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

Rachelle N. Binny\textsuperscript{1,2}, Michael J. Plank\textsuperscript{1,2} and Alex James\textsuperscript{1,2}

\textsuperscript{1}School of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand
\textsuperscript{2}Te Pūnaha Matatini, New Zealand

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.

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 chemo-repellant 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 a 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 minimize their proximity to neighbouring cells. This behaviour can be observed in cell populations cultured \textit{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 insights 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 with 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 [37,40,45,46]. 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 normalized 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.
Figure 1a–c can each be considered as a snapshot in time from a realization of an IBM. Figure 1a 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 (figure 1d). Figure 1b 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 1c. The opposite effect can also arise, whereby cells are less likely to be found close together resulting in a regular pattern (figure 1e). Figure 1f shows that \( C(r) < 1 \) at short distances \( r \) for this type of spatial structure [47].

In previous studies, PCFs have been calculated from experimental images to quantify the extent of spatial structure in live cell populations which adopt Poisson [48], cluster or regular patterns [18,49] to varying degrees. For instance, time-lapse imaging of in vitro cell migration assays, such as circular barrier assays [48] and scratch assays [50], generates data in two spatial dimensions. Image analysis techniques can then be employed to measure the distances between cell pairs and these data used to calculate a PCF. PCFs have also been used alongside experimental data to give insights into the mechanisms responsible for pattern formation [17].

Exploring the dynamics of the second moment can provide insights 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 second moment is lost. However, models that 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 [51,52].

The types of local interactions inherent to migrating cell populations are also of importance in other contexts, such as in animal or plant communities. Many of the modelling tools that employ spatial moment theory were developed for ecological problems [53–56]. 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 [40,53], including the relationship between spatial arrangement and plant size distribution [57]. Models for animal populations undergoing movement have been derived both for the case where movement is dependent on local neighbourhood and the independent case [45,58]. These have also been extended to describe the role of spatial structure in predator–prey relationships [58].

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, whereas the second and third moments can be expressed in terms of displacements between pairs of agents as opposed to agent locations [37,47,59]. 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 [36,60]. 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 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. We assume a setting in which spatial structure is homogeneous, although we 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 weighs 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 also demonstrate that in most cases our spatial moment model provides a good approximation to average results obtained from repeated realizations of the IBM.

2. Individual-based model

We consider the collective movement of \( n \) individuals through a one-dimensional continuous finite domain with periodic
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 $\sigma^2 = 0.2$ for different values $\sigma^2 = 0.2$, and $\sigma^2 = 0.02$. The arrow in (a) marks the location of an individual that is biased to move left. (Online version in colour.)

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 $\sigma^2 = 0.2$ for different values $\sigma^2 = 0.2$, and $\sigma^2 = 0.02$. The arrow in (a) marks the location of an individual that is biased to move left. (Online version in colour.)

boundaries at $x = x_l$ and $x = x_u$. Our model is a continuous time Markov process model in which the state of the system $x(t)$ at time $t$ is

$$x(t) = (x_1(t), x_2(t), \ldots, x_n(t))^T,$$

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

$$x_i \rightarrow x_i + r.$$  

The rate density (i.e. the rate) of this transition is $\psi(x)\mu(x, x_i + r)$, where $\psi$ is the movement rate per unit time of cell $i$ and $\mu(x, x_i + r)$ 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 $\psi_i$ and $\mu(x, x_i + r)$, however, these can be easily adapted to suit different experimental situations.

The movement rate $\psi_i$ 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

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

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

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

where $\alpha$ and $\sigma^2$ 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 $\psi_i$ 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 that 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 $\psi_i$. 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

$$\mu(x, y) = \frac{\lambda}{2} \exp(-\lambda|y - x|),$$

where the mean step length taken by a cell is $1/\lambda$. 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 the 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 $b(z)$ such that the presence of neighbouring cells affects the direction of movement of an individual at $x_i$. Each neighbour contributes $v'(z)$ to the movement of the individual as follows

$$b(z) = \sum_{j=1}^{n} v'(z_j - x),$$

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

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

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 visualize the total neighbour-dependent effect in $b(z)$ 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 2a. 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, whereas 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 \( x \).

Owing 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 \( |y - x| \), whereas for larger distances, the effect is negligible. The strength of interaction is determined by the constant \( \beta \). The variance \( \sigma_v^2 \) 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 \( \sigma_v^2 \). When \( \sigma_v^2 \) is large, \( v(z) \) has a wide spread that will influence outlying cells as shown in figure 2a. On the other hand, for small \( \sigma_v^2 \), \( v(z) \) is a narrow kernel and only neighbouring cells in close proximity to the individual will be affected by its presence (figure 2b).

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

\[
p_r(b) = \frac{1}{1 + \exp(-b)} \tag{2.8}
\]

so that for large \( b(x) > 0 \) a cell at \( x \) is strongly biased to move right, whereas 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).

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

\[
\mu(x, y) = \begin{cases} 
\lambda \exp(-\lambda |y - x|) & \text{if } y - x > 0 \\
\lambda \exp(-\lambda |y - x|) \left(1 - p_r(b(x))\right) & \text{if } y - x < 0,
\end{cases}
\]

with dimensions \( t^{-1} \).

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 \( \beta \) 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 that 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 chemotactant [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 \( \int \mu(x, y) \, dy = 1 \).

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. [47]. 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 notation and method are consistent with the generalized derivation proposed by Plank & 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 \( \delta x, \delta y \) and \( \delta z \) 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 \( O(h^3) \). For ease of visualization these regions are depicted in figure 3a in two-dimensional space; however, the same principles apply in one dimension (figure 3b).

The spatial moments are functions of time as well as space, but for now, we 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 \( \delta x \), centred on \( x \) and of size \( h \), at time \( t \) as follows

\[
Z_1(x) = \lim_{h \to 0} \frac{P(U(x) = 1)}{h} \tag{3.1}
\]
\( I(x) \) is an indicator variable such that \( I(x) = 1 \) if there is a cell in \( \delta x \) centred on \( x \) and \( I(x) = 0 \) if there is no cell in \( \delta x \).

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 \( \delta x \) and \( \delta y \) as follows

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

(3.2)

For simplicity, we assume that \( \delta x \) and \( \delta y \) cannot overlap and, so (3.2) 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 & Law [37].

The second spatial moment has a twofold symmetry such that

\[
Z_2(x, y) = Z_2(y, x).
\]

(3.3)

The third spatial moment, the density of triplets in the small regions \( \delta x \), \( \delta y \), and \( \delta z \), is defined as

\[
Z_3(x, y, z) = \lim_{h \to 0} \frac{P(I(x) = 1 \& I(y) = 1 \& I(z) = 1)}{h^3}.
\]

(3.4)

excluding the cases where \( x = y \) and \( x = z \) as we assume that \( \delta x \), \( \delta y \), and \( \delta z \) cannot overlap. Again, a more detailed description which allows for such non-distinct triplets is given by Plank & Law [37]. The third moment has been shown to have a sixfold symmetry [51]. Similarly, we can define the \( n \)th spatial moment \( Z_n \) as the expected number of \( n \)-tuples of cells per unit (length)\(^n\), for a \( D \)-dimensional space.

It is also useful to define some conditional probabilities. The probability of a cell being found in \( \delta y \) conditional on the presence of a cell in \( \delta x \) is \( P(I(y) = 1 | I(x) = 1) \). We can use the fact that \( P(A|B) = P(AB)/P(B) \) along with equations (3.1)–(3.3) to rewrite this conditional probability as follows

\[
P(I(y) = 1 | I(x) = 1) = \frac{Z_2(x, y) h}{Z_2(x)} + O(h^2).
\]

(3.4)

Similarly, we can write the probability of a cell being found in \( \delta x \) conditional on the presence of a cell in \( \delta x \) and a cell in \( \delta y \) as

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

(3.5)

### 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 solve only 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 expressions for movement rate \( \phi \) 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 equation (3.4), the expected movement rate (from hereafter simply referred to as movement rate) for a single cell at \( x \) is

\[
M_1(x) = m + \frac{\partial}{\partial x} \int Z_2(x, y) \, dy.
\]

(3.6)

The maximum formula that ensured a non-negative movement rate in (2.3) 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

\[
\mu_1(x, y) = \left\{ \begin{array}{ll} \lambda_x \exp(-\lambda_x |y-x|) p_b(b_x(y)) & \text{if } y - x > 0 \\ \lambda_x \exp(-\lambda_x |y-x|) (1 - p_b(b_x(y))) & \text{if } y - x < 0. \end{array} \right.
\]

(3.7)

The neighbour-dependent bias term \( \mu_1(x, y) \) 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 \( v'(y - x) \)

\[
b_1(x) = \int v'(y - x) Z_2(x, y) \, dy.
\]

(3.8)

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

\[
P(\text{cell in } \delta x \text{ at } t + \delta t) = P(\text{cell in } \delta x \text{ at } t) P(\text{cell waited in } \delta x \text{ in } [t, t + \delta t]) + P(\text{cell absent in } \delta x \text{ at } t) P(\text{cell moved into } \delta x \text{ in } [t, t + \delta t]).
\]

(3.9)

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

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

(3.10)

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

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

(3.11)

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 (3.1), (3.10) and (3.11) to write (3.9) as

\[
\frac{dZ_1(x, t)}{dt} = -M_1(x) Z_1(x, t) + \int \mu_1(u, x) M_1(u) Z_1(u, t) \, du.
\]

(3.12)
Equation (3.12) depends on the second spatial moment, incorporated in the movement rate term \( M_2(x, y) \). The first term in (3.12) describes movement out of \( x \), whereas 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 equation (3.5) to write \( M_2(x, y) \) as follows

\[
M_2(x, y) = m + \int \frac{Z_0(x, y, z)}{Z_2(x, y)} dz + w(x - y).
\]  

(3.13)

The third term here accounts for the direct effect of the cell at \( y \) on the cell at \( x \). Because the regions \( \delta y \) and \( \delta z \) do not overlap, the third moment does not account for the case where \( y = z \), 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 \( y \) 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 \( z \)

\[
\mu_2(x, y, z) = \begin{cases} \lambda \exp(-\lambda |y - x|) p_1(b_2(z, y)) & \text{if } y - x > 0 \\ \lambda \exp(-\lambda |y - x|)/(1 - p_1(b_2(z, y))) & \text{if } y - x < 0. \end{cases}
\]

(3.14)

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

\[
b_2(x, y) = \int v'(z - x) \frac{Z_0(x, y, z)}{Z_2(x, y)} dz + v'(y - x).
\]

(3.15)

As in equation (3.13), 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 \( Z_2(x, y) \), we consider the probability of finding a cell in \( \delta y \) at time \( t \) and a cell in \( \delta x \) at time \( t + \delta t \)

\[
P \left( \begin{array}{c}
\text{cell in } \delta x \text{ & cell in } \delta y \\
\text{at } t \text{ & at } t + \delta t
\end{array} \right)
= \int_{\delta y} \int_{\delta x} \left[ P \left( \begin{array}{c}
\text{cell in } \delta x \text{ & cell in } \delta y \\
\text{at } t
\end{array} \right) P \left( \begin{array}{c}
\text{cell in } \delta y \text{ & cell moved into } \delta x \\
\text{in } [t, t + \delta t]
\end{array} \right)
\right]
\]

(3.16)

The probability of cells being present in both \( \delta x \) and \( \delta y \) can be written in terms of \( Z_2(x, y) \) from equation (3.2)

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

(3.17)

Using (3.1) and (3.2), the probability of a cell being present in \( \delta y \) and absent from \( \delta x \) is

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

(3.18)

The probability of both cells waiting in \([t, t + \delta t]\) is written in terms of (3.13)

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

(3.19)

This is comparable to equation (3.10) for the first moment dynamics.

The probability of a cell in \( \delta y \) waiting and a cell moving into \( \delta x \) in \([t, t + \delta t]\) is equivalent to the conditional probability that a cell arrives in \( \delta y \) given that there is a cell in \( \delta y \). As in (3.11), we integrate over all possible starting locations \( u \) for the cell arriving in \( \delta x \). However, the probability of a cell being located at \( u \) is conditional on the presence of a cell at \( y \). Therefore, we have

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

(3.20)

Finally, the probability that a cell moved into \( \delta x \) and a cell moved into \( \delta y \) is \( O(\delta t^2) \), because this would involve two Poisson events occurring during \([t, t + \delta t]\). Similarly, the higher-order terms in equations (3.19) and (3.20) arise from probabilities involving more than one cell undergoing a movement event during a time \( \delta t \).

We substitute equations (3.17)–(3.20) into (3.16) and make use of the twofold symmetry of \( Z_2(x, y, t) \). Using a Taylor expansion of \( Z_2(x, y, t + \delta t) \), expanding terms, then letting \( h, \delta t \rightarrow 0 \) which removes the higher-order terms, gives

\[
\frac{dZ_2(x, y, t)}{dt} = - (M_2(x, y) + M_2(x, y)) Z_2(x, y, t)
+ \int \mu_2(u, x, y) M_2(u, y) Z_2(u, y, t) du
+ \int \mu_2(u, y, x) M_2(u, x) Z_2(u, x, t) du.
\]

(3.21)

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 drop the \( t \) from the spatial moment notation.

Equation (3.21) 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

\[
\hat{Z}_3(x, y, z) = \frac{Z_2(x, y) Z_2(x, z) Z_2(y, z)}{Z_1(x) Z_1(y) Z_1(z)},
\]

(3.22)

however, other choices of closure are also possible [51]. For a Poisson spatial pattern, the third moment is \( Z_3(x, y, z) = Z_1^3 \), and the approximation in (3.22) 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 equation (2.3)), because we are solving for a spatially homogeneous case. To obtain spatial information from the IBM, we calculate the spatial moment model (see appendix) by averaging the results of repeated realizations. The PCF predicted by the spatial moment model is given by $C_{SM}(j) = Z_j(j)/Z_j^2$ such that $C_{SM}(j) = 1$ in the absence of spatial structure. The second moment is isotropic (i.e. it has symmetry about the origin) and therefore we show only $C_{SM}(j)$ for $j \geq 0$.

In each realization of the IBM, we distribute the cells at $t = 0$ according to a spatial Poisson process on $[x, y]$ with intensity $n/L$. Therefore, initially, there is no spatial structure present. The corresponding initial condition for the spatial moment model is to set $Z_0(j) = Z_j^2$ 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 $Z_j(0)$ 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 the results for different values of $\alpha$ 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. Owing to the stochastic nature of the IBM, it is possible that negative motility rates may occur by chance; however, the definition of $\psi_i$ given in equation (2.3) ensures that $\psi_i = 0$ for such rare chance events.

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

$C_{SM}(j)$ provides a good approximation to $C_{IBM}(j)$ except for $\alpha < 0$ when $|\alpha|$ is large. For $\alpha = -2$ and $-2.5$, $C_{SM}(j)$ has converged to a steady state by $t = 25$ but $C_{IBM}(j)$ continues to increase over time at short displacements. The discrepancy between $C_{SM}(j)$ and $C_{IBM}(j)$ 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 $\alpha$, 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 \( C_{\text{SM}}(\xi) \) matched \( C_{\text{IBM}}(\xi) \) 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 \( C_{\text{IBM}}(\xi) \), figure 4f shows that even a relatively low incidence of 1% can lead to a mismatch between \( C_{\text{IBM}}(\xi) \) and \( C_{\text{SM}}(\xi) \). Further increasing the magnitude of \( \alpha < 0 \) causes a significant increase in the incidences of \( \psi_s < 0 \), and the fit between \( C_{\text{SM}}(\xi) \) and \( C_{\text{IBM}}(\xi) \) deteriorates to an even greater extent.

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 \( \beta > 0 \). Figure 5a–c shows results for three different values of \( \beta > 0 \). \( C(\xi) \) decreases at small displacements with increasing interaction strength \( \beta \). In figure 5c, for \( \beta = 10 \), there is a peak in both the \( C_{\text{IBM}}(\xi) \) and \( C_{\text{SM}}(\xi) \) 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 figure 5d–f. The magnitude of \( \beta \) required to generate clustering is less than that needed to form a regular spatial pattern.

In figure 5, \( C_{\text{SM}}(\xi) \) provides a good approximation to \( C_{\text{IBM}}(\xi) \). However, greater magnitudes of \( \beta < 0 \) lead to disparities between \( C_{\text{SM}}(\xi) \) and \( C_{\text{IBM}}(\xi) \). For example when \( \beta = -1.0 \), \( C_{\text{IBM}}(0) \approx 3.6 \) at \( t = 25 \) while \( C_{\text{SM}}(0) \approx 11.5 \) and neither PCF has reached steady state. Over time, the cluster pattern becomes stronger, and the disparity between the two approximations deteriorates, because \( C_{\text{SM}}(0) \) is increasing at a faster rate than \( C_{\text{IBM}}(0) \).

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 \( \alpha \) and \( \beta \). We choose values of \( \alpha \) and \( \beta \) 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 §§4.1 and 4.2. Figure 6a,b shows 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 are working in opposition (figure 6b–c), they counteract one another and very little spatial structure develops over time as indicated by $C_{\text{SM}}(\xi) \approx 1$. $C_{\text{SM}}(\xi)$ is a good approximation to $C_{\text{IBM}}(\xi)$ except in figure 6d. In this case, the slight mismatch 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 (approx. 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 owing 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 $\xi$ indicating an absence of spatial structure. The effect of changing the width of the interaction kernels has also been explored. If we interpret $2\sigma$ (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 $\sigma^2$ 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 realizations 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 insights 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 that 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 with 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 and 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 the 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 cannot 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 owing 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. 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 owing 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.

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 strengths 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.

Middleton et al. [46] studied a model incorporating a local directional bias and employing the second spatial moment. Their IBM is based on a Langevin equation defining each cell’s velocity as the sum of the interaction forces from neighbours plus a noise term. Thus, neighbour interactions always affect both speed and direction of movement, whereas, in our model, average speed and direction are treated as independent effects. Despite these differences, we see similar qualitative patterns of spatial structure owing to local repulsive or attractive forces as in Middleton et al. [46].

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 [51]. 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. Stochastic differential equations, such as Langevin-type equations, capture fluctuations arising owing 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 & Chapman [35] employed a perturbation method to describe the dynamics of moving particles by using matched asymptotic expansions in a small parameter $\varepsilon \ll 1$ 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.

Funding statement. This research was supported by the Royal Society of New Zealand Marsden Fund, grant no. 11-UOC-005.
Appendix A

A.1. Calculating a pair correlation function

The PCF $C(r)$ provides a means of extracting information about spatial structure from the configurations of agents that arise in realizations 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

$$
r = \begin{cases}
|\xi| & \text{if } |\xi| < \frac{L}{2} \\
L - |\xi| & \text{if } |\xi| > \frac{L}{2}.
\end{cases} \quad (A.1)
$$

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 $[r, r + \delta r]$. Normalizing by $2\delta r n^2 / L$ ensures $C(r) = 1$ in the absence of spatial structure.

A.2. Numerical methods

To solve equation (3.21) 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 on 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 $\xi$ and the displacement from $x$ to $z$ is denoted $\xi'$. For the movement PDF $\mu_2(u, x, y)$, we denote the displacement from $u$ to $x$ as $\xi''$. The first spatial moment is required for $Z_3(x, y, z)$ and in the homogeneous case $Z_1$ is a constant.

We rewrite (3.21) in terms of the displacements between pairs as follows

$$
\frac{dZ_2(\xi)}{d\xi} = -(M_2(\xi) + M_2(-\xi))Z_2(\xi) + \int_{\mathbb{R}} \mu_2(\xi'', \xi'' + \xi)M_2(\xi'' + \xi)Z_2(\xi'' + \xi) d\xi'' + \int_{\mathbb{R}} \mu_2(\xi'', \xi'' - \xi)M_2(\xi'' - \xi)Z_2(\xi'' - \xi) d\xi''. \quad (A.2)
$$

The movement rate $M_2(x, y)$ of a cell at $x$ in a pair with a cell at $y$ given in (3.13) is now expressed in terms of the displacement $\xi$ between $x$ and $y$

$$
M_2(\xi) = m + \int w(\xi') \frac{Z_3(\xi', \xi')}{Z_3(\xi)} d\xi' + w(\xi). \quad (A.3)
$$

The movement PDF given in (3.14) becomes

$$
\mu_2(\xi, \xi') = \left\{ \begin{array}{ll} 
\lambda_x \exp(-\lambda_x|\xi|) \rho_1(b_2(\xi')) & \text{if } \xi > 0 \\
\lambda_x \exp(-\lambda_x|\xi|) (1 - \rho_1(b_2(\xi')))) & \text{if } \xi < 0,
\end{array} \right. \quad (A.4)
$$

with neighbour-dependent bias

$$
b_2(\xi) = \int \varphi(\xi') \frac{Z_3(\xi, \xi')}{Z_3(\xi)} d\xi' + \varphi(\xi). \quad (A.5)
$$

The interaction kernels were previously expressed in terms of a single variable in (2.4), and (2.7) and these definitions still hold here. The closure for the third moment is

$$
\tilde{Z}_3(\xi, \xi') = \frac{Z_3(\xi)Z_3(\xi')Z_3(\xi' - \xi)}{Z_1^3}. \quad (A.6)
$$

The boundary condition is as follows

$$
Z_2(\xi) \rightarrow Z_1^2 \text{ as } \xi \rightarrow \infty. \quad (A.7)
$$

Equation (A 2) was solved using the method of lines with MATLAB’s in-built ode45 solver. This involved a discretization of $\xi$ with grid spacing $\Delta = 0.1$ over the domain $|\xi| \leq \xi_{\text{max}}$ where $\xi_{\text{max}}$ was large enough so that $Z_2(\xi) \approx Z_1^2$ at $|\xi| = \xi_{\text{max}}$. 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 (A 2) were approximated using the trapezoidal rule with the same discretization. In addition, the PDF for movement $\mu_2(\xi, \xi')$ was normalized numerically using the trapezoidal rule such that $\int \mu_2(\xi, \xi') d\xi = 1$. The results were insensitive to a reduction in grid spacing $\Delta$.

References


