Bulletin of Mathematical Biology (2009) 71: 781–799
DOI 10.1007/s11538-008-9381-7

ORIGINAL ARTICLE

Distinguishing between Directed and Undirected Cell Motility within an Invading Cell Population

Matthew J. Simpson*, Kerry A. Landman, Barry D. Hughes

Department of Mathematics and Statistics, The University of Melbourne, 3010 Victoria, Australia

Received: 23 January 2008 / Accepted: 24 November 2008 / Published online: 7 February 2009

Abstract Cell invasion is the basis of several fundamental biological systems including developmental morphogenesis and disease progression. Invasion processes involve combined cell motility and proliferation. Standard experimental approaches to characterize invasion systems focus on measuring population-level wavespeed data. However, continuum models which incorporate either directed or undirected motility both give rise to traveling wave solutions with a well-defined wavespeed in terms of the motility parameters. Therefore, such population-level models and experimental data cannot be used to determine whether the motility is directed or undirected. This is a major impediment limiting our ability to interpret experimental observations of cell invasion. We demonstrate how to overcome this difficulty using individual-level data and discrete models. This approach can be used to interpret and design time-lapse imaging data to determine whether the cell motility is directed or undirected. Making a distinction between directed and undirected motility has profound implications regarding our ability to design strategies to manage development and disease associated with cell invasion.

Keywords Cell motility · Cell invasion · Multi-scale modelling · Directed motility

1. Introduction

A population of proliferative and motile agents is considered to mimic cell invasion, which is an essential feature of embryonic morphogenesis and disease (Murray, 2002). Cell invasion involves motile cells proliferating to some capacity density. In one-dimensional geometries, combined motility and proliferation give constant speed invasion waves (Fisher, 1937; Murray, 2002). Curiously, the motility mechanism plays a minor role in controlling the formation and characteristics of such waves regardless of whether the motility is directed or undirected (Simpson et al., 2006). This means that experimental

*Corresponding author.
E-mail address: m.simpson@ms.unimelb.edu.au (Matthew J. Simpson).
The authors gratefully acknowledge the support from the Australian Research Council (ARC) Discovery Project DP0662804. MS is an ARC postdoctoral fellow, KL is an ARC professorial fellow.

measurements which characterize the population-level wavespeed alone are insufficient to determine whether the motility is directed or undirected.

An inverse problem is the focus of this work. Given observations of an invasion system (Druckenbrod and Epstein, 2007; Young et al., 2004), can we determine whether the motility is directed or undirected? Understanding these details is the first step toward designing therapeutic strategies to manipulate the system behavior.

We use a cellular automata (CA) algorithm related to earlier investigations (Simpson et al., 2007b) to explore this inverse problem. Given that cell biologists are interested in distinguishing between biased and unbiased motilities (Cai et al., 2007; Druckenbrod and Epstein, 2007; Stein et al., 2006; Young et al., 2004), we examine a biologically inspired *in silico* system with the aim of making a distinction between different motility mechanisms. The analysis indicates that such distinctions are possible when the proliferation rate is sufficiently small compared to the motility rate. We suggest that experimentalists make measurements of (i) the net length of the movement of cells at the leading edge of the population, and (ii) detailed time-lapse data of movements within the leading edge. When the proliferation rate is not sufficiently small compared to the motility rate, it presently appears impossible to distinguish between different motility mechanisms using *in vivo* data, and other experimental conditions may have to be considered.

2. Algorithm

The CA algorithm has been described previously (Simpson et al., 2007b). The motility mechanism is related to a simple exclusion process that has been studied in other contexts (Chowdhury et al., 2005; Liggett, 1999). Briefly, agents move on a finite two-dimensional square lattice, $0 \leq x \leq L_x$ and $0 \leq y \leq L_y$, with lattice spacing Δ and at most one agent per site (x, y). Periodic boundary conditions are imposed at $y = 0$ and $y = L_y$. Reflecting boundary conditions are imposed at $x = 0$ and $x = L_x$. Each site (x, y) is associated with a local density $0 \leq M \leq 8$, defined as the total number of occupied sites of the eight closest sites. Motility and proliferation are contact inhibited due to two effects: (i) all events depend on whether M exceeds a prescribed integer carrying capacity $\kappa \leq 8$, and (ii) events which would place multiple agents per site are not permitted. At time t, with $N(t)$ agents, we attempt $N(t)$ randomly selected motility events and $N(t)$ randomly selected proliferation events, after which the system advances to the next time increment.

For the *motility mechanism*, the probability that an agent can move is P_m. A motile agent at (x, y) steps to $(x, y \pm 1)$ each with probability $1/4$, or to $(x \pm 1, y)$ with probability $(1 \pm a)/4$. The parameter $-1 \leq a \leq 1$ controls the bias. Our earlier work (Simpson et al., 2007b) only considered the unbiased $a = 0$ case. Setting $a \neq 0$ simulates agent motility with a constant drift. This drift is directed to the right when $a > 0$, or to the left when $a < 0$. This uniform drift mimics tactically-directed cell motility that is directed up a uniform gradient of some external attractant species. If the motility step places the agent on a site which is occupied or would have $M > \kappa$ at the selected site, the movement is aborted.

For the *proliferation mechanism*, the probability that an agent at (x, y) can proliferate is P_p. If $M \geq \kappa$ at (x, y), proliferation is aborted. If $M < \kappa$ at (x, y), the agent remains at (x, y) and attempts to proliferate with the daughter agent placed with equal probability

$1/4$ in either $(x \pm 1, y)$ or $(x, y \pm 1)$. If the selected site for the daughter is occupied or proliferation causes $M > \kappa$ at the selected site, proliferation is aborted. This mechanism does not allow agent removal to mimic death processes, which is appropriate for some applications (Gianino et al., 2003; Young et al., 2004). This proliferation mechanism can be easily modified to account for death in other systems.

The relationship between the simulation time and real time is governed by P_p. Assuming all agents have a uniform cell cycle time t_p, we require the CA time step τ to satisfy $P_p = \tau / t_p$, where τ is the real dimensional time step. Since P_p satisfies $0 \leq P_p \leq 1$, then $0 \leq \tau \leq t_p$, meaning that the real time step cannot be larger than the cell cycle time. We make a point that setting $P_p = 1$ does not uniquely specify the cell-cycle time, nor does $P_p = 1$ imply that the cell cycle time is fast. The correct interpretation of setting $P_p = 1$ is simply that the cell cycle time equal the real dimensional time step τ.

The length scale Δ represents the typical distance that an isolated agent moves within time τ. The motility parameter P_m is the probability that an agent will move a distance Δ within time τ. Therefore, setting $P_m = 1$ means that agents always move a distance Δ within time τ. Without loss of generality, we always work with dimensionless simulations setting $\Delta = \tau = 1$.

Although the motility and proliferation mechanisms involve transitions to the four nearest neighboring sites, we define the local density M using the eight closest sites including the diagonals. This has the advantage of allowing greater flexibility in varying κ than would be possible if the four nearest neighbor sites were used to define M. The CA algorithm does not include any agent-agent adhesion mechanism. This is reasonable as the specific details of cell-cell adhesion mechanisms are unknown for many systems of interest (Druckenbrod and Epstein, 2007; Young et al., 2004). We envisage that these details will be incorporated into future applications of the CA algorithm as our knowledge of adhesion mechanisms increases through experimental advances.

3. Results

CA simulations will be related to solutions of various continuum models that have the general form

$$u_t = -J_x + f(u), \tag{1}$$

where u is cell density, J is the cell flux, $f(u)$ is a source term, and subscripts denote partial differentiation. Traditionally, the connection between random walks and continuum partial differential equations has been made with the random walk as the basic model and the partial differential equation as a limiting case (Hughes, 1995). In recent applications, the logic is reversed and a discrete partial differential equation is interpreted as a random walk (Anderson and Chaplain, 1998; Cai et al., 2006).

In certain applications, it may be difficult to deduce any correspondence between the individual-level random walk mechanism and a continuous partial differential equation. This occurs when discrete models of complex biological phenomena are developed (Anderson and Chaplain, 1998; Longo et al., 2004; Marée and Hogeweg, 2001; Merks and Glazier, 2006). Our approach will be to infer a biologically reasonable individual-level random walk mechanism using existing biological observations, such as

time-lapse images. We show, in certain circumstances, that the inferred individual-level random walk model does correspond identically to a partial differential equation. Connecting the individual-level and population-level mechanisms gives us a greater understanding of the processes of interest from both the individual-level and population-level perspectives (Thorne et al., 2007).

For a single isolated nonproliferative agent, CA simulations reduce to a nearest neighbor random walk (Berg, 1983; Hughes, 1995). Standard arguments relate the stochastic motility to a diffusion process in an appropriate limit. If we average over the y-coordinate, the simulated motility is a one-dimensional random walk with mean displacement $a/2$, and mean square displacement $1/2$ per step. For lattice spacing Δ and constant time steps τ, if we hold $(\Delta)^2/\tau$ constant and let Δ and τ tend to zero jointly the relevant diffusion equation for the position probability density function (Hughes, 1995) is Eq. (1) with $f(u) = 0$ and $J = -Du_x + vu$. Here, $D = \lim_{\Delta,\tau\to 0} P_m\Delta^2/(4\tau)$ is the diffusivity and $v = \lim_{\Delta,\tau\to 0} a P_m \Delta/(2\tau)$ is the drift velocity. If the single isolated nonproliferative agent is biased, then the bias parameter a must decrease to zero with $a = \mathcal{O}(\Delta)$ to obtain a well-defined continuum limit. Therefore, setting $D = P_m/4$, $v = aP_m/2$ and $L_y = 1$ gives an exact correspondence between a single nonproliferative CA agent and the linear advection diffusion partial differential equation model in one dimension.

3.1. Nonproliferative populations

Simulations of nonproliferative populations are initiated on an empty lattice, except for $90 \le x \le 110$, where each site is occupied with probability $\kappa/8$. Results are averaged in the y direction by multiplying the total agents per column by $8/(\kappa L_y)$ so that the averaged one-dimensional density, $\bar{C}(x,t)$, is scaled relative to the initial agent density.

The unbiased case ($a = 0$). Snapshots (Fig. 1(a)) show agents dispersing symmetrically. Averaged density profiles constructed from 40 Monte Carlo simulations (Fig. 2(a)–(b)) are Gaussian. The shape of the profiles is insensitive to κ. Results with $\kappa = 8$ (Fig. 2(a)) are smoother than those with $\kappa = 4$ (Fig. 2(b)), due to the difference in the number of agents in these simulations. The $\kappa = 4$ simulations contain half the number of agents as the $\kappa = 8$ simulations, and consequently the averaging is not as effective in producing smooth density profiles.

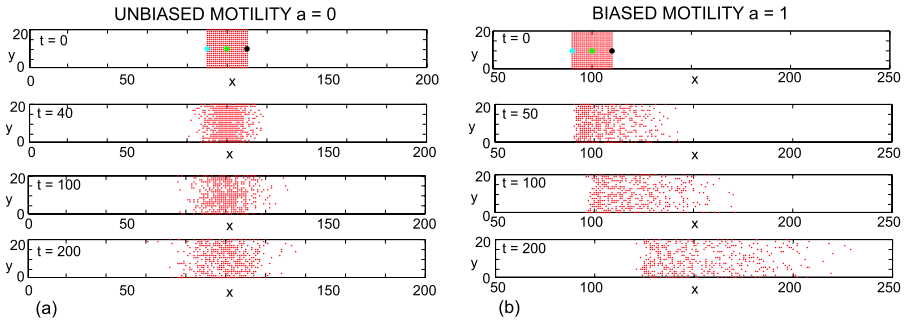

Fig. 1 Dispersion of a dense cluster of nonproliferative agents: (**a**) unbiased $a = 0$ and (**b**) biased $a = 1$ motilities are distinguishable ($\kappa = 8$, $P_m = 1$ and $P_p = 0$).

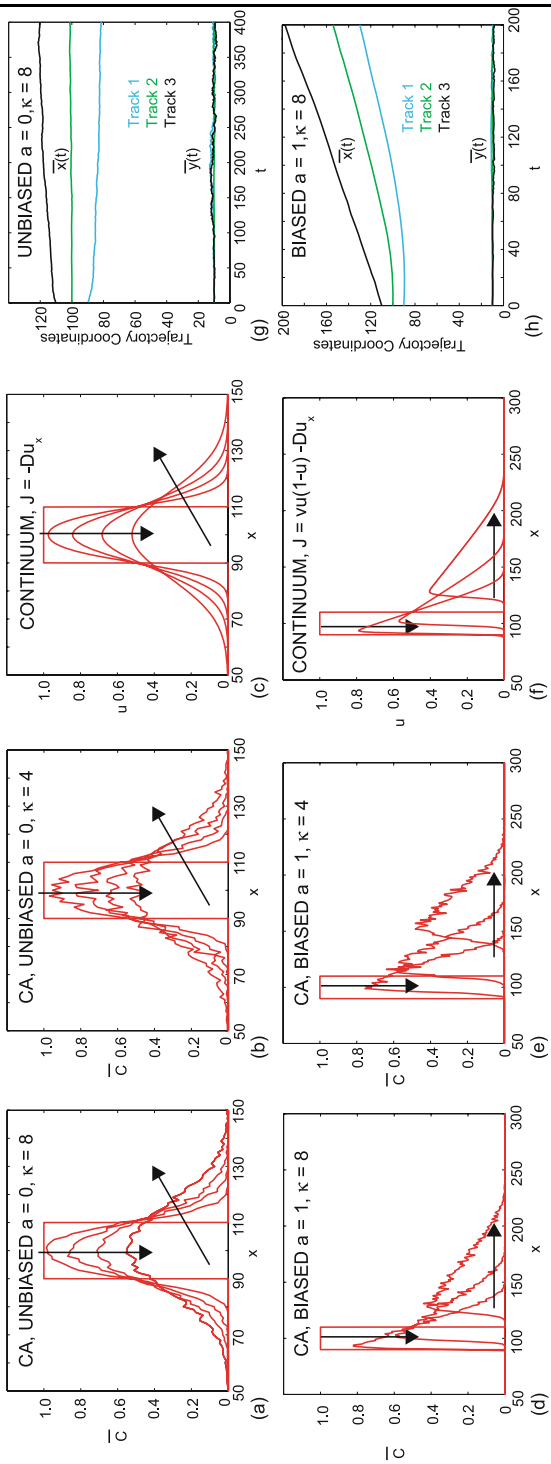

Fig. 2 Population-level data distinguishes between biased and unbiased motility in nonproliferative systems. Simulation agent density profiles $\bar{C}(x,t)$ for $a=0$, $P_m=1$ and $P_p=0$ with **(a)** $\kappa=8$ and **(b)** $\kappa=4$ are compared to continuum density profiles $u(x,t)$ **(c)** Eq. (1) with $f(u)=0$, $J=-Du_x$ and $D=1/4$ at $t=0$, 40, 100, 200 and 400. Simulation agent density profiles $\bar{C}(x,t)$ for $a=1$, $P_m=1$ and $P_p=0$ with **(d)** $\kappa=8$ and **(e)** $\kappa=4$ are compared to continuum density profiles $u(x,t)$ **(f)** Eq. (1) with $f(u)=0$ and $J=-Du_x+vu(1-u)$ and $D=1/4$, $v=1/2$ at $t=0$, 50, 100 and 200. *Arrows* show the direction of increasing time and profiles in corresponding subfigures are given at the same time points. Averaged trajectories within the population for $\kappa=8$ are given for **(g)** unbiased $a=0$ and **(i)** biased $a=1$ motilities. Details of the methods used to solve the partial differential equations are given in Appendix B.

The unbiased CA density profiles correspond to the solution of Eq. (1) with $f(u) = 0$ and $J = -Du_x$. Population-level flux arguments relating the CA mechanism to Eq. (1) are given in Appendix A. The solution of Eq. (1) with $f(u) = 0$, $J = -Du_x$, $u(x, 0) = 1$ where $90 \leq x \leq 110$ and $u(x, 0) = 0$ elsewhere on $-\infty < x < \infty$ is (Crank, 1975, p. 15):

$$2u(x,t) = \operatorname{erf}\big[(110 - x)/(2\sqrt{Dt})\big] + \operatorname{erf}\big[(x - 90)/(2\sqrt{Dt})\big]. \tag{2}$$

Equation (2) with $D = P_m/4$ (Fig. 2(c)) matches the CA simulations identically. Superimposing the profiles in (Fig. 2(a)) and (Fig. 2(c)) or (Fig. 2(b)) and (Fig. 2(c)) reveals a perfect match between the CA and continuum density profiles.

We probe further into the system considering the motion of individuals within the population. The position of three labeled agents (e.g., the blue, green, and black agents in Fig. 1) are tracked during 40 Monte Carlo simulations generating a suite of $(x(t), y(t))$ coordinates that are averaged at each time point. The averaged trajectories (Fig. 2(g)) reveal subtleties that are not apparent at the population level. The averaged $y(t)$ coordinates, $\bar{y}(t)$, remain constant due to the one-dimensional geometry. The averaged $x(t)$ coordinates, $\bar{x}(t)$, change over time. This result is different from a standard unbiased random walk without agent-agent interactions where the averaged x coordinate would remain constant over time (Berg, 1983; Hughes, 1995). The individual tracks indicate that CA agents move with a population-induced bias directed from high to low density. This population-induced bias is due to agent-agent interactions and is density-dependent since agent-agent interactions increase with agent density. The simplest way to represent this kind of population-induced bias using continuum models is by incorporating a density-dependent diffusivity $D(u)$, that decreases with increasing density $D'(u) < 0$.

It is intriguing that while the nonproliferative CA trajectory data (Fig. 2(g)) resembles a random walk associated with a population-level density-dependent diffusion process (Cai et al., 2006), our simulations and analysis (Appendix A) show that the actual population-level density profiles correspond to linear diffusion. Recent work explores this effect in detail (Simpson et al., 2009).

The biased case ($a = 1$). Snapshots (Fig. 1(b)) show agents initially at $x = 110$ moving right very efficiently, as transitions are perfectly biased and unimpeded. Agents initially located at $x = 90$ are unable to move for short times since transitions are impeded. These details are evident in the averaged individual tracks (Fig. 2(h)). Averaged density profiles (Fig. 2(d)–(e)) are triangular and insensitive to κ. Again, the results with $\kappa = 8$ (Fig. 2(d)) are smoother than those with $\kappa = 4$ (Fig. 2(e)), because of the difference in agent numbers. The absence of Gaussian profiles indicates that introducing a motility bias heralds a significant departure from motility governed by linear diffusion.

The biased CA profiles correspond to the solution of Eq. (1) with $f(u) = 0$, $J = -Du_x + vu(1 - u)$ (Appendix A) and appropriate initial conditions (Fig. 2(f)). This motility mechanism is related to standard traffic flow models (Whitham, 1974) with an additional diffusive motility component. Further simulations (not shown) for intermediately right-biased motility $0 < a < 1$ also give right-moving non-Gaussian profiles whose shape becomes more triangular as a increases.

The key point of these nonproliferative simulations is to demonstrate that there are clear distinctions between unbiased and biased motility mechanisms. These differences are dramatic and obvious. For example, the shape of the density profiles for unbiased agent motility (Fig. 2(a)–(b)) is completely different from the shape associated with biased motility (Fig. 2(d)–(e)). The spatial extent of the population spreading is also very

different, with the biased simulations (Fig. 1(b)) spreading over much greater distances than the unbiased simulations (Fig. 1(a)) during the same time period. In summary, these simulations demonstrate that the presence or absence of motility bias in a nonproliferative system is clearly distinguishable. The distinction between biased and unbiased motility can be determined by observing either population-level or individual-level data. We will now show that these differences are no longer obvious when proliferation is introduced into the system.

3.2. Proliferative populations

Traveling wave invasion fronts are supported by continuum invasion models (Eq. (1)) with a proliferation term satisfying $f(u_1) = f(u_2) = 0$, $f(u) > 0$ for $0 \leq u_1 < u < u_2$, $f'(u_1) > 0$ and $f'(u_2) < 0$ (Murray, 2002). Here, u_1 is the uninvaded cell density while u_2 is the invaded cell density. A well-studied example is logistic proliferation $f(u) = \lambda u(1 - u/k)$, where $k > 0$ is the carrying capacity density and $\lambda > 0$ is the proliferation rate.

Cell invasion experiments (Simpson et al., 2007a) and analysis of Eq. (1) with logistic proliferation show that population-level traveling waves are insensitive to the details of the motility flux J (Canosa, 1973; Simpson et al., 2006). Expressions relating the wavespeed c to the motility and proliferation parameters are a key feature of these models. For example, undirected motility $J = -Du_x$ gives $c = 2\sqrt{Df'(u_1)}$ (Murray, 2002). Similar expressions exist for other invasion models with different flux mechanisms such as chemotaxis, where the motility is directed (Landman et al., 2005). Therefore, given estimates of the wavespeed c and the proliferation rate λ, the wavespeed expressions can be used to calibrate the motility coefficient (Maini et al., 2004; Simpson et al., 2006). Inferring the motility mechanism in this way is problematic since there is a range of wavespeed expressions for other J and $f(u)$. Therefore, several different plausible models can predict the population-level speed without necessarily capturing the detailed mechanisms properly (Simpson et al., 2007b). This is a well-known problem called equifinality or multicausality (Beven, 2006).

4. Population-level equifinality

To demonstrate equifinality, both unbiased and biased invasion waves moving to the right are presented in Fig. 3. These simulations indicate that the speed of the invasion wavefront and the agent density behind the wavefront are the same in these two simulations. Therefore, the population-level cell density profiles for these two simulations are virtually identical. Although the motility mechanisms in these simulations are very different, the population-level outcomes are indistinguishable.

In contrast to the nonproliferative simulations (Fig. 1) where differences in biased and unbiased motility mechanisms gave radically different outcomes, we can no longer easily distinguish between biased and unbiased motility when proliferation is introduced. This is a major limitation that impedes our ability to understand and interpret population-level experimental observations of cell invasion. The simplest measurement often made to experimentally characterize an invasion system is the speed of the leading edge (Maini et al., 2004; Young et al., 2004). This measurement alone is inadequate if we are interested in determining the detailed mechanisms of how cells move within the invasion

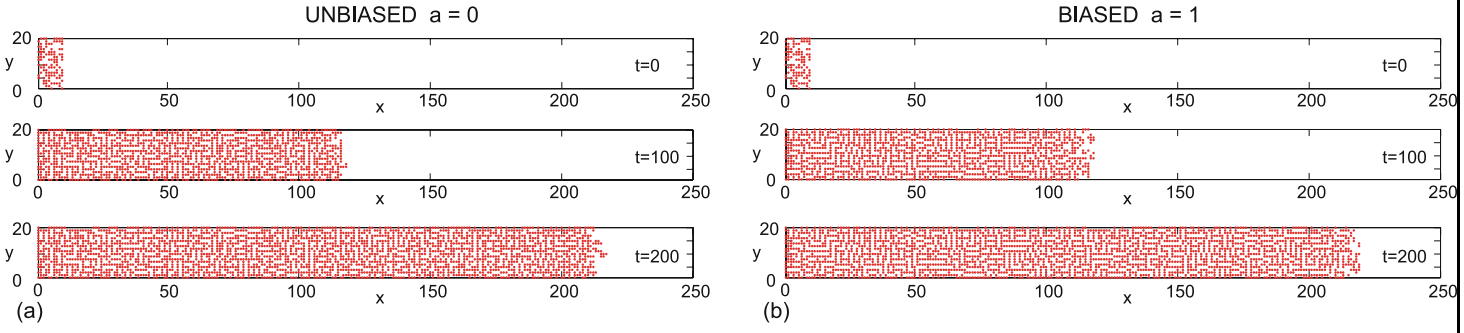

Fig. 3 Population-level data cannot distinguish between biased and unbiased motility in proliferative systems. An (**a**) unbiased and (**b**) biased invasion wave are compared revealing no differences at the population-level, since both waves move with speed $c = 1.05$ and have the same increase in cell number over time. The parameters for the unbiased wave are $P_m = P_p = 1$, $\kappa = 4$ and $a = 0$, while those for the biased wave are $P_m = 1$, $P_p = 0.44$, $\kappa = 4$ and $a = 1$.

system. It is unclear whether other data can be used to distinguish between different motility mechanisms within an invasion system. This motivates the use of a mathematical model to explore other forms of data that experimentalists may collect from invasion systems to help distinguish biased and unbiased motilities within an invasive population.

Clearly one major difference between the systems in Fig. 3 is the difference in proliferation rate. Consequently, one way to distinguish between the two systems would be to measure P_p accurately. While this is possible for *in vitro* conditions (Cai et al., 2007; Sengers et al., 2007), this is not possible for *in vivo* systems. Real *in vivo* observations of cell invasion often occur in complex environments and may depend on the availability of growth factors. These details are often unknown and, therefore, difficult to accurately recreate in an *in vitro* proliferation assay. Therefore, in the absence of any knowledge of P_p, the two systems in Fig. 3 are indistinguishable at the population-level.

5. CA wavespeed characterization

Simulations are used to define how c varies with P_m, P_p and a. We consider a sufficiently long lattice with $L_y = 20$ and each site with $0 \leq x \leq 10$ is initially occupied with probability $\kappa/8$. Simulations are performed for a range of parameters $0 < P_m, P_p \leq 1$, $-1 \leq a \leq 1$ and $\kappa = 4$. For 11 values of a, 25 uniformly distributed combinations of P_m and P_p are simulated 40 times each to generate a suite of data. Simulations are performed for sufficient time that the speed of the leading edge settles to a constant c (Simpson et al., 2007b). We describe c by the sample mean, as variations about the mean are small. The wavespeed can be described by

$$c = \alpha P_p^{\beta} P_m^{\delta}, \tag{3}$$

where α, β, and δ are determined using regression. Equation (3) describes the wavespeed for all parameter combinations. A summary of regression results in Table 1 is consistent with Fig. 3. Full details of the wavespeed data are given in Appendix C. The quality of the regression fit to Eq. (3) is excellent with $R^2 \geq 0.96$ for all cases.

Equation (3) and Table 1 show that it is possible to calibrate the CA algorithm to *any* particular observed population-level wavespeed c simply by choosing an appropriate combination of P_p, P_m, and a. While this procedure can be used to calibrate the

Table 1 Summary of regression data to describe the CA wavespeed using Eq. (3)

$a \geq 0$	α	δ	β	$a < 0$	α	δ	β
1.0	1.470	0.485	0.411	−1.0	0.600	−0.010	0.992
0.8	1.374	0.443	0.441	−0.8	0.697	0.089	0.910
0.6	1.277	0.411	0.469	−0.6	0.771	0.124	0.824
0.4	1.196	0.373	0.516	−0.4	0.834	0.181	0.738
0.2	1.113	0.334	0.572	−0.2	0.936	0.241	0.676
0	1.048	0.271	0.642				

CA algorithm to mimic the invasion wavespeed exactly, it gives no insight into the details of the motility mechanism regarding whether agent motility is directed or undirected.

The propensity to form constant speed invasion waves that are insensitive to the details of the motility mechanism is remarkable. Even simulations with perfectly left-biased motility ($a = -1$) and the same initial conditions and geometry as Fig. 3 give right-moving invasion waves. Regardless of the details of how the agents move in the system, the presence of contact inhibited proliferation masks these details at the population level. The formation of constant speed invasion waves is independent of the magnitude of P_p, so long as $P_p > 0$. The insensitivity to differences in motility bias is even more remarkable given that we have previously demonstrated major differences between biased and unbiased motilities in nonproliferative systems (Fig. 1).

The difficulty in deducing a motility mechanism from invasion wavespeed data alone is illustrated by relating our simulation results to real biological data. For example, Young et al. (2004) generated time-lapse data showing cell invasion within an intact tissue. Young et al. found that the invasion speed of the wavefront was approximately 35 μm/hr. This measurement reflects the rate at which the edge of the invading *population* advances, and does not reflect the speed of *individual* cell movement within that population. In the system that Young et al. (2004) considered, it is unclear whether the cell motility is directed or undirected (Simpson et al., 2007a, 2007b) and accurate measurements of the proliferation rate of the invading cells *in vivo* have not been made. Therefore, it is possible to use Eq. (3) together with Table 1 to fit the CA algorithm with any amount of bias and perfectly match the experimental wavespeed data. This calibration procedure offers no insight into the motility mechanism.

To help alleviate the difficulty of distinguishing between different motility mechanisms within a proliferative population, we investigate individual trajectory data. Focusing on trajectory data is timely as advances in microscopy and cell labeling techniques mean that individual trajectories within invasive populations are now being reported in the experimental literature (Druckenbrod and Epstein, 2007). These experimental advances must be paralleled with equivalent advances in modeling techniques to identify the best way to collect and analyze data to help deduce the mechanisms governing cell invasion.

6. Individual-level data

It is instructive to summarize the physical processes influencing a typical agent in a right-moving invasion wave. A tagged agent born near the leading edge has few neighbors and is initially free to move and proliferate. Neighboring agents also move and proliferate, increasing the local agent density. After sufficient time, the tagged agent becomes unable to move or proliferate due to agent-agent interactions and crowding (Fig. 4(a)–(b)). We refer to an agent that is no longer able to proliferate or move due to crowding as being jammed. The jamming of agents within a network laid down behind the leading edge of the invasion wave is consistent with certain aspects of time-lapse observations of cell invasion (Young et al., 2004). This physical description indicates that two kinds of individual data are relevant. First, we consider net movement from birth until jamming, and refer to this data as *steady state individual data*. Second, temporal details of movement between birth

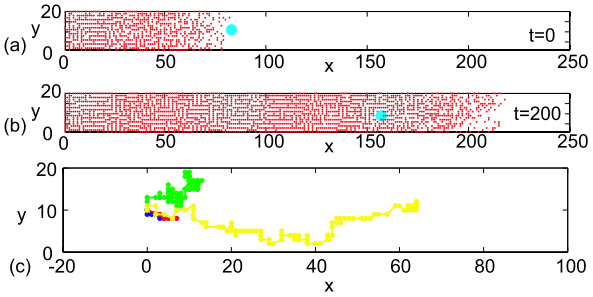

Fig. 4 Individual-level data can distinguish between biased and unbiased motility in proliferative systems. (**a**) An established invasion wave ($a = 1$ and $P_p/P_m = 0.05$) has a blue tagged agent introduced at the leading edge and time is reset to $t = 0$, (**b**) the same system is shown after 200 time steps. (**c**) Four tracks shifted so they start at $(0, 10)$. These tracks correspond to (i) $a = 0$ and $P_p/P_m = 1$ (*dark blue*), (ii) $a = 0$ and $P_p/P_m = 0.05$ (*green*), (iii) $a = 1$ and $P_p/P_m = 1$ (*red*), and (iv) $a = 1$ and $P_p/P_m = 0.05$ (*yellow*). In all cases, $P_m = 1$. Note that (**c**) has different scale to (**a**) and (**b**).

and jamming will be considered and referred to as *transient individual data*. Some illustrative individual data (Fig. 4(c)) shows four trajectories between birth and jamming, each with different P_p/P_m and a. Comparing these trajectories indicates that net displacement decreases with P_p/P_m and increases with a. These trends will be explored in detail.

In this section, we focus on a particular subset of CA parameters chosen to be biologically realistic. Individual-level data will be presented as a function of the proliferation rate relative to the motility rate, P_p/P_m, with τ chosen so that $P_m = 1$. We emphasize that P_p/P_m represents the frequency of proliferation events relative to motility events for an uncrowded agent. For example, if an uncrowded agent were to proliferate once for every 20 movements, then $P_p/P_m = 1/20$. The value of P_p/P_m does not specify the absolute magnitude of the proliferation rate. Although simulations can be performed for any value of P_p/P_m, we focus on cases where $P_p/P_m \leq 1$ since recent time-lapse data (Druckenbrod and Epstein, 2005; Young et al., 2004) indicate that proliferation events occur less frequently than motility events. Unfortunately, the data presented by Druckenbrod and Epstein (2005) and Young et al. (2004) are not suitable to accurately estimate P_p/P_m since these time lapse movies are relatively short and individual cells are not labeled within the population. Nonetheless, we use this experimental information to inform our simulations by focusing on relevant parameter values with $P_p/P_m \leq 1$.

Steady state individual data. We consider a series of *in silico* experiments. A simulation is performed until a constant speed right-moving invasion wave forms on a sufficiently long lattice with $L_y = 20$. Time is reset to $t = 0$ and a tagged agent, which moves and proliferates identically to all others, is placed at the leading edge at position $(x_i, 10)$, with x_i corresponding to the first empty column (counting from the left). As the system evolves, the location of the tagged agent is recorded until it becomes jammed. The agent displacement $R(t) = [(x_t - x_i)^2 + (y_t - 10)^2]^{1/2}$, where (x_t, y_t) are the tagged agent coordinates after t steps, is measured. Representative results are given in Figs. 5(a)–(b). The light blue trajectory in Fig. 5(a) moves 1.4 units before becoming jammed after 10 time steps. Comparing Figs. 5(a) and 5(b) shows that for the same P_p/P_m, biased trajectories are longer and require more time to become jammed than unbiased trajectories. The time taken for an agent to become jammed is denoted by T.

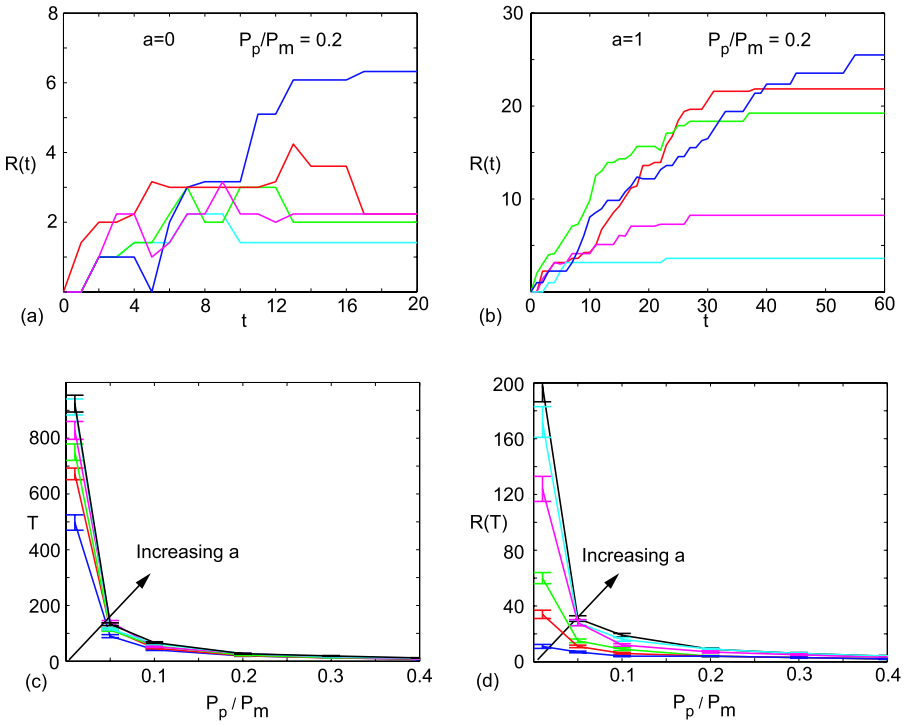

Fig. 5 Five trajectories in (**a**)–(**b**) correspond to unbiased ($a = 0$) and biased ($a = 1$) motilities respectively. An ensemble of steady state individual data in (**c**)–(**d**) shows the time taken until jamming T, and the total displacement $R(T)$ for $a = 0$, 0.2, 0.4, 0.6, 0.8, 1.0 and $0.01 \leq P_p/P_m \leq 0.4$ with $P_m = 1$. Error bars indicate the standard error.

Trends are explored for a range of parameters $0 \leq a \leq 1$ and $0 < P_p/P_m \leq 1$ with $P_m = 1$. For each parameter combination, 40 simulations were used to generate distributions of T and $R(T)$ (Figs. 5(c)–(d)). Several trends are evident. First, T decreases with P_p/P_m, confirming that agent motility is inhibited by progeny from neighboring agents as P_p/P_m increases. Second, differences in the $R(T)$ data for different amounts of bias become apparent as P_p/P_m is reduced, and $R(T)$ increases with a. Although we generated data in the range $0 < P_p/P_m \leq 1$, only a subset of this data in the range $0 < P_p/P_m \leq 0.4$ are presented in Figs. 5(c)–(d) since all data with $P_p/P_m > 0.4$ are indistinguishable for different amounts of bias. These trends indicate that agents undergo larger trajectories when the bias acts in the same direction as the invasion. Since the error bars overlap, data for T do not distinguish clearly between different a. However, data for $R(T)$ is more sensitive to differences in a indicating that it is possible to distinguish between the presence and absence of motility bias, as well as between different levels of bias, provided that P_p/P_m is sufficiently small.

Transient individual data. Dynamic measurements of motility during $0 < t < T$ are considered. Trajectory data (Figs. 5(a)–(b)) shows that $R(t)$ increases when $a = 1$, whereas $R(t)$ has no consistent trend when $a = 0$. These differences are partly caused by differences in the frequency of agent movement in the positive and negative x direc-

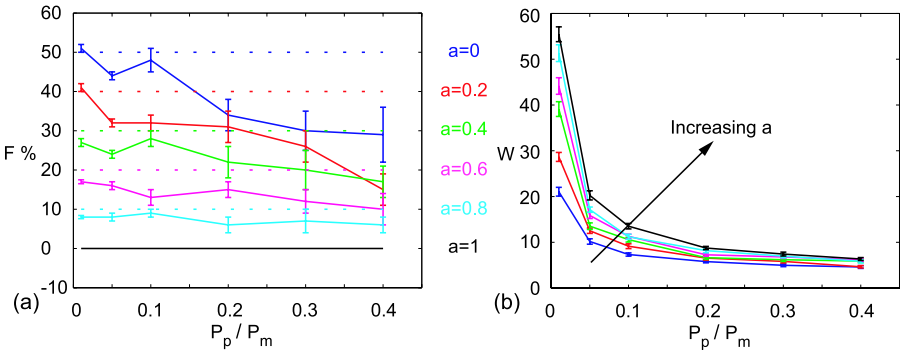

Fig. 6 Differences in the *transient individual data* are related to the shape of the invasion wave. (**a**) Frequency of movements in the negative x direction ($F\%$) are given for a tagged agent at the leading edge. Simulation results (*solid lines*) are compared to theoretical results for a single nonproliferative agent (*dotted lines*). (**b**) Variations in the width W, of the leading edge of the invasion wave. All results are presented for $a = 0$, 0.2, 0.4, 0.6, 0.8, 1.0, and $P_p/P_m \leq 0.4$, since results are indistinguishable for $P_p/P_m > 0.4$. In all cases, $P_m = 1$. Error bars indicate the standard error.

tions. To quantify this, we consider the x components of all trajectories and estimate the frequency of movements in the negative x direction relative to the total number of movements in the x direction during $0 < t < T$. Measurements are made 40 times for each a and P_p/P_m. For an isolated nonproliferative agent, the expected frequency of movement in the negative x direction is $F = 50(1-a)\%$. Simulation results are compared to this limiting case (Fig. 6(a)) showing distinctions between biased and unbiased motility.

The simulated frequency of reverse movements decreases with P_p/P_m, and approaches $50(1-a)\%$ as $P_p/P_m \to 0$. These trends are associated with variations in the shape of the invasion wave's leading edge. For a right-moving invasion wave, we define the width as $W = x_1 - x_2$, where x_1 is the first column, counting from the left, where $\bar{C}(x_1, t) = 0$, and x_2 is the first column, counting from the right, where $\bar{C}(x_2, t) = 1$. Measurements of W are made 40 times for $\kappa = 4$ and a range of P_p/P_m and a. The simulated width of the leading edge decreases with P_p/P_m, similar to continuum invasion models (Murray, 2002, p. 446), and increases with a (Fig. 6(b)). The width data is compatible with the frequency of reverse movement data since an agent at the leading edge of a wave with small P_p/P_m has few neighbors and is relatively unaffected by agent-agent interactions. As P_p/P_m increases and the width of the leading edge narrows, an agent at the leading edge is more likely to interact with others. These interactions reduce the opportunity for movements in the negative x direction.

Data required. The ratio P_p/P_m is critical in determining whether the presence or absence of motility bias in a proliferative population is distinguishable. When P_p/P_m is sufficiently large, proliferation dominates the system and individual-level data cannot distinguish between different amounts of motility bias. Alternatively, when P_p/P_m is sufficiently small, it is possible to distinguish between different amounts of motility bias using individual-level data. Recent time-lapse data demonstrates the dynamics of movement and proliferation within invasive populations in explants of intact tissue (Druckenbrod and Epstein, 2005; Young et al., 2004). Theoretically, estimates of P_p/P_m can be obtained from such data by measuring the ratio of proliferation to motility events for a particular

uncrowded cell over a sufficiently long period of time. To minimize agent-agent interactions, it is preferable to make these measurements in a low density environment such as that demonstrated by Young et al. (2004), where a few cells were isolated from the bulk invading population by sectioning the tissue at the leading edge. The time lapse data reported by Young et al. (2004) indicate that proliferation events occur less frequently than motility events, giving $P_p/P_m \ll 1$ for this system. This is encouraging since our simulations suggest it is possible to distinguish between biased and unbiased motilities under these conditions. We recommend that in future experiments data is collected to accurately estimate P_p/P_m, $R(T)$ and the frequency of reverse movements of cells in the leading edge. This data, together with an appropriately designed discrete model, will help address the question of whether the motility in the system is biased or unbiased.

Although the simulation data in Fig. 5 are given for the range $0.01 \leq P_p/P_m \leq 0.4$, the general trends also apply for lower values of P_p/P_m where it becomes easier to distinguish between biased and unbiased motility mechanisms. We investigated results for $0.01 \leq P_p/P_m \leq 1.0$ and chose to present data for $0.01 \leq P_p/P_m \leq 0.4$ as this interval depicts the critical transition where the effects of the motility bias become apparent as P_p/P_m becomes sufficiently small.

Other forms of individual data are available from the CA algorithm such as (i) the direction of individual trajectories (Simpson et al., 2007b) and (ii) the number of distinct lattice sites visited (Hughes, 1995). We also analysed these data and found that direction data was insensitive to differences in motility bias, while the number of distinct sites visited mirrored the trajectory length data (Fig. 5(d)). These data may be more informative for other applications with alternative CA rules (e.g., allowing agent removal).

7. Conclusions

Population-level wavespeed data cannot distinguish between biased and unbiased motility in an invasion wave of cells. Recognizing and understanding this difficulty is very important since experimental characterization of invasive populations often focuses on population-level data through measuring the speed of invasion (Maini et al., 2004; Young et al., 2004). In order to understand the details of how cells move within an invasive system, that is, to distinguish between random and directed motility, it is insufficient to measure the wavespeed alone. Other data must be considered.

The extreme efficiency with which carrying capacity-limited proliferation dominates and masks the influence of various types of motility within an invasion system is surprising and not intuitive. The dominating influence of carrying capacity-limited proliferation is particularly remarkable given that the presence or absence of motility bias is clearly distinguishable in nonproliferative systems.

The difficulty associated with identifying how to determine whether motility is biased or unbiased in an invasive system can be partly addressed by analyzing individual-level data. We found that measurements of the total displacement of agents and the frequency of reverse movements within the leading edge can distinguish between directed and undirected motility. Making this distinction is only possible when the proliferation rate is sufficiently small relative to the motility rate, in our case we required $P_p/P_m \leq 0.1$. In other words, in order to distinguish between biased and unbiased motility, on average agents must undergo at least ten motility events for each proliferation event. Increasing

our focus on using and understanding discrete invasion models is timely as cell biologists are beginning to generate detailed time-lapse data that reveal individual behavior within invasive populations. Individual-level data are more naturally interpreted using individual-based models, giving access to critical information governing invasion systems that is not available using continuum models in standard ways.

Qualitatively, the trends demonstrated using individual-level data from the CA algorithm are intuitively sound. Agents in the wavefront region can move, and their movement is either biased or unbiased. If the movement is unbiased, agents travel shorter distances than if the movement is biased in the same direction as the net invasion. The faster agents proliferate, the earlier they become jammed in the network of agents laid down by the invading population. While these individual-level trends are intuitively reasonable, our motivation behind focusing on individual-level data rather than population-level data is not immediately obvious. Individual-level data are important when studying the inverse problem of determining whether cell motility within an invasive population is directed or undirected. Standard population-level data, such as wavespeed measurements, are insufficient to address this inverse problem.

Significant challenges using individual-based invasion models remain. Identifying motility mechanisms in systems involving spatially variable motility bias, more complicated agent-agent interactions or invasion on growing substrates are grand challenges. Individual-based invasion modeling techniques are well suited to investigate such problems.

Our work illustrates a key role for mathematical modeling in biology. Making experimental observations naturally leads to the proposition of conjectures about underlying mechanisms. It is possible that several potential mechanisms could explain the same observation and that differentiating between potential explanations may be impossible based on conjecture alone. The beauty of mathematical modeling is that it provides a rational tool which can assist in distinguishing between various explanations of the same outcome.

Acknowledgements

This work was supported by the Australian Research Council (ARC). MS is an ARC postdoctoral fellow and KL is an ARC professorial fellow. We appreciate support from the Particulate Fluids Processing Centre, an ARC Special Research Centre. We are grateful for comments and advice from Heather Young and Donald Newgreen.

Appendix A: Relating individual-level and population-level mechanisms in a nonproliferative population

We consider a population of interacting agents. The relationship between the nonproliferative CA motility mechanism with $P_p = 0$ and the population-level agent flux can be deduced using the nomenclature in Fig. A.1. Each column has height L_y, the number of agents in the ith column is N_i and the probability that any site in the ith column is occupied is $u_i = N_i/L_y$. Two adjacent columns located at x_i and x_{i+1} are considered.

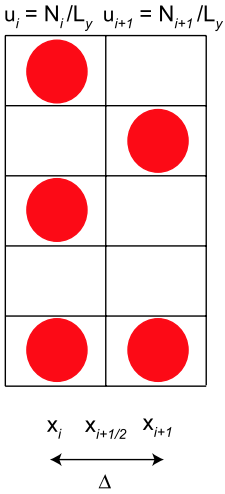

Fig. A.1 Schematic diagram showing two adjacent columns in the CA lattice. The columns are spaced Δ apart and have height L_y. Each site in column x_i is occupied with probability $u_i = N_i/L_y$. Flux arguments can be used to establish a population-level expression for the flux of agents across the interface at $x_{i+\frac{1}{2}}$.

We develop an estimate of the flux of agents across the interface $x_{i+\frac{1}{2}}$ at any time. Physically, this flux corresponds to the number of agents per unit time per unit area that step from x_i to x_{i+1}, less the number of agents per unit time per unit area that step from x_{i+1} to x_i. Mathematically this is equivalent to

$$J_{i+\frac{1}{2}} = \frac{\Delta}{\tau}\left[u_i \frac{P_m(1+a)}{4}(1-u_{i+1}) - u_{i+1}\frac{P_m(1-a)}{4}(1-u_i)\right]. \tag{A.1}$$

The first term on the right of Eq. (A.1) is a product of four factors that can be interpreted as follows: (i) u_i is the number of agents per site in column i, (ii) $P_m(1+a)/4$ is the probability that an agent moves to the right within a time step, (iii) $(1-u_{i+1})$ is the probability that a site in column $i+1$ is vacant and (iv) Δ/τ gives the correct dimension for the flux. A similar interpretation can be given for the second term on the right of Eq. (A.1). Expanding these terms and simplifying gives

$$J_{i+\frac{1}{2}} = \frac{\Delta P_m}{4\tau}\left[-(u_{i+1}-u_i) + a(u_i - 2u_i u_{i+1} + u_{i+1})\right]. \tag{A.2}$$

The terms u_i and u_{i+1} are expanded in a Taylor series about $x_{i+1/2}$,

$$u_i = u_{i+\frac{1}{2}} - \frac{\Delta}{2}(u_x)_{i+\frac{1}{2}} + \mathcal{O}(\Delta)^2, \qquad u_{i+1} = u_{i+\frac{1}{2}} + \frac{\Delta}{2}(u_x)_{i+\frac{1}{2}} + \mathcal{O}(\Delta)^2, \tag{A.3}$$

where subscripts denote partial differentiation. Substituting these Taylor series expansions into Eq. (A.2) gives

$$J_{i+\frac{1}{2}} = \frac{\Delta P_m}{4\tau}\left[-\Delta(u_x)_{i+\frac{1}{2}} + 2a u_{i+\frac{1}{2}}(1-u_{i+\frac{1}{2}})\right] + \mathcal{O}(\Delta)^2. \tag{A.4}$$

We now let Δ and τ tend to zero jointly, while holding $(\Delta)^2/\tau$ constant to obtain the continuous population-level flux that is given by

$$J = -Du_x + vu(1-u), \tag{A.5}$$

where

$$D = \lim_{\Delta,\tau\to 0} \frac{P_m \Delta^2}{4\tau}, \qquad v = \lim_{\Delta,\tau\to 0} \frac{a P_m \Delta}{2\tau}. \tag{A.6}$$

If the motility is biased then the bias parameter a must also decrease to zero with $a = \mathcal{O}(\Delta)$ in order to obtain a well-defined continuum limit.

Although the flux expression given by Eq. (A.5) is valid strictly when $\Delta, \tau, a \to 0$, our results (Fig. 2) show that the solution of Eq. (1) with $f(u) = 0$ and J given by Eq. (A.5) matches the CA density profiles very well for $D = P_m/4$ and $v = a P_m/2$ with finite Δ, τ and a.

Appendix B: Numerical methods

Numerical solutions of Eq. (1) with $f(u) = 0$ are given in Fig. 2 of the main paper. To generate these solutions, we consider a truncated domain $-l_1 < x < l_2$, with l_1 and l_2 chosen large enough to avoid boundary effects. Homogeneous Dirichlet boundary conditions are simulated with $u(-l_1, t) = u(l_2, t) = 0$. Spatial derivatives are replaced with a standard centered in space finite difference approximation on a uniformly discretized grid with spacing δx. The system of semidiscrete nonlinear ordinary differential equations are discretized in time with an implicit backward Euler approximation and constant time step δt. The resulting system of nonlinear equations is linearized with a Picard iterative technique. Iterations are performed until the maximum absolute change in the dependent variable across the grid falls below ϵ per iteration. Constant and sufficiently small δx and δt are used ensuring results are grid independent. Results correspond to $\delta x = 0.10$, $\delta t = 0.05$, $\epsilon = 1 \times 10^{-4}$.

Appendix C: CA wavespeed data

Details of the regression data used to fit the CA wavespeed to Eq. (3) in the main paper are given in Fig. C.1. Individual subfigures correspond to a different value of the bias parameter $-1 \le a \le 1$. The wavespeed data in each subfigure represents 25 combinations of P_p and P_m uniformly distributed within the range $0 < P_p, P_m \le 1$. For each parameter combination, the simulation data is an average of 40 Monte Carlo realizations.

Regression parameters and goodness of fit statistics are given in each subfigure. The regression parameters were obtained by fitting the simulation wavespeed data to Eq. (3) in the main paper using standard nonlinear regression algorithms in MATHEMATICA. The goodness of fit R^2 statistic was evaluated for the simulated data compared to the regression model.

To present the comparison between the regression model and simulation data, the simulation wavespeed data is plotted against the corresponding theoretical regression wavespeed data together with the line $y = x$ which would indicate a perfect match between the simulation and regression data.

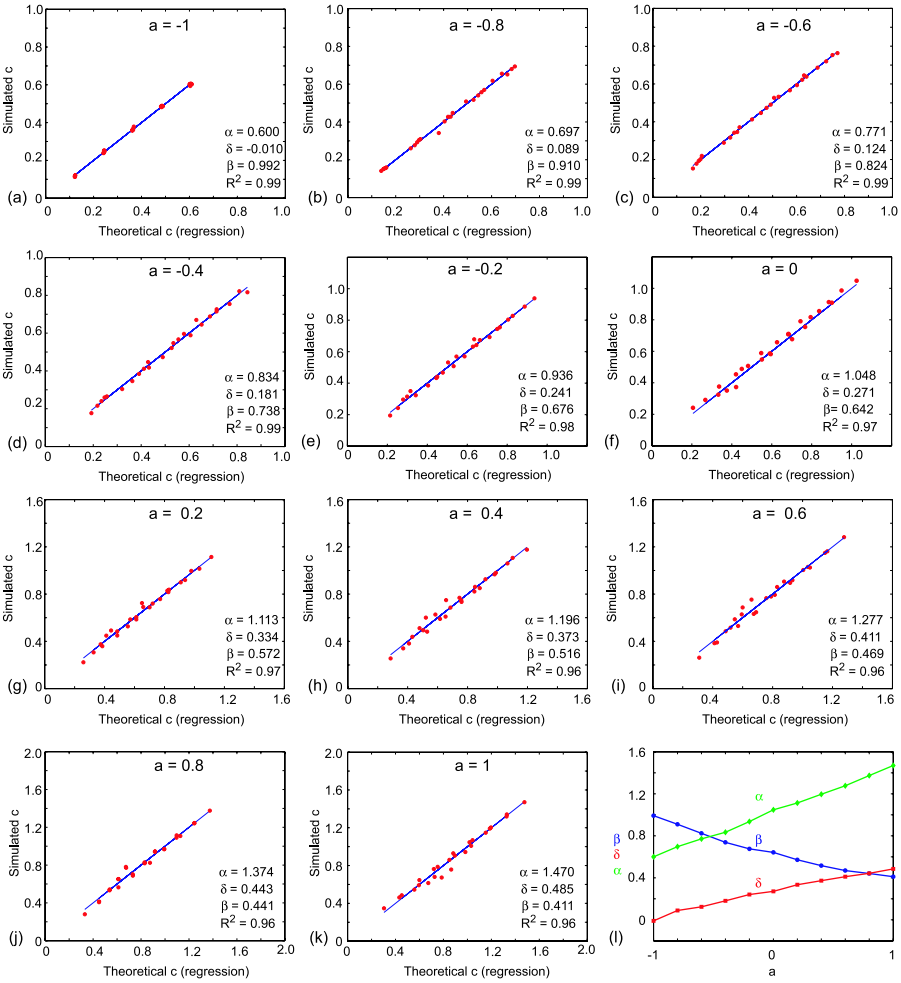

Fig. C.1 Simulated wavespeed data (*red bullets*) superimposed on the line $y = x$ which represents the results from the regression Eq. (3) for various combinations of $0 < P_p, P_m \leq 1$. Data are given for $-1 \leq a \leq 1$ in **(a)**–**(k)** showing the regression parameters α, β, and γ and R^2. The R^2 values represent the goodness of fit. Subfigure **(l)** shows the variation of α, β and δ with a.

References

Anderson, A.R.A., Chaplain, M.A.J., 1998. Continuous and discrete models of tumour-induced angiogenesis. Bull. Math. Biol. 60, 857–900.

Berg, H.C., 1983. Random Walks in Biology. Expanded Edition. Princeton University Press, Princeton.

Beven, K., 2006. A manifesto for the equifinality thesis. J. Hydrol. 320, 18–36.

Cai, A., Landman, K.A., Hughes, B.D., 2006. Modelling directional guidance and motility regulation in cell migration. Bull. Math. Biol. 68, 25–52.

Cai, A., Landman, K.A., Hughes, B.D., 2007. Multi-scale modeling of a wound-healing cell migration assay. J. Theor. Biol. 245, 576–594.

Canosa, J., 1973. On a nonlinear diffusion equation describing population growth. IBM J. Res. Dev. 17, 307–313.
Chowdhury, D., Schadschneider, A., Nishinari, K., 2005. Phys. Life Rev. 2, 318–352.
Crank, J., 1975. The Mathematics of Diffusion, 2nd edn. Oxford University Press, Oxford.
Druckenbrod, N.R., Epstein, M.L., 2005. The pattern of neural crest advance in the cecum and colon. Dev. Biol. 287, 125–133.
Druckenbrod, N.R., Epstein, M.L., 2007. Behavior of enteric neural crest-derived cells varies with respect to the migratory wavefront. Dev. Dyn. 236, 84–92.
Fisher, R.A., 1937. The wave of advance of advantageous genes. Ann. Eugen. 7, 353–369.
Gianino, S., Grider, J.R., Cresswell, J., Enomoto, H., Heuckeroth, R.O., 2003. GDNF availability determines enteric neuron number by controlling precursor proliferation. Development 130, 2187–2198.
Hughes, B.D., 1995. Random Walks and Random Environments, Vol. 1. Oxford University Press, Oxford.
Landman, K.A., Simpson, M.J., Slater, J.L., Newgreen, D.F., 2005. Diffusive and chemotactic cellular migration: Smooth and discontinuous travelling wave solutions. SIAM J. Appl. Math. 65, 1420–1442.
Liggett, T.M., 1999. Stochastic Interacting Systems: Contact, Voter and Exclusion Processes. Springer, Berlin.
Longo, D., Peirce, S.M., Skalak, T.C., Davidson, L., Marsden, M., Dzamba, B., DeSimone, D.W., 2004. Multicellular computer simulation of morphogenesis: blastocoel roof thinning and matrix assembly in *Xenopus laevis*. Dev. Biol. 271, 210–222.
Maini, P.K., McElwain, D.L.S., Leavesley, D.I., 2004. Traveling wave model to interpret a wound-healing cell migration assay for human peritoneal mesothelial cells. Tissue Eng. 10, 475–482.
Marée, A.F.M., Hogeweg, P., 2001. How amoeboids self-organize into a fruiting body: multicellular coordination in *Dictyostelium discoideum*. Proc. Natl. Acad. Sci. 98, 3879–3883.
Merks, R.H.M., Glazier, J.A., 2006. Dynamic mechanisms of blood vessel growth. Nonlinearity. 19, C1–C10.
Murray, J.D., 2002. Mathematical Biology I: An Introduction, 3rd edn. Springer, Heidelberg.
Sengers, B.G., Please, C.P., Oreffo, R.O.C., 2007. Experimental characterization and computational modeling of two-dimensional cell spreading for skeletal regeneration. J. R. Soc. Interface 4, 1107–1117.
Simpson, M.J., Landman, K.A., Hughes, B.D., Newgreen, D.F., 2006. Looking inside an invasion wave of cells using continuum models: proliferation is the key. J. Theor. Biol. 243, 343–360.
Simpson, M.J., Zheng, D.C., Mariani, M., Landman, K.A., Newgreen, D.F., 2007a. Cell prolfieration drives neural crest cell invasion of the intestine. Dev. Biol. 302, 553–568.
Simpson, M.J., Merrifield, A., Landman, K.A., Hughes, B.D., 2007b. Simulating invasion with cellular automata: Connecting cell-scale and population-scale properties. Phys. Rev. E 76, 021918.
Simpson, M.J., Landman, K.A., Hughes, B.D., 2009. Multi-species simple exclusion processes. Physica A 388, 399–406.
Stein, A.M., Demuth, T., Mobley, D., Berens, M., Sander, L.M., 2006. A mathematical model of glioblastoma tumor spheroid invasion in a three-dimensional in vitro experiment. Biophys. J. 92, 356–365.
Thorne, B.C., Bailey, A.M., DeSimone, D.W., Peirce, S.M., 2007. Agent-based modeling of multicell morphogenic processes during development. Birth Defects Res. C 81, 344–353.
Whitham, G.B., 1974. Linear and Nonlinear Waves. Wiley, New York.
Young, H.M., Bergner, A.J., Anderson, R.B., Enomoto, H., Milbrandt, J., Newgreen, D.F., Whitington, P.M., 2004. Dynamics of the neural crest-derived cell migration in the embryonic mouse gut. Dev. Biol. 270, 455–473.