A continuum approximation to an off-lattice individual-cell based model of cell migration and adhesion

Alistair M. Middleton\textsuperscript{1,2}, Christian Fleck\textsuperscript{3} and Ramon Grima\textsuperscript{4,†}

1 University of Heidelberg, Im Neuenheimer Feld 267, 69120 Heidelberg, Germany
2 Center for Biological Systems Analysis, University of Freiburg, Habsburgerstr. 49, 79104 Freiburg, Germany
3 Laboratory for Systems and Synthetic Biology, Dreijenplein 10, 6703HB Wageningen, The Netherlands
4 SynthSys and School of Biological Sciences, University of Edinburgh, EH9 3JR Edinburgh, UK

Abstract

Cell-cell adhesion plays a key role in the collective migration of cells and in determining correlations in the relative cell positions and velocities. Recently, it was demonstrated that off-lattice individual cell based models (IBMs) can accurately capture the correlations observed experimentally in a migrating cell population. However, IBMs are often computationally expensive and difficult to analyse mathematically. Traditional continuum-based models, in contrast, are amenable to mathematical analysis and are computationally less demanding, but typically correspond to a mean-field approximation of cell migration and so ignore cell-cell correlations. In this work, we address this problem by using an off-lattice IBM to derive a continuum approximation which does take into account correlations. We furthermore show that a mean-field approximation of the off-lattice IBM leads to a single partial integro-differential equation of the same form as proposed by Sherratt and co-workers to model cell adhesion. The latter is found to be only effective at approximating the ensemble averaged cell number density when mechanical interactions between cells are weak. In contrast, the predictions of our novel continuum model for the time-evolution of the ensemble cell number density distribution and of the density-density correlation function are in close agreement with those obtained from the IBM for a wide range of mechanical interaction strengths. In particular, we observe ‘front-like’ propagation of cells in simulations using both our IBM and our continuum model, but not in the continuum model simulations obtained using the mean-field approximation.

† Corresponding author:
E-mail address: ramon.grima@ed.ac.uk
Address: C. H. Waddington Building, University of Edinburgh, Mayfield Road EH9 3JD UK
Phone number: (+44) 0131 651 9060
1 Introduction

Mechanical interactions between cells underpin a considerable range of biological processes during tissue development and maintenance. Two pertinent examples are the collective migration of cohesive groups of cells, which occur during wound healing and cancer invasion (see for example [19, 20, 61]), and cell sorting, which may drive certain aspects of self-organisation and tissue pattern during embryo development (see [2] for a review).

A number of mathematical models of cell-cell and cell-ECM adhesion have been developed. Broadly speaking, these can be classified as either continuum-based models or individual cell-(or agent-) based models (IBMs), although some incorporate features from both (and are typically referred to as hybrid models, for example see [32] and [52] for an overview). Traditionally, continuum-based models are derived from a deterministic, top-down perspective, wherein the dynamics of a large population of cells is described in terms of locally averaged properties, such as the spatial distribution of the cell number density (which we shall refer to as the cell density distribution) [3, 35, 41, 66]. To derive these models, various conservation laws, in particular conservation of mass, are exploited (see for example [3, 46]). This means that one must prescribe certain constitutive equations, such as how the flux of cells depends on the system variables [46]. These, in turn, can be obtained either by analogy with well-established physical laws [46], or intuition [3]. For this reason, the connection between the resulting macroscopic continuum-based description and the microscopic cell-scale processes is not always clear. Furthermore such continuum descriptions are typically based on the implicit assumption that statistical correlations between cells are negligible, i.e. a mean-field assumption [59]. However, the behaviour of a cell is often determined by interactions between it and other nearby cells (via signals or mechanical forces), which in turn leads to strong cell-cell correlations. This was observed, for example, during collective cell migration [75], where correlations in the cell velocities are driven in large part by mechanical interactions [65].

In contrast to continuum-based models, IBMs treat cells as discrete entities. At the coarsest scale, each variable in the IBM corresponds to the centre of mass position of an individual cell; for finer-scale descriptions, variables can instead correspond to sub-elements of a cell. The positions of cells or cell elements can either be restricted to specific lattice sites (these being lattice-based models; for example cellular automata [49]), or be off-lattice (i.e. so that cells or cell elements occupy positions in continuous space). In lattice-based models for which a cell only spans a single site, cell movement is captured by prescribing certain transition probabilities (this being the probability of a cell jumping from one site to another). These transition probabilities can be made to depend on whether a particular site is already occupied by another cell. In this way, other biologically relevant effects can be incorporated into the model: volume exclusion is captured by assuming that each lattice site is occupied by at most one cell [36, 68, 69]; cell adhesion by assuming that the probability of a cell vacating a site is lower if another adherent cell occupies a neighbouring site [14, 36, 69]). The coarseness of the lattice means that any two cells can either occupy neighbouring lattice sites (and therefore adhere to one another), or be one or more cell diameters apart. Hence, this approach can only provide a phenomenological description of mechanical interactions between cells. Furthermore, the coarseness of the lattice spacing prevents it from realistically capturing local correlations in the relative positions and velocities of neighbouring cells (although the model may still be an improvement over a mean-field description). To overcome these issues, the model can be altered so that a cell spans multiple lattice points (such as in the Cellular Potts Model or CPM). Alterations in cell size, shape or location (as a consequence of these interaction forces) are...
reflected by changes in the system’s potential energy. The model evolves by moving from one state to another, energetically more favourable state. This approach has been successfully applied to a number of different problems associated with cell-cell adhesion and movement [76, 77], including cell-sorting [25, 27, 45]. Of course, increasing the spatial resolution of the lattice increases the computational cost of the model, although artefacts can still arise as a consequence of imposing a lattice [15, 30].

An alternative to lattice-based models are off-lattice ones. Again, model variables can either correspond to the positions of individual cells (treated as point masses) [17, 51, 55, 64], or subcellular cell elements (as used in the subcellular element model or SEM) [50, 62]. In this case, positions of cells or cell-elements are governed by either stochastic (where stochasticity is adopted to capture random cell movement) or ordinary differential equations. One advantage of an off-lattice model is that, even for the coarsest description where model variables correspond to cell positions, this approach allows one to more realistically capture mechanical interactions between cells than a lattice-based model. In particular, it was recently demonstrated by [64] that a coarse off-lattice model (wherein cells are treated as point masses) could accurately capture the cell-velocity and cell-cell correlations that arise during collective migration of Madin-Darby Canine Kidney (MDCK) cells.

Being able to provide a direct mapping between discrete and continuum based models has a number of clear advantages. In the case where IBMs are stochastic, the model will require many realisations in order to analyse its properties. Moreover, the computational cost of solving an IBM typically scales quadratically with the number of cells or sub-cellular elements being considered, whereas for continuum based models, the problem scales with the required resolution necessary to capture variations of interest. The additional computational cost associated with numerically analysing IBMs is particularly acute when one wants to study effects from varying different parameters in the model (which may be the case, for example, if one wants to fit the model to the available data). Finally, the study of IBMs and continuum-based approximations can help clarify how traditional continuum-based models (derived using a top-down approach) and IBMs relate to one another [48].

Continuum approximations have been derived for coarse-grained lattice-based models of cell adhesion and motility [34, 67]. Importantly, the authors found that the so-called mean field approximation (MFA) is not adequate to describe the behaviour of the IBM when adhesion effects are strong. Instead, to describe the correlations, the authors adopted the so-called Kirkwood Superposition Approximation (KSA). However, coarseness of the lattice meant that cells only span a single lattice point, and so it is unlikely that such models will accurately capture correlations in relative cell neighbour positions. Continuum approximations to CPM-based models (where cells instead span multiple grid points) have been derived [1, 40]. These consider cells that physically interact (via volume exclusion and interaction with the extracellular matrix) and move along a chemical gradient (chemotaxis). Mean-field based approximations to the CPM take into account cell-cell adhesion [1]; however the latter is neglected by current continuum models that go beyond the mean-field approximation [40].

In general, an off-lattice approach offers a better description of the underlying correlations [64]. Previous attempts to relate the behaviour of these models to partial differential equation (PDE) based models include fitting the continuum-based model (and hence estimating parameters) to averaged simulations obtained from an IBM [80]. Of course, this is numerically rather cumbersome; being able to obtain an accurate continuum approximation to an off-lattice model would allow (for example) a rapid exploration of the model’s parameter space. Other attempts to derive continuum
approximations from discrete off-lattice models can be found in the works of Fozard et al. [18] and Murray et al. [47, 48], however these ignore stochasticity.

In this work, we address the above considerations by exploring continuum approximations to an off-lattice IBM that is based on Langevin equations, building on the work of Newman and Grima [51]. First, in Section 2.1, we introduce our IBM of cell migration and adhesion. In Section 2.2, we then derive equations governing the associated probability distribution functions. Namely, these form (for a system of \( N > 2 \) cells) a hierarchy of \( N \) partial integro-differential equations. As is typically the case, it is not practical to solve this hierarchy directly. Rather, in Section 2.3 we consider two different moment closures to truncate the hierarchy, these being the MFA and the KSA. It is worth noting that using the MFA, we arrive at a single partial integro-differential equation which is of the same form as the equations proposed by Sherratt and co-workers [3, 4, 8, 54]. In Section 3 we then obtain numerical solutions to the two alternative moment closures and compare these to ensemble averaged quantities calculated from IBM simulations. We finish by a discussion of our results in Section 4.

2 The model

2.1 The individual-cell based model

In this section we introduce our IBM of cell movement and adhesion, following [51]. The model is intended to be complex enough to capture the motion of cells that interact through physical forces such as cell-cell adhesion, but simple enough so that it is amenable to mathematical analysis. We therefore assume that cells are described as point masses, the motion of which can be described using Newton’s second law of motion, namely

\[
m_i \frac{d^2 x_i}{dt^2} = F_{i}^{\text{visc}} + \sum_{k, k \neq i} F_{i,k}^{\text{int}} + \xi_i, \tag{2.1}
\]

where \( m_i \) is the mass of cell \( i \), \( x_i \) is its position, \( F_{i,j}^{\text{int}} \) is the force generated by interactions between cell \( i \) and cell \( j \), \( F_{i}^{\text{visc}} \) is the viscous force acting on cell \( i \), and \( \xi_i \) is a stochastic force which reflects self-propulsion of the cell. For simplicity \( \xi_i \) is sampled from a Gaussian distribution with zero mean and zero auto-correlation:

\[
\langle \xi_i(t) \xi_j(t') \rangle = 2D \delta_{i,j} \delta(t - t'). \tag{2.2}
\]

The constant of proportionality \( D \) in (2.2) corresponds to the (macroscopic) cell diffusion coefficient [21, 22]. Thus, when isolated, cells will move randomly [51], but when in physical contact, the cells’ motion is influenced by mechanical interactions with their neighbours. The viscous force acting on cell \( i \), which is the result of drag between the cell and its substrate, can be assumed to be proportional to the cell’s velocity (with constant of proportionality \( \mu \)). The low Reynolds number typical at the cellular length scale implies that inertial forces, i.e. the first term on the LHS of (2.1) can be neglected [5]. We further assume that the interaction force \( F_{i,k}^{\text{int}} \) is given by

\[
F_{i,k}^{\text{int}} = \hat{F}_0 F(\parallel x_i - x_k \parallel) \frac{x_i - x_k}{\parallel x_i - x_k \parallel}, \tag{2.3}
\]

where \( F \) is a (dimensionless) scalar function of the separation between pairs of cells, and \( \hat{F}_0 \) is a (dimensional) constant of proportionality (and captures the amplitude of the interaction forces).
This term can be used to capture various mechanical forces, such as cell-cell adhesion or an individual cell’s resistance to deformation (see below for details). Moreover, we are interested in the case where the interaction force is of limited range, so it is possible for cells to ‘detach’ from their neighbours and migrate individually. Based on these considerations, it follows from (2.1) that cell movement can be described by

$$\frac{dx_i}{dt} = F_0 \sum_{k, k \neq i} F(\|x_i - x_k\|) \frac{x_i - x_k}{\|x_i - x_k\|} + \xi_i,$$

(2.4)

where $F_0 = \hat{F}_0/\mu$. Boundary conditions are chosen to be periodic, so that cell movement occurs on an $n$-dimensional torus $x \in [0, L]^n$. Later we will be restricting our attention to the one-dimensional case.

Models of the form (2.4) have been applied to the study of collective cell migration (see, for example, [17, 65, 73, 75]). In particular, Szabo et al. [73] developed a model of cell migration which contains an additional term representing directed (as opposed to purely stochastic) self-propulsion. Such models are related to those appearing in the context of ‘active matter’, wherein particles are subjected to a self-propelling velocity vector which is typically adjusted according to the average velocity of other nearby particles [12, 13, 78, 79].

A number of force laws to model cell-cell interactions have appeared in the literature. The simplest scenario is to assume that both the repulsive and attractive forces can be modelled using linear springs [43, 47]. However, this may not accurately reflect (for example) the resistance to deformation that cells exert. Instead, nonlinear force laws based on phenomenological principles are often adopted in off-lattice models. These can capture the possibility that repulsive forces increase rapidly as cells become compressed. In the extreme case when the repulsive force becomes infinitely strong below a certain critical separation (the cell radius), the model is commonly referred to as a ‘hard-core’ model (see [56]). Examples of nonlinear force laws common in the literature include the Lennard-Jones [32, 72] and the Morse potential [63]. For simplicity, we adopt the latter, given by

$$F(r) = \begin{cases} 
2(\exp(-2a(r - r_c)) - \exp(-a(r - r_c))), & r < \sigma r_c, \\
0, & r \geq \sigma r_c.
\end{cases}$$

(2.5)

A representative plot of $F(r)$ for various values of $a$ is given in Figure 1. Note that $F(r)$ equals zero at $r = r_c$ and as $r \to \infty$. For $r < r_c$, the two cells will exert a repulsive force on one another (corresponding to the cells’ resistance to deformation). For $r > r_c$, cells exert an attractive force on one another (corresponding to cell-cell adhesion). Thus, $r_c$ is the natural radius of a cell. The parameter $a$ controls how rapidly the magnitude of $F(r)$ grows as $r \to 0$, in other words how ‘soft’ or ‘hard’ the cells are. The distance over which cells can mechanically interact is of finite extent (given by $\sigma r_c$; $\sigma$ is the number of natural cell radii over which cells can mechanically interact).

‘Long range’ interactions could occur via protrusions such as filopodia. We note that our implicit assumption that forces between cells are a function of their separation neglects the fact that an intermediary cell may reduce or block this interaction. However filopodia typically have a length of the order of few tens of microns, which is similar to the radius of a eukaryotic cell [42]. For example, the length of filopodia in migrating mouse embryo cells are typically 1-2 times the radius of the associated cell [71]. We therefore consider filopodial interaction between two non-neighbouring cells to be a rare event and we neglect them accordingly. To ensure the biologically plausibility of the long-range interactions we only consider the range $1 < \sigma \leq 3$ (noting that for $\sigma \leq 1$, (2.5) generates repulsive forces only).
We non-dimensionalise the IBM (2.4)-(2.5) as follows. Dimensionless space is based on the characteristic length of the domain ($L$) and dimensionless time on the characteristic timescale for cell movement, namely $L^2/D$. Thus

$$x^* = \frac{x}{L}, \quad t^* = \frac{t}{L^2/D},$$

(2.6)

where the star superscripts are used to indicate that a parameter is dimensionless. Upon substituting (2.6) into (2.4) we obtain

$$\frac{dx_i^*}{dt^*} = F_0^* \sum_{k, k \neq i} F(\parallel x_i^* - x_k^* \parallel) \frac{x_i^* - x_k^*}{\parallel x_i^* - x_k^* \parallel} + \xi_i^*,$$

(2.7)

where $\langle \xi_i^*(t) \xi_j^*(t') \rangle = 2\delta_{i,j} \delta(t^* - t'^*)$ and

$$F(r^*) = \begin{cases} 2 \exp(-2a^*(r^* - \epsilon)/\epsilon) - \exp(-a^*(r^* - \epsilon)/\epsilon) / \epsilon, & r^* < \sigma \epsilon, \\ 0, & r^* \geq \sigma \epsilon, \end{cases}$$

(2.8)

where $F_0^* = F_0 L/D$, $\epsilon = r_c L$, $a^* = ar_c$. (2.9)

In practice, $\epsilon$ is small. We henceforth drop the star superscripts and work exclusively with the dimensionless model.

2.2 Deriving the one and two-cell probability density functions

The behaviour of our (dimensionless) IBM (2.7)-(2.8) can be described in terms of probability density functions (PDFs). We define the one-cell PDF, this being the probability $P_i(x, t)$ that cell $i$ has position $x$ at time $t$, as

$$P_i(x, t) = \langle \delta(x - x_i(t)) \rangle,$$

(2.10)

where the angled brackets indicate averaging over the noise. Similarly, we can define the two-cell PDF (i.e. that cell $i$ is at position $x$ and cell $j$ is at position $x'$ at time $t$) by

$$P_{i,j}(x, x', t) = \langle \delta(x - x_i(t)) \delta(x' - x_j(t)) \rangle,$$

(2.11)

and so on for the $n$-cell PDF (where $n \in \{1, \ldots, N\}$ and $N$ is the number of cells). Following Newman and Grima [51], we derive equations governing the one-cell and two-cells PDFs. For the one-cell PDF, we differentiate both sides of (2.10) with respect to time and obtain

$$\frac{dP_i(x, t)}{dt} = -\nabla \cdot \left( \frac{dx_i}{dt} \delta(x - x_i) \right),$$

(2.12)

and subsequently substitute (2.7) into (2.12) to obtain

$$\frac{dP_i(x, t)}{dt} = -\nabla \cdot \langle \xi_i \delta(x - x_i) \rangle - F_0 \nabla \cdot \left( \sum_{k, k \neq i} F(\parallel x_i - x_k \parallel) \frac{x_i - x_k}{\parallel x_i - x_k \parallel} \delta(x - x_i) \right).$$

(2.13)
Figure 1: (a) Schematic representation of the key processes modelled by the IBM (2.4). Cells have a natural radius $r_c$; for $r < r_c$, cells feel repulsion which reflects the cell’s resistance to deformation, and for $r > r_c$ cells feel an attractive force which models cell-cell adhesion. The attractive forces are mediated by protrusions such as filopodia and are of finite extent (given by $\sigma r_c$). The $i$th cell position is given by $x_i$. The forces are described by an interaction function (2.5) which is shown in (b) for various values of $a$. Here, the natural cell radius $r_c = 0.1$ and $\sigma = 2$. Increasing $a$ causes the repulsive force to increase more rapidly as $r \to 0$, so that cells are stiffer.
Thus, we obtain a local advection term in the governing equation for $P_i$. The second term can be written in terms of a convolution between $F$ and a $\delta$-function centred at $x_k$. Thus:

$$
\frac{\partial P_i(x, t)}{\partial t} = \nabla^2 P_i - F_0 \nabla \left( \sum_{k, k \neq i} \int F ||x_i - x'|| \frac{x_i - x'}{||x_i - x'||} \delta(x - x_i) \delta(x - x_k) dx' \right),
$$

$$
= \nabla^2 P_i - F_0 \nabla \left( \int F ||x - x'|| \frac{x - x'}{||x - x'||} \sum_{k, k \neq i} P_{i,k}(x, x', t) dx' \right),
$$

(2.14)

where we have used (2.11). From (2.14) we observe that the one-cell PDF depends on the two-cell PDF (2.11). To derive the governing equation for the two-cell PDF, we use a similar approach and differentiate (2.11) with respect to $t$ to obtain

$$
\frac{\partial P_{i,j}(x, x', t)}{\partial t} = - \frac{\partial}{\partial x} \left( \frac{dx_i}{dt} \delta(x - x_i) \delta(x' - x_j) \right) - \frac{\partial}{\partial x'} \left( \frac{dx_j}{dt} \delta(x - x_i) \delta(x' - x_j) \right).
$$

(2.15)

Upon substituting (2.7) into (2.15), we obtain

$$
\frac{\partial P_{i,j}(x, x', t)}{\partial t} = - \frac{\partial}{\partial x} \left( \langle \xi(x - x_i) \delta(x' - x_j) \rangle \right) - \frac{\partial}{\partial x'} \left( \langle \xi(x - x_i) \delta(x' - x_j) \rangle \right)
$$

$$
- F_0 \frac{\partial}{\partial x} \left( \sum_{k, k \neq i} F(||x_i - x_k||) \frac{x_i - x_k}{||x_i - x_k||} \delta(x - x_i) \delta(x' - x_j) \right)
$$

$$
- F_0 \frac{\partial}{\partial x'} \left( \sum_{k, k \neq j} F(||x_j - x_k||) \frac{x_j - x_k}{||x_j - x_k||} \delta(x - x_i) \delta(x' - x_j) \right).
$$

(2.16)

Similarly to before, the first two terms (which reflect the random movement of the cells) can be associated with the Laplacian of $P_{i,j}$. The last two terms capture the cell interaction forces. However, unlike before, these terms include expressions for both direct interactions between $i$ and $j$, as well as for interactions between $i$ or $j$ and other cells $k$. Before rewriting these expressions in terms of convolutions, we split the sums into local and non-local parts:

$$
\frac{\partial P_{i,j}(x, x', t)}{\partial t} = \nabla^2 P_{i,j} - F_0 \frac{\partial}{\partial x} \left( \langle \xi(x - x_i) \delta(x' - x_j) \rangle \right)
$$

$$
- F_0 \frac{\partial}{\partial x'} \left( \langle \xi(x - x_i) \delta(x' - x_j) \rangle \right)
$$

$$
- F_0 \frac{\partial}{\partial x} \left( \sum_{k, k \neq i} F(||x_i - x_k||) \frac{x_i - x_k}{||x_i - x_k||} \delta(x - x_i) \delta(x' - x_j) \right)
$$

$$
- F_0 \frac{\partial}{\partial x'} \left( \sum_{k, k \neq j} F(||x_j - x_k||) \frac{x_j - x_k}{||x_j - x_k||} \delta(x - x_i) \delta(x' - x_j) \right).
$$

(2.17)

Thus, we obtain a local advection term in the governing equation for $P_{i,j}$, whereas the last two terms can now (similarly to the one-cell PDF) be written in terms of a convolution between $F$ and a $\delta$-function centred at $x_k$. Thus:

$$
\frac{\partial P_{i,j}(x, x', t)}{\partial t} = \nabla^2 P_{i,j} - F_0 \left( \frac{\partial}{\partial x} \frac{\partial}{\partial x'} \right) \left( F(||x - x'||) \frac{x - x'}{||x - x'||} P_{i,j}(x, x') \right)
$$

$$
- F_0 \frac{\partial}{\partial x} \left( \int F(||x - x''||) \frac{x - x''}{||x - x''||} \sum_{k, k \neq j} P_{i,k}(x, x', x'', t) dx'' \right)
$$

$$
- F_0 \frac{\partial}{\partial x'} \left( \int F(||x' - x''||) \frac{x' - x''}{||x' - x''||} \sum_{k, k \neq j} P_{i,k}(x, x', x'', t) dx'' \right).
$$

(2.18)
where $P_{i,j,k}$ is the three-cell PDF. We can repeat this process to yield a hierarchy of $N - 1$ partial integro-differential equations and one Fokker-Planck equation (this governing the $N$-cell PDF). In particular, solving (2.14) for $P_i$ requires us to solve (2.18) for $P_{i,j}$, which in turn depends on $P_{i,j,k}$ and so on. Since we are interested in the tissue scale dynamics of our IBM, we calculate the following quantities:

\[
p_1(x, t) = \frac{1}{N} \sum_i P_i(x, t), \quad p_2(x, x', t) = \frac{1}{N(N - 1)} \sum_{j \neq i} P_{i,j}(x, x', t), \quad p_3(x, x', x'', t) = \frac{1}{N(N - 1)(N - 2)} \sum_i \sum_{j \neq i} \sum_{k \neq j} P_{i,j,k}(x, x', x'', t),
\]

(2.19)

For a given time $t$, $p_1$ gives the (normalised) cell density distribution and $p_2$ the (normalised) equal time density-density correlation function. Similarly $p_3$ gives the (normalised) correlation of density at three different points in space. The normalisation is such that the integral of $p_i$ over all space equals one. Note also that

\[
p_1(x, t) = \int p_2(x, x', t) dx',
\]

(2.20)
a relationship which we will find numerically convenient later on.

Upon summing over the index $i$ in (2.14), dividing throughout by $N$ and making use of the definitions in (2.19) we obtain an equation for the time-evolution of $p_1(x, t)$. Similarly summing over the indices $i$ and $j$ in (2.18), dividing throughout by $N(N - 1)$ and making use of the definitions in (2.19) we obtain an equation for the time-evolution of $p_2(x, x', t)$. These two time-evolution equations are given by:

\[
\frac{\partial p_1(x, t)}{\partial t} = \nabla^2 p_1(x, t) - F_0(N - 1) \nabla \cdot \int F(\|x - x'\|) \frac{\mathbf{x} - \mathbf{x}'}{\|\mathbf{x} - \mathbf{x}'\|} p_2(x, x', t) dx',
\]

(2.21a)

\[
\frac{\partial p_2(x, x', t)}{\partial t} = \nabla^2 p_2(x, x', t) - F_0 \left( \frac{\partial}{\partial \mathbf{x}} - \frac{\partial}{\partial \mathbf{x}'} \right) \left( \int F(\|x - x'\|) \frac{\mathbf{x} - \mathbf{x}'}{\|\mathbf{x} - \mathbf{x}'\|} p_2(x, x', t) \right) - F_0(N - 2) \frac{\partial}{\partial \mathbf{x}} \int F(\|x - x''\|) \frac{\mathbf{x} - \mathbf{x}''}{\|\mathbf{x} - \mathbf{x}''\|} p_3(x, x', x'', t) dx''
\]

\[
- F_0(N - 2) \frac{\partial}{\partial \mathbf{x}'} \int F(\|x' - x''\|) \frac{\mathbf{x}' - \mathbf{x}''}{\|\mathbf{x}' - \mathbf{x}''\|} p_3(x, x', x'', t) dx'',
\]

(2.21b)

and we recall that boundary conditions are chosen to be periodic.

We note that (2.14) and (2.18) are similar in form to the Bogoliubov-Born-Green-Kirkwood-Yvon (BBGKY) hierarchy [53]. The starting point of the latter is however not a set of Langevin equations; rather it is the Liouville equation for the probability density function of the positions and momenta of a many-particle system in which the motion of the particles is determined by a balance between inertial and interaction forces. In contrast in our case, the motion of cells is determined by a balance between viscous, cell-cell interaction and stochastic forces. Thus the solutions obtained to BBGKY models are rather different to what is presented here. For example, individual particles will typically tend to the same (non-zero) speed and form specific spatial patterns, such as rotating mills or ‘flocks’ (i.e. all particles move in the same direction). For this reason, the BBGKY hierarchy has been extensively applied to problems relating to flocking [11,31].
2.3 Approximations through closure relations

2.3.1 Mean-field approximation (MFA)

The hierarchy of $N$ equations (2.21) are prohibitively difficult to solve numerically for most practical problems. Instead, one can adopt a particular closure relation (see for example [24, 26, 29]) to truncate the hierarchy at some desired level. The simplest scenario is to truncate at the first level by writing down an expression for the density-density correlation function in terms of the cell density distribution. Namely, to adopt the Mean Field Assumption (MFA)

$$p_3(x, x', t) = p_1(x, t)p_1(x', t).$$

(2.22)

This implies that the probability that one cell is at position $x$ at time $t$ is statistically independent of the position of any other cell in the population. Upon substituting (2.22) into (2.21a), we obtain

$$\frac{\partial p_1(x, t)}{\partial t} = \nabla^2 p_1(x, t) - F_0(N - 1)\nabla (p_1(x, t)\phi(x, t)).$$

(2.23)

where

$$\phi(x, t) = \int F(\|x - x'\|) \frac{x - x'}{\|x - x'\|} p_1(x', t)dx'.$$

(2.24)

is the velocity field generated by intercellular mechanical interactions between cells. Equation (2.23) is of the same form used in [3, 4, 8, 28, 54] to provide a mean-field description of cell-cell adhesion. It has also been used to model social dynamics, such as ‘swarming’ [44]. In [3], equation (2.23) was derived from purely deterministic considerations, and in [9] it was found that this expression provides a valid approximation to a deterministic off-lattice model, but only provided that the length-scale over which mechanical interactions occur is large compared to the distance between individual cells. However, we recall that the biologically relevant regime is where mechanical interactions occur over just a few cell diameters, and so we will restrict the sensing range of a cell to $\sigma \epsilon$, with $0 < \sigma < 3$. We will find that in this case, consistent with the findings of [34, 67] for lattice-based models, ignoring correlations (i.e. by adopting the MFA) can have a strong impact on the continuum model’s ability to predict the behaviour of the associated IBM.

2.3.2 Kirkwood superposition approximation

In Section 3 we compare numerical solutions of (2.23) to the analogous ensemble averages obtained from the IBM, and find that in many cases the MFA does not provide an adequate approximation of the cell density distribution. We therefore investigate an alternative closure relation, namely the Kirkwood Superposition Approximation (KSA), which takes the form:

$$p_3(x, x', x'', t) = \frac{p_2(x, x', t)p_2(x, x'', t)p_2(x', x'', t)}{p_1(x, t)p_1(x', t)p_1(x'', t)}.$$  

(2.25)

This closure relation originated as a means to close the BBGKY hierarchy in the theory of simple liquids [37]; it is typically a good approximation in the limit of low densities and can be derived from maximum entropy considerations [70]. It has also been previously used as a closure approximation for lattice-based models of cell-cell adhesion [34, 67] however to our knowledge this is the first time that it is applied to an off-lattice model of cell movement and interaction.
Thus to be explicit, our KSA-based continuum model is given by (2.21a) and the closed form of (2.21b) which reads:

\[
\frac{\partial p_2(x, x', t)}{\partial t} = \nabla^2 p_2(x, x', t) - F_0 \left( \frac{\partial}{\partial x} - \frac{\partial}{\partial x'} \right) \left( F(\|x - x'\|) \frac{x - x'}{\|x - x'\|} p_2(x, x', t) \right)
\]

\[\quad - F_0(N - 2) \frac{\partial}{\partial x} \int F(\|x - x''\|) \frac{x - x''}{\|x - x''\|} p_2(x, x', t) p_2(x', x'', t) \frac{p_2(x', x'', t)}{p_2(x', t) p_1(x', t)} d\mathbf{x}''
\]

\[\quad - F_0(N - 2) \frac{\partial}{\partial x'} \int F(\|x' - x''\|) \frac{x' - x''}{\|x' - x''\|} p_2(x, x', t) p_2(x, x'', t) p_2(x', x'', t) \frac{p_2(x, x', t)}{p_2(x, t) p_1(x, t) p_1(x', t)} d\mathbf{x}''.
\]

We recall that in (2.26), interactions between pairs of cells located at positions $x$ and $x'$ are captured by a local advection term, whereas mechanical interactions from cells at other locations enter through convolutions. In the absence of these local advection terms, (2.26) can be solved using separation of variables, i.e. $p_2(x, y) = p_1(x)p_1(y)$, i.e. we obtain the mean-field approximation. Thus, corrections to the mean-field are manifested through the appearance of the local advection term.

The uniqueness and existence of solutions to partial integro-differential equations of the same form as our MFA-based model (2.23) are discussed in [6, 10, 38], whereas blow-up of solutions is discussed in [6, 7]. In general, these considerations depend in large part on the regularity of the force function $F$ used. Performing a similar analysis on the KSA-based model (2.26), is beyond the scope of the current paper. However we note that in our numerical investigations of the KSA model, no evidence of blow-up could be observed. It has been previously reported that approximations that go beyond the MFA can eliminate blow-up [40], and so it is plausible that this is also the case for the KSA-based model presented here.

3 Numerical investigations

For simplicity, we restrict our attention to the one-dimensional problem. In particular, we compare estimates of the cell density distribution and the density-density correlation function obtained from simulations of the IBM (2.7) to solutions of the same quantities when either the MFA or KSA closure relation is adopted.

3.1 Methods

Both the MFA and the KSA-based models ((2.23) and (2.26) respectively) take the form of advection-diffusion equations. These can be particularly challenging to solve if the advection dominates over diffusion, wherein small errors can rapidly accumulate and lead to instabilities in the numerical scheme [60, 74]. This is particularly problematic in the KSA-based model (2.26), where the appearance of the local advection terms can cause the density-density correlation function to form sharp peaks in the $x - x'$ direction. It is therefore convenient to re-write (2.26) in terms of the centre of
mass \( \eta = x + x' \) (neglecting the factor 1/2) and cell-cell separation \( \xi = x - x' \) to obtain:
\[
\frac{\partial p_2(\eta, \xi, t)}{\partial t} = \nabla^2 p_2(\eta, \xi, t) - F_0 \frac{\partial}{\partial \xi} \left[ F(|\xi|) \frac{\partial p_2(\eta, \xi, t)}{\partial \xi} \right]
\]
\[- \frac{F_0(N-2)}{2} \left( \frac{\partial}{\partial \eta} + \frac{\partial}{\partial \xi} \right) \int F(|\eta + \xi|/2 - x'') \frac{\partial}{\partial \xi} \left[ \frac{\partial p_3(\xi, \eta, x'', t)}{\partial x''} \right] dx''
\]
\[- \frac{F_0(N-2)}{2} \left( \frac{\partial}{\partial \eta} - \frac{\partial}{\partial \xi} \right) \int F(|\eta - \xi|/2 - x'') \frac{\partial}{\partial \xi} \left[ \frac{\partial p_3(\xi, \eta, x'', t)}{\partial x''} \right] dx''
\]
(3.1)

where \( \text{sgn} \) is the sign function. The three cell correlation function \( p_3(\xi, \eta, x'') \) is given by the KSA (2.25), namely
\[
p_3(\xi, \eta, x'') = \frac{p_2((\eta - \xi)/2, (\eta + \xi)/2, x'') p_2((\eta - \xi)/2, \eta''/2, x'', t) p_2((\eta + \xi)/2, \eta''/2, x'', t)}{p_1((\eta - \xi)/2, t) p_1((\eta + \xi)/2, t) p_1(x'', t)}
\]
(3.2)

Numerically, it is more computationally efficient to obtain \( p_1(x, t) \) by integrating \( p_2(x, x', t) \) over \( x' \). In terms of \( \xi \) and \( \eta \), this integration can be written as:
\[
p_1(x, t) = \int_{-\infty}^{\infty} p_2(\eta, 2x - \eta, t) d\eta.
\]
(3.3)

Furthermore, in the original co-ordinates \( 0 < x < 1 \) and \( 0 < x' < 1 \) and we adopt periodic boundary conditions. In terms of \( \xi \) and \( \eta \), the position of these boundaries are now given by
\[
\eta = \begin{cases} 
\xi, & \text{for } 0 < \eta \leq 1, \\
2 - \xi, & \text{for } 1 < \eta \leq 2.
\end{cases}
\]
(3.4)

The MFA and KSA-based continuum models where numerically solved by first discretising space and then integrating the resulting nonlinear ordinary differential equations (ODEs) with independent variable \( t \) using MATLAB routine ode15s for stiff ODEs. Simulations of the IBM (2.7) were computed using the Euler-Maruyama method [58] with step size \( \Delta t = 10^{-7} \) and for \( t \in [0, 10^{-2}] \), although only 51 of these time steps, taken at regular intervals, were actually stored for our analysis. From these, discrete approximations to the cell density distribution \( (\hat{p}_3^{\text{ens}}) \) and the density-density correlation function \( (\hat{p}_3^{\text{ens}}) \) were generated directly from the simulations by taking ensemble averages of solutions obtained from the IBM. In other words, space was divided into \( M \) boxes, and \( \hat{p}_3^{\text{ens}}(i, j) \) was calculated by counting the average number of cells located within a box centred at position \( x_i \) at time \( t_j \). A similar procedure was devised to compute \( \hat{p}_3^{\text{ens}}(i, j, k) \). For each parameter choice, these averages were calculated from \( 10^5 \) simulations. Initial cell positions were sampled from the distribution \( p_0(x) \), where
\[
p_0(x) = \frac{f(x)}{\int_0^1 f(x) dx},
\]
(3.5)

and
\[
f(x) = \frac{1}{2} \left( \tanh(\alpha(x - \theta)) + \tanh(\alpha(1 - \theta - x)) \right),
\]
(3.6)

where \( \theta = 0.2 \) and \( \alpha = 30 \) in all simulations. These initial conditions correspond to the case where cells are densely packed in the middle of the one dimensional space (these initial conditions are analogous to those commonly used in other studies; see for example [75]). To compare solutions to the density-density correlation function obtained by solving (3.1) with those obtained (by taking
ensemble averages) from the IBM (2.7), we rotated the solutions back to their original co-ordinates $x$ and $x'$. The error $E$ between the cell density $p_1(i,j)$ generated by either the MFA-based model (2.23) or the KSA-based one (3.1) and the one obtained by ensemble averaging stochastic simulations of the IBM ($p_{1\text{ens}}(i,j)$) were computed according to

$$E = \frac{1}{MT} \sum_{i,j} (p_1(i,j) - p_{1\text{ens}}(i,j))^2,$$  \hspace{1cm} (3.7)

where the sum takes place over $1 \leq i \leq M$, $1 \leq j \leq T$, where $T = 51$ (see above). The stochastic nature of the IBM means that each realisation is different, and hence to visualise the degree of variability we provide multiple examples of this along the $y$-axis (see Figure 2b for example). Furthermore, the centre of mass of the system will vary for each realisation (recalling in particular that the boundary conditions are periodic), and so for comparison purposes we have rotated the cell positions for each realisation obtained, so that the middle cell (for a given time point) has position $x = 0.5$.

### 3.2 Results

To help characterise the different parameter regimes under consideration, we define $\rho = N\epsilon$, which is the ratio between the cell radius ($\epsilon$) and the characteristic separation distance between cells when their positions are distributed uniformly, namely $1/N$. For small $\rho$ ($< 1$), the typical separation between cells is large relative to their radius ($\epsilon$), and hence cells are sparsely distributed. On the other hand, if $\rho$ is large, cells are densely packed. We discuss each of these cases in turn below.

For sparsely distributed cells ($\rho < 1$) that are relatively soft ($a = 1$) and undergo weak interactions ($F_0$ small), cells spread throughout the domain so that the cell density distribution resembles a Gaussian (Figure 2a). Both the MFA-based and the KSA-based models ((2.23) and (2.26) respectively) provide an excellent agreement to this. The positions of cells in the IBM appear rather disordered (Figure 2b), and as we now explain, this apparent disorder is reflected in the structure of the density-density correlation function (Figure 2c-2d). For $x = x'$, $p_2(x, x')$ remains close to zero for all times, reflecting the fact that two cells feel a repulsive force when their separation is below $\epsilon$ (see (2.8) and Figure 1)). In this sense, cells have some minimal separation distance. Initially, cell separations above the minimal distance are approximately uniformly distributed (Figure 2c). As time evolves (Figure 2d), peaks in the density-density correlation function develop in IBM simulations. Analogous peaks do not appear in the MFA prediction of the density-density correlation function (given by computing (2.22); see Figure 2c-2d), but are captured by the KSA-based model. These structures in the density-density correlation function correspond to the fact that adhesive forces between cells (for separations between $\epsilon$ and $\sigma\epsilon$; Figure 1) pull cells together and cause correlations in their separation distance to appear. We note that for the case considered here, there is only one peak that forms for $x > x'$ axis (together with its reflection for $x < x'$), implying that the separation between a cell and its nearest-neighbour is correlated, but no such correlations are observed between cells and their next-nearest neighbour (which would lead to the formation of additional peaks, see below).

Importantly, we find that the strength of the correlation between pairs of cells depends on their position in the tissue. For example, consider Figure 2d. For a cell positioned in the centre ($x = 0.5$), there are two approximately equal maxima in $p_2$ at $x' = 0.4$ and $x' = 0.6$ (so that $p_2(0.5, 0.4) \approx p_2(0.5, 0.6) \approx 2.85$). These two maxima correspond to the two cells positioned on each side of the one at $x = 0.5$; the maxima in $p_2$ are of equal size indicating that the cell is equally
strongly correlated with its neighbours. However this is not the case for cells located towards the edge of the tissue. For example for \( x = 0.4 \), we find that \( p_2 \) has maxima at \( x' = 0.3 \) and \( x' = 0.5 \), but with \( p_2(0.4, 0.3) \approx 2.29 \) and \( p_2(0.4, 0.5) \approx 2.80 \); this indicates that a cell at \( x = 0.4 \) is more strongly correlated with its neighbour at \( x = 0.5 \) rather than its neighbour at \( x = 0.3 \). Generally we find that a cell’s position is more highly correlated with neighbours that are located toward the centre of a tissue, where interaction forces are highest, than with cells located further toward the tissue’s leading edge.

The KSA provides an excellent approximation to the density-density correlation function obtained from ensemble averaging the IBM. However, since intercellular forces have been chosen to be weak, the peaks in the density-density correlation function are not particularly sharp. This reflects the apparent disorder in individual IBM realisations (Figure 2b). Hence, the weak intercellular forces (small \( F_0 \)) mean that the discrepancy between the actual density-density correlation function (as measured by ensemble averages obtained from the IBM) and the MFA solution is relatively small, which incidentally is why the MFA provides such an excellent approximation to the cell density distribution (Figure 2a).

Increasing \( F_0 \) from 5 to 60, so that interaction forces are stronger, causes the peaks in the density-density correlation function to increase considerably in sharpness. Initially (Figure 3c), only one peak (together with its reflection) forms for \( x > x' \), with cells having a clearer separation distance than before (compare with Figure 2c). However, at later times (Figure 3d), additional ‘ripples’ can be observed, corresponding to weak correlations between cells and their next-nearest neighbours. The sharpness of these peaks reflects the fact that cell positions in the IBM are far more ordered than when cell-cell interactions were weak (compare Figure 2b and Figure 3b). Stronger correlations in turn lead to greater discrepancies in the cell density distribution predicted by the MFA-based model (compare Figures 2a and 3a). In particular, we found that for an intermediate interaction range (\( \sigma = 2 \)) and low densities, the MFA predicts that increases in \( F_0 \) will cause cells to disperse more rapidly, whereas in fact the opposite is true (Figure 3a). In contrast, the KSA correctly captures the time evolution of the IBM.

To explore the differences between the cell density distribution as generated by the MFA-based and KSA-based models, and that calculated directly from IBM simulations, we compute the error according to (3.7) for various sensing ranges (\( \sigma \)), intercellular force strengths (\( F_0 \)) and relative densities \( \rho \) (Figure 4). While the KSA-based model provides an excellent approximation to the cell density distribution for the various parameter regimes considered, the MFA only provides a good approximation for weakly interacting cells (for both sparse and dense populations). The discrepancy between the MFA and the ensemble averages from the IBM increases significantly with increasing \( F_0 \).

For sufficiently long-range interactions (in this case \( \sigma = 3 \)) and strong intercellular interactions (\( F_0 = 10 \)), the cell density distribution displays aggregation, this being a phenomenon that both the KSA and MFA based models can capture (Figure 5a). The fact that the MFA-based model is successful in predicting this behaviour is because increasing \( \rho \) and \( \sigma \) leads to greater compression of individual cells, effectively reducing the effect of the local advection term in (2.26), which is from where the correction to the MFA manifests (Figure 5b-5c). However, this rather extreme form of compression may reflect a biologically unrealistic scenario. This occurs for \( a = 1 \), corresponding to the case where cells have been chosen so that they are rather ‘soft’.

Increasing \( a \) from 1 to 2 abrogates the aggregation behaviour presented in Figure 5. The behaviour of the model is insensitive to variations in the range of interaction (\( \sigma \)), since the repulsive
Figure 2: Comparisons between various model versions (IBM, MFA and the KSA-based models) when cell densities are low ($\rho = 0.5$), mechanical interactions are relatively weak ($F_0 = 5$) and cells are soft ($a = 1$). (a) Comparison between solutions to the cell density distribution at various time points ($t = 0, 0.002, 0.004, 0.006, 0.008$) as obtained using either ensemble averages from the IBM (dots) or the MFA/KSA based continuum models (solid lines). (b) Multiple realisations obtained from the IBM for two time points (see Section 3). (c)-(d) Solutions to the density-density correlation function as obtained by either the MFA, the KSA or ensemble averaging the IBM for two time points. Remaining parameter values were chosen so that $N = 5$, $\epsilon = 0.1$ and $\sigma = 2$. 
Figure 3: Comparisons between various model versions (IBM, MFA and the KSA-based models) when cell densities are low ($\rho = 0.5$), mechanical interactions are relatively strong ($F_0 = 60$) and cells are soft ($a = 1$). (a) Comparison between solutions to the cell density distribution at various time points ($t = 0, 0.002, 0.004, 0.006, 0.008$) as obtained using either ensemble averages from the IBM (dots) or the MFA/KSA-based continuum models (solid lines). (b) Multiple realisations obtained from the IBM for two time points (see Section 3). (c)-(d) Solutions to the density-density correlation function as obtained by either the MFA, the KSA or ensemble averaging the IBM for two time points. Remaining parameter values were chosen so that $N = 5$, $\epsilon = 0.1$ and $\sigma = 2$. 
Figure 4: Relative errors between MFA (solid lines) and KSA-based continuum models (dashed lines) and ensemble-averaged IBM simulations as a function of the interaction range ($\sigma$), for low densities (left panel) and high densities (right panel), and for different interaction strengths ($F_0$). Points represent actual computed errors; lines are provided for visual clarity. The error increases with $F_0$ in each case. MFA-based model performs well for weak interaction strengths; in contrast the KSA-based model performs well for all interaction strengths. Errors were calculated as described in Section 3. Remaining parameter values were $\epsilon = 0.1$, $a = 1$ and $N = 5$ (low density) or $N = 20$ (high density).

force generated by each cell is now significantly stronger relative to the attractive part (Figure 1 and Section 2.1). As before, the KSA-based model accurately captures the cell density distribution obtained from ensemble averages (Figure 6a). Here, cells are under compression ($\rho = 2$) and the profile of the cell density distribution no longer resembles a Gaussian; rather the initial shape of the front is now maintained by intercellular forces as cells propagate throughout the domain (Figure 6c). This behaviour is captured by the KSA, but not by the MFA-based model. This ‘front-like’ propagation is accompanied by strong correlations between cells. We note that these correlations do not only occur at the position of a cell’s nearest neighbour (as was the case when $a = 1$), but extend to strong correlations between multiple neighbouring cells. These correlations reflect the increased “order” in individual IBM realisations (compare for example Figure 2b with Figure 6b) which more likely reflects the composition of a biologically realistic epithelial tissue. We also checked whether the behaviour of our model was dependent on box size; we therefore repeated simulations, but with the length of the domain chosen to be double that of the original. We found no discernible differences from our previous simulations (see Figure 7). This confirms that results presented here are independent of box size.

4 Discussion

In this paper we have derived the first continuum approximation of an off-lattice model where strong cell-cell adhesion (and hence cell correlations) are captured. Our finding that the KSA-based approach is better suited to capturing tissue-scale effects from cell-cell adhesion than the MFA-based models is in broad agreement with previous lattice-based studies [34,67]; however our
Figure 5: Comparisons between various model versions (IBM, MFA and the KSA-based models) when cell densities are high ($\rho = 2$), mechanical interactions are relatively strong ($F_0 = 10$) and cells are soft ($a = 1$). (a) Comparison between solutions to the cell density distribution at various time points ($t = 0, 0.002, 0.004, 0.006, 0.008$) as obtained using either ensemble averages from the IBM (dots) or the MFA/KSA-based continuum models (solid lines). (b) Multiple realisations obtained from the IBM for two time points (see Section 3). (c)-(d) Solutions to the density-density correlation function as obtained by either the MFA, the KSA or ensemble averaging IBM simulations for two time points. Remaining parameter values were chosen so that $N = 5, \epsilon = 0.1$ and $\sigma = 3$. 
Figure 6: Comparisons between various model versions (IBM, MFA and the KSA-based models) when cell densities are high ($\rho = 2$), mechanical interactions are relatively strong ($F_0 = 10$) and cells are hard ($a = 2$). Increasing $a$ from 1 (as used in Figures 2-5) to 2 causes cells to propagate much faster, and so simulations are computed only up to time $t = 0.0016$. (a) Comparison between solutions to the cell density distribution at time points ($t = 0, 0.0004, 0.0008, 0.0012, 0.0016$) as obtained using either ensemble averages from the IBM (dots) or the MFA/KSA-based models (solid lines). (b) Multiple realisations obtained from the IBM for two time points (see Section 3). (c) Blow up of the comparison between the KSA, MFA and ensemble averaged IBM simulations shown in (a). (d) Solutions to the density-density correlation function as obtained by either the MFA, the KSA or ensemble averaging the IBM. Remaining parameter values were chosen so that $N = 10$, $\epsilon = 0.1$ and $\sigma = 2$. 

19
methodology is off-lattice and hence provides an additional level of realism which is missing from the current literature. As in [18, 34, 47, 48, 67], we have restricted our model to one dimension, this being particularly relevant for cases where cell migration is axially symmetric, such as scratch (or wound healing) assays [57, 64, 65, 75]. Currently, most experimental data on scratch assays are obtained at the tissue scale, and come in the form of bulk quantities such as cell density. However in recent years it has also become possible to collect data at the cell-level such as cell position and velocity. The speciality of the description of cell motility and adhesion presented in this paper is that it connects the tissue-level observables with the cell-level parameters. This was not previously possible since derivations either neglected cell-cell correlations [18, 47, 48] or were based on artificial lattices [34, 67]. Thus, our model’s biological utility is twofold: (i) it can be used to predict tissue-level phenomena from knowledge of cell-cell interactions. (ii) it can be used to infer cell-level parameters from experimental tissue-level cell density data. The model also can be seen as a foundational generic framework for modelling cell movement to which one can incorporate other cell-level phenomena, e.g., chemotaxis, haptotaxis, as desired.

To illustrate the above points, we draw on the work of [57] (see also [64]). Here, it was shown that MDCK cells (which strongly adhere to one another as they migrate) have strong spatio-temporal correlations in the bulk velocities and positions of cells and that these extend over many cell diameters. The authors found that this was in direct contrast to Normal Rat Kidney (NDK) cells, which have much weaker intercellular bonds, and hence a far shorter correlation length. The KSA-based model provides a natural framework to explore the differences between these two cell types. In particular, one can predict directly from our work that MDCK cells (in addition to the experimentally observed strong correlations in relative cell position) will collectively propagate in a front-like manner (see Figure 6). On the other hand, the propagation of the weaker adhering NDK cells will more closely resemble diffusion. This is an example of how the model can be used to predict tissue-level phenomena from limited experimental knowledge of cell-cell interactions. However, there are other such phenomena, such as metastasis in cancer cells (whereby a few cells break away from the bulk tissue) for which an IBM may be more suitable than a continuum-based approach. The problem in this case is parameterising the IBM when only tissue-level data is available. Our KSA-based approach can be utilised to circumvent this problem by fitting experimental tissue-level data, e.g., cell density versus time, to the KSA model solutions, from which one can subsequently infer the non-dimensional cell-level parameters defining the IBM. Cell-level phenomena could then be studied further by running simulations of the parameterised IBM. Thus, the framework presented here provides a practical means to move between different scales in a systematic manner, and so develop a far deeper understanding of how collective cell migration is orchestrated.

There are a number of interesting challenges and extensions to this work that, while beyond the scope of this paper, are worth discussing. Although our comparison of the IBM with the KSA and MFA-based continuum models was in one spatial dimension, the framework we developed is already formulated for general number of spatial dimensions. The anticipated major additional computational challenge in higher dimensions than one, stems from an increase in the number of equations to be solved for the KSA-based continuum model. In this case, efficient solution of the model may require the development of more sophisticated numerical approaches. Second, one could investigate the integration of more sophisticated force laws. An example of this comes from Johnson-Kendall-Roberts (JKR) theory [33] and is adopted for IBMs in [16, 56]. JKR theory models interactions between adhesive elastic spheres (in contrast to the Hertz force law which is used to model interactions between non-adhesive spheres [39]). In particular, it captures the hysteresis that
occurs when two cells first come into contact, and then are pulled apart. Thus, unlike the force laws described in Section 2.1, which are all single valued functions of the separation between two cell centres, application of JKR theory to our IBM would entail using a multivalued force function.

Finally, we note that in the IBM used in this paper, cells are described as point masses. In reality, cells can adopt complex morphologies, with the shape of the membrane strongly influencing the cells’ mechanical properties. Other IBM-based approaches, such as the CPM (lattice-based) or SEM (off-lattice), can readily be used to capture a cell’s geometry. Recently, a number of continuum approximations to CPM based models have appeared in the literature [1, 23, 40]. The methodology developed in the present paper can be extended to capture effects due to changes in the shape of a cell’s membrane by using as a starting point the Langevin equations defining the SEM [50]. In the latter cells are composed of multiple “effective particles”, with the equations of motion being similar in form to (2.7), but with separate terms reflecting intracellular and intercellular forces. Our procedure applied to the SEM would lead to partial differential equations for the number density distribution of “effective particles” composing cells and for the correlations between them.

In summary, the KSA-based model (2.26) presented here can be directly related to an IBM which incorporates the stochastic motion emanating from individual cell propulsion, together with the mechanical forces that arise due to cell-cell interactions. The additional flexibility that comes from adopting a stochastic framework, that provides information on the underlying cell-cell correlations, and is valid for both small and large cell populations, may prove beneficial if one is to compare the model outputs to real data. Solving the KSA-based model is considerably faster that obtaining ensemble averages from multiple simulations of the IBM. For example, on a Macbook Pro with 2.2 GHz Intel Core i7 processor, it take approximately five minutes to solve the KSA-based model (regardless of the number of cells or other parameters in the model). In contrast, to obtain one hundred simulations from the IBM, for twenty cells, takes approximately 3 hours. The duration of these simulations reflect that we are probing the long-term behaviour of the cells’ migration. Since the simulations are stochastic, it might be that many more realisations from the IBM are required to obtain reasonable estimates for the cell density distribution or the density-density correlation function, depending on parameter values (for example, we used one hundred thousand simulations in the comparisons presented here). Concluding, the KSA-based continuum model derived in this paper provides high accuracy at a modest computational effort and hence provides a powerful novel framework to model cell movement and interaction.

Acknowledgements

R. G. acknowledges support from SULSA (Scottish Universities Life Science Alliance) and from FRIAS (Freiburg Institute for Advanced Studies). AMM and CF acknowledge support from the BMBF Freiburg Initiative in Systems Biology Grant 0313921 (FRISYS)

References

Figure 7: Comparisons between various model versions (IBM, MFA and the KSA-based models) using the same parameter values as in Figure 6 ($F_0 = 10$, $a = 2$, $N = 10$, $\epsilon = 0.1$ and $\sigma = 2$), except that the size of the box ($L$) is now doubled. There are no discernible qualitative differences between the model results presented here and in Figure 6, indicating that box size does not play a significant role in the behaviour of the system or the performance of the MFA and KSA-based models.


