Spatial moment dynamics for collective cell movement incorporating a neighbour-dependent directional bias

Rachelle N. Binny[1,2], Michael J. Plank[1,2], Alex James[1,2]
[1] *School of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand.*
[2] *Te Pūnaha Matatini, New Zealand.*

Abstract

The ability of cells to undergo collective movement plays a fundamental role in tissue repair, development and cancer. Interactions occurring at the level of individual cells may lead to the development of spatial structure which will affect the dynamics of migrating cells at a population level. Models that try to predict population-level behaviour often take a mean-field approach, which assumes that individuals interact with one another in proportion to their average density and ignores the presence of any small-scale spatial structure. In this work, we develop a lattice-free individual-based model (IBM) that uses random walk theory to model the stochastic interactions occurring at the scale of individual migrating cells. We incorporate a mechanism for local directional bias such that an individual's direction of movement is dependent on the degree of cell crowding in its neighbourhood. As an alternative to the mean-field approach, we also employ spatial moment theory to develop a population-level model which accounts for spatial structure and predicts how these individual-level interactions propagate to the scale of the whole population. The IBM is used to derive an equation for dynamics of the second spatial moment (the average density of pairs of cells) which incorporates the neighbour-dependent directional bias and we solve this numerically for a spatially homogeneous case.

Keywords

Collective cell movement; Individual-based model; Spatial moment dynamics; Directed movement

1 Introduction

The ability of cells to migrate as a collective, either to local sites or distant parts of the body, is fundamental for tissue repair [1], development [2] and the immune response [3]. Pathologies such as cancer can arise when the regulatory mechanisms controlling movement are disrupted [4]. A desire to understand how large numbers of individuals are able to coordinate their movement has fuelled extensive studies into the interactions occurring between migrating cells [5–7]. **Some interactions act as attractive forces to drive cells towards one another, for example the physical coupling of neighbouring cells [8] or the release and detection of diffusible chemoattractant signals which give rise to chemotaxis [9,10]. Alternatively, movement in response to a cell-secreted chemorepellant can have a repulsive effect where cells are biased to move away from their nearest neighbours [11, 12]. Other interactions affecting cell motility include crowding effects which can occur at high cell densities. One such effect is contact inhibition of locomotion whereby, after colliding with another individual, a moving cell will slow down then alter its direction of movement in an attempt to avoid future collisions [5, 13].**

The short-range interactions experienced by cells often lead to self-generated spatial structure which can in turn have a significant impact on the dynamics of the cell population [14–16]. For instance, many cell types are known to form clusters or aggregates as a result of attractive interactions [17, 18]. Examples include breast cancer cells [19] and hepatocyte-stellate aggregates [17]. Others, such as retinal neurons [12, 20], arrange themselves into patterns that minimise their proximity to neighbouring cells. This behaviour can be observed in cell populations cultured *in vitro*, however it is not always obvious which underlying mechanisms are responsible for pattern formation, particularly when multiple types of interaction are involved [17]. Therefore there is good motivation for the development of techniques that give more insight into the effects of these mechanisms.

Mathematical modelling can offer explanations to problems for which an experimental approach alone is insufficient [21–23]. The strategy of using random walks [24] to describe cell movement at the scale of individual cells has been discussed extensively in the literature [25–28]. Stochastic models for simulating the movement of large numbers of individuals have been developed. These include individual-based models (IBMs) or agent-based models where each cell is represented as an individual agent and the movements of all agents are tracked over time [29, 30]. Factors such as cell-cell adhesion [19] or a directional bias [11] can also be incorporated. Lattice-free IBMs allow cells to wander freely across a continuous space, thus avoiding the constraints associated with a lattice-based framework where agent locations are restricted to discrete grid sites. For instance, lattice-free IBMs have been shown to result in more realistic spatially-irregular configurations of cells than in equivalent lattice-based approaches [31,32].

Recent research has highlighted the importance of volume exclusion, the concept that the cells themselves occupy space in the domain and may obstruct other individuals from occupying the same space [33, 34]. In lattice-free models volume-exclusion can be incorporated in a number of ways, for example by defining individuals as hard spheres with fixed diameter around which may lie an exclusion area that other individuals cannot occupy [31, 35].

Simulations of IBMs produce synthetic data that can be compared to experimental images [36] and may shed some light on the underlying mechanisms responsible for emerging spatial structure [17], however they are quite intractable mathematically. Deriving a formal

mathematical representation gives more insight into the population-level dynamics of such systems and provides scope for a more rigorous analysis [37]. For simplicity, models describing collective movement at the scale of a population often neglect the effects of spatial structure. They typically deal with a density of individuals that has been averaged over space and explore the evolution of this average density over time. Such models are termed 'mean-field' and assume that individuals are well-mixed or undergo long-range interactions. 'Local mean-field' models, such as reaction-diffusion equations, allow the average density of individuals to be expressed as a function of the location in space however they still tend to ignore the effects of small-scale spatial structure on the population [37]. For example, the Fisher-Kolmogorov equation [38, 39] has been used to describe both cell migration, incorporated in a diffusion term, and proliferation in the form of a logistic growth function [21, 36].

As mean-field models do not account for local interactions they do not always provide a good representation of real behaviour [40]. **Spatial moment theory, originally developed in statistical physics [41–44], can be used to investigate the effect of spatial structure on population-level dynamics [37, 40, 45].** The average density of individuals dealt with in mean-field models is the first spatial moment which holds no information on small-scale spatial structure. One way to access such information is to consider the second spatial moment, the average density of pairs of cells, expressed as a function of the distance r between them. The second moment is often dealt with as a pair correlation function (PCF) $C(r)$ in which it is normalised by the square of the first moment such that in the absence of spatial structure $C(r) \approx 1$. Figure 1 shows the PCF for three spatial point patterns. Figures 1(a)-(c) can each be considered as a snapshot in time from a realisation of an IBM. Figure 1(a) describes a spatial Poisson point process (sometimes referred to as complete spatial randomness) in which all locations of individuals are independent of one another. For this case $C(r) \approx 1$ and no spatial structure is present (Fig. 1(d)). Figure 1(b) shows a cluster pattern, in which pairs of cells are more likely to be found in close proximity. This corresponds to $C(r) > 1$ for short distances r as shown in Figure 1(e). The opposite effect can also arise, whereby cells are less likely to be found close together resulting in a regular pattern (Fig. 1(c)). Figure 1(f) shows that $C(r) < 1$ at short distances r for this type of spatial structure [46].

In previous studies, PCFs have been calculated from experimental images to quantify the extent of spatial structure in live cell populations which adopt Poisson [47], cluster or regular patterns [18,48] to varying degrees. For instance, time-lapse imaging of *in vitro* cell migration assays, such as circular barrier assays [47] and scratch assays [49], generates data in two spatial dimensions. Image analysis techniques can then be employed to measure the distances between cell pairs and this data used to calculate a PCF. PCFs have also been used alongside experimental data to give insight into the mechanisms responsible for pattern formation [17].

Exploring the dynamics of the second moment can provide insight into how the spatial structure is changing over time and whether the state of the system converges. The dynamics of the third moment, the average density of triplets, can be derived to provide further information still, and so on up to the n-th moment however the descriptions of the dynamics become increasingly complex for higher moments [40, 45]. In order to solve the dynamical system a suitable closure is also required because the dynamics of each moment depend on the next moment in the hierarchy. Mean-field models employ a first order closure (the mean-field assumption) in which the second moment is assumed to equal the first moment squared. In other words, it is assumed that individuals encounter one another in proportion to their average density. Thus, in the mean-field assumption any spatial information that was held in the

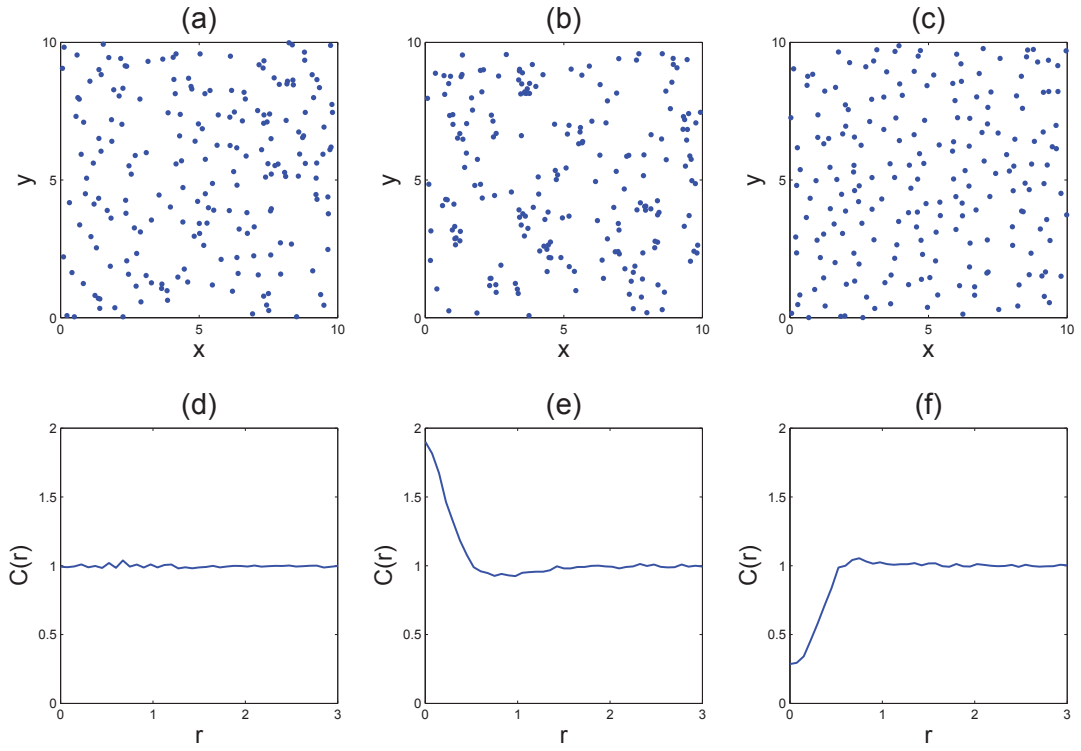

Figure 1: Three different spatial patterns in two-dimensional homogeneous space (a)-(c) and the pair correlation functions $C(r)$ for each pattern (d)-(f). (a) Poisson spatial pattern (or complete spatial randomness) in which there is no spatial structure present; (b) Cluster spatial pattern; (c) Regular spatial pattern.

second moment is lost. However, models which close the dynamics at higher orders retain the spatial information held by the second moment. At second order a number of different closures are possible, for example the Kirkwood superposition approximation [50, 51].

The types of local interactions inherent to migrating cell populations are also of importance in other contexts, for example in animal or plant communities. Many of the modelling tools that employ spatial moment theory were developed for ecological problems [52–55]. Models for dynamics of the second moment which incorporate mechanisms for birth, growth, death and movement (either in isolation or combination) have been derived. In particular they have been used to explore the effects of small-scale spatial structure on plant populations [40, 52], including the relationship between spatial arrangement and plant size distribution [56]. Models for animal populations undergoing movement have been derived both for the case where movement is dependent on local neighbourhood and the independent case [45,57]. These have also been extended to describe the role of spatial structure in predator-prey relationships [57].

Spatial moment models often assume a homogeneous spatial distribution (the pattern is stationary over space). Here we use the term spatially homogeneous to refer to a situation where the probability of finding an individual in a given small region does not depend on the location in space. This is the same as assuming that the spatial structure observed in a small window within a larger space is independent of the position of the window, i.e. it has translational invariance. In terms of spatial moments this corresponds to a first moment that is constant over space, while the second and third moments can be expressed in terms of displacements between pairs of agents as opposed to agent locations [37, 46, 58].

In some cases of collective cell movement it is necessary to consider a non-homogeneous setting, where the average density of cells is higher or lower in certain regions. For example, a non-homogeneous initial condition would be required for the modelling of cell invasion assays in which moving fronts of cells are formed [36, 59]. However, while moment models incorporating terms for density-dependent birth, death and movement have been derived for a spatially non-homogeneous case, solving the dynamics up to at least second order is more complicated and as a result has received significantly less attention than simpler homogeneous systems [37].

In this paper we will describe a lattice-free one-dimensional IBM for collective cell movement. **We incorporate short-range interactions by allowing an individual's rate and direction of movement to depend on the degree of crowding in its neighbourhood. This local directional bias is representative of attractive or repulsive forces occurring between cells, such as in response to a chemoattractant or repellant, and generates spatial structure in the population.** Finally, we derive a corresponding description in terms of the dynamics of spatial moments. Cell movement models incorporating a local directional bias, or similar crowding effects, have previously been discussed in the literature both in a lattice-based framework [10] and lattice-free [31,34]. Similarly, the application of spatial moment theory to modelling lattice-free population-level dynamics of moving individuals has been explored [37,45,57]. **However, the incorporation of a neighbour-dependent directional bias into a second order spatial moment model for lattice-free cell movement was not considered until recently [60] and there is scope for further work in this area.** We assume a setting in which spatial structure is homogeneous, although we will derive equations for the first and second moment which could be applied in a non-homogeneous case.

In reality, motile cell types possess dynamic cytoskeletons which enable them to change their shape and flex around neighbouring cells [13, 61]. For this reason, cells rarely form perfect spheres and it can be difficult to accurately estimate their average diameter. Therefore, we choose not to use a hard-core approach to account for volume exclusion but instead represent the location of a cell by its coordinates in space. Rather than explicitly excluding neighbours from the space surrounding an individual, we consider a kernel (a Gaussian function) which weights the strength of an individual's interaction with its neighbours. The kernel width corresponds to the range over which an individual will affect other cells in its neighbourhood and can be considered a proxy for average cell diameter.

The majority of cell biology experiments are carried out in two or three spatial dimensions. However, numerically solving the moment dynamics up to second order can become quite complicated in higher dimensions and so here we consider a simpler case of movement through one-dimensional space. We show that the one-dimensional model can still capture the qualitative traits of spatial structure inherent to populations in which short-range interactions are important, **i.e. clustering and regular patterns observable in cell populations cultured *in vitro*.** We will also demonstrate that in most cases our model provides a good approximation to the behaviour that is predicted by averaging results obtained by running repeated realisations of the IBM.

2 Individual-Based Model

We consider the collective movement of n individuals through a one-dimensional continuous finite domain with periodic boundaries at $x = x_l$ and $x = x_r$. Our model is a continuous time Markov process model

in which the state of the system $\mathbf{x}(t)$ at time t is

$$\mathbf{x}(t) = (x_1(t), x_2(t), \ldots, x_n(t))^T, \tag{1}$$

where x_i is a coordinate representing the location of cell i. A movement event occurs to cell i as

$$x_i \mapsto x_i + r. \tag{2}$$

The rate density (i.e. the rate) of this transition is $\psi_i(\mathbf{x})\mu(x_i, x_i + r)$, where ψ_i is the movement rate per unit time of cell i and $\mu(x_i, x_i + r)$ is a probability density function (PDF) for movement by a distance r. We simulate this stochastic process using the Gillespie algorithm [62]. In the following description we make choices for the functions ψ_i and $\mu(x_i, x_i + r)$, however these can be easily adapted to suit different experimental situations.

The movement rate ψ_i has dimensions T^{-1} and comprises two terms: an intrinsic motility rate m, i.e. the rate at which an isolated cell would move, and a neighbour-dependent component. The latter term sums a contribution $w(z)$ from each of the other cells in the population, where $w(z)$ is a kernel weighting the strength of interaction between a pair of cells displaced by z. Therefore the movement rate for an individual at x_i with n neighbours at x_j is

$$\psi_i = \max(0,\ m + \sum_{\substack{j=1 \\ i \neq j}}^{n} w(x_j - x_i)). \tag{3}$$

This definition ensures that $\psi_i \geq 0$. For simplicity, the interaction kernel $w(z)$ is a Gaussian function

$$w(z) = \alpha \exp\left(-\frac{z^2}{2\sigma_1^2}\right), \tag{4}$$

where α and σ_1^2 determine strength and range of interaction respectively.

This choice of kernel means that cells interact strongly with near neighbours but are not influenced by those further afield. For $\alpha > 0$, cell i's motility ψ_i is increased by the presence of close-lying neighbours. This type of interaction is relevant from a biological perspective, for example in collective movement involving cell types which release motility-enhancing diffusible signalling factors into their environment. The high concentrations of signals found at high cell densities can result in increased motility rates for cells in crowded regions [63]. On the other hand if $\alpha < 0$ then the presence of close-lying neighbours will reduce ψ_i. For instance, crowding effects such as contact inhibition of locomotion reduce motility at high local cell densities [13, 64].

When a cell undergoes a movement event it takes a step of displacement r from x to y, drawn from a movement PDF $\mu(x, y)$. In the unbiased case where an individual's direction of movement is not affected by the presence of neighbouring cells, we define $\mu(x, y)$ to be a Laplace distribution

$$\mu(x, y) = \frac{\lambda_r}{2} \exp\left(-\lambda_r |y - x|\right), \tag{5}$$

where the mean step length taken by a cell is $1/\lambda_r$. This means cells are more likely to take short steps than undergo large jumps across the space and so is biologically reasonable [5].

The model described so far allows simulation of collective movement in which an individual's motility is influenced by the cell density in its local neighbourhood, as can be observed experimentally [5, 22]. We now incorporate a directional bias $b(x)$ such that the presence of neighbouring cells affects the direction

of movement of an individual at x. Each neighbour contributes $v'(z)$ to the movement of the individual as follows:

$$b(x) = \sum_{j=1}^{n} v'(x_j - x) \ , \tag{6}$$

where $v'(z)$ denotes the derivative of $v(z)$ with respect to z. In theory $v'(z)$ could be replaced by any real-valued kernel which weights the strength of interaction between a cell pair displaced by z. We choose $v(z)$ to be a Gaussian function

$$v(z) = \beta \exp\left(-\frac{z^2}{2\sigma_2^2}\right), \tag{7}$$

with dimension L. This means $v'(z)$ has positive and negative values across its domain and the distinction in sign determines direction of movement.

In order to visualise the total neighbour-dependent effect in $b(x)$ more easily, consider the example in Figure 2 where $\beta > 0$. It shows the total effect of interactions $\sum_{j=1}^{10} v(x_j - x)$, from 10 neighbours located at x_j on a cell at x. To understand why we take a derivative of the interaction kernel $v(z)$ it helps to think of the total weighting function as a 'crowding surface' which a cell at x can use as a means of measuring the extent of crowding in its neighbourhood. In Figure 2, $-b(x)$ is the gradient of this 'surface' and cells are biased to move down the gradient in the direction of reduced crowding. Consider, for example, the arrangement of cells shown in Figure 2(a). Say the cell indicated by the arrow is about to undergo a movement event. At this location the gradient is positive so $b(x) < 0$ which corresponds to a bias for movement in the left direction, away from the crowded region on the individual's right. Thus, the sign of the gradient holds information about the direction in which crowded regions exist. In addition, steep gradients occur at locations on the edges of clusters while shallow or zero gradients occur either within clusters or in sparsely occupied regions. Therefore the magnitude of the gradient provides a measure for the degree of crowding in a location x. The bias $b(x)$ allows us to tap into the information held by the gradient of a 'crowding surface', for a particular arrangement of cells, and use it to determine the direction of movement for an individual at x.

Due to our choice of $v(z)$, the effect of a neighbour located at y on the direction of movement for a cell at x is greater for small distances $|y - x|$, while for larger distances the effect is negligible. The strength of interaction is determined by the constant β. The variance σ_2^2 is a measure of spread for $v(z)$, affecting the range of displacements over which a pair of cells interact. In Figure 2 we consider two different values of σ_2^2. When σ_2^2 is large, $v(z)$ will have a wide spread that will influence outlying cells as shown in Figure 2(a). On the other hand for small σ_2^2, $v(z)$ will be a narrow kernel and only neighbouring cells in close proximity to the individual will be affected by its presence (Fig. 2(b)).

As a means of relating the bias to an individual's direction of movement, we use $b(x)$ to determine the probability of moving right $p_r(b)$ for a cell at x. Its complement $(1 - p_r(b))$ determines the probability of moving left. For simplicity, we define $p_r(b)$ to be a logistic function

$$p_r(b) = \frac{1}{1 + \exp(-b)}. \tag{8}$$

so that for large $b(x) > 0$ a cell at x is strongly biased to move right, while for large $b(x) < 0$ the bias to move left is strong. When $b(x) = 0$ there is no bias from neighbours (i.e. the cell is either isolated or in the centre of a cluster).

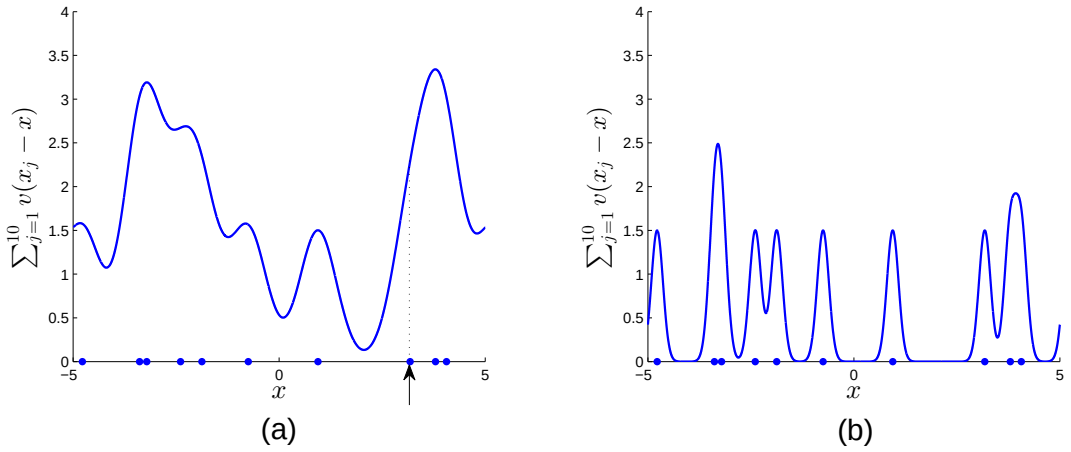

Figure 2: A function $\sum_{j=1}^{10} v(x_j - x)$ (blue line) for the total weighting on a cell at x, of interactions from 10 neighbours located at x_j for $j = 1, \ldots, 10$ (blue dots). The gradient of this function is $-b(x)$ and cells are biased to move down the gradient. The interaction strength is $\beta = 1.5$ and we compare σ_2^2 for different values (a) $\sigma_2^2 = 0.2$, and (b) $\sigma_2^2 = 0.02$. The arrow in (a) marks the location of an individual that is biased to move left.

Finally, we incorporate the directional bias into the movement PDF $\mu(x, y)$ to give a piecewise function

$$\mu(x,y) = \begin{cases} \lambda_r \exp\left(-\lambda_r |y-x|\right)\ p_r(b(x)) & \text{if } y - x > 0 \\ \lambda_r \exp\left(-\lambda_r |y-x|\right)\ (1 - p_r(b(x))) & \text{if } y - x < 0, \end{cases} \tag{9}$$

with dimensions L^{-1}.

In a biological context $v(z)$ could be representative of, say, the extent to which an individual responds to a concentration of chemical signal secreted by a neighbouring cell. Then $b(x)$ would describe the total strength of a cell's response to signals from all neighbours and $p_r(b)$ the mechanism by which these interactions change the cell's direction of movement. The sign of β determines the nature of the directional bias. When $\beta > 0$, as shown in Figure 2, cells are biased to move away from close-lying neighbours. This type of behaviour facilitates movement of individuals out of crowded regions. For example some cell types release chemorepellents which have a repulsive effect on neighbouring cells [11]. Conversely, when $\beta < 0$ the directional bias will drive cells towards one another as may occur in the presence of a cell-secreted chemoattractant [10]. If we set $\beta = 0$ the resulting probability of moving right is $1/2$ and the direction of movement is unbiased. As $\mu(x, y)$ is a PDF we have the constraint that $\int \mu(x, y) \mathrm{d}y = 1$.

3 Spatial Moment Model

The local interactions taking place between cells at the level of individuals give rise to larger scale effects at the population level. In the following sections we introduce a description of the first, second and third moments in terms of the probabilities of individuals being found in given regions. The definitions for the moments given here are equivalent to those given by **Illian, et al** [46]. We then use our IBM to derive a population-level model in terms of the dynamics of the first and second spatial moments. The following

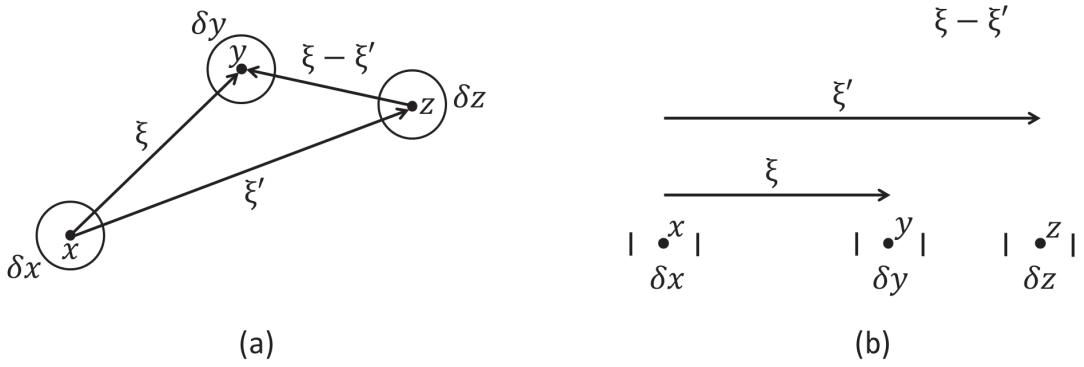

Figure 3: First, second and third spatial moments in (a) two-dimensional space, and (b) one-dimensional space. Small regions δx, δy and δz, of **size** h **(length** h **in one dimension and area** h **in two dimensions)**, are centred on locations x, y and z, respectively. As we are considering a case where cells are distributed homogeneously throughout space, the second and third moment can be expressed in terms of the displacements $\xi = y - z$ and $\xi' = z - x$.

notation and method are consistent with the generalised derivation proposed by **Plank and Law** [37], however we have derived new terms to describe the effect of a neighbour-dependent directional bias.

3.1 Spatial Moments

The first, second and third spatial moments are the average densities of single cells, pairs and triplets, respectively. The concept is better explained by considering the geometry of three small regions δx, δy and δz centred on x, y and z, respectively. Each region has size h **(length** h **in one dimension, area** h **in two dimensions, and volume** h **in three dimensions)** and it is assumed that the probability of finding multiple cells within a single region is $O(h^2)$. For ease of visualisation these regions are depicted in Figure 3(a) in two-dimensional space, however the same principles apply in one dimension (Fig. 3(b)).

The spatial moments are functions of time as well as space but for now we will drop the argument t for ease of notation. The first spatial moment $Z_1(x)$ is expressed in terms of the probability of a cell being found in a small region δx, centred on x and of size h, at time t as follows:

$$Z_1(x) = \lim_{h \to 0} \frac{P(I(x) = 1)}{h} \ . \tag{10}$$

$I(x)$ is an indicator variable such that $I(x) = 1$ if there is a cell in δx centred on x and $I(x) = 0$ if there is no cell in δx.

The second spatial moment $Z_2(x, y)$, the average density of cell pairs, involves the probability of cells being found in the small regions δx and δy as follows:

$$Z_2(x, y) = \lim_{h \to 0} \frac{P(I(x) = 1 \ \& \ I(y) = 1)}{h^2}. \tag{11}$$

For simplicty, we assume that δx **and** δy **cannot overlap and so (11) excludes the case where** $x = y$**. A more rigorous definition which accounts for and removes the effect of such self-pairs (that would otherwise create a Dirac-delta peak in** $Z_2(x, y)$ **at** $x = y$**) is discussed by Plank and Law [37]**. The second spatial moment has a two-fold symmetry such that $Z_2(x, y) = Z_2(y, x)$ [37].

The third spatial moment, the density of triplets in the small regions δx, δy and δz, is defined as

$$Z_3(x,y,z) = \lim_{h\to 0} \frac{P(I(x)=1 \ \& \ I(y)=1 \ \& \ I(z)=1)}{h^3}, \tag{12}$$

excluding the cases where $x = y$, $x = z$ and $y = z$ as we assume that δx, δy and δz cannot overlap. Again, a more detailed description which allows for such non-distinct triplets is given by **Plank and Law** [37]. The third moment has been shown to have a six-fold symmetry [50]. Similarly, we can define the nth spatial moment Z_n as the expected number of n-tuples of cells per unit (length)^{Dn}, for a D-dimensional space.

It is also useful to define some conditional probabilities. The probability of a cell being found in δy conditional on the presence of a cell in δx is $P(I(y) = 1|I(x) = 1)$. We can use the fact that $P(A|B) = P(A\&B)/P(B)$ along with Eqs. (10), (11) and (12) to rewrite this conditional probability as follows:

$$P(I(y)=1|I(x)=1) = \frac{Z_2(x,y)h}{Z_1(x)} + O(h^2). \tag{13}$$

Similarly, we can write the probability of a cell being found in δz conditional on the presence of a cell in δx and a cell in δy as

$$P(I(z)=1|I(x)=1 \ \& \ I(y)=1) = \frac{Z_3(x,y,z)h}{Z_2(x,y)} + O(h^2). \tag{14}$$

3.2 First Spatial Moment

The following derivation can be used to describe moment dynamics in a non-homogeneous space, where the first moment is dependent on x. While the equations are relatively simple to derive, solving them numerically for the non-homogeneous case is not straightforward and so we only solve for a homogeneous space. In addition, as the IBM does not incorporate cell proliferation or death events the first moment is also stationary in time, i.e. its rate of change is zero. However, deriving the equation for the first moment dynamics acts as a good stepping stone to the more complicated second moment dynamics and we include its derivation here.

We derive corresponding descriptions for movement rate ψ_i and PDF $\mu(x, y)$ in terms of spatial moments. In the IBM, the movement rate of individuals comprises an intrinsic component and a neighbour-dependent component which describes the contribution of neighbouring cells to a cell's motility. In the spatial moment dynamics this corresponds to an integration over y of the probability of a cell at y conditional on a cell being present at x. Using the conditional probability in Eq. (13), the expected movement rate (from hereafter simply referred to as movement rate) for a single cell at x is

$$M_1(x) = m + \int w(y-x)\frac{Z_2(x,y)}{Z_1(x)}\mathrm{d}y \ . \tag{15}$$

The maximum formula which ensured a non-negative movement rate in (3) is not incorporated in the spatial moment description because we only consider solutions in which negative expected movement rates do not arise.

When a cell moves, it travels from an original location x to a destination y drawn from the PDF

$$\mu_1(x,y) = \begin{cases} \lambda_r \exp\left(-\lambda_r|y-x|\right) \ p_r(b_1(x)) & \text{if } y-x>0 \\ \lambda_r \exp\left(-\lambda_r|y-x|\right) \ (1-p_r(b_1(x))) & \text{if } y-x<0. \end{cases} \tag{16}$$

The neighbour-dependent bias term $b_1(x)$ sums a contribution $v'(y-x)$ from all possible neighbours at y to the direction of movement of the cell at x. We therefore describe $b_1(x)$ as an integration over y of the probability of a neighbour at y conditional on the presence of a cell at x, weighted by an interaction kernel $v'(y-x)$:

$$b_1(x) = \int v'(y-x)\frac{Z_2(x,y)}{Z_1(x)}\mathrm{d}y. \tag{17}$$

When solving for the spatially homogeneous case, $M_1(x)$ is a constant and $\mu_1(x,y)$ can be expressed in terms of the displacement from x to y.

3.3 Dynamics of the First Spatial Moment

For the dynamics of the first spatial moment $Z_1(x)$ we consider the probability that a cell will be present in the small region δx centred on x at a time $t+\delta t$, where δt is a short period of time. For this situation to arise, a cell could have been present in δx at time t and waited. Alternatively, a cell located elsewhere in the space could have moved into δx. Movement events occur over time as an inhomogeneous Poisson process and so the probability of more than one event occurring in δt is $O(\delta t^2)$. We can combine these possibilities into a single statement

$$\begin{aligned} P(\text{cell in } \delta x \text{ at } t+\delta t) \; &= P(\text{cell in } \delta x \text{ at } t)P(\text{cell waited in } [t,t+\delta t]) \\ &\quad + P(\text{cell absent in } \delta x \text{ at } t)P(\text{cell moved into } \delta x \text{ in } [t,t+\delta t]) \; . \end{aligned} \tag{18}$$

The probability that a cell waited in $[t,t+\delta t]$ is

$$P(\text{cell waited in } \delta x \text{ in } [t,t+\delta t]) = 1 - M_1(x)\delta t + O(\delta t^2) \; . \tag{19}$$

The probability that a cell moved into δx in $[t,t+\delta t]$ can be written as a probability that a cell moved from u into δx, integrated over all possible starting locations u as follows:

$$P(\text{cell moved into } \delta x \text{ in } [t,t+\delta t]) = h\delta t \int \mu_1(u,x)M_1(u)Z_1(u,t)\mathrm{d}u + O(\delta t^2) \; , \tag{20}$$

where $M_1(u)Z_1(u,t)$ is the movement rate per unit area at location u. By making use of the Taylor expansion of $Z_1(x,t+\delta t)$ then taking the limit $h,\delta t \to 0$, we can use (10), (19) and (20) to write (18) as

$$\frac{\mathrm{d}Z_1(x,t)}{\mathrm{d}t} \;=\; -M_1(x)Z_1(x,t) + \int \mu_1(u,x)M_1(u)Z_1(u,t)\mathrm{d}u \tag{21}$$

This equation depends on the second spatial moment, incorporated in the movement rate term $M_1(x)$. The first term in (21) describes movement out of x while movement into x is accounted for in the second term as an integration over all possible starting locations u.

3.4 Second Spatial Moment

For the second moment dynamics, we describe a movement rate function $M_2(x,y)$ for a cell at x conditional on the presence of a cell at y. Recall that the neighbour-dependent component of movement rate for a single cell $M_1(x)$ was conditional on the presence of a second cell. Similarly, the neighbour-dependent component for $M_2(x,y)$ is conditional on the presence of a third cell at z and requires the third spatial moment. We use Eq. (14) to write $M_2(x,y)$ as follows:

$$M_2(x,y) = \; m + \int w(z-x)\frac{Z_3(x,y,z)}{Z_2(x,y)}\mathrm{d}z + w(y-x). \tag{22}$$

The third term here accounts for the direct effect of the cell at y on the cell at x. Because the regions δz and δy do not overlap, the third moment does not account for the case where $z = y$ and we add this interaction as a separate term.

In the dynamics of the second moment, a cell at x moves to a new location at y drawn from a PDF $\mu_2(x, y, z)$, where the third argument accounts for the fact that x is in a pair with a cell at z:

$$\mu_2(x,y,z) = \begin{cases} \lambda_r \exp\left(-\lambda_r |y-x|\right)\ p_r(b_2(x,z)) & \text{if } y - x > 0 \\ \lambda_r \exp\left(-\lambda_r |y-x|\right)\ (1 - p_r(b_2(x,z))) & \text{if } y - x < 0. \end{cases} \tag{23}$$

The neighbour-dependent bias term $b_2(x, y)$ represents the contribution of all possible neighbours to the direction of movement of the cell at x in a pair with a cell at y. It is an integration over z of the probability of a third neighbour at z conditional on the presence of a cell at x and a cell at y, weighted by the kernel $v'(z - x)$. Thus, we have

$$b_2(x,y) = \int v'(z-x) \frac{Z_3(x,y,z)}{Z_2(x,y)} \mathrm{d}z + v'(y-x). \tag{24}$$

As in Eq. (22), the direct effect of a cell at y on a cell at x must be added as a separate term because the third moment does not account for the degenerate case where $y = z$.

3.5 Dynamics of the Second Spatial Moment

To derive an equation for the rate of change of the $Z_2(x, y)$, we consider the probability of finding a cell in δx and a cell in δy at a time $t + \delta t$:

$$\begin{aligned} P\left(\begin{array}{c}\text{cell in } \delta x \text{ \& cell} \\ \text{in } \delta y \text{ at } t+\delta t\end{array}\right) &= P\left(\begin{array}{c}\text{cell in } \delta x \text{ \& cell} \\ \text{in } \delta y \text{ at } t\end{array}\right) P\left(\begin{array}{c}\text{both cells waited} \\ \text{in } [t, t+\delta t]\end{array}\right) \\ &+ P\left(\begin{array}{c}\text{cell in } \delta y \text{ but} \\ \text{not in } \delta x \text{ at } t\end{array}\right) P\left(\begin{array}{c}\text{cell in } \delta y \text{ waited \&} \\ \text{cell moved into } \delta x\end{array}\right) \\ &+ P\left(\begin{array}{c}\text{cell in } \delta x \text{ but} \\ \text{not in } \delta y \text{ at } t\end{array}\right) P\left(\begin{array}{c}\text{cell in } \delta x \text{ waited \&} \\ \text{cell moved into } \delta y\end{array}\right) \\ &+ P\left(\begin{array}{c}\text{cell absent at} \\ \delta x \text{ and } \delta y \text{ at } t\end{array}\right) P\left(\begin{array}{c}\text{cell moved into } \delta x \text{ \&} \\ \text{cell moved into } \delta y\end{array}\right). \end{aligned} \tag{25}$$

The probability of cells being present in both δx and δy can be written in terms of $Z_2(x, y)$ from Eq. (11):

$$P(\text{cell in } \delta x \text{ \& cell in } \delta y \text{ at } t) = Z_2(x,y,t)h^2 + O(h^3). \tag{26}$$

Using (10) and (11), the probability of a cell being present in δx and absent from δy is

$$P(\text{cell in } \delta y \text{ but not in } \delta x \text{ at } t) = Z_1(y,t)h - Z_2(x,y,t)h^2 + O(h^3). \tag{27}$$

The probability of both cells waiting in $[t, t + \delta t]$ in written in terms of (22):

$$P\left(\begin{array}{c}\text{both cells waited} \\ \text{in } [t, t+\delta t]\end{array}\right) = 1 - (M_2(x,y) + M_2(y,x))\delta t + O(\delta t^2). \tag{28}$$

This is comparable to Eq. (19) for the first moment dynamics.

The probability of a cell in δy waiting and a cell moving into δx in $[t, t+\delta t]$ is equivalent to the conditional probability that a cell arrives in δx given that there is a cell in δy. As in (20), we integrate over all possible starting locations u for the cell arriving in δx. However, the probability of a cell being located at u is conditional on the presence of a cell at y. Therefore we have

$$P\left(\begin{array}{c}\text{cell in } \delta y \text{ waited \&}\\ \text{cell moved into } \delta x\end{array}\right) = h\delta t \int \mu_2(u,x,y) M_2(u,y) \frac{Z_2(u,y,t)}{Z_1(y,t)} \mathrm{d}u + O(\delta t^2). \tag{29}$$

Finally, the probability that a cell moved into δx and a cell moved into δy is $O(\delta t^2)$ because this would involve two Poisson events occurring during $[t, t+\delta t]$. Similarly, the higher order terms in Eqs. (28) and (29) arise from probabilities involving more than one cell undergoing a movement event in a time δt.

We substitute Eqs. (26)-(29) into (25) and make use of the 2-fold symmetry of $Z_2(x,y,t)$. Using a Taylor expansion of $Z_2(x,y,t+\delta t)$, expanding terms, then letting $h, \delta t \to 0$ which removes the higher order terms, gives

$$\begin{aligned}\frac{\mathrm{d}Z_2(x,y,t)}{\mathrm{d}t} &= -(M_2(x,y)+M_2(y,x))Z_2(x,y,t)\\ &\quad + \int \mu_2(u,x,y)M_2(u,y)Z_2(u,y,t)\mathrm{d}u\\ &\quad + \int \mu_2(u,y,x)M_2(u,x)Z_2(u,x,t)\mathrm{d}u.\end{aligned} \tag{30}$$

Here, the first negative term describes movement out of x, conditional on the presence of a cell at y. The first integral term represents movement into x from a starting location u, conditional on the presence of a cell at y. The remainder are symmetric terms for movement out of and into y. For notational simplicity, from here on we will drop the t from the spatial moment notation.

Equation (30) depends on the third moment and we need to close the system before solving. To achieve this we use the Kirkwood superposition approximation given by

$$\tilde{Z}_3(x,y,z) = \frac{Z_2(x,y)Z_2(x,z)Z_2(y,z)}{Z_1(x)Z_1(y)Z_1(z)}, \tag{31}$$

however other choices of closure are also possible [50]. For a Poisson spatial pattern the third moment is $Z_3(x,y,z) = Z_1^3$ and the approximation in (31) has perfect accuracy.

4 Results

We now compare some numerical results to measure how effectively our spatial moment model approximates the behaviour predicted by the IBM. Numerical techniques are described in the Appendix. This includes a description of how the spatial moments can be expressed in terms of pair displacements (as in Fig. 3) because we are solving for a spatially homogeneous case. To obtain spatial information from the IBM we calculate a PCF $C_{\mathrm{IBM}}(\xi)$ (see Appendix) by averaging the results of repeated realisations. The PCF predicted by the spatial moment model is given by $C_{\mathrm{SM}}(\xi) = Z_2(\xi)/Z_1^2$ such that $C_{\mathrm{SM}}(\xi) = 1$ in the absence of spatial structure. The second moment is isotropic (i.e. it has symmetry about the origin) and therefore we only show $C_{\mathrm{SM}}(\xi)$ for $\xi \geq 0$.

In each realisation of the IBM we distribute the cells at $t = 0$ according to a spatial Poisson process on $[x_l, x_r]$ with intensity n/L. Therefore, initially there is no spatial structure present. The corresponding initial condition for the spatial moment model is to set $Z_2(\xi) = Z_1^2$ at $t = 0$. Results from both models

are compared at time $t = 25$, by which point the system has converged to steady state in the majority of cases.

In the complete absence of interactions, movement rate is determined by the intrinsic component alone and direction of movement is unbiased. It is straightforward to show analytically that the steady state solution for $Z_2(\xi)$ is a constant for this case. Numerical solutions and IBM simulations confirm this.

4.1 Neighbour-Dependent Motility

We first consider a case with neighbour-dependent motility but in the absence of neighbour-dependent directional bias. Figure 4 shows results for different values of α where interaction strength increases from left to right. We choose $\alpha < 0$ to be sufficiently small such that the sum of the motility rate's intrinsic and neighbour-dependent components will give rise to $\psi_i > 0$ with high probability. Due to the stochastic nature of the IBM it is possible that negative motility rates may occur by chance, however the definition of ψ_i given in Eq. (3) ensures that $\psi_i = 0$ for such rare chance events.

In Figures 4(a)-(c) for $\alpha > 0$, $C(\xi) < 1$ at short displacements which corresponds to a regular spatial pattern. When $\alpha < 0$ (Figs. 4(d)-(f)), $C(\xi) > 1$ at short displacements indicating a cluster spatial pattern. Increasing the magnitude of α (i.e. the strength of interaction) increases the extent of spatial structure. The magnitude of α required to generate clustering is less than that needed to form a regular pattern. For example, $\alpha = 10$ gives $C(0) \approx 0.6$ (Fig. 4(c)) while $\alpha = -2.5$ gives $C(0) \approx 1.4$ (Fig. 4(f)), a similar magnitude of departure from a Poisson spatial pattern at $C(\xi) = 1$.

$C_{\mathrm{SM}}(\xi)$ provides a good approximation to $C_{\mathrm{IBM}}(\xi)$ except for $\alpha < 0$ when $|\alpha|$ is large. For $\alpha = -2$ and $\alpha = -2.5$, $C_{\mathrm{SM}}(\xi)$ has converged to a steady state by $t = 25$ but $C_{\mathrm{IBM}}(\xi)$ continues to increase over time at short displacements. The discrepancy between $C_{\mathrm{SM}}(\xi)$ and $C_{\mathrm{IBM}}(\xi)$ can likely be attributed to the increased occurrence of negative motility rates (set to $\psi_i = 0$ as previously discussed), which can accumulate during simulation of the IBM when the magnitude of $\alpha < 0$ is sufficiently large. The chance occurrence of many pairs being found at short displacements, while reasonably rare for the chosen values of α, may cause a positive feedback reaction whereby the motility rate is reduced for these pairs to an extent where they are very unlikely to undergo further movements. Any cells that move into the resulting cluster will also have their motility rates drastically reduced causing the effect to propagate. The spatial moment model does not account for these rare events as it deals only with average behaviour.

For instance, in the stochastic simulations for $\alpha = -1$ none of the motility rates that arose were negative and $C_{\mathrm{SM}}(\xi)$ matched $C_{\mathrm{IBM}}(\xi)$ well. However for $\alpha = -2$ and $\alpha = -2.5$ at $t = 25$ negative motility rates represented 0.1% and 1% of all motility rates respectively. Increasing t beyond this time caused the incidences to further increase. In contrast, the average motility rate $M_2(\xi)$ predicted by the spatial moment model remained positive for all time. While the 0.1% incidence when $\alpha = -2$ was sufficiently low as to be of little or no consequence for $C_{\mathrm{IBM}}(\xi)$, Figure 4(f) shows that even a relatively low incidence of 1% can lead to a mis-match between $C_{\mathrm{IBM}}(\xi)$ and $C_{\mathrm{SM}}(\xi)$. Further increasing the magnitude of $\alpha < 0$ causes a significant increase in the incidences of $\psi_i < 0$ and the fit between $C_{\mathrm{SM}}(\xi)$ and $C_{\mathrm{IBM}}(\xi)$ deteriorates to an even greater extent.

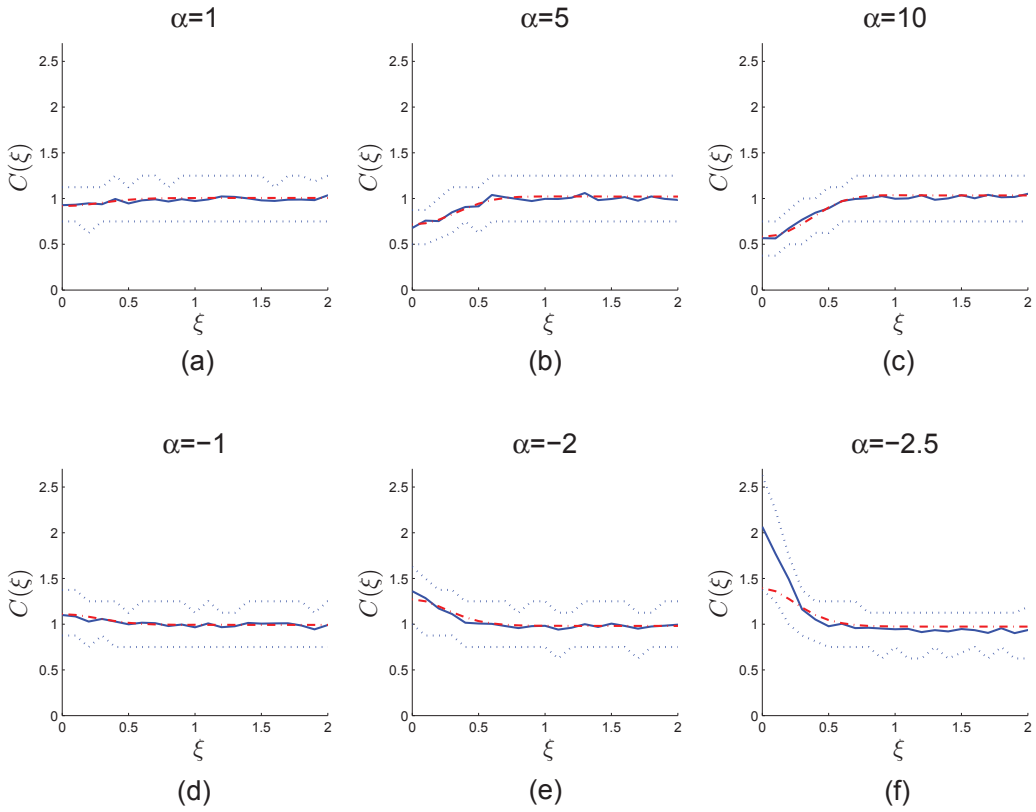

Figure 4: Migration with neighbour-dependent motility but in the absence of neighbour-dependent directional bias ($\beta = 0$). Collective movement of 200 cells in a domain of length $L = 500$ therefore $Z_1 = 0.4$. The PCF $C_{\mathrm{IBM}}(\xi)$ (blue solid) for 300 averaged realisations of the IBM is plotted against the PCF $C_{\mathrm{SM}}(\xi)$ (red broken) predicted by the spatial moment model at time $t = 25$. **Blue dotted lines indicate the interquartile range of IBM results, i.e. 50 % of realisations yield a PCF in the region between the blue dotted lines.** $\sigma_1^2 = \sigma_2^2 = 0.08, \lambda_r = 5, m = 10$. (a) $\alpha = 1$, (b) $\alpha = 5$, (c) $\alpha = 10$, (d) $\alpha = -1$, (e) $\alpha = -2$, and (f) $\alpha = -2.5$.

4.2 Neighbour-Dependent Directional Bias

We now consider a case of migration in the absence of neighbour-dependent motility but in the presence of neighbour-dependent directional bias. We first assume that cells are biased to move away from crowded regions which corresponds to $\beta > 0$. Figures 5(a)-(c) show results for three different values of $\beta > 0$. $C(\xi)$ decreases at small displacements with increasing interaction strength β. In Figure 5(c) for $\beta = 10$ there is a peak in both the $C_{\mathrm{IBM}}(\xi)$ and $C_{\mathrm{SM}}(\xi)$ around $\xi = 1$. This peak arises because the strong directional bias is forcing cells to be displaced as far as possible from their nearest neighbours. This leads to an extreme case of regular spatial pattern in which nearly all cells are separated by approximately the same displacement; the peak in $C(\xi)$ corresponds to this common displacement. The effect of setting $\beta < 0$ such that cells are biased to move towards their neighbours is shown in Figures 5(d)-(f). The magnitude of β required to generate clustering is less than that needed to form a regular spatial pattern.

In Figure 5, $C_{\mathrm{SM}}(\xi)$ provides a good approximation to $C_{\mathrm{IBM}}(\xi)$. However, greater magnitudes of $\beta < 0$ lead to disparities between $C_{\mathrm{SM}}(\xi)$ and $C_{\mathrm{IBM}}(\xi)$. For example when $\beta = -0.5$, $C_{\mathrm{IBM}}(0) \approx 3.6$ at

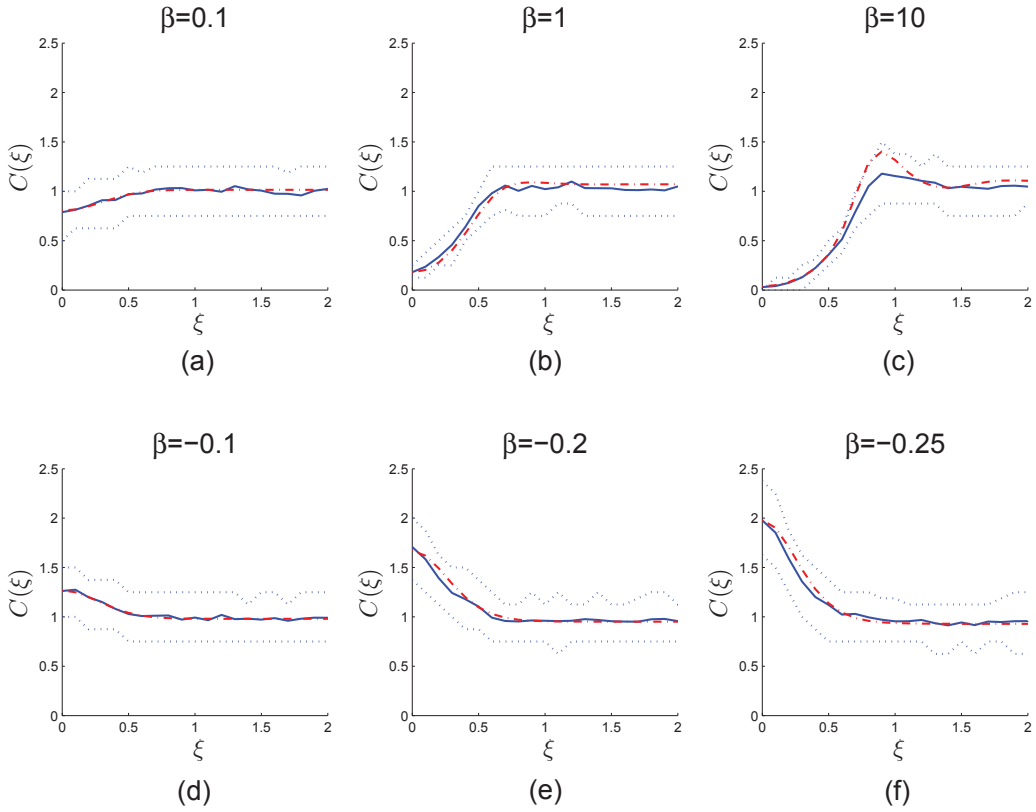

Figure 5: Migration with neighbour-dependent directional bias but in the absence of neighbour-dependent motility ($\alpha = 0$). Collective movement of 200 cells in a domain of length $L = 500$ therefore $Z_1 = 0.4$. The PCF $C_{\mathrm{IBM}}(\xi)$ (blue solid) for 300 averaged realisations of the IBM is plotted against the PCF $C_{\mathrm{SM}}(\xi)$ (red broken) predicted by the spatial moment model at time $t = 25$. **Blue dotted lines indicate the interquartile range of IBM results, i.e. 50 % of realisations yield a PCF in the region between the blue dotted lines.** $\sigma_1^2 = \sigma_2^2 = 0.08, \lambda_r = 5, m = 10$. (a) $\beta = 0.1$, (b) $\beta = 1$, (c) $\beta = 10$, (d) $\beta = -0.1$, (e) $\beta = -0.2$, (f) $\beta = -0.25$.

$t = 25$ while $C_{\mathrm{SM}}(0) \approx 11.5$ and neither PCF has reached steady state. Over time, the cluster pattern becomes stronger and the disparity between the two approximations deteriorates because $C_{\mathrm{SM}}(0)$ is increasing at a faster rate than $C_{\mathrm{IBM}}(0)$.

4.3 Neighbour-Dependent Motility and Directional Bias

Now that we have a better understanding of the independent effects of neighbour-dependent motility and directional bias we will consider the case where both are incorporated together. Figure 6 shows results for four different combinations of α and β. We choose values of α and β that would lead to approximately the same magnitude of departure from a Poisson spatial pattern ($C(\xi) = 1$) at $\xi = 0$ if the neighbour-dependent effects were acting in isolation as in sections 4.1 and 4.2. Figures 6(a) and 6(d) show that when the neighbour-dependent motility and directional bias are working cooperatively to promote spatial structure this results in a greater magnitude of departure from a Poisson spatial pattern than would occur when considering either interaction in isolation. However, when the neighbour-dependent interactions

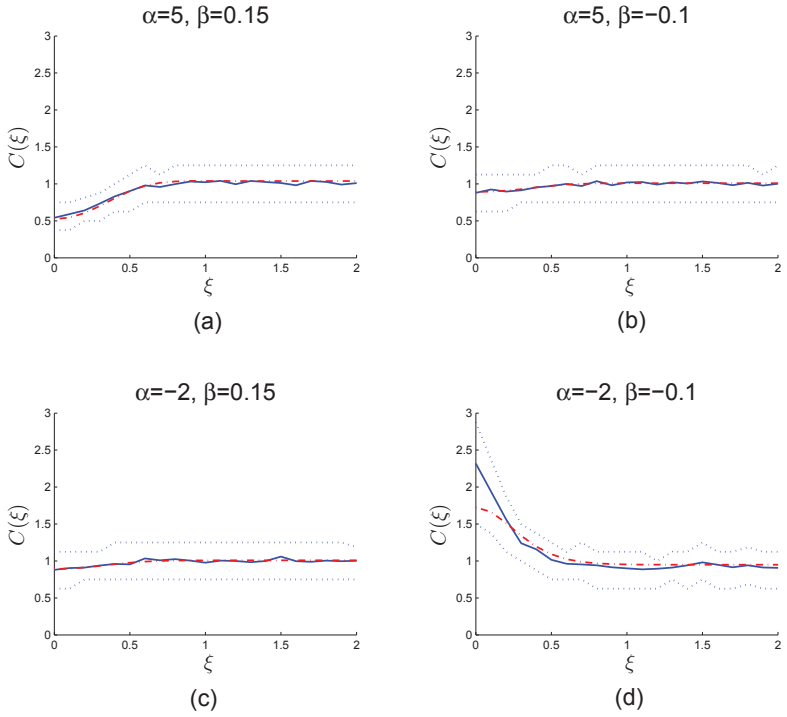

Figure 6: Migration with both neighbour-dependent motility and a directional bias. Collective movement of 200 cells in a domain of length $L = 500$ therefore $Z_1 = 0.4$. The PCF $C_{\mathrm{IBM}}(\xi)$ (blue solid) for 300 averaged realisations of the IBM is plotted against the PCF $C_{\mathrm{SM}}(\xi)$ (red broken) predicted by the spatial moment model at time $t = 25$. **Blue dotted lines indicate the interquartile range of IBM results, i.e. 50 % of realisations yield a PCF in the region between the blue dotted lines.** $\sigma_1^2 = \sigma_2^2 = 0.08, \lambda_r = 5, m = 10$. (a) $\alpha = 5$, $\beta = 0.15$, (b) $\alpha = 5$, $\beta = -0.1$, (c) $\alpha = -2$, $\beta = 0.15$, (d) $\alpha = -2$, $\beta = -0.1$.

are working in opposition (Figs. 6(b)-(c)), they counteract one another and very little spatial structure develops over time as indicated by $C_{\mathrm{SM}}(\xi) \approx 1$. $C_{\mathrm{SM}}(\xi)$ is a good approximation to $C_{\mathrm{IBM}}(\xi)$ except in Figure 6(d). In this case, the slight mis-match near $\xi = 0$ is likely due to the fact that the two forms of interaction working together promote clustering to such a degree that incidences of negative motility rate (approximately 0.8%) in the IBM become important.

The results discussed so far have explored collective movement for $Z_1 = 0.4$. However, the spatial structure that arises due to short-range interactions will depend largely on this average density. We have explored the effect that changing average density Z_1 has on the dynamics of spatial structure. Increasing Z_1 leads to a decrease in the magnitude of local spatial structure. For very high densities, $C(\xi) \approx 1$ for all values of ξ indicating an absence of spatial structure. The effect of changing the width of the interaction kernels has also been explored. If we interpret 2σ (two standard deviations) as the approximate range over which a cell interacts, this can be used as a proxy for the space occupied by a cell and we can give a sense of scale to the spatial domain. It can be shown analytically that there is an equivalence between varying kernel width σ^2 and varying Z_1. A horizontal stretch in the kernels by a factor c leads to the same spatial structure as would increasing Z_1 by a factor c. The second moment $Z_2(\xi)$ is increased by a factor c^2 and horizontally stretched by a factor c. Thus, increasing the range of cell-cell interactions is

equivalent to increasing the average density.

5 Discussion

The IBM enables us to simulate the stochastic behaviour of cells undergoing collective movement and our numerical results demonstrate how individual-level interactions give rise to the development of spatial structure in the population. To obtain an accurate description of average behaviour we either simulate movement for a large number of cells or average the results from many realisations of the model. This approach becomes computationally intensive when cell abundance is high because the interactions between each individual and all of its neighbours must be calculated before every movement event. **In addition, IBMs are not directly amenable to mathematical analysis. Therefore there is good motivation for a population-level model in terms of spatial moment dynamics which provides mechanistic insight into how local directional bias gives rise to spatial structure and creates scope for a more formal analysis of the underlying stochastic process [37].**

Unlike models which employ the mean-field assumption, our spatial moment model takes into account the effects of local spatial structure on the dynamics of the population. **The second moment predicted by the model, expressed as a PCF, provides a measure of this structure and can be directly compared to PCFs calculated from images of *in vitro* cell migration experiments [18].** Our numerical results show that the moment model can provide a good approximation of the spatial structure predicted by the IBM when the distribution of cells is homogeneous throughout space. In the case where interactions affect neighbour-dependent motility but not the direction of movement, the two models mostly match very well both when motility rate is increased in close proximity to neighbours or when it is decreased. However, when interactions that decrease cell motility are strong the moment model tends to under-predict the second moment. This is likely due to it not accounting for the higher incidence of negative motility rates that can arise by chance in the IBM.

When interactions determine only direction of movement and do not affect motility rate the two models again correspond well except when cells are strongly biased to move towards one another. In this case, the spatial moment model over-predicts the second moment. As the motility rate is constant, a rise in negative motility rates can not be causing the disparity in this case. Instead, it is possible that our choice of closure might not be suitable for approximating the third moment when the second moment is large at short displacements. If this is the case, the spatial moment model may also be over-predicting the second moment in the case where strong interactions with close neighbours reduce motility rate and generate clustering. However it is possible that the high average pair densities that arise in the IBM due to the increased incidences of negative motilities could be masking the effect. Using a different closure, for example a power-2 closure, may improve our approximation of the second moment.

When interactions influence both motility rate and directional bias, the results from the IBM and the spatial moments still correspond well in most cases. The two models only start to disagree when the second moment is large for short displacements. In this case the mechanisms that we have seen cause disparity between $C_{\mathrm{SM}}(\xi)$ and $C_{\mathrm{IBM}}(\xi)$ when interactions affecting motility rate and directional bias are considered in isolation, may both contribute to the mis-match in results. However, we have shown that in general our spatial moment model provides a good approximation to the underlying IBM and only starts to break down when the spatial pattern becomes strongly clustered.

As we are primarily interested in the long-term effects of interactions, our main focus is with the

spatial structure of the system at steady state. The time taken to reach steady state depends on both the movement rate and the initial distribution of cells. For ease of comparison between the different types and strength of interactions we considered only a case where cells are initially distributed according to a spatial Poisson process. However, choosing an alternative initial condition, for example a strongly clustered or regular spatial pattern, does not affect the steady state solution approximated by either model.

A similar cell movement model incorporating a local directional bias and employing the second spatial moment was previously derived by Middleton et al. [60]. Their IBM is also lattice-free but is based on Langevin equations describing individual cell velocities. The way in which the directional bias acts is also slightly different. A cell's velocity is taken to be the sum of the interaction forces from all neighbours plus a noise term. In contrast, in our IBM the sum of interactions with neighbours dictates the probability of a cell moving in a certain direction and its movement rate can be determined by an independent mechanism. The choice of interaction kernel is also different to that employed here. Despite these differences our results correspond well with those of Middleton et al. [60], i.e. we see the same qualitative trends in the second moment due to local repulsive or attractive forces.

The closure for the third moment is only an approximation and different closures may perform better under different conditions. For instance, power-2 closures generally perform well but have the potential to violate the positivity constraint which is required because an average density of triplets can never be negative [50]. While it is important to keep this in mind, an exhaustive analysis of moment closures is outside the scope of this work. **Other methods for describing the dynamics of spatial point processes, which do not require a closure assumption, are also discussed in the literature. For instance stochastic differential equations, such as Langevin-type equations, capture fluctuations arising due to short-range interactions via a noise term and can be used to investigate spatio-temporal patterns at different scales [65]. Blath et al. [66] analysed a stochastic, lattice-based model using stochastic differential equations to explore whether spatial structure could give rise to coexistence between two competing species. A closed system of equations for the whole hierarchy of moments was derived by Ovaskainen et al. [67] using techniques from Markov evolutions and a perturbation expansion around the spatial mean-field model. Bruna and Chapman [35] employed a perturbation method to describe the dynamics of moving particles by using matched asymptotic expansions in a small parameter $\epsilon \ll 1$ to derive a nonlinear diffusion equation.**

As our model does not incorporate volume exclusion, for example through the representation of cells as hard objects, there is the possibility that cell locations may arise in very close proximity in the IBM. The use of an interaction kernel which is concentrated around short pair displacements provides a mechanism for generating a regular spatial pattern and thus allows us to reduce the likelihood of two cells being found close together. However, this approach is probabilistic and does not altogether exclude the possibility of such an occurrence.

It is appealing to consider the collective movement of cells in one dimension from a theoretical perspective, in particular because it simplifies the derivation and numerical solution of the spatial moments description. Solving the differential equation in two dimensions is considerably more computationally intensive as it involves double integrations in both the x- and y- direction. While the majority of experimental data is two- or three-dimensional, our results suggest that a one-dimensional model could

still prove useful for quantifying the behaviour of moving cells. In one dimension we observe traits in the second moment that we would expect to see in a live population of cells, namely the development of clusters or regular spatial patterns. However, as a cell moving through two-dimensional space is interacting with neighbours in all directions, not just those on either side, it is possible that this could have a more profound impact on spatial structure than is predicted by our one-dimensional model. Therefore, an important goal of future work will be to extend our model to two dimensions.

Acknowledgements

The authors thank Profs. Matthew Simpson and Richard Law for many helpful discussions. This research was supported by the Royal Society of New Zealand Marsden Fund, Grant Number 11-UOC-005.

Appendix A

Calculating a PCF

The PCF $C(r)$ provides a means of extracting information about spatial structure from the configurations of agents that arise in realisations of an IBM. To calculate the PCF for a particular configuration of agents, a reference agent at x_i is chosen and the distance between x_i and a neighbour at x_j is measured, for n neighbours. We measure across periodic boundaries such that the distance between a pair of agents displaced by $\xi = x_j - x_i$ is

$$r = \begin{cases} |\xi| & \text{if } |\xi| < \frac{L}{2} \\ L - |\xi| & \text{if } |\xi| > \frac{L}{2}. \end{cases} \tag{32}$$

Another reference agent is then chosen and the process repeated until each agent in the population has been selected as a reference once. Once all possible pair distances, excluding self-pairs, have been measured $C(r)$ can be generated by counting the number of distances that fall within an interval $[r, r+\delta r]$. Normalising by $2\delta r n^2/L$ ensures $C(r) = 1$ in the absence of spatial structure.

Numerical Methods

To solve Eq. (30) for the dynamics of the second moment numerically it is beneficial to reduce the number of variables. For a spatially homogeneous distribution of cells, the second moment $Z_2(x, y)$ depends only the displacement $y - x$ which can now be treated as a single variable. As shown in Figure 3, the displacement from x to y is denoted ξ and the displacement from x to z is denoted ξ'. For the movement PDF $\mu_2(u, x, y)$, we denote the displacement from u to x as ξ''. The first spatial moment is required for $\tilde{Z}_3(x, y, z)$ and in the homogeneous case Z_1 is a constant.

We rewrite (30) in terms of the displacements between pairs as follows:

$$\begin{aligned} \frac{\mathrm{d}Z_2(\xi)}{\mathrm{d}t} &= -(M_2(\xi) + M_2(-\xi))Z_2(\xi) \\ &\quad + \int \mu_2(\xi'', \xi'' + \xi) M_2(\xi'' + \xi) Z_2(\xi'' + \xi) \mathrm{d}\xi'' \\ &\quad + \int \mu_2(\xi'', \xi'' - \xi) M_2(\xi'' - \xi) Z_2(\xi'' - \xi) \mathrm{d}\xi''. \end{aligned} \tag{33}$$

The movement rate $M_2(x, y)$ of a cell at x in a pair with a cell at y given in (22) is now expressed in terms of the displacement ξ between x and y:

$$M_2(\xi) \;=\; m + \int w(\xi')\frac{Z_3(\xi, \xi')}{Z_2(\xi)}\mathrm{d}\xi' + w(\xi). \tag{34}$$

The movement PDF given in (23) becomes

$$\mu_2(\xi, \xi') = \begin{cases} \lambda_r \exp\left(-\lambda_r|\xi|\right)\ p_r(b_2(\xi')) & \text{if } \xi > 0 \\ \lambda_r \exp\left(-\lambda_r|\xi|\right)\ (1 - p_r(b_2(\xi'))) & \text{if } \xi < 0, \end{cases} \tag{35}$$

with neighbour-dependent bias

$$b_2(\xi) = \int v'(\xi')\frac{Z_3(\xi, \xi')}{Z_2(\xi)}\mathrm{d}\xi' + v'(\xi). \tag{36}$$

The interaction kernels were previously expressed in terms of a single variable in (4) and (7) and these definitions still hold here. The closure for the third moment is

$$\tilde{Z}_3(\xi, \xi') = \frac{Z_2(\xi)Z_2(\xi')Z_2(\xi' - \xi)}{Z_1^3}. \tag{37}$$

The boundary condition is as follows:

$$Z_2(\xi) \to Z_1^2 \text{ as } \xi \to \infty. \tag{38}$$

Equation (33) was solved using the method of lines with MATLAB's in-built ode45 solver. This involved a discretisation of ξ with grid spacing $\Delta = 0.1$ over the domain $|\xi| \leq \xi_{max}$, where ξ_{max} was large enough so that $Z_2(\xi) \approx Z_1^2$ at $|\xi| = \xi_{max}$. Required values of $Z_2(\xi)$ that lay outside of the computable domain were set to the value of $Z_2(\xi)$ at the boundary. The integral terms in (33) were approximated using the trapezoidal rule with the same discretisation. In addition, the PDF for movement $\mu_2(\xi, \xi')$ was normalised numerically using the trapezoidal rule such that $\int \mu_2(\xi, \xi')\mathrm{d}\xi = 1$. **The results were insensitive to a reduction in grid spacing Δ.**

References

[1] Martin, P. (1997) Wound healing–aiming for perfect skin regeneration. *Science*, **276**, 75–81.

[2] Kurosaka, S. and Kashina, A. (2008) Cell biology of embryonic migration. *Birth Defects Res C Embryo Today*, **84**, 102–122.

[3] Rørth, P. (2009) Collective cell migration. *Annu. Rev. Cell Dev. Biol.*, **25**, 407–29.

[4] Friedl, P. and Wolf, K. (2003) Tumour-cell invasion and migration: diversity and escape mechanisms. *Nat. Rev. Cancer*, **3**, 362–74.

[5] Vedel, S., Tay, S., Johnston, D. M., Bruus, H., and Quake, S. R. (2013) Migration of cells in a social context. *Proc. Natl. Acad. Sci. U. S. A.*, **110**, 129–34.

[6] Angelini, T. E., Hannezo, E., Trepat, X., Marquez, M., Fredberg, J., and Weitz, D. A. (2011) Glass-like dynamics of collective cell migration. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 4714–9.

[7] Trepat, X., Wasserman, M. R., Angelini, T. E., Millet, E., Weitz, D. A., Butler, J. P., and Fredberg, J. J. (2009) Physical forces during collective cell migration. *Nat. Phys.*, **5**, 426–430.

[8] Tambe, D. T., et al. (2011) Collective cell guidance by cooperative intercellular forces. *Nat. Mater.*, **10**, 469–75.

[9] Raz, E. and Mahabaleshwar, H. (2009) Chemokine signaling in embryonic cell migration: a fisheye view. *Development*, **136**, 1223–9.

[10] Painter, K. J. and Hillen, T. (2002) Volume-filling and quorum-sensing in models for chemosensitive movement. *Can. Appl. Math. Quart*, **10**, 501–543.

[11] Cai, A. Q., Landman, K. A., and Hughes, B. D. (2006) Modelling directional guidance and motility regulation in cell migration. *Bull. Math. Biol.*, **68**, 25–52.

[12] Kay, J. N., Chu, M. W., and Sanes, J. R. (2012) MEGF10 and MEGF11 mediate homotypic interactions required for mosaic spacing of retinal neurons. *Nature*, **483**, 465–9.

[13] Abercrombie, M. (1979) Contact inhibition and malignancy. *Nature*, **281**, 259–262.

[14] Markham, D. C., Baker, R. E., and Maini, P. K. (2014) Modelling collective cell behaviour. *Discret. Contin. Dyn. Syst.*, **34**, 5123–5133.

[15] Green, J. E. F., Waters, S. L., Whiteley, J. P., Edelstein-Keshet, L., Shakesheff, K. M., and Byrne, H. M. (2010) Non-local models for the formation of hepatocyte-stellate cell aggregates. *J. Theor. Biol.*, **267**, 106–20.

[16] Diggle, P. J., Eglen, S. J., and Troy, J. B. (2006) Modelling the bivariate spatial distribution of amacrine cells. Baddelely, A., Gregori, P., Mateu, J., Stoica, R., and Stoyan, D. (eds.), *Case Studies in Spatial Point Process Modeling, Lecture Notes in Statistics*, vol. 185, pp. 215–233, Springer.

[17] Agnew, D. J. G., Green, J. E. F., Brown, T. M., Simpson, M. J., and Binder, B. J. (2014) Distinguishing between mechanisms of cell aggregation using pair-correlation functions. *J. Theor. Biol.*, **352**, 16–23.

[18] Binder, B. J. and Simpson, M. J. (2013) Quantifying spatial structure in experimental observations and agent-based simulations using pair-correlation functions. *Phys. Rev. E*, **88**, 022705.

[19] Johnston, S. T., Simpson, M. J., and Plank, M. J. (2013) Lattice-free descriptions of collective motion with crowding and adhesion. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.*, **88**, 062720.

[20] Keeley, P. W., Zhou, C., Lu, L., Williams, R., Melmed, S., and Reese, B. E. (2014) Pituitary tumor-transforming gene 1 regulates the patterning of retinal mosaics. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, 9295–300.

[21] Maini, P. K., McElwain, D. L. S., and Leavesley, D. I. (2004) Traveling wave model to interpret a wound-healing cell migration assay for human peritoneal mesothelial cells. *Tissue Eng.*, **10**, 475–82.

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

[23] Alarcón, T., Byrne, H. M., and Maini, P. K. (2004) Towards whole-organ modelling of tumour growth. *Prog. Biophys. Mol. Biol.*, **85**, 451–72.

[24] Codling, E. A., Plank, M. J., and Benhamou, S. (2008) Random walk models in biology. *J. R. Soc. Interface*, **5**, 813–34.

[25] Peterson, S. C. and Noble, P. B. (1972) A two-dimensional random-walk analysis of human granulocyte movement. *Biophys. J.*, **12**, 1048–55.

[26] Hall, R. L. (1977) Amoeboid movement as a correlated walk. *J. Math. Biol.*, **4**, 327–335.

[27] Alt, W. (1980) Biased random walk models for chemotaxis and related diffusion approximations. *J. Math. Biol.*, **9**, 147–177.

[28] Stokes, C. L., Lauffenburger, D. A., and Williams, S. K. (1991) Migration of individual microvessel endothelial cells: stochastic model and parameter measurement. *J. Cell Sci.*, **99**, 419–30.

[29] Simpson, M. J., Landman, K. A., and Hughes, B. D. (2010) Cell invasion with proliferation mechanisms motivated by time-lapse data. *Phys. A Stat. Mech. its Appl.*, **389**, 3779–3790.

[30] Walker, D. C., Hill, G., Wood, S. M., Smallwood, R. H., and Southgate, J. (2004) Agent-based computational modeling of wounded epithelial cell monolayers. *IEEE Trans. Nanobioscience*, **3**, 153–163.

[31] Plank, M. J. and Simpson, M. J. (2012) Models of collective cell behaviour with crowding effects: comparing lattice-based and lattice-free approaches. *J. R. Soc. Interface*, **9**, 2983–96.

[32] Plank, M. J. and Simpson, M. J. (2013) Lattice-free models of cell invasion: discrete simulations and travelling waves. *Bull. Math. Biol.*, **75**, 2150–66.

[33] Dyson, L., Maini, P. K., and Baker, R. E. (2012) Macroscopic limits of individual-based models for motile cell populations with volume exclusion. *Phys. Rev. E*, **86**, 031903.

[34] Dyson, L. and Baker, R. E. (2014) The importance of volume exclusion in modelling cellular migration. *J. Math. Biol.*.

[35] Bruna, M. and Chapman, S. J. (2012) Excluded-volume effects in the diffusion of hard spheres. *Phys. Rev. E*, **85**, 011103.

[36] Simpson, M. J., Treloar, K. K., Binder, B. J., Haridas, P., Manton, K. J., Leavesley, D. I., McElwain, D. L. S., and Baker, R. E. (2013) Quantifying the roles of cell motility and cell proliferation in a circular barrier assay. *J. R. Soc. Interface*, **10**, 20130007.

[37] Plank, M. J. and Law, R. (2014) Spatial point processes and moment dynamics in the life sciences: a parsimonious derivation and some extensions. *Bull. Math. Biol.*.

[38] Fisher, R. A. (1937) The wave of advance of advantageous genes. *Ann. Eugen.*, **7**, 355–369.

[39] Kolmogorov, A. N., Petrovsky, I. G., and Piskunov, N. S. (1937) Étude de léquation de la diffusion avec croissance de la quantité de matière et son application à un problème biologique. *Moscow Univ Bull Math*, **1**, 1–25.

[40] Law, R., Murrell, D. J., and Dieckmann, U. (2003) Population growth in space and time: spatial logistic equations. *Ecology*, **84**, 252–262.

[41] Bogoliubov, N. N. (1946) Kinetic Equations. *Acad Sci USSR J. Phys.*, **10**, 265–274.

[42] Born, M. and Green, H. S. (1946) A General Kinetic Theory of Liquids. I. The Molecular Distribution Functions. *Proc. R. Soc. A Math. Phys. Eng. Sci.*, **188**, 10–18.

[43] Kirkwood, J. G. (1946) The Statistical Mechanical Theory of Transport Processes I. General Theory. *J. Chem. Phys.*, **14**, 180.

[44] Yvon, J. (1935) *La théorie statistique des fluides et l'équation d'état, Actual. Sci. & Indust. No. 203*. Hermann.

[45] Murrell, D. J. and Law, R. (2000) Beetles in fragmented woodlands: a formal framework for dynamics in ecological landscapes of movement. *J. Anim. Ecol.*, **69**, 471–483.

[46] Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. (2008) *Statistical analysis and modelling of spatial point patterns.*. Wiley.

[47] Treloar, K. K., Simpson, M. J., Binder, B. J., McElwain, D. L. S., and Baker, R. E. (2014) Assessing the role of spatial correlations during collective cell spreading. *Sci. Rep.*, **4**, 5713.

[48] Binder, B. J. and Simpson, M. J. (2015) Spectral analysis of pair-correlation bandwidth: application to cell biology images. *R. Soc. Open Sci.*, **2**, 140494.

[49] Johnston, S. T., Simpson, M. J., McElwain, D. L. S., Binder, B. J., and Ross, J. V. (2014) Interpreting scratch assays using pair density dynamics and approximate Bayesian computation. *Open Biol.*, **4**, 140097.

[50] Murrell, D. J., Dieckmann, U., and Law, R. (2004) On moment closures for population dynamics in continuous space. *J. Theor. Biol.*, **229**, 421–32.

[51] Kirkwood, J. G. (1935) Statistical mechanics of fluid mixtures. *J. Chem. Phys.*, **3**, 300–313.

[52] Bolker, B. and Pacala, S. W. (1997) Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. *Theor. Popul. Biol.*, **52**, 179–97.

[53] Law, R. and Dieckmann, U. (2000) Moment approximations of individual-based models. Dieckmann, U., Law, R., and Metz, J. A. J. (eds.), *The Geometry of Ecological Interactions: Simplifying Spatial Complexity*, chap. 14, pp. 252–270, Cambridge University Press.

[54] Lewis, M. A. (2000) Spread rate for a nonlinear stochastic invasion. *J. Math. Biol.*, **41**, 430–454.

[55] Lewis, M. A. and Pacala, S. (2000) Modeling and analysis of stochastic invasion processes. *J. Math. Biol.*, **41**, 387–429.

[56] Adams, T. P., Holland, E. P., Law, R., Plank, M. J., and Raghib, M. (2013) On the growth of locally interacting plants: differential equations for the dynamics of spatial moments. *Ecology*, **94**, 2732–2743.

[57] Murrell, D. J. (2005) Local spatial structure and predator-prey dynamics: counterintuitive effects of prey enrichment. *Am. Nat.*, **166**, 354–67.

[58] Law, R., Illian, J., Burslem, D. F. R. P., Gratzer, G., Gunatilleke, C. V. S., and Gunatilleke, I. A. U. N. (2009) Ecological information from spatial patterns of plants: insights from point process theory. *J. Ecol.*, **97**, 616–628.

[59] Simpson, M. J. and Baker, R. E. (2011) Corrected mean-field models for spatially dependent advection-diffusion-reaction phenomena. *Phys. Rev. E*, **83**, 051922.

[60] Middleton, A. M., Fleck, C., and Grima, R. (2014) A continuum approximation to an off-lattice individual-cell based model of cell migration and adhesion. *J. Theor. Biol.*, **359**, 220–232.

[61] Le Clainche, C. and Carlier, M. (2008) Regulation of actin assembly associated with protrusion and adhesion in cell migration. *Physiol. Rev.*, **88**, 489–513.

[62] Gillespie, D. T. (1977) Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, **81**, 2340–2361.

[63] Barbaric, I., Biga, V., Gokhale, P. J., Jones, M., Stavish, D., Glen, A., Coca, D., and Andrews, P. W. (2014) Time-lapse analysis of human embryonic stem cells reveals multiple bottlenecks restricting colony formation and their relief upon culture adaptation. *Stem Cell Reports*, **3**, 142–55.

[64] Cai, A. Q., Landman, K. A., and Hughes, B. D. (2007) Multi-scale modeling of a wound-healing cell migration assay. *J. Theor. Biol.*, **245**, 576–94.

[65] Cooper, F., Ghoshal, G., and Pérez-Mercader, J. (2013) Composite bound states and broken U(1) symmetry in the chemical-master-equation derivation of the Gray-Scott model. *Phys. Rev. E*, **88**, 042926.

[66] Blath, J., Etheridge, A., and Meredith, M. (2007) Coexistence in locally regulated competing populations and survival of branching annihilating random walk. *Ann. Appl. Probab.*, **17**, 1474–1507.

[67] Ovaskainen, O., Finkelshtein, D., Kutoviy, O., Cornell, S., Bolker, B., and Kondratiev, Y. (2014) A general mathematical framework for the analysis of spatiotemporal point processes. *Theor. Ecol.*, **7**, 101–113.