Quantifying spatial structure in experimental observations and agent-based simulations using pair-correlation functions

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

Matthew J. Simpson
Mathematical Sciences, Queensland University of Technology, Brisbane, Queensland, Australia
and Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia
(Received 21 February 2013; revised manuscript received 15 May 2013; published 12 August 2013)

We define a pair-correlation function that can be used to characterize spatiotemporal patterning in experimental images and snapshots from discrete simulations. Unlike previous pair-correlation functions, the pair-correlation functions developed here depend on the location and size of objects. The pair-correlation function can be used to indicate complete spatial randomness, aggregation, or segregation over a range of length scales, and quantifies spatial structures such as the shape, size, and distribution of clusters. Comparing pair-correlation data for various experimental and simulation images illustrates their potential use as a summary statistic for calibrating discrete models of various physical processes.

DOI: 10.1103/PhysRevE.88.022705 PACS number(s): 87.17.−d, 87.18.−h, 87.85.−d

I. INTRODUCTION

Spatiotemporal patterning is observed in images of various processes, including (i) an approximately uniform distribution of cocultured cells that form multicellular aggregates [1], as shown in Figs. 1(a) and 1(b), (ii) a segregated mixture of steel and glass beads that is vibrated to produce a less segregated mixture [2], as shown in Figs. 1(c) and 1(d), and (iii) aggregates of dye placed on a viscous fluid that is stirred into folds of laminar patterns [3], as shown in Figs. 1(e) and 1(f). The evolution of such physical systems can be simulated using a range of frameworks including discrete random-walk-based methods [4–10] and other techniques [2,11–17]. Quantifying the spatial structure predicted by such modeling frameworks is essential when we consider comparing the model prediction with an experimental image.

In this work, we define a discrete pair-correlation function [18–24] to characterize spatial structure in the two-dimensional Cartesian plane. We assume that the spatial domain can be represented by a two-dimensional integer lattice with each site either being vacant or occupied by a single unit square area agent (an area exclusion process [25,26]). The area agent may represent an entire object, part of an object, or a pixel in an experimental image [27–38]. The pair-correlation function is formulated by normalizing the counts of the pair distances, in either Cartesian direction, between the area agents. For a domain populated uniformly at random, at the exclusion complete spatial randomness (ECSR) state [39–41], the normalization ensures that the expected value of the pair-correlation functions is unity for all pair distances. When the pair-correlation functions are greater than unity we have aggregation, and when the pair-correlation functions are less than unity we have segregation [18,19]. We analyze these signals over short, intermediate, and long length scales. Our work extends previous studies that have focused on point processes and considered short length scales and isotropic distributions [18,19].

We consider spatial data sets generated by a discrete model of agent proliferation and motility [4,7–9,38,42,43]. This modeling framework has been influential in quantifying the role of cell proliferation and cell motility in many different cell biology applications [38,43,44]. However, many important questions about the interpretation and quantification of spatial data sets remain unanswered, such as developing methods that can distinguish between different kinds of spatial patterning [45]. Our work confirms that the pair-correlation functions can be used to distinguish between three basic types of spatial patterning signals: (i) ECSR, (ii) aggregation, and (iii) segregation [18,19]. Furthermore, we show that more detailed signals are possible, such as wavelike oscillations indicating aggregation and segregation at multiple length scales. We demonstrate how these more detailed signals can be quantified using the pair-correlation framework by analyzing images from three different experiments [27,33,38].

II. PAIR-CORRELATION FUNCTION

Unlike previous studies [18–24] we consider an exclusion process where objects cannot overlap. We represent this using a two-dimensional integer lattice, with each site being vacant or occupied by, at most, a single square of unit area [25,26]. Focusing on an exclusion process is relevant for the analysis of experimental images that have been converted into a standard black and white format where black pixels represent the occupied area and white pixels represent the vacant area. The exclusion process framework is also relevant when we analyze snapshots from discrete models where agent exclusion is enforced [4,46].

We consider an $X \times Y$ rectangular lattice of integers (x,y) with unit spacing. The occupancy of the lattice can be represented by a matrix, given by

$$M_{xy} = \begin{cases} 0, & \text{if } (x,y) \text{ is vacant,} \\ 1, & \text{if } (x,y) \text{ is occupied.} \end{cases} \quad (1)$$

*benjamin.binder@adelaide.edu.au

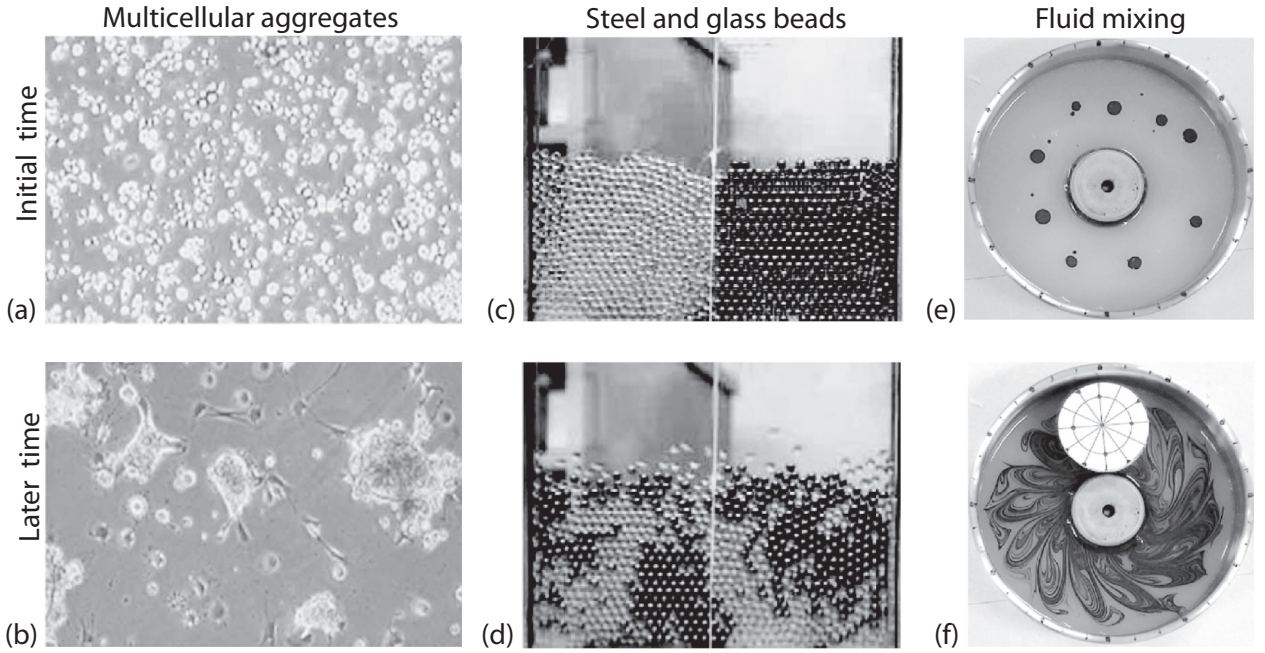

FIG. 1. Images of spatial patterning. (a),(b) An initially uniform coculture of cells forms multicellular aggregates (reprinted from Thomas *et al*. [1], Fig. 3, with permission from Eur. Cells Mater.). (c),(d) A segregated mixture of steel and glass beads evolves to a less segregated mixture when vibrated (reprinted from Yang [2], Fig. 2, with permission from Powder Technol.). (e),(f) Aggregates of dye on a viscous fluid evolve to a more regular pattern when the fluid is stirred [reprinted from Kobayashi *et al*. [3], picture 4.1(1a and 1d), with permission from Topol. Appl.].

The total number of area agents is

$$n = \sum_{x=1}^{X} \sum_{y=1}^{Y} M_{xy} \leqslant XY. \tag{2}$$

In either Cartesian direction, the nonperiodic counts of the pair distances between area agents can be described using set notation. We define the set of paired area agents as

$$\psi_{ab} = \{(a,b) \mid a = (x_a,y_a), b = (x_b,y_b), a \neq b, M_{x_a,y_a} = M_{x_b,y_b} = 1, x_a, x_b \in X, y_a, y_b \in Y\}. \tag{3}$$

The subsets of agent pairs at distance $i \in X$ or $j \in Y$ are

$$S_i = \{(a,b) \mid |x_a - x_b| = i, (a,b) \in \psi_{ab}\} \quad \text{and} \tag{4}$$

$$S_j = \{(a,b) \mid |y_a - y_b| = j, (a,b) \in \psi_{ab}\}. \tag{5}$$

The numbers of elements in the subsets S_i and S_j indicate the counts of pair distances

$$c_x(i) = |S_i| \quad \text{for} \quad i = 1, \ldots, X \quad \text{and} \tag{6}$$

$$c_y(j) = |S_j| \quad \text{for} \quad j = 1, \ldots, Y. \tag{7}$$

As an example, the arrows in Fig. 2(a) connect pairs of agents at (1,2), (3,1), and (4,2) on a lattice with $X = 4$ and $Y = 2$. The counts of pair distances in the x direction are $c_x(1) = 1$, $c_x(2) = 1$, $c_x(3) = 1$, and $c_x(4) = 0$. In general, the counts of pair distances can be binned in the following way:

$$C_x(k\Delta_x) = \sum_{i=(k-1)\Delta_x+1}^{k\Delta_x} c_x(i) \quad \text{for} \quad k = 1, \ldots, X/\Delta_x \quad \text{and} \tag{8}$$

$$C_y(l\Delta_y) = \sum_{j=(l-1)\Delta_y+1}^{l\Delta_y} c_y(j) \quad \text{for} \quad l = 1, \ldots, Y/\Delta_y, \tag{9}$$

where Δ_x and Δ_y are the bin widths. Note that $X \equiv 0 \bmod \Delta_x$ and $Y \equiv 0 \bmod \Delta_y$, as we require that X and Y are divisible by Δ_x and Δ_y without any remainder. In the case that $\Delta_x = \Delta_y = 1$, Eqs. (8) and (9) are equivalent to Eqs. (6) and (7). Introducing variable bin widths allows us to analyze the spatial distribution of objects over a range of length scales.

Figure 3(b) shows C_x and C_y for the random distribution of objects in Fig. 3(a), indicating that we have a linear relationship between the bin counts and pair distance. When there is no spatial structure and we are at the ECSR state, as in Fig. 3(a), we define a set of pair-correlation functions that normalize C_x and C_y so that their expected values are unity for all pair distances [18–24]. The normalization factor can be interpreted as the probability that we choose a pair of agents, on a randomly occupied domain, separated by a particular pair distance. This probability is the product of three terms: (i) the probability of selecting an agent, (ii) the probability of selecting a second distinct agent, and (iii) the number of combinations of pairs of sites separated by a particular pair distance. We now present the details of how to derive this normalizing factor by considering the x direction only and note that a similar argument applies in the y direction.

The probability of selecting the first agent is the mean-field density, $\rho = n/XY$, and the probability of selecting a second distinct agent is given by

$$\tilde{\rho} = \frac{n-1}{XY-1} = \frac{\rho - 1/XY}{1 - 1/XY}. \tag{10}$$

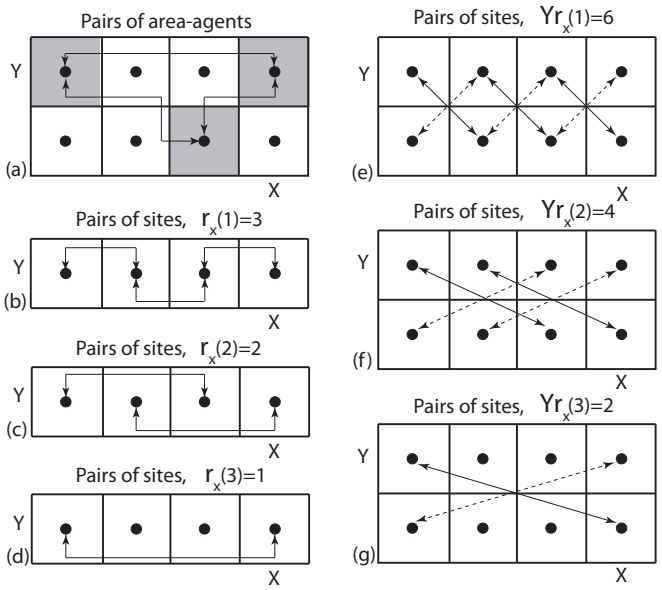

FIG. 2. (a) Gray squares represent sites that are occupied by unit sized area agents on a lattice with $X = 4$ and $Y = 2$. The counts of the pair distances between area agents in the x direction are indicated with arrows. There is one pair of agents separated by a distance of one $[c_x(1) = 1]$, one pair of agents separated by a distance of two $[c_x(2) = 1]$, one pair of agents separated by a distance of three $[c_x(3) = 1]$, and no pairs of agents separated by a distance of four $[c_x(4) = 0]$. (b)–(d) Counts of pairs of sites on a lattice with $X = 4$ and $Y = 1$. In (b) we have three pairs of sites separated by a distance of one $[r_x(1) = 3]$. In (c) we have two pairs of sites separated by a distance of two $[r_x(2) = 2]$. In (d) we have one pair of sites separated by a distance of three $[r_x(3) = 1]$. (e)–(g) Counts of pairs of lattice sites on a lattice with $X = 4$ and $Y = 2$. In these subfigures we have used a combination of solid and dashed arrows to make the counts of pairs of sites clear. In (e) we have six pairs of sites separated by a distance of one. In (f) we have four pairs of sites separated by a distance of two. In (g) we have two pairs of sites separated by a distance of three.

We now derive an expression describing the number of combinations of pairs of lattices sites separated by a particular pair distance. When the domain is populated at random the area-agent counts are expected to be proportional to the counts between lattice sites. We begin with reference to Figs. 2(b)–2(d) by considering the counts between lattice sites in a single row of a lattice, which is given by

$$r_x(i) = (X - i) \quad \text{for} \quad i = 1, \ldots, X. \tag{11}$$

For Y such rows, we have $Yr_x(i)$ combinations of pair distances. If we consider any two distinct rows in the lattice, the number of combinations of pair distances between two sites, with each site belonging to a different row, is $2r_x(i)$, as illustrated in Figs. 2(e)–2(g). To account for all possible combinations of pairs of sites from any two distinct rows of a lattice containing a total of Y rows, we introduce a binomial coefficient. Combining these considerations gives us an expression for the counts of pair distances between lattice sites,

$$d_x(i) = Yr_x(i) + 2r_x(i)\binom{Y}{2} = Y^2(X - i) \quad \text{for} \quad i = 1, \ldots, X. \tag{12}$$

The normalization term to define our pair-correlation function is given by

$$\begin{aligned}\hat{c}_x(i) &= d_x(i)\rho\tilde{\rho} \\ &= Y^2(X - i)\rho\tilde{\rho} \quad \text{for} \quad i = 1, \ldots, X\end{aligned} \tag{13}$$

for the counts of the pair distances in Eq. (6), when $\Delta_x = 1$. A similar expression holds for $\hat{c}_y(j)$ when $\Delta_y = 1$ and the normalization values can be binned according to

$$\hat{C}_x(k\Delta_x) = \sum_{i=(k-1)\Delta_x+1}^{k\Delta_x} \hat{c}_x(i) \quad \text{for} \quad k = 1, \ldots, X/\Delta_x \quad \text{and} \tag{14}$$

$$\hat{C}_y(l\Delta_y) = \sum_{j=(l-1)\Delta_y+1}^{l\Delta_y} \hat{c}_y(j) \quad \text{for} \quad l = 1, \ldots, Y/\Delta_y. \tag{15}$$

To verify our arguments we plot $\hat{C}_x$ and $\hat{C}_y$ in Fig. 3(b), which provides an excellent match to the binned counts of C_x and C_y for the distribution shown in Fig. 3(a). Therefore, the two pair-correlation functions are given by

$$P_x(k\Delta_x) = \frac{C_x(k\Delta_x)}{\hat{C}_x(k\Delta_x)} \quad \text{for} \quad k = 1, \ldots, X/\Delta_x \quad \text{and} \tag{16}$$

$$P_y(l\Delta_y) = \frac{C_y(l\Delta_y)}{\hat{C}_y(l\Delta_y)} \quad \text{for} \quad l = 1, \ldots, Y/\Delta_y. \tag{17}$$

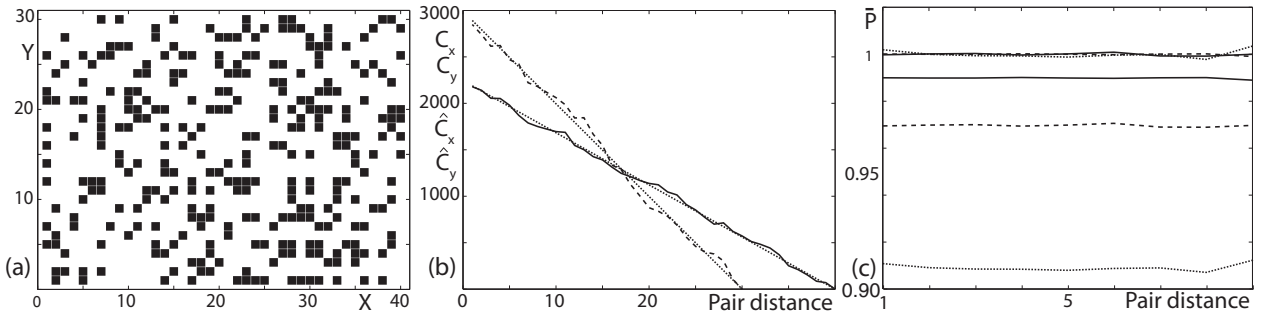

FIG. 3. Domain populated uniformly at random with (black) unit square area agents. (a)–(c) $XY = 1200$, $\rho = 0.25$, and $\Delta_x = \Delta_y = 1$. (a) Typical realization. (b) Counts of pair distances C_x and C_y (solid and dashed curves). The dotted lines are for the normalization values $\hat{C}_x$ and $\hat{C}_y$. (c) Average pair-correlation function $\bar{P}$, $N = 10000$, $XY = 100$, and $\Delta = 1$. The dotted, dashed, and solid curves are for mean field densities $\rho = \{0.10, 0.25, 0.5\}$. The upper three curves are for $\tilde{\rho}$, given by Eq. (10). The lower three curves are for $\tilde{\rho} = \rho$.

The pair-correlation function signals, P_x and P_y, for a uniformly random distribution of area agents in Fig. 3(a), fluctuate around unity as expected. We can average P_x and P_y over N identically prepared realizations to give $\bar{P}_x(k\Delta_x)$ and $\bar{P}_y(l\Delta_y)$. For an isotropic distribution, we can further average to obtain $\bar{P}(m\Delta) = [P_x(k\Delta_x) + P_y(l\Delta_y)]/2$, for $X = Y$, $\Delta = \Delta_x = \Delta_y$, and $m = k = l$.

Results in Fig. 3(c) show $\bar{P}(m\Delta)$ for a randomly populated domain at three different mean densities, $\rho = 0.10$, 0.25, and 0.50. The significance of accounting for exclusion is illustrated by comparing the three upper curves with the three lower curves in Fig. 3(c). The three upper curves correspond to our approach for the analysis of exclusion processes with $\rho = 0.10$, 0.25, and 0.50, and we have $\bar{P}(m\Delta) \approx 1$ for all ρ considered. The three lower curves show $\bar{P}(m\Delta)$ for $\rho = 0.10$, 0.25, and 0.50, where we made no distinction between $\tilde{\rho}$ and ρ, such as in the case in a standard nonexclusion point process. The three lower curves show that $\bar{P}(m\Delta) \neq 1$, confirming that significant differences between the standard approach to analyze point processes and our approach for dealing with exclusion processes can occur.

We will now implement a discrete model to generate various biologically inspired data sets, that are not necessarily spatially uniform, and analyze these patterns using our discrete pair-correlation functions. We note that all the theory developed in Sec. II is for a single snapshot of a spatial process. Since we will consider a discrete model we will apply the results from Sec. II to snapshots from the model at the end of each discrete time step to give an estimate of how the pair-correlation functions change with time. When we evaluate $\bar{P}$, we choose N, the number of realizations, to ensure that the fluctuations in $\bar{P}$ are sufficiently small and we evaluate this based on a visual examination of the results in each case.

III. SPATIAL ANALYSIS OF DISCRETE SIMULATIONS

We consider a discrete model of biological cell proliferation and motility [4,7–9,43] on a nondimensional square lattice of unit spacing, $\delta = 1$. At time t the domain contains $\hat{n}(t)$ square cellular agents, each of area $s = (1 + 2\alpha)^2$, where α is a non-negative integer and s denotes the number of unit square area agents contained in a cellular agent so that the total number of area agents is $n(t) = s\hat{n}(t)$. If $\alpha = 0$ with $s = 1$, each cellular agent is represented by a single area agent. When $\alpha > 0$ the configuration of the cellular agent can be described as follows. If the central area agent of the cellular agent is at (x, y), then the four area agents at the corners of the cellular agent are located at $(x \pm \alpha, y \pm \alpha)$. The remaining $s - 5$ area agents occupy lattice sites to form a single square cellular agent. For example, a square cellular agent with $\alpha = 1$ has nine area agents at $(x, y \pm 1)$,$(x \pm 1, y)$, (x, y), and $(x \pm 1, y \pm 1)$.

The temporal evolution of the discrete model is simulated with a random sequential update method [25,26] with nondimensional time steps of unit duration, $\tau = 1$. On average, each cellular agent is given the opportunity to undergo one unbiased motility event during each time step and the motility event will occur with probability $\mathbb{P}_m \in [0,1]$. For example, in a successful motility event, the central area agent at (x, y) of a particular cellular agent would attempt to move to $(x \pm 1, y \pm 1)$, with the direction of movement chosen at random. Since the model is an exclusion process, any potential motility event that would place an area agent on a lattice site that is occupied by another cellular agent is aborted. Once the $\hat{n}(t)$ potential motility events have been assessed we then allow, on average, each cellular agent the opportunity to undergo a proliferation event with probability $\mathbb{P}_p \in [0,1]$ per time step, and any new daughter agent that is produced is placed adjacent to the mother agent. We consider an unbiased proliferation mechanism where the direction of the target site is chosen at random, and any potential proliferation events that would place an area agent on an occupied site is aborted. Our discrete simulations are nondimensional in the sense that we consider a square lattice with unit spacing, $\delta = 1$, and time steps of unit duration, $\tau = 1$ [26]. This nondimensional framework can be applied to any particular biological application by rescaling, using appropriate length and time scales, L and T, as discussed previously [26]. The parameters in the discrete model are related to standard measures of cell diffusivity, $D = \mathbb{P}_m\delta^2/(4\tau)$, and the cell proliferation rate, $\lambda = \mathbb{P}_p/\tau$ [38]. The relationship between the parameters in the discrete model and the cell diffusivity and cell proliferation rate have been analyzed previously [26].

To mimic the way that experimental images are recorded and analyzed [27–38] we always consider a central portion of the domain so that the pair-correlation functions are unaffected by the boundary conditions. To test the sensitivity of our results to differences in the boundary conditions we used both periodic and reflecting boundary conditions in our simulations and found that the pair-correlation signals were insensitive to these differences provided that the domain was sufficiently large relative to the size of the central region of the domain that we analyzed.

Snapshots in Figs. 4(a) and 4(b) correspond to proliferative and immotile ($\mathbb{P}_p > 0, \mathbb{P}_m = 0$) unit square agents where the domain is initially populated by sixteen evenly spaced agents. The pair-correlation signals P_x, P_y, and $\bar{P}$ in Fig. 4(c) are all similar for the isotropic spatial pattern in Fig. 4(b). All signals contain four maxima indicating aggregation on four different length scales. The first maxima at pair distance $k\Delta_x = l\Delta_y = m\Delta \approx 1$ corresponds to short scale aggregation within each aggregate and the first minima at $k\Delta_x = l\Delta_y = m\Delta \approx 10$ indicates segregation at a length scale corresponding to the average aggregate length. The three remaining maxima indicate intermediate and long scale aggregation between pairs of nearest aggregates and pairs of next nearest aggregates. The distance separating the adjacent maxima and minima, $k\Delta_x = l\Delta_y = m\Delta \approx 20$, corresponds to the distance between the central location of aggregates.

We now consider a more stochastic pattern by analyzing randomly initialized domains containing proliferative and immotile agents ($\mathbb{P}_p > 0, \mathbb{P}_m = 0$), with $\rho(0) = 0.001$. Typical snapshots are given in Figs. 5(a) and 5(b), with $\bar{P}$ shown in Fig. 5(c) containing one maxima which indicates short scale aggregation. The pair distance that the signals intersect with unity gives us a measure of the average aggregate length. At larger pair distances, the pair-correlation functions fluctuate around constant values, less than unity, indicating intermediate and long scale segregation. However, there is no other spatial structure at these larger pair distances, which indicates that the aggregates are distributed uniformly at random throughout the

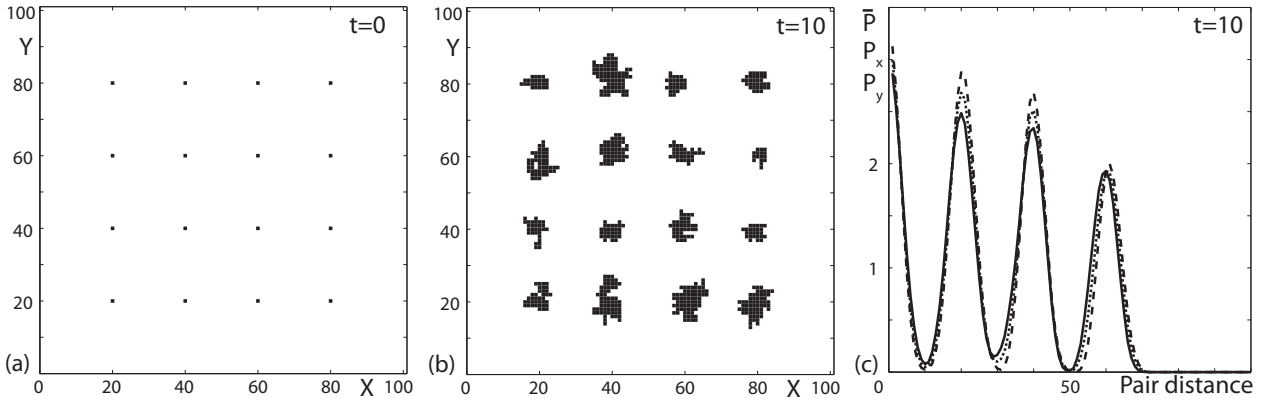

FIG. 4. Cellular-agent aggregates generated by the discrete proliferation mechanism, $XY = 100\,000$, $\mathbb{P}_p = 1.0$, $\mathbb{P}_m = 0$, and $s = 1$. (a) Initial condition. (b) Typical realization at $t = 10$. (c) Pair-correlation functions P_x (solid) and P_y (dashed) for (b), at $t = 10$ and $\Delta_x = \Delta_y = 1$. The dotted curve is the average pair-correlation function $\bar{P}$, $N = 1$, and $\Delta = 1$.

domain. This is expected as the domain was initially occupied at random and the pair-correlation signal at $t = 0$ in Fig. 5(c) fluctuates around unity.

The stochasticity in the simulations can be further increased by considering randomly initialized simulations of proliferative and motile agents ($\mathbb{P}_p > 0, \mathbb{P}_m > 0$). A snapshot of such a process in Fig. 6(a) suggests that, unlike the results in Fig. 5(a), it is very difficult for us to visually assess whether there is any aggregation present in the system. The corresponding pair-correlation signal in Fig. 6(b) indicates that there is short scale aggregation and we confirm this by considering an independent measure of the spatial structure using an index introduced by Phelps and Tucker [14]. The index, $I(t)$, is a scaled variance obtained by counting the number of agents in B equally sized bins. For a lattice-based exclusion processes, with $s = 1$, the ECSR limit [41,43,47,48] is

$$I(t)_{\text{ecsr}} = \frac{1}{n(t)} - \frac{1}{XY}. \tag{18}$$

Comparing averaged values of $I(t)$ and $I(t)_{\text{ecsr}}$ gives us an independent test for the presence of spatial structure. Results in Fig. 6(c) show that $I(t) > I(t)_{\text{ecsr}}$, which indicates that the spatial distribution of objects in Fig. 6(a) is not at the ECSR state. We note that the Phelps and Tucker index provides us with relatively little information about the spatial structure since this index only provides a test of whether a particular distribution of objects is uniform. Since we have $I(t) > I(t)_{\text{ecsr}}$, the index indicates that the system is not uniform; however, this test provides no further information. In comparison, our pair-correlation data in Fig 6(b) reveals far more information since it indicates that the distribution is aggregated over a length scale of approximately $m\Delta \approx 5$ and is distributed uniformly at larger length scales.

Our results so far have concentrated on simulations where $s = 1$ and each square cellular agent is equivalent to a single area agent. We now consider simulations with $s > 1$ so that each cellular agent consists of several area agents. Interpreting the pair-correlation signals in this more general case is relevant when analyzing experimental images where a single *image agent* consists of several pixels (or area agents). We begin by examining a domain randomly populated with square cellular agents, each of area $s = 169$ and length $\sqrt{169} = 13$, as shown in Fig. 7(a). Results in Fig. 7(b) show $\bar{P}$ for all the area agents, as well as showing $\bar{P}$ for the central area agents of

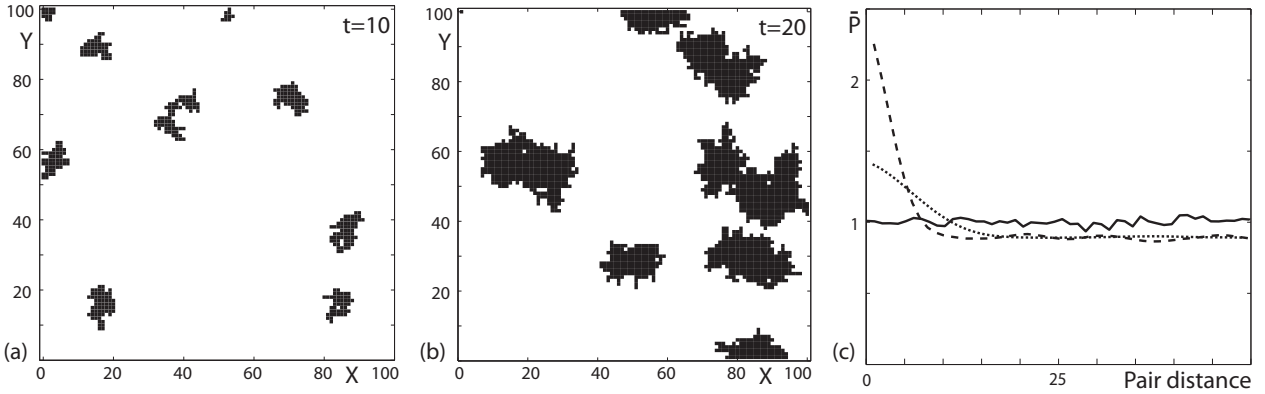

FIG. 5. Cellular-agent aggregates generated by the discrete proliferation mechanism, $XY = 100\,000$, $\mathbb{P}_p = 1.0$, $\mathbb{P}_m = 0$, and $s = 1$. Initially the domain was populated uniformly at random with unit square cellular agents and $\rho(0) = 0.001$. (a),(b) Typical snapshots at $t = 10$ and 20, respectively. (c) $\bar{P}$ at $t = 10$ (dashed) and $t = 20$ (dotted) with $N = 100$ and $\Delta = 1$. The solid curve is for the initial distribution of agents with $N = 1000$ and $\Delta = 1$.

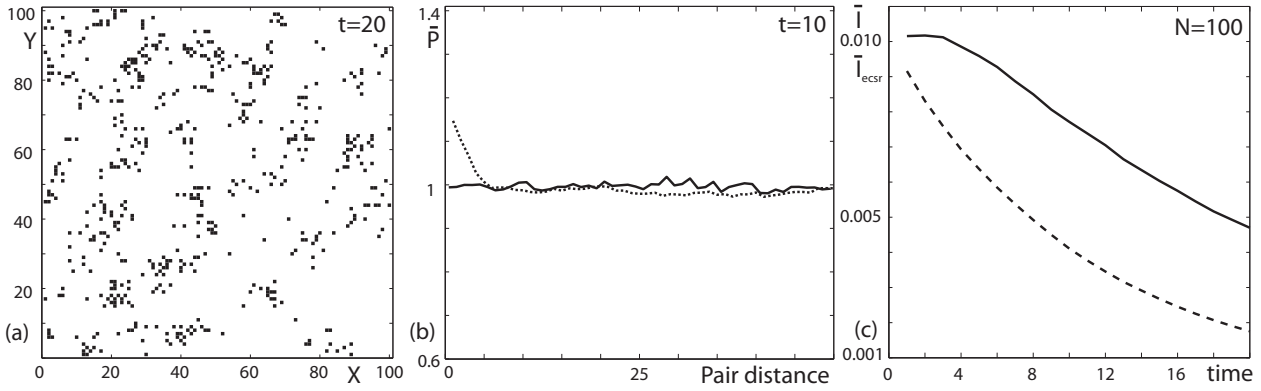

FIG. 6. Aggregation patterns generated by the discrete proliferation and motility mechanism, $XY = 100\,000$, $\mathbb{P}_p = 0.1$, and $\mathbb{P}_m = 1.0$. Initially the domain was populated uniformly at random with unit square cellular agents at $\rho(0) = 0.01$. (a) Typical realization at $t = 20$. (b) Average pair-correlation functions $\bar{P}$ at $t = 10$ and $\Delta = 1$. The dotted curve is for all the agents within the domain, $N = 100$. The solid curve is for the agents that initially seeded the domain, $N = 200$. (c) Average index $\bar{I}$ (solid curve) for $B = 400$ and $N = 100$. The dashed curve represents the average ECSR limiting value $\bar{I}_{\text{ecsr}}$.

each cellular agent. Results in Fig. 7(c) are equivalent to those in Fig. 7(b) except that a different bin width was used. For $\Delta = 1$, both signals in Fig. 7(b) accurately predict the length of the cellular agents $m\Delta \approx 13$. The pair correlation for the area agents shows short scale aggregation indicating the size of the cellular agents, whereas the pair correlation for the central area agents of each cellular agent indicates short scale segregation which is consistent with the fact that the distance between any pair of central area agents cannot be less than $m\Delta = 13$. Both pair-correlation signals in Fig. 7(b) confirm a lack of spatial structure at larger length scales indicating that the distribution is uniform at these larger distances. The pair-correlation signals in Fig. 7(c) obtained with a larger bin width indicate that some of the short scale information present in Fig. 7(b) is lost. Regardless of these details, if we compare the pair-correlation signals for Figs. 3(a), 5(a), and 7(a), we see that these different types of spatial patterns produce distinguishable signals.

The snapshot in Fig. 8(a) corresponds to a typical simulation of proliferative and immotile ($\mathbb{P}_p > 0, \mathbb{P}_m = 0$) cellular agents (each of area $s = 49$ and length $\sqrt{49} = 7$), for a domain that is initial populated uniformly at random. The patterning is isotropic and $\bar{P}$ for the central area agents contains multiple maxima indicating aggregation at multiple length scales, similar to Fig. 4(a), but distinguishable from Fig. 5(b). The periodic maxima in Fig. 8(b) reflects the increased frequency of pair distances between central area agents as proliferation events lead to regular clustering. The emergence of new maxima with time can be interpreted as a measure of generation time. For example, the five maxima in Fig. 8(b) at $t = 5$ correspond to approximately five generations in this case. The pair-correlation curve for all the area agents in Fig. 8(b) indicates that there is one aggregation scale which corresponds to the average length of the cellular-agent aggregates. Both pair-correlation signals in Fig. 8(b) approach the same constant level after a sufficiently large pair distance indicating that the cellular agents are distributed uniformly throughout the domain at large distances. This multiscale spatiotemporal pattern, illustrated by the two signals, provides a measure of cellular-aggregate size, with the dotted signal giving additional information on both cellular-agent size and generation time.

We now consider an anisotropic spatial pattern from a discrete simulation that mimics a scrape wound assay, which is a standard experiment in the cell biology literature [33]. The initial condition in Fig. 9(a) consists of two horizontal regions

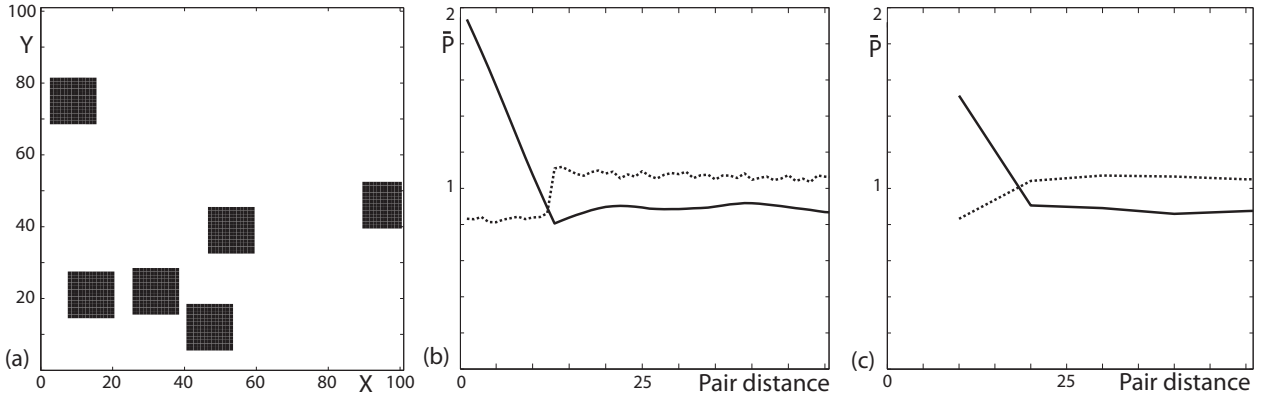

FIG. 7. Domain populated uniformly at random with square cellular agents, $s = 169$, $XY = 10\,000$, and $\rho = 0.1$. (a) Typical realization. (b),(c) Average pair-correlation functions $\bar{P}$. The solid curves are for all the area agents in the domain, $N = 100$. The dotted curves are for the central area agents within each of the cellular agents, $N = 10\,000$. (b) $\Delta = 1$. (c) $\Delta = 10$.

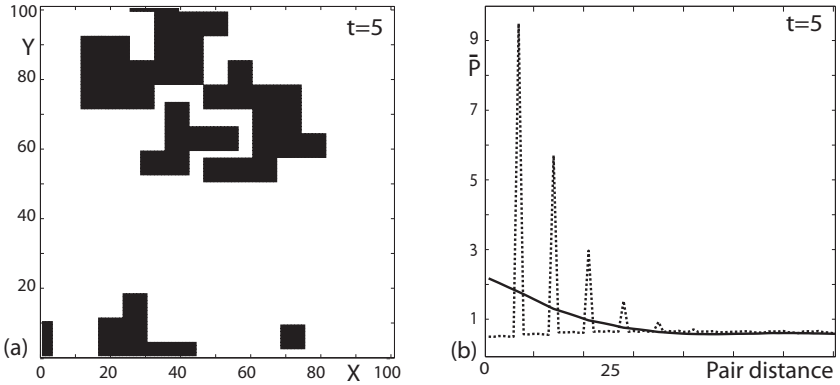

FIG. 8. Aggregation patterns generated by the discrete proliferation mechanism, $XY = 100\,000$, $\mathbb{P}_p = 1.0$, and $\mathbb{P}_m = 0$. (a) Snapshot of a realization at $t = 5$ for a domain initially populated uniformly at random with cellular agents of size $s = 49$ and $\rho(0) = 0.01$. (b) Average pair-correlation functions $\bar{P}$ for (a). The dotted curve is for the central area agents within each of the cellular agents, $N = 10000$ and $\Delta = 1$. The solid curve is for all the area agents in the domain, $N = 100$ and $\Delta = 1$.

populated uniformly at random with unit square cellular agents. These two regions are separated by a third initially vacant region which represents the wound. We consider motile and proliferative agents ($\mathbb{P}_m > 0, \mathbb{P}_p > 0$), and typical snapshots of the healing process are shown in Figs. 9(c) and 9(e). The evolution of P_x and P_y in Figs. 9(b), 9(d), and 9(f) shows $P_x(m\Delta) \approx 1$, confirming the absence of spatial structure in the x direction for all $t > 0$. The P_y signal for the initial condition in Fig. 9(b) indicates aggregation up to $m\Delta \approx 5$, which is the half-width of the vacant region in Fig. 9(a). For larger pair distances P_y indicates intermediate scale segregation. The minimum in the P_y signal occurs at pair distance $m\Delta \approx 10$, which is the height of the vacant strip in Fig. 9(a). For $m\Delta > 15$, P_y indicates long scale aggregation corresponding to the geometry of the initial condition. For $m\Delta \approx 20$, P_y reaches a constant maximum value indicating a lack of any other spatial structuring for $m\Delta > 20$. As the simulation proceeds we observe that the shape of the P_y signal remains similar with time but that we have $P_y \to 1$ as $t \to \infty$ and the domain becomes uniformly populated as the wound closes. We note that applying the pair-correlation signals to this kind of anisotropic process could be useful to indicate the presence of short-scale aggregation, such as we observed previously in Fig. 6 for the isotropic uniform initial condition, which can be subtle and not always discernible using visual inspection. Understanding whether or not short-scale aggregation plays a role in such processes can have major implications with regard to the use of standard continuum models of such processes [42].

IV. SPATIAL ANALYSIS OF EXPERIMENTAL IMAGES

We conclude by considering three different images from cell biology experiments: a scrape wound assay [33], multicellular aggregates [27], and randomly distributed cells [38], as shown in Figs. 10–12. Our pair-correlation functions can be applied to two-dimensional digital images where a pixel represents the smallest element in the image. We used

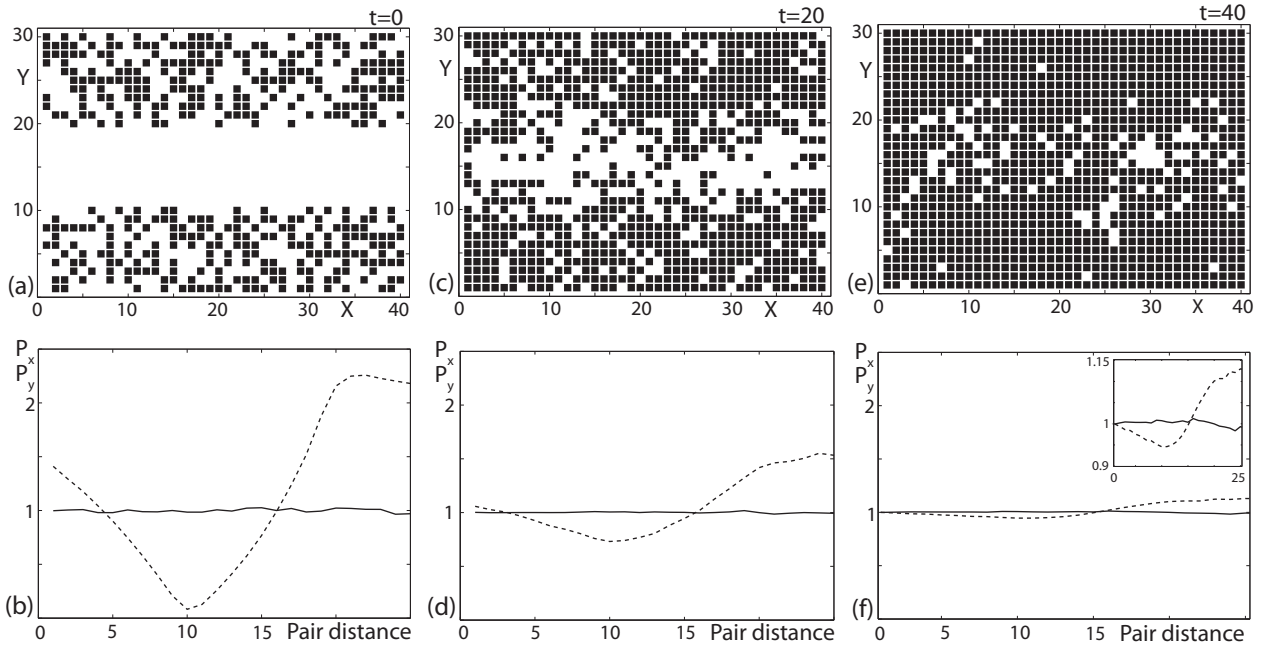

FIG. 9. Simulation of a scrape wound assay, using the discrete proliferation and motility mechanism, $s = 1$, $XY = 1200$, $\mathbb{P}_p = 0.1$, and $\mathbb{P}_m = 1.0$. (a) Initial condition at $t = 0$. The uniform density in each populated strip is $\rho_s = 0.5$. (c) and (e) Snapshots of the simulation at $t = 20$ and $t = 40$. (b), (d), and (f) Pair-correlation functions P_x (solid curves) and P_y (dashed curves) for (a), (c), and (e), $N = 1$ and $\Delta_x = \Delta_y = 1$.

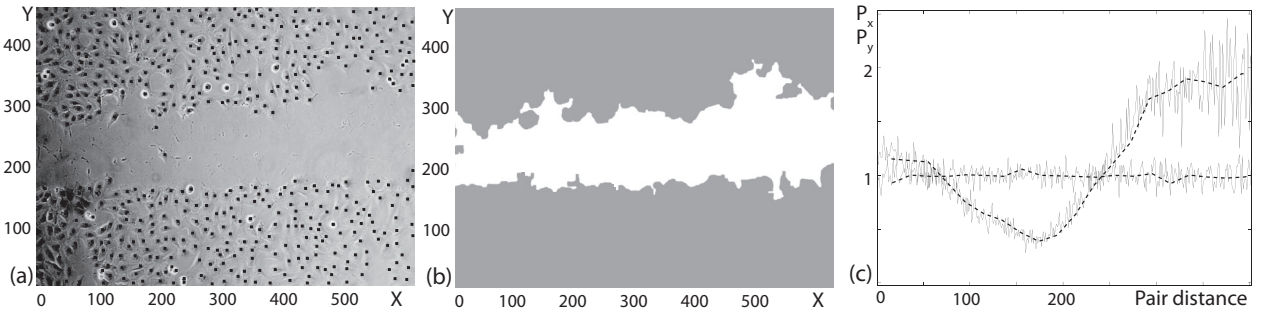

FIG. 10. Spatial analysis of a scrape wound assay, $XY = 281\,190$. (a) Experimental image [reprinted from Glab *et al.* [33], Fig. 1(a), with permission from Pattern Recogn.]. The black markers indicate the location of each cell, $n = 554$. (b) Overestimate of the total area (gray) of the domain that is occupied by cells, $S = 205\,000$ [reprinted from Glab *et al.* [33], Fig. 4(c), with permission from Pattern Recogn.]. (c) P_x and P_y with $\Delta_x = \Delta_y = 1$ (solid curves) and $\Delta_x = \Delta_y = 20 \approx \sqrt{S/n}$ (dashed curves). The curves that fluctuate around unity are for P_x, while the curves that deviate from unity are for P_y.

MATLAB's Image Processing Toolbox [49] to convert the images into black and white, and we represent a black pixel with a unit square area agent. We define a single (square) image agent as a collection of adjacent unit square area agents (pixels). Our aim now is to demonstrate that the pair-correlation signals characterize the features in these images. In all our analyses we choose a pixel to represent a unit of length that is equivalent to the unit of pair distance in Figs. 10–12.

Similar to the simulation results in Figs. 7 and 8, the pair-correlation signals for the experimental images can be calculated using either the center of individual cells or the area of the cells. The position of individual cells is represented by a single (unit square) area agent located approximately at the center of the cell as indicated by the markers in Fig. 10(a) and Fig. 12(a), with the area of the cells being represented by a number of adjacent image agents such as those shown in Fig. 11(b) and Fig. 12(b). We prefer to calculate both forms of the pair-correlation signals where possible; however, for some images, such as Fig. 11(a), where we cannot distinguish between different cells within each multicellular aggregate, this is not possible.

The pair-correlation signals for the scrape wound assay in Fig. 10(a) are generated by identifying the locations of n individual cells in the image. We calculated P_x and P_y for two bin widths to give the results in Fig. 10(c). The largest bin width approximates the average cell length $\sqrt{S/n}$, where S is an upper bound estimate of the total area occupied by cells that is given in Fig. 10(b). As the bin width increases, the fluctuations in P_x and P_y decrease, but both sets of signals reflect the same characteristics. For example, the two P_y signals give approximately the same estimate of (i) the pair distance where the signal intersects with unity, (ii) the pair distance at which P_y is a maximum, and (iii) the same constant value of the signal for pair distances $l\Delta_y > 300$. Moreover, the interpretation of these signals is similar to the discussion about the results in Fig. 9 where the bin width was equal to the length of the cellular agents (i.e., $\Delta_x = \Delta_y = \sqrt{s} = 1$).

To analyze the multicellular aggregates [27] in Fig. 11(a), we processed the image so that the aggregates could be represented by image agents of size $s = 25$, as shown in Fig. 11(b). Visually we see that the size of the image agents is less than the average cell size. We are unable to assess whether the spatial patterning is isotropic by comparing $\bar{P}_x$ and $\bar{P}_y$ since we only have access to the $N = 1$ image. However, there is no reason to anticipate any directional bias in the experimental procedure [27] so we calculated $\bar{P}$, given in Fig. 11(c) for $N = 1$, in the usual way.

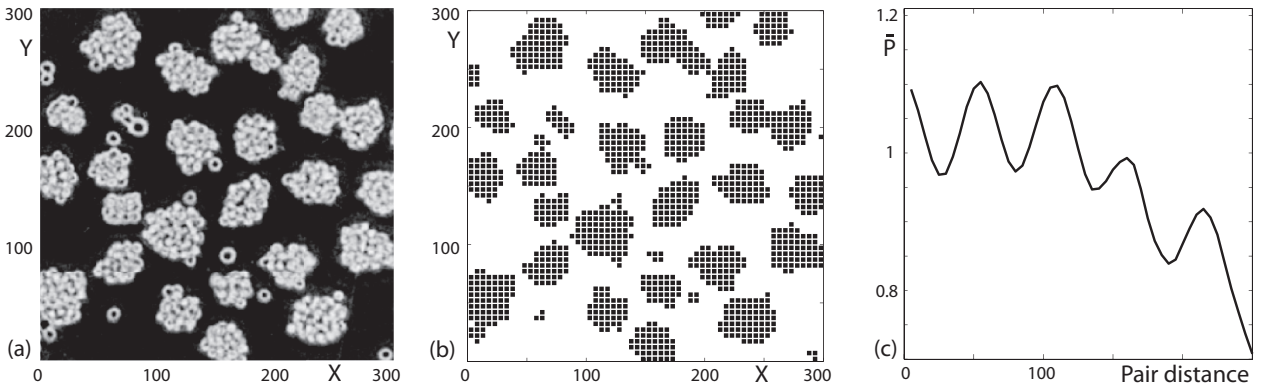

FIG. 11. Spatial analysis of multicellular aggregates, $XY = 90\,000$. (a) Experimental image [reprinted from Shimoyama *et al.* [27], Fig. 4(L9-30), with permission from Biochem. J.]. (b) Image agents that approximate the area occupied with cells in (a), $S = 31950$, $\hat{n} = 1278$, and $s = 25$. (c) Average pair-correlation function $\bar{P}$ for (b), $\Delta = 5$ and $N = 1$.

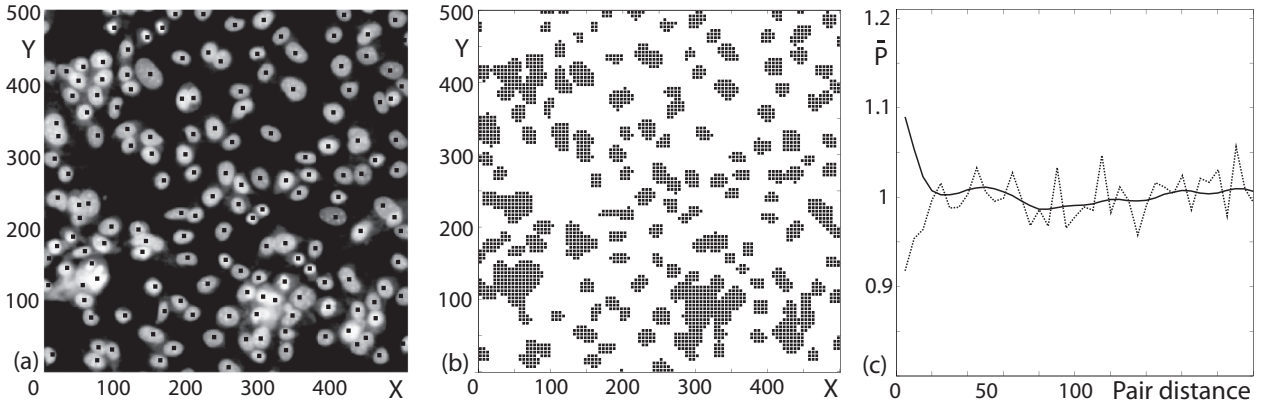

FIG. 12. Spatial analysis of evenly distributed cells, $XY = 250\,000$. (a) Experimental image used in the analysis [38]. The superimposed markers represent the location of each cell, $n = 163$. (b) Image agents that approximate the area occupied with cells in (a), $S = 67050$, $\hat{n} = 2682$, and $s = 25$. (c) Average pair-correlation functions $\bar{P}$, $N = 10$ and $\Delta = 5$. The solid curve is for image agents that approximate the area occupied with cells [e.g., (b)]. The dotted curve is for the locations of the cells, for example, the superimposed markers in (a).

The five maxima in $\bar{P}$ in Fig. 11(c) indicate short, intermediate, and long scale aggregation which corresponds to aggregation within each of the cellular aggregates, between pairs of nearest cellular aggregates, between pairs next nearest cellular aggregates, and so on. The signal is similar to Fig. 4(c) and provides quantitative information about the average cellular aggregate size and the average distance between the aggregates.

The image in Fig. 12(a) shows a population of cells in a barrier assay, which is an experimental technique used to quantify the spreading characteristics of different cell types. The experiment was performed with great care to ensure that the cells were placed within the barrier so that they were distributed as evenly as possible [38,50]. To analyze this distribution we used $N = 10$ identically prepared experimental images. For each image we recorded the number of cells and the number of image agents required to approximate the area occupied by cells. Using this information we estimated the average area $\bar{a} \approx 400$ and average diameter of a cell $\sqrt{\bar{a}} \approx 20$. The corresponding $\bar{P}$ signal, shown in Fig. 12(c), was calculated by considering both the central area agent and the image agents that approximate the area occupied by cells in these images. The pair-correlation signal for the image agents indicates short scale aggregation, while the pair-correlation signal for the central agents indicates short scale segregation. Both signals fluctuate around constant values, close to unity, for length scales that are greater than the average length of a cell, $m\Delta > \sqrt{\bar{a}} \approx 20$, confirming that the cells are distributed evenly.

V. DISCUSSION

We have derived a discrete pair-correlation function that depends on the location and size of objects in an exclusion process. We have demonstrated how to analyze spatial patterns in images produced by discrete models and from cell biology experiments, and our results indicate that these pair-correlation functions can characterize and quantify different types of spatial patterning over multiple length scales.

Pair-correlation functions have been used previously to analyze point processes with various applications including plant ecology [18–20]. We extend these previous studies by using pair-correlation functions to account for both the location and size of objects, such as biological cells or particles in a granular material, as illustrated in Fig. 1. Our pair-correlation functions explicitly incorporate area exclusion, and this allows us to study more detailed spatial patterns. For example, the pair-correlation signals from some discrete simulations and experiment images showed both short-scale segregation and short scale aggregation depending on whether we consider the center of the object or the area of the object. These details cannot be deduced using previous pair-correlation functions, which only considered point processes without any consideration of object size. For some of our proliferative simulation data we observed multiple maxima in the pair-correlation signals indicating aggregation over multiple length scales. We anticipate that this kind of signal might be relevant to images of certain types of cells that form clusters as a result of being highly proliferative but relative immotile [42].

One application of our exclusion pair-correlation functions could be as a summary statistic for parametrizing discrete models to mimic a set of experimental observations such as describing the closure of a scrape wound assay (see Figs. 9 and 10). Such applications could involve using inference methods, such as approximate Bayesian computation [51,52], to estimate parameters in a discrete model [38]. Another topic for future research is to use the pair-correlation function to attempt to distinguish between different mechanisms that give rise to the formation of multicellular aggregates, such as making a distinction between aggregation due to high cell proliferation rates and aggregation due to strong cell-to-cell adhesion [46].

ACKNOWLEDGMENTS

This work was supported by the Australian Research Council, DP120100551. We thank Parvathi Haridas for generating the experimental image Fig. 12(a).

[1] R. J. Thomas, A. Bennett, B. Thomson, and K. M. Shakesheff, Eur. Cells Mater. **11**, 16 (2006).
[2] S. C. Yang, Powder Technol. **164**, 65 (2006).
[3] T. Kobayashi and S. Umeda, Topol. Appl. **157**, 280 (2010).
[4] M. J. Simpson, A. Merrifield, K. A. Landman, and B. D. Hughes, Phys. Rev. E **76**, 021918 (2007).
[5] B. J. Binder, K. A. Landman, M. J. Simpson, M. Mariani, and D. F. Newgreen, Phys. Rev. E **78**, 031912 (2008).
[6] B. J. Binder and K. A. Landman, J. Theor. Biol. **259**, 541 (2009).
[7] D. Zhang, I. M. Brinas, B. J. Binder, K. A. Landman, and D. F. Newgreen, Dev. Biol. **339**, 280 (2010).
[8] M. J. Simpson, K. A. Landman, and B. D. Hughes, Physica A **388**, 399 (2009).
[9] R. E. Baker and M. J. Simpson, Physica A **391**, 3729 (2012).
[10] H. Jonsson and A. Levchenko, Multiscale Model. Simul. **3**, 346 (2005).
[11] P. A. Cundall and O. D. L. Strack, Geotechnique **29**, 47 (1979).
[12] S. S. Hsiau and S. C. Yang, Chem. Eng. Sci. **58**, 339 (2003).
[13] C. Henrique, G. Batrouni, and D. Bideau, Phys. Rev. E **63**, 011304 (2000).
[14] J. H. Phelps and C. L. Tucker, Chem. Eng. Sci. **61**, 6826 (2006).
[15] B. J. Binder and S. M. Cox, Fluid Dynam. Res. **40**, 34 (2008).
[16] B. J. Binder, Phys. Lett. A **374**, 3483 (2010).
[17] L. Rossi, D. Doorly, and D. Kustrin, Phys. Rev. E **86**, 026313 (2012).
[18] M. Raghib, N. A. Hill, and U. Dieckmann, J. Math. Biol. **62**, 605 (2011).
[19] U. Dieckmann, R. Law, and J. A. J. Metz, *The Geometry of Ecological Interactions: Simplifying Spatial Complexity* (Cambridge University Press, Cambridge, UK, 2000).
[20] R. Law, D. J. Murrel, and U. Dieckmann, Ecology **84**, 252 (2003).
[21] E. Hinde, F. Cardarelli, M. A. Digman, and E. Gratton, Proc. Natl. Acad. Sci. USA **107**, 16560 (2010).
[22] H. P. Zhang, A. Be'er, E.-L. Florin, and H. L. Swinney, Proc. Natl. Acad. Sci. USA **107**, 13626 (2010).
[23] F. Cardarelli and E. Gratton, PLoS ONE **5**, e10475 (2010).
[24] W. R. Young, A. J. Roberts, and G. Stuhne, Nature (London) **412**, 328 (2001).
[25] D. Chowdhury, A. Schadschneider, and K. Nishinari, Phys. Life Rev. **2**, 318 (2005).
[26] M. J. Simpson, K. A. Landman, and B. D. Hughes, Physica A **389**, 3779 (2010).
[27] Y. Shimoyama, G. Tsulimoto, M. Kitajima, and M. Natori, Biochem. J. **349**, 159 (2000).
[28] S. A. Kim, C.-Y. Tai, L.-P. Mok, E. A. Mosser, and E. M. Schuman, Proc. Natl. Acad. Sci. USA **108**, 9857 (2011).
[29] S.-Y. Park, M.-Y. Jung, and I.-S. Kim, FEBS Lett. **583**, 1375 (2009).
[30] C. L. Bauwens, R. Peerani, S. Niebruegge, K. A. Woodhouse, E. Kumacheva, M. Husain, and P. W. Zandstra, Stem Cells **26**, 2300 (2008).
[31] M. C. Garcia, J. T. Lee, C. B. Ramsook, D. Alsteens, Y. F. Dufrêne, and P. N. Lipke, PLoS ONE **6**, e17632 (2011).
[32] Z. Jevnikar, N. Obermajer, M. Bogyo, and J. Kos, J. Cell Sci. **121**, 2652 (2008).
[33] M. Glab, B. Möller, A. Zirkel, K. Wächter, S. Hüttelmaier, and S. Posch, Pattern Recogn. **45**, 3154 (2012).
[34] C.-C. Liang, A. Y. Park, and J.-L. Guan, Nat. Protoc. **2**, 329 (2007).
[35] J.-C. Lin, S.-K. Liao, E.-H. Lee, M.-S. Hung, Y. Sayion, H.-C. Chen, C.-C. Kang, L.-S. Huang, and J.-M. Cherng, J. Biomed. Sci. **16**, 105 (2009).
[36] E. Scoditti, N. Calabriso, M. Massaro, M. Pellegrino, C. Storelli, G. Martines, R. De Caterina, and M. A. Carluccio, Arch. Biochem. Biophys. **557**, 81 (2012).
[37] J. C. Yarrow, Z. E. Perlman, N. J. Westwood, and T. J. Mitchison, BMC Biotechnol. **4**, 1472 (2004).
[38] M. J. Simpson, K. K. Treloar, B. J. Binder, P. Haridas, K. J. Manton, D. I. Leavesley, D. L. S. McElwain, and R. E. Baker, J. R. Soc. Interface **10**, 20130007 (2013).
[39] B. D. Ripley, *Spatial Statistics* (Wiley, New York, 1981).
[40] P. J. Diggle, *Statistical Analysis of Spatial Point Patterns* (Academic Press, London, 1983).
[41] B. J. Binder and K. A. Landman, Phys. Rev. E **83**, 041914 (2011).
[42] M. J. Simpson, B. J. Binder, P. Haridas, B. K. Wood, K. K. Treloar, D. L. S. McElwain, and R. E. Baker, Bull. Math. Biol. **75**, 871 (2013).
[43] B. J. Binder, K. A. Landman, D. F. Newgreen, J. E. Simkin, Y. Takahashi, and D. Zhang, Bull Math Biol. **74**, 474 (2012).
[44] K. A. Landman, B. J. Binder, and D. F. Newgreen, Lect. Notes Comput. Sc. **7495**, 405 (2012).
[45] M. J. Simpson, K. A. Landman, B. D. Hughes, and A. E. Fernando, Physica A **389**, 1412 (2010).
[46] M. J. Simpson, C. Towne, D. L. Sean McElwain, and Z. Upton, Phys. Rev. E **82**, 041901 (2010).
[47] E. J. Hackett-Jones, K. J. Davies, B. J. Binder, and K. A. Landman, Phys. Rev. E **85**, 061908 (2012).
[48] B. J. Binder, E. J. Hackett-Jones, J. Tuke, and K. A. Landman, ANZIAM J. **53**, 122 (2011).
[49] Mathworks 2013. Image Aquisition Toolbox. Users Guide R2012b. Available from http://www.mathworks.com.au/products/image/ (2013).
[50] K. K. Treloar and M. J. Simpson, PLoS ONE **8**, e67389 (2013).
[51] D. B. Rubin, Ann. Stat. **12**, 1151 (1984).
[52] M. A. Beaumont, W. Zhang, and D. J. Balding, Genetics **162**, 2025 (2002).