bioRxiv preprint doi: https://doi.org/10.1101/811257; this version posted October 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Identifying density-dependent interactions in collective cell behaviour

Alexander P Browning[1,2*], Wang Jin[1], Michael J Plank[3,4], and Matthew J Simpson[1]

[1]*School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia*
[2]*ARC Centre of Excellence for Mathematical and Statistical Frontiers, QUT, Australia*
[3]*Biomathematics Research Centre, University of Canterbury, Christchurch, New Zealand*
[4]*Te Pūnaha Matatini, a New Zealand Centre of Research Excellence, New Zealand*

October 18, 2019

Abstract

Scratch assays are routinely used to study collective cell behaviour *in vitro*. Typical experimental protocols do not vary the initial density of cells, and typical mathematical modelling approaches describe cell motility and proliferation based on assumptions of linear diffusion and logistic growth. Jin *et al.* (2016) find that the behaviour of cells in scratch assays is density dependent, and showed that standard modelling approaches cannot simultaneously describe data initiated across a range of initial densities. To address this limitation, we calibrate an individual based model to scratch assay data across a large range of initial densities. Our model allows proliferation, motility and a direction bias to depend on interactions between neighbouring cells. By considering a hierarchy of models where we sequentially remove interactions, we perform model selection analysis to identify the minimum interactions required for the model to simultaneously describe data across all initial densities. The calibrated model is able to match the experimental data across all densities and captures details about the spatial structure of cells. Our primary findings provide strong evidence to suggest that motility is density-dependent in these experiments. On the other hand, we do not see the effect of crowding on proliferation in these experiments. These results are significant as they are precisely the opposite of the assumptions in standard continuum models, such as the Fisher-Kolmogorov equation and generalisations of the Fisher-Kolmogorov equation.

Keywords: individual based model | cell migration | scratch assay | interactions | approximate Bayesian computation | model selection

*Corresponding author. E-mail: ap.browning@qut.edu.au

bioRxiv preprint doi: https://doi.org/10.1101/811257; this version posted October 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

1 Introduction

Scratch assays are routinely used to study collective cell behaviour *in vitro* [1–5]. These experiments are conducted by placing a monolayer of cells on a two-dimensional substrate and creating an artificial wound, or *scratch*, in the centre of the population (figure 1*a–d*) [3]. Typical experimental protocols do not vary the initial density of cells between experiments, which limits the amount of information that can be obtained about any density-dependence of cell migration or proliferation. We consider novel experimental data where we deliberately vary the initial density of cells between experiments. The variation in the initial cell density in our experiments is large: the initial population in the highest density experiment is greater than the the population after 36 h in the lowest density experiment.

Logistic growth and linear diffusion are often assumed to be the key mechanisms governing collective cell behaviour in a range of *in vitro* and *in vivo* conditions [3,6–11]. Mean-field mathematical models that incorporate one or both of these mechanisms are routinely used to model tumour spheroids [12]; cells in living tissues [13, 14]; and simple *in vitro* experiments such as scratch [3], migration [15] and proliferation [16] assays. While calibrating these models to experimental data often leads to a good match [6], these models make the standard assumption that the parameters are independent of both initial condition and cell density. For example, the Fisher-Kolmogorov equation

$$\frac{\partial c}{\partial t} = \underbrace{D\nabla^2 c}_{\substack{\text{Linear}\\ \text{diffusion}}} + \underbrace{\lambda c\left(1 - \frac{c}{K}\right)}_{\substack{\text{Logistic}\\ \text{growth}}}, \tag{1}$$

is commonly used to model scratch assay experiments [3, 17] and describes density-independent motility, characterised by a constant diffusivity D; and density-dependent proliferation, characterised by a constant proliferation rate λ and a constant carrying capacity K. Jin *et al.* [3] find that calibrating equation (1) to scratch assay data yields vastly different estimates of D for each initial condition considered. The assumption of density-independent motility may be, therefore, inappropriate.

In this work we describe the cell behaviour with a lattice-free individual based model (IBM) [16,18,19]. The IBM represents cells as *agents* that take locations in continuous space, and so we can specify the initial agent locations in the model to precisely match the initial cell locations in the experiments. This choice also allows the model to capture local details—such as spatial structure and clustering—which are neglected by standard continuum modelling approaches [3, 9]. The agents in the IBM undergo random proliferation and movement events, the rates of which we assume depend explicitly on interactions between neighbouring agents. We quantify these interactions with kernels that depend on the distance between pairs of cells. Directional bias is also incorporated, so that agents are more likely to move either away from, or towards, regions of high density [18,20]. A key advantage of the IBM is its flexibility: it is trivial to add and remove mechanisms, which we do to study the interactions required for the model to simultaneously match all experiments. Finally, the IBM is stochastic and so naturally describes the variation between experiments.

The primary goal of this study is to identify interactions which enable the model to simultaneously describe experimental data across a range of initial densities. We take a Bayesian approach to parameter estimation [16,21–23], and identify interactions using model selection [21]. We always force the model to simultaneously match data from all

bioRxiv preprint doi: https://doi.org/10.1101/811257; this version posted October 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Experiment	$N(0)$	$N(18)$	$N(36)$	Fold change
1	183	322	652	3.56
2	299	528	927	3.10
3	354	549	893	2.52
4	404	636	1121	2.77
5	427	657	1077	2.52
6	522	849	1414	2.71
7	677	1200	2013	2.97
8	692	1285	1853	2.67
9	731	1155	1974	2.70

Table 1: Summary of experimental data showing the cell count, $N(t)$, and fold change, $N(36)/N(0)$, for all nine experiments.

nine experiments. The mathematical model is always initiated using the initial configuration of cells in each experiment, and we compare simulated and experimental data at 18 h and 36 h, the latter which corresponds to the duration of the experiment. The calibrated model is able to replicate the experimental data, and we find evidence of density dependent motility, which is contrary to the usual assumption of linear diffusion. Additionally, experimentation with summary statistics confirms the importance of spatial structure, which is neglected by standard modelling and model calibration approaches.

2 Materials and methods

2.1 Experimental methods

We consider a series of scratch assay experiments using the PC-3 prostate cancer cell line [24], where we deliberately vary the initial number of cells. For full details of the experimental technique, see Jin *et al.* [3]. In summary, a population of cells is seeded at a density of approximately 8000, 10000 and 12000 cells in a 9000 µm diameter well within a 96-well plate (figure 1*a*,*b*). Cells are grown overnight to create a spatially uniform monolayer before a scratch is created (figure 1*c*). Images of the central 1440×1900 µm of each well are captured over a period of 48 hours after the monolayer is scratched (figure 1*d*).

ImageJ [25] is used to determine the approximate coordinates of individual cells in each image, this data is given in the supporting material. We exclude the first 12 hours of experimental data from our analysis [16] to ensure that sufficient time has passed so that the cells are migrating and proliferating after the scratch has been made. We then record experimental images and we treat this as the beginning of the experiment, $t = 0$ h. The variability in initial cell number is high: despite an initial seeding density of approximately 8000–12000 cells peer well, which corresponds to expected initial number of cells within the field-of-view of 344–516, we find that the initial number of cells within the field-of-view at $t = 0$ h ranges from 183 to 731 (figure 1*f*–*i*). This variation is also high between experiments of the same seeding density [26], due to the fact that our field-of-view is relatively small and so that fluctuations about the expected values are relatively large. We summarise the experimental data in figure 2 and table 1.

bioRxiv preprint doi: https://doi.org/10.1101/811257; this version posted October 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

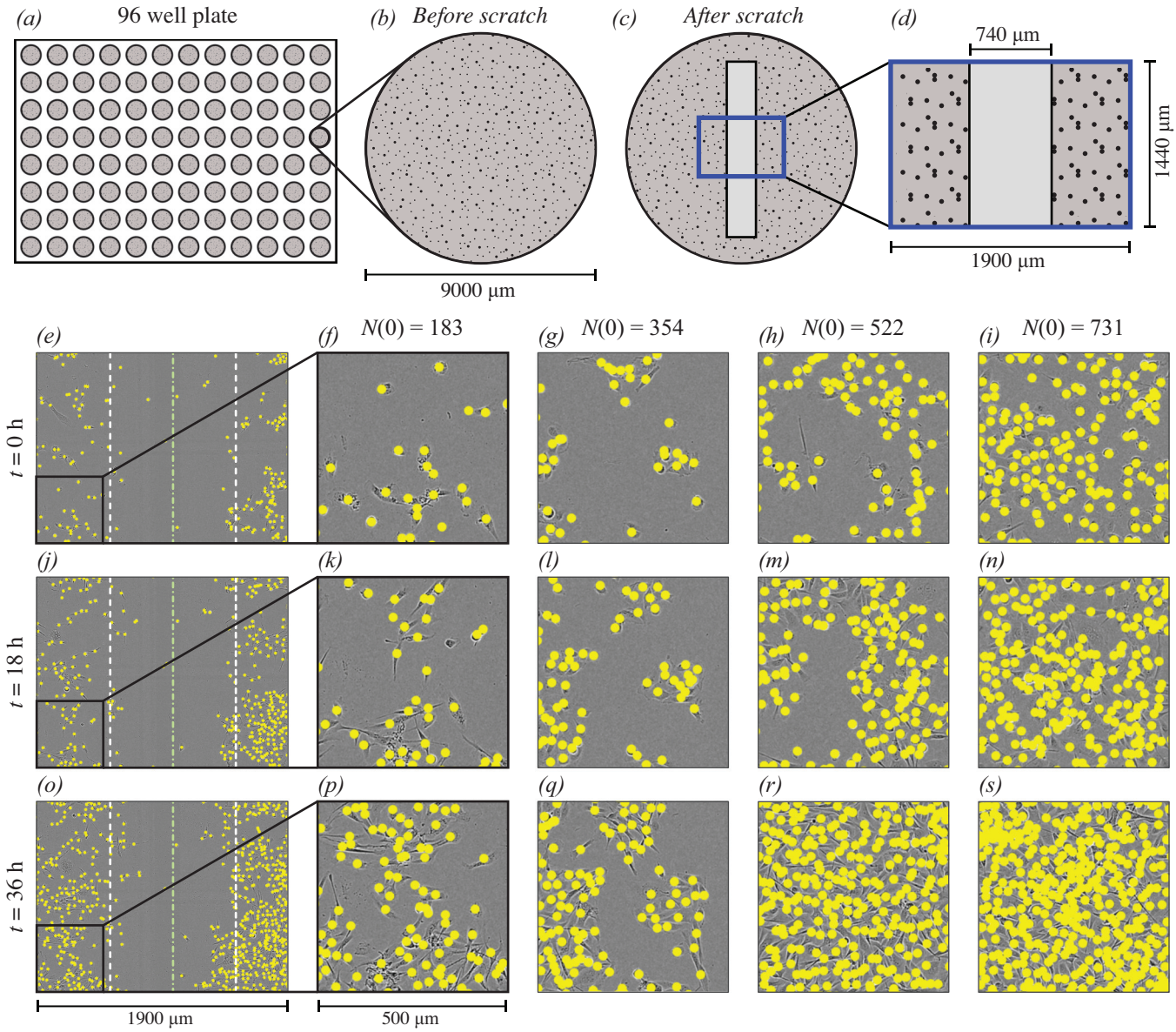

Figure 1: (a)–(d) Schematic of the experimental geometry. (a) 96 well plate. (b) Each assay was performed by distributing a monolayer of cells in a well of diameter 9000 µm. (c) An artificial wound (light region, not to scale) created within the monolayer of cells. (d) Field-of-view of the experimental data, which is much smaller than each well (not to scale). (e),(j),(o) Experimental data for the lowest density replicate (where $N(0) = 183$) at 0 h, 18 h and 36 h, respectively. In (e),(j) and (o) the green dash-dot line represents the approximate centre of the scratch at $t = 0$ h; and, the white dashed lines represent the approximate edge of the scratch at $t = 0$ h. Insets in (f),(k) and (p) show the lower-left region of respective images in (e),(j) and (o). The height and width of the field-of-view in the insets is 500 µm. Subsequent columns show insets for experimental data at increasing densities where $N(0) = 354$, 522 and 731. In each image, the location of each cell is indicated with a yellow marker with diameter $\varphi = 24$ µm (to scale).

2.2 Mathematical model

We use a lattice-free individual based model (IBM) [16, 18] which we simulate with the Gillespie algorithm [27]. The model includes density-dependent proliferation and movement events, but does not consider death, which is not observed in the experiments. To be consistent with previous experimental observations [20], the model incorporates a bias mechanism so that cells both move, and disperse daughter agents during proliferation, in a direction either towards, or away, from crowded regions.

The field-of-view of the experimental data is rectangular, with dimensions 1440 × 1900 µm (figure 1*d*), and we replicate this by using the same geometry in the model. As the well in the tissue culture plate is much larger than this field-of-view, we apply

bioRxiv preprint doi: https://doi.org/10.1101/811257; this version posted October 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

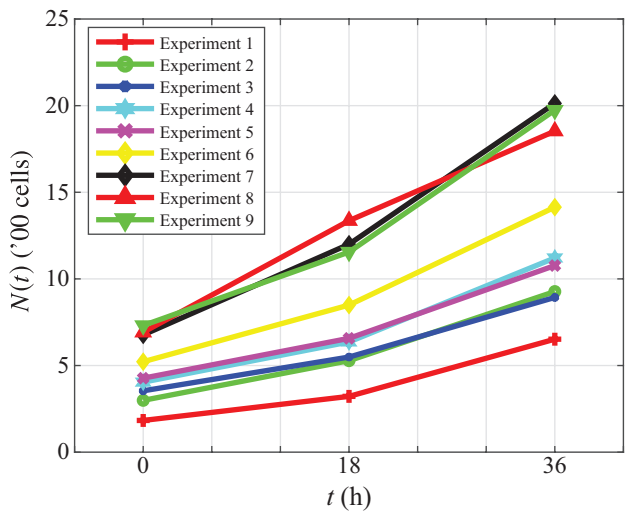

Figure 2: Summary of experimental data showing the cell count, $N(t)$, for all nine experiments.

periodic boundary conditions [16] (indicated in blue in figure 1*c*,*d*). Cells are modelled as *agents* that have a point location but no physical size. In our previous work, we find that, on average, these PC-3 prostate cancer cells have an area that corresponds to a disc of diameter $\varphi = 24\,\mu\text{m}$ [16]. The interaction mechanisms we model are not based on volume exclusion, but rather depend agent separation in such a way that configurations wherein two agent centres are very close are unlikely. We denote the agent locations $\mathbf{x}_n = (x, y)$, $n \in \{1, ..., N(t)\}$, where $N(t)$ denotes the number of agents in the simulation. We specify the initial agent locations in each simulation to match the experimental images at $t = 0\,\text{h}$.

Directional bias

We quantify crowding by placing a *bias kernel* at the location of each agent to form a *crowding surface*, $B(\mathbf{x})$, as shown in figure 3*c*,*d* for the configuration of cells in figure 3*a*,*b*. Mathematically, this is given by

$$B(\mathbf{x}) = \sum_{i=1}^{N(t)} w^{(b)}(\|\mathbf{x} - \mathbf{x}_i\|) \tag{2}$$

and describes a measure of local crowding at $\mathbf{x}$, where $w^{(b)}(r)$ is the bias kernel. The contributions of each agent to $B(\mathbf{x})$ depend on the distance between $\mathbf{x}$ and the location of the ith agent, $\mathbf{x}_i$, given by $r = \|\mathbf{x} - \mathbf{x}_i\|$. In this study, we choose $w^{(b)}(r)$ to be a Gaussian [28] of spread σ with an extremum of γ_b so that

$$w^{(b)}(r) = \gamma_b \exp\left(-\frac{r^2}{2\sigma^2}\right). \tag{3}$$

For computational efficiency, we truncate the kernel to zero for $r \geq 3\sigma$ [28].

For $\gamma_b > 0$, agents prefer to move and disperse daughter agents in the direction of steepest descent on the crowding surface, which corresponds to regions of lower density (setting $\gamma_b < 0$ has the opposite effect). This preference depends on the steepness, so that agents close to highly crowded regions are more likely to move and disperse daughter agents in their preferred direction, demonstrated in figure 3*e*,*f*, where the red agent has a stronger bias strength than the green agent. To do this, we define the bias vector of agent n as

$$\mathbf{B}_n = -\nabla B(\mathbf{x}_n), \tag{4}$$

bioRxiv preprint doi: https://doi.org/10.1101/811257; this version posted October 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

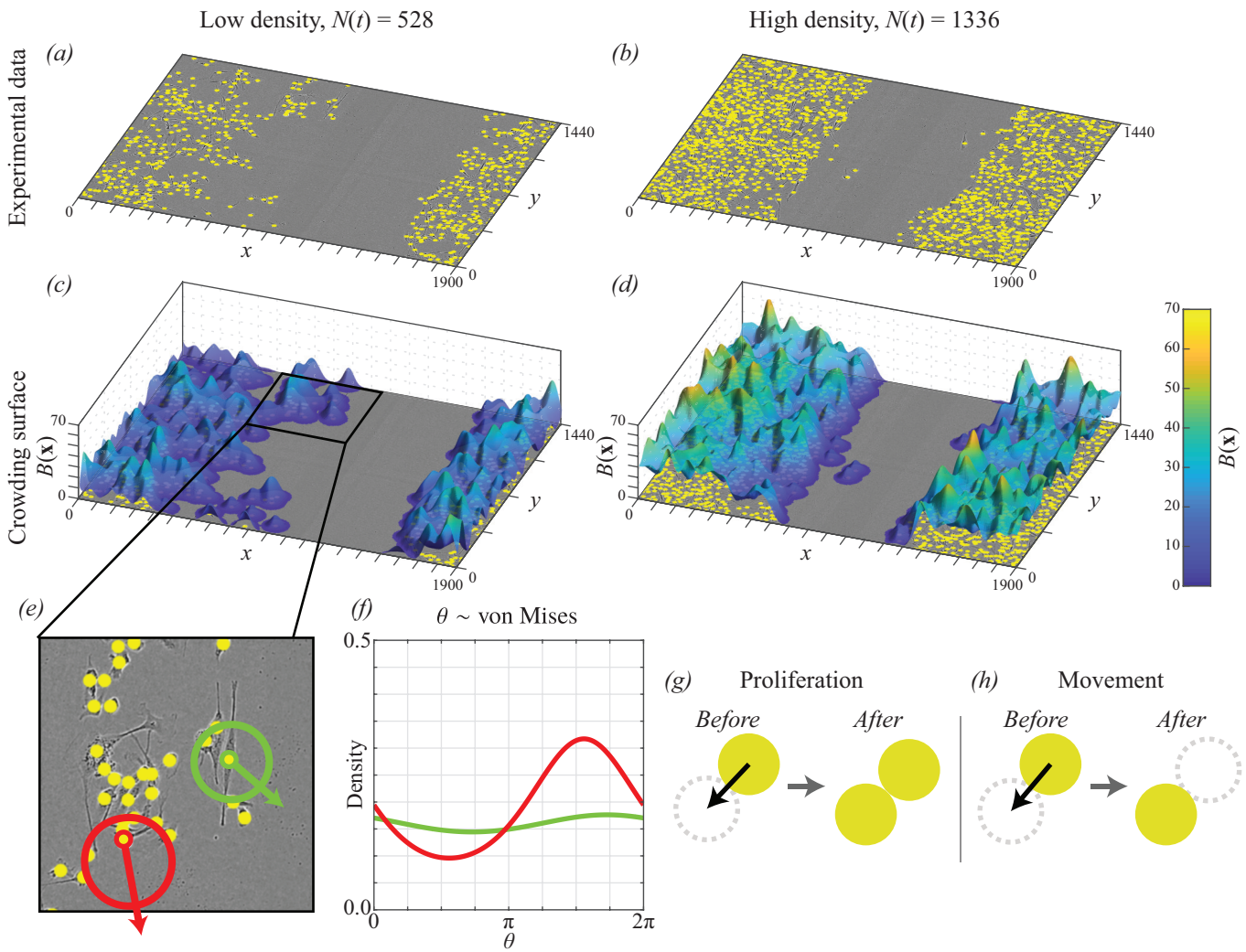

Figure 3: (a)–(b) Experimental data at (a) low density; and, (b) high density. The location of each cell is indicated with a yellow marker of diameter $\varphi = 24\,\mu\text{m}$ (to scale). (c)–(d) Example crowding surface, where agent locations are taken from experimental data in (a) and (b), respectively. (e) Two-dimension inset of experimental image in (c), showing the bias distribution for two agents in radial coordinates centred at each agent. The off-centredness of each circle therefore represents the strength of the bias, which is stronger for the red cell than the green cell. (f) The bias distributions in (e) shown in Euclidean coordinates for clarity. (g)–(h) Schematic of proliferation and movement events, respectively, where the black arrow indicates the sampled direction of each cell. For a proliferation event in (g), the daughter cell is placed a distance of φ from the mother cell. For a movement event in (h), the cell is moved a distance of φ.

which gives the magnitude and direction of steepest descent. The movement and proliferation directions are then sampled from the von Mises distribution [29]

$$\text{von Mises}(\arg(\mathbf{B}_n), \|\mathbf{B}_n\|). \tag{5}$$

The expected and most likely direction is, therefore, $\arg(\mathbf{B}_n)$. The direction distribution becomes increasingly concentrated around $\arg(\mathbf{B}_n)$ as $\|\mathbf{B}_n\|$ becomes large, and approaches a uniform distribution on $[0, 2\pi)$ as $\|\mathbf{B}_n\| \to 0$.

We illustrate the directional bias mechanism in figure 3*c*–*f*. The crowding surface is constructed with a Gaussian kernel placed at the location of each agent (figure 3*c*,*d*). In figure 3*e* we show the bias distribution and preferred direction for an agent in a low (green) and high (red) density region. For each agent, the arrow shows the preferred direction, and the corresponding von Mises distribution is plotted in radial coordinates centred at the location of each agent. In figure 3*f* we show these distributions are shown as a function of the angle, $\theta \in [0, 2\pi)$, for clarity.

bioRxiv preprint doi: https://doi.org/10.1101/811257; this version posted October 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Proliferation and movement

Proliferation and movement events occur according to a Poisson process [30] with density-dependent rates $P_n \geq 0$ and $M_n \geq 0$, respectively. These rates comprise constant intrinsic rates $p > 0$ and $m > 0$, that are modified by interactions with neighbouring agents. We quantify these interactions using kernels, $w^{(\cdot)}(r)$, that depend on the separation distance, $r \geq 0$, between an agent and its neighbours, such that

$$P_n = \max\left(0, p - \sum_{i \neq n}^{N(t)} w^{(p)}(r)\right), \tag{6}$$

$$\text{and} \qquad M_n = \max\left(0, m - \sum_{i \neq n}^{N(t)} w^{(m)}(r)\right). \tag{7}$$

Again, we choose the kernels to be Gaussian, with spread σ, so that

$$w^{(p)}(r) = \gamma_p \exp\left(-\frac{r^2}{2\sigma^2}\right), \tag{8}$$

$$\text{and} \qquad w^{(m)}(r) = \gamma_m \exp\left(-\frac{r^2}{2\sigma^2}\right). \tag{9}$$

Here, γ_p and γ_m are the extrema of the proliferation and movement kernels, respectively. A value of $\gamma < 0$ means that crowding increases motility or proliferation; a value of $\gamma > 0$ means that crowding decreases motility or proliferation; and, a value of $\gamma = 0$ means that motility or proliferation is independent of local density. Again, for computational efficiency, we truncate the kernels to zero for $r \geq 3\sigma$ [28].

When an agent at $\mathbf{x}_n$ proliferates, the daughter agent is dispersed to a location of distance φ (approximately one cell diameter) from $\mathbf{x}_n$, with the direction sampled from the bias distribution for that agent (figure 3*e,f*). This is demonstrated in figure 3*g*.

When an agent at $\mathbf{x}_n$ moves, it is moved to a location of distance φ (approximately one cell diameter) from $\mathbf{x}_n$, with the direction sampled from the bias distribution for that agent (figure 3*e,f*). This is demonstrated in figure 3*h*.

2.3 Summary statistics

To match model simulations to the experimental data, we record the locations of agents at both $t = 18$ h and $t = 36$ h. We denote the experimental data at both time points from experiment $i \in \{1, ..., 9\}$ as $\mathbf{X}_{\text{obs}}^{(i)}$, and simulation data from experiment i as $\mathbf{X}_{\text{sim}}^{(i)}$. In this section, we detail how we summarise the high dimensional data $\mathbf{X}$ into lower dimensional summary statistics. This allows us to define a distance function, $d(\mathbf{X}_{\text{obs}}, \mathbf{X}_{\text{sim}})$, that represents the distance between experimental and simulation data.

We aim to capture three key pieces of information in the experiments: (1) the population size; (2) the spatial structure; and, (3) the density profile. The first two pieces of information are related to the first two spatial moments [28], and the last piece of information relates to the wound closure, total population and the spatial distribution of cells. The first spatial moment, the average density, is the number of agents in the population, $N(t)$. The second spatial moment describes the spatial distribution of agents, often characterised by a pair correlation function [2, 18, 28]. In summary, the pair correlation function describes the number of pairs of agents separated by a distance r, relative to if the population were uniformly distributed. Since the data are discrete, we define the pair correlation, $\mathcal{P}(j, t)$, $j \in \mathbb{N}$, which describes the relative number of pairs

bioRxiv preprint doi: https://doi.org/10.1101/811257; this version posted October 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

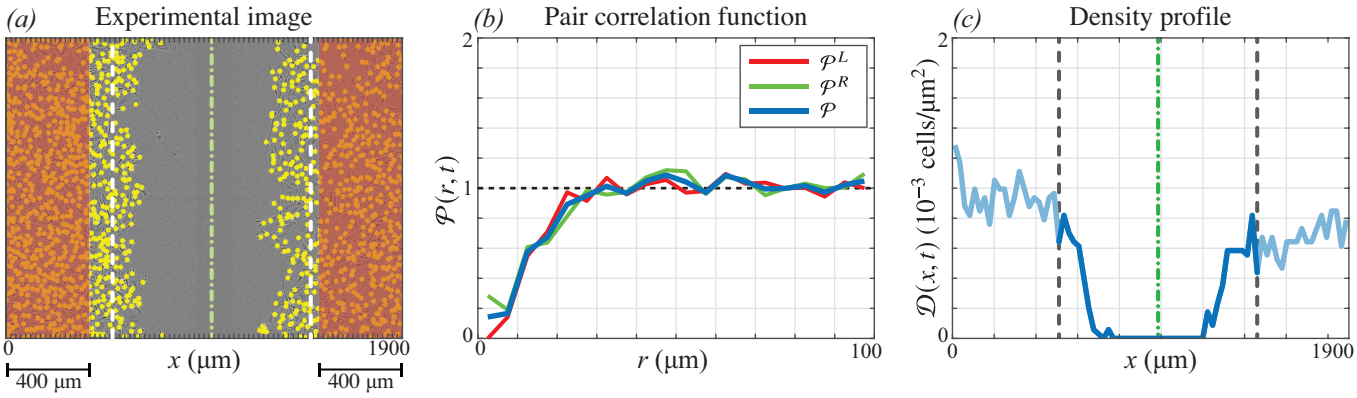

Figure 4: (a) Experimental image. The location of each cell is indicated with a yellow marker of diameter $\varphi = 24\,\mu$m (to scale). (b) Pair correlation function calculated from the distribution in cells in (a). $\mathcal{P}$ is the average of $\mathcal{P}^L$ and $\mathcal{P}^R$, calculated using the agents in the 400 μm to the far-left and far-right (red region) of the experimental data in (a), respectively. Therefore, $\mathcal{P}$ contains information about the spatial distribution of cells, but not about the scratched region. (c) The density profile, calculated by counting the number of cells in subregions of width 1900/80 μm and dividing by the area of each subregion. The sub-regions are indicated as the axis ticks in (a). Only the central 41 bins are used to compare experimental and simulated data. Green dash-dot lines in (a) and (c) indicate the approximate centre of the scratch at $t = 0$ h and dashed lines indicated the approximate boundary of this region at $t = 0$ h.

separated by a distances ranging from $(j-1)\Delta r < r < j\Delta r$. This is given by

$$\mathcal{P}(j,t) = \frac{LW}{N(t)^2 \pi \Delta r (2j + \Delta r)} \sum_{n=1}^{N(t)} \sum_{\substack{i=1 \\ i \neq n}}^{N(t)} \mathbb{1}_{(j-1)\Delta r \leq \|\mathbf{x}_n - \mathbf{x}_i\| < j\Delta r}, \tag{10}$$

where L and W are the dimensions of the region and $\mathbb{1}$ is the indicator function. In this study, we choose $\Delta r = 5\,\mu$m, and consider the pair correlation up to a distance of 100 μm such that $j \leq 20$. Smaller values of Δr lead to a noisier pair correlation function, and larger values of Δr hide information.

In a scratch assay the central region of the experimental field-of-view is approximately devoid of agents (figure 4*a*). To account for this, we calculate pair correlation functions for sub-region of width 400 μm in the far-left, and far-right, of the domain (figure 4*a*, indicated in red) denoted $\mathcal{P}^{(L)}(j,t)$ and $\mathcal{P}^{(R)}(j,t)$, respectively. We apply periodic boundary conditions on these sub-regions, so that the separation of a pair of agents is the smallest possible distance accounting for the periodic boundary conditions. The pair correlation function that summarises the entire experiment is $\mathcal{P}(j,t) = (\mathcal{P}^{(L)}(j,t) + \mathcal{P}^{(R)}(j,t))/2$ (figure 4*b*).

The final piece of information, the density profile, describes the wound closure, total population and spatial structure. We subdivide the field-of-view in figure 4*a* into 80 vertical sub-regions, each of width $\Delta x = 1900/80 = 23.75$ μm. We define the density profile $\mathcal{D}(j,t)$ to be the number of agents with an x-coordinate between $(j-1)\Delta x$ and $j\Delta x$, divided by the area of the sub-region, giving the density. This density profile is shown in figure 4*c*. To avoid capturing excessive noise in our measurement of wound closure, we do not include the entire density profile in the distance metric. Rather, we manually approximate the x-coordinate of the centre of the scratch at $t = 0$ h for each experiment, denoting $I_{\text{mid}}^{(i)}$ as the bin index of the centre the scratch in experiment i. We include the central 41-subregions which, in effect, surround the initially scratched region of each experiment. This region is indicated in figure 4*c* and avoids the fluctuations in density outside this region.

bioRxiv preprint doi: https://doi.org/10.1101/811257; this version posted October 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

The distance metric, $d(\mathbf{X}_{\text{obs}}, \mathbf{X}_{\text{sim}})$, is given by

$$d(\mathbf{X}_{\text{obs}}, \mathbf{X}_{\text{sim}}) = \sum_{t\in\{18,30\}} \Bigg(\frac{[N_{\text{sim}}(t) - N_{\text{obs}}(t)]^2}{N_{\text{obs}}(t)^2} + \frac{\sum_{j=1}^{20}[\mathcal{P}_{\text{sim}}(j,t) - \mathcal{P}_{\text{obs}}(j,t)]^2}{\sum_{j=1}^{20}\mathcal{P}_{\text{obs}}(j,t)^2} + \frac{\sum_{j=I_{\text{mid}}-20}^{I_{\text{mid}}+20}[\mathcal{D}_{\text{sim}}(j,t) - \mathcal{D}_{\text{obs}}(j,t)]^2}{\sum_{j=I_{\text{mid}}-20}^{I_{\text{mid}}+20}\mathcal{D}_{\text{obs}}(j,t)^2} \Bigg), \tag{11}$$

and includes information from all three summary statistics, at $t = 18$ h and $t = 30$ h. Therefore, $d(\mathbf{X}_{\text{obs}}, \mathbf{X}_{\text{sim}})$ is the relative square error of the simulation from the experiment. For $\mathcal{P}$ and $\mathcal{D}$, the contributions to $d(\mathbf{X}_{\text{obs}}, \mathbf{X}_{\text{sim}})$ approximate the relative square error in the integral of each summary statistic, given the spatial discretisation we have applied to each.

2.4 Approximate Bayesian computation and model selection

We consider a hierarchy of models. The full model, which we denote as Model 1, contains the five unknown parameters $\boldsymbol{\theta}_1 = (m, p, \gamma_m, \gamma_p, \gamma_b)$. Models 2 to 5 are subsets of the full model, where we progressively restrict various combinations of the interaction strength parameters γ_m, γ_p and γ_b to be zero, effectively removing them from the model. We summarise these five models in table 2, where we denote $\boldsymbol{\theta}_k$ as the unknown parameter combination for Model k.

	$\boldsymbol{\theta}_k$	Density Dependence
Model 1	$(m, p, \gamma_m, \gamma_p, \gamma_b)$	Proliferation, Motility and Direction
Model 2	$(m, p, \gamma_p, \gamma_b)$	Proliferation and Direction
Model 3	$(m, p, \gamma_m, \gamma_p)$	Proliferation and Motility
Model 4	(m, p, γ_p)	Proliferation only (Fisher-Kolmogorov)
Model 5	(m, p)	None (Skellam's model [31])

Table 2: The hierarchy of models considered. The full model (Model 1) contains a parameter governing the: motility rate, m; proliferation rate, p; motility interaction strength, γ_m; proliferation interaction strength, γ_p; and, directional bias strength, γ_b. In subsequent models, we restrict various combinations of the parameters to zero, effectively removing them from the model.

We treat the unknown parameters in each model as a random variable, $\boldsymbol{\theta}$. In the absence of experimental observations, our knowledge of $\boldsymbol{\theta}$ is characterised by specified prior distributions. When included in the model, the priors were chosen to be independent and are as follows: $\pi(m) = U(0, 10)/\text{h}$; $\pi(p) = U(0.02, 0.05)/\text{h}$; $\pi(\gamma_m) = U(-2, 2)/\text{h}$; $\pi(\gamma_p) = U(0, 0.02)/\text{h}$; and $\pi(\gamma_b) = U(0, 100)$ µm. Initially, we also treat σ as an unknown parameter where $\pi(\sigma) = U(2, 30)$ µm. This initial analysis provides strong evidence for the value of σ, so we set $\sigma = \varphi/2 = 12\,\mu\text{m}$ to decrease the dimensionality of the parameter space. In the supporting material, we also investigate $\sigma = \varphi = 24\,\mu\text{m}$, since this is a natural choice in a lattice-based framework where the migration distance and dispersal distance are also the same as the average agent diameter. We apply approximate Bayesian computation (ABC) [14, 16, 21, 23] to update our knowledge of the parameters using experimental observations, $\mathcal{X}_{\text{obs}}$, from all nine experiments, to produce posterior distributions, $\pi(\boldsymbol{\theta}|\mathcal{X}_{\text{obs}})$. Since this model is known to be computationally expensive [16] and we have a high-dimensional parameter space, we apply an ABC method based on sequential Monte-Carlo (SMC) [21, 23, 32].

bioRxiv preprint doi: https://doi.org/10.1101/811257; this version posted October 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

In this study, we aim to find parameter combinations that simultaneously match all nine experimental data sets, such that $\mathcal{X}_{\text{obs}} = \{\mathbf{X}_{\text{obs}}^{(i)}\}_{i=1}^{9}$. For each prior sample in the ABC rejection algorithm we simulate a model realisation using each experimental initial condition, to obtain $\mathcal{X}_{\text{sim}} = \{\mathbf{X}_{\text{sim}}^{(i)}\}_{i=1}^{9}$. We then compare observed data, $\mathcal{X}_{\text{obs}}$, to simulated data, $\mathcal{X}_{\text{sim}}$, using the discrepancy measure

$$\rho(\mathcal{X}_{\text{obs}}, \mathcal{X}_{\text{sim}}) = \sum_{i=1}^{9} d\left(\mathbf{X}_{\text{obs}}^{(i)}, \mathbf{X}_{\text{sim}}^{(i)}\right), \tag{12}$$

where $d(\cdot,\cdot)$ is given in equation (11). In ABC techniques, we accept a proposal as a posterior sample if $\rho(\mathcal{X}_{\text{obs}}, \mathcal{X}_{\text{sim}}) < \varepsilon$ for some threshold ε. As $d(\cdot,\cdot) \geq 0$, the sum in equation (12) is non-decreasing in i. We therefore implement early rejection [33] by sequentially producing model realisations for $i \in \{1, ..., 9\}$. If, at any time, the partial sum up to a value i exceeds the threshold ε, we immediately reject the sample. In practise, this saves considerable computation time by reducing the number of times the model must be simulated using high-density initial conditions.

The principle behind ABC SMC is to propagate a series of prior samples, called *particles*, through a sequence of distributions $\pi(\boldsymbol{\theta}|\rho(\mathcal{X}_{\text{obs}}, \mathcal{X}_{\text{sim}}) < \varepsilon_u)$, $u = \{1, ..., U\}$ [21, 23, 32]. The thresholds ε_u satisfy $\varepsilon_u > \varepsilon_{u+1}$, so that the distribution gradually evolves to the target distribution $\pi(\boldsymbol{\theta}|\rho(\mathcal{X}_{\text{obs}}, \mathcal{X}_{\text{sim}}) < \varepsilon_U) \approx \pi(\boldsymbol{\theta}|\mathcal{X}_{\text{obs}})$. To obtain a sequence of thresholds, and an estimate of the smallest discrepancy possible in all models, we first perform a pilot run using ABC rejection [16, 23] with Model 1 (supporting material, section 1.1). From 100,000 prior samples, this provides an estimate of the probabilities $\Pr(\rho(\mathcal{X}_{\text{obs}}, \mathcal{X}_{\text{sim}}) < \varepsilon_u)$, given $\boldsymbol{\theta}$ is simulated from the prior. We choose the sequence $\{\varepsilon_u\}_{u=1}^{U}$ by examining a quantile plot (supporting material, section 3). We choose ε_U to corresponds to an acceptance rate of approximately 1% under ABC rejection. The sequence of discrepancies, and details of the ABC rejection and SMC algorithms are given in the supporting material (sections 1 and 3).

We follow the ABC SMC algorithm of Toni *et al.* [21] to perform parameter inference and model selection. Under this algorithm, we place a prior distribution on the model index, $\pi(M_k)$, which we choose to be a discrete uniform distribution so that each model is equiprobable. ABC SMC is then used to estimate the posterior probability of each model, $\pi(M_k|\mathcal{X}_{\text{obs}})$. We detail this algorithm in the supporting material (section 1.2). A key feature of this technique is to implicitly penalise models with a higher number of parameters. We compare models by computing the Bayes factor, $\mathcal{B}_k$ [34], which describe the *evidence* in favour of Model k over the full model, Model 1. As a uniform prior is placed on the model index, the Bayes factor is given by

$$\mathcal{B}_k = \frac{\pi(M_k|\mathcal{X}_{\text{obs}})}{\pi(M_1|\mathcal{X}_{\text{obs}})}. \tag{13}$$

Here, $\pi(M_k|\mathcal{X}_{\text{obs}})$ denotes the marginal posterior density of M_k (Model k). A value $\mathcal{B}_k > 1$ indicates evidence in favour of Model k compared to the full model, and vice-versa for $\mathcal{B}_k < 1$. The Bayes factor is therefore simply the ratio of the posterior density for Models k and 1, and provides evidence to compare models in a similar way to that used in frequentist hypothesis testing.

bioRxiv preprint doi: https://doi.org/10.1101/811257; this version posted October 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

3 Results and Discussion

Common mean-field models, such as the Fisher-Kolmogorov equation and generalisations of the Fisher-Kolmogorov equation, are not able to simultaneously describe collective cell behaviour in scratch assay experiments across a range of initial densities [3]. This suggests density-dependent behaviour in these experiments that is not captured by linear diffusion. Our model allows interactions between cells to affect proliferation, movement and direction. To identify the importance of each of these interactions, we simultaneously calibrate our model to nine scratch assay experiments which we initiate across a wide range of initial densities.

Our first result is to identify the distance over which these interactions occur. We quantify interactions using Gaussian kernels dependent on the distance between pairs of agents [28], and characterised by a spread parameter σ (equations (3), (8) and (9)). The interaction between a pair of agents separated by more than approximately 3σ is, therefore, negligible. We expect σ to be of the same order of magnitude as $\varphi = 24\,\mu\text{m}$, which is the approximate cell diameter [16]. We perform ABC rejection where σ is sampled from the prior $U(2, 30)$ (supporting material, section 1.1). These results suggest that $\sigma \approx \varphi/2 = 12\,\mu\text{m}$, and we fix this for the rest of the study to reduce the number of unknown parameters. This result suggests that interactions between cells occurs over a relatively short distance, since the model predicts interactions between cells separated by more than $3\sigma = 36\,\mu\text{m}$ is negligible.

One of the most important aspects of the lattice-free IBM is its ability to describe, in fine detail, the spatial structure of cells in the experiments, which we quantify using the pair correlation function. In contrast, mean-field models consider only average properties of the cell population [17] and lattice-based methods [19, 22] are not able to precisely capture the initial agent configuration from the experiments. Lattice-based methods also, by definition, constrain the separation of agents to take discrete values, and typically agents in these models cannot lie closer than one cell diameter. The pair correlation describes the probability of finding pairs of agents separated by each distance, and hence can provide information about the effect of interactions on the dynamics. To show this, we repeat ABC rejection but exclude the pair correlation function from the distance metric (supporting material, section 2.3). These results show that the posterior distributions change significantly in this case, verifying that the pair-correlation function contains a significant amount of information about these interactions.

To quantitatively determine the importance of each interaction, we consider a hierarchy of models where we successively set interaction strength parameters (γ_m, γ_p and γ_b) to zero to remove the corresponding interaction from the model. We use the model selection algorithm of Toni *et al.* [21], and compare the evidence in favour of each model over the full model (Model 1) using Bayes factors [21, 34]. We show the posterior density for each model in figure 5*a*, and summarise the Bayes factors and evidence in table 3. Overall, we find that Model 1 has the highest posterior density (figure 5*a*). We find positive evidence in favour of Model 1 over Model 2 (where $\gamma_m = 0$ and so motility is density-independent); and weak evidence in favour of Model 1 over Model 3 (where $\gamma_b = 0$ and so there is no directional bias). Importantly, we find that Models 4 and 5, where $\gamma_m, \gamma_b = 0$ and $\gamma_m, \gamma_p, \gamma_b = 0$, respectively, cannot match the experimental data ($\mathcal{B}_4 = \mathcal{B}_5 = 0$). Contrary to assumptions that are commonly made in models such as the Fisher-Kolmogorov equation, these results provide evidence to suggest that motility is density-dependent, as either a density dependent movement rate must be included (Models 1 and 3) or a directional bias (Models 1 and 2).

bioRxiv preprint doi: https://doi.org/10.1101/811257; this version posted October 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

	$\mathcal{B}_k$	Evidence
Model 1	1.000	–
Model 2	0.248	Positive
Model 3	0.482	Weak
Model 4	0.000	Very Strong
Model 5	0.000	Very Strong

Table 3: Bayes factor, $\mathcal{B}_k$, for each model, which describes the evidence in favour of Model k over Model 1, or vice versa. We also indicate the strength of the evidence in favour of Model 1 over Model k [21].

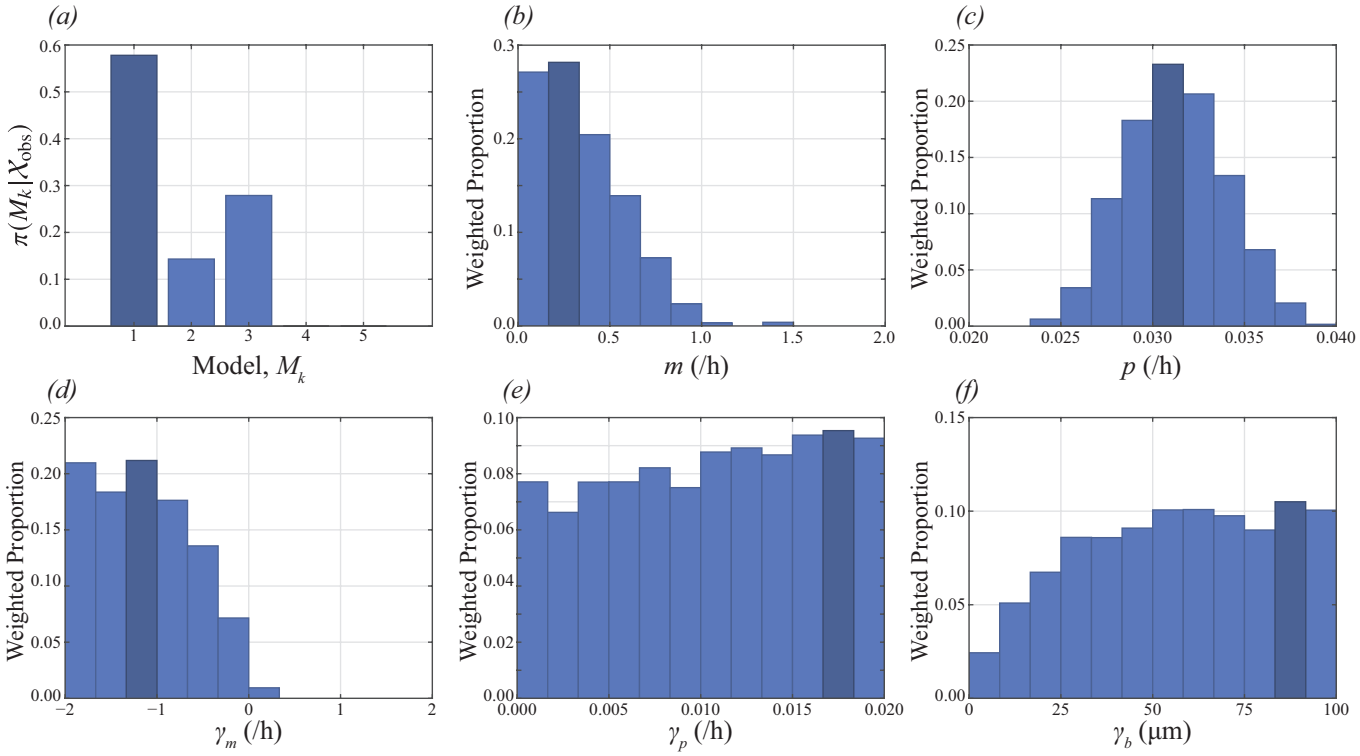

Figure 5: (a) Posterior for the model index, $\pi(M_k|\mathcal{X}_{\text{obs}})$, showing that Model 1 (the full model) is the posterior mode. (b)–(f) Marginal posterior distributions for each parameter in Model 1, shown as weighted histograms. In all cases, the posterior mode is indicated in dark blue.

We now focus on results for the full model (Model 1), which has the highest posterior density. In figure 5*b*–*f* we show marginal posterior distributions for each parameter in Model 1, and in figure 6 we compare the experimental data from four of the nine experiments to the calibrated model (in the supporting material, we show these results for all nine experiments). Overall, we find an excellent match between the model and experimental data, which has not been seen across a range of initial densities for this kind of experimental data. In addition to matching the density profile (figure 6*m*–*p*) and population (figure 6*u*–*x*), we find that the calibrated IBM is able to capture information about the spatial structure of cells, specifically, the pair correlation function (figure 6*q*–*t*). We perform a posterior predictive check for each summary statistic by producing 50% and 95% prediction intervals (PI) that characterise both the parameter uncertainty and stochasticity described by the model. The summary statistics produced from the experimental data almost always lie completely within the 95% PI, further indicating that the calibrated model is consistent with the experimental data across the range of initial densities. While we have not presented these results for Models 2 and 3, which have non-zero posterior density, the nature of ABC means that all accepted samples lie a similar distance to the experimental data.

bioRxiv preprint doi: https://doi.org/10.1101/811257; this version posted October 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

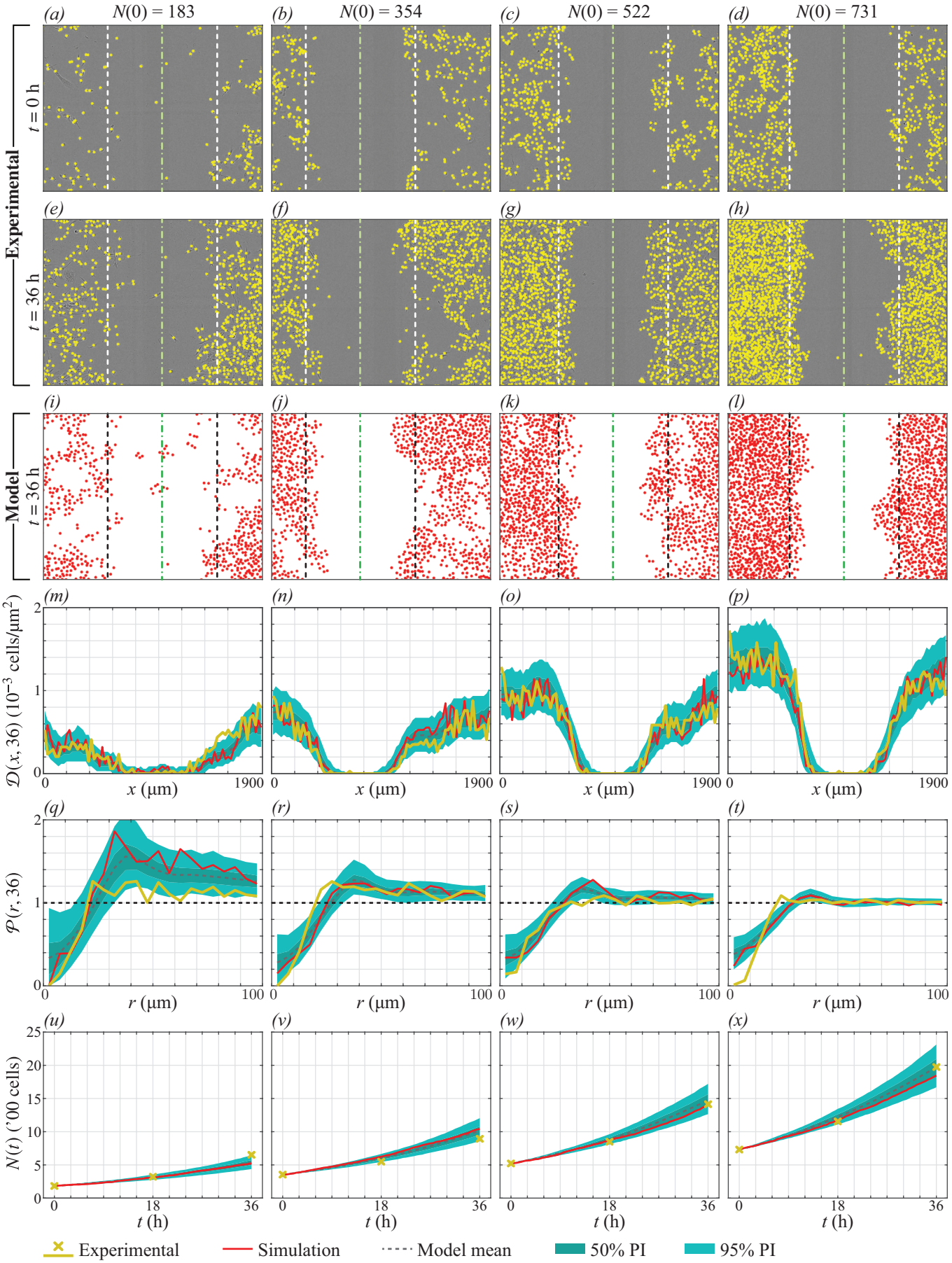

Figure 6: Comparison between experimental data and model estimate. Dashed lines in (a)–(l) give the scratched region that was used to compare the observed and simulated density profile, where green dash-dot lines give the approximate centre of the scratch at $t = 0$ h. Simulation results in (i)–(x) refer to a single realisation of the IBM using the parameter combination that gave the smallest discrepancy. In (m)–(x) the model mean, 50% prediction interval (PI) and 95% PI were produced using 500 realisations of the IBM using parameters resampled from the posterior distribution. Results for all nine experiments are provided in the supporting material (section 6).

bioRxiv preprint doi: https://doi.org/10.1101/811257; this version posted October 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Results in figure 5*c* suggest that $\gamma_m < 0$, so that crowding *increases* motility. This is consistent with mean-field models such as the porous Fisher equation [35] where the diffusivity monotonically increases with local density, but contrasts to other non-linear diffusion models where cell motility decreases with crowding [36]. This observation also explains why model realisations with small values of the motility rate, m, are able to match the data (this is seen in figure 5*b*), since a value $\gamma_m < 0$ allows motility in crowded regions if $m \ll 1$. Interestingly, these results are less clear in the case where the pair correlation function is neglected (supporting material, section 2.3), which again highlights the importance of considering spatial structure when studying these interactions. The increase of motility due to crowding may correspond to mechanical interactions such as volume exclusion in very high density regions. It is trivial to add mechanisms to the IBM, and future work may examine γ_m in the case volume exclusion [19], or other kinds of mechanical interactions [37–39], are included as additional mechanisms. Alternatively, the inclusion of non-monotonic interaction kernels [40] may allow movement to increase for agents close together, and decrease in crowded regions.

An interesting result is that the directional bias is included in the models with the highest posterior density (Models 1 and 3), but examining the marginal posterior for γ_b (figure 5*f*), we see that the strength of this bias may not be identifiable: the posterior distribution is relatively flat without a clear mode. These results might suggest that, past a certain point, increasing the strength of the directional bias as negligible effect. We verify these observations by widening the prior distribution for γ_b by a factor of two in the supporting material (section 2.2). To obtain more information about the strength of the directional bias, representing the propensity of cells to move away from proximate neighbours, more detailed data, such as cell tracking data, may be required [39].

Results in figure 5*e* indicate that the proliferation interaction strength, characterised by γ_p, may also be unidentifiable. Figure 6*u–x* shows that population growth appears to be exponential, and so we do not see crowding effects on proliferation in these experiments. Early time data is often exponential for a variety of growth laws and experiments must be run for a longer period of time to identify the appropriate growth function [22]. We verify this by performing model selection with three additional models (Models 6–8) that respectively correspond to Models 1–3 with $\gamma_p = 0$ (supporting material, section 4). These additional results show that the distributions for Models 6–8 are similar to those for Models 1–3 and confirm that crowding effects on proliferation are simply not seen in these experiments.

4 Conclusion

The ability of common mean-field models, such as the Fisher-Kolmogorov equation and generalisations of the Fisher-Kolmogorov equation, to match experimental data across a range of densities is rarely tested as typical experimental protocols do not vary the initial number of cells. These models typically assume either or both density-dependent proliferation and density-independent motility [3,6–11]. By modelling density-dependent interactions which effect motility, proliferation and directional bias, we calibrate a mathematical model that simultaneously describes scratch assay data across a range of densities. Using model selection, we quantitatively assess which interactions are most important. We find, in contrast to common mean-field models such as the

bioRxiv preprint doi: https://doi.org/10.1101/811257; this version posted October 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Fisher-Kolmogorov equation and generalisations of the Fisher-Kolmogorov equation, evidence to suggest that movement is density-dependent, while there is little evidence of density-dependent proliferation. Additionally, our results confirm the importance of spatial structure, which is neglected by standard modelling approaches.

In this study we study density-dependent interactions that effect proliferation, motility and directional bias. Applying SMC, which penalises models with high dimensionality of the unknown parameters, our study results in the minimal model required to match the experimental data. Two of the primary advantages of the IBM is its ability to precisely replicate the initial condition from experimental data; and, the ease of which new mechanisms can be added. Our approach can, therefore, be applied to quantify experimental evidence for more complex mechanisms including chemotaxis [14, 41]; mechanotaxis [42]; generalised growth laws [43], to name a few. However, we do not pursue such extensions here since we find that our simpler modelling framework can provide a good match to all our experimental data without including more complex mechanisms.

Acknowledgements

M.J.S. is supported by the Australian Research Council, M.J.P. is partly supported by Te Pūnaha Matatini, a New Zealand Centre of Research Excellence, and W.J. is supported by a QUT Vice Chancellor's Research Fellowship. We thank David Warne for technical advice. Computational resources and services used in this work were provided by the HPC and Research Support Group, Queensland University of Technology, Brisbane, Australia.

Authors' contributions

A.P.B. performed the research and wrote the paper. A.P.B. and W.J. processed the experimental data. All authors provided feedback and gave approval for final publication.

Data accessibility

All experimental data are available as electronic supporting material. Unprocessed experimental data, and code used in this work, are available on Github at `github.com/ap-browning/scratchIBM`.

Competing interests

We have no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/811257; this version posted October 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

References

[1] Liang CC, Park AY, Guan JL. 2007 *In vitro* scratch assay: a convenient and inexpensive method for analysis of cell migration *in vitro*. *Nat Protoc* **2**, 329–333.

[2] Johnston ST, McElwain DLS, Binder BJ, Ross JV, Simpson MJ. 2014 Interpreting scratch assays using pair density dynamics and approximate Bayesian computation. *Open Biol* **4**, 140097.

[3] Jin W, Shah ET, Penington CJ, McCue SW, Chopin LK, Simpson MJ. 2016 Reproducibility of scratch assays is affected by the initial degree of confluence: Experiments, modelling and model selection. *J Theor Biol* **390**, 136–145.

[4] Beaumont KA, Mohana-Kumaran N, Haass NK. 2013 Modeling melanoma *in vitro* and *in vivo*. *Healthcare* **2**, 27–46.

[5] Bobadilla AVP, Arévalo J, Sarró E, Byrne HM, Maini PK, Carraro T, Balocco S, Meseguer A, Alarcón T. 2019 *In vitro* cell migration quantification method for scratch assays. *J R Soc Interface* **16**, 20180709.

[6] Sherratt JA, Murray JD. 1990 Models of epidermal wound healing. *Proc Royal Soc B* **241**, 29–36.

[7] Murray JD. 2002 *Mathematical Biology, Third Edition*. Berlin: Springer-Verlag.

[8] Swanson KR, Bridge C, Murray J, Alvord EC. 2003 Virtual and real brain tumors: using mathematical modeling to quantify glioma growth and invasion. *J Neurol Sci* **216**, 1–10.

[9] Maini PK, McElwain DLS, Leavesley DI. 2004 Traveling wave model to interpret a wound-healing cell migration assay for human peritoneal mesothelial cells. *Tissue Eng* **10**, 475–482.

[10] Jackson PR, Juliano J, Hawkins-Daarud A, Rockne RC, Swanson KR. 2015 Patient-specific mathematical neuro-oncology: using a simple proliferation and invasion tumor model to inform clinical practice. *Bull Math Biol* **77**, 846–856.

[11] Rockne RC, Trister AD, Jacobs J, Hawkins-Daarud AJ, Neal ML, Hendrickson K, Mrugala MM, Rockhill JK, Kinahan P, Krohn KA, Swanson KR. 2015 A patient-specific computational model of hypoxia-modulated radiation resistance in glioblastoma using ^{18}F-FMISO-PET. *J R Soc Interface* **12**, 20141174.

[12] Ward JP, King JR. 1997 Mathematical modelling of avascular-tumour growth. *Math Med Biol* **14**, 39–69.

[13] Browning AP, Haridas P, Simpson MJ. 2019 A Bayesian sequential learning framework to parameterise continuum models of melanoma invasion into human skin. *Bull Math Biol* **81**, 676–698.

[14] Lambert B, MacLean AL, Fletcher AG, Combes AN, Little MH, Byrne HM. 2018 Bayesian inference of agent-based models: a tool for studying kidney branching morphogenesis. *J Math Biol* **76**, 1673–1697.

[15] Nardini JT, Chapnick DA, Liu X, Bortz DM. 2016 Modeling keratinocyte wound healing dynamics: Cell–cell adhesion promotes sustained collective migration. *J Theor Biol* **400**, 103–117.

[16] Browning AP, McCue SW, Binny RN, Plank MJ, Shah ET, Simpson MJ. 2018 Inferring parameters for a lattice-free model of cell migration and proliferation using experimental data. *J Theor Biol* **437**, 251–260.

[17] Fadai NT, Baker RE, Simpson MJ. 2019 Accurate and efficient discretizations for stochastic models providing near agent-based spatial resolution at low computational cost. *J R Soc Interface* **16**, 20190421.

[18] Binny RN, James A, Plank MJ. 2016 Collective cell behaviour with neighbour-dependent proliferation, death and directional bias. *Bull Math Biol* **78**, 2277–2301.

[19] Plank MJ, Simpson MJ. 2012 Models of collective cell behaviour with crowding effects: comparing lattice-based and lattice-free approaches. *J R Soc Interface* **9**, 2983–2996.

[20] Cai AQ, Landman KA, Hughes BD. 2007 Multi-scale modeling of a wound-healing cell migration assay. *J Theor Biol* **245**, 576–594.

[21] Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MPH. 2009 Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J R Soc Interface* **6**, 187–202.

[22] Browning AP, McCue SW, Simpson MJ. 2017 A Bayesian computational approach to explore the optimal duration of a cell proliferation assay. *Bull Math Biol* **10**, 1888–1906.

[23] Warne DJ, Baker RE, Simpson MJ. 2019 Simulation and inference algorithms for stochastic biochemical reaction networks: from basic concepts to state-of-the-art. *J R Soc Interface* **16**, 20180943.

[24] Kaighn ME, Narayan KS, Ohnuki Y, Lechner JF, Jones LW. 1979 Establishment and characterization of a human prostatic carcinoma cell line (PC-3). *Invest Urol* **17**, 16–23.

[25] Schneider CA, Rasband WS, Eliceiri KW. 2012 NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671–675.

[26] Jin W, Shah ET, Penington CJ, McCue SW, Maini PK, Simpson MJ. 2017 Logistic proliferation of cells in scratch assays is delayed. *Bull Math Biol* **79**, 1028–1050.

bioRxiv preprint doi: https://doi.org/10.1101/811257; this version posted October 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

[27] Gillespie DT. 1977 Exact stochastic simulation of coupled chemical reactions. *J Phys Chem A* **81**, 2340–2361.

[28] Law R, Murrell DJ, Dieckmann U. 2003 Population growth in space and time: spatial logistic equations. *Ecology* **84**, 252–262.

[29] Forbes C, Evans M, Hastings N, Peacock B. 2011 *Statistical Distributions, Third Edition*. New Jersey: Wiley.

[30] Binder BJ, Simpson MJ. 2013 Quantifying spatial structure in experimental observations and agent-based simulations using pair-correlation functions. *Phys Rev E* **88**, 022705.

[31] Skellam JG. 1951 Random dispersal in theoretical populations. *Biometrika* **38**, 196–218.

[32] Filippi S, Barnes CP, Cornebise J, Stumpf MPH. 2013 On optimality of kernels for approximate Bayesian computation using sequential Monte Carlo. *Stat Appl Genet Mol Biol* **12**, 87–107.

[33] Prangle D. 2016 Lazy ABC. *Stat Comput* **26**, 171–185.

[34] Kass RE, Raftery AE. 1995 Bayes factors. *J Am Stat Assoc* **90**, 773–795.

[35] Warne DJ, Baker RE, Simpson MJ. 2019 Using experimental data and information criteria to guide model selection for reaction-diffusion problems in mathematical biology. *Bull Math Biol* **19**, 1760–1804.

[36] Tremel A, Cai A, Tirtaatmadja N, Hughes B, Stevens G, Landman K, O'Connor A. 2009 Cell migration and proliferation during monolayer formation and wound healing. *Chem Eng Sci* **64**, 247–253.

[37] Murphy RJ, Buenzli PR, Baker RE, Simpson MJ. 2019 A one-dimensional individual-based mechanical model of cell movement in heterogeneous tissues and its coarse-grained approximation. *Proc Royal Soc A* **475**, 20180838.

[38] Stichel D, Middleton AM, Müller BF, Depner S, Klingmüller U, Breuhahn K, Matthäus F. 2017 An individual-based model for collective cancer cell migration explains speed dynamics and phenotype variability in response to growth factors. *NPJ Syst Biol Appl* **3**, 5.

[39] Nan P, Walsh DM, Landman KA, Hughes BD. 2018 Distinguishing cell shoving mechanisms. *PLoS One* **13**, e0193975.

[40] Binny RN, Law R, Plank MJ. 2019 Living in groups: spatial-moment dynamics with neighbour-biased movements. *Ecol Model*.

[41] Codling EA, Plank MJ, Benhamou S. 2008 Random walk models in biology. *J R Soc Interface* **5**, 813–834.

[42] Ozcelikkale A, Dutton JC, Grinnell F, Han B. 2017 Effects of dynamic matrix remodelling on *en masse* migration of fibroblasts on collagen matrices. *J R Soc Interface* **14**, 20170287.

[43] Sarapata EA, de Pillis LG. 2014 A Comparison and Catalog of Intrinsic Tumor Growth Models. *Bull Math Biol* **76**, 2010–2024.