Continuum approximations for lattice-free multi-species models of collective cell migration

Oleksii M Matsiaka[1], Catherine J Penington[1], Ruth E Baker[2], *Matthew J Simpson[1]

[1] *School of Mathematical Sciences, Queensland University of Technology (QUT) Brisbane, Queensland, Australia.*

[2] *Mathematical Institute, University of Oxford, Radcliffe Observatory Quarter, Woodstock Road, Oxford, United Kingdom.*

Abstract

Cell migration within tissues involves the interaction of many cells from distinct subpopulations. In this work, we present a discrete model of collective cell migration where the motion of individual cells is driven by random forces, short range repulsion forces to mimic crowding, and longer range attraction forces to mimic adhesion. This discrete model can be used to simulate a population of cells that is composed of $K \geq 1$ distinct subpopulations. To analyse the discrete model we formulate a hierarchy of moment equations that describe the spatial evolution of the density of agents, pairs of agents, triplets of agents, and so forth. To solve the hierarchy of moment equations we introduce two forms of closure: (i) the mean field approximation, which effectively assumes that the distributions of individual agents are independent; and (ii) a moment dynamics description that is based on the Kirkwood superposition approximation. The moment dynamics description provides an approximate way of incorporating spatial patterns, such as agent clustering, into the continuum description. Comparing the performance of the two continuum descriptions confirms that both perform well when adhesive forces are sufficiently weak. In contrast, the moment dynamics description outperforms the mean field model when adhesive forces are sufficiently large. This is a first attempt to provide an accurate continuum description of a lattice-free, multi-species model of collective cell migration.

Key words: Cell migration; mean field approximation; moment dynamics approximation; Kirkwood superposition approximation.

1 Introduction

In vivo cell migration involves many different cell types interacting with each other. For example, tumour invasion involves malignant cancer cells moving through normal surrounding tissues (Weinberg, 2009). Interactions between different cell types are also captured in certain *in vitro* experiments, such as the migration of malignant melanoma cells, which is thought to be enhanced when these cells are moving amongst skin cells (Eves et al., 2003). Multiple species of cells can also be created in experiments where some subpopulation of cells, amongst an otherwise identical subpopulation, are labelled and tracked over time (Simpson et al. 2006; Simpson et al., 2007). While some mathematical models explicitly account for interactions between different subpopulations of cells (Painter and Sherratt, 2003), most mathematical models deal with a single population of cells only (Sherratt and Murray, 1990; Maini et al., 2004).

A common approach to modelling cell migration is to use a lattice-based random walk model. This approach captures details of the motion of individual cells, which is attractive because this kind of information can be linked to time lapse images from experiments. The continuum-limit description of such a lattice-based model can also be used to study the group behaviour. Although some previous lattice-based models account for interactions between different types of cells (Simpson et al., 2009; Penington et al., 2011), these lattice-based models are unrealistic because real cells do not move on regular lattice-based structures. Other limitations of lattice-based models include restrictions on cell size. For example, the diameter of a typical melanoma cell is approximately 18 μm (Treloar et al., 2013) whereas the diameter of a typical skin cell is approximately 25 μm (Simpson et al., 2013). In a model with both types of cells present, it is not possible to accommodate these differences in cell size

* Corresponding author
Email address: matthew.simpson@qut.edu.au, *Telephone* + 617 3138 5241, *Fax* + 617 3138 2310 (*Matthew J Simpson[1]).

if we use a standard lattice-based approach where each cell occupies a single lattice site (Binder and Simpson, 2016).

To address these limitations, we define a lattice-free model that can be used to describe the migration of a population of cells that is composed of many potentially distinct subpopulations. We adopt a modelling framework that is an extension of previous approaches by Newman and Grima (2004) and Middleton and co-workers (2014). The work by Newman and Grima considered a stochastic model of individual cell migration, with chemotactic effects, and they described the continuum limit using a Langevin formulation. The work of Newman and Grima (2004) was then extended by Middleton and co-workers (2014) who also considered a stochastic model of individual cell migration in terms of a Langevin formulation, however they considered both a traditional mean field continuum approximation as well as a more sophisticated moment closure continuum approximation that accounts for the spatial and temporal dynamics of pairs of agents. A key feature of both these previous models is that they are appropriate for studying the collective migration of a single populations of cells. However, many practical problems in development and disease progression involves multiple interacting subpopulations of cells. Therefore, the main aim of the current study is to develop a discrete model of collective migration where the total population of cells consists of an arbitrary number of interacting subpopulations. Our discrete model incorporates random cell motility, adhesion between cells and finite size effects (crowding). We allow for differences in cell size, cell motility and cell adhesion between the different subpopulations. In addition to producing stochastic realisations of the discrete model, we also analyse the continuum limit using both a standard mean field approximation and a more sophisticated moment dynamics approximation. Comparing averaged behaviour from the discrete simulations with the solution of the continuum models confirms that the mean field approach can be inaccurate when adhesion is sufficiently strong. This is important because almost all mathematical models of collective cell migration invoke the mean field approximation (Sherratt and Murray, 1990; Painter and Sherratt, 2003; Maini et al., 2004).

This manuscript is organised in the following way. In Section 2 we describe the discrete model. In Section 3.1, we analyse the discrete model, showing how

we can obtain a continuum description of the average behaviour of the discrete model. In particular, we focus on two different continuum descriptions: (i) a mean field approximation; and (ii) a higher-order moment dynamics approximation. Results in Sections 3.2-3.3 compare solutions of both continuum approximations and averaged discrete results for problems involving one and two interacting subpopulations, with additional comparisons presented in the Supplementary Material. In Section 3.4 we investigate how the accuracy of the MFA and KSA approximations depends on the choice of model parameters. Finally, in Section 4, we summarise our work and highlight opportunities for future investigation.

2 Discrete model

We consider a population of N cells that is composed of an arbitrary number of subpopulations, $K \geq 1$. Illustrative schematics showing interactions between individuals in a population with $K = 1$ and $K = 2$ subpopulations are given in Figure 1(a)-(b), respectively.

We begin by assuming that each individual cell is a point mass and that its movement can be described by an equation of motion. For simplicity, from this point on, we restrict our attention to a one-dimensional geometry, and in Section 4 we discuss how the framework can be adapted to higher dimensions. To begin describing the collective motion, we assume that the motion of each cell is governed by Newton's second law,

$$m_i \frac{\mathrm{d}^2 x_i}{\mathrm{d}t^2} = \mathcal{V}_i + \sum_{j \neq i} R_{ij} + \zeta_i, \qquad i = 1, \ldots, N, \tag{2.1}$$

where x_i is the position of the ith cell, m_i is its mass, and R_{ij} is an interaction force between the ith and jth cells. $\mathcal{V}_i$ is the viscous force between the cell and the surrounding medium, and ζ_i is the stochastic force associated with random Brownian motion. According to Stokes' law, the viscous force on a small spherical particle moving in a viscous fluid is given by

$$\mathcal{V}_i = -\mu \frac{\mathrm{d}x_i}{\mathrm{d}t}, \tag{2.2}$$

where $\mu > 0$ is the drag coefficient. If we neglect inertial forces and invoke

Stokes' law (Middleton et al., 2014), we arrive at a system of Langevin stochastic differential equations (SDEs) given by

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = \sum_{j \neq i} F_{ij} + \xi_i, \qquad i = 1, \ldots, N, \tag{2.3}$$

where $R_{ij} = \mu F_{ij}$ and $\zeta_i = \mu \xi_i$.

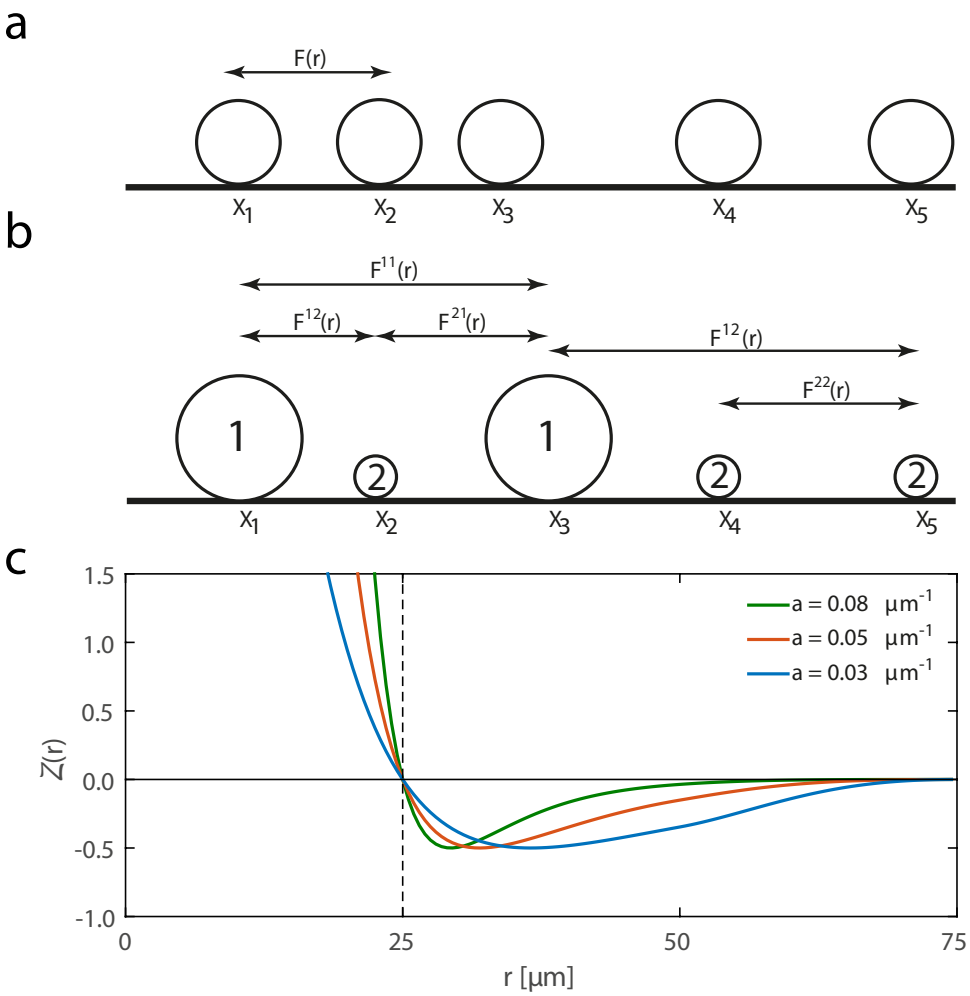

Fig. 1. (a)-(b) Representative plot of single- and multi-species systems of cells, respectively. In (a) we show the intraspecies force, $F(r)$, and in (b) we show both intraspecies forces, $F^{11}(r)$ and $F^{22}(r)$, and interspecies forces, $F^{12}(r)$ and $F^{21}(r)$. Here, r is the distance between cells. (c) Dimensionless force law function $\mathcal{Z}(r)$, given by Equation (2.5), for various values of a. Here, $\delta = 25$ μm corresponds to a typical cell diameter.

In summary, according to Equation (2.3), the collective migration of cells is determined by a balance between cell-to-cell interactions (short-range crowding

and longer range adhesion), stochastic forces, and viscous forces. Collective cell migration that is driven by unbiased stochastic forces is thought to be relevant in many applications, such as collective cell spreading in many single-species *in vitro* experiments (Simpson et al., 2013). Therefore, we focus on unbiased stochastic forces by sampling ξ_i from a Gaussian distribution with zero mean and zero auto-correlation (Middleton et al. 2014).

It is biologically reasonable to model the interaction forces between cells, F_{ij}, to have different amplitudes for subpopulations of cells. This is relevant if we wish to specify different adhesion forces between different subpopulations (Steinberg, 1996). For simplicity, we assume $F_{ij} = F_{ji}$, and we specify the interaction force to be

$$F_{ij} = f_0 \, \mathcal{Z}(r) \, \mathrm{sgn}(x_i - x_j), \tag{2.4}$$

where f_0 is the dimensional amplitude of the interaction force, $\mathcal{Z}(r)$ is the dimensionless force law function that depends on the separation distance, and $r = |x_i - x_j|$. The function sgn is the *signum* function. The particular choice of $\mathcal{Z}(r)$ depends on phenomenological cellular behaviour we wish to model. Several force laws have been suggested, including a linear spring model (Murray et al., 2009) and non-linear force laws such as Morse (Middleton et al., 2014) or Lennard-Jones (Jeon et al., 2010) potentials. In this work we adopt a modified Morse potential force law, so that

$$\mathcal{Z}(r) = \begin{cases} 2\Big(\exp(-2a\,(r-\delta)) - \exp(-a\,(r-\delta)) \Big), & r < 2\delta, \\ 2\Big(\exp(-2a\,(r-\delta)) - \exp(-a\,(r-\delta)) \Big)\, g(r), & 2\delta \le r \le 3\delta, \\ 0, & r > 3\delta, \end{cases} \tag{2.5}$$

where a is a parameter that controls the shape of the force function, as illustrated in Figure 1(c), and δ is the cell diameter. The distance $r = \delta$ corresponds to the case where two cells are just in contact with each other. In this model we have $\mathcal{Z}(\delta) = 0$ and $\mathcal{Z}(r) = 0$ for $r > 3\delta$. Introducing the cell diameter δ in Equation (2.5) allows us to more realistically model the behaviour of multi-species populations of cells with different diameters.

Equation (2.5) incorporates the Tersoff cut-off function, $g(r)$ (Tersoff, 1988),

to capture a finite interaction range between cells. This function is given by

$$g(r) = \frac{1}{2}\left(1 - \sin\left(\pi\frac{2r-\delta}{2\delta}\right)\right). \tag{2.6}$$

The interaction range has been chosen to be 3δ (Srinivas et al., 2004).

A representative plot of $\mathcal{Z}(r)$ for different values of a is given in Figure 1(c). The force function consists of two regimes: short-range repulsion and longer range attraction. The repulsive term mimics crowding effects while the attractive tail models adhesion. While all of the results presented in this work are for this particular choice of force law, it is straightforward to incorporate other choices of $\mathcal{Z}(r)$.

3 Results and discussion

3.1 Mathematical model for an arbitrary number of subpopulations

We consider a total population of N cells that come from K subpopulations of cells, so that $N = \sum_{k=1}^{K} n_k$, where n_k is the number of cells in subpopulation k. This framework can be used to model both situations where each subpopulation is distinct (Eves et al. 2003) and situations where each subpopulation is composed of tagged, but otherwise identical cells (Simpson et al. 2006; Simpson et al. 2007). In addition, these distinct subpopulations may differ in many ways, such as differences in diameter, motility rates, or interaction forces and they can be arbitrarily arranged in space.

We define the one-cell probability density function (PDF), $P_1^i(x,t)$, as the probability that the position of cell i is in the small neighbourhood $[x, x+\mathrm{d}x]$ at time t. Similarly, we define the two-cell PDF, $P_2^{ij}(x,y,t)$, as the probability that cells i and j lie in $[x, x+\mathrm{d}x]$ and $[y, y+\mathrm{d}y]$, respectively, at time t. At present, we do not specify which of the subpopulations these cells belong to.

Given that the motile behaviour of cells is governed by Equation (2.3), we can

relate the PDFs to the position of cells as follows (van Kampen, 1975),

$$P_1^i(x,t) = \langle \delta^{(N)}(x - x_i(t)) \rangle, \tag{3.1}$$

$$P_2^{ij}(x,y,t) = \langle \delta^{(N)}(x - x_i(t))\, \delta^{(N)}(y - y_j(t)) \rangle, \tag{3.2}$$

where $x_i(t)$ and $y_j(t)$ are the positions of cells given by Equation (2.3). The angled brackets indicate an average over a sufficiently large number of identically prepared initial conditions and a sufficiently large number of realisations of the stochastic force. Further background explanation about Equations (3.1) and (3.2) is given in the Supplementary Material.

The time evolution of $P_1^i(x,t)$ is governed by a Fokker-Planck equation (Supplementary Material),

$$\frac{\partial P_1^i(x,t)}{\partial t} = D\Delta P_1^i(x,t) - \nabla \left(f_i\, P_1^i(x,t) \right), \tag{3.3}$$

which describes the motion of particles under the influence of random forces, proportional to the diffusivity, D, and directed drift forces, f_i. The force f_i acting on cell i in subpopulation l may be expressed as the sum of two types of forces: intraspecies forces exerted by other members of subpopulation l, and interspecies forces exerted by cells from all other subpopulations, giving

$$f_i = \sum_{j \neq i} F_{ij}^{ll} \; + \sum_{k \neq l}^{K} \sum_{j \in k} F_{ij}^{lk}. \tag{3.4}$$

Combining Equations (3.1), (3.3) and (3.4), and taking the convolution of the interaction force and a δ-function centred at y_j, we obtain

$$\begin{aligned} \frac{\partial P_1^i(x,t)}{\partial t} &= D\Delta P_1^i(x,t) \\ &\quad - \nabla \Big\langle \sum_{j \in l, j \neq i} \int_\Omega F^{ll}(x_i - y)\, \delta(x - x_i(t))\, \delta(y - y_j(t))\, \mathrm{d}y \Big\rangle \\ &\quad - \sum_{k \neq l}^{K} \nabla \Big\langle \sum_{j \in k} \int_\Omega F^{lk}(x_i - y)\, \delta(x - x_i(t))\, \delta(y - y_j(t))\, \mathrm{d}y \Big\rangle, \end{aligned} \tag{3.5}$$

where Ω denotes the domain. The second and third terms on the right hand side of Equation (3.5) are advection terms that incorporate intraspecies and interspecies forces, respectively. Combining Equations (3.2) and (3.5), and

interchanging summation and integration, we obtain

$$\frac{\partial P_1^i(x,t)}{\partial t} = D\Delta P_1^i(x,t) - \nabla \int_\Omega F^{ll}(x-y) \sum_{j\in l, j\neq i} P_2^{ij}(x,y,t)\,\mathrm{d}y - \sum_{k\neq l}^{K} \nabla \int_\Omega F^{lk}(x-y) \sum_{j\in k} P_2^{ij}(x,y,t)\,\mathrm{d}y, \tag{3.6}$$

where, from this point forward, we drop the subscript i on x_i.

To make the transition from individual level behaviour in a discrete simulation to the population level dynamics, we define the following quantities,

$$p_1^l(x,t) = \frac{1}{n_l} \sum_{i\in\mathcal{L}} P_1^i(x,t), \tag{3.7}$$

$$p_2^{ll}(x,y,t) = \frac{1}{n_l(n_l-1)} \sum_{i\in\mathcal{L}} \sum_{j\in\mathcal{L}, j\neq i} P_2^{ij}(x,y,t), \tag{3.8}$$

$$p_2^{lm}(x,y,t) = \frac{1}{n_l n_m} \sum_{i\in\mathcal{L}} \sum_{j\in\mathcal{L}} P_2^{ij}(x,y,t), \qquad l\neq m, \tag{3.9}$$

where $p_1^l(x,t)$ is the normalised one-cell density distribution of subpopulation l, $p_2^{ll}(x,y,t)$ is the density-density correlation function that captures intraspecies correlations, and $p_2^{lm}(x,y,t)$ is the density-density correlation function that captures interspecies correlations.

To proceed, we sum over the index i in Equation (3.6) and apply the definitions given in Equations (3.7)-(3.9). We repeat this procedure K times for each subpopulation to yield a system of K non-linear integro partial differential equations (IPDEs), that can be written as

$$\frac{\partial p_1^l(x,t)}{\partial t} = D\Delta p_1^l(x,t) - (n_l-1)\nabla\Big(\int_\Omega F^{ll}(x-y)\, p_2^{ll}(x,y,t)\,\mathrm{d}y\Big) - \sum_{k\neq l}^{K} n_k\, \nabla\Big(\int_\Omega F^{lk}(x-y)\, p_2^{lk}(x,y,t)\,\mathrm{d}y\Big), \tag{3.10}$$

for each subpopulation l. We define the PDF of the total population of N cells as a weighted sum of the individual distributions,

$$p_1^{\text{total}}(x,t) = \frac{1}{N} \sum_{k=1}^{K} n_k\; p_1^k(x,t). \tag{3.11}$$

Equation (3.10) shows that the evolution of $p_1^l(x,t)$ depends on $p_2^{ll}(x,y,t)$. To derive an evolution equation for $p_2^{ll}(x,y,t)$ we begin with the two-cell Fokker-

177 Planck equation,

$$\frac{\partial P_2^{ij}(x,y,t)}{\partial t} = D\Delta P_2^{ij}(x,y,t) - \frac{\partial}{\partial x}\left(f_i\, P_2^{ij}(x,y,t)\right) - \frac{\partial}{\partial y}\left(f_j\, P_2^{ij}(x,y,t)\right), \tag{3.12}$$

178 where cells i and j both belong to subpopulation l. The forces f_i and f_j, applied
179 to cells i and j, can be written as the sum of intraspecies and interspecies
180 forces. For example, the force on an arbitrary cell z can be written as

$$f_z = \sum_{s\neq z} F_{zs}^{ll} + \sum_{k\neq l}^{K}\sum_{s\in k} F_{zs}^{lk}. \tag{3.13}$$

181 Adopting the interaction force law, Equation (2.4), using the definition of
182 the two-cell PDF as given by Equation (3.2), and evaluating the required
183 convolutions, we can rewrite Equation (3.12) as

$$
\begin{aligned}
\frac{\partial P_2^{ij}(x,y,t)}{\partial t} &= D\Delta P_2^{ij}(x,y,t) \\
&\quad - \frac{\partial}{\partial x}\left\langle F^{ll}(x-y)\,\delta(x-x_i(t))\,\delta(y-y_j(t))\right\rangle \\
&\quad - \frac{\partial}{\partial y}\left\langle F^{ll}(y-x)\,\delta(x-x_i(t))\,\delta(y-y_j(t))\right\rangle \\
&\quad - \frac{\partial}{\partial x}\left\langle \sum_{g\in l, g\neq i,j}\int_\Omega F^{ll}(x-z)\,\delta(x-x_i(t))\,\delta(y-y_j(t))\,\delta(z-z_g(t))\,\mathrm{d}z\right\rangle \\
&\quad - \frac{\partial}{\partial y}\left\langle \sum_{g\in l, g\neq i,j}\int_\Omega F^{ll}(y-z)\,\delta(x-x_i(t))\,\delta(y-y_j(t))\,\delta(z-z_g(t))\,\mathrm{d}z\right\rangle \\
&\quad - \sum_{k\neq l}^{K}\frac{\partial}{\partial x}\left\langle \sum_{g\in k}\int_\Omega F^{lk}(x-z)\,\delta(x-x_i(t))\,\delta(y-y_j(t))\,\delta(z-z_g(t))\,\mathrm{d}z\right\rangle \\
&\quad - \sum_{k\neq l}^{K}\frac{\partial}{\partial y}\left\langle \sum_{g\in k}\int_\Omega F^{lk}(y-z)\,\delta(x-x_i(t))\,\delta(y-y_j(t))\,\delta(z-z_g(t))\,\mathrm{d}z\right\rangle, \qquad (3.14)
\end{aligned}
$$

where the second and third terms on the right hand side of Equation (3.14) represent interactions between cells i and j, the fourth and fifth terms on the right hand side of Equation (3.14) represent interactions between cells i and j and other cells within subpopulation l, and the sixth and seventh terms on the right hand side of Equation (3.14) represent interactions between cells i and j and cells in other subpopulations.

The three-particle normalised density functions can be defined as,

$$p_3^{lms}(x,y,z,t) = \frac{1}{n_l n_m n_s} \sum_{i\in l} \sum_{j\in m} \sum_{g\in s} P_3^{ijg}(x,y,z,t), \tag{3.15}$$

$$p_3^{lls}(x,y,z,t) = \frac{1}{n_l(n_l-1)n_s} \sum_{i\in l} \sum_{j\in l, j\neq i} \sum_{g\in s} P_3^{ijg}(x,y,z,t), \tag{3.16}$$

$$p_3^{lll}(x,y,z,t) = \frac{1}{n_l(n_l-1)(n_l-2)} \sum_{i\in l} \sum_{j\in l, j\neq i} \sum_{g\in l, g\neq i,j} P_3^{ijg}(x,y,z,t). \tag{3.17}$$

We therefore require a definition for the three particle PDF, $P_3^{ijg}(x,y,z,t)$, similar to Equation (3.2),

$$P_3^{ijg}(x,y,z,t) = \langle \delta(x - x_i(t))\, \delta(y - y_j(t))\, \delta(z - z_g(t)) \rangle. \tag{3.18}$$

To proceed we divide Equation (3.14) by $n_l(n_l-1)$, and combine Equations (3.14)-(3.18), summing over the indices i and j, to obtain an expression for the evolution of $p_2^{ll}(x,y,t)$,

$$\begin{aligned}
\frac{\partial p_2^{ll}(x,y,t)}{\partial t} = {} & D\Delta p_2^{ll}(x,y,t) \\
& - \frac{\partial}{\partial x}\Big(F^{ll}(x-y)\, p_2^{ll}(x,y,t) \Big) \\
& - \frac{\partial}{\partial y}\Big(F^{ll}(y-x)\, p_2^{ll}(x,y,t) \Big) \\
& - (n_l-2)\, \frac{\partial}{\partial x} \int_\Omega F^{ll}(x-z)\, p_3^{lll}(x,y,z,t)\, \mathrm{d}z \\
& - (n_l-2)\, \frac{\partial}{\partial y} \int_\Omega F^{ll}(y-z)\, p_3^{lll}(x,y,z,t)\, \mathrm{d}z \\
& - \sum_{k\neq l}^{K} n_k\, \frac{\partial}{\partial x} \int_\Omega F^{lk}(x-z)\, p_3^{llk}(x,y,z,t)\, \mathrm{d}z \\
& - \sum_{k\neq l}^{K} n_k\, \frac{\partial}{\partial y} \int_\Omega F^{lk}(y-z)\, p_3^{llk}(x,y,z,t)\, \mathrm{d}z.
\end{aligned} \tag{3.19}$$

The total system of equations governing the evolution of the density-density correlation functions for K subpopulations consists of K equations in the form of Equation (3.19), and $K!$ equations for the interspecies density-density correlation functions, $p_2^{kl}(x,y,t)$.

This procedure for deriving evolution equations for the density and density-density correlation functions can be repeated to yield a hierarchy of $N-1$

systems of IPDEs and a system of Fokker-Planck equations that govern the N-level density. At each level d, where $d \in [1, N]$, the d-density function p_d depends on the next order, p_{d+1}. This means that the full hierarchy of equations is, in general, both analytically and numerically intractable. Therefore, we must invoke some approximations to proceed, and we will now discuss two different approximations.

3.1.1 Mean field approximation

The simplest way to approximate the hierarchy is to truncate it at the first level by writing the density-density correlation function in terms of the one-cell density functions (Baker and Simpson, 2010),

$$p_2^{lm}(x, y, t) = p_1^l(x, t)\, p_1^m(y, t). \tag{3.20}$$

This approximation, often called the Mean Field Approximation (MFA), implies that the probability of finding one cell at $[x, x + \mathrm{d}x]$ at time t is independent of the probability of finding another cell at $[y, y + \mathrm{d}y]$ at the same time. MFA-based equations are, by far, the most popular way to describe collective cell migration (Sherratt and Murray, 1990; Painter and Sherratt, 2003; Maini et al., 2004).

We now present MFA equations for the cases relevant to both monoculture ($K = 1$) and co-culture ($K = 2$) experiments. First, for $K = 1$, substituting Equation (3.20) into Equation (3.10), we obtain

$$\frac{\partial p_1(x, t)}{\partial t} = D\Delta p_1(x, t) - (N - 1)\nabla(p_1(x, t)\, V(x, t)), \tag{3.21}$$

where

$$V(x, t) = \int_\Omega F(x - y)\, p_1(y, t)\, \mathrm{d}y, \tag{3.22}$$

is the velocity field induced by interactions between cells. Second, for $K = 2$,

the MFA leads to two coupled equations,

$$\frac{\partial p_1^1(x,t)}{\partial t} = D\Delta p_1^1(x,t) - (n_1-1)\nabla\Big(p_1^1(x,t)\,V^{11}(x,t)\Big) - n_2\nabla\Big(p_1^1(x,t)\,V^{12}(x,t)\Big), \tag{3.23}$$

$$\frac{\partial p_1^2(x,t)}{\partial t} = D\Delta p_1^2(x,t) - (n_2-1)\nabla\Big(p_1^2(x,t)\,V^{22}(x,t)\Big) - n_1\nabla\Big(p_1^2(x,t)\,V^{21}(x,t)\Big), \tag{3.24}$$

$$V^{lm}(x,t) = \int_\Omega F^{lm}(x-y)\,p_1^m(y,t)\,\mathrm{d}y, \tag{3.25}$$

where indices $l, m = 1, 2$.

3.1.2 Moment dynamics approximation

A more sophisticated approach is to use a closure relation to write for the three-particle correlation function in terms of the two-particle correlation function (Baker and Simpson, 2010; Middleton et al., 2014). A commonly-used closure relations is the Kirkwood superposition approximation (KSA) (Kirkwood, 1935), which can be written as

$$p_3^{lms}(x,y,z,t) = \frac{p_2^{lm}(x,y,t)\,p_2^{ls}(x,z,t)\,p_2^{ms}(y,z,t)}{p_1^l(x,t)\,p_1^m(y,t)\,p_1^s(z,t)}, \tag{3.26}$$

where the subpopulations l, m and s are not necessarily distinct.

For monoculture experiments with $K = 1$, the KSA continuum model can be

231 written as

$$\frac{\partial p_1(x,t)}{\partial t} = D\Delta p_1(x,t) - (N-1)\nabla\Big(\int_\Omega F(x-y)\,p_2(x,y,t)\,\mathrm{d}y\Big), \tag{3.27}$$

$$\begin{aligned}\frac{\partial p_2(x,y,t)}{\partial t} &= D\Delta p_2(x,y,t)\\ &\quad - \frac{\partial}{\partial x}\Big(F(x-y)\,p_2(x,y,t)\Big)\\ &\quad - \frac{\partial}{\partial y}\Big(F(y-x)\,p_2(x,y,t)\Big)\\ &\quad - (N-2)\,\frac{\partial}{\partial x}\int_\Omega F(x-z)\,\frac{p_2(x,y,t)\,p_2(x,z,t)\,p_2(y,z,t)}{p_1(x,t)\,p_1(y,t)\,p_1(z,t)}\,\mathrm{d}z\\ &\quad - (N-2)\,\frac{\partial}{\partial y}\int_\Omega F(y-z)\,\frac{p_2(x,y,t)\,p_2(x,z,t)\,p_2(y,z,t)}{p_1(x,t)\,p_1(y,t)\,p_1(z,t)}\,\mathrm{d}z.\end{aligned} \tag{3.28}$$

232 It is useful to note that there is more than one way to solve a problem with
233 $K = 1$ using the KSA framework. One approach would be to solve Equations
234 (3.27) and (3.28) simultaneously. However, it is more computationally efficient
235 to solve Equation (3.28) to give $p_2(x,y,t)$, and then to obtain $p_1(x,t)$ by
236 numerical integration

$$p_1(x,t) = \int_\Omega p_2(x,y,t)\,\mathrm{d}y. \tag{3.29}$$

237 For co-culture experiments with $K = 2$, the KSA continuum model can be
238 written as

$$\begin{aligned}\frac{\partial p_1^1(x,t)}{\partial t} &= D\Delta p_1^1(x,t) - (n_1-1)\nabla\Big(\int_\Omega F^{11}(x-y)\,p_2^{11}(x,y,t)\,\mathrm{d}y\Big)\\ &\quad - n_2\,\nabla\Big(\int_\Omega F^{12}(x-y)\,p_2^{12}(x,y,t)\,\mathrm{d}y\Big),\end{aligned} \tag{3.30}$$

$$\begin{aligned}\frac{\partial p_1^2(x,t)}{\partial t} &= D\Delta p_1^2(x,t) - (n_2-1)\nabla\Big(\int_\Omega F^{22}(x-y)\,p_2^{22}(x,y,t)\,\mathrm{d}y\Big)\\ &\quad - n_1\,\nabla\Big(\int_\Omega F^{21}(x-y)\,p_2^{21}(x,y,t)\,\mathrm{d}y\Big),\end{aligned} \tag{3.31}$$

$$
\begin{aligned}
\frac{\partial p_2^{11}(x,y,t)}{\partial t} &= D\Delta p_2^{11}(x,y,t) \\
&\quad - \frac{\partial}{\partial x}\Big(F^{11}(x-y)\, p_2^{11}(x,y,t)\Big) \\
&\quad - \frac{\partial}{\partial y}\Big(F^{11}(y-x)\, p_2^{11}(x,y,t)\Big) \\
&\quad - (n_1-2)\frac{\partial}{\partial x}\int_\Omega F^{11}(x-z)\,\frac{p_2^{11}(x,y,t)\, p_2^{11}(x,z,t)\, p_2^{11}(y,z,t)}{p_1^1(x,t)\, p_1^1(y,t)\, p_1^1(z,t)}\,\mathrm{d}z \\
&\quad - (n_1-2)\frac{\partial}{\partial y}\int_\Omega F^{11}(y-z)\,\frac{p_2^{11}(x,y,t)\, p_2^{11}(x,z,t)\, p_2^{11}(y,z,t)}{p_1^1(x,t)\, p_1^1(y,t)\, p_1^1(z,t)}\,\mathrm{d}z \\
&\quad - n_2\frac{\partial}{\partial x}\int_\Omega F^{12}(x-z)\,\frac{p_2^{11}(x,y,t)\, p_2^{12}(x,z,t)\, p_2^{12}(y,z,t)}{p_1^1(x,t)\, p_1^1(y,t)\, p_1^2(z,t)}\,\mathrm{d}z \\
&\quad - n_2\frac{\partial}{\partial y}\int_\Omega F^{12}(y-z)\,\frac{p_2^{11}(x,y,t)\, p_2^{12}(x,z,t)\, p_2^{12}(y,z,t)}{p_1^1(x,t)\, p_1^1(y,t)\, p_1^2(z,t)}\,\mathrm{d}z,
\end{aligned}
\tag{3.32}
$$

$$
\begin{aligned}
\frac{\partial p_2^{22}(x,y,t)}{\partial t} &= D\Delta p_2^{22}(x,y,t) \\
&\quad - \frac{\partial}{\partial x}\Big(F^{22}(x-y)\, p_2^{22}(x,y,t)\Big) \\
&\quad - \frac{\partial}{\partial y}\Big(F^{22}(y-x)\, p_2^{22}(x,y,t)\Big) \\
&\quad - (n_2-2)\frac{\partial}{\partial x}\int_\Omega F^{22}(x-z)\,\frac{p_2^{22}(x,y,t)\, p_2^{22}(x,z,t)\, p_2^{22}(y,z,t)}{p_1^1(x,t)\, p_1^1(y,t)\, p_1^1(z,t)}\,\mathrm{d}z \\
&\quad - (n_2-2)\frac{\partial}{\partial y}\int_\Omega F^{22}(y-z)\,\frac{p_2^{22}(x,y,t)\, p_2^{22}(x,z,t)\, p_2^{22}(y,z,t)}{p_1^1(x,t)\, p_1^1(y,t)\, p_1^1(z,t)}\,\mathrm{d}z \\
&\quad - n_1\frac{\partial}{\partial x}\int_\Omega F^{21}(x-z)\,\frac{p_2^{22}(x,y,t)\, p_2^{21}(x,z,t)\, p_2^{21}(y,z,t)}{p_1^1(x,t)\, p_1^1(y,t)\, p_1^2(z,t)}\,\mathrm{d}z \\
&\quad - n_1\frac{\partial}{\partial y}\int_\Omega F^{21}(y-z)\,\frac{p_2^{22}(x,y,t)\, p_2^{21}(x,z,t)\, p_2^{21}(y,z,t)}{p_1^1(x,t)\, p_1^1(y,t)\, p_1^2(z,t)}\,\mathrm{d}z,
\end{aligned}
\tag{3.33}
$$

$$
\begin{aligned}
\frac{\partial p_2^{12}(x,y,t)}{\partial t} &= D\Delta p_2^{12}(x,y,t) \\
&\quad - \frac{\partial}{\partial x}\Big(F^{12}(x-y)\, p_2^{12}(x,y,t)\Big) \\
&\quad - \frac{\partial}{\partial y}\Big(F^{12}(y-x)\, p_2^{12}(x,y,t)\Big) \\
&\quad - (n_2-1)\frac{\partial}{\partial x}\int_\Omega F^{12}(x-z)\,\frac{p_2^{12}(x,y,t)\, p_2^{12}(x,z,t)\, p_2^{22}(y,z,t)}{p_1^1(x,t)\, p_1^2(y,t)\, p_1^2(z,t)}\,\mathrm{d}z \\
&\quad - (n_1-1)\frac{\partial}{\partial x}\int_\Omega F^{11}(x-z)\,\frac{p_2^{12}(x,y,t)\, p_2^{11}(x,z,t)\, p_2^{21}(y,z,t)}{p_1^1(x,t)\, p_1^2(y,t)\, p_1^1(z,t)}\,\mathrm{d}z \\
&\quad - (n_1-1)\frac{\partial}{\partial y}\int_\Omega F^{21}(y-z)\,\frac{p_2^{12}(x,y,t)\, p_2^{11}(x,z,t)\, p_2^{21}(y,z,t)}{p_1^1(x,t)\, p_1^2(y,t)\, p_1^1(z,t)}\,\mathrm{d}z \\
&\quad - (n_2-1)\frac{\partial}{\partial y}\int_\Omega F^{22}(y-z)\,\frac{p_2^{12}(x,y,t)\, p_2^{12}(x,z,t)\, p_2^{22}(y,z,t)}{p_1^1(x,t)\, p_1^2(y,t)\, p_1^2(z,t)}\,\mathrm{d}z,
\end{aligned}
\tag{3.34}
$$

$$
\begin{aligned}
\frac{\partial p_2^{21}(x,y,t)}{\partial t} &= D\Delta p_2^{21}(x,y,t) \\
&\quad - \frac{\partial}{\partial x}\Big(F^{21}(x-y)\, p_2^{21}(x,y,t)\Big) \\
&\quad - \frac{\partial}{\partial y}\Big(F^{21}(y-x)\, p_2^{21}(x,y,t)\Big) \\
&\quad - (n_1-1)\frac{\partial}{\partial y}\int_\Omega F^{21}(y-z)\,\frac{p_2^{21}(x,y,t)\, p_2^{21}(x,z,t)\, p_2^{11}(y,z,t)}{p_1^2(x,t)\, p_1^1(y,t)\, p_1^1(z,t)}\,\mathrm{d}z \\
&\quad - (n_2-1)\frac{\partial}{\partial y}\int_\Omega F^{22}(y-z)\,\frac{p_2^{21}(x,y,t)\, p_2^{22}(x,z,t)\, p_2^{12}(y,z,t)}{p_1^2(x,t)\, p_1^1(y,t)\, p_1^2(z,t)}\,\mathrm{d}z \\
&\quad - (n_2-1)\frac{\partial}{\partial x}\int_\Omega F^{12}(x-z)\,\frac{p_2^{21}(x,y,t)\, p_2^{22}(x,z,t)\, p_2^{12}(y,z,t)}{p_1^2(x,t)\, p_1^1(y,t)\, p_1^2(z,t)}\,\mathrm{d}z \\
&\quad - (n_1-1)\frac{\partial}{\partial x}\int_\Omega F^{11}(x-z)\,\frac{p_2^{21}(x,y,t)\, p_2^{21}(x,z,t)\, p_2^{11}(y,z,t)}{p_1^2(x,t)\, p_1^1(y,t)\, p_1^1(z,t)}\,\mathrm{d}z. \qquad (3.35)
\end{aligned}
$$

Again, there are multiple strategies for solving the KSA equations when $K = 2$. Here, we solve Equations (3.32) and (3.33) to give $p_2^{11}(x,y,t)$ and $p_2^{22}(x,y,t)$, respectively. Using these results we calculate $p_1^1(x,t)$ and $p_1^2(x,t)$ by numerical integration, similar to Equation (3.29). To obtain $p_2^{12}(x,y,t)$ and $p_2^{21}(x,y,t)$, we use $p_2^{12}(x,y,t) = p_1^1(x,t)p_1^2(y,t)$ and $p_2^{21}(x,y,t) = p_1^2(x,t)p_1^1(y,t)$, respectively.

Now that we have documented both the MFA and KSA continuum approximations for both single species monoculture ($K = 1$) and two-species co-culture ($K = 2$) experiments, we will now solve these governing equations for both cases and compare results with averaged data from discrete simulations.

3.2 Application to monoculture experiments, $K = 1$

We first consider the situation where we have one population of cells, $K = 1$. In all of our numerical results we always fix the diffusivity to be $D = 300\ \mu\text{m}^2\text{h}^{-1}$ (Treloar et al. 2013). To emphasize the importance of non mean-field effects, all simulation results in the main paper involve strong adhesion, where f_0 is sufficiently large. This situation is relevant when we apply our model to mimic the collective migration of epithelial cells (Treloar et al., 2013). In contrast, if the models are applied to deal with the collective migration of mesenchymal cells, without significant adhesion (Simpson et al., 2013), then additional results in the Supplementary Material document with reduced f_0 are more relevant.

Since we consider unbiased random forces, we sample ξ_i from a Gaussian distribution with zero mean and zero auto-correlation

$$\left\langle \xi_i(t)\xi_j(t') \right\rangle = \frac{2D}{\Delta t}\delta_{ij}\delta_{tt'}, \tag{3.36}$$

which is a white noise limit (Supplementary Material). The variance of ξ_i is given by

$$\text{Var}(\xi_i) = \frac{2D}{\Delta t}, \tag{3.37}$$

where Δt is the duration of the time step used in the discrete simulations.

The initial distribution of cells in the monoculture simulations is given by

$$\alpha(x) = \begin{cases} 0, & 0\ \mu\text{m} \leq x < 600\ \mu\text{m}, \\ 25 \times 10^{-3}, & 600\ \mu\text{m} \leq x \leq 1400\ \mu\text{m}, \\ 0, & 1400\ \mu\text{m} < x \leq 2000\ \mu\text{m}, \end{cases} \tag{3.38}$$

on $0 \leq x \leq 2000\ \mu$m, which is a typical length scale for an *in vitro* cell migration experiment (Jin et al. 2016). Here, $\alpha(x)$ is a function of position, and we sample from this function to define the initial distribution of cells in the

discrete simulations. This initial distribution corresponds to a confined group of cells in the centre of the domain. When presenting results from simulations we refer to both the dimensional density of cells, $p_1(x,t)$ [cells/μm], as well as the non-dimensional density of cells relative to the carrying capacity density, $p_1(x,t)/C$, where C is the carrying capacity density that is given by $C = N\delta/L$, where N is the maximum number of cells of diameter δ that can be distributed along a domain of length L without compression. Periodic boundary conditions are imposed for all simulations.

To solve the MFA model, we set $p_1(x,0) = \alpha(x)$, and to solve the KSA model, we note that since cells are randomly placed according to Equation (3.38), there are no spatial correlations in the initial positions of the cells. Therefore, the initial conditions for the KSA model are given by $p_1(x,0) = \alpha(x)$ and $p_2(x,y,0) = \alpha(x)\alpha(y)$. With this information, Equations (3.21) and (3.28) are solved using the method of lines with spatial and temporal discretisations chosen to be sufficiently fine that the numerical solutions are grid independent. The discrete model, Equation (2.3), is numerically integrated using a fourth order Runge-Kutta (RK4) method (Press et al., 2007) and density distributions are obtained by considering a large number of identically prepared simulations. Results in Figure 2 compare numerical solutions of the MFA and KSA continuum descriptions with averaged data from discrete simulations. Snapshots of the discrete simulations are shown in Figure 2(a)-(b). A comparison of the ensemble averaged data and the solution of the MFA and KSA models are given in Figure 2(c) and Figure 2(e), respectively. To clearly compare the performance of the MFA and KSA models near the position of the spreading profile, we show a magnified region of the profiles in Figures 2(d) and Figure 2(f).

In summary, we see that both the KSA and MFA models capture the overall spreading behaviour of the collective migration reasonably well, as shown in Figure 2(c) and Figure 2(e). However, when we examine the performance of MFA model more closely, as illustrated in Figure 2(d), we see that the solution of the MFA continuum model is not as steep as the discrete density data. In contrast, the performance of the KSA model, as shown in Figure 2(f), provides an improved match to the averaged discrete data. We now examine the relative performance of the MFA and KSA approaches for two multi-species problems.

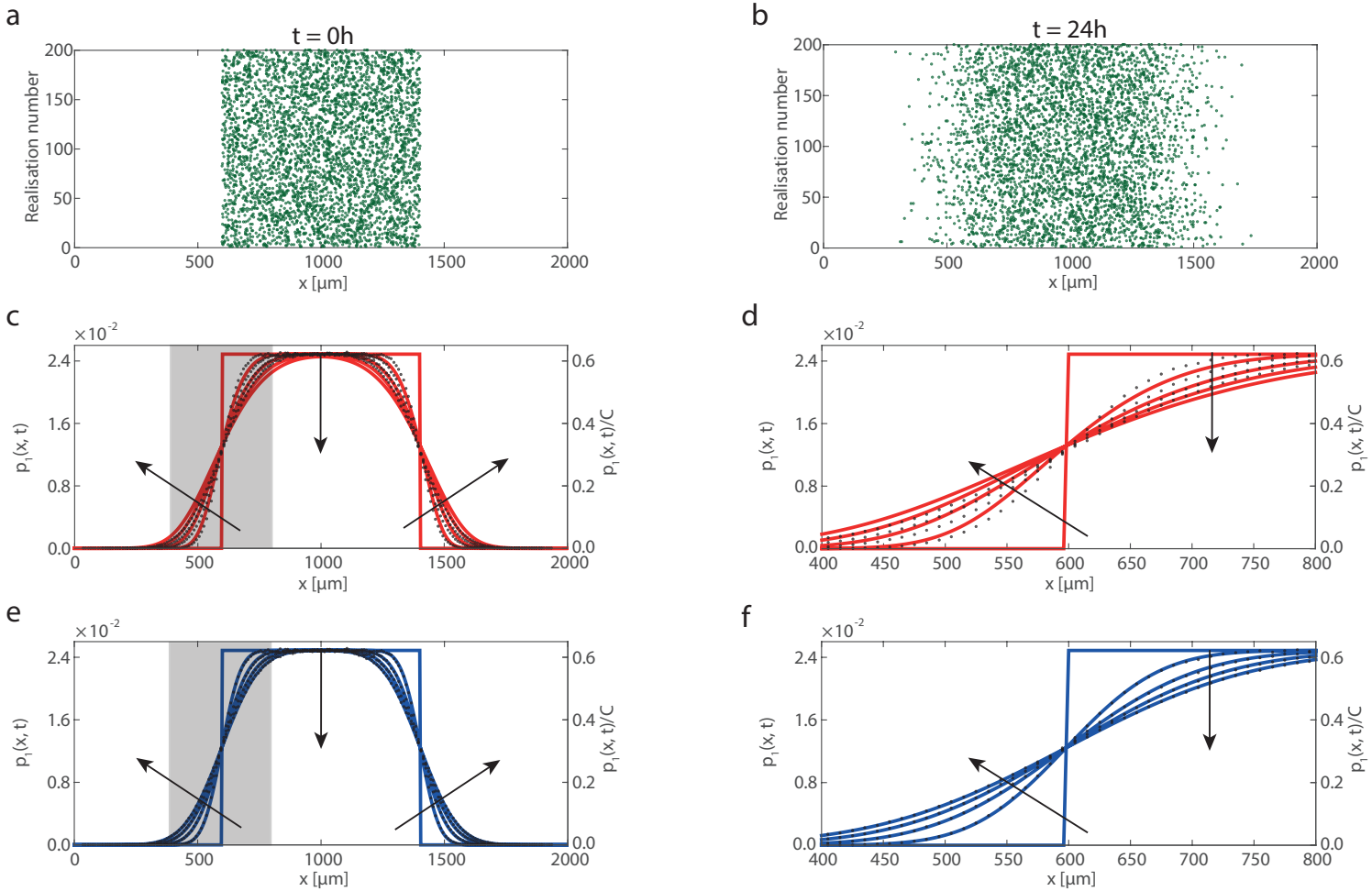

Fig. 2. Comparison of ensemble averages of stochastic simulations and solutions of the MFA and KSA continuum models for a single-species population of cells on a one-dimensional domain with $0 \leq x \leq 2000$ μm. Snapshots in (a)-(b) show 200 realisations of the discrete model at $t = 0$ and $t = 24$ hours, respectively. The population of cells (green) initially occupies the central region, which is 800 μm wide, and has an initial density 25×10^{-3} cells/μm. Results in (c)-(f) show the cell density profiles obtained using an ensemble of 5×10^5 simulations (black dots) with binsize of 10 μm. These results are compared to solutions of the MFA model, Equation (3.21) (red lines), and the KSA model, Equation (3.28) (blue lines). Profiles are given at $t = 0, 6, 12, 18$, and 24 h, with the arrows indicating the direction of increasing t. In (c)-(f) the cell density is reported in terms of the dimensional cell density, $p_1(x, t)$, as well as the dimensionless cell density, $p_1(x, t)/C$, where $C = 40 \times 10^{-3}$ cells/μm. Equation (2.3) is integrated with $\Delta t = 5 \times 10^{-2}$ h, Equation (3.21) is integrated with $\Delta x = 4$ μm and $\Delta t = 10^{-2}$ h, and Equation (3.28) is integrated with $\Delta x = \Delta y = 4$ μm and $\Delta t = 10^{-2}$ h. The remaining parameters are $N = 20$, $a = 0.08$ μm^{-1}, $\delta = 25$ μm, $f_0 = 0.2$ μm/h.

3.3 Application to co-culture experiments, $K = 2$

We now consider the evolution of two types of two-species problems. These two problems involve different experimental designs. In both cases we choose the size of the cells in subpopulations 1 and 2 to be different. Here, the diameter of cells in the first subpopulation is $\delta_1 = 18$ μm, and the diameter of cells in the second subpopulation is $\delta_2 = 25$ μm. We also introduce differing interspecies interaction parameters such as the interspecies force amplitude, f_0^{12}, shape parameter, a_{12}, and the interspecies diameter, δ_{12}, which corresponds to the average radius of the different cell types.

The first experiment involves one population of cells spreading through another background population of cells, and this mimics the way that an initially confined population of tumour cells might spread through surrounding healthy tissue (Eves et al. 2003). To specify the initial condition for this problem we must describe the initial location of both subpopulations,

$$\alpha_1(x) = \begin{cases} 0, & 0 \ \mu\text{m} \leq x < 600 \ \mu\text{m}, \\ 25 \times 10^{-3}, & 600 \ \mu\text{m} \leq x \leq 1400 \ \mu\text{m}, \\ 0, & 1400 \ \mu\text{m} < x \leq 2000 \ \mu\text{m}, \end{cases} \tag{3.39}$$

$$\alpha_2(x) = \begin{cases} 10.8 \times 10^{-3}, & 0 \ \mu\text{m} \leq x < 600 \ \mu\text{m}, \\ 0, & 600 \ \mu\text{m} \leq x \leq 1400 \ \mu\text{m}, \\ 10.8 \times 10^{-3}, & 1400 \ \mu\text{m} < x \leq 2000 \ \mu\text{m}, \end{cases} \tag{3.40}$$

where $\alpha_1(x)$ is a function of position that describes the initial location of cells from the first subpopulation, and $\alpha_2(x)$ is a function of position that describes the initial location of cells from the second subpopulation. This initial condition corresponds to the situation where the region $600 \leq x \leq 1400$ μm is relatively densely occupied by subpopulation 1, and the remaining space is less densely populated by subpopulation 2. To initialise the discrete simulations we sample from $\alpha_1(x)$ and $\alpha_2(x)$, and snapshots showing 200 realisations of discrete model are given in Figure 3(a)-(c) at $t = 0$, 12 and 24 hours, showing how the two subpopulations mix.

The second experiment that we consider corresponds to two initially adjacent subpopulations of cells. The initial location of both subpopulations is given

by

$$\alpha_1(x) = \begin{cases} 0, & 0\ \mu\text{m} \leq x < 600\ \mu\text{m}, \\ 25 \times 10^{-3}, & 600\ \mu\text{m} \leq x < 1000\ \mu\text{m}, \\ 0, & 1000\ \mu\text{m} < x \leq 2000\ \mu\text{m}, \end{cases} \tag{3.41}$$

$$\alpha_2(x) = \begin{cases} 0, & 0\ \mu\text{m} \leq x < 1000\ \mu\text{m}, \\ 25 \times 10^{-3}, & 1000\ \mu\text{m} \leq x \leq 1400\ \mu\text{m}, \\ 0, & 1400\ \mu\text{m} < x \leq 2000\ \mu\text{m}. \end{cases} \tag{3.42}$$

To initialise the discrete simulations we sample from $\alpha_1(x)$ and $\alpha_2(x)$, and snapshots showing 200 realisations of discrete model for the second initial condition are given in Figure 5(a)-(c) at $t = 0$, 12 and 24 hours. Here we see that the two subpopulations mix near $x = 1000$ μm. Furthermore, we also see that both subpopulations spread into the initially vacant surrounding regions.

To obtain continuum results for the two-species problems, the MFA and KSA continuum models, given by Equations (3.23)-(3.24) and Equations (3.32)-(3.33), respectively, are solved using the method of lines with spatial and temporal discretisations chosen to be sufficiently fine that the numerical solutions are grid independent (Supplementary Material). Results in Figures 3 and 5 compare the performance of the MFA approach with the averaged discrete data. Since these simulations involve significant interaction forces, we see that the solution of the MFA model does not always accurately capture the details of how the subpopulations spread and interact with each other. Results in Figures 4 and 6 compare the performance of the KSA approach with the averaged discrete data. Comparing the performance of the KSA and MFA models confirms that, similar to our results for the single-species problem in Figure 2, the KSA approach outperforms the MFA model.

3.4 Parameter sensitivity

In this section we investigate how the accuracy of the both continuum approximations depends on the choice of the model parameters. To explore this question we re-examine the results of the first co-culture experiment, as illustrated in Figures 3–4, and we quantify how the accuracy of the KSA and MFA continuum models depends on the strength of adhesion and the ratio

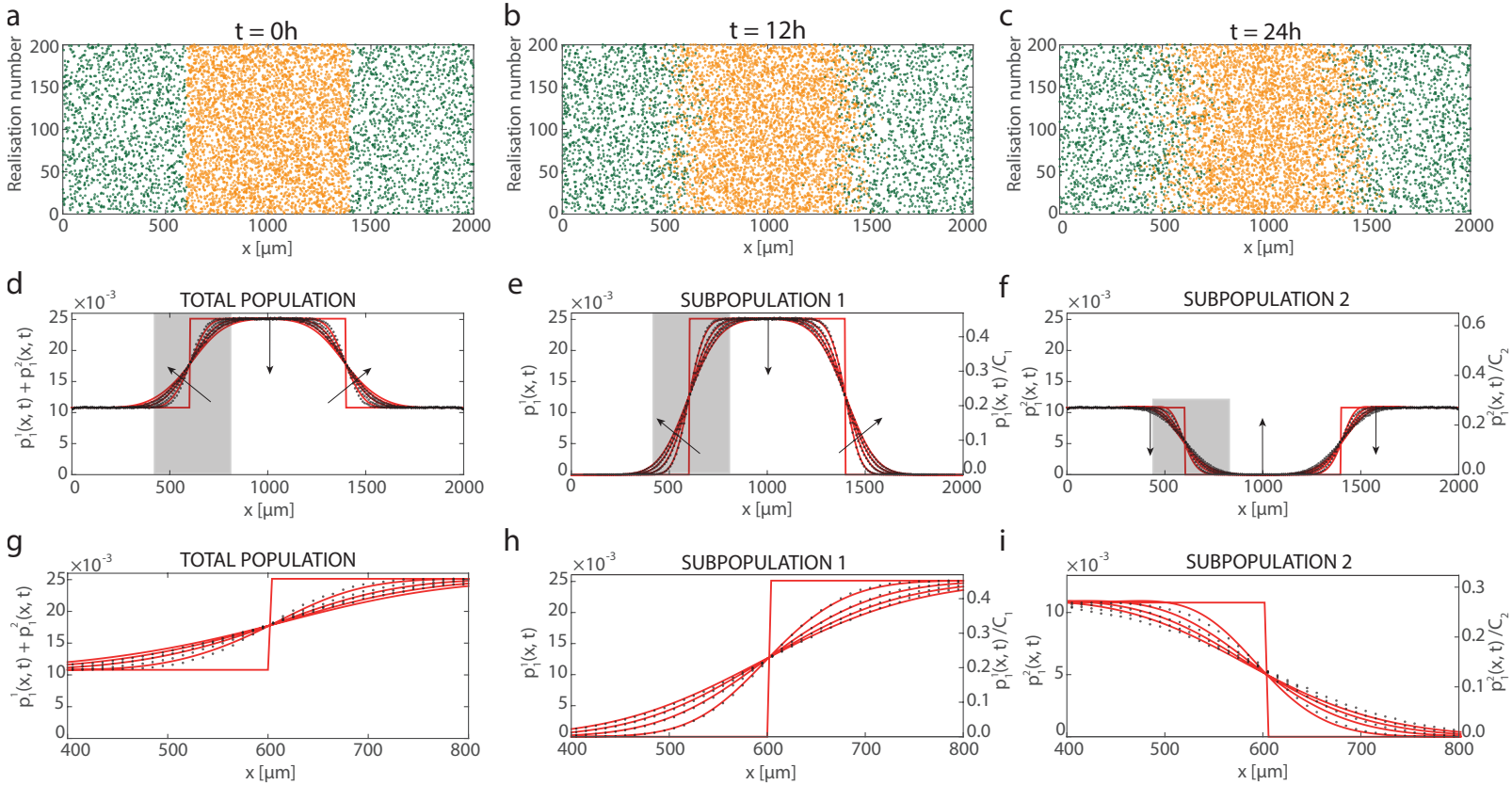

Fig. 3. Comparison of ensemble averages of stochastic simulations and solutions of the MFA continuum model, given by Equations (3.23)-(3.24), for a two-species population of cells on a one-dimensional domain with $0 \leq x \leq 2000$ μm. Snapshots in (a)-(c) show 200 realisations of the discrete model at $t = 0, 12$, and 24 hours, respectively. Subpopulation 1 (orange) initially occupies the central region at a density of 25×10^{-3} cells/μm, and subpopulation 2 (green) initially occupies two outer regions at a density of 10.8×10^{-3} cells/μm. Results in (d)-(i) show the density profiles obtained using an ensemble of 5×10^5 simulations (black dots) with binsize of 10 μm, and the solutions of the MFA model (red lines) at $t = 0, 6, 12, 18$ and 24 h, with the arrows indicating the direction of increasing t. Results in (d)-(i) are shown in terms of the total population density, the density of subpopulation 1, and the density of subpopulation 2, as indicated. Density profiles are reported in terms of the dimensional cell densities, $p_1^1(x,t)$ and $p_1^2(x,t)$, as well as the dimensionless cell densities, $p_1^1(x,t)/C_1$ and $p_1^2(x,t)/C_2$, where $C_1 = 55.5 \times 10^{-3}$ cells/μm, $C_2 = 40 \times 10^{-3}$ cells/μm. The MFA model is integrated with $\Delta x = 4$ μm and $\Delta t = 5 \times 10^{-3}$ h. The remaining parameters are $n_1 = 20$, $n_2 = 13$, $a_1 = 0.08$ μm^{-1}, $a_2 = 0.06$ μm^{-1}, $a_{12} = 0.07$ μm^{-1}, $\delta_1 = 18$ μm, $\delta_2 = 25$ μm, $\delta_{12} = 21.5$ μm, $f_0^{11} = 1.5$ μm/h, $f_0^{22} = 1$ μm/h, $f_0^{12} = 1.25$ μm/h.

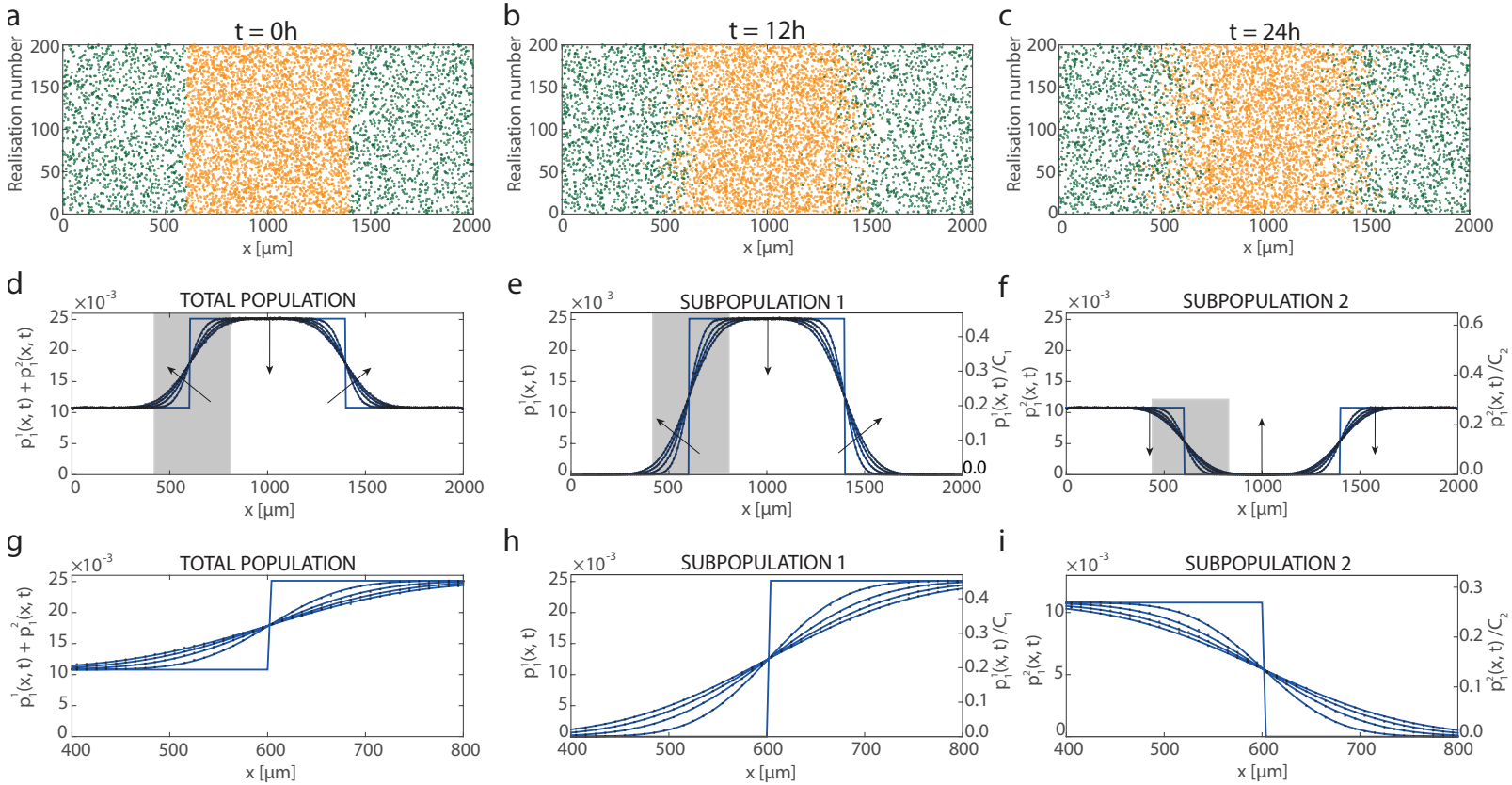

Fig. 4. Comparison of ensemble averages of stochastic simulations and solutions of the KSA continuum model, given by Equations (3.32)-(3.33), for a two-species population of cells on a one-dimensional domain with $0 \leq x \leq 2000$ μm. Snapshots in (a)-(c) show 200 realisations of the discrete model at $t = 0, 12$, and 24 hours, respectively. Subpopulation 1 (orange) initially occupies the central region at a density of 25×10^{-3} cells/μm, and subpopultation 2 (green) initially occupies two outer regions at a density of 10.8×10^{-3} cells/μm. Results in (d)-(i) show the density profiles obtained using an ensemble of 5×10^5 simulations (black dots) with binsize of 10 μm, and the solutions of the KSA model (red lines) at $t = 0, 6, 12, 18$ and 24 h, with the arrows indicating the direction of increasing t. Results in (d)-(i) are shown in terms of the total population density, the density of subpopulation 1 and the density of subpopulation 2, as indicated. Density profiles are reported in terms of the dimensional cell densities, $p_1^1(x,t)$ and $p_1^2(x,t)$, as well as the dimensionless cell densities, $p_1^1(x,t)/C_1$ and $p_1^2(x,t)/C_2$, where $C_1 = 55.5 \times 10^{-3}$ cells/μm, $C_2 = 40 \times 10^{-3}$ cells/μm. The KSA model is integrated with $\Delta x = \Delta y = 4$ μm and $\Delta t = 5 \times 10^{-3}$ h. The remaining parameters are $n_1 = 20$, $n_2 = 13$, $a_1 = 0.08$ μm^{-1}, $a_2 = 0.06$ μm^{-1}, $a_{12} = 0.07$ μm^{-1}, $\delta_1 = 18$ μm, $\delta_2 = 25$ μm, $\delta_{12} = 21.5$ μm, $f_0^{11} = 1.5$ μm/h, $f_0^{22} = 1$ μm/h, $f_0^{12} = 1.25$ μm/h.

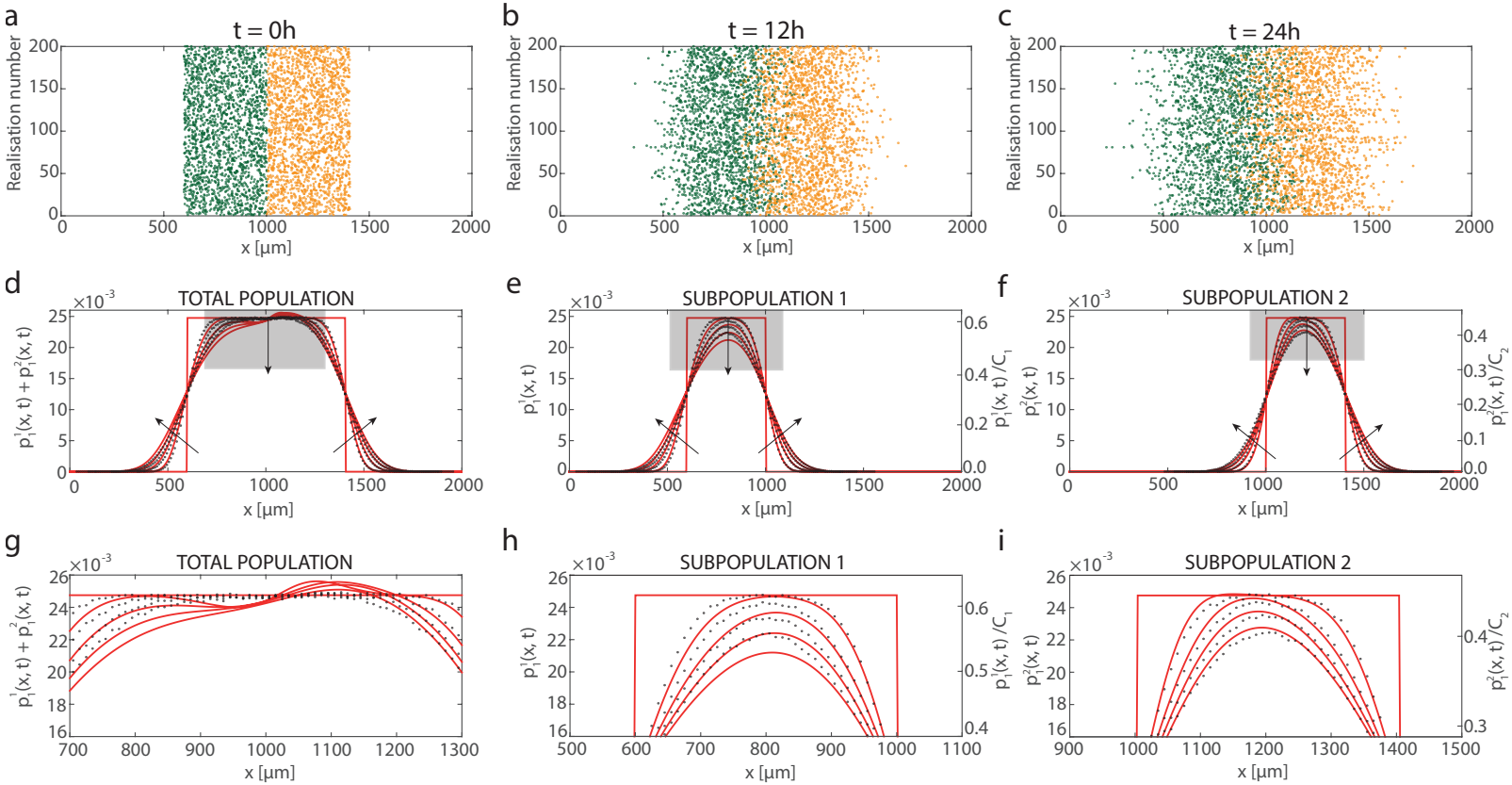

Fig. 5. Comparison of ensemble averages of stochastic simulations and solution of the MFA continuum model, given by Equations (3.23)-(3.24), for a two-species population of cells on a one-dimensional domain with $0 \leq x \leq 2000$ μm. Snapshots in (a)-(c) show 200 realisations of the discrete model at $t = 0, 12$, and 24 hours, respectively. Subpopulations 1 (green) and 2 (orange) initially occupy adjacent regions at a density of 25×10^{-3} cells/μm. Results in (d)-(i) show the density profiles obtained using an ensemble of 5×10^5 simulations (black dots) with binsize of 10 μm, and the solutions of the MFA model (red lines) at $t = 0, 6, 12, 18$ and 24 h, with the arrows indicating the direction of increasing t. Results in (d)-(i) are shown in terms of the total population density, the density of subpopulation 1 and the density of subpopulation 2, as indicated. Density profiles are reported in terms of the dimensional cell densities, $p_1^1(x,t)$ and $p_1^2(x,t)$, as well as the dimensionless cell densities, $p_1^1(x,t)/C_1$ and $p_1^2(x,t)/C_2$, where $C_1 = 40 \times 10^{-3}$ cells/μm, $C_2 = 55.5 \times 10^{-3}$ cells/μm. The MFA model is integrated with $\Delta x = 4$ μm and $\Delta t = 5 \times 10^{-3}$ h. The remaining parameters are $n_1 = 10$, $n_2 = 10$, $a_1 = 0.06$ μm^{-1}, $a_2 = 0.08$ μm^{-1}, $a_{12} = 0.07$ μm^{-1}, $\delta_1 = 25$ μm, $\delta_2 = 18$ μm, $\delta_{12} = 21.5$ μm, $f_0^{11} = 1.5$ μm/h, $f_0^{22} = 2$ μm/h, $f_0^{12} = 1.75$ μm/h.

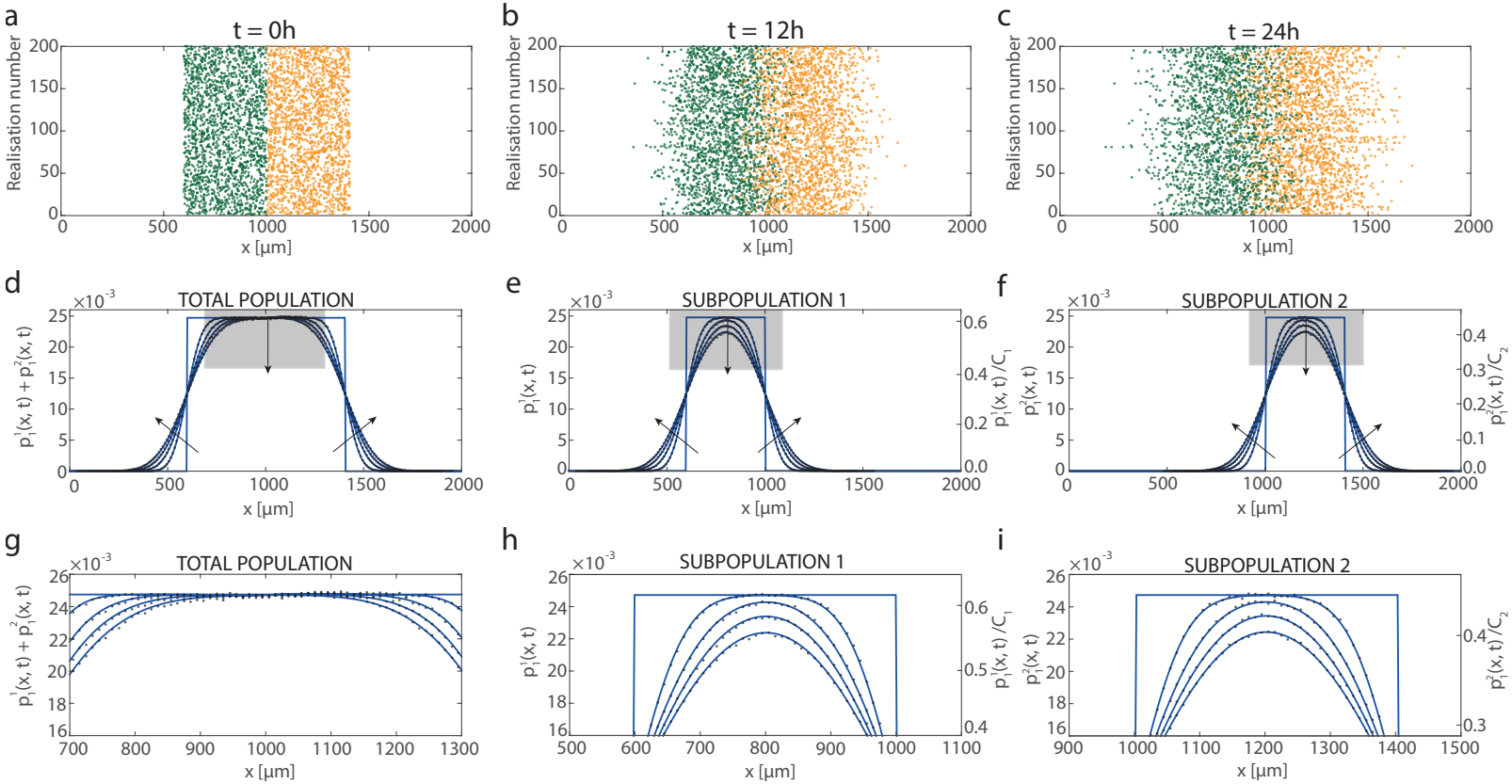

Fig. 6. Comparison of ensemble averages of stochastic simulations and solutions of the KSA continuum model, given by Equations (3.32)-(3.33), for a two-species population of cells on a one-dimensional domain with $0 \leq x \leq 2000$ μm. Snapshots in (a)-(c) show 200 realisations of the discrete model at $t = 0, 12$, and 24 hours, respectively. Subpopulations 1 (green) and 2 (orange) initially occupy adjacent regions at a density of 25×10^{-3} cells/μm. Results in (d)-(i) show the density profiles obtained using an ensemble of 5×10^5 simulations (black dots) with binsize of 10 μm, and the solutions of the KSA Equations (blue lines) at $t = 0, 6, 12, 18$ and 24 h, with the arrows indicating the direction of increasing t. Results in (d)-(i) are shown in terms of the total population density, the density of subpopulation 1 and the density of subpopulation 2, as indicated. Density profiles are reported in terms of the dimensional cell densities, $p_1^1(x,t)$ and $p_1^2(x,t)$, as well as the dimensionless cell densities, $p_1^1(x,t)/C_1$ and $p_1^2(x,t)/C_2$, where $C_1 = 40 \times 10^{-3}$ cells/μm, $C_2 = 55.5 \times 10^{-3}$ cells/μm. The KSA model is integrated with $\Delta x = \Delta y = 4$ μm and $\Delta t = 5 \times 10^{-3}$ h. The remaining parameters are $n_1 = 10$, $n_2 = 10$, $a_1 = 0.06$ μm^{-1}, $a_2 = 0.08$ μm^{-1}, $a_{12} = 0.07$ μm^{-1}, $\delta_1 = 25$ μm, $\delta_2 = 18$ μm, $\delta_{12} = 21.5$ μm, $f_0^{11} = 1.5$ μm/h, $f_0^{22} = 2$ μm/h, $f_0^{12} = 1.75$ μm/h.

of the two cell sizes in the co-culture experiment. To explore this we repeat the discrete simulations and vary the force amplitude f_0^{11}, which determines strength of the cell-to-cell adhesion, as well as varying the ratio δ_1/δ_2. To keep our analysis as straightforward as possible, we vary these two quantities separately.

To quantify the accuracy of both the MFA and KSA continuum approximations we define the following quantities,

$$\mathrm{E}_{\mathrm{MFA}}(t) = \frac{1}{I}\sum_{i=1}^{I}[S_{\mathrm{MFA}}(i,t) - S_{\mathrm{discrete}}(i,t)]^2, \tag{3.43}$$

$$\mathrm{E}_{\mathrm{KSA}}(t) = \frac{1}{I}\sum_{i=1}^{I}[S_{\mathrm{KSA}}(i,t) - S_{\mathrm{discrete}}(i,t)]^2, \tag{3.44}$$

where $\mathrm{E}_{\mathrm{MFA}}(t)$ and $\mathrm{E}_{\mathrm{KSA}}(t)$ indicate mean squared error associated with the MFA and KSA approximations, respectively. The index i denotes the spatial node, and $I = 200$ is the total number of spatial nodes across the domain. To construct these mean squared errors we compare the total density profiles so that $S_{\mathrm{MFA}}(i,t) = p_1^1(i,t) + p_1^2(i,t)$ is the total population density predicted by the MFA continuum approximation, $S_{\mathrm{KSA}}(i,t) = p_1^1(i,t) + p_1^2(i,t)$ is the total population density predicted by the KSA continuum approximation, and $S_{\mathrm{discrete}}(i,t) = p_1^1(i,t) + p_1^2(i,t)$ is the total population density obtained by considering an ensemble average of the discrete model.

Results in Figure 7 show $\mathrm{E}_{\mathrm{MFA}}(t)$ and $\mathrm{E}_{\mathrm{KSA}}(t)$ as a function of δ_1/δ_2 and f_0^{11}. The vertical lines correspond to choices of δ_1/δ_2 and f_0^{11} that are identical to the parameter values used to construct the results in Figures 3-4. Overall, the results in Figure 7 show three main trends: (i) for all parameter choices considered in the sensitivity analysis, the KSA approximation outperforms the MFA approximation; (ii) the accuracy of both the MFA and KSA approximations decrease with both δ_1/δ_2 and f_0^{11}; and (iii) the accuracy of both the MFA and KSA approximations is more sensitive to changes in δ_1/δ_2 than changes in f_0^{11} for the range of parameters considered.

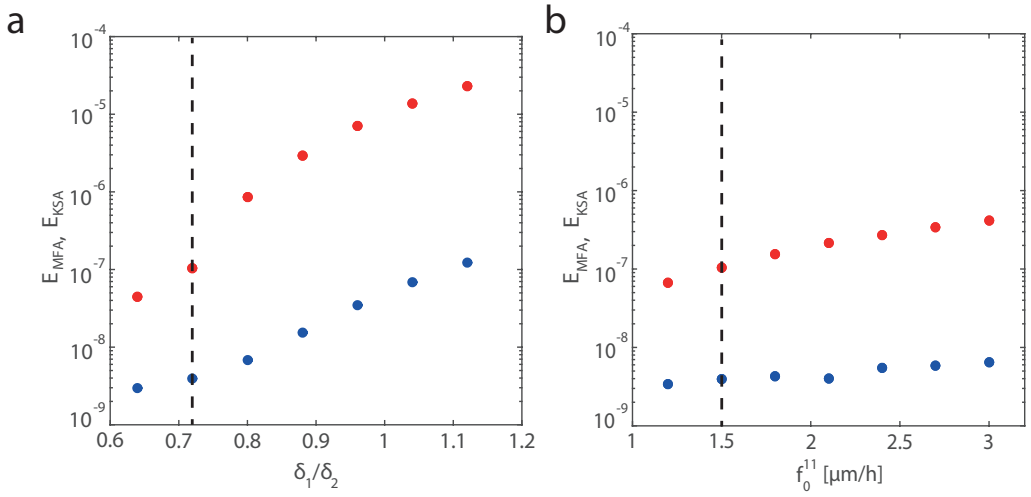

Fig. 7. (a) Comparison of the accuracy of the MFA and KSA continuum approximations as a function of δ_1/δ_2 at time $t = 24\,\mathrm{h}$ for the first co-culture experiment. All results in (a) correspond to $\delta_2 = 25\,\mu\mathrm{m}$, and the ratio δ_1/δ_2 is varied by altering δ_1. (b) Comparison of the accuracy of the MFA and KSA continuum approximations as a function of f_0^{11} at time $t = 24\,\mathrm{h}$ for the first co-culture experiment. All data in (b) correspond to a fixed choice of $f_0^{22} = 1.0\,\mu\mathrm{m/h}$. Both subfigures show $\mathrm{E_{MFA}}(t)$ (red dots) and $\mathrm{E_{KSA}}(t)$ (blue dots), and the dashed vertical line indicates the parameter values presented previously in Figures 3-4. All continuum models are solved numerically with $\Delta x = 4\,\mu\mathrm{m}$ and $\Delta t = 5 \times 10^{-3}\,\mathrm{h}$.

4 Conclusions

In this work, we develop a discrete multi-species model of collective cell migration. Our framework is very general, and can deal with genuine multi-species problems where the subpopulations are distinct (Eves et al., 2003), as well as other types of experiments where an otherwise identical subpopulation of cells is labelled (Simpson et al., 2007). Our discrete modelling framework can include various effects such as: random unbiased stochastic motion of individual cells; short range finite size effects to account for crowding interactions; longer range adhesive forces; as well as dealing with subpopulations of cells that have different cell diameters.

To analyse the discrete model, we derive a hierarchy of continuum moment equations to describe the spatial dynamics of agents, pairs of agents, triplets of agents, and so forth. We then develop two different approximate solutions of the hierarchy of moment equations. Firstly, using the MFA, and secondly,

using the KSA. We compare both continuum approximations with ensemble averages from discrete simulations.

Overall, both continuum approximations match the broad features of the discrete results reasonably well. When there is little or no adhesion, both continuum models match the averaged discrete results extremely well. However, once the adhesive force is sufficiently strong, the KSA continuum model matches the averaged discrete results much better than MFA model. This difference is the consequence of adhesion causing correlations in the positions of agents in the discrete simulations (Baker and Simpson, 2010). These effects are neglected in the MFA model, however the KSA model explicitly includes the effects of pairwise correlations, $p_2(x, y, t)$.

There are many potential extensions which we leave for future analysis. All our analysis has been in one dimension, but many biological experiments are in two or three dimensions (Treloar et al., 2013; Eves et al., 2003). It is relatively straightforward to apply our continuum models to higher dimensional problems, however we choose to take the most fundamental approach here and focus on one dimension only. As it stands, isolated individual cells in our discrete model move due to unbiased random motion. However, in many applications cells move with a bias, such as in chemotaxis (Keller and Segel, 1971). To extend our model to deal with chemotaxis we would need to introduce an evolution equation for some kind of nutrient, and to allow individual cells to move with some bias in response to the spatial gradient of the nutrient (Keller and Segel, 1971). We also note that all non-MFA results are obtained by approximately closing the system of continuum equations using the KSA, however other kinds of closure relations could also be used (Murrell et al., 2004; Frasca and Sharkey, 2016).

Acknowledgements

This work is supported by the Australian Research Council (FT130100148, DP170100474). Computational resources are provided by the High Performance Computing and Research Support Group at QUT. We appreciate the helpful comments and suggestions from the referee.

References

[1] Baker RE, Simpson MJ (2010). Correcting mean-field approximations for birth-death-movement processes. Physical Review E. 82: 041905.

[2] Binder BJ, Simpson MJ (2016). Cell density and cell size dynamics during *in vitro* tissue growth experiments: Implications for mathematical models of collective cell behaviour. Applied Mathematical Modelling. 40: 3438-3446.

[3] Eves P, Katerinaki E, Simpson C, Layton C, Dawson R, Evans G, MacNeil S (2003). Melanoma invasion in reconstructed human skin is influenced by skin cells – investigation of the role of proteolytic enzymes. Clinical & Experimental Metastasis. 20(8): 685-700.

[4] Frasca M, Sharkey KJ (2016) Discrete-time moment closure models for epidemic spreading in populations of interacting individuals. Journal of Theoretical Biology. 399: 13-21.

[5] Jeon J, Quaranta V, Cummings PT (2010). An off-lattice hybrid discrete-continuum model of tumor growth and invasion. Biophysical Journal. 98 (1): 37-47.

[6] Jin W, Shah ET, Penington CJ, McCue SW, Chopin LK, Simpson MJ (2016). Reproducibility of scratch assays is affected by the initial degree of confluence: Experiments, modelling and model selection. Journal of Theoretical Biology. 390: 136-145.

[7] Johnston ST, Simpson MJ, Baker RE (2015). Modelling the movement of interacting cell populations: A moment dynamics approach. Journal of Theoretical Biology. 370: 81-92.

[8] Keller EF, Segel LA (1971). Traveling bands of chemotactic bacteria: a theoretical analysis. Journal of Theoretical Biology. 30:235.

[9] Kirkwood JG (1935). Statistical mechanics of fluid mixtures. Journal of Chemical Physics. 3 (5): 300-313.

[10] Maini PK, McElwain DL, Leavesley DI (2004). Traveling wave model to interpret a wound-healing cell migration assay for human peritoneal mesothelial cells. Tissue Engineering. 10 (3-4): 475-82.

[11] Middleton AM, Fleck C, Grima R (2014). A continuum approximation to an off-lattice individual-cell based model of cell migration and adhesion. Journal of Theoretical Biology. 359: 220-232.

[12] Murrell DJ, Dieckmann U, Law R (2004). On moment closures for population dynamics in continuous space. Journal of Theoretical Biology. 229: 421-432.

[13] Murray PJ, Edwards CM, Tindall MJ, Maini PK (2009). From a discrete to a continuum model of cell dynamics in one dimension. Physical Review E. 80 (3): 031912.

[14] Newman TJ, Grima R (2004). Many-body theory of chemotactic cell-cell interactions. Physical Review E. 70: 051916.

[15] Painter KJ, Sherratt JA (2003). Modelling the movement of interacting cell populations. Journal of Theoretical Biology. 225: 327-339.

[16] Penington CJ, Hughes BD, Landman KA (2011). Building macroscale models from microscale probabilistic models: A general probabilistic approach for nonlinear diffusion and multispecies phenomena. Physical Review E. 84: 041120.

[17] Press WH, Flannery BP, Teukolsky SA, Vetterling WT (2007). Numerical recipes: The art of scientific computing (3rd ed.). Cambridge University Press. Cambridge. UK.

[18] Sherratt JA, Murray JD (1990). Models of epidermal wound healing. Proceedings of the Royal Society of London B. 241: 29-36.

[19] Simpson MJ, Landman KA, Hughes BD, Newgreen DF (2006). Looking inside an invasion wave of cells using continuum models: Proliferation is the key. Journal of Theoretical Biology, 243: 343-360.

[20] Simpson MJ, Zhang DC, Landman KA, Newgreen DF (2007). Cell proliferation drives neural crest cell invasion of the intestine. Developmental Biology, 302: 553-568.

[21] Simpson MJ, Landman KA, Hughes BD (2009). Multi-species simple exclusion processes. Physica A: Statistical Mechanics and its Applications. 388: 399-406.

[22] Simpson MJ, Binder BJ, Haridas P, Wood BK, Treloar KK, McElwain DLS, Baker RE (2013). Experimental and modelling investigation of monolayer development with clustering. Bulletin of Mathematical Biology. 75: 871-889.

[23] Srinivas S, Rodriguez T, Clements M, Smith JC, Beddington RSP (2004). Active cell migration drives the unilateral movements of the anterior visceral endoderm. Development. 131 (5): 1157.

[24] Steinberg MS (1996). Adhesion in development: An historical review. Developmental Biology. 180: 377-388.

[25] Tersoff J (1987). New empirical approach for the structure and energy of covalent systems. Physical Review B. 37(12): 6991-7000.

[26] Treloar KK, Simpson MJ, Haridas P, Manton KJ, Leavesley DI, McElwain DLS, Baker RE (2013). Multiple types of data are required to identify the mechanisms influencing the spatial expansion of melanoma cell colonies. BMC Systems Biology. 7:137.

[27] Van Kampen NG (1976). Stochastic differential equations. Physics Reports. 24c (3): 171-228.

[28] Weinberg RA (2009). The biology of cancer. Second edition. Garland Science. New York. USA.