(XY)$ and $\bar{s}$ is the area of an average object.

IV. DISCUSSION

We have developed new measures which indicate whether or not a spatial data set, generated from an exclusion process,

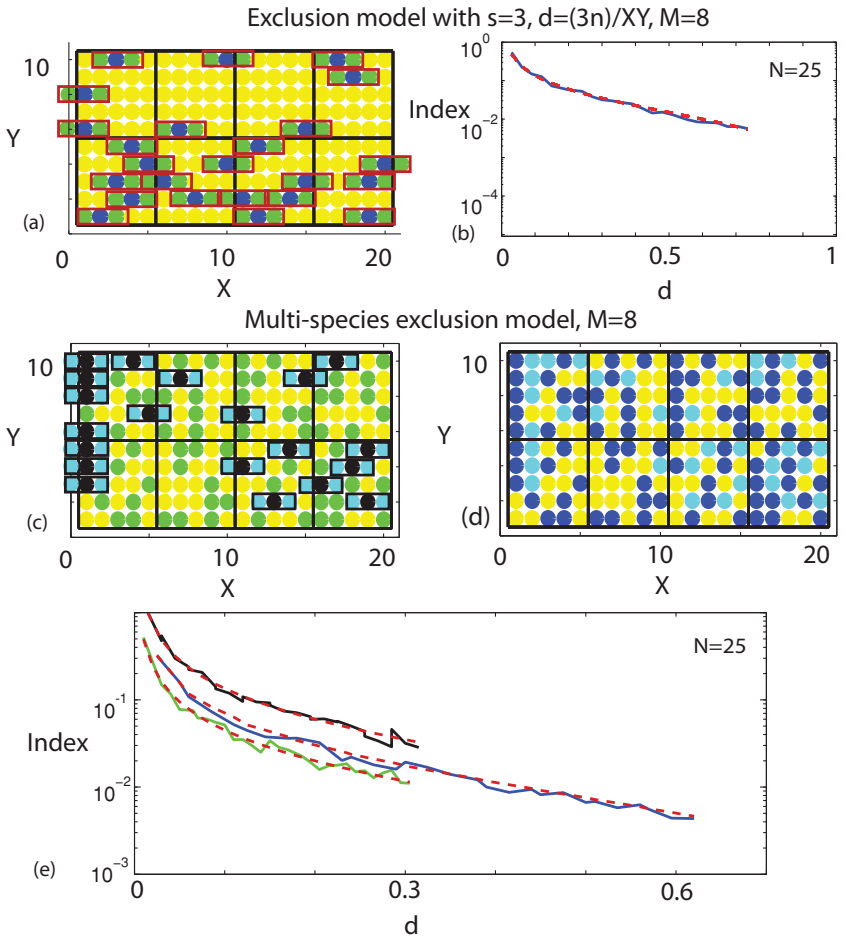

FIG. 7. (Color online) Exclusion models with different sized objects. (a) Typical realization with $d = 0.36$ and $M = 8$. (b) Calculated values of the average index [blue (dark gray)line], $N = 25$. The dashed red curves are the limiting values of I_{xcsr}. (c) Typical realization with multiple species with $d_1(s = 3) = 0.3$ (black markers) $d_2(s = 1) = 0.3$ [green (light gray) markers], where $n_2 = 3n_1$. (d) The same realization as in (c), but all markers now blue (dark gray and lighter gray) with $\bar{d} = d_1 + d_2 = 0.6$. (e) Calculated values of the average index for black objects ($s = 3$) (black, top curve), green (light gray, lowest curve) objects ($s = 1$), and totality of objects ($\bar{d} = 0.6, \bar{s} = 1.5$) [blue (dark gray), middle curve], $N = 25$. The dashed red curves are the limiting values of I_{xcsr}. Here, where $n_2 = 3n_1$.

is at its most evenly distributed state, the complete spatial randomness (CSR) state. This generalizes the work of Phelps and Tucker [6] for chaotic laminar flows, which is a nonexclusion process. The fundamental difference between the distribution of objects from a nonexclusion and an exclusion process is characterized by their limiting CSR values. There are fewer ways in which objects can organize themselves (spatially) in an exclusion process than for a nonexclusion process, which leads to the corresponding CSR limit being equal to the evenly distributed state when the density $d = 1$. Nonexclusion processes only approach the evenly distributed state when the number of objects $n \to \infty$ since there is no (exclusion) spatial restriction imposed on them.

Using three case studies, we have demonstrated the ease in which our measures can be implemented. For the nonexclusion process, we considered a chaotic laminar flow, as did Phelps and Tucker [6]. We have extended their work from a square to a circular domain, either with or without molecular diffusion, and used the index measures and limits to determine the time evolution of the mixing state. For the exclusion processes, the demonstrated versatility of the index in determining the spatial state opens up many new areas. For example, in future studies of the CA model, we will extend the derivations of the indexes by incorporating additional mechanisms such as domain growth and agent proliferation [28]. In addition, refinements to the formulations accounting for cell-size variation and cells that occupy more than one bin (e.g., *E. coli* cell study) is an interesting topic for future research.

ACKNOWLEDGMENTS

This work was supported by the Australian Research Council Discovery Grant (Kerry Landman). We also thank Chris Baker, Kelli Francis-Staite, and Joshua Ross for their assistance.

[1] H. Aref, J. Fluid Mech. **143**, 1 (1984).
[2] H. Aref and S. W. Jones, Phys. Fluids A **1**, 470 (1989).
[3] S. W. Jones, Phys. Fluids A **3**, 1081 (1991).
[4] M. D. Finn, S. M. Cox, and H. M. Byrne, J. Fluid Mech. **493**, 345 (2003).
[5] M. D. Finn, S. M. Cox, and H. M. Byrne, J. Eng. Math. **48**, 129 (2004).
[6] J. H. Phelps and C. L. Tucker, Chem. Engng Sci. **61**, 6826 (2006).
[7] B. J. Binder and S. M. Cox, Fluid Dynam. Res. **40**, 34 (2008).
[8] B. J. Binder, Phys. Lett. A **374**, 3483 (2010).
[9] M. J. Simpson, A. Merrifield, K. A. Landman, and B. D. Hughes, Phys. Rev. E **76**, 021918 (2007).
[10] M. J. Simpson, K. A. Landman, and B. D. Hughes, Physica A **388**, 399 (2008).
[11] B. J. Binder, K. A. Landman, M. J. Simpson, M. Mariani, and D. F. Newgreen, Phys. Rev. E **78**, 031912 (2008).
[12] D. Zhang, I. M. Brinas, B. J. Binder, K. A. Landman, and D. F. Newgreen, Dev. Biol. **339**, 280 (2010).
[13] E. Khain, L. M. Sander, and C. M. Schneider-Mizell, J. Stat. Phys. **128**, 209 (2007).
[14] H. Enderling, L. Hlatky, and P. Hahnfeldt, Brit. J. Cancer **100**, 1917 (2009).
[15] C. A. Chung, T.-H. Lin, S.-D. Chen, and H. I. Huang, J. Theor. Biol. **262**, 267 (2010).
[16] D. Volfson, S. Cookson, J. Hasty, and L. S. Tsimring, Proc. Natl. Acad. Sci. **105**, 15346 (2008).
[17] H. Cho *et al.*, PLoS Biol. **5**(11), e302 (2007).
[18] H. Jonsson and A. Levchenko, Multiscale Model. Simul. **3**, 346 (2005).
[19] F. Hammad, R. Watling, and D. Moore, Mycol. Res. **97**, 275 (1993).
[20] B. D. Ripley, *Spatial Statistics* (Wiley, New York, 1981).
[21] P. J. Diggle, *Statistical Analysis of Spatial Point Patterns* (Academic Press, London, 1983).
[22] T. Li and I. Manas-Zloczower, International Polymer Processing **10**, 314 (1995).
[23] D. Chowdhury, A. Schadschneider, and K. Nishinari, Phys. Life. Rev. **2**, 318 (2005).
[24] N. L. Johnson, A. W. Kemp, and S. Kotz, *Univariate Discrete Distributions*, 3rd ed., (Wiley, New York, 2005).
[25] A. W. Drake, *Fundamentals of Applied Probability Theory* (McGraw-Hill, New York, 1967).
[26] M. Liu, F. J. Muzzio, and R. L. Peskin, Chaos Solitons Fractals **4**, 869 (1994).
[27] G. Cheng, B. B. Youssef, P. Markenscoff, and K. Zygourakis, Biophys. J. **90**, 713 (2006).
[28] B. J. Binder and K. A. Landman, J. Theor. Biol. **259**, 541 (2009).