Physica A 391 (2012) 3729–3750

Contents lists available at SciVerse ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Models of collective cell motion for cell populations with different aspect ratio: Diffusion, proliferation and travelling waves

Ruth E. Baker [a,*], Matthew J. Simpson [b,c]

[a] Centre for Mathematical Biology, Mathematical Institute, University of Oxford, 24-29 St Giles', Oxford, OX1 3LB, UK

[b] Mathematical Sciences, Queensland University of Technology, Brisbane, Australia

[c] Tissue Repair and Regeneration Program, Institute of Health and Biomedical Innovation (IHBI), Queensland University of Technology, Brisbane, Australia

ARTICLE INFO

Article history:
Received 26 May 2011
Received in revised form 29 December 2011
Available online 11 January 2012

Keywords:
Cell motility
Cell shape
Exclusion process
Travelling waves
Cellular automata

ABSTRACT

Continuum, partial differential equation models are often used to describe the collective motion of cell populations, with various types of motility represented by the choice of diffusion coefficient, and cell proliferation captured by the source terms. Previously, the choice of diffusion coefficient has been largely arbitrary, with the decision to choose a particular linear or nonlinear form generally based on calibration arguments rather than making any physical connection with the underlying individual-level properties of the cell motility mechanism. In this work we provide a new link between individual-level models, which account for important cell properties such as varying cell shape and volume exclusion, and population-level partial differential equation models. We work in an exclusion process framework, considering aligned, elongated cells that may occupy more than one lattice site, in order to represent populations of agents with different sizes. Three different idealisations of the individual-level mechanism are proposed, and these are connected to three different partial differential equations, each with a different diffusion coefficient; one linear, one nonlinear and degenerate and one nonlinear and nondegenerate. We test the ability of these three models to predict the population-level response of a cell spreading problem for both proliferative and nonproliferative cases. We also explore the potential of our models to predict long time travelling wave invasion rates and extend our results to two-dimensional spreading and invasion. Our results show that each model can accurately predict density data for nonproliferative systems, but that only one does so for proliferative systems. Hence great care must be taken to predict density data with varying cell shape.

1. Introduction

Mathematical models of collective motion have a wide range of applications including cell motility [1], animal swarms [2] and pedestrian traffic [3]. Typically, models of collective motion are based on a macroscopic mean-field approach using partial differential equations (PDEs), or a microscopic approach using discrete random walk processes. For these problems it is relevant to ask how a particular microscopic model relates to a particular macroscopic description of the system. Having the ability to represent a particular collective motility process from both the macroscopic and microscopic points of view is extremely important, especially for applications in cell biology where experimental data is often collected at both the macroscopic and microscopic scales [4,5]. Under these conditions, mathematical models ought to replicate and predict observations at all relevant scales.

* Corresponding author. Tel.: +44 0 1865 283893; fax: +44 0 1865 283882.
E-mail addresses: ruth.baker@maths.ox.ac.uk (R.E. Baker), matthew.simpson@qut.edu.au (M.J. Simpson).

0378-4371/$ – see front matter
doi:10.1016/j.physa.2012.01.009

Discrete random walk models of collective cell motion are often formulated using an exclusion process where each cell is represented by an agent on a lattice, and each lattice site can be occupied by, at most, one single agent. Movement of agents is generally represented by a nearest neighbour random walk [1,8] and proliferation of agents by the deposition of new agents on a nearest neighbour lattice site [9,10]. Interactions between the agents are naturally encoded into the exclusion process since any attempted motility or proliferation event that would place an agent on an occupied lattice site would be aborted. Most previous applications of exclusion processes to problems from cell biology have represented biological cells as uniformly-sized round agents on a square or hexagonal lattice [8,11]. Traditionally the lattice spacing, Δ, is chosen to coincide with the average cell diameter so that each lattice site can be occupied by a single agent. This approach is well-suited to simulating round cells, however for more complicated cell shapes, such as rod-shaped or elongated cells, a different approach is required [12]. We note that cells are often rod-shaped and elongated. For example, the populations of cells in Fig. 1(a)–(b) show two different scrape wound assays where the cells are rod-shaped and the ratio of the longitudinal length scale to the transverse length scale is approximately four.

To motivate our work, we first consider a population of idealised rod-shaped agents, aligned along their longest side, such as the population shown in Fig. 1(c), where the longitudinal length scale of each agent is twice the transverse length scale of each agent. To describe the collective motion of these agents using a regular square lattice, we choose a lattice with lattice spacing Δ that coincides with the smallest length scale of each agent. In this case, each agent occupies two horizontally-adjacent lattice sites. For example, the central agent in row j in Fig. 1(c) occupies sites (i, j) and $(i + 1, j)$. Although we are motivating our work by considering a population of agents where the ratio of the horizontal length scale to the vertical length scale is two, we will later consider the more general case with a system of agents where the ratio of the horizontal length scale to the vertical length scale is $L \geq 1$ which would be more appropriate for the experimental images shown in Fig. 1(a)–(b). In general we will refer to the shape of the agents relative to the lattice spacing. For example, the physical dimensions of the agents in Fig. 1(c) are $2\Delta \times \Delta$, and we will refer to the these agents as having size 2×1 or length 2. In general we will deal with populations of agents with physical dimensions of $L\Delta \times \Delta$, and we will refer to the these agents as having size $L \times 1$ or length L.

We first consider allowing the system depicted in Fig. 1(c) to evolve according to an unbiased simple exclusion process without proliferation. With time uniformly discretised into increments of duration δt, each discrete time step advances the system from time t to time $t+\delta t$. We begin by assuming that an isolated agent attempts to step a distance Δ with probability P_m per increment δt and that the direction of motion is unbiased. At first, for simplicity, we will analyse a one-dimensional motility mechanism where motility events only take place in the horizontal direction. Later we will relax this assumption and consider two-dimensional motion. For one-dimensional motion, the central agent in row j of Fig. 1(c) will attempt to step left or right with probability $P_m/2$ during a time step. If this particular agent attempts to step left this event will be allowed and the agent would then occupy sites $(i - 1, j)$ and (i, j) at the end of the motility event. If this particular agent attempts to step right then this motility event will be aborted since there is another agent already occupying sites $(i + 2, j)$ and $(i + 3, j)$. This particular example highlights the effects of exclusion since the presence of any agent on a lattice site will prevent all other agents in the system from occupying that same lattice site at the same time.

The focus of our work is to construct an appropriate mean-field description of certain exclusion process-based random walk models that describe the movement and proliferation of a population of elongated agents. Traditionally, in the case where $L = 1$ and each agent occupies a single lattice site, conservation arguments are applied to describe the time evolution of the average occupancy of each lattice site [8,9,13,14]. The average occupancy is obtained by considering a sufficiently large number of identically prepared realisations of the microscopic process and taking an average across these different realisations. By considering appropriate limits as $\delta t \to 0$ and $\Delta \to 0$ simultaneously it is possible to show that the average density profile will evolve according to a PDE description [8,9,13,14]. For example, Liggett showed that a one-dimensional unbiased exclusion process is related to the linear diffusion equation whereas a one-dimensional biased exclusion process is related to Burgers' equation [15]. Similarly, Deroulers showed that a two-dimensional exclusion process with agent-to-agent contact effects is related to a nonlinear diffusion equation [8], whereas Simpson showed that a two-dimensional exclusion process with agent proliferation is related to a reaction–diffusion equation that is a generalisation of Fisher's equation [9]. All of these previous studies have only considered the simplest possible case where the agents are round ($L = 1$) and each agent occupies a single lattice site. In this work we will explore several options for constructing continuum models of elongated agent movement and proliferation such as those shown in the schematic in Fig. 1(c).

Having the ability to propose a realistic microscopic model of cell behaviour and to determine the governing macroscopic PDE description of that collective motion is very important in applications relating to cell biology. Several previous modelling studies highlight the importance of taking a multiscale approach. For example, Sherratt and Murray [16] modelled a set of wound closure experiments using two different PDE models. One model involved linear diffusion and the other model involved nonlinear diffusion. Sherratt and Murray [16] showed that both models could replicate their experimental observations despite the two models being very different. This outcome meant that it was impossible to uniquely identify which of these PDE models was best able to capture the underlying physics of the wound healing process. Similarly, Sengers et al. [17] modelled a set of cell invasion assays using two different skeletal cell types: HBMSC and MG63 cells. Detailed experimental data describing the evolution of cell density profiles was collected, and the solution of a reaction–diffusion equation was fitted to the experimental data. Sengers found that the cell density profiles of the MG63 population were best described by a degenerate nonlinear diffusion mechanism whereas the cell density profiles of the HBMSC population were best described by a linear diffusion mechanism. Sengers' work did not explain why one population of cells was apparently

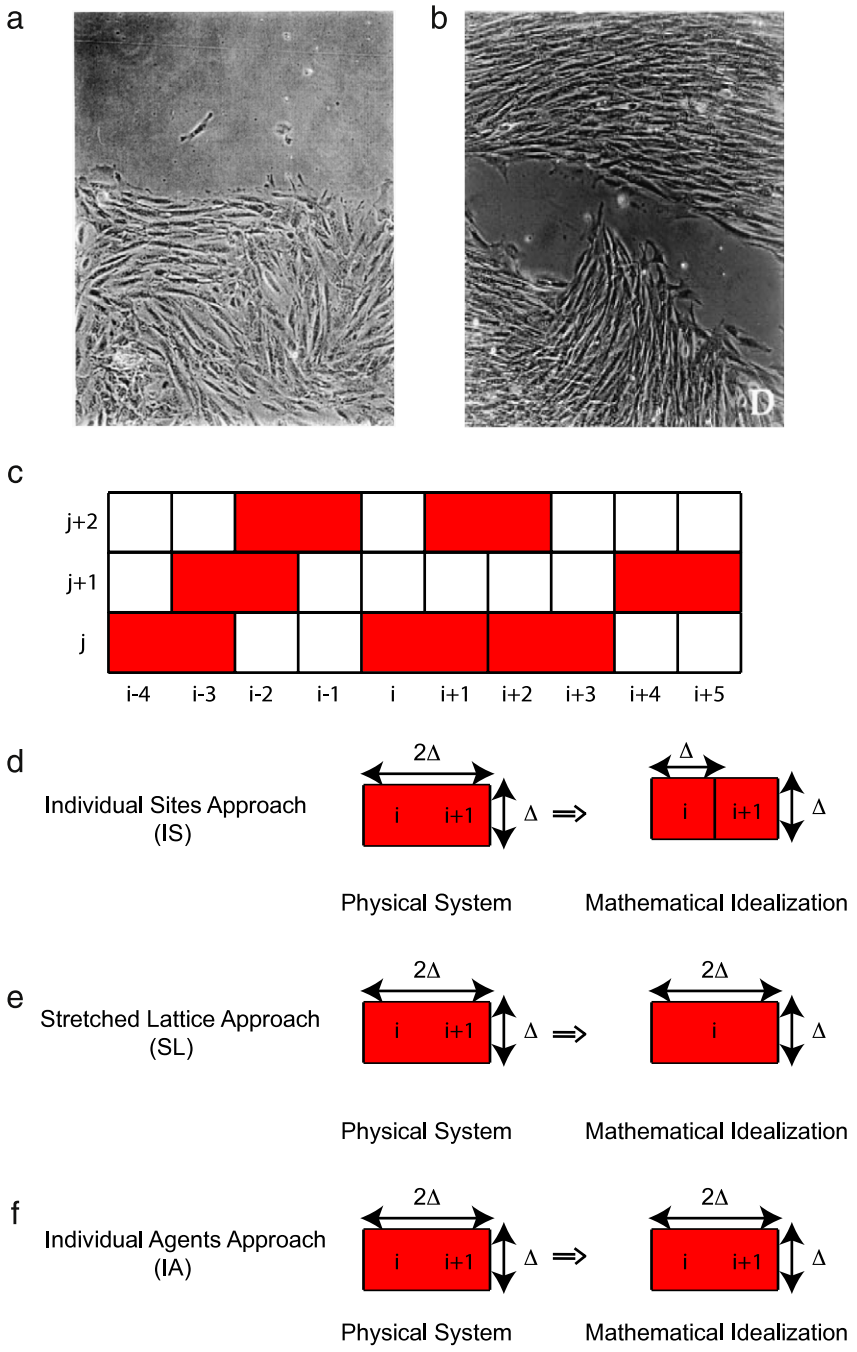

Fig. 1. (Color online). (a) An image from an *in vitro* wound healing scratch assay showing a population of human peritoneal mesothelial cells. During the assay cells are motile and proliferative and the leading edge of the population advances to close the wound space. Full details of this experiment are described in Ref. [6]. Reprinted from Appl. Math. Lett., Maini PK et al., Travelling waves in a wound healing assay, 17, 575–580 (2004) with permission from Elsevier. (b) An image from an *in vitro* wound healing scratch assay showing a population of myoblasts. During the assay cells are motile and proliferative and the leading edge of the population advances to close the wound space. Full details of this experiment are described by Huttenlocher et al. [7]. Reprinted from J. Cell Biol., Huttenlocher et al., Integrin and cadherin synergy regulates contact inhibition of migration and motile activity, 141, 515–526 (1998) with permission from Rockefeller University Press. In both (a) and (b), the longitudinal length scale of the cells is approximately four times the transverse length scale giving $L = 4$. (c) A portion of a lattice that is populated with uniformly-aligned agents, each with $L = 2$. (d) To derive the PDE model for the IS approach, we idealise the occupancy of L adjacent sites by considering the product of L independent probability distribution functions. (e) To derive the PDE model for the SL approach, we approximate the discrete mechanism on a stretched lattice where the horizontal lattice spacing is $L\Delta$. (f) To derive the PDE models for the IA approach, we introduce a single probability distribution function that represents the position and size of each agent.

governed by a linear diffusion mechanism whereas the other population was governed by a nonlinear diffusion mechanism. These two previous studies highlight an important difficulty in applying PDE models to describe collective cell motion. It is often straightforward to propose a particular PDE to represent the motion and spreading of a particular cell population, and it may even be possible to select the parameters so that the solution of this PDE can match the experimental system. This process, however, does not guarantee that the chosen PDE model actually describes the processes of interest. To help overcome this limitation, recent multiscale approaches have used experimental observations to motivate discrete random walk models of cell behaviour which are thought to best describe the system of interest [8,9,13,14]. The particular random walk model can then be converted into a mean-field description using ensemble averaging arguments. In this new work we show that the process of deriving a PDE model from a particular random walk for elongated agents ($L > 1$) is far more challenging than previous studies where only round agents ($L = 1$) have been considered.

In summary, our work demonstrates that there are at least three physically reasonable approaches to derive a mean-field model describing the movement and proliferation of elongated, volume-excluding agents in the context of an exclusion process. The first approach (Fig. 1(d)) is an extension of our previous work [12] where we suppose that the occupancy of each lattice site is completely independent of all other lattice sites. This assumption leads us to represent the presence of an agent with $L = 2$ that occupies sites (i, j) and $(i+1, j)$ as the product of two probability distribution functions $P(A_{i,j}, t)P(A_{i+1,j}, t)$. Here, $P(A_{i,j}, t)$ is the probability that site (i, j) will be occupied by an agent (or part of an agent) at time t. The second approach (Fig. 1(e)) uses an anisotropic lattice to accommodate the elongated agents. For $L = 2$ we simply stretch the horizontal lattice spacing to be 2Δ while maintaining the vertical spacing as Δ. Thirdly, we consider representing the presence of an agent of length two at sites (i, j) and $(i+1, j)$ by a single univariate distribution function, $P^2(A_{i,j}, t)$, which represents the probability of finding an agent of length two occupying sites (i, j) and $(i + 1, j)$ at time t. The superscript indicates the length of the agent.

Using these three approaches, in the first part of this work we construct conservation arguments that incorporate the relevant exclusion effects accounting for the location and length of that particular agent. In the absence of proliferation, we show that these three approaches lead to three different diffusion equations describing the distribution of averaged agent densities. Comparing simulation data with the numerical solution of these diffusion equations shows that all three approaches are able to predict the averaged simulation data reasonably well. This is a surprising result: the three PDE models are very different yet the solution of each equation provides us with a reasonable match to the simulation data. We explain this counterintuitive observation by showing that the mean action times for these three different diffusive processes are identical [18–20].

In the second part of this work we extend our analysis to include agent proliferation so that our discrete models can be applied to cell invasion problems. Although the three diffusion equations derived in the first part of the manuscript can describe averaged simulation data without proliferation, we show that this observation no longer holds when we introduce proliferation. Each approach to deriving the governing PDE gives a different reaction–diffusion equation, and comparisons between numerical solutions of these PDEs and averaged simulation data show that only the third approach (Fig. 1(f)) is able to provide an accurate description of the system. The reaction–diffusion equations derived here are consistent with previous studies of "round" agents ($L = 1$) since the source term in each case relaxes to a logistic model in this case [10]. The agreement between the averaged discrete results and numerical solution of the reaction–diffusion PDE decreases as the proliferation rate P_p increases [10]. We have recently shown that this is due to the failure of independence assumptions underlying the discrete conservation statements that lead to the PDE models [21,22].

Finally, several extensions, such as the consideration of long time travelling wave behaviour and two-dimensional motility and proliferation processes, are presented and discussed.

2. Simulation and analysis of collective motion in one-dimension

To illustrate the influence of varying the shape of agents in the simplest possible way, we first consider a one-dimensional lattice that is occupied by a population of uniformly-aligned agents where each agent has the same shape, $L \times 1$. We suppose that all agents are parallel to the x axis [12] such as the schematic shown in Fig. 1(c) where $L = 2$. In our system all the agents are motile and each agent is able to move in the x direction only.

2.1. Discrete simulations

The motility of a population of N agents is simulated as follows: during each time step of duration δt, N agents are selected independently at random, one at a time, and given the opportunity to move. When chosen, an agent attempts a motility event with probability P_m, where $P_m \in [0, 1]$. We interpret P_m as the probability that an agent will attempt to move a distance Δ in the time interval δt. If, during the attempted motility event, any target site is occupied by any agent other than the agent attempting to move, then that motility event is aborted (see Ref. [12] for more details).

Using this motility mechanism, with a population of horizontally-aligned agents each with a fixed length, L, we performed many simulations on a one-dimensional lattice. Each site is indexed i, where $i \in \mathbb{Z}^+$, and each site has position $x_i = \Delta i$. Simulations were performed on a lattice with $1 \leq i \leq 1000$ with reflecting boundary conditions imposed at $i = 1$ and $i = 1000$. This means, for example, that if a cell has its left-most end at site 1, then any attempted move to the left will

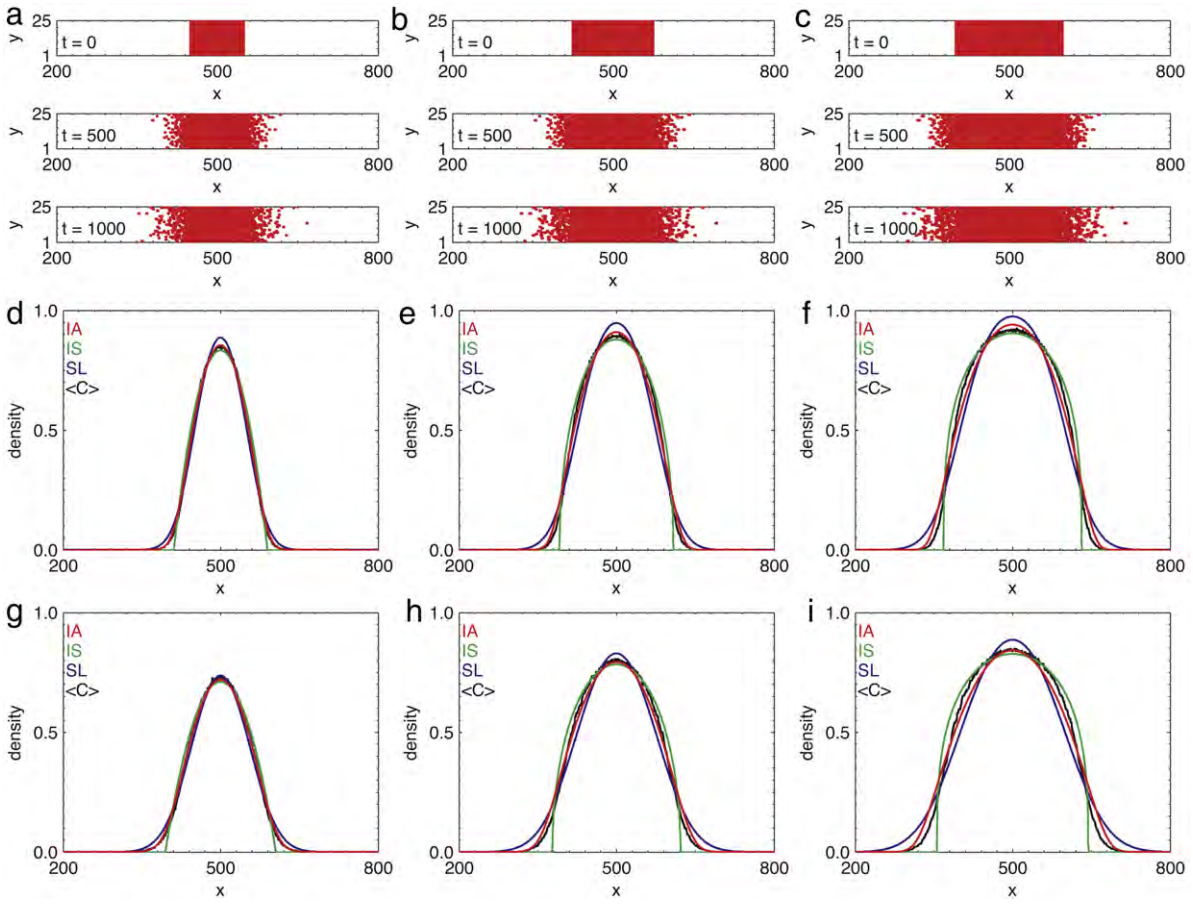

Fig. 2. (Color online). Without proliferation: a series of one-dimensional simulations comparing the effects of different diffusion coefficients with the discrete data for $P_m = 1.0$ and $P_p = 0.0$. (a) Simulations for agents with $L = 2$. Initially all sites with $450 \leq i < 550$ are occupied by 50 adjacent columns of non-overlapping, horizontally-aligned agents. (b) Simulations for agents with $L = 3$ and initially 50 columns of adjacent agents filling sites $425 \leq i < 575$. (c) Simulations for agents with $L = 4$ and initially 50 columns of adjacent agents filling sites $400 \leq i < 600$. In each case, snapshots are shown at $t = 0, 500, 1000$ for $\delta t = \Delta = 1.0$. Results in (d)–(f) and (g)–(i) show averaged agent density (black lines) for a series of corresponding simulations, similar to those shown in (a)–(c), with the data averaged over $M = 1000$ realisations at times $t = 500$ ((d)–(f)) and $t = 1000$ ((g)–(i)). These averaged discrete profiles are superimposed on the numerical solutions of Eqs. (5), (11) and (20) shown in red, blue and green, respectively. The PDEs are solved numerically using the method outlined in the main text with $\Delta x = 0.1 - 0.5$ and $\Delta t = 0.001 - 0.1$ and $\epsilon = 10^{-6}$. The root mean square error in each case is given in Appendix A.

be aborted. Similarly, for a cell with its right-most end at site 1000 any attempted jumps to the right will be aborted. In all simulations shown in this work $\delta t = \Delta = 1$. The central part of this lattice is shown in Fig. 2(a)–(c), where we consider three different simulations and each simulation corresponds to a population with different sized agents, $L = 2, 3, 4$, respectively. To visualise the simulation data we present 25 realisations of the model in the 25 rows of the lattices shown in Fig. 2(a)–(c). The snapshots in each row of Fig. 2(a)–(c) are independent realisations of the same one-dimensional stochastic process. Presenting these snapshots side-by-side conveniently illustrates the stochastic nature of the random walk algorithm since the distribution of agents in each row of the lattice can be quite different. In each simulation, the central lattice sites were initially occupied so that the initial distribution of agents was composed of several columns of adjacent, non-overlapping agents. For simulations with different sized agents, L, we always initiated each simulation with the same number of agents. The system was allowed to evolve and we observe snapshots of the spreading populations in Fig. 2(a)–(c) at $t = 0$, 500, 1000, respectively. To complement these single snapshots, we also generated averaged agent density data in Fig. 2(d)–(i) for the same problems at $t = 500$, 1000. This averaged data, denoted $\langle C(x_i, t)\rangle$, was obtained by repeating the simulations in Fig. 2, averaging across $M = 1000$ identically prepared realisations of the same one-dimensional problem [9,13,23]:

$$\langle C(x_i, t)\rangle = \frac{1}{M}\sum_{m=1}^{M} c^m(i, t), \tag{1}$$

where $c^m(i, t)$ is the occupancy of lattice site i at time t during the mth identically prepared realisation. In any single realisation $c^m(i) = 0$ if site i is unoccupied, while $c^m(i) = 1$ if site i is occupied. The data in Fig. 2 shows how the density

profile evolves as the agents spread away from their initial closely-packed distribution. We will now derive three PDE models that can replicate and predict the spreading density profiles for these simulations.

2.2. Mean-field models describing one-dimensional collective motility

To connect the discrete mechanism with a PDE model we form a discrete conservation statement describing the change in average occupancy of site i (or the change in average occupancy of groups of sites, $[i, i+1, i+2, \ldots, i+L-1]$) during the time interval from t to $t+\delta t$. To match the one-dimensional density profiles in Fig. 2 we begin by considering a population of horizontally-aligned agents each with the same size, L, that move only in the horizontal direction. Later we will generalise these arguments to two-dimensional motion.

2.2.1. Individual sites

We first suppose that the average occupancy of lattice site i at time t is given by $P(A_i, t) \in [0, 1]$, and that the averaged occupancies of all lattice sites are independent [12]. This approach, outlined schematically in Fig. 1(d), amounts to constructing a conservation argument by considering individual lattice sites, and so we refer to this approach as the *individual sites* (IS) approach. Using these assumptions we can write down a conservation statement describing the change in average occupancy of a general lattice site during any time step. To illustrate this conservation statement we begin by considering the case where $L = 2$:

$$P(A_i, t+\delta t) - P(A_i, t) = \frac{P_m}{2} P(A_{i-2}, t) P(A_{i-1}, t)\,(1 - P(A_i, t)) + \frac{P_m}{2} P(A_{i+1}, t) P(A_{i+2}, t)\,(1 - P(A_i, t)) - \frac{P_m}{2} P(A_i, t) P(A_{i-1}, t)\,(1 - P(A_{i-2}, t)) - \frac{P_m}{2} P(A_i, t) P(A_{i+1}, t)\,(1 - P(A_{i+2}, t))\,, \tag{2}$$

where $P(A_i, t)$ is the probability of site i being occupied at time t. Positive terms on the right of Eq. (2) represent motility events that increase the average occupancy of site i while negative terms on the right of Eq. (2) represent motility events that decrease the average occupancy of site i. Each of these terms has a physical interpretation. For example, the first term on the right of Eq. (2) accounts for the case where we have an agent occupying sites $i-2$ and $i-1$ and this agent attempts to move in the positive x direction. To account for this motility event, the term is proportional to (i) $P_m/2$, which is the probability than an isolated agent attempts to step in the positive x direction during the time interval δt, (ii) $P(A_{i-2}, t) P(A_{i-1}, t)$ which is the probability that sites $i-2$ and $i-1$ are occupied, and (iii) $(1 - P(A_i, t))$ which is the probability that the target site, i, is vacant. We interpret the product of these three probabilities as a net transition probability that governs the averaged behaviour of the system. All other terms on the right of Eq. (2) can be interpreted in a similar way.

As we have shown previously [12], the discrete conservation statement is related to a PDE model in the appropriate limit as $\Delta \to 0$ and $\delta t \to 0$ simultaneously, and discrete values of $P(A_i, t)$ are written in terms of a continuous variable $C(x, t)$. To see this relationship, all terms in Eq. (2) and are expanded in a Taylor series about site i. The Taylor series are truncated by neglecting all terms of $\mathcal{O}(\Delta^3)$ and smaller, and we then take the limit as $\Delta \to 0$ and $\delta t \to 0$ simultaneously to give rise to the PDE

$$\frac{\partial C}{\partial t} = D_0 \frac{\partial}{\partial x}\left(4C \frac{\partial C}{\partial x}\right), \quad D_0 = \lim_{\Delta \to 0, \delta t \to 0}\left(\frac{\Delta^2 P_m}{2\delta t}\right). \tag{3}$$

This PDE is a degenerate nonlinear diffusion equation with a nonlinear diffusivity function $D(C) = 4C$. This particular form of the nonlinear diffusivity function has important consequences for the solution of this PDE model and we will elaborate on this detail later when we compare solutions of this PDE with simulation data.

It is possible to generalise the conservation arguments given in Eq. (2) to describe the conservation of occupancy of any site on a lattice that supports the movement of a population of agents with any integer length $L \geq 1$. The general conservation statement is given by

$$P(A_i, t+\delta t) - P(A_i, t) = \frac{P_m}{2} \prod_{s=1}^{L} P(A_{i-s}, t)\,(1 - P(A_i, t)) + \frac{P_m}{2} \prod_{s=1}^{L} P(A_{i+s}, t)\,(1 - P(A_i, t)) - \frac{P_m}{2} \prod_{s=0}^{L-1} P(A_{i-s}, t)\,(1 - P(A_{i-L}, t)) - \frac{P_m}{2} \prod_{s=0}^{L-1} P(A_{i+s}, t)\,(1 - P(A_{i+L}, t))\,, \tag{4}$$

which, in the appropriate limits as $\Delta \to 0$ and $\delta t \to 0$ simultaneously, gives rise to the PDE

$$\frac{\partial C}{\partial t} = D_0 \frac{\partial}{\partial x}\left(D^{\mathrm{IS}}(C) \frac{\partial C}{\partial x}\right), \tag{5}$$

with

$$D^{\mathrm{IS}}(C) = L^2 C^{L-1}, \qquad D_0 = \lim_{\Delta \to 0, \delta t \to 0}\left(\frac{\Delta^2 P_m}{2\delta t}\right). \tag{6}$$

As discussed in Ref. [12], the limiting PDE description is a nonlinear degenerate diffusion equation with $D^{\mathrm{IS}}(0) = 0$ for $L \geq 2$. We remark that, for the simple case with $L = 1$, we obtain $D^{\mathrm{IS}}(C) = 1$ which is the usual linear diffusion equation described previously for round agents [13]. This confirms that our conservation arguments for agents of arbitrary length $L \geq 2$ relax to the previously established result when $L = 1$.

Since the governing PDE model is a degenerate nonlinear diffusion equation for $L \geq 2$ the solution of the PDE can contain sharp interfaces, beyond which the population density is zero [24]. This means that the solution of Eq. (6), evolving from an initial condition with compact support, will always have compact support, with sharp interfaces that separate those regions of the solution where $C(x, t) = 0$ and $C(x, t) > 0$. This feature is of immediate interest in this context since the density profiles we obtained from our averaged simulation data in Fig. 2 do not appear to contain sharp interfaces.

2.2.2. Stretched lattice

Instead of formulating our conservation arguments on a regular square lattice with spacing Δ, it is also possible to consider deriving a conservation statement on an anisotropic lattice so that each anisotropic lattice site can accommodate a single anisotropic agent. To do this we formulate our conservation statement on a lattice where the horizontal lattice spacing is stretched with lattice spacing $L\Delta$. We call this approach the *stretched lattice* (SL) approach. Note that we initially indexed lattice sites by i so that position $x_i = i\Delta$. In the SL model, the index i now refers to being in position $x_i = iL\Delta$. In our simulations, each agent steps a distance Δ with probability P_m per increment δt. To develop a conservation statement on a stretched lattice, we consider the probability of an isolated agent stepping a distance $L\Delta$ per time step δt. One way to do this is to scale the intrinsic motility probability P_m so that the probability of an agent stepping a distance $L\Delta$ per increment δt is $\bar{P}_m$ where $\bar{P}_m = P_m/L$. This ensures that the average displacement per unit time is the same as for an agent which can only move a distance Δ per motility event [25]. For the particular case that $L = 2$, we define $P_2(A_i, t)$ as the probability of finding an agent of length two occupying site i on the stretched lattice. The subscript refers to the length of the agent. With these definitions we can form a discrete conservation statement on the stretched lattice:

$$P_2(A_i, t+\delta t) - P_2(A_i, t) = \frac{\bar{P}_m}{2}\left[P_2(A_{i-1}, t) + P_2(A_{i+1}, t)\right](1 - P_2(A_i, t)) - \frac{\bar{P}_m}{2}\left[(1 - P_2(A_{i-1}, t)) + (1 - P_2(A_{i+1}, t))\right]P_2(A_i, t), \tag{7}$$

where $\bar{P}_m = P_m/2$. Expanding all terms in a truncated Taylor series about the point i, and taking the limit as $\Delta \to 0$ and $\delta t \to 0$ simultaneously, gives the PDE

$$\frac{\partial C}{\partial t} = D_0 \frac{\partial}{\partial x}\left(D^{\mathrm{SL}}(C)\frac{\partial C}{\partial x}\right), \tag{8}$$

$$D^{\mathrm{SL}}(C) = 2, \qquad D_0 = \lim_{\Delta\to 0, \delta t \to 0}\left(\frac{\Delta^2 P_m}{2\delta t}\right), \tag{9}$$

which, for $L = 2$, is a linear diffusion equation where the linear diffusivity is twice the diffusivity for "round" agents with $L = 1$.

The discrete conservation statement and the resulting PDE model can be generalised to account for the collective movement of a population of agents with length L on an appropriately stretched lattice, and is given by

$$P_L(A_i, t+\delta t) - P_L(A_i, t) = \frac{\bar{P}_m}{2}\left[P_L(A_{i-1}, t) + P_L(A_{i+1}, t)\right](1 - P_L(A_i, t)) - \frac{\bar{P}_m}{2}\left[(1 - P_L(A_{i-1}, t)) + (1 - P_L(A_{i+1}, t))\right]P_L(A_i, t), \tag{10}$$

where $P_L(A_i, t)$ represents the probability of having an agent of length L occupying site i on the stretched lattice. The discrete conservation statement can be related to a PDE model in the appropriate limit as $\Delta \to 0$ and $\delta t \to 0$ simultaneously, where we identify $P_L(A_i, t)$ with $C(x, t)$ in this limit. This gives

$$\frac{\partial C}{\partial t} = D_0 \frac{\partial}{\partial x}\left(D^{\mathrm{SL}}(C)\frac{\partial C}{\partial x}\right), \tag{11}$$

with

$$D^{\mathrm{SL}}(C) = L, \qquad D_0 = \lim_{\Delta\to 0, \delta t \to 0}\left(\frac{\Delta^2 P_m}{2\delta t}\right). \tag{12}$$

This PDE model is a linear diffusion equation where the diffusivity changes with L. For the simple case that $L = 1$, we obtain $D^{\mathrm{SL}}(C) = 1$ with $D_0 = \lim_{\Delta\to 0, \delta t\to 0}(\Delta^2 P_m/2\delta t)$, which is the same as the linear equation derived previously for "round" agents, and confirms that our conservation arguments on the stretched lattice are consistent with previously established work when $L = 1$ [13].

A key feature of the linear diffusion equation model is that the solution is smooth without any sharp interfaces, unlike the nonlinear degenerate diffusion model derived by considering the IS approach. This is a promising feature, since our averaged discrete simulations are smooth-fronted.

2.2.3. Individual agents

Our final approach to deriving a PDE model to describe the collective motion of a population of agents is to consider improving how we represent agents with $L \geq 2$ on a regular square lattice with lattice spacing Δ. When we presented the IS approach, we argued that, by assuming average occupancies of lattice sites are independent [12], the presence of an agent at sites $i-1$ and $i-2$ could be represented by the product of two probability distribution functions, $P(A_{i-2}, t)P(A_{i-1}, t)$. The key weakness is that we would like to describe the probability of a single agent residing on the sites $i-1$ and $i-2$. Unfortunately the expression $P(A_{i-2}, t)P(A_{i-1}, t)$ describes the probability of sites $i-1$ and $i-2$ being occupied, but this does not necessarily imply that the same agent is occupying both sites. For example, if we consider the case where there is an agent with $L = 2$ residing on sites $i-3$ and $i-2$ and that this agent is adjacent to another agent with $L = 2$ that occupies sites $i-1$ and i, then our independence assumption means that this configuration is indistinguishable from the case where a single agent resides at sites $i-1$ and $i-2$. More generally, if we have a horizontal agent with $L = 2$ and part of this agent resides on site i, then it must also occupy one of sites $i \pm 1$. This detail was neglected in Section 2.2.1 when we considered deriving the conservation statement for individual sites, and we will now attempt to improve on this assumption by developing an alternative conservation statement that considers the variable $P^L(A_i, t)$, the probability of having an agent of length L with its left-most end occupying site i at time t. This approach is based on considering the location and length of an individual agent and so we call this approach the *individual agents* (IA) approach. We note that the assumption is now that individual agents are independent and we average over the occupancy of sites by the left-most end of an agent. Further, in taking the limit that the lattice spacing goes to zero, the agent size also tends to zero, and so we will identify $P^L(A_i, t)$ with density, $C(x, t)$, in the continuum limit.

For demonstrative purposes, we outline a conservation statement for the case with $L = 2$, which is given by

$$\begin{aligned} P^2(A_i, t+\delta t) - P^2(A_i, t) = {} & \frac{P_m}{2}\left[P(0_i; A_{(i+1,i+2)}, t) + P(A_{(i-1,i)}; 0_{i+1}, t)\right] \\ & - \frac{P_m}{2}\left[P(A_{(i,i+1)}; 0_{i+2}, t) + P(0_{i-1}; A_{(i,i+1)}, t)\right], \end{aligned} \tag{13}$$

where, for example, $P(0_i; A_{(i+1,i+2)}, t)$ represents the probability of site i being vacant and an agent with $L = 2$ occupying sites $i+1$ and $i+2$ at time t. To close the system we use conditional probability arguments on the right-hand side to specify the agent location and enforce exclusion principles. This allows us to write, for example,

$$P(0_i; A_{(i+1,i+2)}, t) = P(0_i, t|A_{(i+1,i+2)}, t)P^2(A_{i+1}, t) = (1 - P^2(A_{i-1}, t))P^2(A_{i+1}, t). \tag{14}$$

The second line enforces the relevant agent exclusion principle for this particular sized agent: if an agent with $L = 2$ is situated such that it occupies sites $i+1$ and $i+2$, it is impossible for another agent to reside any closer to the original agent other than occupying sites $i-1$ and i. This detail was not accounted for in the IS approach. By closing all terms on the right of Eq. (13) using conditional probability statements we obtain

$$\begin{aligned} P^2(A_i, t+\delta t) - P^2(A_i, t) = {} & \frac{P_m}{2}\left[(1 - P^2(A_{i-1}, t))P^2(A_{i+1}, t) + (1 - P^2(A_{i+1}, t))P^2(A_{i-1}, t)\right] \\ & - \frac{P_m}{2}\left[(1 - P^2(A_{i+2}, t))P^2(A_i, t) + (1 - P^2(A_{i-2}, t))P^2(A_i, t)\right]. \end{aligned} \tag{15}$$

To arrive at the PDE model, all terms in Eq. (15) are expanded in a truncated Taylor series about site i and we consider the limit as $\Delta \to 0$ and $\delta t \to 0$ simultaneously to give

$$\frac{\partial C}{\partial t} = D_0 \frac{\partial}{\partial x}\left((1 + 2C)\frac{\partial C}{\partial x}\right), \quad D_0 = \lim_{\Delta \to 0, \delta t \to 0}\left(\frac{\Delta^2 P_m}{2\delta t}\right). \tag{16}$$

This PDE model is nonlinear, but not degenerate since $D(0) = 1$.

Arguments for the specific case where $L = 2$ can be generalised for arbitrary L as follows,

$$\begin{aligned} P^L(A_i, t+\delta t) - P^L(A_i, t) = {} & \frac{P_m}{2}\left[P(0_i; A_{(i+1,\ldots,i+L)}, t) + P(A_{(i-1,\ldots,i+L-2)}; 0_{i+L-1}, t)\right] \\ & - \frac{P_m}{2}\left[P(A_{(i,\ldots,i+L-1)}; 0_{i+L}, t) + P(0_{i-1}; A_{(i,\ldots,i+L-1)}, t)\right], \end{aligned} \tag{17}$$

where, for example, $P(0_i; A_{(i+1,\ldots,i+L)}, t)$ represents the probability of site i being vacant and an agent of length L occupying sites $i+1, \ldots, L$ at time t. Again, we close the system by using conditional probabilities on the right-hand side to account for the agent shape and associated exclusion details. For example, we now have

$$P(0_i; A_{(i+1,\ldots,i+L)}) = P(0_i, t|A_{(i+1,\ldots,i+L)}, t)P^L(A_{i+1}, t) = (1 - P^L(A_{i-L+1}, t))P^L(A_{i+1}, t). \tag{18}$$

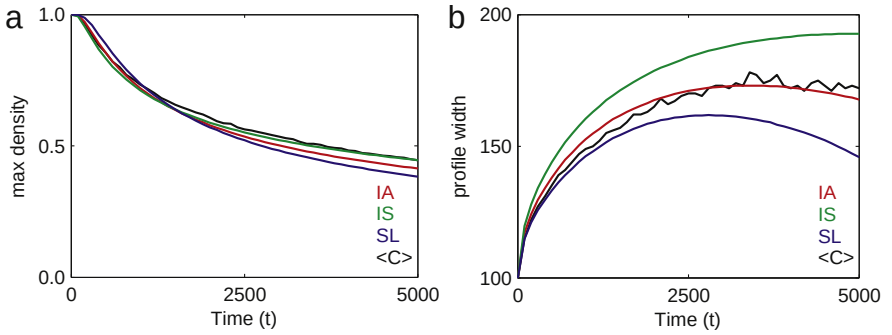

Fig. 3. (Color online). Without proliferation: the change in (a) density profile maximum and (b) width over time with $L = 2$ and the same initial and boundary conditions as in Fig. 2. The width of the region for which $C > 0.3$ is used as a measure of profile width, and all other details are as in Fig. 2.

Using these conditional probability statements, Eq. (17) then gives

$$P^L(A_i, t+\delta t) - P^L(A_i, t) = \frac{P_m}{2}\left[(1 - P^L(A_{i-L+1}, t))P^L(A_{i+1}, t) + (1 - P^L(A_{i+L-1}, t))P^L(A_{i-1}, t)\right] - \frac{P_m}{2}\left[(1 - P^L(A_{i+L}, t))P^L(A_i, t) + (1 - P^L(A_{i-L}, t))P^L(A_i, t)\right]. \tag{19}$$

This discrete conservation statement can be related to a PDE model in the appropriate limit as $\Delta \to 0$ and $\delta t \to 0$ simultaneously where we now recognise $P^L(A_i, t)$ as $C(x, t)$ in this limit. The resulting PDE is

$$\frac{\partial C}{\partial t} = D_0 \frac{\partial}{\partial x}\left(D^{\mathrm{IA}}(C)\frac{\partial C}{\partial x}\right), \tag{20}$$

with

$$D^{\mathrm{IA}}(C) = 1 + 2(L-1)C, \qquad D_0 = \lim_{\Delta \to 0}\left(\frac{\Delta^2 P_m}{2\delta t}\right). \tag{21}$$

Thus, under these assumptions we arrive at a nonlinear diffusion equation. However, this nonlinear diffusion equation is nondegenerate as $D^{\mathrm{IA}}(0) = 1$ which means that the solution of this equation will be smooth. For the simple case that $L = 1$ we obtain $D^{\mathrm{IA}}(C) = 1$, which is the same as the usual linear equation derived previously for round agents and confirms that our conservation arguments based on using conditional probabilities to account for detailed agent shape and exclusion processes relax to the previously established PDE model when $L = 1$.

2.2.4. Comparing discrete simulation data with the mean-field models

To compare the averaged density profiles with the solution of the mean-field models, we must solve Eqs. (5)–(6), (11)–(12) and (20)–(21) using the same domain, boundary conditions and initial condition as in the discrete simulations given in Fig. 2. We solve the PDEs numerically using a finite difference approximation with constant grid spacing δx and implicit Euler time stepping with constant time steps of duration h. Picard linearisation, with absolute error tolerance ϵ, is used to solve the resulting nonlinear algebraic systems [26]. We note that the reflecting boundary conditions of the discrete simulations correspond with zero flux boundary conditions in our PDE models.

The numerical solutions of the mean-field models for the IS, SL and IA approaches are superimposed on the averaged density profiles obtained from the discrete simulations in Fig. 2, and we compare results for $L = 2, 3, 4$. Several comments can be made about the comparison between the simulation data and the solutions of the PDE models. First, all three mean-field models provide a reasonable match to the simulation data since all three models are able to capture the basic shape of the spreading density profile and the spatial extent of spreading at short times. Fig. 3 shows how each of the mean-field models predicts the maximum density and width of the spreading population over time. We quantified the error between each mean-field model and the averaged discrete results by calculating the root mean square error (RMSE) and the results are presented in Appendix A. We note that at long times the quality of this match decreases (see Fig. 3) but the more agents on the lattice, the longer the mean-field approximation holds (results not shown). Second, we observe minor differences in the quality of the discrete-continuum match right at the leading edge of the density profiles. For example, the solution of the degenerate nonlinear PDE associated with the IS approach leads to sharp-fronted profiles [24] that always slightly underestimate the spatial extent of the spreading. Conversely, the SL approach predicts a smooth profile that slightly overestimates the spatial extent of the spreading while the IA approach gives the closest match to the simulation density profiles and is able to capture the shape and location of the leading edge of the population density quite accurately.

In general, we see that the quality of the discrete-continuum match becomes poorer as L increases. This trend is observed for all three mean-field models and can be attributed to the decreasing validity of our assumptions as L increases. For

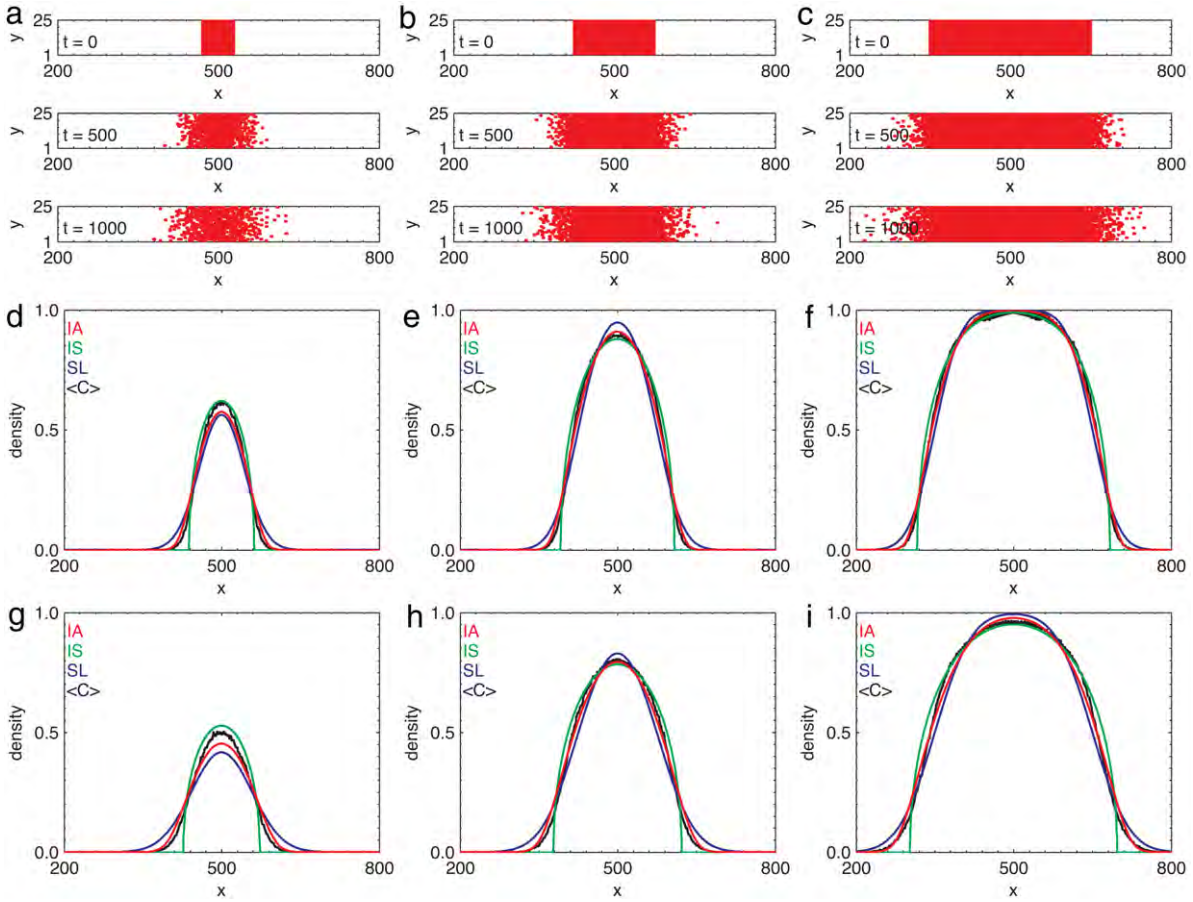

Fig. 4. (Color online). Without proliferation: a series of one-dimensional simulations comparing different diffusion coefficients with the discrete data for $P_m = 1.0$, $P_p = 0.0$ and different numbers of discrete agents with $L = 3$ and $\delta t = \Delta = 1.0$. (a) Initially all sites with $470 \leq i < 530$ were occupied by 20 adjacent columns of non-overlapping, horizontally-aligned agents. (b) Initially all sites with $425 \leq i < 575$ were occupied by 50 adjacent columns of non-overlapping, horizontally-aligned agents. (b) Initially all sites with $350 \leq i < 650$ were occupied by 100 adjacent columns of non-overlapping, horizontally-aligned agents. Results in (d)–(f) and (g)–(i) show averaged agent density (black lines) for a series of corresponding simulations similar to those shown in (a)–(c), with the data averaged over $M = 1000$ realisations at times $t = 500$ ((d)–(f)) and $t = 1000$ ((g)–(i)). These averaged discrete profiles are compared with the solutions of Eqs. (5), (20) and (11) shown in red, blue and green, respectively. The PDEs are solved numerically using the method outlined in the main text with $\Delta x = 0.1 - 0.5$ and $\Delta t = 0.001 - 0.1$ and $\epsilon = 10^{-6}$. The root mean square error in each case is given in Appendix A.

example, in the IS model the assumption that occupancies of individual sites are independent becomes less valid as L increases. To investigate the robustness of the discrete-continuum comparisons in Fig. 2 where we studied simulations that had a fixed number of agents but we varied the agent size, L, we now present a complementary set of discrete-continuum comparisons in Fig. 4 where we consider discrete simulations for three systems each with a fixed agent shape ($L = 3$) but with a varying number of agents. In general we see that the quality of the discrete-continuum comparison increases as we consider systems with increasing numbers of agents. Similar to the results shown in Fig. 2, we observe that all three mean-field models are able to capture the density profiles reasonably well since all three capture the shape of the spreading density profile and the spatial extent of spreading. We also observe some discrepancy between the simulation data and the PDE solutions at the leading edge of the spreading population where the IS model underestimates the extent of spreading, the SL model overestimates the extent of spreading and the IA approach gives the best match.

2.2.5. Mean action time

We now briefly discuss some mathematical features of the three different diffusion PDE models that we derived in Section 2.2. The three approaches for constructing the mean-field descriptions of the system involve making certain assumptions about how we represent the averaged behaviour of the system. Intuitively we expect that there should be some equivalence between the mean-field models since they all aim to describe the same behaviour. At first glance, however, the three mean-field models given by Eqs. (5)–(6), (11)–(12) and (20)–(21) are very different. For example, Eqs. (5)–(6) is a nonlinear degenerate diffusion equation, Eqs. (11)–(12) is a linear diffusion equation and Eqs. (20)–(21) is a nonlinear nondegenerate diffusion equation. At first there appears to be no relationship between these three different diffusion equations. However, when we compare the solutions of these equations to the discrete density data we see that the three

diffusion models are all able to describe the basic features of the collective motion. This is reassuring since all three diffusion models attempt to describe the same system. Moreover, we note that there is a formal mathematical connection between these three diffusion models which gives a rigorous explanation as to why these seemingly different diffusion equations produce solutions that look very similar.

The mean action time (MAT) is a measure of the time required for a disturbance at the boundary of a diffusive problem to reach a particular observation point. The theory behind the MAT was developed by McNabb and coworkers [18,19] to quantify the behaviour of nonlinear diffusion problems, which may or may not contain a moving front, in various geometries. For a general nonlinear diffusion process, described by

$$\frac{\partial C}{\partial t} = D_0 \nabla \cdot [D(C) \nabla C], \tag{22}$$

the MAT, $T(\mathbf{r})$, can be written as

$$T(\mathbf{r}) = \frac{\int_0^\infty t \frac{\partial}{\partial t} \left(\int_{C_0}^{C(\mathbf{r},t)} D(C)\, \mathrm{d}C \right) \mathrm{d}t}{\int_0^\infty \frac{\partial}{\partial t} \left(\int_{C_0}^{C(\mathbf{r},t)} D(C)\, \mathrm{d}C \right) \mathrm{d}t}, \tag{23}$$

where $\mathbf{r}$ is the coordinate of the fixed observation point. For certain linear diffusion problems $T(\mathbf{r})$ satisfies a Poisson equation [20]. Furthermore, it has been established that two diffusion problems, with the same boundary conditions and initial condition, will have the same MAT provided that the quantity

$$\int_{C_0}^{C_\infty} D(C)\, \mathrm{d}C, \tag{24}$$

is the same in both problems. Here it is assumed that the system is finite and initially uniform with $C(\mathbf{r}, 0) \equiv C_0$. The value of C is then changed at an outer boundary at time $t = 0$ such that $C(\mathbf{r}, t)$ tends towards a new equilibrium value, $\lim_{t\to\infty} C(\mathbf{r}, t) = C_\infty$. These features of the MAT have been exploited to provide insight into many applications including filtration problems and modelling water uptake during cereal production [20]. For example, Landman and collaborators show that pressure filtration of flocculated suspensions can be modelled using a one-dimensional nonlinear diffusion equation [20,27]. Due to the nonlinearity in the governing equation it was impossible to obtain any analytical insight into the solution of the problem. This difficulty was overcome by using the concept of MAT to replace the governing nonlinear equation with a simpler equation that permitted analysis [20,27]. The key assumption used by Landman was that two diffusion problems are similar provided the MAT is the same for both problems.

By considering the MAT for our three diffusion models of collective cell migration we can see that the quantity

$$\int_0^1 D(C)\, \mathrm{d}C, \tag{25}$$

with $D(C) = D_0 D^{\mathrm{IS}}(C)$, $D(C) = D_0 D^{\mathrm{SL}}(C)$ or $D(C) = D_0 D^{\mathrm{IA}}(C)$, is the same for all three approaches regardless of the value of L:

$$\int_0^1 D_0 D^{\mathrm{IS}}(C)\, \mathrm{d}C = \int_0^1 D_0 D^{\mathrm{SL}}(C)\, \mathrm{d}C = \int_0^1 D_0 D^{\mathrm{IA}}(C)\, \mathrm{d}C = D_0 L. \tag{26}$$

This gives a mathematical explanation for our observations. Although our three diffusion equations appear to be very different they have the same MAT for any choice of L. This explains why the solutions of the different diffusion equations are similar, each equation providing a good approximation to the simulation data, comparable except right at the leading edge of the spreading population.

This work is the first to make a connection between the exclusion process literature and the theory of MAT. This is an important connection to make since many research groups are currently working to derive mean-field PDE descriptions for many different kinds of random walk models based on exclusion processes [1,8,12]. Given that MAT has been used previously to provide powerful insight into various applied transport processes, we believe that MAT will also provide new and important insight into the relationship between exclusion processes and mean-field PDE models in the future.

3. Simulation and analysis of collective motion and proliferation in one-dimension

We now generalise our analysis to include both motility events and proliferation events in the discrete model. Proliferation events will occur with probability $P_p \in [0, 1]$ per time step δt and will attempt to place a daughter agent

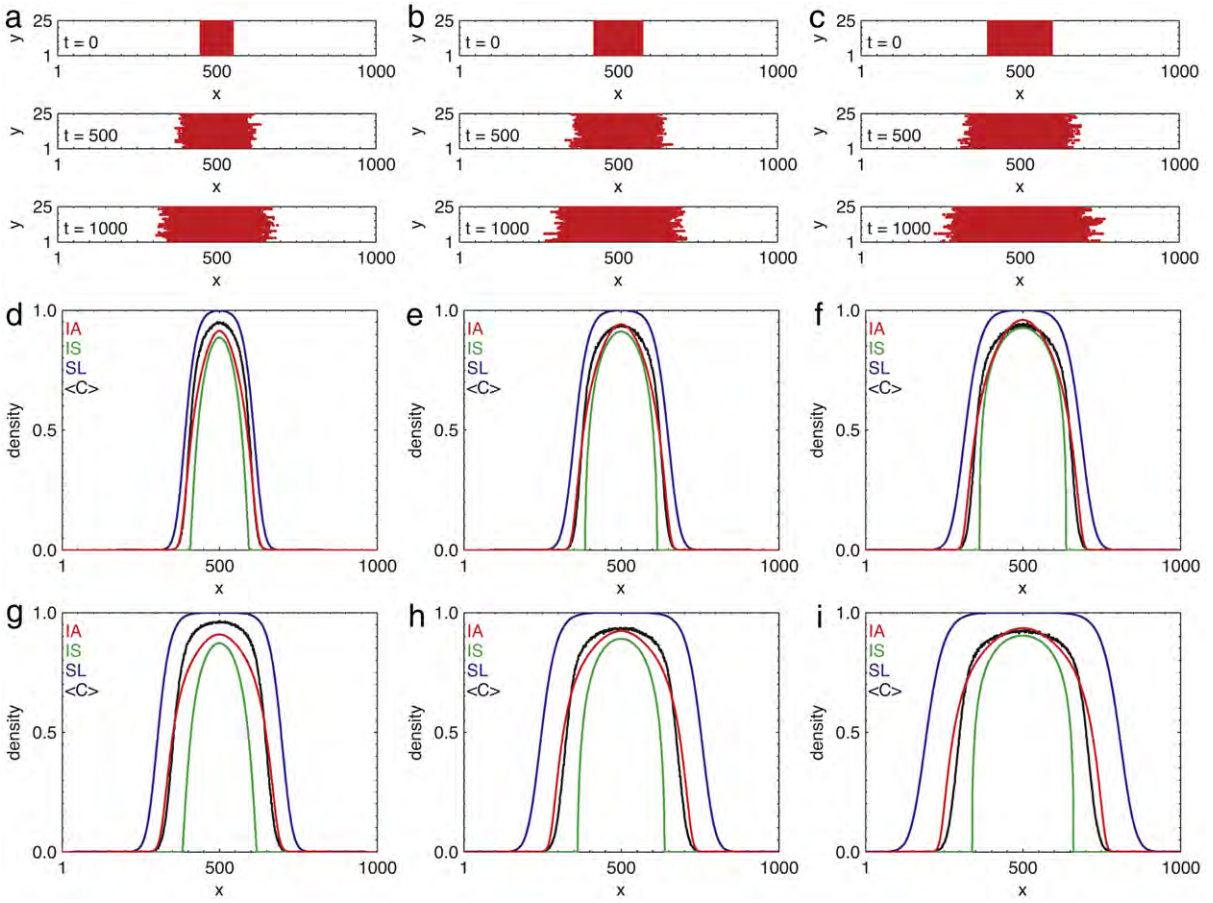

Fig. 5. (Color online). With proliferation: a series of one-dimensional simulations comparing the effects of different diffusion coefficients and source terms with the discrete data for $P_m = 1.0$ and $P_p = 0.01$. (a) Simulations for agents with $L = 2$. Initially all sites with $450 \leq i < 550$ are occupied by 50 adjacent columns of non-overlapping, horizontally-aligned agents. (b) Simulations for agents with $L = 3$ and initially 50 columns of adjacent agents filling sites $425 \leq i < 575$. (c) Simulations for agents with $L = 4$ and initially 50 columns of adjacent agents filling sites $400 \leq i < 600$. Simulation results are given at $t = 0$, 500, 1000 for $\delta t = \Delta = 1.0$. Results in (d)–(f) and (g)–(i) show averaged agent density (black lines) for a series of corresponding simulations similar to those shown in (a)–(c), with the data averaged over $M = 1000$ realisations at times $t = 500$ ((d)–(f)) and $t = 1000$ ((g)–(i)). These averaged discrete profiles are compared with the solutions of Eqs. (5), (20) and (11) shown in red, blue and green, respectively. The PDEs are solved numerically using the method outlined in the main text with $\Delta x = 0.1 - 0.5$ and $\Delta t = 0.001 - 0.1$ and $\epsilon = 10^{-6}$. The root mean square error in each case is given in Appendix A.

adjacent to the proliferating agent provided that the target site is unoccupied. We begin by outlining the method used to incorporate proliferation events in the discrete model, and then we incorporate these proliferation events into our three different approaches for deriving the continuum models.

3.1. Discrete simulations

In line with previous approaches [9,13,23], we adapt our discrete simulations to include proliferation events as follows. We record the total number of agents at time t as $N(t)$, and update the discrete simulations with time steps of duration δt. In any time step we first select $N(t)$ agents at random and independently, one at a time, and give these agents the opportunity to move. When chosen, an agent attempts to move using the same algorithm previously outlined. Second, $N(t)$ agents are selected independently at random, one at a time, and given the opportunity to proliferate. When chosen, an agent attempts to proliferate with probability P_p where $P_p \in [0, 1]$. The daughter agent is placed on the L adjacent sites to the agent, with equal probability of $1/2$ of choosing to place the daughter agent to the left or to the right. If, during the attempted proliferation event, any target site is occupied, then the proliferation event is aborted. At the end of the time step we update $N(t)$ to reflect the addition of new agents from successful proliferation events. This approach is appropriate for small values of P_p where the increase in $N(t)$ per time step is small.

Using this discrete mechanism, with a population of horizontally-aligned agents with varying length, L, we performed many simulations on a one-dimensional lattice. Simulations were performed on a lattice with $1 \leq i \leq 1000$ with reflecting boundary conditions imposed at $i = 1$ and $i = 1000$. The central part of this lattice is shown in Fig. 5(a)–(c) for $L = 2$, 3, 4,

respectively. The system was allowed to evolve and we observe snapshots of the populations of agents as they spread and increase in size in Fig. 5(a)–(c) at $t = 0, 500, 1000$, respectively. To complement these single snapshots, we also generated averaged agent density data in Fig. 5(d)–(i) for the same problems at $t = 500, 1000$ by averaging across $M = 1000$ identically prepared realisations. The data in Fig. 5 show how the density profile evolves as the agents spread away from their initial closely-packed distribution. If we compare these new simulation results with the previous results without proliferation (Fig. 2) then we can clearly see the effect of the proliferation events since the area under the density profiles increases with time which indicates that the total size of the population is increasing owing to the effects of proliferation. In each case the population changes by approximately a factor of two over the course of the simulation. We now aim to derive PDE models that can predict the density profiles with proliferation.

3.2. Mean-field models for one-dimensional collective motility and proliferation

We now discuss how to include proliferation in the different continuum population descriptions outlined in Sections 2.2.1 and 2.2.2. For simplicity we will present the new terms in the discrete conservation statement that arise due to proliferation events without including the motility terms given in Sections 2.2.1 and 2.2.2.

3.2.1. Individual sites model with proliferation

With proliferation events included, two additional terms arise in the discrete conservation equation (2) as agents may place the daughter agent either to the left or to the right. For the IS approach these additional terms can be written as

$$\frac{P_p}{2}\prod_{s=1}^{L} P(A_{i-s}, t)\,(1 - P(A_{i+s-1}, t)) + \frac{P_p}{2}\prod_{s=1}^{L} P(A_{i+s}, t)\,(1 - P(A_{i-s+1}, t))\,. \tag{27}$$

If we include the motility terms in the discrete conservation statement, then expand all terms in a truncated Taylor series about the point i, and consider the limit as $\Delta \to 0$ and $\delta t \to 0$ simultaneously we obtain

$$\frac{\partial C}{\partial t} = D_0 \frac{\partial}{\partial x}\left(D^{\mathrm{IS}}(C)\frac{\partial C}{\partial x}\right) + \rho\, C^L (1 - C)^L, \tag{28}$$

with D_0 and $D^{\mathrm{IS}}(C)$ given as in Eq. (6) and $\rho = \lim_{\delta t \to 0}(P_p/\delta t)$. This means that the discrete system has a well-defined continuum limit in the case that P_p scales linearly with δt [9,28]. This seems intuitive as it amounts to supposing, for example, that if P_p is the probability of proliferation in a time increment δt then the probability of proliferation in time increment $\delta t/2$ is $P_p/2$. Note that this reaction–diffusion equation reduces to Fisher's equation in the simple case that $L = 1$ which is consistent with previous studies that have considered round agents only [10].

3.2.2. Stretched lattice model with proliferation

With proliferation events included, the two additional terms in the discrete conservation equation for the SL approach can be written as

$$\frac{P_p}{2}P_L(A_{i-1}, t)(1 - P_L(A_i, t)) + \frac{P_p}{2}P_L(A_{i+1}, t)(1 - P_L(A_i, t)), \tag{29}$$

and the corresponding PDE takes the form

$$\frac{\partial C}{\partial t} = D_0 \frac{\partial}{\partial x}\left(D^{\mathrm{SL}}(C)\frac{\partial C}{\partial x}\right) + \rho\, C(1 - C), \tag{30}$$

with D_0 and $D^{\mathrm{SL}}(C)$ given as in Eq. (12) and $\rho = \lim_{\delta t \to 0}(P_p/\delta t)$. Once again, this is a reaction–diffusion equation which is valid whenever P_p scales linearly with δt. The system reduces to Fisher's equation in the simple case that $L = 1$, again consistent with previous studies that have considered round agents only [10]. We note that the reaction terms in the governing equation are independent of L.

3.2.3. Individual agents model with proliferation

With proliferation events included, the two additional terms in the discrete conservation equation for the IA approach can be written as

$$\frac{P_p}{2}P(A_{(i-L,\ldots,i-1)}; 0_i; \ldots; 0_{i+L-1}, t) + \frac{P_p}{2}P(0_i; \ldots; 0_{i+L-1}; A_{(i+L,\ldots,i+2L-1)}, t), \tag{31}$$

where, for example, $P(A_{(i-L,\ldots,i-1)}; 0_i; \ldots; 0_{i+L-1}, t)$ denotes the probability of having an agent with its left-most end at site $i - L$ and sites $i, \ldots, i + L - 1$ vacant at time t. We use a recursive argument to reduce our system to include only terms of the form $P^L(A_i, t)$. For example, for $L = 2$ we may write

$$P(A_{(i-2,i-1)}; 0_i; 0_{i+1}) = P(0_i; 0_{i+1}, t|A_{(i-2,i-1)}, t)P^2(A_{i-2}, t), \tag{32}$$

together with

$$P(0_i; 0_{i+1}, t|A_{(i-2,i-1)}, t) = 1 - P(A_{(i,i+1)}, t|A_{(i-2,i-1)}, t) - P(0_i; A_{(i+1,i+2)}, t|A_{(i-2,i-1)}, t),$$
$$= 1 - P^2(A_i, t) - P^2(A_i, t)P^2(A_{i+1}, t). \quad (33)$$

The first line of Eq. (33) comes from considering the possible occupancies of sites i and $i+1$ given we have an agent on sites $i-2$ and $i-1$. We could have either: (i) an agent on sites i and $i+1$; (ii) site i empty and an agent on sites $i+1$ and $i+2$; or (iii) both sites i and $i+1$ empty. On the second line we have assumed that, on average, the locations of agents are independent. The case for general L is a simple extension of this argument.

In general, for a population of fixed but arbitrary length, $L \geq 1$, agents this procedure gives rise to the PDE

$$\frac{\partial C}{\partial t} = D_0 \frac{\partial}{\partial x}\left(D^{\mathrm{IA}}(C)\frac{\partial C}{\partial x}\right) + \rho\, C(1-C)^L, \quad (34)$$

with D_0 and $D^{\mathrm{IA}}(C)$ as given in Eq. (21) and $\rho = \lim_{\delta t \to 0}(P_p/\delta t)$. Once again, this is a reaction–diffusion equation which is valid whenever P_p scales linearly with δt and the governing equation reduces to Fisher's equation in the simple case that $L = 1$.

3.3. Comparing discrete simulation data and mean-field models with proliferation

To compare the averaged density profiles with solutions of the mean-field models, we must solve Eqs. (28), (30) and (34) using the same domain, boundary conditions and initial condition as in the discrete simulations given in Fig. 5. We solve the PDEs numerically using the finite difference technique outlined previously except that we incorporate discretised source terms to account for proliferation.

Numerical solutions of Eqs. (28), (30) and (34) are superimposed on the discrete profiles in Fig. 5 and this reveals several key differences relative to the discrete-continuum comparison for the nonproliferative simulations shown previously in Figs. 2–4. Firstly we see that the PDE solutions corresponding to the IS approach and the SL approach give very poor approximations to the discrete data. In general we see that the IS PDE underestimates the spatial extent of the invasion profile whereas the SL PDE overestimates the spatial extent of the invasion profile. These trends are consistent with the nonproliferative results in Figs. 2–4 except that the discrepancy in the nonproliferative systems was localised to the leading edge only. The IA PDE provides an excellent match with the proliferative density profiles since it captures both the shape and location of the density profiles.

The difference between the discrete-continuum comparisons for the proliferative and nonproliferative simulations is physically intuitive. We know that the IS PDE cannot predict the smooth leading edge that we observed in Figs. 2–4. In the discrete simulations we know that proliferation requires sufficient space and so we expect that the majority of the proliferation events will occur at the leading edge of the population. This means that any discrepancy between the location and shape of the leading edge in the nonproliferative case will be amplified when we include proliferation. This explains why the IA PDE gives the best match to the proliferative profiles since it was the only PDE model that was able to capture the true shape and location of the leading edge in the nonproliferative simulations.

4. Travelling waves

Previously, in Section 3.3, we examined how the three continuum models are able to predict the collective spreading of proliferative systems and showed that the IA model gives the best approximation over relatively short time scales. To apply our models in the context of cell biology and, in particular, to apply our models to represent cell invasion waves prompts us now to consider the discrete–continuum comparison over much longer time scales and, therefore, to examine travelling wave behaviour. It is particularly interesting to investigate the appearance of travelling waves in the agent density profile and to determine which of the three continuum models can best approximate the shape and speed of the invasion fronts that arise in the discrete simulations.

4.1. Discrete simulations

In order to generate the discrete data, we performed $M = 1000$ identically prepared simulations of the discrete system on a lattice with $1 \leq i \leq 10,000$ for various values of P_p and a fixed agent shape $L = 2$. In all cases we used reflecting boundary conditions and started with only the left-most 200 lattice sites filled. In Fig. 6(a)–(b) we show the average wavefront shape for $P_p = 0.01$ and $P_p = 0.001$, respectively. In each case we generated the averaged discrete data by finding, in each simulation, the position of the agent that was fourth from the leading edge of the population. The data series in each realisation was translated so that these identified agents were at the same location. We then averaged the density data using Eq. (1). We used a similar approach to estimate the travelling wave speeds in Fig. 6(c). In each of the $M = 1000$ simulations, we identified the time history of the position of the agent that was fourth from the leading edge of the population and used a least-squares method to fit a straight line to the time history data. Once sufficient time has passed so that the constant speed travelling wave had appeared, the slope of the straight line was taken as an estimate of the travelling wave speed. The mean, standard deviation and standard error of the wave speed was calculated. We note that the wave speed results were

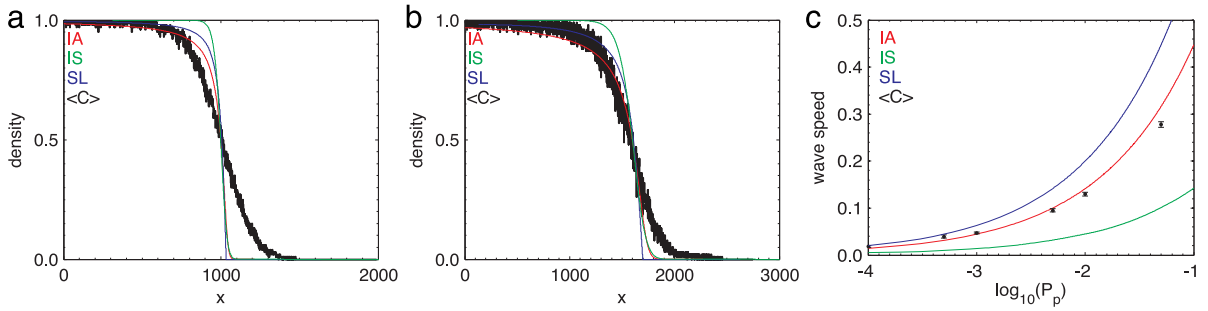

Fig. 6. (Color online). With proliferation: a series of one-dimensional simulations with $L = 2$ comparing travelling waves arising from the use of different diffusion coefficients and source terms with the discrete data for $P_m = 1.0$ and various values of P_p. (a)–(b) Comparison of the shape of the wavefronts for $P_p = 0.01$ and $P_p = 0.001$, respectively. In each case the results have been translated along the x axis so that the mid-points of the waves coincide. The averaged discrete results are constructed using $M = 1000$ realisations and $\delta t = \Delta = 1.0$. (c) Comparison of the travelling wave speeds for a range of values of P_p. The error bars indicate the standard deviation that was calculated from 100 identically prepared discrete simulations. The PDEs are solved numerically using the method outlined in the main text with $\Delta x = 0.1 - 0.5$, $\Delta t = 0.001 - 0.1$ and $\epsilon = 10^{-6}$.

insensitive to the choice of the tracked agent since tracking the position of the mth agent gave virtually indistinguishable results for $m = 1, \ldots, 10$ (results not shown).

4.2. Mean-field models

Once again, to compare the averaged density profiles with solutions of the mean-field models, we solve Eqs. (28), (30) and (34) using the same domain, boundary conditions and initial condition as in the discrete simulations. We solve the PDEs numerically using the finite difference technique outlined previously for a sufficiently long time to allow the evolution of travelling wave profiles. In the IA and SL models we can estimate the minimum wave speeds using the phase plane and standard linear analysis techniques to give

$$c_{\mathrm{SL}} = 2\sqrt{\rho D_0 D^{\mathrm{SL}}(0)} \quad \text{and} \quad c_{\mathrm{IA}} = 2\sqrt{\rho D_0}. \tag{35}$$

For initial conditions with compact support we expect numerical solutions of our PDE models to move with these minimum wave speeds [29], and simulations confirm that this is the case (data not shown). For simulations with $L = 2$ and $P_m = 1$, we have $c_{\mathrm{SL}} = 2\sqrt{\rho}$ and $c_{\mathrm{IA}} = \sqrt{2\rho}$. For the IS model it is not possible to determine the wave speed using the phase plane and standard linear analysis, and so we use generated numerical solutions of the PDE for $\rho = 0.1, 0.07, 0.05, 0.01, 0.005, 0.001, 0.0005, 0.0001$ and used these solutions to estimate the wave speed. Using these numerical approximations, we interpolated to produce a continuous estimate of the wave speed for $P_p \in [0.0001, 0.1]$.

4.3. Comparing discrete and continuum travelling waves

In Fig. 6 we present results that compare the shapes and speeds of the travelling wavefronts in the three mean-field models with averaged discrete data for $L = 2$. Fig. 6(a)–(b) compare the discrete and continuum shapes of the travelling wavefronts for $L = 2$. For $P_p = 0.01$ we see that none of the three models reproduce the shape of the wave particularly accurately, while for $P_p = 0.001$ we see a better match with the shape of the wave in the IA model.

More promising results come to light when we consider the travelling wave speed. Our results are presented in Fig. 6(c). All three PDE models provide a reasonably accurate prediction of the travelling wave speed and we see that the quality of the discrete-continuum comparison increases as P_p decreases relative to P_m. For small values for P_p both the SL and the IA approaches accurately estimate the travelling wave speed. However, for larger values of P_p we see that the IA model is the only PDE model that provides a good match to the discrete wave speed data. In fact, the IA wave speed estimate is at least an order of magnitude more accurate than the SL model for $P_p \in [0.005, 0.05]$. Note that $\delta t = 1$ and so $P_p = \rho$ here.

5. Simulation and analysis of collective motion and proliferation in two-dimensions

Finally, we present results that extend our analysis to two space dimensions. This is important as many cell populations, both *in vivo* and *in vitro* are 2D or 3D entities (see for example the neuronal explants shown in Fig. 7), and extension to higher dimensions is nontrivial. We consider only the case $L = 2$ as we have already demonstrated that we expect results to be similar for other values of L.

5.1. Discrete simulations

We extend our discrete simulations by allowing our agents to both move and proliferate in two dimensions. When undergoing a motility event, an agent can move in the positive and negative x and y directions with equal probability of $1/4$ and, as usual, attempted moves that place the agent on already occupied sites are aborted. Consider the agent occupying

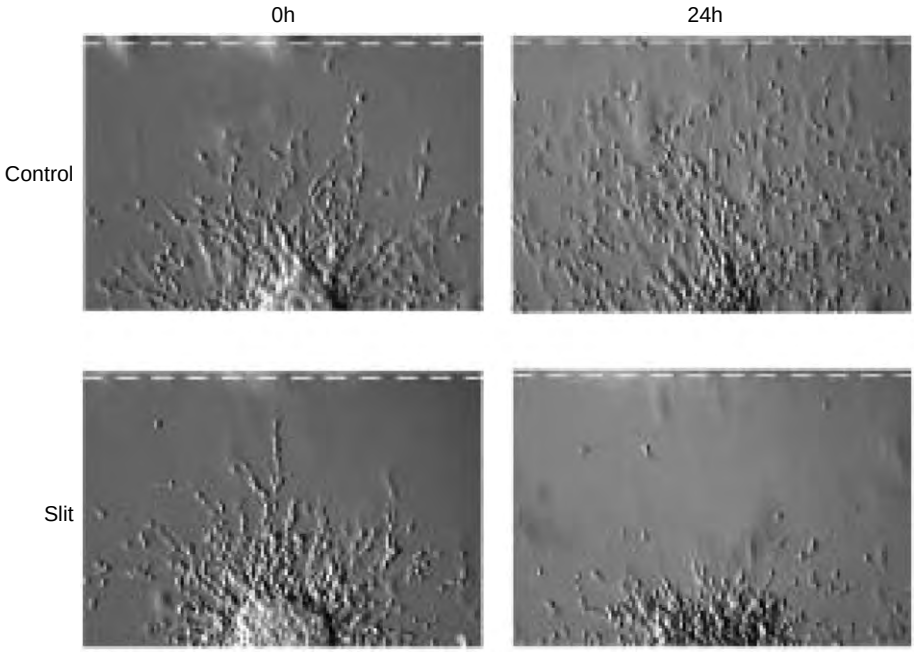

Fig. 7. Images from neuronal explant experiments designed to determine whether Slit controls cell directionality or motility illustrate the 2D nature of many cell populations. Explants were first cultured for 24 h, and then for another 24 h either in the presence of absence of Slit. Full details are described in Ref. [30].
Source: Reproduced from Ref. [30] with permission of Society for Neuroscience.

sites $(i-3, j+1)$ and $(i-2, j+1)$ in Fig. 1(c): as neither of the sites to the left or right of the agent are occupied, potential movement events in either of these directions would be permitted. However, attempted vertical movements, either in the positive or negative y direction, would be aborted as sites $(i-2, j+2)$ and $(i-3, j)$ are occupied. Proliferation is dealt with in a similar manner, placing agents on the nearest neighbour lattice sites whenever they are vacant. For example, the agent in Fig. 1(c) which occupies sites (i, j) and $(i+1, j)$ is able to proliferate in the positive y direction by placing a daughter agent on sites $(i, j+1)$ and $(i+1, j+1)$, however a proliferation event that would attempt to place a daughter agent in the positive x direction would be aborted due to exclusion. All simulations are performed with reflecting boundary conditions along all four boundaries of the lattice.

Fig. 8 shows the results of simulations of the discrete system. Column (a) shows the results of two single stochastic realisations, with the initial condition at the top and results for $P_p = 0.0$ and $P_p = 0.001$ at $t = 2000$ shown in the centre and bottom rows, respectively. In each case the resulting population distribution is anisotropic since the exclusion principles and the alignment of the agents make it easier for agents to move in the x direction compared to the y direction. The proliferative simulation with $P_p = 0.001$ is performed for a sufficiently long period of time that the number of agents approximately doubles in size, from 5000 agents at $t = 0$ to, on average, approximately 10,100 agents by $t = 2000$.

5.2. Mean-field models describing two-dimensional collective motility and proliferation

To connect the discrete mechanism with a mean-field model we form a discrete conservation statement describing the change in average occupancy of site (i, j) (or the change in average occupancy of groups of sites, $[(i, j), (i+1, j), (i+2, j), \ldots, (i+L-1, j)]$) during the time interval from t to $t+\delta t$. We note that there are now four possible movement directions and four possible directions in which to place a daughter agent when a proliferation event occurs. To match the two-dimensional density profiles in Fig. 6, we consider a population of horizontally-aligned agents, each with $L = 2$, that move and proliferate in both the x and y directions.

5.2.1. Individual sites

The conservation statement describing the IS model is given in Eq. (B.1) of Appendix B.1 and gives rise to the PDE

$$\frac{\partial C}{\partial t} = D_0 \frac{\partial}{\partial x}\left(D_x^{\mathrm{IS}}(C)\frac{\partial C}{\partial x}\right) + D_0 \frac{\partial}{\partial y}\left(D_y^{\mathrm{IS}}(C)\frac{\partial C}{\partial y}\right) + 2\rho\, C^2(1-C)^2, \tag{36}$$

with

$$D_x^{\mathrm{IS}}(C) = 4C, \qquad D_y^{\mathrm{IS}}(C) = 4C(1-C), \qquad D_0 = \lim_{\Delta\to 0,\delta t\to 0}\left(\frac{\Delta^2 P_m}{4\delta t}\right), \qquad \rho = \lim_{\delta t\to 0}\left(\frac{P_p}{\delta t}\right). \tag{37}$$

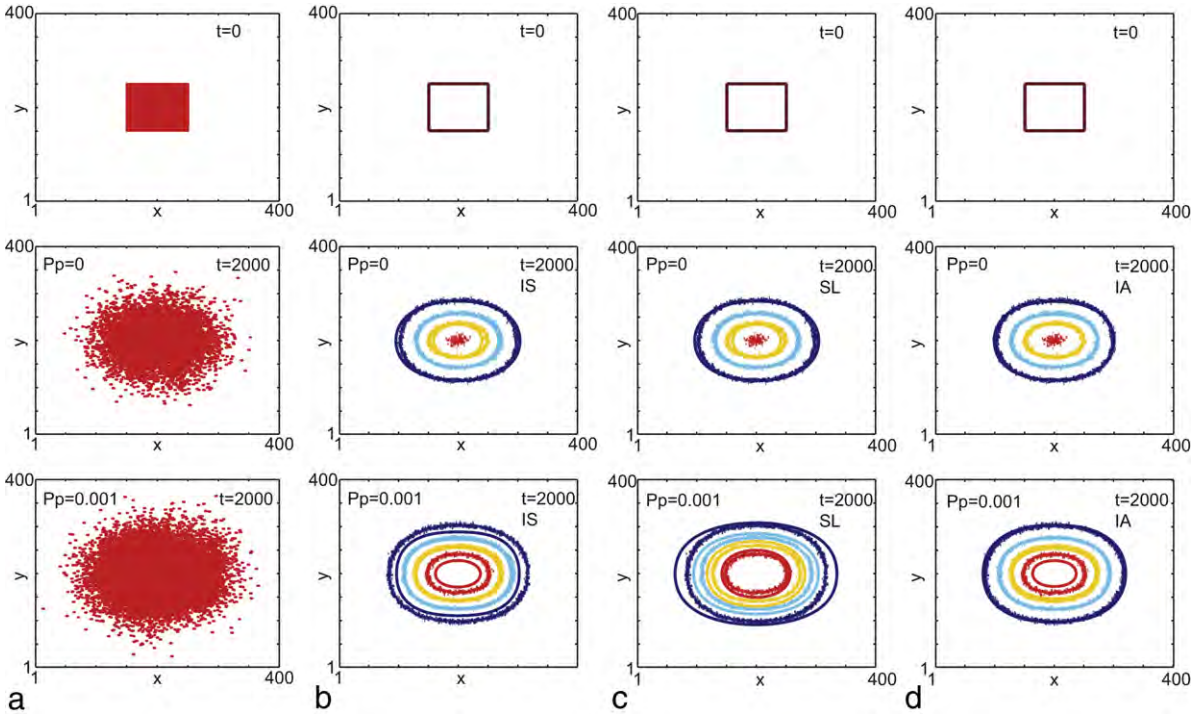

Fig. 8. (Color online). A series of two-dimensional simulations comparing the effects of different diffusion coefficients and source terms with averaged discrete data for $P_m = 1.0$ and $L = 2$. In column (a) we show snapshots from single discrete simulations showing the initial condition, the distribution of agents in a nonproliferative ($P_p = 0$) simulation, and the distribution of agents in a proliferative ($P_p = 0.001$) simulation. In all simulations $\delta t = \Delta = 1.0$, and the initial distribution of agents has all sites with $151 \leq i < 250$ occupied by 50 adjacent columns of non-overlapping, horizontally-aligned agents so that all rows with $151 \leq i < 250$ are occupied. The total number of agents initially on the lattice is 5000. Results in columns (b)–(d) show the discrete-continuum comparison for the IS, SL and IA approaches, respectively. We compare the PDE models with the discrete simulations by constructing averaged simulation data using Eq. (42) and 1000 identically prepared realisations. The PDEs are solved numerically using the method outlined in the main text with $\Delta x = \Delta y = 1.0$, $\Delta t = 0.25$ and $\epsilon = 10^{-6}$. All results are presented as contours that were generated using standard linear interpolation routines in MATLAB for $C = 0.1$ (dark blue), $C = 0.3$ (light blue), $C = 0.5$ (yellow) and $C = 0.7$ (red).

Once again, the limiting PDE is a nonlinear degenerate diffusion equation with $D_x^{\mathrm{IS}}(0) = D_y^{\mathrm{IS}}(0) = 0$. This means that the solution of the PDE can contain sharp interfaces.

5.2.2. Stretched lattice

The conservation statement describing the SL model is given in Eq. (B.2) of Appendix B.2 and gives rise to the PDE

$$\frac{\partial C}{\partial t} = D_0 \frac{\partial}{\partial x}\left(D_x^{\mathrm{SL}}(C)\frac{\partial C}{\partial x}\right) + D_0 \frac{\partial}{\partial y}\left(D_y^{\mathrm{SL}}(C)\frac{\partial C}{\partial y}\right) + \rho\, C(1-C), \tag{38}$$

with

$$D_x^{\mathrm{SL}}(C) = \frac{\bar{P}_m^x}{P_m}, \qquad D_y^{\mathrm{SL}}(C) = \frac{\bar{P}_m^y}{P_m}, \qquad D_0 = \lim_{\Delta\to 0,\delta t\to 0}\left(\frac{\Delta^2 P_m}{4\delta t}\right), \qquad \rho = \lim_{\delta t\to 0}\left(\frac{P_p}{\delta t}\right). \tag{39}$$

This is consistent with our previous one-dimensional models since the PDE has a linear diffusion mechanism and the source term is logistic. To determine $\bar{P}_m^x$ and $\bar{P}_m^y$, we consider the two-dimensional motion of a horizontally-aligned agent with length L on a stretched lattice, and compare this with the motion of a simple round agent with $L = 1$ on the regular square lattice with spacing Δ. An isolated horizontally-aligned agent with length L on a stretched lattice will step in the vertical direction in exactly the same way as the round agent, so we set $\bar{P}_m^y = P_m^y$. An isolated horizontally-aligned agent with length L on a stretched lattice will step in the horizontal direction with a reduced probability compared to the round agent so we set $\bar{P}_m^y = P_m^y/L$. These physical considerations give us $D_x^{\mathrm{SL}}(C) = 2$ and $D_y^{\mathrm{SL}}(C) = 1$ for $L = 2$.

5.2.3. Individual agents

The conservation statement describing the IA model is given in Eq. (B.3) of Appendix B.3 and gives rise to the PDE

$$\frac{\partial C}{\partial t} = D_0 \frac{\partial}{\partial x}\left(D_x^{\mathrm{IA}}(C)\frac{\partial C}{\partial x}\right) + D_0 \frac{\partial}{\partial y}\left(D_y^{\mathrm{IA}}(C)\frac{\partial C}{\partial y}\right) + \frac{P_p}{2}C(1-C)^2 + \frac{P_p}{2}C(1-3C+C^2), \tag{40}$$

with

$$D_x^{\mathrm{IA}}(C) = 1 + 2C, \qquad D_y^{\mathrm{IA}}(C) = 1 - C^2, \qquad D_0 = \lim_{\Delta\to 0,\delta t\to 0}\left(\frac{\Delta^2 P_m}{4\delta t}\right), \qquad \rho = \lim_{\delta t\to 0}\left(\frac{P_p}{\delta t}\right). \tag{41}$$

Similar to our one-dimensional results the PDE is a nonlinear, nondegenerate diffusion equation. We note that here the proliferation term becomes negative for $C > 0.5$. This is an artefact of the anisotropy of the agent shape; there are fewer opportunities for the agents to proliferate in the vertical direction than in the horizontal direction.

5.3. Comparing discrete simulation data and mean-field models in 2D

To compare the averaged density profiles with solutions of the mean-field models, we must solve Eqs. (36)–(37), (38)–(39) and (40)–(41) numerically using the same domain, boundary conditions and initial condition as in the discrete simulations. We solve the PDEs using the finite difference technique outlined previously. To generate the averaged agent density data from the two-dimensional simulations, we take a two-dimensional average over many identically prepared realisations, given by

$$\langle C(i,j,t)\rangle = \frac{1}{M}\sum_{m=1}^{M} c^m(i,j,t). \tag{42}$$

To present the two-dimensional data, we contour both the averaged discrete data and the numerical solution of the PDE to generate the contours shown in Fig. 8. For these results we constructed the contours using $C = 0.1, 0.3, 0.5, 0.7$.

Comparing the numerical solutions of Eqs. (36)–(37) and the discrete contours in Fig. 8 reveals certain trends that are consistent with our one-dimensional results. Results in columns (b)–(d) show the discrete–continuum comparison for the IS, SL and IA models, respectively. The contours in the middle row correspond to the nonproliferative simulations and, in each case, we observe that the discrete–continuum match is good. In comparison, the contours in the lower row correspond to the proliferative simulations. We can clearly see the effect of proliferation since the contours show that the population density has spread further in the proliferative simulations compared to the nonproliferative simulations. For the proliferative simulations we observe that the discrete-continuum match is relatively poor for both the IS and the SL models whereas the quality of the match is higher for the IA model. These effects are demonstrated clearly by comparing the discrete–continuum match for the $C = 0.1$ contour where the IA PDE matches the discrete data, the IA model underestimates the extent of the spreading and the SL model overestimates the extent of spreading. These trends are entirely consistent with our one-dimensional results shown in Fig. 5.

6. Discussion, conclusion and outlook

The development of multiscale models of cell biology processes such as cell motility and proliferation is vital if we are to accurately describe the collective motion of spreading cell populations. Experimental data are generally collected on both the population-level and individual-level and it is difficult to reconcile observations on the two levels without models which can explicitly link how processes at the individual-level affect the population-level behaviour of the system. As a result, a great deal of recent modelling work as been concerned with investigating the links between the individual-level and population-level, see, for example Refs. [8,31–33].

Most previous work has considered the motion and proliferation of round cells, where the longitudinal and transverse length scales of each cell are equal. However, cells are often elongated or rod-shaped (see Fig. 1) and it is unclear how the detailed shape of the cells impacts motility and proliferation rates on the population-level. This work investigates motility and proliferation mechanisms for uniformly sized populations of cells, with a view to understanding the population-level response to changes in cell shape.

To model the collective motion of a population of interacting cells, we have considered a generalised exclusion process model where each cell is represented by an agent on a lattice. To model varying cell shape, each agent can occupy $L \geq 1$ adjacent lattice sites, and agents are aligned with their longest sides horizontal. Here we have considered three different individual-level descriptions of our agents, in each case deriving a conservation statement that describes the occupancy probability of groups of lattice sites and a corresponding population-level PDE. Numerical simulation shows that, without proliferation, each of the mean-field models provides a good approximation to the averaged discrete data. This is, in part, an expected result since all three approaches aim to describe the same process. Simulation data and mathematical arguments based on the MAT confirm that all three mean-field models can describe the collective motion of a population of nonproliferative agents with $L = 2, 3, 4$. Our results show that there is a small discrepancy between the continuum and discrete data right at the leading edge of the population for the SL and IS models and we observe that the IA approach gives the best match to the discrete data at the leading edge. We reason that the IA model provides the best match to the discrete data since this is the most comprehensive approach for deriving mean-field models which account for varying the agent size and the associated exclusion effects.

The discrete-continuum match for proliferative systems shows that the density profiles based on the IS and SL approaches are a very poor approximation to the discrete data whereas the density profiles based on the IA approach provides a good approximation to the data over the entire domain. Proliferation events in the discrete simulations always occur at the lead-

ing edge of the population. Accordingly, we reason that the small discrepancies at the leading edge of the nonproliferative density profiles for the IS and SL models are amplified when we introduce proliferation. This explains why the IS and SL models do not predict the density profiles for proliferative populations while the IA model gives a much better result. The trends observed in the discrete–continuum match are similar when we consider proliferative simulations over longer time scales in order to study the travelling wave behaviour of the models. We simulated each discrete model for a sufficiently long period of time until a travelling wave formed and compared continuum and discrete wave speed and wave shape data. Once again, this comparison revealed that the IA model provided the best description of the averaged discrete data. Two-dimensional results were also generated, and the discrete–continuum match showed similar trends that were observed in the one-dimensional case.

This work represents a important step forward in bridging the gap between discrete, agent-based models of cell processes that can only be subject to repeated simulation, and continuum PDE models that can be explored using a range of analytical techniques. Our work demonstrates how both linear and nonlinear diffusion can play a role in describing the collective motion of cell populations. This work provides a link between the use of linear and nonlinear diffusion mechanisms and the physical processes associated with cell migration rather than relying on model calibration arguments or subjectively choosing a particular form of a diffusion coefficient.

The significance of this work can be best understood by considering how it advances recent research into the development of mean-field PDE models for exclusion processes. All previous investigators have used approximate averaging arguments and an independence assumption to derive approximate mean-field PDE models for exclusion process with round agents ($L = 1$), see for example Refs. [1,8,15]. In all these previous studies, the transition probability associated with an unbiased motility event where an agent steps from site i to site $i + 1$, would be proportional to $A_i(1 - A_{i+1})$. The product of the two terms, $A_i(1 - A_{i+1})$, represents the probability that site i is occupied and the probability that site $i + 1$ is vacant. By interpreting the product of these two terms as a net transition probability, all previous studies made the implicit assumption that the occupancy of site i is independent of the occupancy of site $i + 1$. This is a standard assumption made by us [12] and others [1,8,15]. We believe that our work is the first to pioneer the development of mean-field PDE models for more practical problems where we consider the motion of agents with different shapes and sizes ($L > 1$). Repeating the standard conservation arguments with agents of different sizes leads to the IS model and here we show that the IS model captures nonproliferative simulation data reasonably accurately, but fails to capture proliferative simulation data. This shows that a new approach is required and motivates the development of the SL and IA approaches. In general, our work shows that the relationship between the exclusion process random walk and the associated mean-field PDE description becomes very complicated when we consider agents of different shapes and sizes. Many further details, such as considering agents that dynamically change shape or populations of agents with a distribution of agent shapes and sizes, warrants further study. The work outlined in the current manuscript lays the foundation upon which these future challenges will be met, and we are making progress toward meeting them.

We conclude with a final remark; that our work provides a novel link between exclusion process models, diffusion models (both linear and nonlinear) and the concept of MAT. Although we have showed that, in one-dimension, our three diffusion equations for the nonproliferative systems have the same MAT for any agent length L, our work also suggests that further developments to extend the concept of MAT are warranted. The theoretical basis for MAT was originally proposed to study nonlinear diffusion equations with isotropic nonlinear diffusivity and no source term in the PDE [18,19]. Here we show that, for applications involving collective cell migration with proliferation in two or three dimensions, we are interested in reaction–diffusion PDEs with anisotropic diffusivity. Under these conditions it would be useful to have a more broad definition of MAT so that we could replace one anisotropic reaction–diffusion PDE with another equivalent reaction–diffusion PDE. This could, perhaps, build on our physical arguments and provide a mathematical explanation as to why the IS and SL models fail to match the discrete data for proliferative simulations.

Appendix A. Root mean square errors

RMSE for the case of no proliferation as presented in Fig. 2.

Model, $L = 2$	RMSE at $t = 500$	RMSE at $t = 1000$
IA	0.0003	0.0003
IS	0.0005	0.0006
SL	0.0006	0.0005
Model, $L = 3$	RMSE at $t = 500$	RMSE at $t = 1000$
IA	0.0005	0.0005
IS	0.0008	0.0010
SL	0.0010	0.0009
Model, $L = 4$	RMSE at $t = 500$	RMSE at $t = 1000$
IA	0.0008	0.0008
IS	0.0011	0.0013
SL	0.0013	0.0013

RMSE for the case of no proliferation with $L = 3$ and different initial conditions as presented in Fig. 4.

Model, 20 columns	RMSE at $t = 500$	RMSE at $t = 1000$
IA	0.0004	0.0005
IS	0.0007	0.0009
SL	0.0008	0.0009
Model, 50 columns	RMSE at $t = 500$	RMSE at $t = 1000$
IA	0.0005	0.0005
IS	0.0008	0.0010
SL	0.0010	0.0009
Model, 100 columns	RMSE at $t = 500$	RMSE at $t = 1000$
IA	0.0005	0.0006
IS	0.0010	0.0013
SL	0.0009	0.0010

RMSE for the case with proliferation as presented in Fig. 5.

Model, $L = 2$	RMSE at $t = 500$	RMSE at $t = 1000$
IA	0.0013	0.0016
IS	0.0039	0.0075
SL	0.0026	0.0067
Model, $L = 3$	RMSE at $t = 500$	RMSE at $t = 1000$
IA	0.0010	0.0016
IS	0.0038	0.0072
SL	0.0040	0.0094
Model, $L = 4$	RMSE at $t = 500$	RMSE at $t = 1000$
IA	0.0011	0.0023
IS	0.0039	0.0072
SL	0.0050	0.0113

Appendix B. Conservation statements for the two-dimensional derivations

B.1. Individual sites

The two-dimensional conservation statement for the IS model in two dimensions can be written

$$\begin{aligned}
P(A_{i,j}, t+\delta t) - P(A_{i,j}, t) = {} & \frac{P_m}{4} P(A_{i-2,j}, t) P(A_{i-1,j}, t) \left(1 - P(A_{i,j}, t)\right) + \frac{P_m}{4} P(A_{i+1,j}, t) P(A_{i+2,j}, t) \left(1 - P(A_{i,j}, t)\right) \\
& + \frac{P_m}{4} P(A_{i,j-1}, t) P(A_{i+1,j-1}, t) \left(1 - P(A_{i,j}, t)\right) \left(1 - P(A_{i+1,j}, t)\right) \\
& + \frac{P_m}{4} P(A_{i,j+1}, t) P(A_{i+1,j+1}, t) \left(1 - P(A_{i,j}, t)\right) \left(1 - P(A_{i+1,j}, t)\right) \\
& + \frac{P_m}{4} P(A_{i,j-1}, t) P(A_{i-1,j-1}, t) \left(1 - P(A_{i,j}, t)\right) \left(1 - P(A_{i-1,j}, t)\right) \\
& + \frac{P_m}{4} P(A_{i,j+1}, t) P(A_{i-1,j+1}, t) \left(1 - P(A_{i,j}, t)\right) \left(1 - P(A_{i-1,j}, t)\right) \\
& - \frac{P_m}{4} P(A_{i,j}, t) P(A_{i-1,j}, t) \left(1 - P(A_{i-2,j}, t)\right) - \frac{P_m}{4} P(A_{i,j}, t) P(A_{i+1,j}, t) \left(1 - P(A_{i+2,j}, t)\right), \\
& - \frac{P_m}{4} P(A_{i,j}, t) P(A_{i+1,j}, t) \left(1 - P(A_{i,j-1}, t)\right) \left(1 - P(A_{i+1,j-1}, t)\right) \\
& - \frac{P_m}{4} P(A_{i,j}, t) P(A_{i+1,j}, t) \left(1 - P(A_{i,j+1}, t)\right) \left(1 - P(A_{i+1,j+1}, t)\right) \\
& - \frac{P_m}{4} P(A_{i,j}, t) P(A_{i-1,j}, t) \left(1 - P(A_{i,j-1}, t)\right) \left(1 - P(A_{i-1,j-1}, t)\right)
\end{aligned}$$

$$
\begin{aligned}
&- \frac{P_m}{4} P(A_{i,j}, t) P(A_{i-1,j}, t) \left(1 - P(A_{i,j+1}, t)\right) \left(1 - P(A_{i-1,j+1}, t)\right) \\
&+ \frac{P_p}{4} P(A_{i-2,j}, t) P(A_{i-1,j}, t) \left(1 - P(A_{i,j}, t)\right) \left(1 - P(A_{i+1,j}, t)\right) \\
&+ \frac{P_p}{4} P(A_{i+1,j}, t) P(A_{i+2,j}, t) \left(1 - P(A_{i,j}, t)\right) \left(1 - P(A_{i-1,j}, t)\right) \\
&+ \frac{P_p}{4} P(A_{i,j+1}, t) P(A_{i+1,j+1}, t) \left(1 - P(A_{i,j}, t)\right) \left(1 - P(A_{i+1,j}, t)\right) \\
&+ \frac{P_p}{4} P(A_{i,j+1}, t) P(A_{i-1,j+1}, t) \left(1 - P(A_{i,j}, t)\right) \left(1 - P(A_{i-1,j}, t)\right) \\
&+ \frac{P_p}{4} P(A_{i,j-1}, t) P(A_{i+1,j-1}, t) \left(1 - P(A_{i,j}, t)\right) \left(1 - P(A_{i+1,j}, t)\right) \\
&+ \frac{P_p}{4} P(A_{i,j-1}, t) P(A_{i-1,j-1}, t) \left(1 - P(A_{i,j}, t)\right) \left(1 - P(A_{i-1,j}, t)\right) \\
&+ \frac{P_p}{4} P(A_{i+2,j}, t) P(A_{i+3,j}, t) \left(1 - P(A_{i,j}, t)\right) \left(1 - P(A_{i+1,j}, t)\right) \\
&+ \frac{P_p}{4} P(A_{i-3,j}, t) P(A_{i-2,j}, t) \left(1 - P(A_{i,j}, t)\right) \left(1 - P(A_{i-1,j}, t)\right). \qquad \text{(B.1)}
\end{aligned}
$$

B.2. Stretched lattice

The two-dimensional conservation statement for the SL model in two dimensions can be written

$$
\begin{aligned}
P_2(A_{i,j}, t + \delta t) - P_2(A_{i,j}, t) = {} & \frac{\bar{P}_m^x}{4} \left[P_2(A_{i-1,j}, t) + P_2(A_{i+1,j}, t)\right] (1 - P_2(A_{i,j}, t)) \\
&+ \frac{\bar{P}_m^y}{4} \left[P_2(A_{i,j-1}, t) + P_2(A_{i,j+1}, t)\right] (1 - P_2(A_{i,j}, t)) \\
&- \frac{\bar{P}_m^x}{4} \left[(1 - P_2(A_{i-1,j}, t)) + (1 - P_2(A_{i+1,j}, t))\right] P_2(A_{i,j}, t) \\
&- \frac{\bar{P}_m^y}{4} \left[(1 - P_2(A_{i,j-1}, t)) + (1 - P_2(A_{i,j+1}, t))\right] P_2(A_{i,j}, t) \\
&+ \frac{P_p}{4} P_2(A_{i-1,j}, t)(1 - P_2(A_{i,j}, t)) + \frac{P_p}{4} P_2(A_{i+1,j}, t)(1 - P_2(A_{i,j}, t)) \\
&+ \frac{P_p}{4} P_2(A_{i,j-1}, t)(1 - P_2(A_{i,j}, t)) + \frac{P_p}{4} P_2(A_{i,j+1}, t)(1 - P_2(A_{i,j}, t)). \qquad \text{(B.2)}
\end{aligned}
$$

B.3. Individual agents

The two-dimensional conservation statement for the IA model in two dimensions can be written

$$
\begin{aligned}
P^2(A_{i,j}, t + \delta t) - P^2(A_{i,j}, t) = {} & \frac{P_m}{4} P(0_{i,j}; A_{(i+1,j),(i+2,j)}, t) + \frac{P_m}{4} P(A_{(i-1,j),(i,j)}; 0_{i+1,j}, t) \\
&+ \frac{P_m}{4} P(0_{i,j}; 0_{i+1,j}; A_{(i,j-1),(i+1,j-1)}, t) + \frac{P_m}{4} P(0_{i,j}, ; 0_{i+1,j}; A_{(i,j+1),(i+1,j+1)} t) \\
&- \frac{P_m}{4} P(A_{(i,i+1),j}; 0_{i+2,j}, t) - \frac{P_m}{4} P(0_{i-1,j}; A_{(i,j),(i+1,j)}, t) \\
&- \frac{P_m}{4} P(0_{i,j-1}; 0_{i+1,j-1}; A_{(i,j),(i+1,j)}, t) - \frac{P_m}{4} P(0_{i,j+1}, ; 0_{i+1,j+1}; A_{(i,j),(i+1,j)} t) \\
&+ \frac{P_p}{4} P(A_{(i-2,j),(i-1,j)}; 0_{i,j}; 0_{i+1,j}, t) + \frac{P_p}{4} P(0_{i,j}; 0_{i+1,j}; A_{(i+2,j),(i+3,j)}, t) \\
&+ \frac{P_p}{4} P(A_{(i,j-1),(i+1,j-1)}; 0_{i,j}; 0_{i+1,j}, t) + \frac{P_p}{4} P(0_{i,j}; 0_{i+1,j}; A_{(i,j+1),(i+1,j+1)}, t). \qquad \text{(B.3)}
\end{aligned}
$$

As in the one-dimensional case, we condition on occupancy and assume independence of agents to write, for example, for the vertical proliferation events,

$$\begin{aligned} P(0_{i,j};\, 0_{i+1,j};\, A_{(i,j+1),(i+1,j+1)},\, t) &= P(0_{i,j};\, 0_{i+1,j},\, t|A_{(i,j+1),(i+1,j+1)},\, t)P^2(A_{i,j+1},\, t), \\ &= \left[1 - P^2(A_{i,j},\, t) - P^2(A_{i-1,j},\, t)P^2(A_{i+1,j},\, t) - P^2(A_{i-1,j},\, t)(1 - P^2(A_{i+1,j},\, t))\right. \\ &\quad \left. -\ P^2(A_{i+1,j},\, t)(1 - P^2(A_{i-1,j},\, t))\right] P^2(A_{i,j+1},\, t), \end{aligned} \tag{B.4}$$

where we get from the first line to the second by considering all the possible configurations of agents that could occupy either or both of the sites (i, j) and $(i + 1, j)$. In this case we could have: (i) an agent with its left-most end at (i, j); (ii) an agent with its left-most end at $(i - 1, j)$ and $(i + 1, j)$ empty; (iii) (i, j) empty and an agent with its left-most end at $(i + 1, j)$; (iv) agents with their left-most sites at $(i - 1, j)$ and $(i + 1, j)$; or (v) both sites (i, j) and $(i + 1, j)$ empty. For the horizontal proliferation events we apply the same argument to get

$$\begin{aligned} P(0_{i,j};\, 0_{i+1,j};\, A_{(i+2,j),(i+3,j)},\, t) &= P(0_{i,j};\, 0_{i+1,j},\, t|A_{(i+2,j),(i+3,j)},\, t)P^2(A_{i+2,j},\, t), \\ &= \left[1 - P^2(A_{i,j},\, t) - P^2(A_{i-1,j},\, t)(1 - P^2(A_{i,j},\, t))\right] P^2(A_{i+2,j},\, t). \end{aligned} \tag{B.5}$$

References

[1] E. Khain, M. Katakowski, S. Hopkins, A. Szalad, X. Zheng, F. Jiang, M. Chopp, Collective behavior of brain tumor cells: the role of hypoxia, Phys. Rev. E 83 (3) (2011) 031920.
[2] C. Topaz, A.L. Bertozzi, Swarming patterns in a two-dimensional kinematic model for biological groups, SIAM J. Appl. Math. 65 (1) (2004) 152–174.
[3] D. Yanagisawa, K. Nishinari, Mean-field theory for pedestrian outflow through an exit, Phys. Rev. E 76 (2) (2007) 061117.
[4] H.M. Young, A.J. Bergner, R.B. Anderson, H. Enomoto, J. Milbrandt, D.F. Newgreen, P.M. Whitington, Dynamics of neural crest-derived cell migration in the embryonic mouse gut, Dev. Biol. 270 (2) (2004) 455–473.
[5] P.M. Kulesa, J.M. Teddy, D.A. Stark, S.E. Smith, R. McLennan, Neural crest invasion is a spatially-ordered progression into the head with higher cell proliferation at the migratory front as revealed by the photoactivatable protein, KikGR, Dev. Biol. 316 (2) (2008) 275–287.
[6] P.K. Maini, D.L.S. McElwain, D.I. Leavesley, Travelling waves in a wound healing assay, Appl. Math. Lett. 17 (5) (2004) 575–580.
[7] A. Huttenlocher, M. Lakonishok, M. Kinder, S. Wu, T. Truong, K. Knudsen, A. Horwitz, Integrin and cadherin synergy regulates contact inhibition of migration and motile activity, J. Cell Biol. 141 (2) (1998) 515–526.
[8] C. Deroulers, M. Aubert, M. Badoual, B. Grammaticos, Modeling tumor cell migration: from microscopic to macroscopic models, Phys. Rev. E 79 (3) (2009) 031917.
[9] M.J. Simpson, K.A. Landman, B.D. Hughes, Cell invasion with proliferation mechanisms motivated by time-lapse data, Physica A 389 (18) (2010) 3779–3790.
[10] T. Callaghan, E. Khain, L.M. Sander, R.M. Ziff, A stochastic model for wound healing, J. Stat. Phys. 122 (5) (2006) 909–924.
[11] A.E. Fernando, K.A. Landman, M.J. Simpson, Nonlinear diffusion and exclusion processes with contact interactions, Phys. Rev. E 81 (1) (2010) 011903.
[12] M.J. Simpson, R.E. Baker, S.W. McCue, Models of collective cell spreading with variable cell aspect ratio: a motivation for degenerate diffusion models, Phys. Rev. E 83 (2) (2011) 021901.
[13] M.J. Simpson, K.A. Landman, B.D. Hughes, Multi-species simple exclusion processes, Physica A 388 (4) (2009) 399–406.
[14] M.J. Simpson, C. Towne, D.L.S. McElwain, Z. Upton, Migration of breast cancer cells: understanding the roles of volume exclusion and cell-to-cell adhesion, Phys. Rev. E 82 (4) (2010) 041901.
[15] T.M. Liggett, Stochastic Interacting Systems: Contact, Voter and Exclusion Processes, Springer-Verlag, 1999.
[16] J.A. Sherratt, J.D. Murray, Models of epidermal wound healing, Proc. R. Soc. Lond. Ser. B 241 (1300) (1990) 29–36.
[17] B.G. Sengers, C.P. Please, R.O.C. Oreffo, Experimental characterization and computational modelling of two-dimensional cell spreading for skeletal regeneration, J. Roy. Soc. Interface 4 (17) (2007) 1107–1117.
[18] A. McNabb, G.C. Wake, Heat conduction and finite measures for transition times between steady states, IMA J. Appl. Math. 47 (47) (1991) 193–206.
[19] A. McNabb, Mean action times, time lags, and mean first passage times for some diffusion problems, Math. Comput. Modelling 18 (10) (1993) 123–129.
[20] K.A. Landman, M. McGuinness, Mean action time for diffusive processes, J. Appl. Math. Decis. Sci. 4 (4) (2000) 125–141.
[21] R.E. Baker, M.J. Simpson, Correcting mean-field approximations for birth–death-movement processes, Phys. Rev. E 82 (4) (2010) 041905.
[22] M.J. Simpson, R.E. Baker, Corrected mean-field models for spatially-dependent advection–diffusion–reaction phenomena, Phys. Rev. E 83 (5) (2011) 051922.
[23] M.J. Simpson, K.A. Landman, B.D. Hughes, A.E. Fernando, A model for mesoscale patterns in motile populations, Physica A 389 (7) (2010) 1412–1424.
[24] G.I. Barenblatt, Scaling, Cambridge University Press, 2003.
[25] M.J. Simpson, K.A. Landman, B.D. Hughes, Pathlines in exclusion processes, Phys. Rev. E 79 (3) (2009) 031920.
[26] M.J. Simpson, K.A. Landman, T.P. Clement, Assessment of a non-traditional operator split algorithm for simulation of reactive transport, Math. Comput. Simulation 70 (1) (2005) 44–60.
[27] K.A. Landman, L. White, Predicting filtration time and maximizing throughput in a pressure filter, AIChE J. 43 (12) (1997) 3147–3160.
[28] E.A. Codling, M.J. Plank, S. Benhamou, Random walk models in biology, J. Roy. Soc. Interface 5 (25) (2008) 813–934.
[29] J.D. Murray, Mathematical Biology I: An Introduction, third ed., vol. I, Springer-Verlag, 2003.
[30] M. Ward, C. McCann, M. DeWulf, J.Y. Wu, Y. Rao, Distinguishing between directional guidance and motility regulation in neuronal migration, J. Neurosci. 23 (12) (2003) 5170–5177.
[31] M.J. Simpson, A. Merrifield, K.A. Landman, B.D. Hughes, Simulating invasion with cellular automata: connecting cell-scale and population-scale properties, Phys. Rev. E 76 (2 Pt 1) (2007) 021918.
[32] P.J. Murray, C.M. Edwards, M.J. Tindall, P.K. Maini, From a discrete to a continuum model of cell dynamics in one dimension, Phys. Rev. E 80 (3) (2009) 031912.
[33] J.A. Fozard, H.M. Byrne, O.E. Jensen, J.R. King, Continuum approximations of individual-based models for epithelial monolayers, Math. Med. Biol. 27 (1) (2010) 39–74.