Continuum descriptions of spatial spreading for heterogeneous cell populations: theory and experiment

Oleksii M Matsiaka[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

Variability in cell populations is frequently observed in both *in vitro* and *in vivo* settings. Intrinsic differences within populations of cells, such as differences in cell sizes or differences in rates of cell motility, can be present even within a population of cells from the same cell line. We refer to this variability as cell *heterogeneity*. Mathematical models of cell migration, for example, in the context of tumour growth and metastatic invasion, often account for both undirected (random) migration and directed migration that is mediated by cell-to-cell contacts and cell-to-cell adhesion. A key feature of standard models is that they often assume that the population is composed of identical cells with constant properties. This leads to relatively simple single-species *homogeneous* models that neglect the role of heterogeneity. In this work, we use a continuum modelling approach to explore the role of heterogeneity in spatial spreading of cell populations. We employ a three-species heterogeneous model of cell motility that explicitly incorporates different types of experimentally-

motivated heterogeneity in cell sizes: (i) monotonically decreasing; (ii) uniform; (iii) non-monotonic; and (iv) monotonically increasing distributions of cell size. Comparing the density profiles generated by the three-species heterogeneous model with density profiles predicted by a more standard single-species homogeneous model reveals that when we are dealing with some forms of heterogeneity a simple and computationally efficient single-species homogeneous model can be remarkably accurate in describing the evolution of a heterogeneous cell population. In contrast, we also identify other forms of heterogeneity for which the simpler single-species homogeneous model performs relatively poorly. Additional results for heterogeneity in parameters describing both undirected and directed cell migration are also considered, and we find that similar results apply.

Key words: Cell migration; heterogeneity; continuum description; scratch assay

---

* Corresponding author
Email address: matthew.simpson@qut.edu.au, *Telephone* + 617 31385241, *Fax* + 617 3138 2310 (*Matthew J Simpson[1]).

1 Introduction

In vitro cell migration experiments play an important role in the discovery and testing of putative drug treatments, the study of malignant tumour growth and metastasis, as well as tissue regeneration and repair (Savla et al., 2004; Sengers et al., 2007; Tremel et al., 2009; Sarapata and de Pillis, 2010; Gerlee, 2013; Edmondson et al., 2014; Shah et al., 2016). Mathematical models of many biological processes involved in these experiments normally require certain assumptions to make the problem mathematically and computationally tractable. When modelling large populations of cells, one of the most intuitive approaches is to assume that all cells have fixed properties, such as assuming all cells have constant size and constant diffusivity (Sherratt and Murray, 1990; Galle et al., 2005; Simpson et al., 2013). In this framework a cell population is considered to be a *homogeneous* population, and single-species homogeneous models are routinely invoked (Maini et al., 2004a; Maini et al., 2004b; Sepulveda et al., 2013; Simpson et al., 2013; George et al., 2017). Single-species homogeneous models are much less computationally expensive than more elaborate multi-species heterogeneous models and, as a result, are frequently used relative to multi-species counterparts. In addition, multi-species frameworks usually involve a significantly larger number of free model parameters that we may have little prior knowledge about and so the process of calibrating multi-species heterogeneous models to match experimental observations is significantly more challenging than calibrating single-species homogeneous models. This is an important consideration because it is well-known that parameterising mathematical models of biological processes can be challenging, often requiring computationally-intensive methods (Pozzobon

and Perré, 2018; Warne et al. 2019).

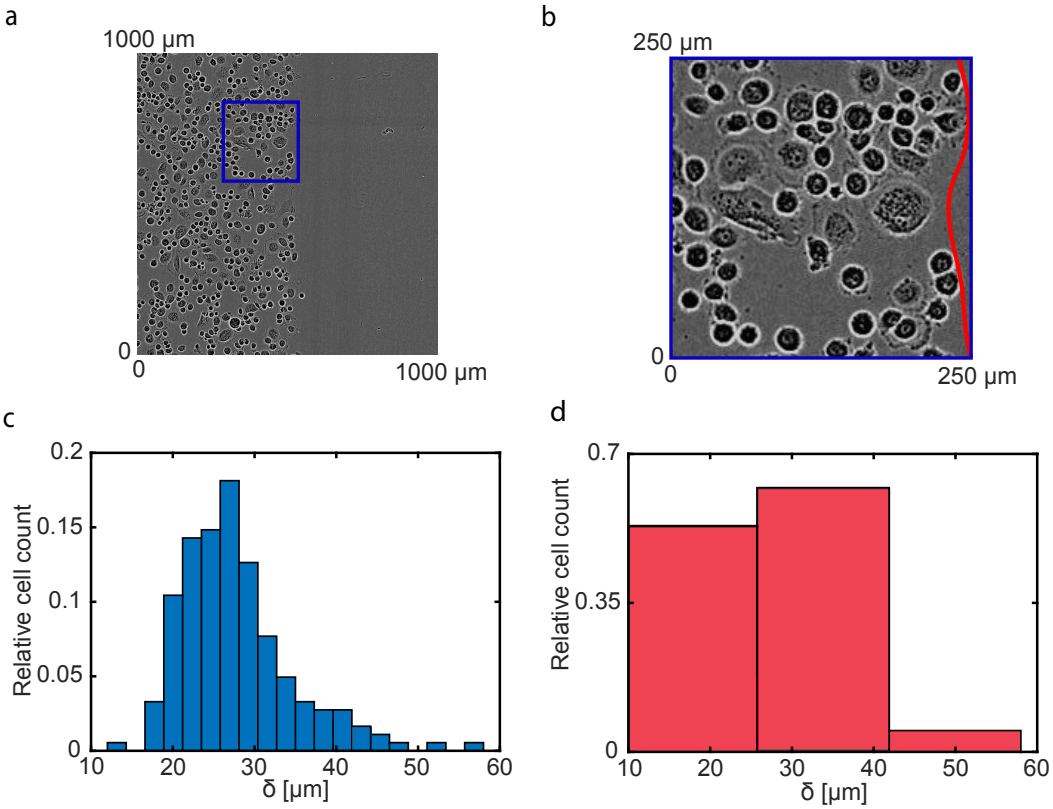

Fig. 1. Heterogeneity in a population of PC-3 prostate cancer cells (Kaighn et al., 1979). (a) Experimental image of an advancing cell population and corresponding cell size distribution. The red solid line denotes position of the leading edge. (b) Detailed image of the subregion denoted in the blue rectangle in Figure 1(a). (c) Cell size distribution with a bin size of 15 μm. The cell size distribution is obtained from the sample of 184 cells randomly selected from the population. (d) Cell size distribution with a bin size of 2.3 μm. The histogram in Figure 1(d) is constructed using the same sample of 184 cells.

Although *heterogeneity* in cell populations is frequently observed in experiments, there is relatively little guidance or consensus in the literature about how to incorporate such heterogeneity into the mathematical models used to replicate and predict such experiments (An et al., 2001; Altschuler et al., 2010). Figure 1(a)-(b) shows a typical experiment where we can clearly visually observe cells of different sizes. The measured cell size distribution in Figure 1(c) quantifies this heterogeneity in cell sizes and raises the question if

the most straightforward approach of applying a single-species homogeneous model can be reasonably used to predict the spatial spreading of this clearly heterogeneous population. In addition to the clear visual heterogeneity in cell sizes, it could be relevant to consider that cells of different sizes can exhibit different behaviour such as different rates of motility, or different mechanical properties including resistance to deformation and adhesion. Therefore, it could be possible that there are multiple types of heterogeneity acting in even this very simple experiment. Previously, heterogeneity in cell populations has been introduced in both discrete and continuum models of cell motility (Simpson et al., 2014; Jin et al., 2016b; Sundstrom et al., 2016; Matsiaka et al., 2017). However, these previous modelling studies do not address the basic question of identifying whether it is absolutely necessary to apply a multi-species heterogeneous models to obtain a faithful description of the behaviour of the heterogeneous population and whether different forms of heterogeneity affect the answer to this fundamental question.

In our work we use an experimentally-motivated approach to investigate the role of heterogeneity in two-dimensional scratch assays, and we compare the performance of a single-species homogeneous model relative to a heterogeneous multi-species model. We use numerical solutions of the multi-species heterogeneous model to produce synthetic test data that we use to investigate the performance of a simpler single-species homogeneous model. To mimic experimental data, such as depicted in Figure 1, we use the multi-species continuum approach introduced by Matsiaka et al. (2017). To keep our work tractable, we describe the heterogeneity by dividing the total population into three subpopulations with varying properties. The choice of working with three subpopulations allows us to keep the model computationally tractable while

capturing important differences in the population properties, as illustrated in Figure 1(d). Throughout this work we consider four distinct distributions of cell sizes: (i) monotonically decreasing; (ii) uniform; (iii) non-monotonic; and (iv) monotonically increasing distributions of cell size. The monotonically decreasing distribution, as shown in Figure 3(a), is a fairly accurate approximation of the experimentally observed cell size distribution in Figure 1(d). The other three kinds of distributions are included in our work for completeness. Our findings suggest that, for certain cell size distributions, namely monotonically decreasing and uniform distributions, the single-species homogeneous model performs remarkably well with an excellent match between the density profiles generated by the three-species heterogeneous model and density profiles predicted by its single-species homogeneous analogue. Therefore, our results imply that applying a single-species homogeneous model to describe experiments with monotonically decreasing or uniform cell size distributions might be sufficient for accurately predicting population-level behaviour. In contrast, the data with non-monotonic and monotonically increasing cell size distributions might require the application of multi-species models to account for differences in population.

This manuscript is organised in the following way. In Section 2 we describe experimental data for a series of two-dimensional scratch assays that clearly involve a significant level of heterogeneity among the population. In Section 3 we introduce a mathematical model of the cell motility and adhesion. In particular, we focus on two analogues of the mathematical model: (i) a three-species heterogeneous model of cell motility where parameters including cell size, cell diffusivity and cell adhesion strength can vary between the subpopulations; and (ii) a more traditional single-species homogeneous model of cell motility

where all cells in the population are treated as having the same cell size, cell diffusivity and cell adhesion strength. Results in Section 4 compare performance of the single-species homogeneous model as applied to data generated using the three-species heterogeneous model for different cell size distributions. Additional results exploring the role of heterogeneity in both the undirected (diffusive) and directed (adhesion/cell-to-cell contacts) migration are detailed in the Supplementary Material. Finally, in Section 5 we summarise our result and propose potential extensions.

2 Experimental data

Monolayer scratch assays are performed using the IncuCyte ZOOM^{TM} system (Essen BioScience). In all experiments we use the PC-3 prostate cancer cell line (Kaighn et al., 1979) from the American Type Culture Collection (ATCC^{TM}, Manassas, USA). After growing, cells are removed from the flask using TrypLE^{TM} (ThermoFisher Scientific) in phosphate buffered saline, resuspended in growth medium and seeded at a density of 20,000 cells per well in 96-well ImageLock plates (Essen BioScience). The diameter of each individual well is 9000 μm.

Mitomycin-C is added at a concentration of 10 g/mL for two hours before a scratch is made in the monolayer of cells (Sadeghi et al., 1998). Mitomycin-C is a chemotherapy drug that blocks DNA replication and, consequently, stops proliferation. As a result of treatment the number of cells in the assay remains approximately constant since cells neither proliferate or die on the timescale of the experiment. Often scratch assays are performed using mitomycin-C treated cells so that the experiment focuses only upon the role of cell migration as

opposed to the combined effects of cell migration and cell proliferation. A WoundMaker^{TM} (Essen BioScience) is used to create identical scratches in the uniformly distributed populations. Medium is aspirated after scratching; each well is washed twice and refilled with fresh medium (100 μL). Plates are incubated in the IncuCyte ZOOM^{TM} and photographed every two hours for 48 hours. In total, these experiments are performed in eight of the 96 wells on the 96-well plate. In our work we use one of the experimental replicates at $t = 0\,\text{h}$, Figure 1, to quantify the heterogeneity in a cell population.

To quantify the heterogeneity in cell size we randomly select 184 cells from the experimental image in Figure 1(a) at $t = 0\,\text{h}$. Assuming each cell can be treated as a disc, we estimate the equivalent diameter of each individual cell using the following approach. First, we use the histogram tool in Photoshop CS5 to count a number of pixels in the area occupied by each individual cell. The pixel count is converted to an area, A. Second, we estimate the equivalent diameter, $\delta = \sqrt{4A/\pi}$ and use this data to produce histograms to illustrate and visualise the variability in cell size within the experiment. The resulting cell size distribution, presented as a histogram constructed with bin width $2.3\,\mu\text{m}$, is shown in Figure 1(d). The bin width $2.3\,\mu\text{m}$ is chosen to demonstrate the fine structure within the cell population that is not normally incorporated in mathematical models of cell migration. However, the computational simulation of a population with the cell size distribution shown in Figure 1(c) is impractical since it would require significant computational resources to simulate the dynamics of 17 distinct subpopulations. As a compromise, we increase the bin width to reduce the number of distinct subpopulations while still retaining a sufficient number of bins to allow us to broadly characterise the heterogeneity in the population. Figure 1(d) demonstrates the histogram

of cell sizes constructed using the same sample of cells with a larger bin size width of 15 μm. Here, we have three subpopulations that capture the key trends in the heterogeneity in Figure 1(c) without needing to deal with 17 distinct subpopulations.

3 Mathematical model

Discrete, stochastic models are often used to describe the spatial spreading of a population of cells, especially when the population of cells is not too large. Here, cells move and interact with each other via predefined force function, as illustrated schematically in Figure 2 (Newman and Grima, 2004; Callaghan et al., 2006; Hasenauer et al., 2011; Frascoli et al. 2013; Osborne et al., 2017). This approach is *individual based* in the sense that knowledge about the movement of each individual is essential to infer the evolution of a density on the population-level scale. One of the most popular individual based modelling approaches makes the assumption that the motion of each cell can be described by a Langevin stochastic differential equation (Newman and Grima, 2004; Middleton et al., 2014). As such, the system of N cells is described by a system of N stochastic differential equations of the form

$$\frac{\mathrm{d}\vec{x}_i}{\mathrm{d}t} = \sum_{i \neq j} \vec{F}_{ij} + \vec{\xi}_i, \tag{3.1}$$

where $\vec{x}_i$ is the position vector of the ith cell, $\vec{F}_{ij}$ is the interaction force between cells i and j, and $\vec{\xi}_i$ is the random stochastic force acting upon cell i (Middleton et al., 2014; George et al., 2017; Osborne et al., 2017). The interaction force, $\vec{F}_{ij}$, can be used to parametrise various features of cell populations,

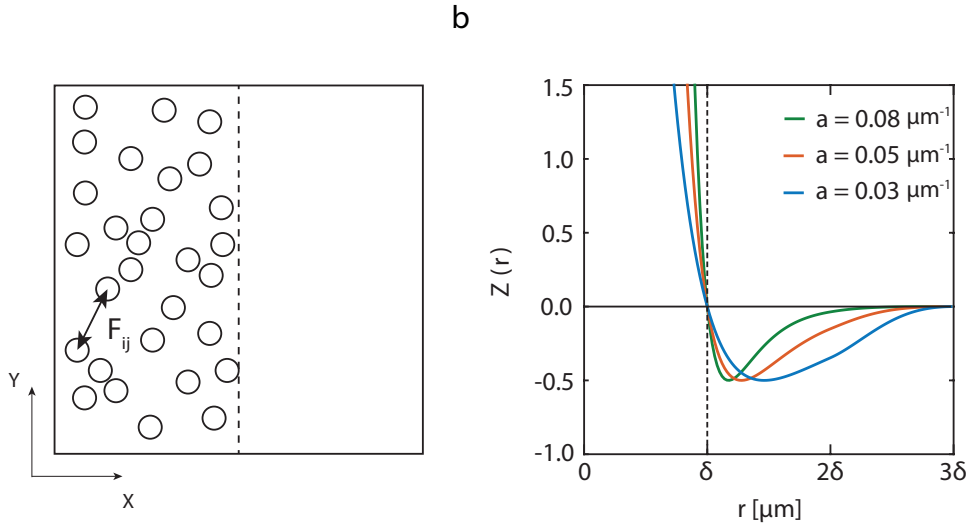

Fig. 2. (a) An idealisation of the front-like distribution of cells in the experimental design shown in Figure 1(a). Here all cells are of constant size. F_{ij} is the interaction force between cell i and cell j. The vertical dashed line represents the approximate leading edge of the population. (b) A typical cell-to-cell interaction force function in the form of the modified Morse potential, $Z(r)$, (Equation (3.7)) used to mimic adhesion and repulsion between individual cells. The vertical dashed line represents the diameter of individual agents, δ. The horizontal line at $Z(r) = 0$ shows the change from long-range attraction ($Z(r) < 0$ for $r > \delta$) to short-range repulsion ($Z(r) > 0$ for $r < \delta$).

including heterogeneity. In fact, it is relatively straightforward to model heterogeneity in cell sizes in a discrete framework since the interaction force, $\vec{F}_{ij}$, can be chosen to explicitly include the cell size as a parameter (Matsiaka et al., 2018). Here we can easily differentiate the population into an arbitrary number of subpopulations by assigning the value of the cell size to each member of the population. Despite the many advantages of this kind of individual based modelling approach, such individual based models are computationally inefficient as the number of cells, N, increases. This is because the computation time required to simulate such models increases with N.

In contrast, continuum models based on partial differential equations (PDEs) are much more convenient to model large cell populations because the time

taken to solve continuum PDE models is independent of the size of the population (Sherratt and Murray, 1990; Sheardown and Cheng, 1995; Cai et al., 2007; Wise et al., 2008). Often, PDE models are derived using continuum-limit approximations of underlying discrete models and are able to retain certain features of discrete (Middleton et al., 2014; O'Dea and King, 2012). In this work we focus on a continuum model that is derived by taking the limit of a three-species heterogeneous individual based model (Matsiaka et al., 2017). This approach allows us to conceptually incorporate key features of the heterogeneous cell populations into a discrete modelling framework, and then using a computationally efficient approach to solve the resulting continuum-limit PDE description of the underlying heterogeneous model.

We note that, due to the geometry of experiments presented in Figure 1, we are interested in the net movement of cells in only one direction, in this case the horizontal direction (Jin et al., 2016a). This is due to the fact that the net flux of cells in the vertical direction is, on average, zero because of the symmetry in the initial conditions of a scratch assay. Consequently, we focus on a one-dimensional continuum model and consider the evolution of the total cell population in the horizontal direction only. The use of a one-dimensional framework to describe two-dimensional scratch assays has been previously demonstrated to be a convenient approach to reduce the computational complexity while still describing the key features of the experiment (Matsiaka et al., 2018).

Here we employ a mean field model describing the spatial spreading of a population of cells composed of three distinct subpopulations. In one-dimension,

192 the model can be written as

$$\frac{\partial p^{(1)}(x,t)}{\partial t} = D_1 \Delta p^{(1)}(x,t) + \nabla(p^{(1)}(x,t)\, V^{(1,1)}(x,t)) - \sum_{m=1}^{3} n_m \nabla(p^{(1)}(x,t)\, V^{(1,m)}(x,t)), \tag{3.2}$$

$$\frac{\partial p^{(2)}(x,t)}{\partial t} = D_2 \Delta p^{(2)}(x,t) + \nabla(p^{(2)}(x,t)\, V^{(2,2)}(x,t)) - \sum_{m=1}^{3} n_m \nabla(p^{(2)}(x,t)\, V^{(2,m)}(x,t)), \tag{3.3}$$

$$\frac{\partial p^{(3)}(x,t)}{\partial t} = D_3 \Delta p^{(3)}(x,t) + \nabla(p^{(3)}(x,t)\, V^{(3,3)}(x,t)) - \sum_{m=1}^{3} n_m \nabla(p^{(3)}(x,t)\, V^{(3,m)}(x,t)), \tag{3.4}$$

$$V^{(l,m)}(x,t) = \int_{\Omega} F^{(l,m)}(x-y)\, p^{(m)}(y,t)\, \mathrm{d}y, \tag{3.5}$$

193 where $p^{(1)}(x,t)$, $p^{(2)}(x,t)$, and $p^{(3)}(x,t)$ are the cell densities associated with
194 each subpopulation and depend on position x and time t. In this heterogeneous
195 model, D_1, D_2, and D_3 are diffusivities of subpopulations 1, 2, and 3, n_1, n_2,
196 and n_3 are the numbers of cells in each subpopulation, and $V^{(l,m)}(x,t)$ is the
197 velocity field of subpopulation l induced by subpopulation m (Matsiaka et
198 al., 2017). The diffusivity constants parameterise the undirected migration of
199 each subpopulation and the velocity fields describe the directed migration of
200 each subpopulation that is driven by a combination of cell-to-cell adhesion
201 and crowding effects.

202 The interaction force between subpopulations l and m that describes directed
203 migration is given by

$$F^{(l,m)}(x-y) = f_0 \mathcal{Z}(r)\, \mathrm{sgn}(x-y), \tag{3.6}$$

204 where f_0 is the dimensional amplitude of the interaction force, $\mathcal{Z}(r)$ is a di-

mensionless function that parametrises different features of the cell-to-cell interactions, and *sgn* is the signum function. We choose to include long-range attraction that models cell-to-cell adhesion, and a short-range repulsion that reflects volume exclusion effects (Frascoli et al., 2013). A number of different phenomenological laws, $\mathcal{Z}(r)$, are used to model repulsive and adhesive intercellular forces (Murray et al., 2009; Jeon et al., 2010; Middleton et al., 2014). In our work we adopt modified Morse potential in the form

$$\mathcal{Z}(r) = \begin{cases} 2\Big(\exp[-2a\,(r-\delta)] - \exp[-a\,(r-\delta)]\Big), & r < 2\delta, \\ 2\Big(\exp[-2a\,(r-\delta)] - \exp[-a\,(r-\delta)]\Big)\, g(r), & 2\delta \leq r \leq 3\delta, \\ 0, & r > 3\delta, \end{cases} \tag{3.7}$$

where a is the shape parameter that controls the shape of the force function, δ is the cell size, and $r = |x - y|$. We fix the value of the shape parameter at $a = 0.08\,\mu\text{m}^{-1}$ (Matsiaka et al., 2017). The function $g(r) = \Big(1 - \sin\Big[(2\pi r - \pi\delta)/2\delta\Big]\Big)/2$ is the Tersoff cut-off function introduced to impose a finite range of intercellular interactions (Tersoff, 1988). A sketch of the potential function given by Equation (3.7) for different values of the parameter a is shown in Figure 2(b) confirming that this potential function describes short range repulsion, longer range attraction and no interactions at over much longer distances. In summary, the key parameters in our model are: (i) cell size, δ; (ii) the cell diffusivity, D; and (iii) the amplitude of interaction forces, f_0. In this work we will systematically explore how heterogeneity in each of these three key parameters influences whether we need to consider a complex heterogeneous multi-species model or whether we can describe the spatial spreading of a cell population using a relatively simple homogeneous, single-species models. Since our experimental data in Figure 1 allows us to

explicitly characterise the heterogeneity in cell size, all results in the main document focus on cell size. Additional results in the Supplementary Material focus on heterogeneity in D and f_0 to provide additional insight into the role of heterogeneity in these kinds of experiments.

We define the total density of the heterogeneous population as

$$\mathcal{P}(x,t) = \sum_{i=1}^{3} \Big[p^{(i)}(x,t) \Big], \tag{3.8}$$

where $p^{(i)}(x,t)$ is the cell density of subpopulation $i = 1, 2, 3$ predicted by Equations (3.2)-(3.4), and $\mathcal{P}(x,t)$ is the total cell density. It is important to interpret the solutions of Equations (3.2)-(3.4) in terms of total cell density since standard experimental protocols do not normally facilitate the measurement of spatial and temporal distributions of various subpopulations (Cai et al., 2007; Treloar et al., 2014).

We can reduce the three-species heterogeneous system of equations, Equations (3.2)-(3.4), to obtain a single-species homogeneous model in the form,

$$\frac{\partial P(x,t)}{\partial t} = D\Delta P(x,t) - (N-1)\nabla\Big(P(x,t)\, V(x,t) \Big), \tag{3.9}$$

where $P(x,t)$ is the cell density of the total population, $N = \sum_{i=1}^{3} n_i$ is the total number of cells in the population. Here we assume that the cell size, diffusivity and strength of the interaction force for each population is constant, giving $\delta_i = \delta$ and $D_i = D$ for $i = 1, 2, 3$ with f_0 also being treated as a constant across the population. The key differences between the homogeneous single-species model, Equation (3.9), and the three-species heterogeneous model, Equations (3.2)-(3.4) are: (i) the three-species heterogeneous model incorpo-

rates three advection-diffusion equations while the single-species homogeneous model is given by a single advection-diffusion equation; (ii) the three-species heterogeneous model contains up to nine free parameters as opposed to three parameters in the single-species homogeneous model.

The initial conditions in all simulations are chosen to mimic a cell front, such as that shown in our experimental data set, Figure 1(a). As such, we adopt an initial cell distribution in the form of the one-dimensional step function,

$$P(x,0) = \mathcal{P}(x,0) = \begin{cases} 23.9 \times 10^{-3}\,\text{cells}/\mu\text{m}, & 0\,\mu\text{m} < x < 1000\,\mu\text{m}, \\ 0 \qquad\qquad\quad \text{cells}/\mu\text{m}, & 1000\,\mu\text{m} < x < 2000\,\mu\text{m}, \end{cases} \tag{3.10}$$

on $0 < x < 2000\,\mu$m, which is consistent with a length-scale of a typical *in vitro* experiment (Jin et al., 2016a). The initial cell distribution in the heterogeneous model is given by the sum of initial densities of three subpopulations, $\mathcal{P}(x,0) = \sum_i p^{(i)}(x,0)$, where the density of each subpopulation, $p^{(i)}(x,0)$, varies between each cell size distribution and can be inferred from the histograms in Figure 3(a). The value of the initial density of the total population is chosen to represent fairly confluent population of cells. For example, the simulation of the three-species population with the monotonically decreasing cell size distribution, Set Ia, is initiated with the confluence level of approximately 65% of maximum packing density, which is fairly typical for scratch assay experiments (Jin et al., 2016; Matsiaka et al., 2017). We note that the boundary of the experimental image in Figure 1(a) is not a physical boundary and cells can freely move across this boundary because the image captures only a small fraction of a much larger experimental domain (Simpson et al., 2018). During the experiment, cells freely migrate, in each direction, across

the boundary. However, since the density of cells away from the scratch is spatially uniform, the net flux of cells across the boundary of the image is zero. To capture this situation we impose zero net flux boundary conditions at $x = 0\,\mu$m and $x = 2000\,\mu$m.

All continuum results for single-species homogeneous and three-species heterogeneous models, given by Equation (3.9) and Equations (3.2)-(3.4), respectively, are solved numerically using the method of lines with $\Delta x = 4\,\mu$m and $\Delta t = 0.005\,$h on $0 < x < 2000\,\mu$m (Matsiaka et al., 2017). We find that this choice of spatial and temporal discretisations are sufficiently fine to produce grid independent results. The detailed discretisation scheme used in this work is presented in the Supplementary material.

4 Results and Discussion

To investigate the ability of a single-species homogeneous model to capture the behaviour of the three-species heterogeneous analogue, we consider a series of case studies. In these case studies we vary only one parameter at a time to simplify our analysis and to focus on the impact of each individual parameter. Another approach would be to use the mathematical models to explore heterogeneity multiple parameter at the same time. However, in this first instance, we prefer to take a more fundamental approach and examine the role of heterogeneity in each parameter separately. In the first set of experiments, Set I, we vary the cell size, δ, while keeping D and f_0 fixed at $D = 250\,\mu\text{m}^2/\text{h}$ and $f_0 = 1.0\,\mu\text{m/h}$, respectively (Matsiaka et al., 2019). These estimates of D and f_0 are based on detailed experimental measurements reported previously (Matsiaka et al., 2019).

There are number of ways to quantify performance of the single-species homogeneous model in our framework. The position of the leading edge of the spreading population is routinely used by experimentalists to provide quantitative insights into the rate of spatial spreading of a cell population (Treloar and Simpson, 2013; Johnston et al., 2014; Nardini et al., 2016; Bobadilla et al., 2019). Therefore, we quantify the discrepancy between the solution of the heterogeneous three-species model and the homogeneous single-species model using an error measure, $E(\delta)$, associated with the position of the leading edge,

$$E(\delta) = \frac{1}{\alpha} \sum_{j} \Big[\mathcal{S}(t_j) - S(t_j) \Big]^2, \tag{4.1}$$

where $\mathcal{S}(t_j)$ is the position of the leading edge according to the three-species heterogeneous model at time t_j, $S(t_j)$ is the position of the leading edge predicted by the single-species homogeneous model, and α is a number of discrete time points that we used to compute $E(\delta)$. In both scenarios the position of the leading edge is computed as the coordinate on the one-dimensional domain where the density is 1% of the initial density (Treloar and Simpson, 2013).

In the first set of results we present, called Set I, we choose to fix D and f_0, and vary δ between each subpopulation in the three-species heterogeneous model. The experimental distribution of cell sizes in Figure 1(d) provides insights into potential choices of the cell size distribution in Equations (3.2)-(3.4). Here we define three subpopulations based on the equivalent cell size: small (18 μm), medium (34 μm), and large cells (50 μm). For simplicity, we set the fractions of small and medium cells to be equal and refer to this distribution as a monotonically decreasing distribution of cell sizes (Set Ia, Figure 3). After considering the experimentally-motivated monotonically decreasing distribution, we then

systematically explore: (i) uniform (Set Ib, Figure 4), (ii) non-monotonic (Set Ic, Figure 5), and (iii) monotonically increasing distributions (Set Id, Figure 6).

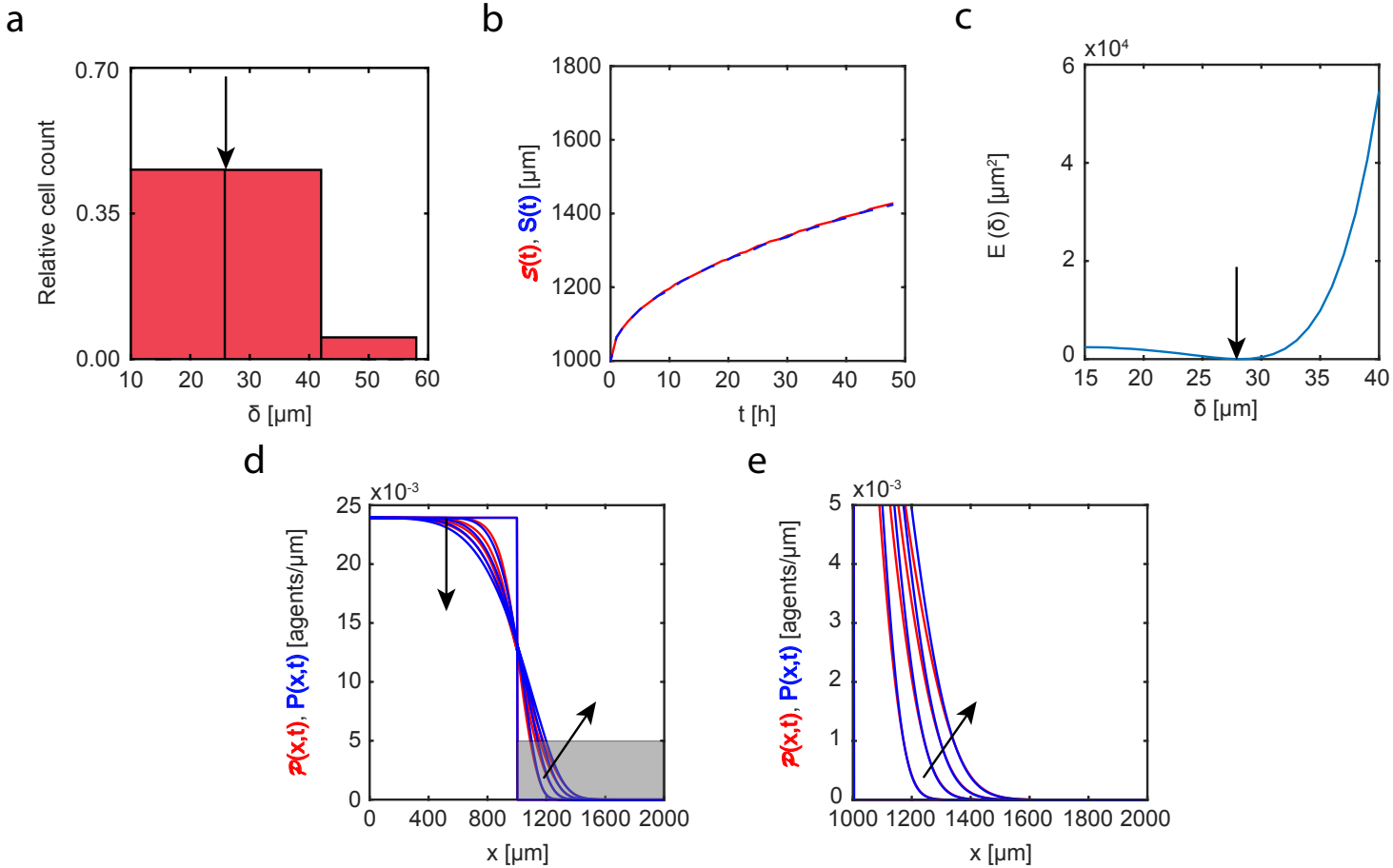

Fig. 3. Set Ia. Heterogeneity in cell sizes: monotonically decreasing distribution. (a) Cell size distribution adopted in the three-species heterogeneous model, Equations (3.2)-(3.4). (b) Leading edge as predicted by the three-species heterogeneous model, $\mathcal{S}(t)$ (solid red), and the best-fit approximation given by the single-species homogeneous model, $S(t)$ (blue dashed). (c) Error measure, $E(\delta)$, between the density profiles given by the three-species heterogeneous model and the profiles predicted by the single-species homogeneous model as a function of cell size, δ. The black arrow denotes the best-fit value of cell size, $\delta = 28\,\mu$m. (d)-(e) Cell density profiles predicted by the three-species heterogeneous model, $\mathcal{P}(x,t)$ (solid red), superimposed with density profiles given by the single-species homogeneous model calibrated with the best-fit value of δ, $P(x,t)$ (solid blue). The continuum results for both models are presented at $t = 0, 12, 24, 36,$ and 48 h. Black arrows denote the direction of increasing time. Results in (e) show a close-up comparison right near the leading edge, denoted by the gray shaded region in (d).

Figure 3(b) compares the leading edge prediction, $\mathcal{S}(t)$, given by the three-species heterogeneous model with the associated best-fit match, $S(t)$, predicted by the single-species homogeneous model. Our systematic computation of the error measure, $E(\delta)$, demonstrates a clear minimum which ensures the unique choice of a best-fit cell size, δ. Results in Figure 3(d) superimposes the solution of the three-species heterogeneous model with the solution of the single-species homogeneous model parameterised with the best fit cell size. Comparing the time evolution of the spreading density profiles in Figure 3(d) (with additional details at the leading edge shown in the magnified region in Figure 3(e)) we see that the appropriately parameterised single-species homogeneous model captures the temporal evolution of the spreading profile given by the heterogeneous model remarkably accurately. In particular, the density profiles predicted by the single-species homogeneous model match both the position and shape of the density profiles generated by the three-species heterogeneous model. These results imply that in this case it would be reasonable to use a much simpler single-species homogeneous model to describe and predict this spatial spreading.

Visual inspection of the results in Figures 3 - 6 suggests that we can always find a unique, well-defined value of the cell size in the single-species homogeneous model to provide an accurate prediction of the temporal evolution of the position of a leading edge of the spreading heterogeneous cell populations regardless of the underlying cell size distribution in the three-species heterogeneous model (Figures 3(b)-6(b)). In contrast, the quality of match between the shape of the shape of the density profiles for the three-species heterogeneous model and the single-species homogeneous model varies significantly between different cell size distributions. For example, the experimentally mo-

tivated distribution in Figure 3(a) leads to a remarkably good match between the three-species heterogeneous model and the single-species homogeneous model. In contrast, the shape of the density profiles associated with the non-monotonic cell size distribution (Figure 5) and monotonically increasing cell size distribution (Figure 6) show a relatively poor match. In these cases, it would seem prudent not to use a simpler single-species homogeneous model to simulate and predict these experiments.

The values of the cell size, δ, that produce best match between the single-species homogeneous and three-species heterogeneous models vary significantly between different cell size distributions. For example, the best-fit value of the cell size for the uniform distribution (Figure 4), $\delta = 36\,\mu$m, is quite close to the weighted average value of $\delta = 34\,\mu$m for the distribution in Figure 4(a). This indicates that the choice of a simple weighted average of the cell sizes might be a reasonable way to to parameterise the single-species homogeneous model if the experimentally observed distribution is close to uniform. We observe similar agreement for best-fit values of the cell size in the case of monotonically decreasing and monotonically increasing cell size distributions (Figure 3, Figure 6). In contrast, the best-fit value of cell size for the non-monotonic distribution, $\delta = 40\,\mu$m, differs significantly from the weighted average of $\delta = 34\,\mu$m. Therefore, these results suggest that great care out to be exercised when taking a distribution of parameter values and attempting to select the most appropriate single representative value of that parameter.

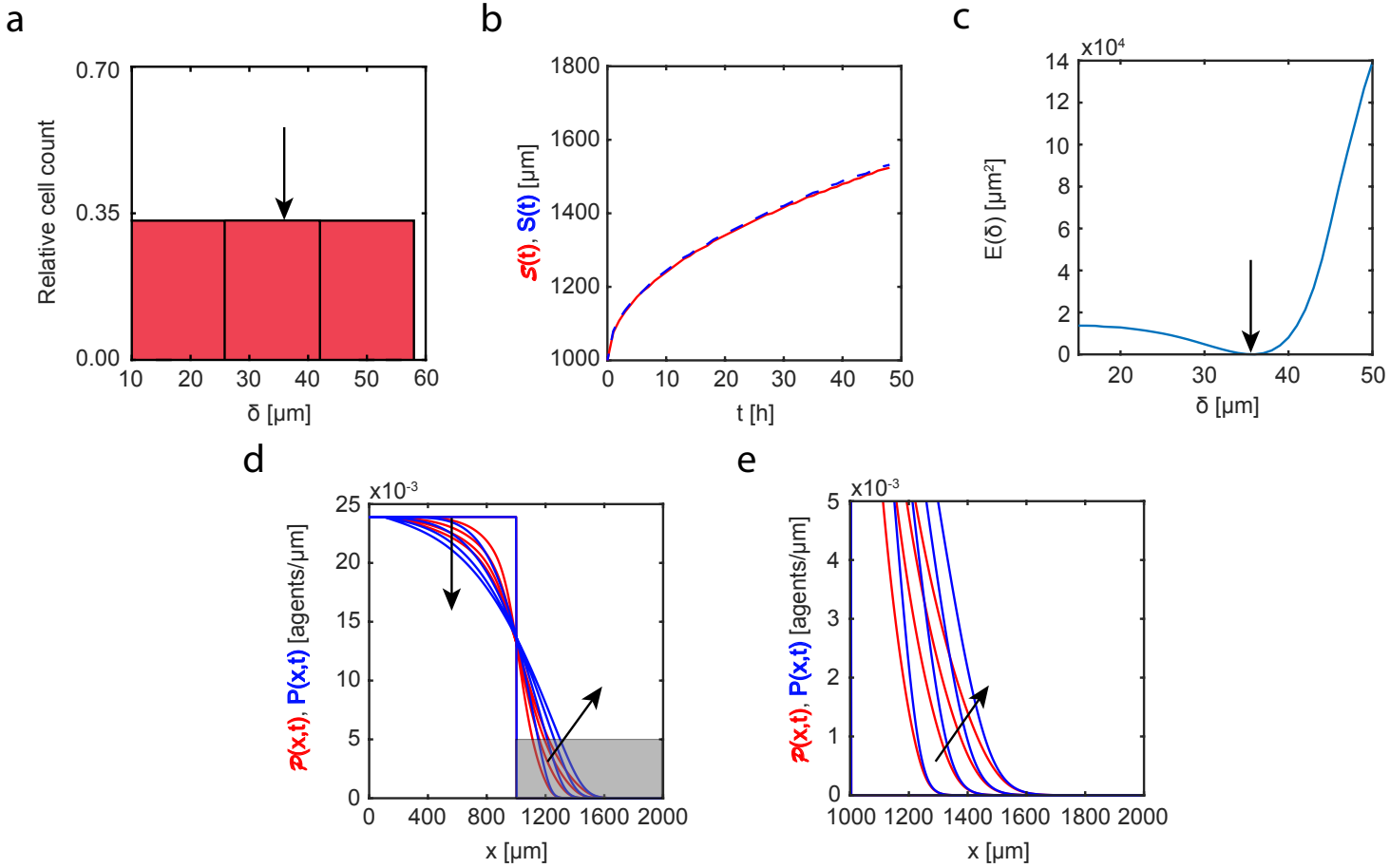

Fig. 4. Set Ib. Heterogeneity in cell sizes: uniform distribution. (a) Cell size distribution adopted in the three-species heterogeneous model, Equations (3.2)-(3.4). (b) Leading edge as predicted by the three-species heterogeneous model, $\mathcal{S}(t)$ (solid red), and the best-fit approximation given by the single-species homogeneous model, $S(t)$ (blue dashed). (c) Error measure, $E(\delta)$, between the density profiles given by the three-species heterogeneous model and the profiles predicted by the single-species homogeneous model as a function of cell size, δ. The black arrow denotes the best-fit value of cell size, $\delta = 36\,\mu$m. (d)-(e) Cell density profiles predicted by the three-species heterogeneous model, $\mathcal{P}(x,t)$ (solid red), superimposed with density profiles given by the single-species homogeneous model calibrated with the best-fit value of δ, $P(x,t)$ (solid blue). The continuum results for both models are presented at $t = 0, 12, 24, 36$, and 48 h. Black arrows denote the direction of increasing time. Results in (e) show a close-up comparison right near the leading edge, denoted by the gray shaded region in (d).

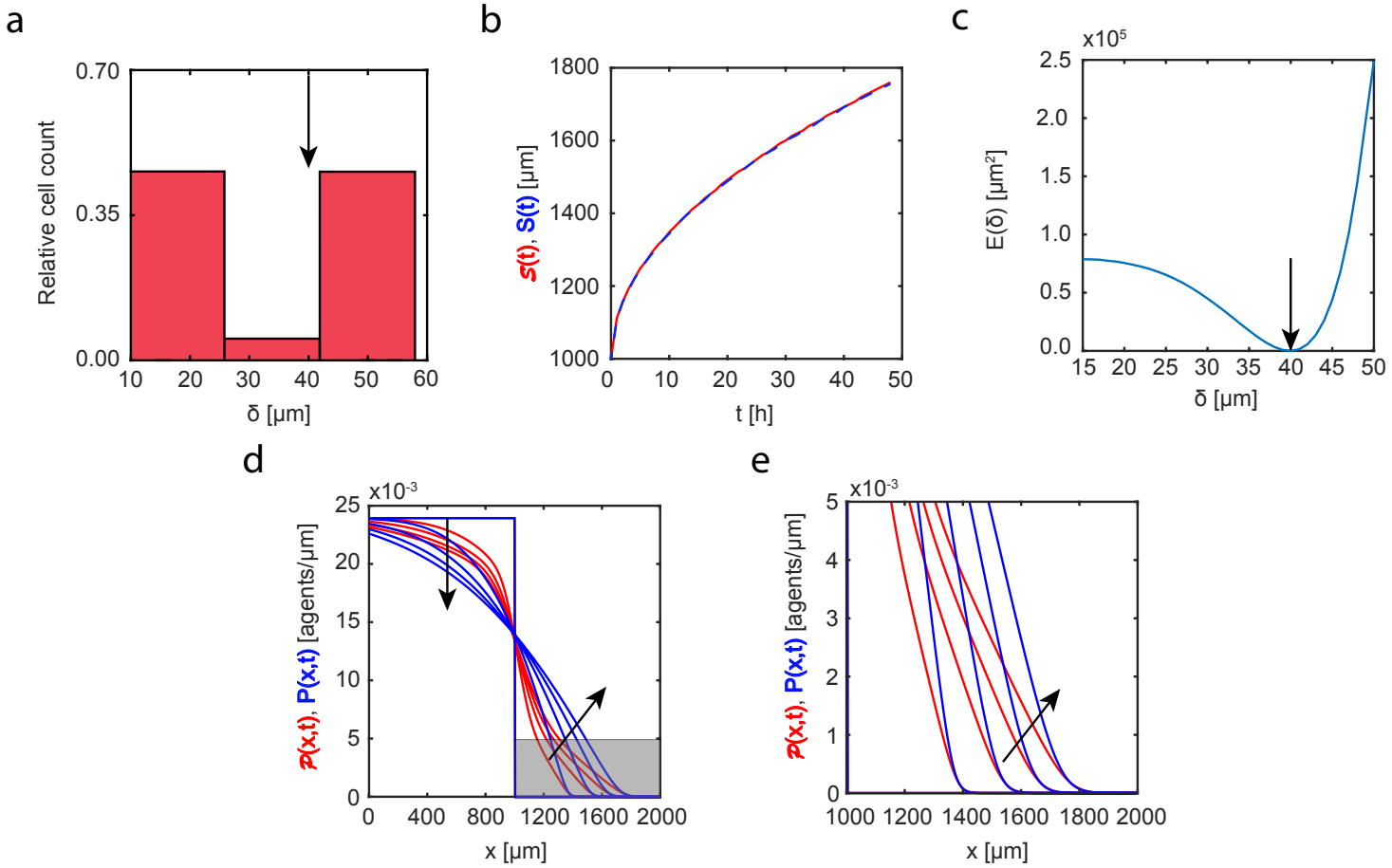

Fig. 5. Set Ic. Heterogeneity in cell sizes: non-monotonic distribution. (a) Cell size distribution adopted in the three-species heterogeneous model, Equations (3.2)-(3.4). (b) Leading edge as predicted by the three-species heterogeneous model, $\mathcal{S}(t)$ (solid red), and the best-fit approximation given by the single-species homogeneous model, $S(t)$ (blue dashed). (c) Error measure, $E(\delta)$, between the density profiles given by the three-species heterogeneous model and the profiles predicted by the single-species homogeneous model as a function of cell size, δ. The black arrow denotes the best-fit value of cell size, $\delta = 40\,\mu$m. (d)-(e) Cell density profiles predicted by the three-species heterogeneous model, $\mathcal{P}(x,t)$ (solid red), superimposed with density profiles given by the single-species homogeneous model calibrated with the best-fit value of δ, $P(x,t)$ (solid blue). The continuum results for both models are presented at $t = 0, 12, 24, 36$, and 48 h. Black arrows denote the direction of increasing time. Results in (e) show a close-up comparison right near the leading edge, denoted by the gray shaded region in (d).

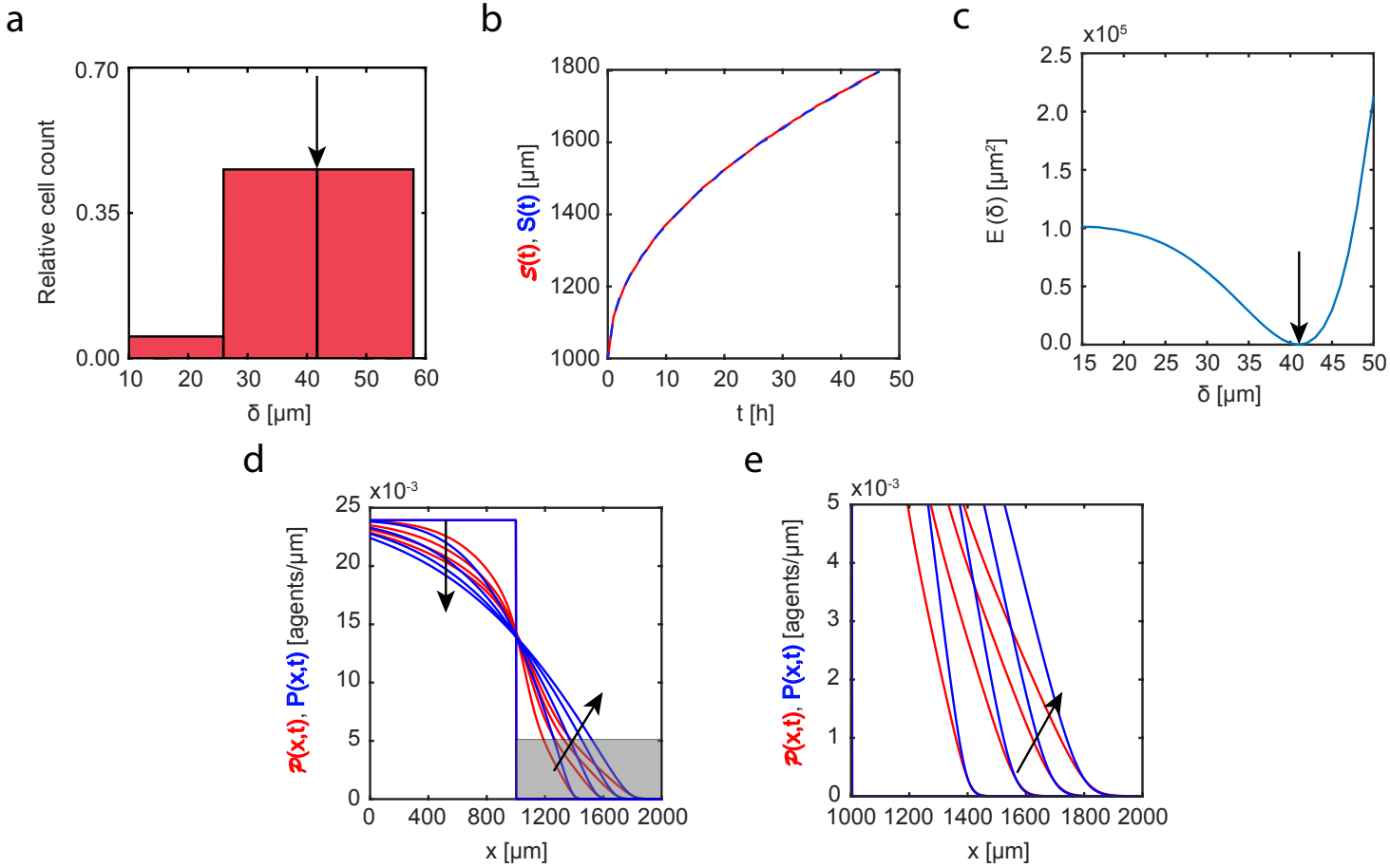

Fig. 6. Set Id. Heterogeneity in cell sizes: monotonically increasing distribution. (a) Cell size distribution adopted in the three-species heterogeneous model, Equations (3.2)-(3.4). (b) Leading edge as predicted by the three-species heterogeneous model, $\mathcal{S}(t)$ (solid red), and the best-fit approximation given by the single-species homogeneous model, $S(t)$ (blue dashed). (c) Error measure, $E(\delta)$, between the density profiles given by the three-species heterogeneous model and profiles predicted by the single-species homogeneous model as a function of cell size, δ. The black arrow denotes the best-fit value of cell size, $\delta = 41\,\mu$m. (d)-(e) Cell density profiles predicted by the three-species heterogeneous model, $\mathcal{P}(x,t)$ (solid red), superimposed with density profiles given by the single-species homogeneous model calibrated with the best-fit value of δ, $P(x,t)$ (solid blue). The continuum results for both models are presented at $t = 0, 12, 24, 36,$ and 48 h. Black arrows denote the direction of increasing time. Results in (e) show a close-up comparison right near the leading edge, denoted by the gray shaded region in (d).

In addition to the results in Figures 3 - 6 exploring the role of heterogeneity in cell size, we present an additional suite of results where we systematically explore the role of heterogeneity in D (Set II) and f_0 (Set III) while keeping δ constant. These additional results are presented in the Supplementary Material document. Both Set II and Set III data sets demonstrate exceptional quality of match between the three-species heterogeneous simulation data and its best-fit single-species homogeneous equivalent. Again, these additional results provide guidance about when it is reasonable to approximate a more complicated heterogeneous mathematical model with a simpler single-species homogeneous model.

5 Conclusions

In this work, we explore the role of heterogeneity in the context of studying how an initially confined population of cells can spread into surrounding initially unoccupied regions, as in the case of a scratch assay. We use a three-species heterogeneous model of cell motility, account for undirected cell motility, short range repulsion (crowding) and longer range adhesion, to capture experimentally observed heterogeneity in cell sizes from a new experimental data set from a two-dimensional scratch assay as shown in Figure 1. Our continuum models account for the undirected random motility, cell-to-cell adhesion, and cell crowding. The single-species homogeneous model is applied to each set of three-species heterogeneous simulation data in an attempt to match cell density profiles.

To analyse the performance of the single-species homogeneous model to capture data from our three-species heterogeneous model we consider four dif-

ferent cell size distributions: (i) monotonically decreasing distribution, (ii) uniform distribution, (iii) non-monotonic distribution, and (iv) monotonically increasing distribution. Overall, for a set of experimentally-motivated parameter combinations, we find that the standard single-species homogeneous model is able to accurately predict the position of the leading edge for all case studies presented. However, the quality of the match between the shape of the density profiles varies significantly depending on the details of the form of the heterogeneity present. For example, the monotonically decreasing distribution (Set Ia) demonstrates remarkable goodness of fit between the two sets of density profiles, as shown in Figure 3(d). This result is important because the monotonically decreasing cell size distribution is chosen to mimic the distribution of the cell sizes observed in our new experimental data set, shown in Figure 1. Similarly, the homogeneous distribution, Figure 4, shows that single-species homogeneous model is able to accurately replicate the three-species heterogeneous model results. This is an expected result because the uniform population of cells does not involve any heterogeneity and so we might anticipate that the simpler homogeneous model would perform well in this case. In contrast, the single-species homogeneous model does not perform so well when applied to both non-monotonic and monotonically increasing distributions in Figures 5-6, respectively. Additionally we explore potential heterogeneity in diffusivity and amplitude of the cell-to-cell interactions (Supplementary material). Overall, our results suggest that for certain cell size distributions, a simple and computationally efficient single-species homogeneous model is preferable over a thee-species heterogeneous model.

There are number of ways this work can be extended which we leave for future analysis. All our simulations and analysis focus on treating the heterogeneity

in the population of cells by considering the total population to be composed of three distinct subpopulations. For more extreme forms for heterogeneity, such as multi-modal distributions, the results presented in this work could be extended by considering additional subpopulations. Another avenue for further exploration would be to consider heterogeneity in more than one parameter at a time, whereas in this work we have taken the most fundamental approach and examined heterogeneity in just one parameter in isolation from the others. For both of these extensions, the modelling framework presented in this study can be extended to explore these additional features, and we leave such extensions for future consideration. Another option for extending the work would be to consider further details in the mathematical models, such as the effects of combined cell migration and combined cell proliferation. Here we have not pursued this approach because our experimental data set has been carefully prepared to exclude the effects of proliferation so that we can focus just on cell migration and heterogeneity in cell migration alone.

Acknowledgements

This work is supported by the Australian Research Council (DP170100474). Computational resources are provided by the High Performance Computing and Research Support Group at QUT. REB is a Royal Society Wolfson Research Merit Award holder, would like to thank the Leverhulme Trust for a Research Fellowship and also acknowledges the BBSRC for funding via grant no. BB/R000816/1.

References

[1] Altschuler, S.J., Wu, L.F., 2010. Cellular heterogeneity: when do differences make a difference? Cell 141(4), 559-563.

[2] An, F.-Q., Matsuda, M., Fujii, H., Tang, R.-F., Amemiya, H., Dai, Y.-M., Matsumoto Y., 2001. Tumor heterogeneity in small hepatocellular carcinoma: Analysis of tumor cell proliferation, expression and mutation of p53 and β-Catenin. International Journal of Cancer 93, 468-474.

[3] Baker, R.E., Simpson, M.J., 2010. Correcting mean-field approximations for birth-death-movement processes. Physical Review E 82, 041905.

[4] Binder, B.J., Simpson, M.J., 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.

[5] Bobadilla, A.V.P., Arevalo. J., Sarro, E., Byrne, H.M., Maini, P.K., Carraro, T., Balocco, S., Meseguer, A., Alarcon, T., 2019. In vitro cell migration quantification method for scratch assays. Journal of the Royal Society Interface 16, 20180709.

[6] Cai, A.Q., Landman, K.A., Hughes, B.D., 2007. Multi-scale modeling of a wound-healing cell migration assay. Journal of Theoretical Biology 245(3), 576-594.

[7] Callaghan, T., Khain, E., Sander, L.M., Ziff, R.M., 2006. A stochastic model for wound healing. Journal of Statistical Physics 122(5), 909-924.

[8] Edmondson, R., Broglie, J.J., Adcock, A.F., Yang, L., 2014. Three-dimensional cell culture systems and their applications in drug discovery and cell-based biosensors. ASSAY and Drug Development Technologies 12(4), 207-218.

[9] Essen BioScience: IncuCyte ZOOM live cell imaging.

Available from: http://www.essenbioscience.com/essen-products/incucyte/ (Accessed: May 2019).

[10] Frascoli, F., Hughes, B.D., Zaman, M.H., Landman, K.A., 2013. A computational model for collective cellular motion in three dimensions: General framework and case study for cell pair dynamics. PLoS ONE 8(3), e59249.

[11] Galle, J., Loeffler, M., Drasdo D., 2005. Modeling thee effect of deregulated proliferation and apoptosis on the growth dynamics of epithelial cell populations *in vitro*. Biophysical Journal 88, 62-75.

[12] Gerlee, P., 2013. The model muddle: In search of tumor growth laws. Cancer Research 73(8), 2407-2411.

[13] George, M., Bullo F., Campas O., 2017. Connecting individual to collective cell migration. Scientific Reports 7, 9720.

[14] Hasenauer, J., Waldherr, S., Doszczak, M., Scheurich, P., Radde, N., Allgwer, F. 2011. Analysis of heterogeneous cell populations: A density-based modeling and identification framework. Journal of Process Control 21, 1417-1425.

[15] 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.

[16] Jin, W., Shah, E.T., Penington, C.J., McCue, S.W., Chopin, L.K., Simpson, M.J., 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.

[17] Jin, W., McCue, S.W., Simpson, M.J., 2018. Extended logistic growth model for heterogeneous populations. Journal of Theoretical Biology 445, 51-61.

[18] Johnston, S.T., Simpson, M.J., McElwain, D.L.S., 2014. How much information can be obtained from tracking the position of the leading edge in a scratch assay? Journal of the Royal Society Interface 11, 20140325.

[19] Kaighn, M.E., Narayan, K.S., Ohnuki, Y., Lechner, J.F., Jones, L.W., 1979.

Establishment and characterization of a human prostatic carcinoma cell line (PC-3). Investigative urology 17(1), 16-23.

[20] Maini, P.K., McElwain, D.L.S., Leavesley, D., 2004. Travelling waves in a wound healing assay. Applied Mathematics Letters 17(5), 575-580.

[21] Maini, P.K., McElwain, D.L.S., Leavesley, D., 2004. Traveling wave model to interpret a wound-healing cell migration assay for human peritoneal mesothelial cells. Tissue Engineering 10(3/4), 475-482.

[22] Matsiaka, O.M., Penington, C.J., Baker, R.E., Simpson, M.J., 2017. Continuum approximations for lattice-free multi-species models of collective cell migration. Journal of Theoretical Biology 422, 1-11.

[23] Matsiaka, O.M., Penington, C.J., Baker, R.E., Simpson, M.J., 2018. Discrete and continuum approximations for collective cell migration in a scratch assay with cell size dynamics. Bulletin of Mathematical Biology 80(4), 738-757.

[24] Matsiaka, O.M., Baker, R.E., Simpson, M.J., 2019. Mechanistic and experimental models of cell migration reveal the importance of intercellular interactions in cell invasion. Biomedical Physics and Engineering Express. In press https://doi.org/ 10.1088/2057- 1976/ab1b01.

[25] Middleton, A.M., 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.

[26] Murray, P.J., Edwards, C.M., Tindall, M.J., Maini, P.K., 2009. From a discrete to a continuum model of cell dynamics in one dimension. Physical Review E 80(3), 031912.

[27] Nardini, J.T., Chapnick, D.A., Liu, X., Bortz, D.M., 2016 Modeling keratinocyte wound healing dynamics: Cell-cell adhesion promotes sustained collective migration. Journal of Theoretical Biology 400, 103-117.

[28] Newman, T.J., Grima, R., 2004. Many-body theory of chemotactic cell-cell interactions. Physical Review E 70, 051916.

[29] O'Dea, R.D., King, J.R., 2012. Continuum limits of pattern formation in hexagonal-cell monolayers. Journal of Mathematical Biology 64, 579-610.

[30] Osborne, J.M., Fletcher, A.G., Pitt-Francis, J.M., Maini, P.K., Gavaghan, D.J., 2017. Comparing individual-based approaches to modelling the self-organization of multicellular tissues. PLOS Computational Biology 13(2), e1005387.

[31] Painter, K.J., Sherratt, J.A., 2003. Modelling the movement of interacting cell populations. Journal of Theoretical Biology 225(3), 327-339.

[32] Pozzobon, V., Perre, P., 2018. Han's model parameters for microalgae grown under intermittent illumination: Determined using particle swarm optimization. Journal of Theoretical Biology 437, 29-35.

[33] Sadeghi, H.M., Seitz, B., Hayashi, S., LaBree, L., McDonnell, P.J., 1998. In vitro effects of Mitomycin-C on human keratocytes. Journal of Refractive Surgery 14, 534-40.

[34] Sarapata, E.A., de Pillis, L.G., 2010. A comparison and catalog of intrinsic tumor growth models. Bulletin of Mathematical Biology 76, 2010-2024.

[35] Savla, U., Olson, L.E., Waters, C.M., 2004. Mathematical modeling of airway epithelial wound closure during cyclic mechanical strain. Journal of Applied Physiology 96, 566-574.

[36] Sengers, B.A., Please, C.P., Oreffo, R.O.C., 2007. Experimental characterization and computational modelling of two-dimensional cell spreading for skeletal regeneration. Journal of the Royal Society Interface 4, 1107-1117.

[37] Sepulveda, N., Petitjean, L., Cochet, O., Grasland-Mongrain, E., Silberzan, P., Hakim, V., 2016. Collective cell motion in an epithelial sheet can be quantitatively

described by a stochastic interacting particle model. PLOS Computational Biology 9(3), e1002944

[38] Shah, E.T., Upadhyaya, A., Philp, L.K., Tang, T., Skalamera, D., Gunter, J., Nelson, C.C., Williams, E.D., Hollier, B.G., 2016. Repositioning "old" drugs for new causes: identifying new inhibitors of prostate cancer cell migration and invasion. Clinical & Experimental Metastasis 33, 385-399.

[39] Sheardown, H., Cheng, Y.-L., 1995. Mechanisms of corneal epithelial wound healing. Chemical Engineering Science 51(19), 4517-4529.

[40] Sherratt, J.A., Murray, J.D., 1990. Models of epidermal wound healing. Proceedings of the Royal Society B 241, 29-36.

[41] Sundstrom, A., Bar-Sagi, D., Mishra, B., 2016. Simulating heterogeneous tumor cell populations. PLoS ONE 11(12), e0168984.

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

[43] Simpson, M.J., Haridas P., McElwain D.L.S., 2014. Do pioneer cells exist? PLoS ONE 9(1), e85488.

[44] Simpson M.J., Jin, W., Vittadello, S.T., Tambyah, T.A., Ryan, J.M., Gunasingh, G., Haass, N.K., McCue, S.W., 2018. Stochastic models of cell invasion with fluorescent cell cycle indicators. Physica A 510, 375-386.

[45] Tersoff, J., 1988. New empirical approach for the structure and energy of covalent systems. Physical Review B 37(12), 6991-7000.

[46] ThermoFisher Scientific. Available from: https://www.thermofisher.com/au/en/home/life-science/cell-

565 culture/mammalian-cell-culture/reagents/trypsin/tryple-express.html (Accessed:
566 May, 2019).

567 [47] Treloar, K.K., Simpson, M.J. 2013. Sensitivity of edge detection methods for
568 quantifying cell migration assays. PLoS ONE 8(6), e67389.

569 [48] Tremel, A., Cai, A., Tirtaatmadja, N., Hughes, B.D., Stevens, G.W., Landman,
570 K.A., O'Connor, A.J., 2009. Cell migration and proliferation during monolayer
571 formation and wound healing. Chemical Engineering Science 64, 247-253.

572 [49] Warne, D.J., Baker, R.E., Simpson, M.J., 2019. Simulation and inference
573 algorithms for stochastic biochemical reaction networks: from basic concepts to
574 state-of-the-art. Journal of the Royal Society Interface 16, 20180943.

575 [50] Wise, S.M., Lowengrub, J.S., Frieboes, H.B., Cristini, V., 2008. Three-
576 dimensional multispecies nonlinear tumor growth–I Model and numerical method.
577 Journal of Theoretical Biology 253(3), 524-543.