AMSI
AUSTRALIAN MATHEMATICAL SCIENCES INSTITUTE

National Collaboration in the Mathematical Sciences

Collective motion of a population of elongated cells

Melanie Kaasinen
Supervisors: A/Profs. Matthew Simpson and Scott McCue

Mathematical Sciences, Queensland University of Technology, Brisbane, Australia

February 2014

Abstract

The collective spreading of cell populations is often described using continuum diffusion models. Most previous studies use linear diffusion models to describe collective cell migration while some prefer to use a degenerate nonlinear diffusion model as it provides a better match to experimental data and supports sharp-fronted solutions. There is no guidance in the cell modelling literature with regard to which approach is more suitable for representing the collective spreading of cell populations under which circumstances. Here, a generalised lattice-based exclusion process is used to simulate the collective motion of a populations of interacting agents of various cell length, in one dimension. Rather than relying on model calibration, the relationship between the discrete motility mechanisms and PDE models is analysed by considering the continuum limit of the processes. Results confirm that "round" agents occupying a single lattice site, obey the linear diffusion equation whereas elongated agents, occupying more than one lattice site, obey a degenerate nonlinear diffusion equation related to the porous medium equation (PME) where the exponent in the nonlinear diffusivity function is dependent upon the length of the agents. Furthermore, the validity of the assumption that the PME solutions are totally differentiable was analysed by comparing discrete and continuum models at steady state. Initial results suggest that the subsequent use of a truncated Taylor series expansion to relate the occupancy of lattice sites was the main cause of the discrepancy between the discrete and continuum models for elongated agents.

1 Introduction

1.1 Biological Motivation

Cell motility plays a crucial role in many biological processes in the human body, including during tissue regeneration and embryological development. The collective migration of tissue cells, such as fibroblasts, is of particular importance in the applications such as wound healing [1], biomaterial development [2] and cancer biology [3]. Although many cell types deform during movement [4] fibroblasts do not, instead "gliding" over their substrate. When released from tissue and placed in culture they migrate randomly and display contact inhibition [4], a physical process that prohibits cells in contact from continued movement over the surface of the other [5]. These characteristic properties of fibroblastic cell motility define the PDE models used to describe collective cell spreading.

Most existing models neglect to take into account cell morphology [6–13], describing the motility of perfectly round or spherical cells. However, many of the applications which require such models involve fibroblastic cells with an elongated form. The population of human peritoneal mesothelial cells in [1], used in wound healing applications, clearly falls within this category. Furthermore, Smith, Marra and Marshall [14] showed that cells in live melanoma tumours undergo changes in shape from round to more elongated, thereby affecting their motility. Such examples illustrate the importance of developing models which provide a link between motility and morphology. This project aims to address that need by exploring models of collective cell spreading for cells of varying shape.

1.2 Modelling Collective Cell Spreading

Collective cell migration is most commonly modelled through continuum models based on linear Fickian diffusion, resulting in cell density descriptions of the form,

$$\frac{\partial C}{\partial t} = D_0 \nabla^2 C \,, \tag{1}$$

[1, 15, 17–19] where C is the probability of a cell being located at a certain position (x, y, z) at time t. The linear diffusion equation is one of the most commonly used partial differential equations (PDEs) in applied mathematics and is considered the single most important prototype for parabolic PDEs, finding applications in a wide range of fields including physics and biology. A more generalised nonlinear model,

$$\frac{\partial C}{\partial t} = D_0 \nabla \cdot [D(C) \nabla C] \,, \tag{2}$$

with $D(C) = C^n$ and $n > 0$, has also been suggested by several authors [1, 16–18, 20] to describe collective cell migration. This parabolic differential equation is known as the porous medium equation (PME) [21]. It is degenerate as $D(0) = 0$ and thus $D(C) = C^n$ does not satisfy the condition for classical diffusion equations, $D(C) > 0$ [22]. The PME is less well known than Eq. (1) and more difficult to deal with, given its non-linearity, with applications including fluid flow, heat transfer and diffusion [21]. Several ecological and biological uses have also arisen such as the representation of population pressure in biological systems [23].

Although numerous examples exist where the PME is applied to cell motility problems, there is little coherence between approaches. Some authors favour the use of the PME as the solutions often have distinct boundaries known as interfaces [21], thought to represent sharp-fronted cell density profiles [1]. Sherrat and Murray [17] and Sengers et al [18] justify the use of the non-linear diffusion equation through the fit to experimental data. No widely accepted mathematical or physical justification yet exists for the appropriate use of the PME and choice of the exponent, n. This report explores the approach taken by Simpson, Baker and McCue [16], where the use of the PME to describe the collective spreading of cells of varying aspect ratio is justified by considering the continuum limit of discrete motility mechanisms.

Collective cell spreading can be represented through the use of discrete interacting random walk models known as exclusion processes [16]. Simulations of these lattice-based random walk models provide a microscopic set of data equivalent to a photograph of the distribution of cells. The average of many such microscopic realisations provides a continuous description of the system which can be related to the solution of a PDE. Most existing applications of exclusion process models consider each cell to be represented by a single agent occupying a single lattice site [6–13]. Given the uniform nature of such lattices, these models are only appropriate for perfectly "round" cells where the longitudinal and transverse dimensions are equal. Yet, many of the cells relevant in wound healing applications and cancer biology are elongated or rod-shaped.

In Sec. 2 the collective motion of an initially adjacent population of cells is considered. Exclusion process models are developed in Sec. 2.1 for populations of cells with various lengths. Note that since the models are in 1D it is implied that the cells have a unit "thickness" with length denoted by L. The exclusion process models are simulated on a lattices with spacing given by Δ. Each lattice site is indexed i where $i \in \mathbb{N}$, and each site has a position $x = i\Delta$. The lattice's length is defined by the x coordinate where $1 \leq x \leq x_{\max}$. A population of identical interacting agents is considered in which all agents occupy a number of lattice sites given by their length, L. For example, an agent with $L = 2$ may occupy the lattice sites i and $i + 1$ as shown in Fig. 1. PDE models are developed in Sec. 2.2 by considering the continuum limit of the discrete models, providing a motivation for the use

of the PME. The discrete models and solutions to the PDEs are compared in Sec. 2.3.

The use of the Taylor series in developing the continuum models in Sec. 2.2 is analysed in Sec. 3 by considering the steady state of the spreading agent population. Steady state solutions to the continuum model are presented in Sec. 3.1 while the steady state exclusion process model is discussed in Sec. 3.2. The two approaches are compared in Sec. 3.3 and the effect of the Taylor series expansion is discussed.

Although modelling collective cell migration in one dimension may appear trivial, such models still have realistic applications. Fibroblast movement in tissue culture is often seen to be focused along grooves [4] which can in the simplest case be modelled by a one dimensional lattice.

2 Collective motion of an initially contiguous population of agents

2.1 Discrete simulations

To simulate the collective spreading of cells, a random walk model was created for agents with $L = 1, 2, 3$ and 4. Each simulation begins with a contiguous population of N identical agents occupying the central 40 lattice sites where the number of lattice sites occupied by a single agent depends upon L. Fig. 1 illustrates the number of sites occupied by agents with $L = 1$ and 2 respectively. Initially, the agent with $L = 1$ occupies lattice site i whilst the agent with $L = 2$ occupies i and $i + 1$.

A random sequential update method is applied where for each time step, of duration

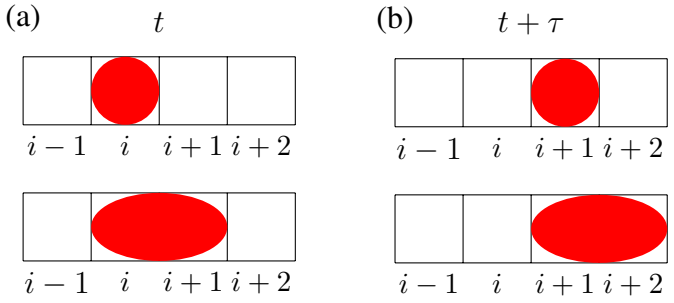

Figure 1: Schematic of the lattice positions occupied by an agents with $L = 1$ and 2. (a) Agents positions at time t before the motility event. (b) New agent positions after one time step of duration τ, given the opportunity to move right, unhindered by other agents.

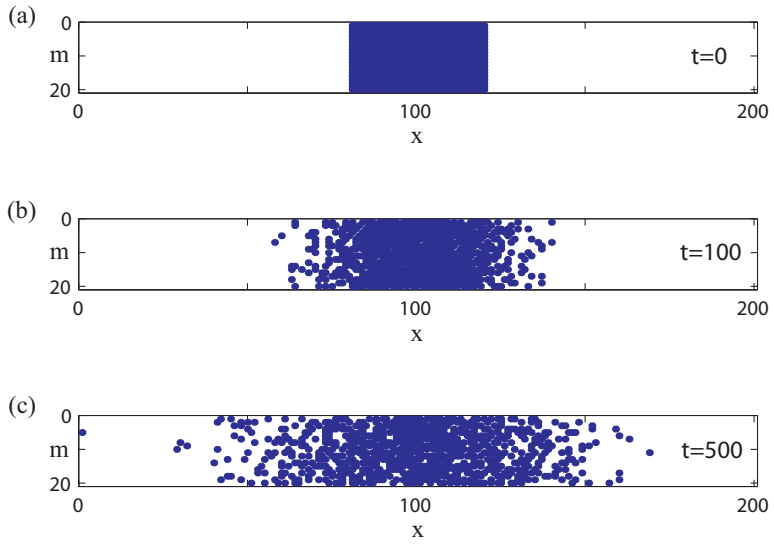

Figure 2: Simulations of agent motility showing the horizontal positions of agents with $L = 1$ along a lattice of length 200 for $t = 0, 100$ and 500. Each plot has a height of 20 indicating the number of identically prepared realisations, M where each new row shows the mth realisation. (a) Simulations begin with a contiguous population of 40 agents per row, occupying the centre of the lattice. (b)-(c) Simulation results are given at $t = 100$ and 500 respectively for $P_m = \Delta = \tau = 1$.

τ, N agents are selected and given the opportunity to move in the positive or negative x direction. The direction is determined at random for each agent by generating a random value from the standard uniform distribution, p_d. Agents attempt to move in the negative x direction when $p_d < 0.5$ and in the positive x direction when $p_d \geq 0.5$. When chosen, an agent attempts to move a distance Δ, with probability $P_m \in [0, 1]$ in the time interval τ. All simulations are conducted with $P_m = \Delta = \tau = 1$. Fig. 1 shows the positions of agents with $L = 1$ and 2 after one time step, given the opportunity to move to the right. Due to the contact inhibition of cells, the motility event is aborted when the target site is occupied. Thus, the initial configuration of the system yields only two possible motility events. Either the agent on the left boundary moves left or the agent on the right boundary moves to the right. Since the lattices used for the simulations are long enough that the agents do not reach the boundaries in the simulated time-frames, they can essentially be considered infinite.

Fig. 2 shows a set of simulations for a lattice $1 \leq x \leq 200$ occupied by agents with

$L = 1$. Results are shown for 20 independent 1D lattices, stacked vertically, to give the appearance of two dimensions. Thus each new "row" shows the results of a separate microscopic realisation, m, given the same length of time and initial conditions. The total number of realisations for each set of simulations is given by $M = 20$. All lattice sites between 81 and 220 were initially occupied as shown in (a). The system was allowed to evolve, resulting in a spreading of the population over time, as would be expected from a population of cells.

From Fig. 2 it is evident that each microscopic realisation results in a different distribution of agents. In order to obtain any meaningful results describing the collective motion of the cells, a smooth distribution of the occupancy of sites is required. Suppose $C^m(i)$ is the occupancy of site i during the mth realisation, where $C^m(i) = 0$ if the lattice site is vacant and $C^m(i) = 1$ if it is occupied. For M realisations, starting from the same initial condition, the average occupancy of a site is,

$$\langle C(x,t)\rangle = \frac{1}{M}\sum_{m=1}^{M} C^m(i)\,. \tag{3}$$

The average occupancy, $\langle C(x,t)\rangle$, describes the probability of position x being occupied after a time t. This raises the question; how many realisations are required such that a continuous probability distribution is achieved? Clearly one realisation will just give a binary plot, from which it would be impossible to deduce anything by inspecting the resulting density profile. Fig. 3 shows the smoothing out of the density profiles for agents with $L = 1$ and $L = 4$ as the number of realisations is increased. For $M = 50$ the density data is highly unreliable, displaying pronounced fluctuations. This effect is particularly noticeable near the leading edges of the profile as there are very few agents in this region. It is clear that a large number of simulations, $M > 5000$, is required to achieve a continuous distribution. The difference between $M = 5000$ and $M = 10000$ is difficult to see at this scale although in (c) and (d) it would appear as though the average density profile for $M = 10000$ is slightly smoother. Further increasing M would yield a smoother distribution, however, the visual improvement would not be significant. Hence, comparisons of the averaged profile and solutions in Sec. 2.3 were conducted using $M = 8000$.

While it is possible to fit a linear diffusion model to the data [17], as each density profile remains symmetric about the central lattice position, this approach fails to explain the difference in motility or address both the increased spread and change in shape of the profiles. Since the agent density profiles are quite different, see Figs. 3 (a) and (b), this model calibration approach would lead to different estimates of the free cell diffusivity, D_0, for every value of L. Rather than using this calibration procedure, in Sec. 2.2 different values of L are related to different diffusion mechanisms with nonlinear diffusivity functions, $D(C)$, while D_0 remains the same.

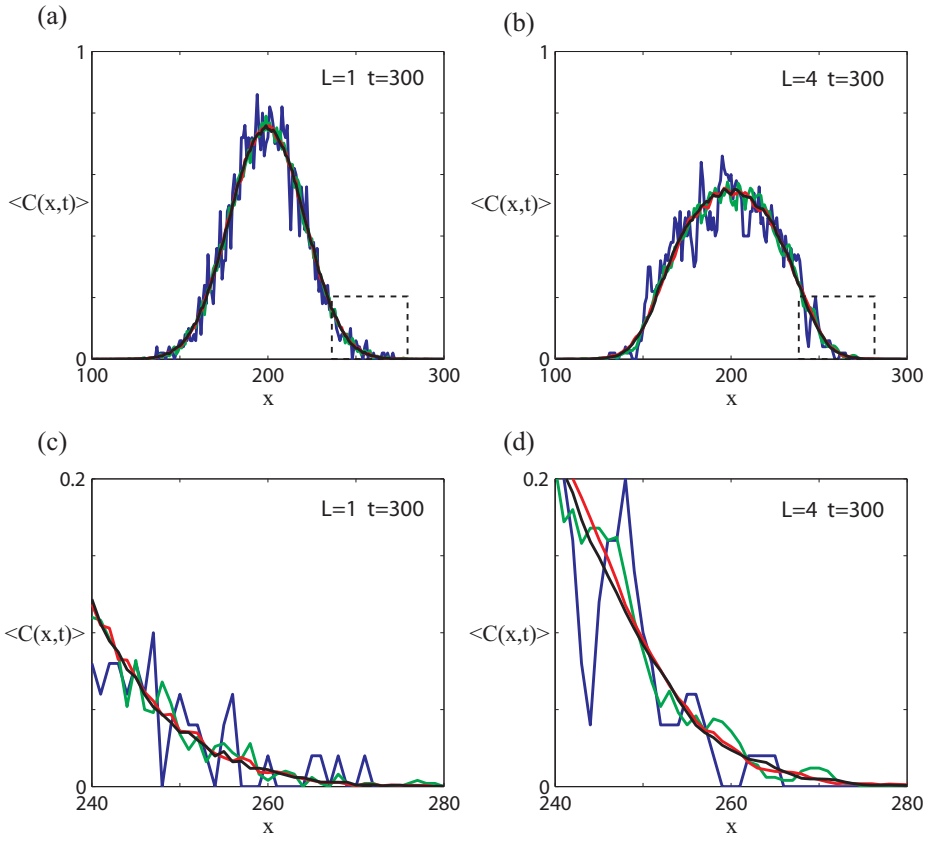

Figure 3: Convergence of the mean agent occupancy profiles obtained using Eq. (3) for a lattice of length 400 with sites $181 \leq x \leq 220$ *initially occupied. All simulations were performed until* $t = 300$ *for* $P_m = \Delta = \tau = 1$*. Four profiles are shown with* $M = 50, 500, 5000$ *and 10000 shown in blue, green, red and black, respectively. In (a) results are provided for a lattice occupied by agents with* $L = 1$ *whereas the lattice corresponding to (b) is occupied by agents with* $L = 4$*. The details of the probability profiles in the tail regions between* $240 \leq x \leq 280$ *are shown in (c) and (d).*

2.2 Continuum models

In Sec. 2.1 the discrete mechanism was connected to a continuous expression by averaging over many statistically identical realisations to obtain $\langle C_i \rangle \in [0, 1]$. This average represented the probability of an agent occupying site i. To develop equations governing the motility for different lengths, discrete conservation statements are constructed describing $\delta\langle C_i \rangle$, the change in average occupancy of a site i during the interval from t to $t + \tau$. Four possible

motility events must be considered in order to form this statement; movement from the left or right into site i and movement to the left or right out of site i. When constructing the discrete conservation statements it is assumed that the occupancy of the lattice sites is independent. Although this assumption seems counter-intuitive, it is the foundation of many mean field models in the exclusion process literature [7, 8, 10, 16].

For a population of agents with $L = 1$ the discrete conservation statement is,

$$\begin{aligned}\delta\langle C_i\rangle =& \frac{P_m}{2}\left[\langle C_{i+1}\rangle(1-\langle C_i\rangle) + \langle C_{i-1}\rangle(1-\langle C_i\rangle) - \langle C_i\rangle(1-\langle C_{i+1}\rangle)\right.\\ &\left.-\langle C_i\rangle(1-\langle C_{i-1}\rangle)\right] .\end{aligned} \tag{4}$$

where $P_m/2$ is the probability per time step that an agent will move in a given direction, $\langle C_{i\pm\beta}\rangle$ is the probability of site $i \pm \beta$ being occupied and $(1 - \langle C_{i\pm\beta}\rangle)$ is the probability that site $i \pm \beta$ is vacant. The four possible motility events are represented by the product of these probabilities, where the sign indicates whether the corresponding motility event increases or decreases the occupancy of site i. The first term on the right of Eq. (4), $\langle C_{i+1}\rangle(1 - \langle C_i\rangle)$, therefore represents the change in average occupancy of site i caused by an agent moving from site $i + 1$ to i. Simplifying Eq. (4) and dividing by the time increment, τ, results in,

$$\frac{\delta\langle C_i\rangle}{\tau} = \frac{P_m}{2\tau}\left[\langle C_{i+1}\rangle - 2\langle C_i\rangle + \langle C_{i-1}\rangle\right] . \tag{5}$$

Taking the limit as $\Delta \to 0$ and $\tau \to 0$ simultaneously, with discrete values of $\langle C_i\rangle$ written in terms of a continuous variable $C(x, t)$ allows the discrete conservation statement to be transformed into a PDE model. To obtain the appropriate relationship, terms in Eq. (5) are expanded in a Taylor series about i according to,

$$C_{i+a} = C_i + (a\Delta)\frac{\partial C_i}{\partial x} + \frac{(a\Delta)^2}{2}\frac{\partial^2 C_i}{\partial^2 x} + \mathcal{O}\left(\Delta^3\right) , \tag{6}$$

where $a \in \mathbb{Z}$. This truncation follows from the fact that terms of $\mathcal{O}\left(\Delta^3\right)$ and higher are negligible. Taking the limits as $\Delta \to 0$ and $\tau \to 0$ simultaneously, with (Δ^2/τ) held constant, results in the expression,

$$\frac{\partial C}{\partial t} = \frac{\Delta^2 P_m}{2\tau}\frac{\partial^2 C}{\partial x^2} \tag{7}$$

which is the linear diffusion equation with constant diffusivity,

$$D_0 = \frac{P_m}{2}\lim_{\Delta,\tau\to 0}\left(\frac{\Delta^2}{\tau}\right) . \tag{8}$$

Postal Address: 111 Barry Street
c/- The University of Melbourne
Victoria 3010 Australia

Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 9349 4106
Web: www.amsi.org.au

The same process is repeated for agents with $L > 1$. Firstly, the case for agents with $L = 2$ is considered. Following the same process as above, again making the assumption that the occupancy status of adjacent sites is independent, the conservation statement may be written as,

$$\begin{aligned}\delta\langle C_i\rangle =& \frac{P_m}{2}\left[(1-\langle C_i\rangle)\langle C_{i+1}\rangle\langle C_{i+2}\rangle + (1-\langle C_i\rangle)\langle C_{i-1}\rangle\langle C_{i-2}\rangle\right.\\ &\left. - (1-\langle C_{i+2}\rangle)\langle C_i\rangle\langle C_{i+1}\rangle - (1-\langle C_{i-2}\rangle)\langle C_i\rangle\langle C_{i-1}\rangle\right] . \end{aligned} \tag{9}$$

Since each agent now occupies two lattice sites, see Fig. 1(b), each term in the sum on the right is now the product of three probabilities of agent occupancy. The term:$(1 - \langle C_i\rangle)\langle C_{i+1}\rangle\langle C_{i+2}\rangle$, therefore represents the probability that an agent occupying sites $i + 1$ and $i + 2$ will move to the left into sites i and $i + 1$. Simplifying and dividing by τ gives,

$$\frac{\delta\langle C_i\rangle}{\tau} = \frac{P_m}{2\tau}\left[\langle C_{i+1}\rangle\langle C_{i+2}\rangle + \langle C_{i-1}\rangle\langle C_{i-2}\rangle - \langle C_i\rangle\langle C_{i+1}\rangle - \langle C_i\rangle\langle C_{i-1}\rangle\right] . \tag{10}$$

Again, the Taylor series expansion in Eq. (6) is used to expand terms on the right of Eq. (10). Taking the limits as $\Delta \to 0$ and $\tau \to 0$ with (Δ^2/τ) held constant and replacing $\langle C_i\rangle$ with $C(x, t)$ gives,

$$\frac{\partial C}{\partial t} = \frac{2\Delta^2 P_m}{\tau}\left[C\frac{\partial^2 C}{\partial x^2} + \left(\frac{\partial C}{\partial x}\right)^2\right] , \tag{11}$$

which may be rewritten as a PME of the form,

$$\frac{\partial C}{\partial t} = \frac{\Delta^2 P_m}{2\tau}\frac{\partial}{\partial x}\left[4C\frac{\partial C}{\partial x}\right] . \tag{12}$$

This time the diffusivity term has both the linear term, D_0, given by (8) and a non-linear diffusivity function, $D(C) = 4C$. The same derivations for agents with $L = 3$ and 4 also results in PMEs with free cell diffusivities given by (8), and nonlinear diffusivity functions $D(C) = 9C^2$ and $D(C) = 16C^3$ respectively. The general form of the PDE, for all L, is

$$\frac{\partial C}{\partial t} = D_0\frac{\partial}{\partial x}\left[D(C)\frac{\partial C}{\partial x}\right] , \tag{13}$$

$$D(C) = L^2 C^{L-1} , \tag{14}$$

where Eq. (13) is a PME with $n = L - 1$.

2.3 Continuum-discrete comparison

To test how the PDE model predicts the movement of a population of agents, the agent density profiles are compared to solutions of Eq. (13). The initial condition, describing the population of contiguous agents in the centre of the lattice, is

$$C(x,0) = \mathrm{H}(x - x_L) - \mathrm{H}(x - x_R)\,, \tag{15}$$

where x_L and x_R represent the left and right boundaries of the initial agent population and H is the *Heaviside* function. When $x_L \leq x \leq x_R$, $\mathrm{H}(x - x_L) = 1$ and $\mathrm{H}(x - x_R) = 0$, giving $C(x,0) = 1$. The boundary values are chosen to match the discrete simulations, therefore $(x_L, x_R) = (181, 220), (181, 220), (181, 119)$ and $(181, 220)$ for $L = 1, 2, 3$ and 4 respectively.

For $L = 1$ the collective migration of agents is described by the linear diffusion equation (Eq. (7)) with initial condition (15). Applying the Fourier transform, solving the resulting ODE and using convolution in conjunction with the inverse transform results in the exact solution,

$$C(x,t) = \frac{1}{2}\left[\mathrm{erf}\left(\frac{x - x_L}{2\sqrt{D_0 t}}\right) - \mathrm{erf}\left(\frac{x - x_R}{2\sqrt{D_0 t}}\right)\right]. \tag{16}$$

For agents with $L > 1$, where the motility is described by a non-linear PDE, Eq. (13), solutions must be found numerically. Applying a first order finite difference approximation to Eq. (13) with constant grid spacing δx and constant time steps δt gives,

$$\frac{C_i^{k+1} - pC_i}{\delta t} = \frac{D_0}{\delta x}\left[D(C_{i+\frac{1}{2}})^k \left.\frac{\partial C^{k+1}}{\partial x}\right|_{i+\frac{1}{2}} - D(C_{i-\frac{1}{2}})^k \left.\frac{\partial C^{k+1}}{\partial x}\right|_{i-\frac{1}{2}}\right], \tag{17}$$

where C_i^k is known value at the next time step, C_i^{k+1} is unknown at the next time step and pC_i is known value at the previous time step. $D(C)$ and $\partial C/\partial x$ are approximated midway between lattice points, as shown in Fig. 4. Thus,

$$D(C_{i\pm\frac{1}{2}}) = \frac{D(C_i) + D(C_{i\pm1})}{2}\,, \tag{18}$$

allowing Eq. (13) to be approximated using,

$$\begin{aligned}\frac{C_i^{k+1} - pC_i}{\delta t} =& \frac{D_0}{2(\delta x)^2}\Big[(D(C_{i+1}) + D(C_i))^k\,(C_{i+1} - C_i)^{k+1} \\ &- (D(C_i) + D(C_{i-1}))^k\,(C_i - C_{i-1})^{k+1}\Big]\,.\end{aligned} \tag{19}$$

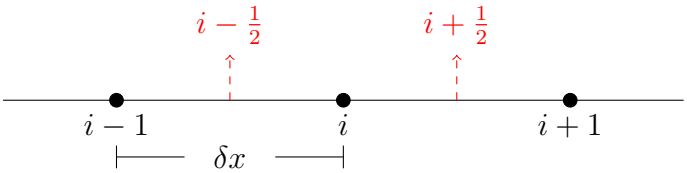

Figure 4: 1D mesh used to approximate $D(C)$ and $\partial C/\partial x$ at points midway between the lattice sites.

Each location on the lattice results in an equation of the form,

$$a(C_{i-1}^k)C_{i-1}^{k+1} + b(C_i^k)C_i^{k+1} + c(C_{i+1}^k)C_{i-1}^{k+1} = d(pC_i)\,. \tag{20}$$

The solution can therefore be obtained by solving the nonlinear algebraic system $\mathbf{A}\mathbf{C}^{m+1} = \mathbf{d}$, where $\mathbf{A}$ is the tridiagonal matrix containing the coefficients a, b and c, and $\mathbf{d}$ is the right hand side vector containing entries of the form $(-pC_i/\delta t)$. Picard linearisation, with 5 iterations per time-step, is used to approximate a, b, c and d for each time step with the approximated linear tridiagonal system solved via the Thomas algorithm.

Comparison of the discrete density profiles in Fig. 5 indicates a pronounced difference in the evolution of the averaged agent density profiles with changing L. The agent density near the centre of the lattice ($x = 200$) decreases faster as L increases. Although more rapid spreading of large agents seems counter-intuitive, the observation can be easily justified as the movement of larger agents will affect the population density over a greater portion of the lattice relative to the same motility events acting on a population of smaller agents. Consider the difference between the attempted movement in the positive x direction of agents with $L = 1$ and $L = 4$. The movement of the agent with $L = 1$, initially at site i, will alter the density of sites i and $i + 1$. In contrast, an agent with $L = 4$, initially occupying sites $i - 3$, $i - 2$, $i - 1$ and i, will alter the density at sites $i + 1$ and $i - 3$, which are much further apart. Although this example illustrates why the density profiles evolve more rapidly with increasing values of L, it does not provide a physical explanation for why cells of greater length would spread more rapidly. Given the dimensionless implementation of the discrete and continuum models, the difference in the evolution of density profiles cannot necessarily be interpreted as "faster" or "slower" in a physical sense.

Fig. 5(a) shows that there is an excellent match between the simulation data for agents with $L = 1$ and the solution to Eq. (13), as has been demonstrated previously [7]. For $L > 1$ the solutions to Eqs. (8), (13) and (14) also appear to match the simulation data reasonably well, especially near the centre of the lattice. However, the match at the boundary of the spreading agent population for $L > 1$ is fairly poor, as the simulation data is smooth whereas

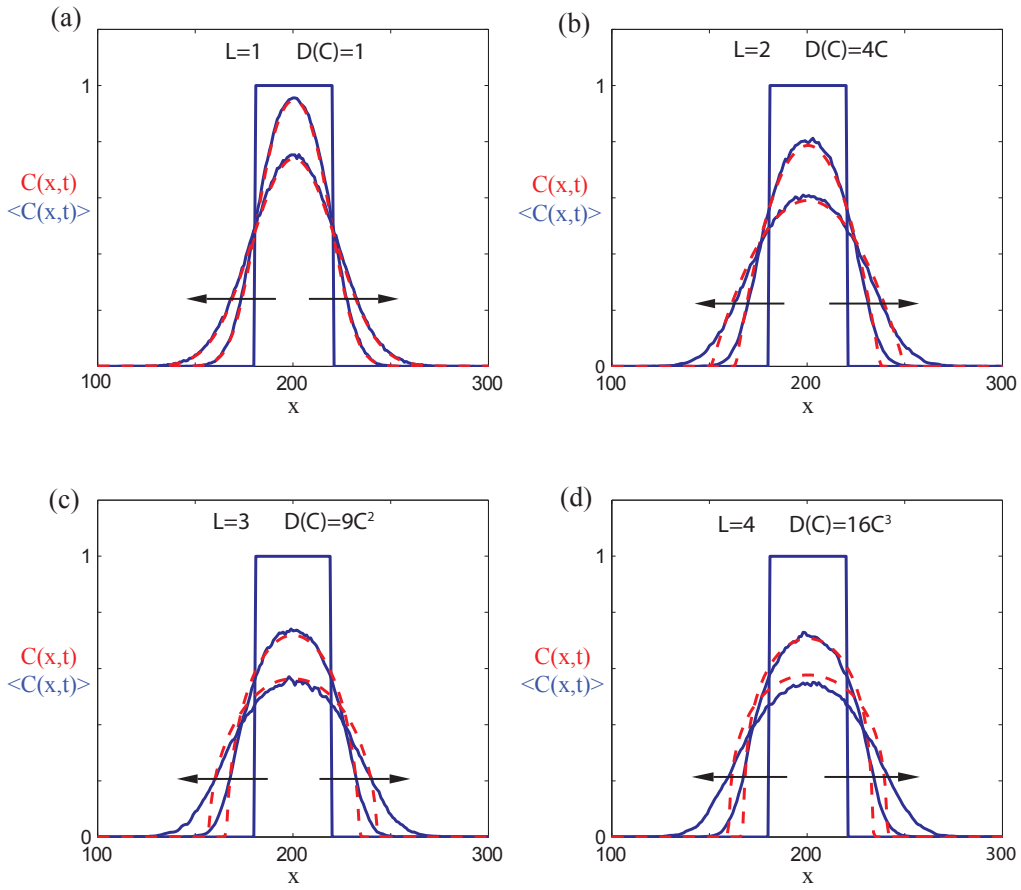

Figure 5: Averaged agent density data for a series of simulations for $L = 1, 2, 3, 4$ respectively. All sites with $181 \leq x \leq 220$, $181 \leq x \leq 220$, $181 \leq x \leq 219$, $181 \leq x \leq 220$ are initially occupied. In each case the agent occupancy, $\langle C(x,t) \rangle$ is shown in blue (solid) at $t = 0, 100, 500$ averaged over $M = 8000$ identically prepared realisations. The simulation profiles are compared with the numerically obtained solution of Eq. (13) *shown in red (dotted), with the arrows indicating the direction of increasing time.*

the solution of the PME (13) is sharp-fronted. This mismatch becomes more pronounced with increasing agent length.

At least two explanations may be provided for the the discrepancy between the continuum and discrete profiles [16] at the peripheries of the spreading agent population. The first stems from the use of the truncated Taylor series expansion, Eq. (6), to relate the occupancy of lattice sites. In doing so it was assumed that the corresponding solution, $C(x,t)$ would

be sufficiently differentiable so that the density profile could be expanded via the Taylor series. Although the solution of the PDE model is differentiable everywhere for $L = 1$, the sharp-fronted solutions of the degenerate nonlinear diffusion equation [21] which result when $L > 1$, are not. Hence this would appear to be a plausible explanation for the discrepancy between the continuum and discrete models. The effect of the Taylor series expansion upon the mismatch between the models is further investigated in Sec. 3.

The second possible cause is the failure of the independence assumptions upon which the discrete conservation statements, Eqs.(4) and (9), are based. Although this assumption, made by others [7, 8, 10, 16], allows the products of the occupancy probabilities to be interpreted as a net transition probability, it becomes unreasonable when $L > 1$. Consider the difference between agents with $L = 1$ and $L = 4$. The agent with $L = 1$ initially occupies site i whereas the agent with $L = 4$ occupies $i - 3$, $i - 2$, $i - 1$ and i. If these agents attempt to move in the positive x direction the discrete conservation statement requires that, when averaged over many identically prepared realisations, for $L = 1$ the occupancy of sites i and $i + 1$ are independent whilst for $L = 4$ the occupancy of sites $i - 3$, $i - 2$, $i - 1$, i and $i + 1$ are independent. Since there are many more ways in which an agents with $L = 4$ may occupy these sites over a large number of realisations this independence assumption is less satisfactory for $L = 4$ than for $L = 1$. In general, the assumption appears to become less appropriate as L increases and may therefore contribute to the increased discrepancy as L increases. Work is being done to understand how to relax the independence assumption for spatially uniform problems [24], however further investigation is required to understand how this assumption may be relaxed.

3 Steady state analysis

3.1 Steady state solutions

Completely smooth solutions to Eqs. (13) and (14) are sought for $L > 1$ in order to determine whether the Taylor series expansion, Eq. (6), is the cause of the discrepancy noted at the boundaries of the spreading agent population. Such solutions are achieved at the steady state of a cell population's collective spreading, as demonstrated by the smoothness of the steady state profiles in Fig. 6. The steady state solutions of Eqs. (13) and (14) are obtained

by setting the right hand side of Eq. (13) equal to zero and solving by direct integration,

$$0 = D_0 \frac{\partial}{\partial x}\left[D(C)\frac{\partial C}{\partial x}\right] \tag{21}$$

$$\int a \, \mathrm{d}x = \int D(C) \, \mathrm{d}C \, . \tag{22}$$

Applying Eq. (14), the resulting steady state distribution is,

$$C(x) = (a_2 x + b_2)^{\frac{1}{L}} \ , \tag{23}$$

where $a_2 = a/L$ and $b_2 = b/L$. Since a spreading agent population, regardless of its initial distribution, will reach the boundaries of a lattice of any length given a long enough time, it is necessary to enforce boundary conditions. This allows the constants a_2 and b_2 to be determined. If $C(0) = 1$ and $C(x_{max}) = d$, then $b_2 = 1$ and $a_2 = 1/x_{max}\left(d^L - 1\right)$.

For the solutions and random walk simulations presented the boundary conditions,

$$C(0) = 1 \tag{24}$$

$$C(x_{max}) = 0.2 \, , \tag{25}$$

are applied, whilst the initial condition considered is,

$$C(x, 0) = 0.6 \, . \tag{26}$$

To determine when the steady state of the spreading agent population is reached, solutions to the IBVP given by Eqs. (13), (24), (25) and (26), are found for a range of times and compared to the steady state solutions. The general solution for $L = 1$ can be solved for exactly, giving,

$$C(x,t) = 1 - \frac{8}{10x_{max}}x + \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{4}{10n}\left((-1)^{n+1} - 1\right)\mathrm{e}^{(n^2\pi^2 D_0 t/x_{max}^2)}\sin\left(\frac{n\pi}{x_{max}}x\right) \ . \tag{27}$$

When $L > 1$ the solutions are found numerically as described in Sec. 2.3, but with the initial and boundary conditions, Eqs. (24)-(26). Although the initial distribution has no bearing upon the steady state solution, it will affect the numerical solutions prior to steady state as well as impacting the rate at which solutions converge to steady state. The longer the lattice the greater the time-frame required to reach steady state.

As illustrated in Fig. 6, for a lattice with $x_{max} = 100$, with boundary conditions and initial conditions (24)-(26), a steady state of the spreading agent population is reached when

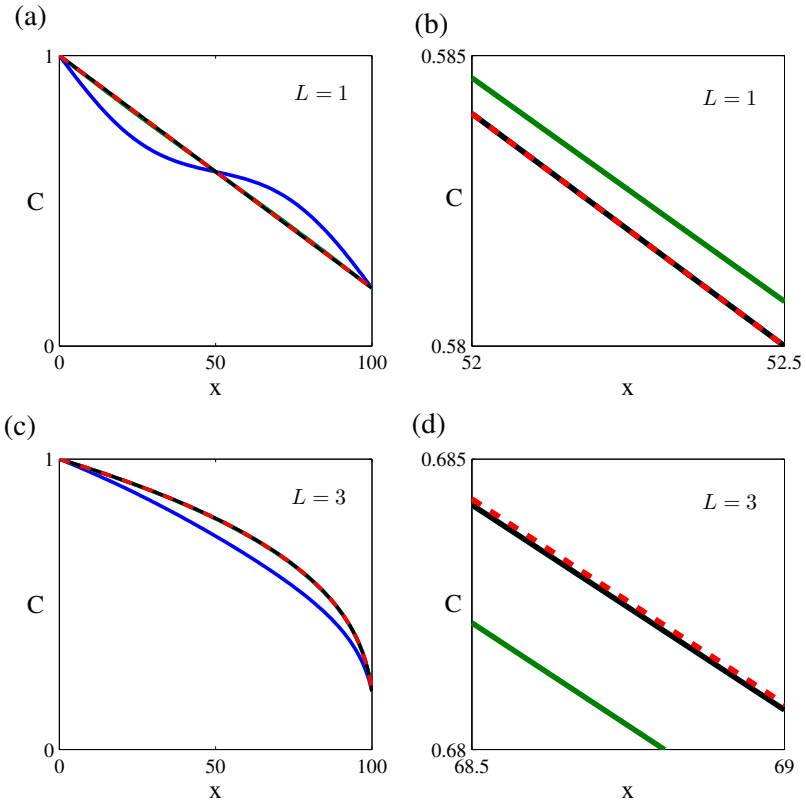

Figure 6: Comparison of the steady state agent density profiles (red dotted) with density profiles at $t = 500, 2000, 8000$ *and* 10000 *(blue, green, cyan and black respectively) for a lattice of length 100. Exact solutions to* (13) *and* (24)-(26) *are displayed in (a) and (b) for* $L = 1$ *while numerical solutions for* $L = 3$ *are shown in (c) and (d). (b) and (c) are magnified views of sections of (a) and (c) respectively.*

$t > 8000$. Even at the high resolution depicted in Figs. 6(b) and 6(d) no difference can be determined between the solution when $t = 8000$ (cyan) and $t = 10000$ (black). For $L = 3$, the solution at $t = 10000$ does not match the steady state solution as well as for $L = 1$. This is because exact solutions are shown for $L = 1$ while for $L = 3$ numerical solutions, obtained using the method described in Sec. 2.3, are shown. A greater number of iterations per time-step during the Picard linearisation would have yielded a more accurate match but since the solutions at $t = 8000$ and $t = 10000$ were indistinguishable it was apparent that a steady state had been reached.

3.2 Discrete steady state simulations

To determine whether each steady state solution, Eq. (23), matches the steady state of the corresponding spreading agent population, the exclusion process model described in Sec. 2.1 must be suitably modified. While the steady state of a spreading agent population has no real physical applications, it is expected to allow the validity of the continuum models presented in Sec. 2.2 to be analysed.

Rather than beginning with a contiguous population of agents in the centre of the lattice, as in Sec. 2.1, the steady state simulations begin with the uniform initial distribution, Eq. (26). Thus, the entire lattice is initially occupied with probability 0.6 i.e. each site has a 60% chance of being occupied by an agent. A less populated initial lattice requires more time steps to reach the steady state while a fully populated initial lattice does not allow enough motility events to be assessed.

To enforce the boundary conditions, (24) and (25), a further condition must be introduced to the random walk simulations. After each time step the occupancy of the left hand boundary position, $i = 1$, is assessed. If site 1 is found to be unoccupied it is occupied with another agent. The right hand side boundary condition is enforced in a similar manner except that for each time step, a random number, p_R, is selected from the standard uniform distribution. If $p_R \leq 0.2$ then site $i = x_{max}$ must be occupied whereas if $p_R > 0.2$ the site will be vacated.

Since agents may attempt to move beyond the lattice boundaries during the random walk simulations, further boundary conditions must be enforced during each time step. Here a choice needs to be made between reflecting and non-reflecting boundary conditions. Reflecting boundary conditions consider the lattice bounds to be impenetrable 'walls', thus when an agent occupies sites 1 or x_{max}, and the exclusion process model selects an attempted movement to a site beyond the lattice, the motility event is aborted. This introduces a motility bias at the lattice bounds. In contrast, non-reflecting boundary conditions would allow agents to move beyond the lattice, resulting in fractions of agents with $L > 1$ occupying the boundaries.

The choice between reflecting and non-reflecting boundaries impacts the way in which boundary conditions (24) and (25) are implemented. Consider the case for reflective boundaries. For agents with $L > 1$, when $C(0) = 1$ is enforced sites 1 to L must be occupied. Thus, if there is another agent occupying any sites from 2 to L it must be removed completely to avoid introducing fractions of agents onto the lattice. Fig. 7 demonstrates how enforcing these boundary conditions for $L = 2$ will introduce an artificial vacancy at site $i = 2$. Implementing the boundary conditions in this way therefore results in pronounced fluctuations in occupancy near the lattice boundaries when $L > 1$. Similarly for the boundary condition at site x_{max}, removing an agent with $L > 1$ when $p_R > 0.2$ results in a decrease in occupancy

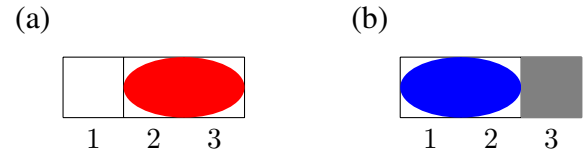

Figure 7: Schematic of the lattice positions occupied after enforcing Eq. (24) *for a lattice with reflective boundaries, when* $L = 2$*. The site in gray is vacated to enforce the boundary condition, causing a marked decrease in the occupancy of this site over many trials. (a) Agent position before implementing the boundary condition. (b) Positions occupied by new agent after enforcing the boundary condition, Eq .*(24)*, with the vacated site shown in gray.*

for sites $x_{max} - L + 1$ to $x_{max} - 1$. The vacancies caused by this approach yield too many fluctuations to accurately compare the simulations and steady state solutions, as is evident from Fig. 8(b). Simulations with non-reflecting boundary conditions are instead considered. Since non-reflecting boundary conditions allow agents to move beyond the lattice, implementing conditions (24) and (25) for $L > 1$ does not require a previously occupied site to be vacated. Instead, fractions of agents may now occupy the lattice at the boundaries avoiding the marked fluctuations at the boundaries. However, enforcing non-reflective boundaries introduces new difficulties. If, during the random walk simulations, only the lattice sites are considered, then when an agent moves beyond the lattice at either end it cannot return onto the lattice. This results in a marked decrease in lattice occupancy for sites close to the boundaries, despite the reintroduction of agents after each time-step, as illustrated by the green plots in Fig. 8. In order simulate the collective motion of the agents, without introducing a bias at either lattice boundary, ghost sites must be introduced at either end. Boundary conditions (24) and (24) are applied not only to the boundary sites but also to the ghost sites beyond the lattice bounds. The introduction of these ghost nodes allows agents to move beyond the lattice bounds during a time-step but also allows agents to move back onto the lattice. Two difficulties now arise with regards to the use of the ghost sites; how many are necessary to obtain a smooth distribution? and what conditions should be enforced for the outermost ghost sites? Again, using non-reflecting conditions at the boundary of the ghost nodes results in agents moving off but not back onto the ghost lattice while reflecting boundaries still introduce some fluctuations. For the following simulations the bounds of the ghost sites are considered to be fixed while the lattice bounds are non-reflecting. At least $4L$ ghost sites must be considered beyond each lattice bound to avoid the fluctuations evident in Fig. 8(b).

Enforcing the right hand boundary condition for $L > 1$ requires yet more consideration. While the left hand boundary condition can be simply enforced by testing the occupancy

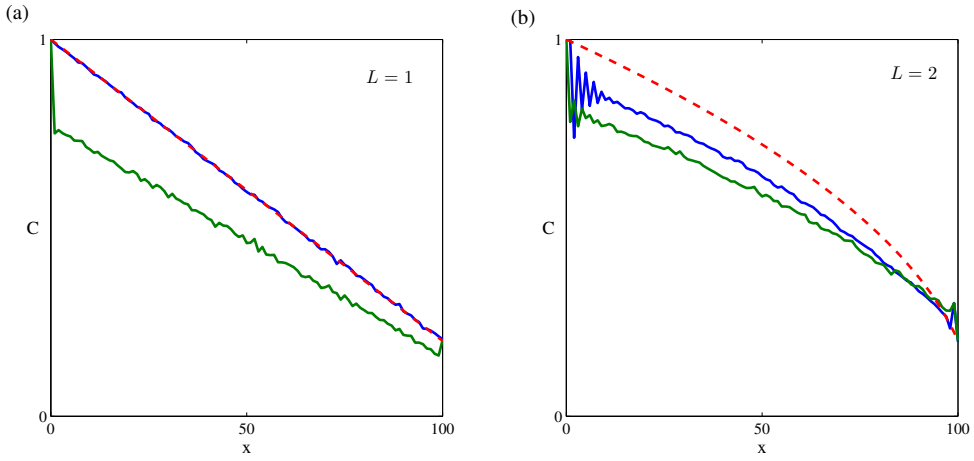

Figure 8: Comparison of the discrete steady state simulations performed without ghost nodes and the steady state solutions (red dotted) for agent populations with $L = 1$ and 2, in (a) and (b) respectively. Discrete simulations were conducted with $M = 10000$, $t = 9000$, $C(x, 0) = 0.6$ and boundary conditions (24) and (25). Simulations with reflecting boundary conditions are shown in blue while simulations with non-reflecting conditions are shown in green.

and introducing an agent to fill only the first site if it is empty, the same simplistic approach can't be used for the right hand boundary. If $p_R > 0.2$ and the right hand boundary position is occupied then the entire agent must be removed. Thus for agents with $L > 1$, sites $x_{max} - L + 1$ to $x_{max} - 1$ may also be vacated. This causes a pronounced drop in occupancy for those sites which is not considered in the solutions to the PDE model. To combat this drop in probability the right hand boundary condition is extended to these sites for a fraction of instances when sites to the left of x_{max} are empty.

3.3 Continuum-discrete comparison at steady state

A comparison of the discrete and continuum steady states is shown in Fig. 9. It is clear that the match between the discrete and continuum models is far better than for the simulations without ghost sites, presented in Sec. 3.2. Again, as would be expected from Sec. 2.3, the discrete and continuum models match perfectly when $L = 1$. The discrete and continuum models also appear to match well for $L = 2$, although a slight discrepancy is noted when $x > 50$. For $L = 3$ the mismatch is more pronounced. Given the previous difficulty with the implementation of boundary conditions this is likely to be result of how the boundary condition at the right hand side is implemented. For $L = 4$ the match again appears very good although there is a pronounced fluctuation at $x = 97$.

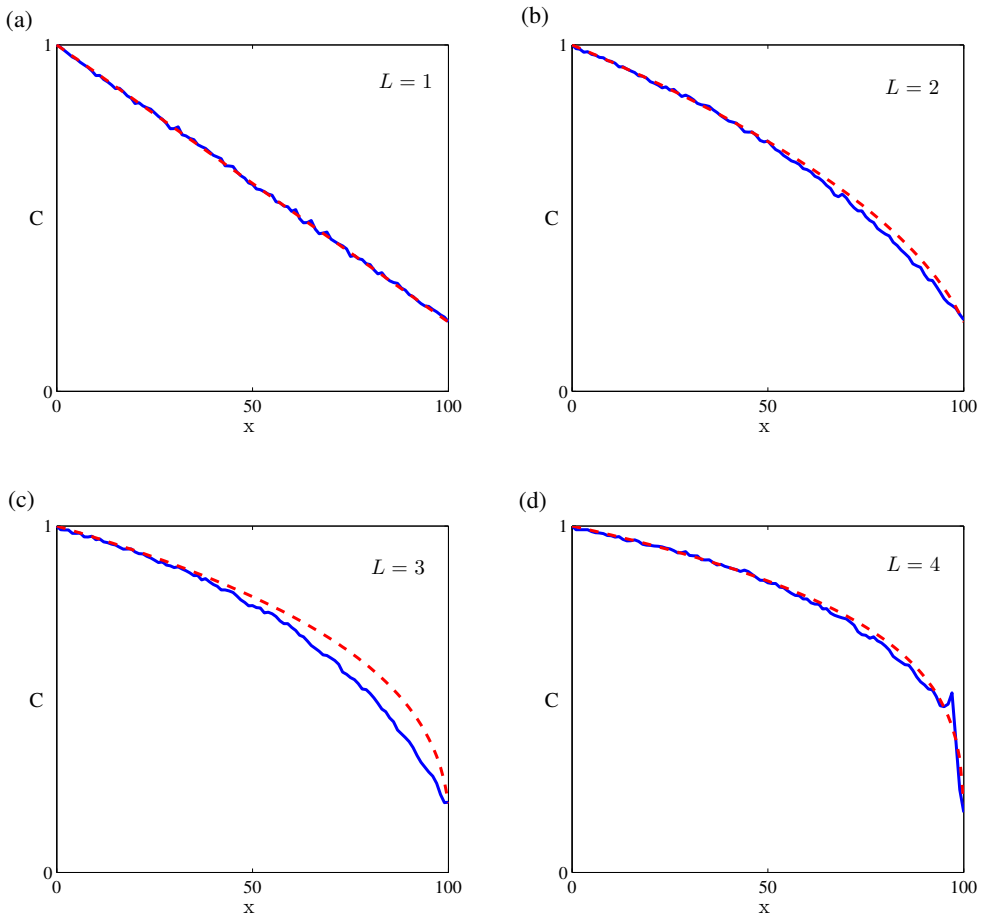

Figure 9: Comparison of the discrete steady state simulations (blue) and steady state solutions (red dotted) for $L = 1, 2, 3$ *and* 4 *((a), (b), (c) and (d) respectively) given the initial and boundary conditions in Eqs. 24-*(26)*. The discrete simulations were conducted with* $t = 9000$, $M = 10000$ *and* $x_{max} = 100$.

Despite the results being non-conclusive, it would appear as though there is minimal discrepancy at steady state between the discrete and continuum models. This would indicate that the use of the Taylor series expansion to relate the occupancy of lattice sites was the main cause for the discrepancy observed at the edges of the spreading agent population in Fig. 5. Thus the independence assumption appears to have had no real impact upon the

validity of the continuum model derived.

4 Discussion and conclusions

Most existing applications of exclusion process models to collective cell spreading rely upon the assumption that cells can be represented as "round" agents occupying a single lattice site [7, 11–13]. However, cells are often rod shaped and elongated [1, 14, 16] and there is evidence suggesting that this elongation has a significant effect upon the motility of the cells [14]. Here, the collective migration of cells is simulated through a generalised exclusion process model where each cell is represented by an agent occupying $L \geq 1$ adjacent lattice sites. A continuum description of a population of such agents is derived by making standard assumptions and forming discrete conservation statements. For "round" agents where $L = 1$, the resulting continuum model has a linear diffusion mechanism whereas for $L > 1$ the continuum limit process resulted in degenerate non-linear diffusion models related to the PME. Although a slight discrepancy is observed between the averaged density data from the simulations and the continuum models at the sharp fronts of the solutions to the PME for $L > 1$, the overall comparison between the continuum and discrete data is good. The results suggest that it is important to take the cell elongation into account when modelling collective cell spreading.

In addition to the work of Simpson, Baker and McCue [16], this report also analyses the assumption that a Taylor series expansion may be used to relate the occupancy of the lattice sites. It therefore represents the first attempt to analyse the impact of this assumption upon the shape dependent continuum model and thereby tests the validity of the models presented in [16]. While no conclusive results were able to be presented it appears as though the Taylor series expansion plays an important role in the discrepancy noted in Sec. 2.3, as there is a good match between the discrete and continuum steady state models. Implementing the boundary conditions required to simulate the steady state distribution of a spreading agent population is neither intuitive nor simple. Further work will be undertaken to improve the steady state simulation approach such that it can more accurately be compared to the solutions presented.

Clearly the 1D models represented in this report are a substantial simplification of the three dimensional nature of collective cell spreading. The extension of exclusion process models and continuum models to two and three dimensions would offer a more thorough description of collective cell migration. Modifying the exclusion process model such that it is suitable in 2D requires separating the randomly determined probability of an agent moving in a certain direction, p_d, into at least four categories; $p_d \leq 0.25$, $0.25 < p_d \leq 0.5$, $0.5 < p_d \leq 0.75$ and $0.75 < p_d \leq 1$, representing the attempted movement in the positive and negative x

and y directions. 2D simulations and continuum models are presented for agents of varying aspect ratio in the work of Simpson, Baker and McCue [16] for a population of uniformly aligned agents and a population consisting of two sub-populations, one horizontally and one vertically aligned. Note that in higher dimensions, agents are also able to rotate, a property which cannot be represented in a one dimensional model such as the one considered.

In addition, the model presented in this report considers an unchanging population of cells which spread without any growth, proliferation or change in shape. Models which include the changing size and shape of cells have particular applications in cancer biology. As highlighted in the work of Smith, Marra and Marshall, the dynamic elongation of live tumour cells can have a pronounced impact upon motility [14]. These dynamical changes may be modelled by including a second probability term, $P_g \in [0, 1]$, describing cell growth or elongation, which would cause agents in the exclusion process model to occupy an increasing or decreasing amount of lattice sites. Consider the 1D case for an agent with $L = 1$ occupying site i, where $P_g = 1$. During the time interval τ the agent may attempt to grow to an agent with $L = 2$, occupying sites i and $i + 1$ (Fig. 1(a)).

Acknowledgements

I would like to extend my sincerest thanks to my supervisors, Associate Professors Matthew J. Simpson and Scott W. McCue for their guidance and support throughout this project as well as their constructive feedback regarding my report. To AMSI and QUT, many thanks for providing me with the opportunity to undertake this research project during the summer vacation period. I would also like to thank both AMSI and CSIRO for providing me with the opportunity to present my project findings to my peers and gain experience in a conference setting. Computational resources and services used in this work were provided by the HPC and Research Support Group, Queensland University of Technology, Brisbane, Australia. I am particularly grateful for the coding advice given by HPC support staff, David Warne and Adam Siliato, which allowed me to obtain a greater amount of useful data.

Melanie Kaasinen received a 2013/14 AMSI Vacation Research Scholarship.

[1] Maini PK, McElwain DLS, Leavesley D. (2004) 'Travelling waves in a wound healing assay'. *Applied Mathematics Letters*. 17(5): 575-580.
[2] Tremel A, Cai A, Tirtaatmadja N, Hughes BD, Stevens GW, Landman KA, OConnor AJ. (2009) 'Cell migration and proliferation during monolayer formation and wound healing'. *Chemical Engineering Science* 64(2): 247-253.
[3] Kalluri R, Zeisberg M. (2006) 'Fibroblasts in cancer'. *Nature Reviews Cancer* 6(5): 392-401.
[4] Stebbings H. (2005) 'Cell Motility'. eLS. doi: 10.1038/npg.els.0003967
[5] Abercrombie M. (1970) 'Contact inhibition in tissue culture'. *In Vitro*, 6(2): 128-142.
[6] Simpson MJ, Merrifield A, Landman KA, Hughes BD. (2007) 'Simulating invasion with cellular automata: Connecting cell-scale and population-scale properties'. *Physical Review E* 76(2): 021918.
[7] Simpson MJ, Landman KA, Hughes BD. (2009) 'Multi-species simple exclusion processes'. *Physica A*, 388(4): 399-406.
[8] Simpson MJ, Landman KA, Hughes BD, Fernando AE. (2010) 'A model for mesoscale patterns in motile populations'. *Physica A*, 389(7): 1412-1424.
[9] Simpson MJ, Landman KA, Hughes BD. (2010) 'Cell invasion with proliferation mechanisms motivated by time-lapse data'. *Physica A* 389(18): 3779-3790.
[10] Deroulers C, Aubert M, Badoual M, Grammaticos B. (2009) 'Modeling tumor cell migration: From microscopic to macroscopic models'. *Physical Review E* 79(3): 031917.
[11] Khain E, Sander LM, Schneider-Mizell CM. (2007) 'The role of cell-cell adhesion in wound healing'. *Journal of Statistical Physics* 128(1/2): 209-218.
[12] Khain E, Sander LM. (2008) 'Generalized Cahn-Hilliard equation for biological applications'. *Physical Review E*, 77(5): 051129.
[13] Khain E, Schneider-Mizell CM, Nowicki MO, Chiocca EA, Lawler SE, Sander LM. (2009) 'Pattern formation of glioma cells: Effects of adhesion'. *Europhysics Letters*, 88(2): 28006.
[14] Smith HW, Marra P, Marshall CJ. (2008) 'uPAR promotes formation of the p130Cas-Crk complex to activate Rac through DOCK180'. *Journal of Cell Biology* 182(4): 777-790.
[15] Simpson MJ, Zhang DC, Mariani M, Landman KA, Newgreen DF. (2007) 'Cell proliferation drives neural crest cell invasion of the intestine'. *Developmental Biology* 302(2): 553-68.
[16] Simpson MJ, Baker RE, McCue SW. (2011) 'Models of collective cell spreading with variable cell aspect ratio'. *Physical Review E*, 83(2): 021901.

[17] Sherratt JA, Murray JD. (1990) 'Models of epidermal wound healing'. *Proceedings of the Royal Society B* 241(1300): 29-36.
[18] Sengers BG, Please CP, Oreffo ROC. (2007) 'Experimental characterization and computational modelling of two-dimensional cell spreading for skeletal regeneration'. *Journal of the Royal Society Interface* 4(17): 1107-17.
[19] Sheardown H, Cheng YL. (1996) 'Mechanisms of corneal epithelial wound healing'. *Chemical Engineering Science* 51(13): 4517-4529.
[20] Sherratt JA, Marchant BP. (1996) 'Nonsharp travelling wave fronts in the Fisher equation with degenerate nonlinear diffusion'. *Applied Mathematics Letters* 9(5): 33-38.
[21] Vázquez JL. (2007) *The Porous Medium Equation*. Oxford Science Publications, Oxford.
[22] Witelski TP. (1997) 'Segregation and mixing in degenerate diffusion in population dynamics'. *Journal of Mathematical Biology* 35(6): 695-712.
[23] Murray JD. (2002) *Mathematical Biology I: An Introduction*. 3rd Ed. Springer-Verlag, Heidelberg.
[24] Baker RE, Simpson MJ. (2012) 'Correcting mean-field approximations for birth-death-movement processes'. *Physical Review E* 82(4): 041905.