New mathematical models for cell biology assays incorporating realistic cell size dynamics

A THESIS SUBMITTED TO THE SCIENCE AND ENGINEERING FACULTY IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Oleksii Matsiaka

School of Mathematical Sciences
Science and Engineering Faculty
Queensland University of Technology

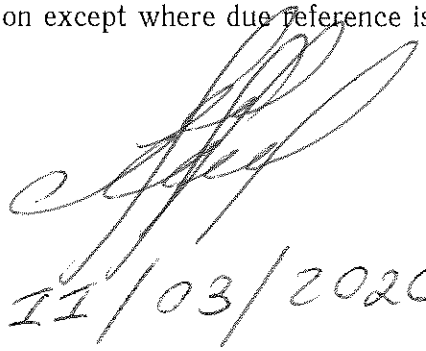
2020

Copyright in Relation to This Thesis

© Copyright 2020 by Oleksii Matsiaka *B.Sc., M.Sc.* All rights reserved.

Statement of Original Authorship

The work contained in this thesis has not been previously submitted to meet requirements for an award at this or any other higher education institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made.

Signature 

Date 11/03/2020

Abstract

The collective migration of cells involves many competing and interplaying mechanisms, such as undirected random motility, proliferation, and biochemical clues. However, an exact contribution of those often remains unclear. In our work we explore the role of previously overlooked cell motility mechanisms caused by the dynamical change in the cell size. Researchers routinely assume that the cell size remains constant throughout the experiment, which allows the use of fairly straightforward mathematical techniques. In our work, we include the dynamical cell size in the discrete and continuum models of cell migration and apply novel experimental data to provide insights into the role of cell size in the cell front expansion. Our work suggests that ignoring the dynamical cell size may result in a poor prediction of the population-level behaviour. Similarly, in the literature it is normally assumed that the cell culture is uniform and contains cells of constant properties. As such, the single-species models are routinely applied to describe and parametrize cell population. However, it has been previously shown that innate intrinsic differences can be even within the same cell line. We explore the experimentally motivated heterogeneity in cell sizes, as well as the heterogeneity in diffusivity and cell-to-cell interaction strength. Numerical simulations demonstrate that, only for certain types of heterogeneity single-species model performs well, while the multi-species models might be required for other types of heterogeneity.

Acknowledgments

Foremost, I would like to say thank you to my supervisor Professor Matthew Simpson for his continuous support and guidance throughout my research journey. You have my sincere gratitude for the years of mentoring and dedication.

I would like to thank Doctor Catherine Penington and Professor Ruth Baker for their assistance over the years of my study. I also appreciate the resources of High Performance Computing center at Queensland University of Technology (QUT). Also I would like to say thank you to all staff in the School of Mathematical Sciences, QUT.

I thank ANZIAM for providing financial assistance to travel in Australia and New Zealand and present my reasearch to a wider community. I would like to thank Australian Research Council for funding my PhD (FT130100148, DP170100474), the Institute of Health and Biomedical Research at QUT, and QUT for providing fee waiver scholarship.

Finally, I would like to express gratitude to friends, family, and my wife Amy for their help and support.

List of publications

1. **Matsiaka Oleksii**, Penington Catherine, Baker Ruth, Simpson Matthew. Continuum approximations for lattice-free multi-species models of collective cell migration. *Journal of Theoretical Biology*, vol. 422, p. 1-11, 2017.

2. **Matsiaka Oleksii**, Penington Catherine, Baker Ruth, Simpson Matthew. Discrete and continuum approximations for collective cell migration in a scratch assay with cell size dynamics *Bulletin of Mathematical Biology*, vol. 80, p. 738-757, 2018.

3. **Matsiaka Oleksii**, Baker Ruth, Shah Esha, Simpson Matthew. Mechanistic and experimental models of cell migration reveal the importance of intercellular interactions in cell invasion. *Biomedical Physics & Engineering Express*, vol. 5, 045009, 2019.

4. **Matsiaka Oleksii**, Baker Ruth, Simpson Matthew. Continuum descriptions of spatial spreading for heterogeneous cell populations: theory and experiment. *Journal of Theoretical Biology*, vol. 482, 109997, 2019.

Contents

Abstract i

Acknowledgments iii

List of publications v

1 Introduction 1

1.1 Background 1

1.2 Research questions 5

1.3 Objectives and outcomes 7

1.4 Structure of the thesis 10

1.5 Statement of joint authorship 12

1.5.1 Chapter 2: Continuum approximations for lattice-free multi-species models of collective cell migration 12

1.5.2 Chapter 3: Discrete and continuum approximations for collective cell migration in a scratch assay with cell size dynamics 12

1.5.3 Chapter 4: Mechanistic and experimental models of cell migration reveal the importance of cell-to-cell pushing in cell invasion . . . 13

1.5.4 Chapter 5: Continuum descriptions of spatial spreading for heterogeneous cell populations: theory and experiment 13

2 Continuum approximations for lattice-free multi-species models of collective cell migration 15

2.1 Abstract 15

2.2 Introduction 16

2.3 Discrete model 17

2.4 Results and discussion 20

2.4.1 Mathematical model for an arbitrary number of subpopulations . . 20

2.4.2 Application to monoculture experiments, $K = 1$ 29

2.4.3 Application to co-culture experiments, $K = 2$ 31

2.4.4 Parameter sensitivity 37

2.5 Conclusions 38

2A Additional results for Chapter 2 **41**

2A.1 Derivation of one- and two-cell PDF and corresponding Fokker-Planck equations 41

2A.2 Additional results for alternative parameter choices 43

2A.3 Discretisation scheme for the MFA model 48

2A.4 Discretisation scheme for the KSA model 49

3 Discrete and continuum approximations for collective cell migration in a scratch assay with cell size dynamics **51**

3.1 Abstract 51

3.2 Introduction 52

3.3 Methods 56

3.3.1 Langevin stochastic model 56

3.3.2 Continuum description 59

3.4 Results and discussion 63

3.5 Conclusions 70

3A Additional results for Chapter 3 **73**

3A.1 Comparing results from the two-dimensional and one-dimensional stochastic models 73

3A.2 Derivation of the $p_1(x,t)$ and $p_2(x,y,t)$ governing equations for the one-dimensional model 75

3A.3 Discretisation scheme for the one-dimensional MFA model 79

3A.4 Discretisation scheme for the one-dimensional KSA model 80

4 Mechanistic and experimental models of cell migration reveal the importance of cell-to-cell pushing in cell invasion **83**

4.1 Abstract 83

4.2 Introduction 84

4.3 Materials and methods 88

4.3.1 IncuCyte ZOOM™ experimental data 88

4.3.2 Image analysis 90

4.3.3 Discrete stochastic model 94

4.4 Results and Discussion 99

4.5 Conclusions 106

4A Additional results for Chapter 4 **109**
4A.1 Images of IncuCyte ZOOM™ scratch assay experiments A, B, C, and D at times t = 0, 12, 24, 36, and 48 h. . . . 109
4A.2 Cell counts for experiments A, B, C, and D at times t = 0, 12, 24, 36, and 48 h. . . . 110
4A.3 Cell density distributions for experiments A, B, C, and D at times t = 0, 12, 24, 36, and 48 h. . . . 111
4A.4 Additional results for Models I-IV 112

5 Continuum descriptions of spatial spreading for heterogeneous cell populations: theory and experiment **119**
5.1 Abstract 119
5.2 Introduction 120
5.3 Experimental data 123
5.4 Mathematical model 124
5.5 Results and Discussion 129
5.6 Conclusions 137

5A Additional results for Chapter 5 **139**
5A.1 Heterogeneity in the cell size assessed using density profile data 139
5A.2 Heterogeneity in the interaction force, f_0 142
5A.3 Heterogeneity in the diffusivity, D 147
5A.4 Discretisation scheme for the single-species homogeneous model and heterogeneous three-species model 152

6 Conclusions **155**
6.1 Summary of the research 155
6.2 Future work 158

Literature cited **161**

Chapter 1

Introduction

1.1 Background

Different types of cells in real biological systems are not isolated from each other but rather form complex aggregates, such as tissues and organs. Interactions between cells of different sizes and functions play a crucial role in most biophysical processes, such as wound healing, tumour growth, and tissue development to name a few (Weinberg, 2009; Weijer, 2009; Rognoni and Watt, 2018; Bocci et al., 2019). The variations between properties of individual cells are observed even within the same cell line (Hastings and Franks, 1983). We refer to these innate differences as population *heterogeneity*. Revealing and quantifying driving mechanisms behind behaviour of heterogeneous populations of cells is crucial for our understanding how complex biological environments function. Figure 1.1 illustrates a typical example of heterogeneous cell population in the form of a co-culture of two distinct cancer cell types: fibrosarcoma and adenocarcina cells. After visually inspecting the image one can notice an immense range of cell sizes in this population. Additionally, different cell types can have different migratory and mechanical properties that are much harder to quantify in comparison to the heterogeneity in cell sizes. Biological systems such as tissues and organs can involve interactions of dozens of different cell types which makes the problem of quantifying the role of heterogeneity even harder. As such, heterogeneity in cell populations yet remains major problem in the contemporary mathematical biology (Menon et al., 2018; Rutter et al., 2018).

The standard approach to emulate real biological systems utilizes *in vitro* cell culture experiments (Edmondson et al., 2014). These experiments often demonstrate significant changes in cell properties over time that constitutes an additional complication on top of the innate heterogeneity. For example, one of the most evident examples is cells increasing their size as they go through different stages of the cell cycle (Amodeo and Skotheim, 2016). Figure 1.2 shows an example of the two-dimensional cell culture experiment where PC-3 cancer cells appear to double in size over the course of an

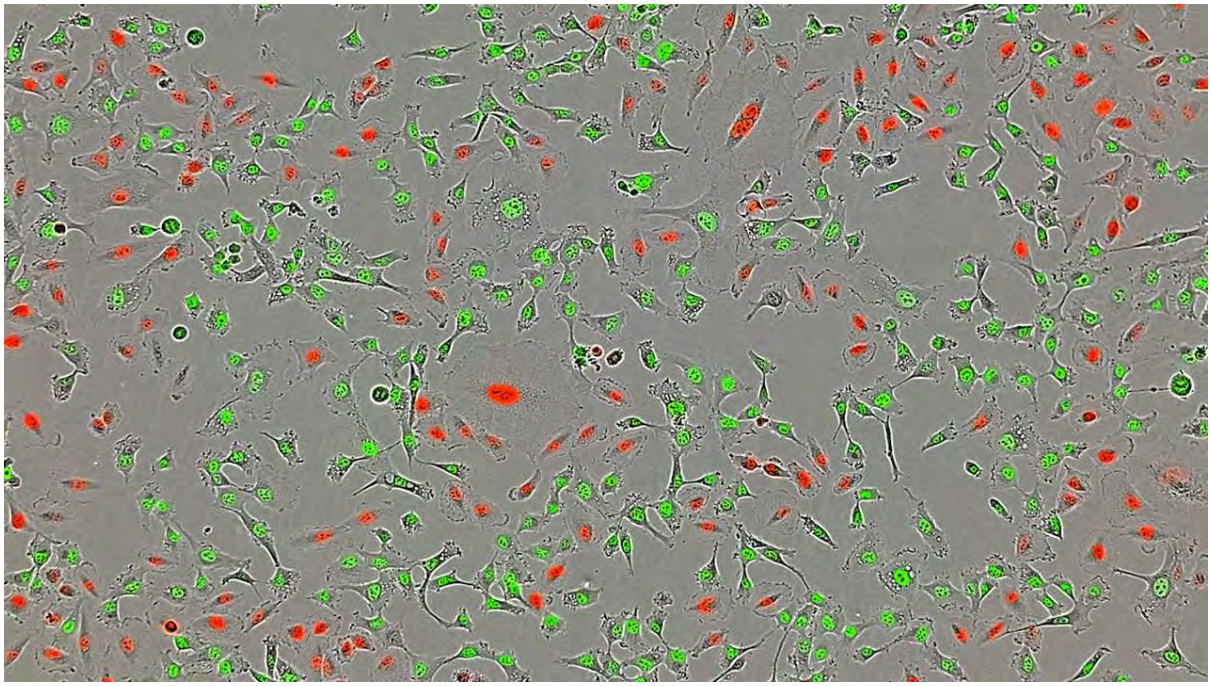


Figure 1.1: Heterogeneous population of cells containing HT1080 fibrosarcoma cells (green nuclei) in co-culture with A549 adenocarcina cells (red nuclei). Fluorescent nuclei are automatically detected using the IncuCyte ZOOM Live Cell Imaging System. Cancer cell nuclei were labelled using the IncuCyte NucLight Lentivirus Reagents. *Credit: Essen BioScience.*

experiment. These dynamical changes in such seemingly fundamental property as the cell size are not normally reflected in the mathematical models.

There is a variety of two-dimensional and three-dimensional cell culture assays of different geometries designed to study cell-to-cell interactions, drug discovery, tissue regeneration, and disease progression (Riss, 2005; Goers et al., 2014; Edmondson et al., 2014). In the most general sense, *in vitro* assay is a cell cultivation set-up, in which population of cells is grown either on a flat substrate in the case of two-dimensional assays or on the more realistic three-dimensional architectures (Griffith and Swartz, 2006; Weigelt et al., 2014). To date, the most popular choice of *in vitro* assays is a traditional two-dimensional monolayer cultures, including circular barrier assays (De Leso and Pei, 2018), microfluidic assays (Regnault et al., 2018), scratch assays (Jin et al., 2016a), and IncuCyte ZOOM™ assays (Essen BioScience, Ann Arbor, MI). Even though two-dimensional cell assays have been recognised as simple and cost-effective tool for biological research, they do not adequately model the real *in vivo* environment where cells are embedded into extracellular matrix (ECM) (Rosso et al., 2004; Edmondson et al., 2014). This drawback may become especially acute if we consider a culture of cells where subpopulations may differ in many aspects which may range from minor differing in phenotype to being different species. In contrast to two-dimensional cultures, three-dimensional cultures provide more realistic modelling of the *in vivo* environment due the fact that cells form aggregates within a matrix which more

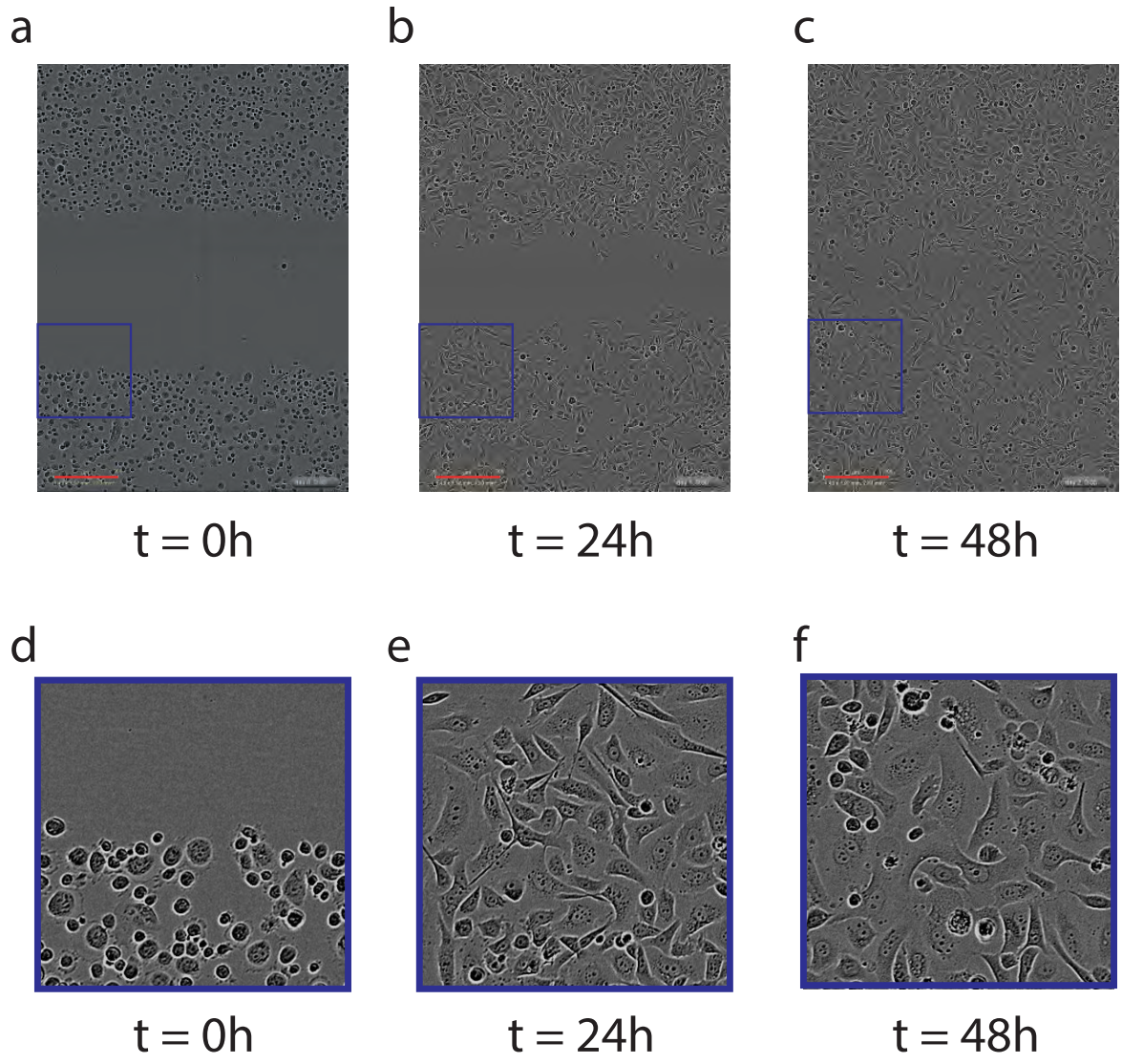


Figure 1.2: (a)-(c) IncuCyte ZOOM™ images of the two-dimensional cell culture containing PC-3 prostate cancer cells (Kaighn et al., 1979) for $t = 0, 24$, and $48\,\mathrm{h}$. A red solid line indicates a scale bar corresponding to $300\ \mu\mathrm{m}$. (d)-(f) Zoomed-in images denoted in blue squares in Figure 1.2 (a)-(c).

closely mimic real cell-ECM relationships (Edmondson et al., 2014; Holle et al., 2016).

These *in vitro* experiments when combined with appropriate mathematical models and parameter estimation techniques may provide powerful tools to describe and predict behaviour of heterogeneous populations of cells. For example, it was shown (Eves et al., 2003) that migration of malignant melanoma cells is enhanced in the presence of healthy skin cells that implies special interaction between these two types of cells. In this example it becomes clear that in order to describe invasive nature of melanoma we should take into account interactions between normal and melanoma skin cells. Different subpopulations of cells are not necessarily distinct but may be just labelled to allow us track individual cells (Progatzky et al., 2013), or different functional cell groups (Trinh et al., 2017).

In a literature, the convenient and intuitive way to simulate a population of cells is to invoke the discrete modelling framework, where cells are represented as discrete agents with behaviour that is determined by predefined rules (Newman and Grima, 2004; Galle et al., 2005; Gardiner et al., 2015). Discrete models can be classified into two major groups depending on the physical domain that contains agents (Osborne et al., 2017). Building a domain on the discrete lattice leads to the on-lattice and cellular automata models which provide convenient tools to mimic key motility mechanisms and allow us to obtain snapshots similar to real experiments. In this framework, cells are treated as constant-size agents that occupy one or a few lattice sites and move only along coordinate axes. Since the cell size is associated with the lattice grid spacing, it is problematic to model population of cells that change in size. Previously, the dynamical cell size has been incorporated into the two-species lattice-based model, where each agent of type 1 is allowed to grow into an agent of type 2 doubling its size (Binder and Simpson, 2016). The cellular Potts model is more advanced type of lattice-based models where cells are modelled as spatially extended objects composed of many discrete agents (Szabó and Merks, 2013). The lattice-based models have been used to simulate multi-species populations of cells, tumor growth, and angiogenesis (Delgado-SanMartin et al., 2017; Pillay et al., 2018), as well as a variety of *in vitro* experiments of different geometries, such as circular barrier assays (Treloar et al., 2014) or scratch assays (Khain et al., 2006). The main drawback of lattice-based models is that cells do not move on a lattice in reality. Although the on-lattice models may be improved by imposing finer lattice grid, we do not use them in this thesis because of their limitations when applied to modelling dynamical cell size.

On the other hand, the off-lattice approach does not feature an artificial spatial discretisation. This provides an improved level of realism in comparison to the lattice-based models (Drasdo et al., 1995; Drasdo and Höhme, 2005; Johnston et al., 2013; Osborne et al., 2017). Consequently, cells are allowed to move in any direction bounded only by naturally imposed boundaries which exist in real experiments. We choose the off-lattice model where migration of the population of cells is governed by the system of Langevin stochastic differential equations (SDEs) (Middleton et al., 2014):

$$\frac{\mathrm{d}\vec{x}_i}{\mathrm{d}t} = \sum_{j \neq i} \vec{F}_{ij} + \vec{\xi}_i, \qquad i = 1, ..., N, \tag{1.1}$$

where index i denotes agent label, $\vec{x}_i$ is the position of the ith agent, $\vec{\xi}_i$ is the stochastic undirected force, and $\vec{F}_{ij}$ is the deterministic interaction forces between agents i and j. Normally, the interaction force, $\vec{F}_{ij}$, is assumed to be a function of distance between agents that suitably mimics certain phenomenological aspects of cell-to-cell interactions (Newman and Grima, 2004; Middleton et al., 2014). In this thesis, we introduce the

novel approach that features the time-dependent cell-to-cell interaction force to model populations of cells that increase in size.

An alternative approach is to invoke a continuum description which is based on applying partial differential equations (PDE) to tissue-level data, such as density profiles. Routinely PDE-based models are obtained in the continuum limit of a grid size tending to zero transforming discrete conservation laws into the continuous form (Murray et al., 2009; Field and Tough, 2010). One of the most popular models for describing populations of cells is the Fisher–KPP (Fisher–Kolmogorov–Petrovsky–Piskunov) reaction-diffusion equation,

$$\frac{\partial p}{\partial t} = \underbrace{D\frac{\partial^2 p}{\partial x^2}}_{\text{diffusion}} + \underbrace{\lambda p\Big(1 - \frac{p}{K}\Big)}_{\text{reaction}}, \tag{1.2}$$

where D is the cell diffusivity, λ is the proliferation rate, K is the carrying capacity density, and the cell density, $p(x,t)$, is a function of space x and time t. The diffusion term represents undirected random motility of cells, and the latter term incorporates cell proliferation.

The Fisher–KPP type models do not include cell size parameter and therefore cannot be used for describing populations of cells with dynamical cell size. Nevertheless, the Fisher–KPP model is used to estimate individual-level parameters, such as diffusion coefficient D and proliferation rate λ (Johnston et al., 2015), and to describe multi-species cell culture experiments (Simpson et al., 2007).

We note that, Fisher–KPP model does not take into account the cell-to-cell adhesive properties. This implies that we are not able to describe the cellular behaviour caused by intercellular interactions. While some cells in the human body are highly adhesive and form strong confluent layer, such as epithelial skin cells, other cells exhibit much weaker bonds with closest neighbours (mesenchymal cells). It is therefore of interest to use alternative continuum description to incorporate the cell-to-cell adhesion.

1.2 Research questions

This thesis explores and answers the following questions:

1. **How can we develop models of collective cell migration and adhesion to describe many, potentially distinct, subpopulations of cells?**

   In the literature when applying mathematical models to describe *in vitro* cell assay experiments, it is common practice to assume that all cells within population have constant properties, such as cell sizes and diffusivities. Some individual cell properties, for example diffusivity, can be fairly hard to measure for every

single population member. On the other hand, cell size can be quantified using semi-automatic image processing tools. In this thesis, we refer to these intrinsic differences as population *heterogeneity* and allow population of cells to contain many different subpopulations with different properties. Many popular mathematical descriptions, discrete or continuous, do not account for innate heterogeneity and routinely use models with only one set of parameters accounting for averaged population-wide characteristics. The way to address this gap is to develop mathematical model of cell motility and adhesion able to parametrise many, potentially distinct, subpopulations. We focus to do so in the first stage of our project (Chapter 2).

2. **How can we incorporate dynamically changing cell size into models of collective cell migration and adhesion?**

   Many cell assay experiments indicate that the size of individual cells can change significantly throughout the duration of the experiment, sometimes doubling over 48-72 hours. Most mathematical descriptions used to model cell assays have limited capacity to deal with dynamical cell size. For example, in popular lattice-based models cells are represented as an agent that moves on a discrete lattice and the cell size is associated with the lattice spacing. As such, the cell size cannot be set to change continuously on the lattice-based domain. Some of the continuous models, such as models based on the Fisher–KPP equation, do not contain cell size as a parameter, thus modelling cells as point objects. We address this issue by extending discrete and continuous models developed in Chapter 2 to include dynamical cell size (Chapter 3).

3. **Can new models provide insights into previously overlooked cell motility mechanisms, such as cell-to-cell pushing?**

   Understanding and quantifying the contribution of different mechanisms that drive the movement of cell fronts is of great interest. Cell fronts are normally occurring in a wide range of biological processes, such as wound healing, cancer invasion, and tumour growth. A precise contribution of different individual-level mechanisms to a population-level behaviour remains an open question. The role of diffusivity and adhesion is well studied and is broadly discussed in the literature. On the other hand, there are few insights into cell-to-cell pushing and its role in the progression of the advancing cell fronts. To address this question mathematical model of cell motility and adhesion with dynamical cell size is connected to a novel experimental data, and contribution of different cell motility mechanisms is assessed in Chapter 4.

4. **Can a cell population consisting of many distinct subpopulations be de-**

scribed by a simple single-species model of cell motility and adhesion?

Cell populations can exhibit significant level of heterogeneity even within the same cell line. The innate differences in cell sizes, diffusivities, or mechanical properties of cells can heavily influence the ability of any given model to describe and predict the collective-level behaviour of a cell population. Intuitive approach to account for heterogeneity in a system of cells is to invoke multi-species framework where whole population is divided into subpopulations with relatively constant properties. The interesting question is – *What differences in a cell population are really important?* In Chapter 5 we explore the heterogeneity in cell sizes and both random and directed cell motilities to explore the ability of a single-species model to predict behaviour of the heterogeneous cell population.

1.3 Objectives and outcomes

The primary goal of this thesis is to develop and validate new mathematical models of cell motility and adhesion that incorporate realistic cell size dynamics.

The specific objectives of this thesis are as follows:

1. Develop novel discrete and continuum multi-species models of cell motility and adhesion for arbitrary number of subpopulations.

2. Extend discrete and continuum models developed in Objective 1 to include dynamical cell size.

3. Validate new models using novel experimental data that demonstrate significant change in cell sizes and quantify the role of the cell-to-cell pushing.

4. Investigate the ability of a single-species model to describe behaviour of a heterogeneous population of cells.

This thesis is presented by publication and incorporates four publications in peer reviewed journals. Three of these publications are published in Q1 journals (Journal of Theoretical Biology and Bulletin of Mathematical Biology), one is published in Q2 journal (Biomedical Physics & Engineering Express). The PhD candidate is the first author and contributed significantly to all four publications. The work presented in this thesis fulfils Queensland University of Technology criteria for the thesis by publication.

This thesis incorporates following publications:

Matsiaka Oleksii, Penington Catherine, Baker Ruth, Simpson Matthew. Continuum approximations for lattice-free multi-species models of collective cell migration. *Journal of Theoretical Biology*, vol. 422, p. 1-11, 2017.

Abstract Cell migration within tissues involves the interaction of many cells from distinct subpopulations. In this chapter, 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.

Matsiaka Oleksii, Penington Catherine, Baker Ruth, Simpson Matthew. Discrete and continuum approximations for collective cell migration in a scratch assay with cell size dynamics *Bulletin of Mathematical Biology*, vol. 80, p. 738-757, 2018.

Abstract Scratch assays are routinely used to study the collective spreading of cell populations. In general, the rate at which a population of cells spreads is driven by the combined effects of cell migration and proliferation. To examine the effects of cell migration separately from the effects of cell proliferation, scratch assays are often performed after treating the cells with a drug that inhibits proliferation. Mitomycin-C is a drug that is commonly used to suppress cell proliferation in this context. However, in addition to suppressing cell proliferation, Mitomycin-C also causes cells to change size during the experiment, as each cell in the population approximately doubles in size as a result of treatment. Therefore, to describe a scratch assay that incorporates the effects of cell-to-cell crowding, cell-to-cell adhesion, and dynamic changes in cell size, we present a new stochastic model that incorporates these mechanisms. Our agent-based stochastic model takes the form of a system of Langevin equations that is the system of stochastic differential equations governing the evolution of the population of agents. We incorporate a time-dependent interaction force that is used to mimic the dynamic increase in size of the agents. To provide a mathematical description of the

average behaviour of the stochastic model we present continuum limit descriptions using both a standard mean-field approximation, and a more sophisticated moment dynamics approximation that accounts for the density of agents and density of pairs of agents in the stochastic model. Comparing the accuracy of the two continuum descriptions for a typical scratch assay geometry shows that the incorporation of agent growth in the system is associated with a decrease in accuracy of the standard mean-field description. In contrast, the moment dynamics description provides a more accurate prediction of the evolution of the scratch assay when the increase in size of individual agents is included in the model.

Matsiaka Oleksii, Baker Ruth, Shah Esha, Simpson Matthew. Mechanistic and experimental models of cell migration reveal the importance of intercellular interactions in cell invasion. *Biomedical Physics & Engineering Express*, vol. 5, 045009, 2019.

Abstract Moving fronts of cells are essential for development, repair and disease progression. Therefore, understanding and quantifying the details of the mechanisms that drive the movement of cell fronts is of wide interest. Quantitatively identifying the role of intercellular interactions, and in particular the role of cell pushing, remains an open question. In this work, we report a combined experimental-modelling approach showing that intercellular interactions contribute significantly to the spatial spreading of a population of cells. We use a novel experimental data set with PC-3 prostate cancer cells that have been pretreated with Mitomycin-C to suppress proliferation. This allows us to experimentally separate the effects of cell migration from cell proliferation, thereby enabling us to focus on the migration process in detail as the population of cells recolonizes an initially-vacant region in a series of two-dimensional experiments. We quantitatively model the experiments using a stochastic modelling framework, based on Langevin dynamics, which explicitly incorporates random motility and various intercellular forces including: (i) long range attraction (adhesion); and (ii) finite size effects that drive short range repulsion (pushing). Quantitatively comparing the ability of this model to describe the experimentally observed population-level behaviour provides us with quantitative insight into the roles of random motility and intercellular interactions. To quantify the mechanisms at play, we calibrate the stochastic model to match experimental cell density profiles to obtain estimates of cell diffusivity, D, and the amplitude of intercellular forces, f_0. Our analysis shows that taking a standard modelling approach which ignores intercellular forces provides a poor match to the experimental data whereas incorporating intercellular forces, including short-range pushing and longer range attraction, leads to a faithful representation of the experimental observations. These results demonstrate a significant role of cell pushing during cell front movement and invasion.

Matsiaka Oleksii, Baker Ruth, Simpson Matthew. Continuum descriptions of spatial spreading for heterogeneous cell populations: theory and experiment. *Journal of Theoretical Biology*, vol. 482, 109997, 2019.

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 monotonically decreasing and uniform distributions 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 find that the simpler single-species homogeneous model performs relatively poorly when applied to non-monotonic and monotonically increasing distributions of cell sizes. Additional results for heterogeneity in parameters describing both undirected and directed cell migration are also considered, and we find that similar results apply.

1.4 Structure of the thesis

Every publication mentioned in Section 1.3 corresponds to a chapter. Since every publication is an independent piece of work there is some overlap between different chapters. For example, numerical scheme used to solve continuum models is presented in both Chapter 2A and Chapter 3A. Team members have read statement of join authorship in Section 1.5 and agreed for inclusion of publications in the thesis.

In Chapter 2 we introduce multi-species discrete and continuum models of cell migration and adhesion that allow us to describe populations of cells consisting of many

subpopulations. In the literature discrete approach is normally invoked to describe populations of cells on the individual-level scale. Here we choose to work with an off-lattice model based on the Langevin stochastic differential equations (SDEs). On the other hand, continuum models are more suitable to describe systems of cells on the population-level scale. We develop the hierarchy of integro-partial differential equations (IPDEs) as the limit of the corresponding discrete model, and then apply two continuum approximations to simplify and solve it. Specifically, mean field approximation and moment dynamics approximation are invoked and accuracy of both approximations is estimated by matching solutions of the discrete model and IPDEs-based model. Different parameter regimes are explored in the attempt to identify suitable continuum approximation.

In Chapter 3 we extend models presented in Chapter 2 to include dynamical change in cell size. Here we decide to work with a fundamental case of one subpopulation only and focus on the effects of increase in cell size. A similar analysis is performed in which the accuracy of mean field approximation and moment dynamics approximation is explored.

In Chapter 4 we explore the role of the cell-to-cell pushing in the cell front spreading. Cell fronts are observed in many biological processes, such as wound healing, morphogenesis, and tumour growth. Understanding the driving mechanisms behind frontal expansion plays vital clinical role in how we approach conditions that involve cell fronts. Previously, cell front spreading has been mostly associated with undirected motility and proliferation, while the roles of mechanical interactions and pushing have received little attention. Here we apply the discrete model of cell motility and adhesion developed in Chapter 3 to a novel experimental data of cell cultures with cells that demonstrate significant increase in size. We show that excluding of the cell-to-cell pushing and dynamical cell size from the model can lead to a poor prediction of population behaviour.

In Chapter 5 multi-species model developed in Chapter 2 is used to provide insights into the role of heterogeneity in cell populations. Specifically, we are interested in the ability of simple, computationally straightforward, single-species continuum model to describe heterogeneous cell populations. We consider heterogeneity in cell sizes, directed and undirected cell migration, and find that for some forms of heterogeneity single-species model is able to accurately describe the evolution of the multi-species cell population.

Finally, in Chapter 6 we outline conclusions of our study and propose avenues for future research.

1.5 Statement of joint authorship

In this section we summarize the contributions of PhD candidate and coauthors to each publication. All coauthors have agreed to include publications in this thesis.

1.5.1 Chapter 2: Continuum approximations for lattice-free multi-species models of collective cell migration

The associated reference for this chapter is:
Matsiaka Oleksii, Penington Catherine, Baker Ruth, Simpson Matthew. Continuum approximations for lattice-free multi-species models of collective cell migration. *Journal of Theoretical Biology*, vol. 422, p. 1-11, 2017.

- **Matsiaka Oleksii (Candidate)** derived the continuum equations of cell motility and adhesion, performed numerical simulations and data analysis, made figures, compiled manuscript, and revised manuscript during peer-review process.
- Penington Catherine supervised the research, oversaw drafting and redrafting of the manuscript, and critically reviewed manuscript during peer-review process.
- Baker Ruth oversaw drafting of the manuscript before submitting for publication, and revised it during peer-review process.
- Simpson Matthew conceptualised the ideas for the manuscript, supervised the research, oversaw drafting and redrafting of the manuscript prior to publication, reviewed manuscript during peer-review process, compiled cover and revision letters and acted as corresponding author.

1.5.2 Chapter 3: Discrete and continuum approximations for collective cell migration in a scratch assay with cell size dynamics

The associated reference for this chapter is:
Matsiaka Oleksii, Penington Catherine, Baker Ruth, Simpson Matthew. Discrete and continuum approximations for collective cell migration in a scratch assay with cell size dynamics *Bulletin of Mathematical Biology*, vol. 80, p. 738-757, 2018.

- **Matsiaka Oleksii (Candidate)** derived the continuum equations of cell motility and adhesion, performed numerical simulations and data analysis, made figures, compiled manuscript, and revised manuscript during peer-review process.

- Penington Catherine supervised the research, oversaw drafting and redrafting of the manuscript, and critically reviewed manuscript during peer-review process.
- Baker Ruth oversaw drafting of the manuscript before submitting for publication, and revised it during peer-review process.
- Simpson Matthew conceptualised the ideas for the manuscript, supervised the research, oversaw drafting and redrafting of the manuscript prior to publication, reviewed manuscript during peer-review process, compiled cover and revision letters and acted as corresponding author.

1.5.3 Chapter 4: Mechanistic and experimental models of cell migration reveal the importance of cell-to-cell pushing in cell invasion

The associated reference for this chapter is:
Matsiaka Oleksii, Baker Ruth, Shah Esha, Simpson Matthew. Mechanistic and experimental models of cell migration reveal the importance of intercellular interactions in cell invasion. *Biomedical Physics & Engineering Express*, vol. 5, 045009, 2019.

- **Matsiaka Oleksii (Candidate)** jointly developed a concept, performed numerical simulations and data analysis, made figures, compiled manuscript, revised manuscript during peer-review process, and compiled revision letter.
- Baker Ruth oversaw drafting of the manuscript before submitting for publication, and revised it during peer-review process.
- Shah Esha performed IncuCyte^{TM} scratch assay experiments.
- Simpson Matthew jointly developed the concept for the manuscript, supervised the research, oversaw drafting and redrafting of the manuscript prior to publication, reviewed manuscript during peer-review process, compiled cover letter and acted as corresponding author.

1.5.4 Chapter 5: Continuum descriptions of spatial spreading for heterogeneous cell populations: theory and experiment

The associated reference for this chapter is:
Matsiaka Oleksii, Baker Ruth, Simpson Matthew. Continuum descriptions of spatial spreading for heterogeneous cell populations: theory and experiment. *Journal of Theoretical Biology*, vol. 482, 109997, 2019.

- **Matsiaka Oleksii (Candidate)** jointly developed a concept, performed numerical simulations and data analysis, made figures, compiled manuscript, and revised it during peer-review process.
- Baker Ruth oversaw drafting of the manuscript before submitting for publication, and revised it during peer-review process.
- Simpson Matthew jointly developed the concept for the manuscript, supervised the research, oversaw drafting and redrafting of the manuscript prior to submitting for publication and during peer-review process, and acted as corresponding author.

Chapter 2

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

A paper published in the *Journal of Theoretical Biology*.

Matsiaka Oleksii, Penington Catherine, Baker Ruth, Simpson Matthew. Continuum approximations for lattice-free multi-species models of collective cell migration. *Journal of Theoretical Biology*, vol. 422, p. 1-11, 2017.

2.1 Abstract

Cell migration within tissues involves the interaction of many cells from distinct subpopulations. In this chapter, 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.

2.2 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., 2004a).

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., 2009a; 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., 2013a) whereas the diameter of a typical skin cell is approximately 25 μm (Simpson et al., 2013a). In a model with both types of cells present, it is not possible to accommodate these differences in cell size 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., 2004a).

This chapter is organised in the following way. In Section 2.3 we describe the discrete model. In Section 2.4.1, 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 2.4.2-2.4.3 compare solutions of both continuum approximations and averaged discrete results for problems involving one and two interacting subpopulations, with additional comparisons presented in Chapter 2A. In Section 2.4.4 we investigate how the accuracy of the MFA and KSA approximations depends on the choice of model parameters. Finally, in Section 2.5, we summarise our work and highlight opportunities for future investigation.

2.3 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 2.1(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 2.5 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. Here we neglect inertial forces, invoke Stokes' law (Middleton et al., 2014), and, consequently, 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 $\zeta_i = \mu \xi_i$.

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., 2013a). 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

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

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 chapter

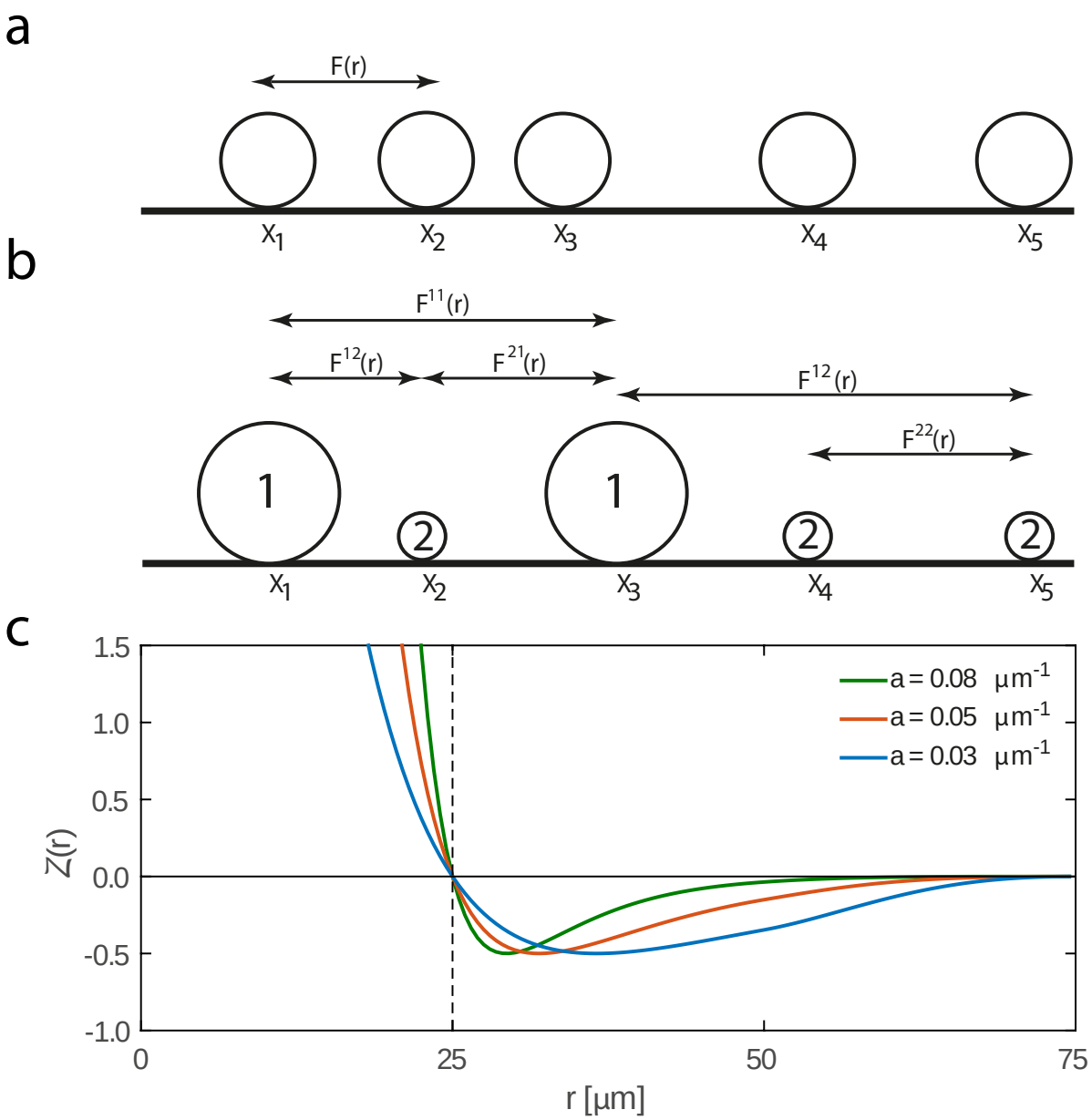


Figure 2.1: (a)-(b) Representative plot of single- and multi-species systems of cells, respectively. In (a) we show the intraspecies force, $F(r)$, and in (b) we show both intraspecies forces, $F^{11}(r)$ and $F^{22}(r)$, and interspecies forces, $F^{12}(r)$ and $F^{21}(r)$. Here, r is the distance between cells. (c) Dimensionless force law function $\mathcal{Z}(r)$, given by Equation (3.6), for various values of a. Here, $\delta = 25$ μm corresponds to a typical cell diameter.

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 2.1(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

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

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 2.1(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 chapter are for this particular choice of force law, it is straightforward to incorporate other choices of $\mathcal{Z}(r)$.

2.4 Results and discussion

2.4.1 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., 2006a; 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, 1976),

$$P_1^i(x,t) = \langle \delta^{(N)}(x - x_i(t))\rangle, \tag{2.7}$$

$$P_2^{ij}(x,y,t) = \langle \delta^{(N)}(x - x_i(t))\, \delta^{(N)}(y - y_j(t))\rangle, \tag{2.8}$$

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 (2.7) and (2.8) is given in Chapter 2A.

The time evolution of $P_1^i(x,t)$ is governed by a Fokker-Planck equation (Chapter 2A),

$$\frac{\partial P_1^i(x,t)}{\partial t} = D\Delta P_1^i(x,t) - \nabla\left(f_i\, P_1^i(x,t)\right), \tag{2.9}$$

which describes the motion of particles under the influence of random forces, proportional to the diffusivity, D, and directed drift forces, f_i. Here operator Δ is defined as $\Delta = \partial^2/\partial x^2$, and $\nabla = \partial/\partial x$. 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

$$f_i = \sum_{j\neq i} F_{ij}^{ll} \; + \sum_{k\neq l}^{K} \sum_{j\in k} F_{ij}^{lk}. \tag{2.10}$$

Combining Equations (2.7), (2.9) and (2.10), and taking the convolution of the interaction force and a δ-function centred at y_j, we obtain

$$\begin{aligned}\frac{\partial P_1^i(x,t)}{\partial t} &= D\Delta P_1^i(x,t) \\ &\quad - \nabla\Big\langle \sum_{j\in l, j\neq i} \int_\Omega F^{ll}(x_i - y)\,\delta(x - x_i(t))\,\delta(y - y_j(t))\,\mathrm{d}y \Big\rangle \\ &\quad - \sum_{k\neq l}^{K} \nabla\Big\langle \sum_{j\in k} \int_\Omega F^{lk}(x_i - y)\,\delta(x - x_i(t))\,\delta(y - y_j(t))\,\mathrm{d}y \Big\rangle,\end{aligned} \tag{2.11}$$

where Ω denotes the domain. The second and third terms on the right hand side of Equation (2.11) are advection terms that incorporate intraspecies and interspecies forces, respectively. Combining Equations (2.8) and (2.11), and interchanging summation and integration, we obtain

$$\begin{aligned}\frac{\partial P_1^i(x,t)}{\partial t} &= D\Delta P_1^i(x,t) - \nabla \int_\Omega F^{ll}(x-y) \sum_{j\in l, j\neq i} P_2^{ij}(x,y,t)\,\mathrm{d}y \\ &\quad - \sum_{k\neq l}^{K} \nabla \int_\Omega F^{lk}(x-y) \sum_{j\in k} P_2^{ij}(x,y,t)\,\mathrm{d}y,\end{aligned} \tag{2.12}$$

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,

$$p_1^l(x,t) = \frac{1}{n_l}\sum_{i\in\mathcal{L}} P_1^i(x,t), \tag{2.13}$$

$$p_2^{ll}(x,y,t) = \frac{1}{n_l(n_l-1)}\sum_{i\in\mathcal{L}}\sum_{j\in\mathcal{L},j\neq i} P_2^{ij}(x,y,t), \tag{2.14}$$

$$p_2^{lm}(x,y,t) = \frac{1}{n_l n_m}\sum_{i\in\mathcal{L}}\sum_{j\in\mathcal{L}} P_2^{ij}(x,y,t), \qquad l\neq m, \tag{2.15}$$

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 (2.12) and apply the definitions given in Equations (2.13)-(2.15). 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

$$\begin{aligned}\frac{\partial p_1^l(x,t)}{\partial t} = D\Delta p_1^l(x,t) - (n_l-1)\nabla\Big(\int_\Omega F^{ll}(x-y)\,p_2^{ll}(x,y,t)\,\mathrm{d}y\Big)\\ -\sum_{k\neq l}^{K} n_k\,\nabla\Big(\int_\Omega F^{lk}(x-y)\,p_2^{lk}(x,y,t)\,\mathrm{d}y\Big),\end{aligned} \tag{2.16}$$

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

$$p_1^{\text{total}}(x,t) = \frac{1}{N}\sum_{k=1}^{K} n_k\; p_1^k(x,t). \tag{2.17}$$

Equation (2.16) 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-Planck equation,

$$\frac{\partial P_2^{ij}(x,y,t)}{\partial t} = D\Delta P_2^{ij}(x,y,t) - \frac{\partial}{\partial x}\left(f_i\,P_2^{ij}(x,y,t)\right) - \frac{\partial}{\partial y}\left(f_j\,P_2^{ij}(x,y,t)\right), \tag{2.18}$$

where cells i and j both belong to subpopulation l, and operator Δ is defined as $\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2$. The forces f_i and f_j, applied to cells i and j, can be written as the sum of intraspecies and interspecies forces. For example, the force on an arbitrary cell z can be written as

$$f_z = \sum_{s\neq z} F_{zs}^{ll} + \sum_{k\neq l}^{K}\sum_{s\in k} F_{zs}^{lk}. \tag{2.19}$$

Adopting the interaction force law, Equation (2.4), using the definition of the two-cell

PDF as given by Equation (2.8), and evaluating the required convolutions, we can rewrite Equation (2.18) as

$$\begin{aligned}\frac{\partial P_2^{ij}(x,y,t)}{\partial t} &= D\Delta P_2^{ij}(x,y,t)\\ &\quad - \frac{\partial}{\partial x}\left\langle F^{ll}(x-y)\,\delta(x-x_i(t))\,\delta(y-y_j(t))\right\rangle\\ &\quad - \frac{\partial}{\partial y}\left\langle F^{ll}(y-x)\,\delta(x-x_i(t))\,\delta(y-y_j(t))\right\rangle\\ &\quad - \frac{\partial}{\partial x}\left\langle \sum_{g\in l, g\neq i,j}\int_\Omega F^{ll}(x-z)\,\delta(x-x_i(t))\,\delta(y-y_j(t))\,\delta(z-z_g(t))\,\mathrm{d}z\right\rangle\\ &\quad - \frac{\partial}{\partial y}\left\langle \sum_{g\in l, g\neq i,j}\int_\Omega F^{ll}(y-z)\,\delta(x-x_i(t))\,\delta(y-y_j(t))\,\delta(z-z_g(t))\,\mathrm{d}z\right\rangle\\ &\quad - \sum_{k\neq l}^{K}\frac{\partial}{\partial x}\left\langle \sum_{g\in k}\int_\Omega F^{lk}(x-z)\,\delta(x-x_i(t))\,\delta(y-y_j(t))\,\delta(z-z_g(t))\,\mathrm{d}z\right\rangle\\ &\quad - \sum_{k\neq l}^{K}\frac{\partial}{\partial y}\left\langle \sum_{g\in k}\int_\Omega F^{lk}(y-z)\,\delta(x-x_i(t))\,\delta(y-y_j(t))\,\delta(z-z_g(t))\,\mathrm{d}z\right\rangle, \end{aligned} \tag{2.20}$$

where the second and third terms on the right hand side of Equation (2.20) represent interactions between cells i and j, the fourth and fifth terms on the right hand side of Equation (2.20) 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 (2.20) represent interactions between cells i and j and cells in other subpopulations.

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

$$p_3^{lms}(x,y,z,t) = \frac{1}{n_l n_m n_s}\sum_{i\in l}\sum_{j\in m}\sum_{g\in s} P_3^{ijg}(x,y,z,t), \tag{2.21}$$

$$p_3^{lls}(x,y,z,t) = \frac{1}{n_l(n_l-1)n_s}\sum_{i\in l}\sum_{j\in l, j\neq i}\sum_{g\in s} P_3^{ijg}(x,y,z,t), \tag{2.22}$$

$$p_3^{lll}(x,y,z,t) = \frac{1}{n_l(n_l-1)(n_l-2)}\sum_{i\in l}\sum_{j\in l, j\neq i}\sum_{g\in l, g\neq i,j} P_3^{ijg}(x,y,z,t). \tag{2.23}$$

We therefore require a definition for the three particle PDF, $P_3^{ijg}(x,y,z,t)$, similar to Equation (2.8),

$$P_3^{ijg}(x,y,z,t) = \langle \delta(x-x_i(t))\,\delta(y-y_j(t))\,\delta(z-z_g(t))\rangle. \tag{2.24}$$

To proceed we divide Equation (2.20) by $n_l(n_l-1)$, and combine Equations (2.20)-(3A.21),

summing over the indices i and j, to obtain an expression for the evolution of $p_2^{ll}(x,y,t)$,

$$\begin{aligned}\frac{\partial p_2^{ll}(x,y,t)}{\partial t} &= D\Delta p_2^{ll}(x,y,t)\\ &\quad - \frac{\partial}{\partial x}\Big(F^{ll}(x-y)\,p_2^{ll}(x,y,t)\Big)\\ &\quad - \frac{\partial}{\partial y}\Big(F^{ll}(y-x)\,p_2^{ll}(x,y,t)\Big)\\ &\quad - (n_l-2)\,\frac{\partial}{\partial x}\int_\Omega F^{ll}(x-z)\,p_3^{lll}(x,y,z,t)\,\mathrm{d}z\\ &\quad - (n_l-2)\,\frac{\partial}{\partial y}\int_\Omega F^{ll}(y-z)\,p_3^{lll}(x,y,z,t)\,\mathrm{d}z\\ &\quad - \sum_{k\neq l}^{K} n_k\,\frac{\partial}{\partial x}\int_\Omega F^{lk}(x-z)\,p_3^{llk}(x,y,z,t)\,\mathrm{d}z\\ &\quad - \sum_{k\neq l}^{K} n_k\,\frac{\partial}{\partial y}\int_\Omega F^{lk}(y-z)\,p_3^{llk}(x,y,z,t)\,\mathrm{d}z. \end{aligned} \tag{2.25}$$

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 (2.25), 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.

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),

$$p_2^{lm}(x,y,t) = p_1^{l}(x,t)\,p_1^{m}(y,t). \tag{2.26}$$

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., 2004a).

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 (2.26) into Equation (2.16), we obtain

$$\frac{\partial p_1(x,t)}{\partial t} = D\Delta p_1(x,t) - (N-1)\nabla(p_1(x,t)\,V(x,t)), \tag{2.27}$$

where

$$V(x,t) = \int_\Omega F(x-y)\,p_1(y,t)\,\mathrm{d}y, \tag{2.28}$$

is the velocity field induced by interactions between cells. Second, for $K = 2$, the MFA leads to two coupled equations,

$$\begin{aligned}\frac{\partial p_1^1(x,t)}{\partial t} &= D\Delta p_1^1(x,t) - (n_1-1)\nabla\Big(p_1^1(x,t)\,V^{11}(x,t)\Big)\\ &\quad - n_2\nabla\Big(p_1^1(x,t)\,V^{12}(x,t)\Big),\end{aligned} \tag{2.29}$$

$$\begin{aligned}\frac{\partial p_1^2(x,t)}{\partial t} &= D\Delta p_1^2(x,t) - (n_2-1)\nabla\Big(p_1^2(x,t)\,V^{22}(x,t)\Big)\\ &\quad - n_1\nabla\Big(p_1^2(x,t)\,V^{21}(x,t)\Big),\end{aligned} \tag{2.30}$$

$$V^{lm}(x,t) = \int_\Omega F^{lm}(x-y)\,p_1^m(y,t)\,\mathrm{d}y, \tag{2.31}$$

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

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

$$p_3^{lms}(x,y,z,t) = \frac{p_2^{lm}(x,y,t)\,p_2^{ls}(x,z,t)\,p_2^{ms}(y,z,t)}{p_1^l(x,t)\,p_1^m(y,t)\,p_1^s(z,t)}, \tag{2.32}$$

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

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

$$\frac{\partial p_1(x,t)}{\partial t} = D\Delta p_1(x,t) - (N-1)\nabla\Big(\int_\Omega F(x-y)\,p_2(x,y,t)\,\mathrm{d}y\Big), \tag{2.33}$$

$$\begin{aligned}\frac{\partial p_2(x,y,t)}{\partial t} &= D\Delta p_2(x,y,t)\\ &\quad - \frac{\partial}{\partial x}\Big(F(x-y)\,p_2(x,y,t)\Big)\\ &\quad - \frac{\partial}{\partial y}\Big(F(y-x)\,p_2(x,y,t)\Big)\\ &\quad - (N-2)\,\frac{\partial}{\partial x}\int_\Omega F(x-z)\,\frac{p_2(x,y,t)\,p_2(x,z,t)\,p_2(y,z,t)}{p_1(x,t)\,p_1(y,t)\,p_1(z,t)}\,\mathrm{d}z\\ &\quad - (N-2)\,\frac{\partial}{\partial y}\int_\Omega F(y-z)\,\frac{p_2(x,y,t)\,p_2(x,z,t)\,p_2(y,z,t)}{p_1(x,t)\,p_1(y,t)\,p_1(z,t)}\,\mathrm{d}z. \end{aligned} \tag{2.34}$$

It is useful to note that there is more than one way to solve a problem with $K = 1$ using the KSA framework. One approach would be to solve Equations (2.33) and (2.34) simultaneously. However, it is more computationally efficient to solve Equation (2.34) to give $p_2(x,y,t)$, and then to obtain $p_1(x,t)$ by numerical integration

$$p_1(x,t) = \int_\Omega p_2(x,y,t)\,\mathrm{d}y. \tag{2.35}$$

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

$$\begin{aligned}\frac{\partial p_1^1(x,t)}{\partial t} &= D\Delta p_1^1(x,t) - (n_1-1)\nabla\Big(\int_\Omega F^{11}(x-y)\,p_2^{11}(x,y,t)\,\mathrm{d}y\Big)\\ &\quad - n_2\,\nabla\Big(\int_\Omega F^{12}(x-y)\,p_2^{12}(x,y,t)\,\mathrm{d}y\Big), \end{aligned} \tag{2.36}$$

$$\begin{aligned}\frac{\partial p_1^2(x,t)}{\partial t} &= D\Delta p_1^2(x,t) - (n_2-1)\nabla\Big(\int_\Omega F^{22}(x-y)\,p_2^{22}(x,y,t)\,\mathrm{d}y\Big)\\ &\quad - n_1\,\nabla\Big(\int_\Omega F^{21}(x-y)\,p_2^{21}(x,y,t)\,\mathrm{d}y\Big), \end{aligned} \tag{2.37}$$

$$
\begin{aligned}
\frac{\partial p_2^{11}(x,y,t)}{\partial t} &= D\Delta p_2^{11}(x,y,t) \\
&\quad - \frac{\partial}{\partial x}\Big(F^{11}(x-y)\, p_2^{11}(x,y,t)\Big) \\
&\quad - \frac{\partial}{\partial y}\Big(F^{11}(y-x)\, p_2^{11}(x,y,t)\Big) \\
&\quad - (n_1-2)\frac{\partial}{\partial x}\int_\Omega F^{11}(x-z)\,\frac{p_2^{11}(x,y,t)\, p_2^{11}(x,z,t)\, p_2^{11}(y,z,t)}{p_1^1(x,t)\, p_1^1(y,t)\, p_1^1(z,t)}\,\mathrm{d}z \\
&\quad - (n_1-2)\frac{\partial}{\partial y}\int_\Omega F^{11}(y-z)\,\frac{p_2^{11}(x,y,t)\, p_2^{11}(x,z,t)\, p_2^{11}(y,z,t)}{p_1^1(x,t)\, p_1^1(y,t)\, p_1^1(z,t)}\,\mathrm{d}z \\
&\quad - n_2\frac{\partial}{\partial x}\int_\Omega F^{12}(x-z)\,\frac{p_2^{11}(x,y,t)\, p_2^{12}(x,z,t)\, p_2^{12}(y,z,t)}{p_1^1(x,t)\, p_1^1(y,t)\, p_1^2(z,t)}\,\mathrm{d}z \\
&\quad - n_2\frac{\partial}{\partial y}\int_\Omega F^{12}(y-z)\,\frac{p_2^{11}(x,y,t)\, p_2^{12}(x,z,t)\, p_2^{12}(y,z,t)}{p_1^1(x,t)\, p_1^1(y,t)\, p_1^2(z,t)}\,\mathrm{d}z,
\end{aligned}
\tag{2.38}
$$

$$
\begin{aligned}
\frac{\partial p_2^{22}(x,y,t)}{\partial t} &= D\Delta p_2^{22}(x,y,t) \\
&\quad - \frac{\partial}{\partial x}\Big(F^{22}(x-y)\, p_2^{22}(x,y,t)\Big) \\
&\quad - \frac{\partial}{\partial y}\Big(F^{22}(y-x)\, p_2^{22}(x,y,t)\Big) \\
&\quad - (n_2-2)\frac{\partial}{\partial x}\int_\Omega F^{22}(x-z)\,\frac{p_2^{22}(x,y,t)\, p_2^{22}(x,z,t)\, p_2^{22}(y,z,t)}{p_1^1(x,t)\, p_1^1(y,t)\, p_1^1(z,t)}\,\mathrm{d}z \\
&\quad - (n_2-2)\frac{\partial}{\partial y}\int_\Omega F^{22}(y-z)\,\frac{p_2^{22}(x,y,t)\, p_2^{22}(x,z,t)\, p_2^{22}(y,z,t)}{p_1^1(x,t)\, p_1^1(y,t)\, p_1^1(z,t)}\,\mathrm{d}z \\
&\quad - n_1\frac{\partial}{\partial x}\int_\Omega F^{21}(x-z)\,\frac{p_2^{22}(x,y,t)\, p_2^{21}(x,z,t)\, p_2^{21}(y,z,t)}{p_1^1(x,t)\, p_1^1(y,t)\, p_1^2(z,t)}\,\mathrm{d}z \\
&\quad - n_1\frac{\partial}{\partial y}\int_\Omega F^{21}(y-z)\,\frac{p_2^{22}(x,y,t)\, p_2^{21}(x,z,t)\, p_2^{21}(y,z,t)}{p_1^1(x,t)\, p_1^1(y,t)\, p_1^2(z,t)}\,\mathrm{d}z,
\end{aligned}
\tag{2.39}
$$

$$
\begin{aligned}
\frac{\partial p_2^{12}(x,y,t)}{\partial t} &= D\Delta p_2^{12}(x,y,t) \\
&\quad - \frac{\partial}{\partial x}\Big(F^{12}(x-y)\, p_2^{12}(x,y,t)\Big) \\
&\quad - \frac{\partial}{\partial y}\Big(F^{12}(y-x)\, p_2^{12}(x,y,t)\Big) \\
&\quad - (n_2-1)\frac{\partial}{\partial x}\int_\Omega F^{12}(x-z)\,\frac{p_2^{12}(x,y,t)\, p_2^{12}(x,z,t)\, p_2^{22}(y,z,t)}{p_1^1(x,t)\, p_1^2(y,t)\, p_1^2(z,t)}\,\mathrm{d}z \\
&\quad - (n_1-1)\frac{\partial}{\partial x}\int_\Omega F^{11}(x-z)\,\frac{p_2^{12}(x,y,t)\, p_2^{11}(x,z,t)\, p_2^{21}(y,z,t)}{p_1^1(x,t)\, p_1^2(y,t)\, p_1^1(z,t)}\,\mathrm{d}z \\
&\quad - (n_1-1)\frac{\partial}{\partial y}\int_\Omega F^{21}(y-z)\,\frac{p_2^{12}(x,y,t)\, p_2^{11}(x,z,t)\, p_2^{21}(y,z,t)}{p_1^1(x,t)\, p_1^2(y,t)\, p_1^1(z,t)}\,\mathrm{d}z \\
&\quad - (n_2-1)\frac{\partial}{\partial y}\int_\Omega F^{22}(y-z)\,\frac{p_2^{12}(x,y,t)\, p_2^{12}(x,z,t)\, p_2^{22}(y,z,t)}{p_1^1(x,t)\, p_1^2(y,t)\, p_1^2(z,t)}\,\mathrm{d}z,
\end{aligned}
\tag{2.40}
$$

$$
\begin{aligned}
\frac{\partial p_2^{21}(x,y,t)}{\partial t} &= D\Delta p_2^{21}(x,y,t) \\
&\quad - \frac{\partial}{\partial x}\Big(F^{21}(x-y)\, p_2^{21}(x,y,t)\Big) \\
&\quad - \frac{\partial}{\partial y}\Big(F^{21}(y-x)\, p_2^{21}(x,y,t)\Big) \\
&\quad - (n_1-1)\frac{\partial}{\partial y}\int_\Omega F^{21}(y-z)\,\frac{p_2^{21}(x,y,t)\, p_2^{21}(x,z,t)\, p_2^{11}(y,z,t)}{p_1^2(x,t)\, p_1^1(y,t)\, p_1^1(z,t)}\,\mathrm{d}z \\
&\quad - (n_2-1)\frac{\partial}{\partial y}\int_\Omega F^{22}(y-z)\,\frac{p_2^{21}(x,y,t)\, p_2^{22}(x,z,t)\, p_2^{12}(y,z,t)}{p_1^2(x,t)\, p_1^1(y,t)\, p_1^2(z,t)}\,\mathrm{d}z \\
&\quad - (n_2-1)\frac{\partial}{\partial x}\int_\Omega F^{12}(x-z)\,\frac{p_2^{21}(x,y,t)\, p_2^{22}(x,z,t)\, p_2^{12}(y,z,t)}{p_1^2(x,t)\, p_1^1(y,t)\, p_1^2(z,t)}\,\mathrm{d}z \\
&\quad - (n_1-1)\frac{\partial}{\partial x}\int_\Omega F^{11}(x-z)\,\frac{p_2^{21}(x,y,t)\, p_2^{21}(x,z,t)\, p_2^{11}(y,z,t)}{p_1^2(x,t)\, p_1^1(y,t)\, p_1^1(z,t)}\,\mathrm{d}z.
\end{aligned}
\tag{2.41}
$$

Again, there are multiple strategies for solving the KSA equations when $K = 2$. Here, we solve Equations (2.38) and (2.39) 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 (2.35). 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 ($K = 2$) experi-

ments, we will now solve these governing equations for both cases and compare results with averaged data from discrete simulations.

2.4.2 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., 2013a). To emphasize the importance of non mean-field effects, all simulation results in Chapter 2 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., 2013a). In contrast, if the models are applied to deal with the collective migration of mesenchymal cells, without significant adhesion (Simpson et al., 2013a), then additional results in Chapter 2A 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

$$\left\langle \xi_i(t)\xi_j(t')\right\rangle = \frac{2D}{\Delta t}\delta_{ij}\delta_{tt'}, \tag{2.42}$$

which is a white noise limit (Chapter 2A). The variance of ξ_i is given by

$$\text{Var}(\xi_i) = \frac{2D}{\Delta t}, \tag{2.43}$$

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

$$\alpha(x) = \begin{cases} 0, & 0\ \mu\text{m} \leq x < 600\ \mu\text{m}, \\ 25\times 10^{-3}, & 600\ \mu\text{m} \leq x \leq 1400\ \mu\text{m}, \\ 0, & 1400\ \mu\text{m} < x \leq 2000\ \mu\text{m}, \end{cases} \tag{2.44}$$

on $0 \leq x \leq 2000\ \mu\text{m}$, which is a typical length scale for an *in vitro* cell migration experiment (Jin et al., 2016a). 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

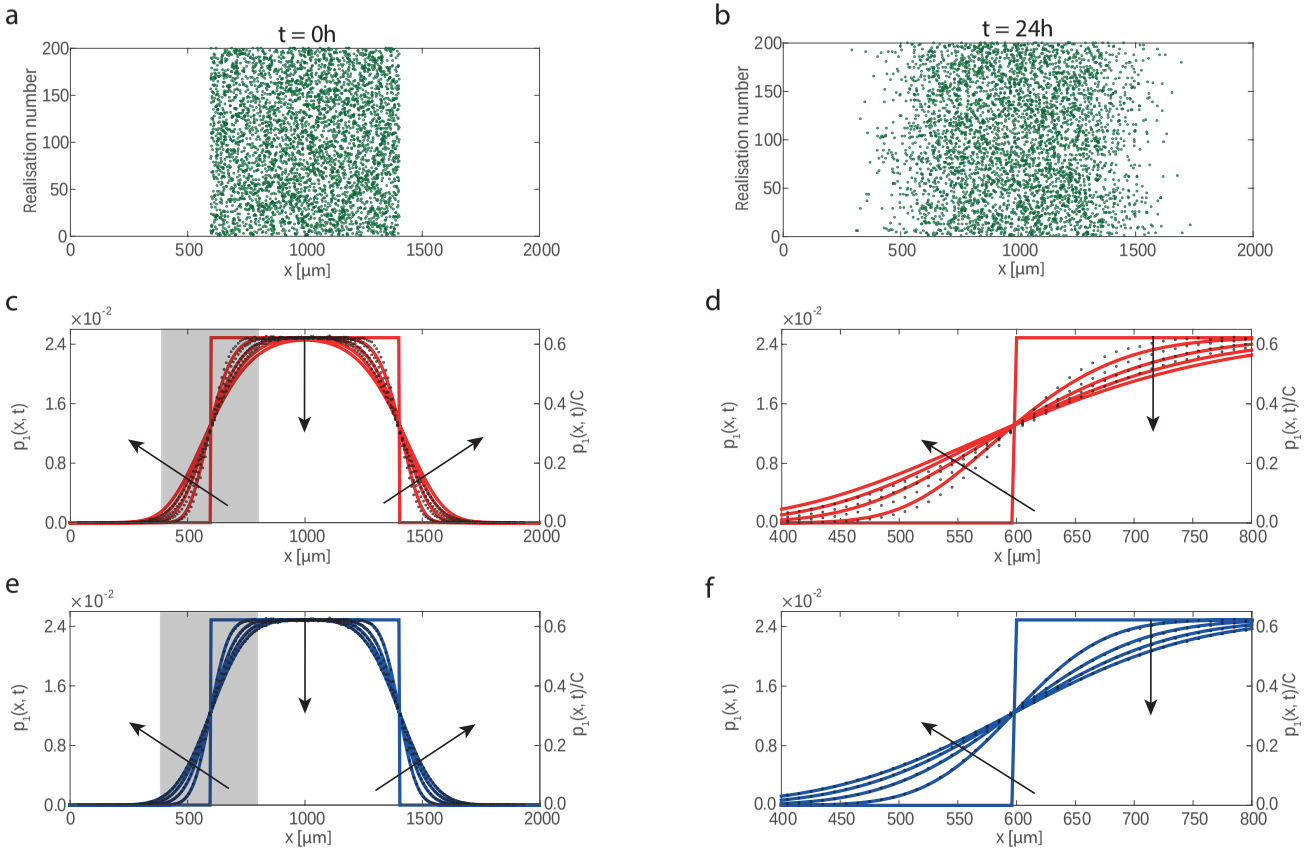


Figure 2.2: Comparison of ensemble averages of stochastic simulations and solutions of the MFA and KSA continuum models for a single-species population of cells on a one-dimensional domain with $0 \leq x \leq 2000$ μm. Snapshots in (a)-(b) show 200 realisations of the discrete model at $t = 0$ and $t = 24$ 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 (2.27) (red lines), and the KSA model, Equation (2.34) (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 (2.27) is integrated with $\Delta x = 4$ μm and $\Delta t = 10^{-2}$ h, and Equation (2.34) 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.

placed according to Equation (2.44), 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 (2.27) and (2.34) 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 2.2 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 2.2(a)-(b). A comparison of the ensemble averaged data and the solution of the MFA and KSA models are given in Figure 2.2(c) and Figure 2.2(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 2.2(d) and Figure 2.2(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 2.2(c) and Figure 2.2(e). However, when we examine the performance of MFA model more closely, as illustrated in Figure 2.2(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 2.2(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.

2.4.3 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 $\delta_1 = 18\ \mu$m, and the diameter of cells in the second subpopulation is $\delta_2 = 25\ \mu$m. 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,

$$\alpha_1(x) = \begin{cases} 0, & 0\ \mu\text{m} \le x < 600\ \mu\text{m}, \\ 25 \times 10^{-3}, & 600\ \mu\text{m} \le x \le 1400\ \mu\text{m}, \\ 0, & 1400\ \mu\text{m} < x \le 2000\ \mu\text{m}, \end{cases} \tag{2.45}$$

$$\alpha_2(x) = \begin{cases} 10.8 \times 10^{-3}, & 0\ \mu\text{m} \le x < 600\ \mu\text{m}, \\ 0, & 600\ \mu\text{m} \le x \le 1400\ \mu\text{m}, \\ 10.8 \times 10^{-3}, & 1400\ \mu\text{m} < x \le 2000\ \mu\text{m}, \end{cases} \tag{2.46}$$

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 $600 \le x \le 1400\ \mu$m 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 2.3(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

$$\alpha_1(x) = \begin{cases} 0, & 0\ \mu\text{m} \le x < 600\ \mu\text{m}, \\ 25 \times 10^{-3}, & 600\ \mu\text{m} \le x < 1000\ \mu\text{m}, \\ 0, & 1000\ \mu\text{m} < x \le 2000\ \mu\text{m}, \end{cases} \tag{2.47}$$

$$\alpha_2(x) = \begin{cases} 0, & 0\ \mu\text{m} \le x < 1000\ \mu\text{m}, \\ 25 \times 10^{-3}, & 1000\ \mu\text{m} \le x \le 1400\ \mu\text{m}, \\ 0, & 1400\ \mu\text{m} < x \le 2000\ \mu\text{m}. \end{cases} \tag{2.48}$$

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 2.5(a)-(c) at $t = 0$, 12 and 24 hours. Here we see that the two subpopulations mix near $x = 1000\ \mu$m. 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 (2.29)-(2.30) and Equations (2.38)-(2.39), 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 (Chapter 2A). Results in Figures 2.3 and 2.5 compare the performance of the MFA approach with the

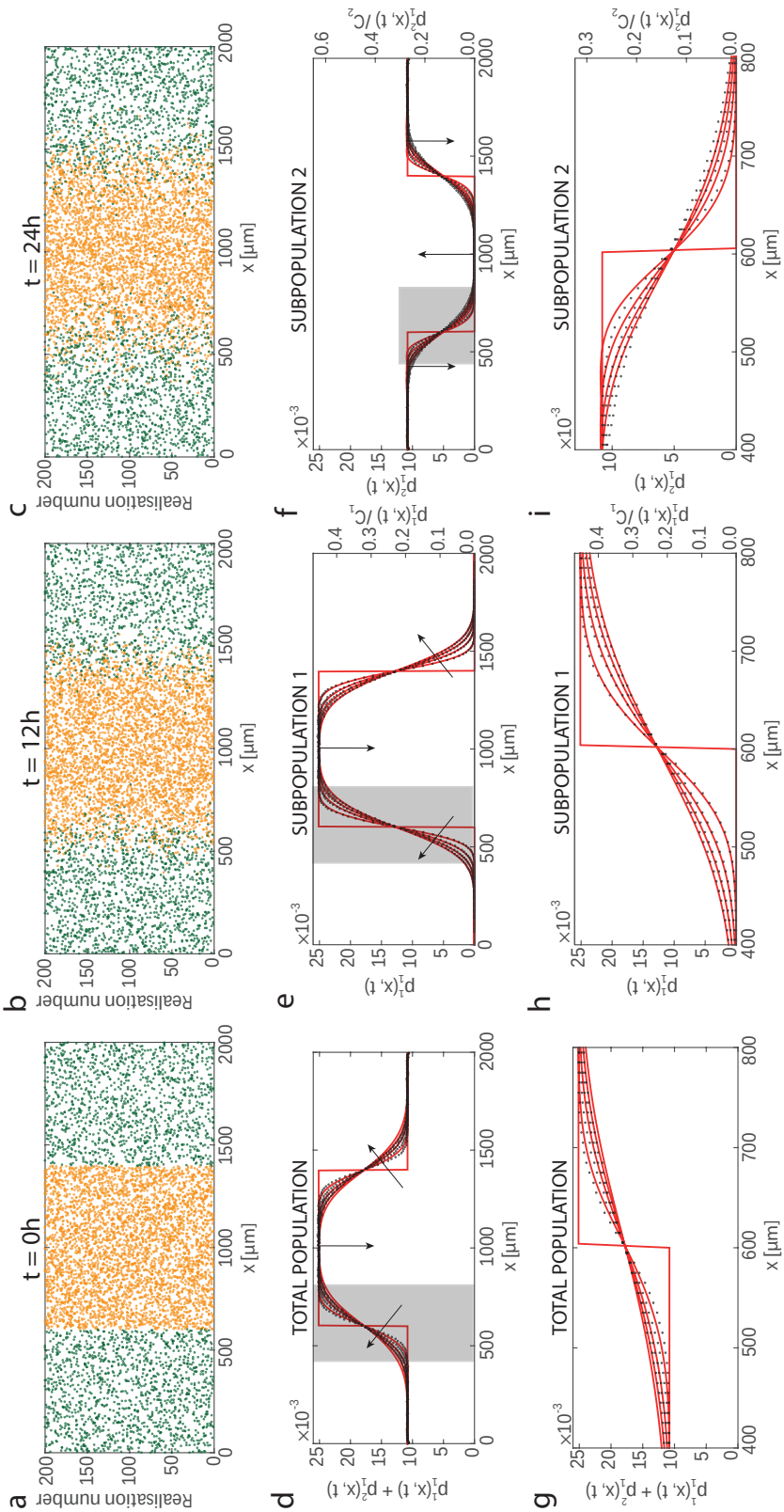


Figure 2.3: Comparison of ensemble averages of stochastic simulations and solutions of the MFA continuum model, given by Equations (2.29)-(2.30), for a two-species population of cells on a one-dimensional domain with $0 \leq x \leq 2000$ μm. Snapshots in (a)-(c) show 200 realisations of the discrete 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.

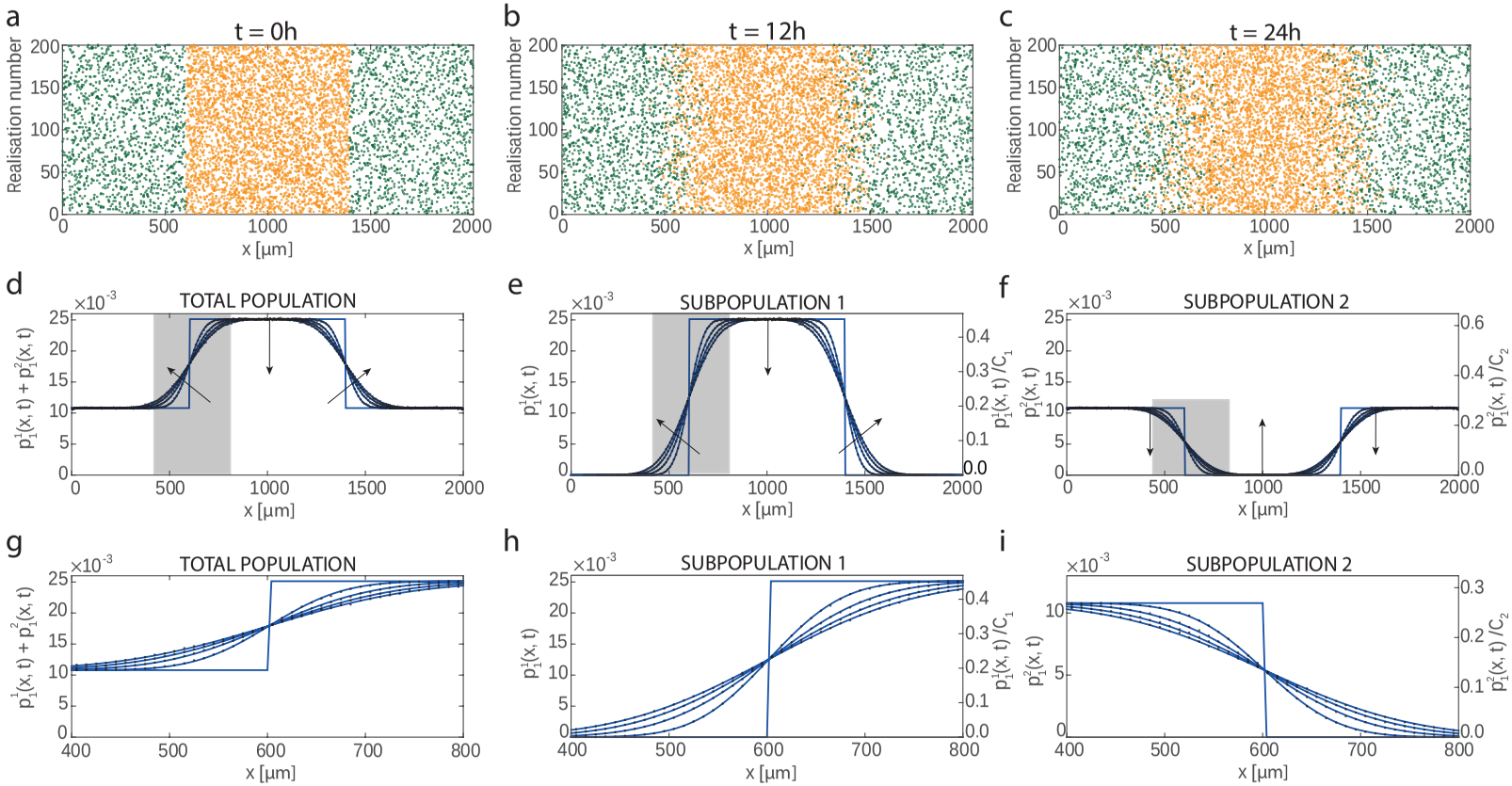


Figure 2.4: Comparison of ensemble averages of stochastic simulations and solutions of the KSA continuum model, given by Equations (2.38)-(2.39), for a two-species population of cells on a one-dimensional domain with $0 \leq x \leq 2000$ μm. Snapshots in (a)-(c) show 200 realisations of the discrete 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.

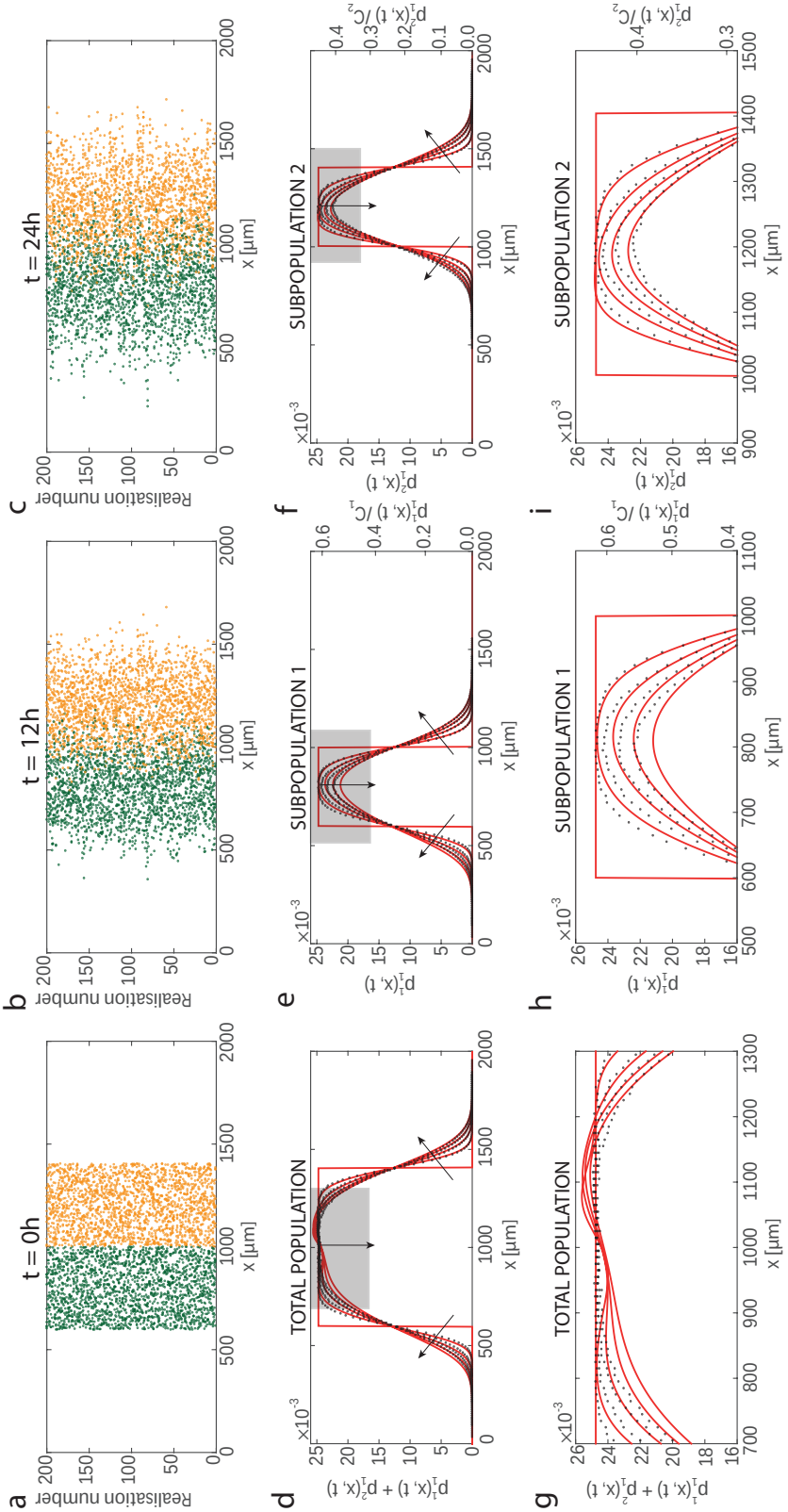


Figure 2.5: Comparison of ensemble averages of stochastic simulations and solution of the MFA continuum model, given by Equations (2.29)-(2.30), for a two-species population of cells on a one-dimensional domain with $0 \leq x \leq 2000\ \mu$m. Snapshots in (a)-(c) show 200 realisations of the discrete 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 $\Delta x = 4\ \mu$m and $\Delta t = 5 \times 10^{-3}$ h. The remaining parameters are $n_1 = 10$, $n_2 = 10$, $a_1 = 0.06\ \mu\text{m}^{-1}$, $a_2 = 0.08\ \mu\text{m}^{-1}$, $a_{12} = 0.07\ \mu\text{m}^{-1}$, $\delta_1 = 25\ \mu$m, $\delta_2 = 18\ \mu$m, $\delta_{12} = 21.5\ \mu$m, $f_0^{11} = 1.5\ \mu$m/h, $f_0^{22} = 2\ \mu$m/h, $f_0^{12} = 1.75\ \mu$m/h.

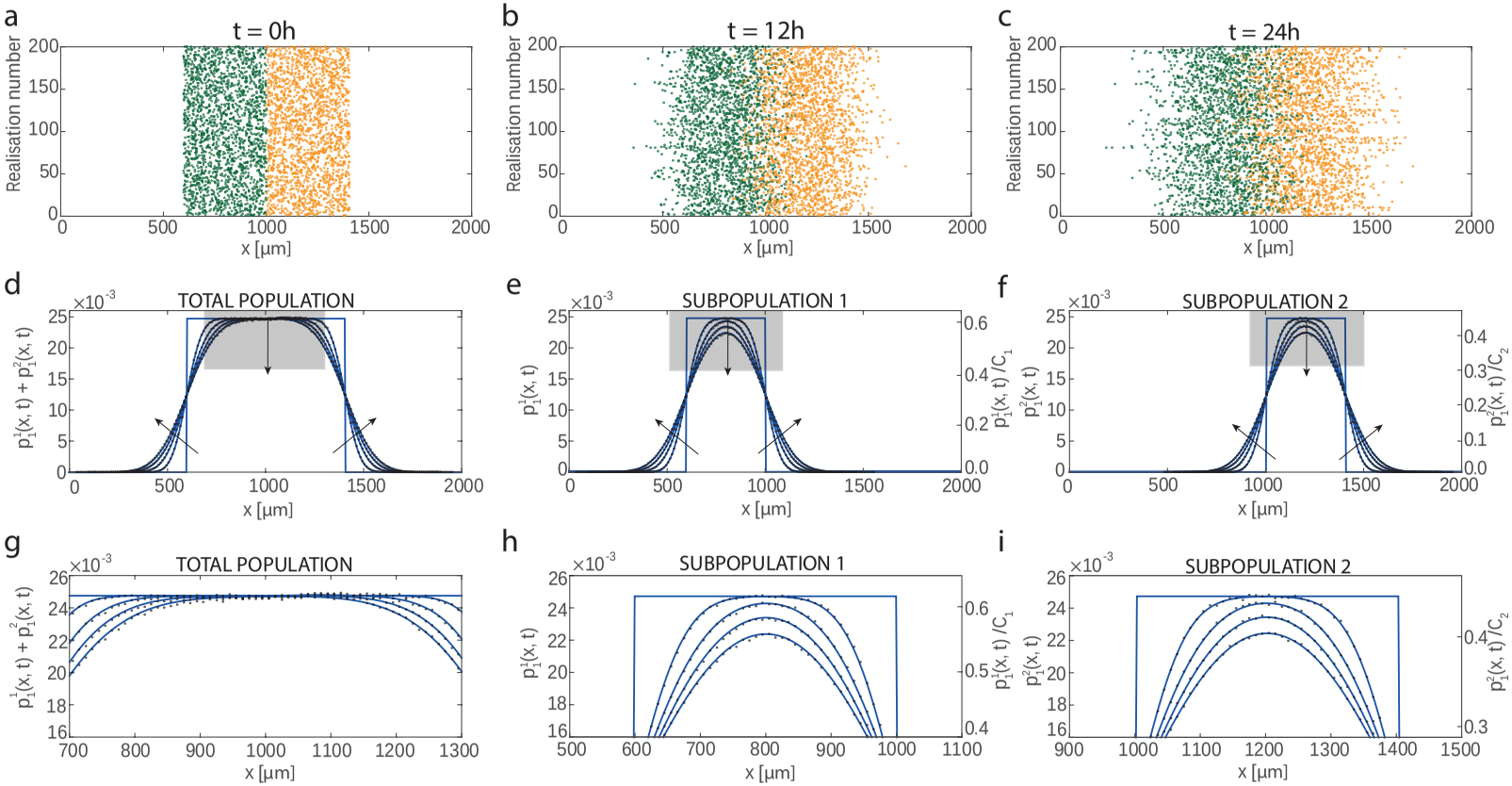


Figure 2.6: Comparison of ensemble averages of stochastic simulations and solutions of the KSA continuum model, given by Equations (2.38)-(2.39), for a two-species population of cells on a one-dimensional domain with $0 \leq x \leq 2000$ μm. Snapshots in (a)-(c) show 200 realisations of the discrete 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.

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 2.4 and 2.6 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 2.2, the KSA approach outperforms the MFA model.

2.4.4 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 2.3–2.4, and we quantify how the accuracy of the KSA and MFA continuum models depends on the strength of adhesion and the ratio 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,

$$\mathrm{E}_{\mathrm{MFA}}(t) = \frac{1}{I}\sum_{i=1}^{I}[S_{\mathrm{MFA}}(i,t) - S_{\mathrm{discrete}}(i,t)]^2, \tag{2.49}$$

$$\mathrm{E}_{\mathrm{KSA}}(t) = \frac{1}{I}\sum_{i=1}^{I}[S_{\mathrm{KSA}}(i,t) - S_{\mathrm{discrete}}(i,t)]^2, \tag{2.50}$$

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 2.7 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 2.3-2.4. Overall, the results in Figure 2.7 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.

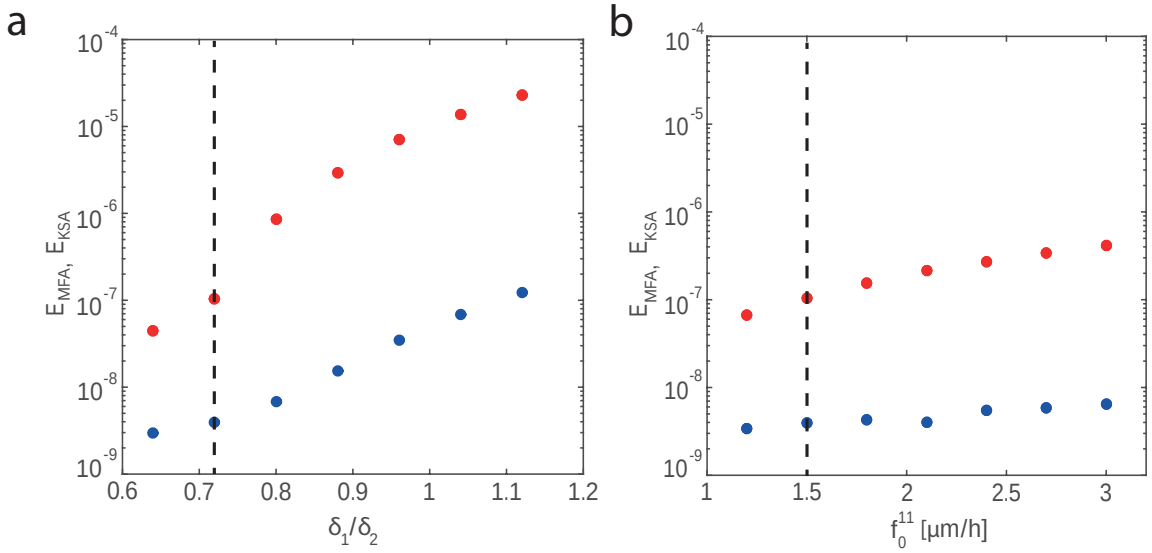


Figure 2.7: (a) Comparison of the accuracy of the MFA and KSA continuum approximations as a function of δ_1/δ_2 at time $t = 24\,\mathrm{h}$ for the first co-culture experiment. All results in (a) correspond to $\delta_2 = 25\,\mu\mathrm{m}$, and the ratio δ_1/δ_2 is varied by altering δ_1. (b) Comparison 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 $E_{MFA}(t)$ (red dots) and $E_{KSA}(t)$ (blue dots), and the dashed vertical line indicates the parameter values presented previously in Figures 2.3-2.4. All continuum models are solved numerically with $\Delta x = 4\,\mu\mathrm{m}$ and $\Delta t = 5 \times 10^{-3}\,\mathrm{h}$.

2.5 Conclusions

In this chapter, 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., 2013a; 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).

Chapter 2A

Additional results for Chapter 2

2A.1 Derivation of one- and two-cell PDF and corresponding Fokker-Planck equations

To write down explicit expressions for $P_1(x,t)$ and $P_2(x,y,t)$ we start with the Langevin equations, given by Equation (2.3). We introduce an agent density function, $\rho(x,t)$. The evolution of the agent density is given by

$$\frac{\partial \rho(x,t)}{\partial t} = -\frac{\partial J}{\partial x}, \tag{2A.1}$$

where J is the flux of agents. Assuming $J = \rho\, \mathrm{d}x/\mathrm{d}t$, we have

$$\frac{\partial \rho(x,t)}{\partial t} = -\sum_{i=1}^{N} \frac{\partial}{\partial x_i}\Big(\sum_{j\neq i} F_{ij} + \xi_i\Big)\rho(x,t), \tag{2A.2}$$

where $\sum_{j\neq i} F_{ij} + \xi_i$ is the right hand side of Equation (2.3) and ξ_i is treated as a fixed parameter.

Suppose that ξ_i is a random variable, and we have obtained a solution to Equation (2A.2) with the initial conditions

$$\rho(x,0) = \prod_{i=1}^{N} \delta(x - x_i(0)) = \delta^{(N)}(x - x_i(0)), \tag{2A.3}$$

where δ is the Dirac delta-function, and $x_i(0)$ is the initial position of the ith cell. For any initial density distribution, $\rho(x,0)$, we assume that we can obtain $\langle \rho(x,t)\rangle_\xi$, where the average is taken over many different realisations of the stochastic force ξ_i.

Let $P_1^i(x,t)$ be the probability density distribution for an individual agent. Then the

one-cell PDF is given by (van Kampen, 1976)

$$P_1^i(x,t) = \langle\langle \rho(x,t)\rangle_\xi\rangle_{\mathrm{IC}}, \tag{2A.4}$$

where the average is taken over different realisations of the initial distribution and the random variable ξ_i. Since the averaged local density can be expressed as

$$\langle\langle \rho(x,t)\rangle_\xi\rangle_{\mathrm{IC}} = \langle\langle \delta^{(N)}(x - x_i(t))\rangle_\xi\rangle_{\mathrm{IC}}, \tag{2A.5}$$

we can define the one-cell PDF in the following way,

$$P_1^i(x,t) = \langle\langle \delta^{(N)}(x - x_i(t))\rangle_\xi\rangle_{\mathrm{IC}}, \tag{2A.6}$$

where ξ_i, and $x_i(t)$ is the position of the ith cell at time t, as given by Equation (2.3).

Similarly, we define the two-cell PDF by the following

$$P_2^{ij}(x,y,t) = \langle\langle \delta^{(N)}(x - x_i(t))\, \delta^{(N)}(y - y_j(t))\rangle_\xi\rangle_{\mathrm{IC}}. \tag{2A.7}$$

For the present, we begin with the general case of having space-time correlated stochastic movement of cells. This is a generalisation of Brownian motion. In this general case we have an autocorrelation function for the stochastic force, ξ_i, in the form of

$$\langle \xi_i(x,t)\xi_j(y,t')\rangle_\xi = D\,\omega\left(\frac{x-y}{\lambda}; \frac{|t-t'|}{\tau}\right), \tag{2A.8}$$

where the function ω describes correlations in the stochastic force, D is the diffusivity, and λ and τ describe the spatial and temporal correlation lengths. In the case of Brownian motion, when there are no correlation in either space and time, the ω takes the form of the product of two δ functions. This is known as the white noise limit. In the study of collective cell migration we are interested in the case where isolated cells undergo an unbiased random walk, which corresponds to vanishing space-time correlations of the stochastic force, ξ_i, such that $\lambda \to 0$ and $\tau \to 0$. Under these conditions, the evolution of $P_1^i(x,t)$ and $P_2^{ij}(x,y,t)$ are given by (García-Ojalvo and Sancho, 1999)

$$\frac{\partial P_1^i(x,t)}{\partial t} = D\Delta P_1^i(x,t) - \frac{\partial}{\partial x}\left(f_i\, P_1^i(x,t)\right), \tag{2A.9}$$

$$\frac{\partial P_2^{ij}(x,y,t)}{\partial t} = D\Delta P_2^{ij}(x,y,t) - \frac{\partial}{\partial x}\left(f_i\, P_2^{ij}(x,y,t)\right) - \frac{\partial}{\partial y}\left(f_j\, P_2^{ij}(x,y,t)\right). \tag{2A.10}$$

2A.2 Additional results for alternative parameter choices

We now present additional results that compare the performance of the MFA and KSA approximations for different choices of the force parameters for the problems in Chapter 2 where $K = 2$. The only differences between these additional comparisons and the comparisons in Chapter 2 are changes in the values of f_0 and a. The main point of these additional comparisons is to illustrate the improved performance of the MFA approximation when f_0 is sufficiently small. Results in Figures (2A.1)-(2A.4) show that the MFA and KSA models produce almost identical results in these additional cases.

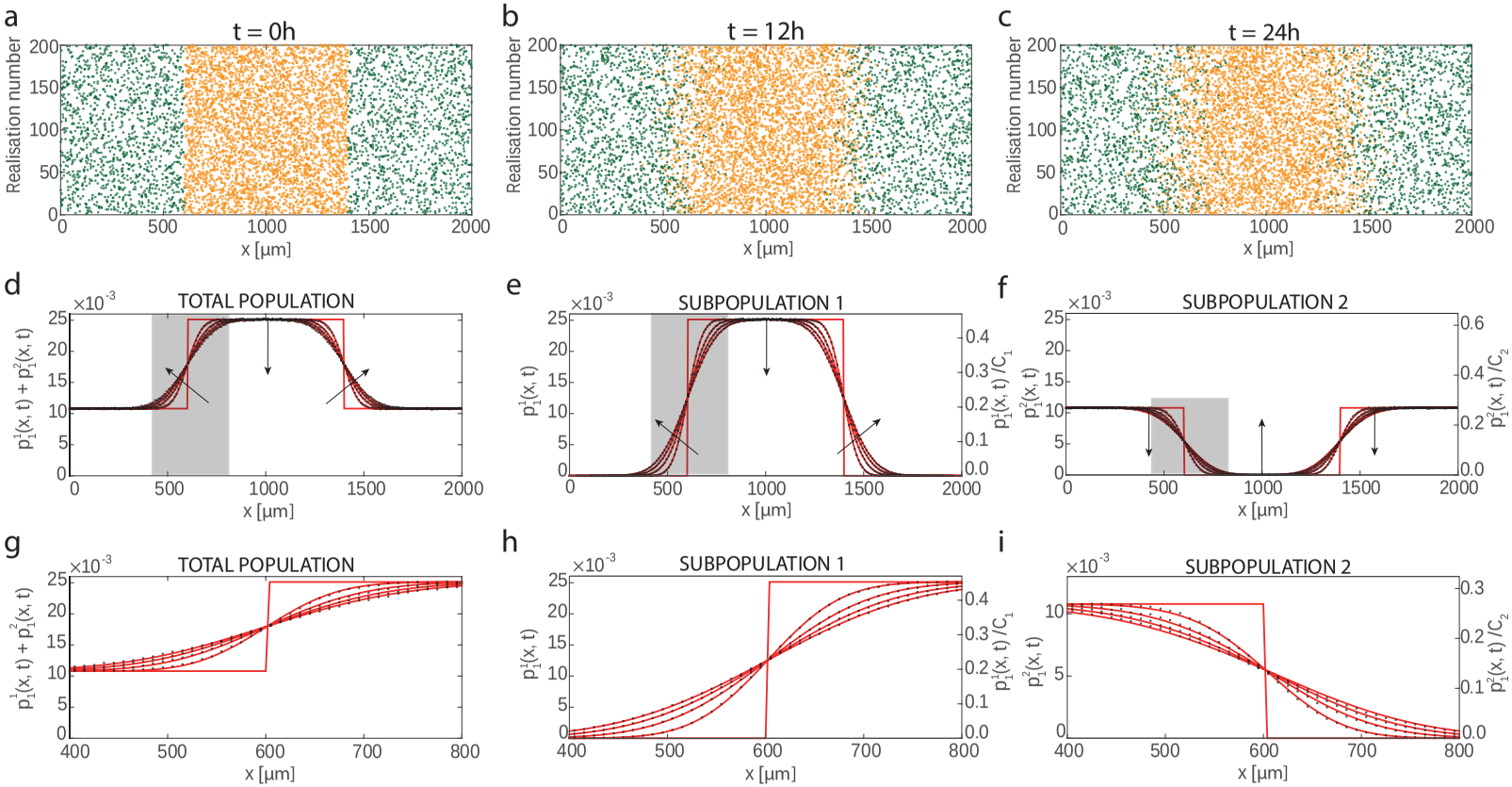


Figure 2A.1: Comparison of ensemble averages of stochastic simulations and solutions of the MFA continuum model, given by Equations (2.29)-(2.30), for a two-species population of cells on a one-dimensional domain with $0 \leq x \leq 2000$ μm. Snapshots in (a)-(c) show 200 realisations of the discrete 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 $p_1^2(x,t)$, and 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 $n_1 = 20$, $n_2 = 13$, $a_1 = 0.05$ μm^{-1}, $a_2 = 0.03$ μm^{-1}, $a_{12} = 0.04$ μm^{-1}, $\delta_1 = 18$ μm, $\delta_2 = 25$ μm, $\delta_{12} = 21.5$ μm, $f_0^{11} = 0.15$ μm/h, $f_0^{22} = 0.1$ μm/h, $f_0^{12} = 0.25$ μm/h.

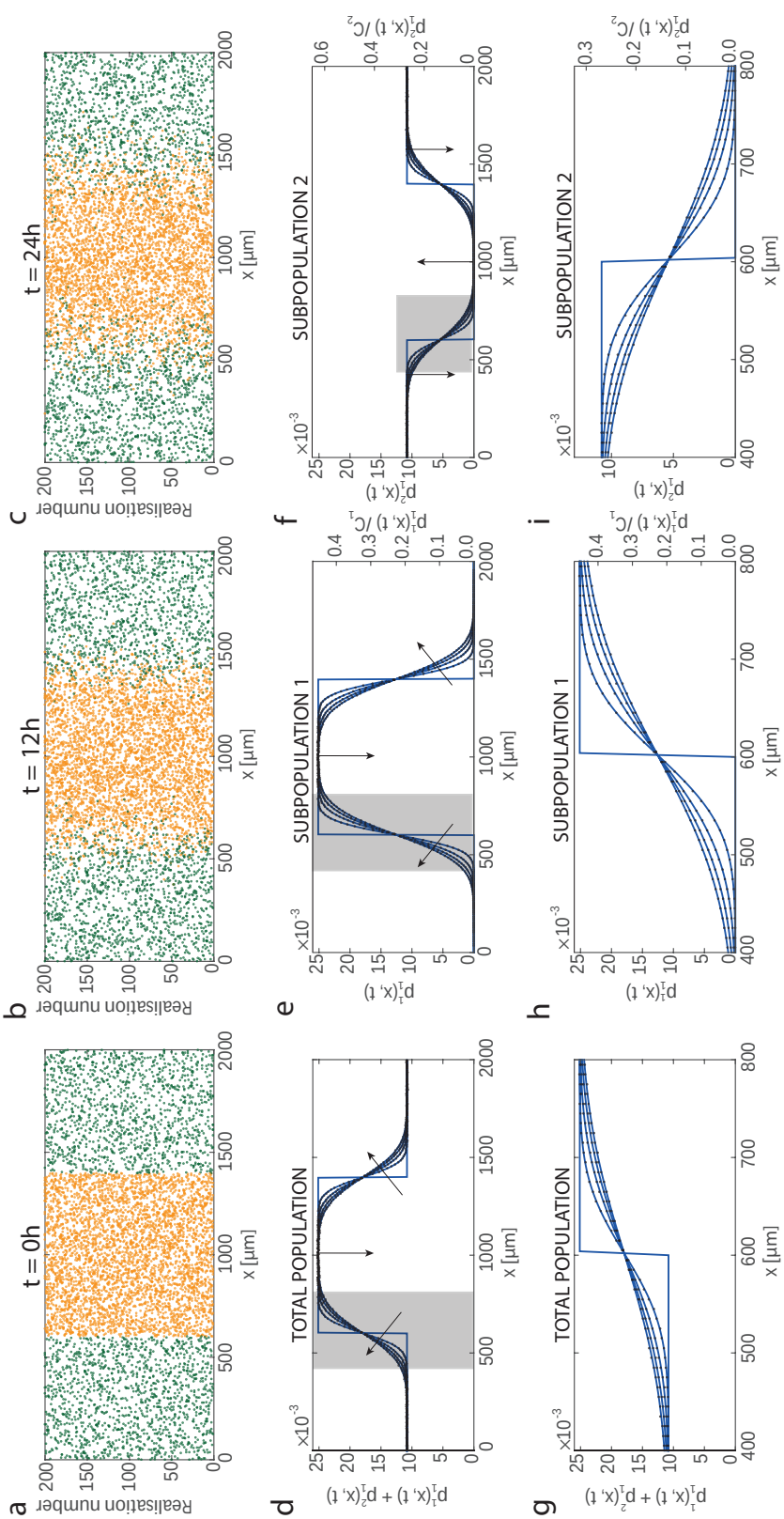


Figure 2A.2: Comparison of ensemble averages of stochastic simulations and solutions of the KSA continuum model, given by Equations (2.38)-(2.39), for a two-species population of cells on a one-dimensional domain with $0 \leq x \leq 2000$ μm. Snapshots in (a)-(c) show 200 realisations of the discrete 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 $p_1^2(x,t)$, and 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 $n_1 = 20$, $n_2 = 13$, $a_1 = 0.05$ μm^{-1}, $a_2 = 0.03$ μm^{-1}, $a_{12} = 0.04$ μm^{-1}, $\delta_1 = 18$ μm, $\delta_2 = 25$ μm, $\delta_{12} = 21.5$ μm, $f_0^{11} = 0.15$ μm/h, $f_0^{22} = 0.1$ μm/h, $f_0^{12} = 0.25$ μm/h.

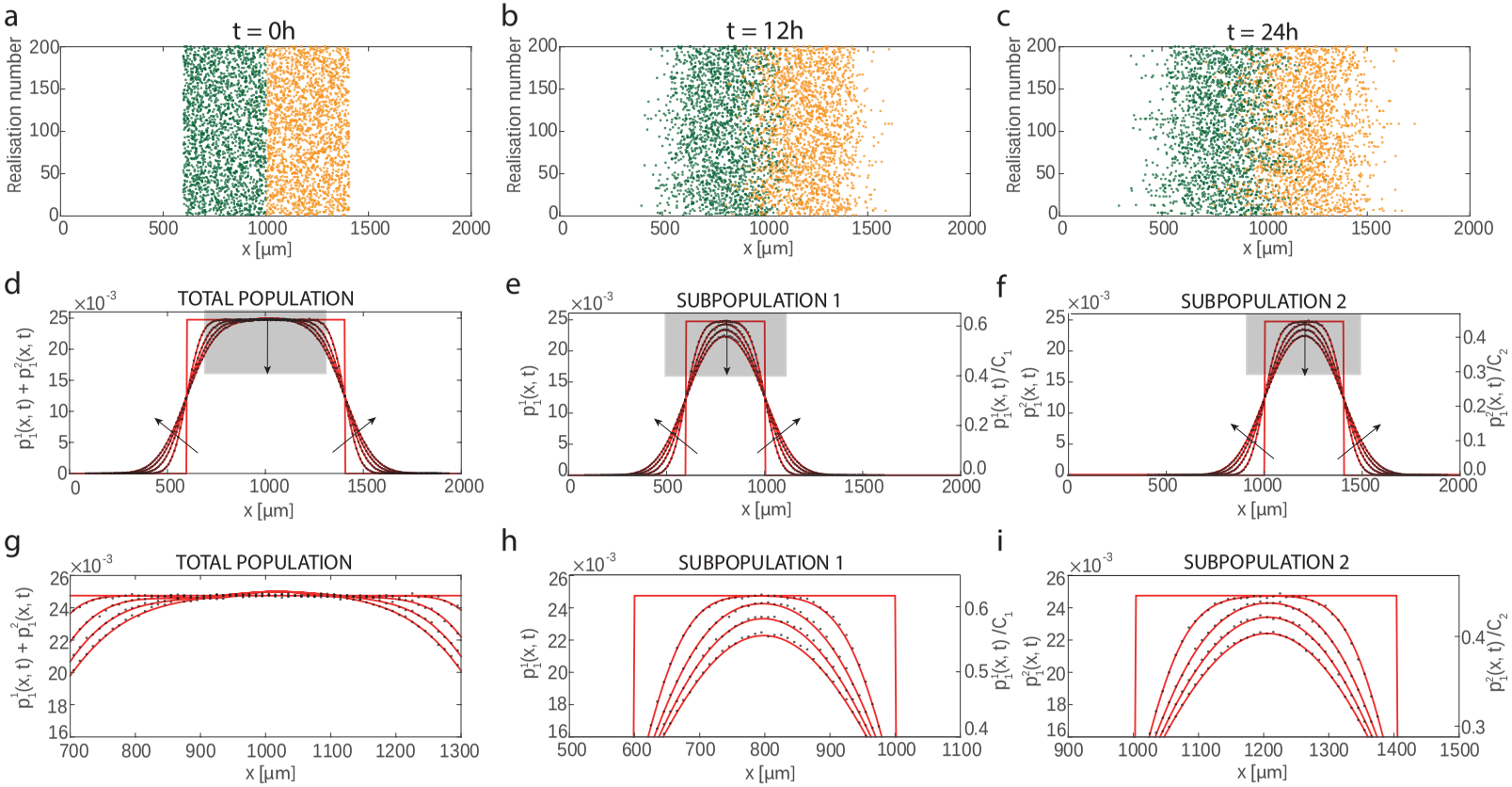


Figure 2A.3: Comparison of ensemble averages of stochastic simulations and solutions of the MFA continuum model, given by Equations (2.29)-(2.30), for a two-species population of cells on a one-dimensional domain with $0 \leq x \leq 2000$ μm. Snapshots in (a)-(c) show 200 realisations of the discrete 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 $n_1 = 10$, $n_2 = 10$, $a_1 = 0.05$ μm^{-1}, $a_2 = 0.03$ μm^{-1}, $a_{12} = 0.04$ μm^{-1}, $\delta_1 = 25$ μm, $\delta_2 = 18$ μm, $\delta_{12} = 21.5$ μm, $f_0^{11} = 0.15$ μm/h, $f_0^{22} = 0.1$ μm/h, $f_0^{12} = 0.25$ μm/h.

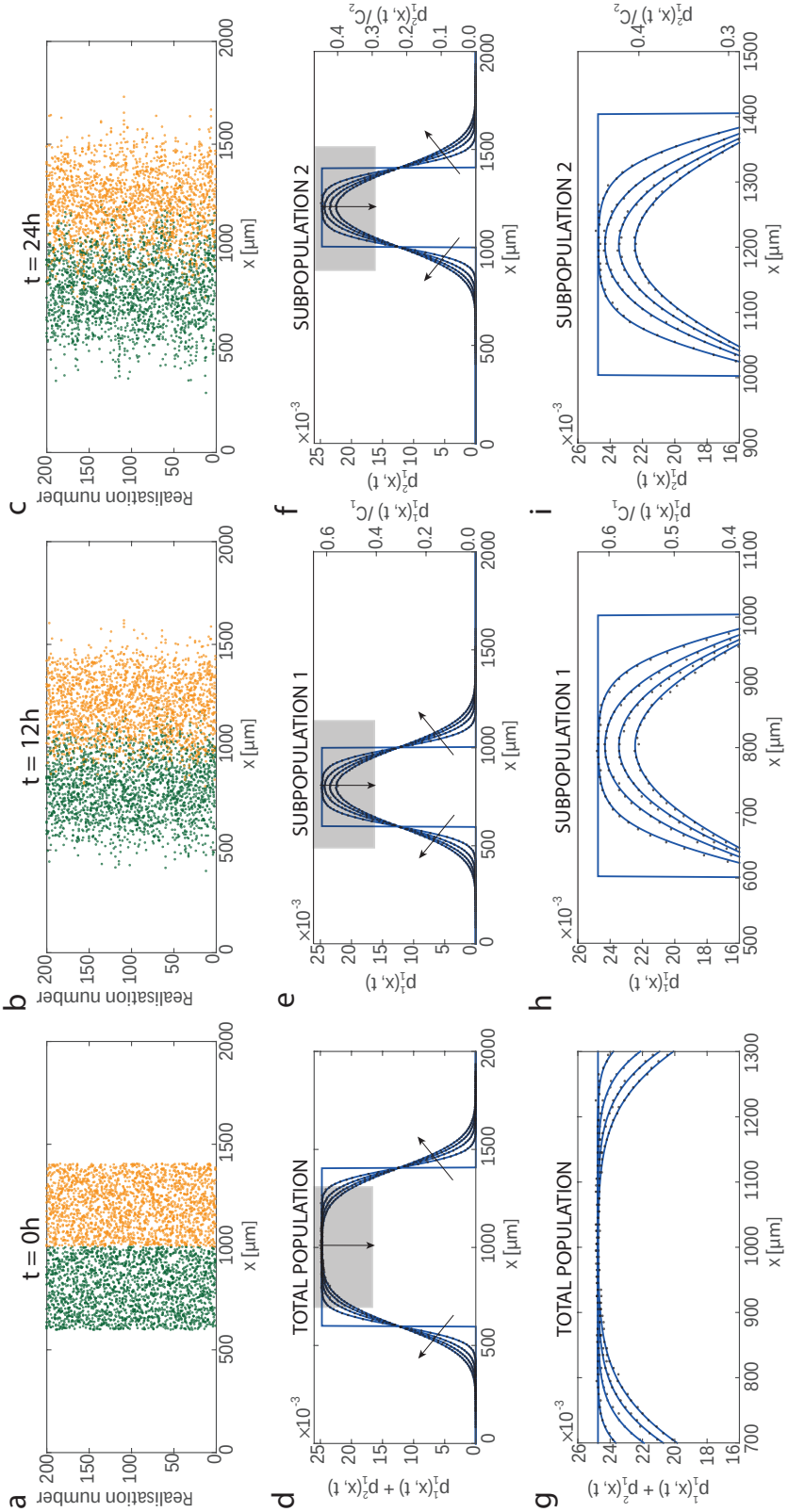


Figure 2A.4: Comparison of ensemble averages of stochastic simulations and solutions of the KSA continuum model, given by Equations (2.38)-(2.39), for a two-species population of cells on a one-dimensional domain with $0 \leq x \leq 2000$ μm. Snapshots in (a)-(c) show 200 realisations of the discrete 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 $n_1 = 10$, $n_2 = 10$, $a_1 = 0.05$ μm^{-1}, $a_2 = 0.03$ μm^{-1}, $a_{12} = 0.04$ μm^{-1}, $\delta_1 = 25$ μm, $\delta_2 = 18$ μm, $\delta_{12} = 21.5$ μm, $f_0^{11} = 0.15$ μm/h, $f_0^{22} = 0.1$ μm/h, $f_0^{12} = 0.25$ μm/h.

2A.3 Discretisation scheme for the MFA model

In this section we present the discretisation scheme used to obtain the numerical solution of the equation governing the evolution of $p_1(x,t)$ in the MFA framework. Here we deal only with the case where we have $K = 1$, and the extension of this method to deal with the multi-species problems follows from the results we present here. In summary, the governing equation that we consider is as follows,

$$\frac{\partial p_1(x,t)}{\partial t} = D\Delta p_1(x,t) - (N-1)\nabla(p_1(x,t)\,V(x,t)), \tag{2A.11}$$

where

$$V(x,t) = f_0 \int_0^L \mathcal{Z}(|x-y|)\,\mathrm{sgn}(x-y)\,p_1(y,t)\,\mathrm{d}y. \tag{2A.12}$$

To present the numerical method as succinctly as possible, we define

$$\beta(x,y,t) = f_0\,\mathcal{Z}(|x-y|)\,\mathrm{sgn}(x-y)\,p_1(y,t), \tag{2A.13}$$

$$\begin{aligned} I_l &= p_1(x_l,t)\int_0^L \beta(x_l,y)\,\mathrm{d}y \\ &= p_1(x_l,t)\frac{h}{2}\sum_s \Big[\beta(x_l,y_{s+1}) + \beta(x_l,y_s)\Big] + O(h^2), \end{aligned} \tag{2A.14}$$

where the trapezoidal rule with step h is used for numerical integration, and indices l and s denote the equally-spaced spatial mesh nodes.

Using the definitions in Equations (2A.13)-(2A.14), we apply the method of lines to Equation (2A.11) and obtain the following system of coupled ordinary differential equations,

$$\frac{\mathrm{d}p_i^1}{\mathrm{d}t} = \frac{D}{h^2}\Big[p_{i+1} - 2p_i + p_{i-1}\Big] - (N-1)\frac{1}{2h}\Big[I_{i+1} - I_{i-1}\Big], \tag{2A.15}$$

where index i denotes a spatial mesh node. Since we consider periodic boundary conditions, this general expression at an interior node, i, can be adapted at the boundaries of the domain. This systems of ordinary differential equations is solved using a fourth order Runge-Kutta algorithm with constant time steps of duration Δt. The values of spatial discretisation, h, and time step, Δt, are chosen to be sufficiently small so that numerical solutions of Equation (2A.11) using h and Δt are visually indistinguishable from solutions obtained using discretisations $h/2$ and $\Delta t/2$.

2A.4 Discretisation scheme for the KSA model

We now present the related discretisation scheme used to obtain the numerical solution to the equation governing the evolution of $p_2(x,y,t)$ in the KSA framework. We note that we only solve the equation for $p_2(x,y,t)$ and obtain $p_1(x,t)$ by numerical integration. Again, we deal only with the case where we have $K = 1$, and we note that the extension of these methods to deal with multi-species problems follows from the results we present here.

The governing equation that we consider is as follows,

$$\begin{aligned}\frac{\partial p_2(x,y,t)}{\partial t} &= D\Delta p_2(x,y,t) \\ &\quad - f_0\Big(\frac{\partial}{\partial x} - \frac{\partial}{\partial y}\Big)\Big(\mathcal{Z}(|x-y|)\,\mathrm{sgn}(x-y)\,p_2(x,y,t)\Big) \\ &\quad - f_0(N-2)\frac{\partial}{\partial x}\int_0^L \mathcal{Z}(|x-z|)\,\mathrm{sgn}(x-z)\,\frac{p_2(x,y,t)\,p_2(x,z,t)\,p_2(y,z,t)}{p_1(x,t)\,p_1(y,t)\,p_1(z,t)}\,\mathrm{d}z \\ &\quad - f_0(N-2)\frac{\partial}{\partial y}\int_0^L \mathcal{Z}(|y-z|)\,\mathrm{sgn}(y-z)\,\frac{p_2(x,y,t)\,p_2(x,z,t)\,p_2(y,z,t)}{p_1(x,t)\,p_1(y,t)\,p_1(z,t)}\,\mathrm{d}z.\end{aligned} \tag{2A.16}$$

To present the numerical method as succinctly as possible, we define

$$\gamma(x,y,t) = f_0\,\mathcal{Z}(|x-y|)\,\mathrm{sgn}(x-y)\,p_2(x,y,t), \tag{2A.17}$$

$$\phi(x,y,z,t) = f_0\,\mathcal{Z}(|x-z|)\,\mathrm{sgn}(x-z)\,\frac{p_2(x,z,t)\,p_2(y,z,t)}{p_1(z,t)}, \tag{2A.18}$$

$$\psi(x,y,z,t) = f_0\,\mathcal{Z}(|y-z|)\,\mathrm{sgn}(y-z)\,\frac{p_2(x,z,t)\,p_2(y,z,t)}{p_1(z,t)}. \tag{2A.19}$$

Upon substituting Equations (2A.17)-(2A.19) into Equation (2A.16), the evolution equation for $p_2(x,y,t)$ can be written as

$$\begin{aligned}\frac{\partial p_2(x,y,t)}{\partial t} &= D\Delta p_2(x,y,t) - \frac{\partial}{\partial x}\gamma(x,y,t) + \frac{\partial}{\partial y}\gamma(x,y,t) \\ &\quad - (N-2)\frac{\partial}{\partial x}\Bigg[\frac{p_2(x,y,t)}{p_1(x,t)\,p_1(y,t)}\int_0^L \phi(x,y,z,t)\,\mathrm{d}z\Bigg] \\ &\quad - (N-2)\frac{\partial}{\partial y}\Bigg[\frac{p_2(x,y,t)}{p_1(x,t)\,p_1(y,t)}\int_0^L \psi(x,y,z,t)\,\mathrm{d}z\Bigg].\end{aligned} \tag{2A.20}$$

We now introduce the discretised quantities

$$\begin{aligned} Q_{l,k} &= \frac{p_2(x_l, y_k, t)}{p_1(x_l, t)\, p_1(y_k, t)} \int_0^L \phi(x_l, y_k, z, t)\, \mathrm{d}z \\ &= \frac{p_2(x_l, y_k, t)}{p_1(x_l, t)\, p_1(y_k, t)} \frac{h}{2} \sum_s \Big[\phi(x_l, y_k, z_{s+1}, t) + \phi(x_l, y_k, z_s, t)\Big] + O(h^2), \end{aligned} \tag{2A.21}$$

$$\begin{aligned} Y_{l,k} &= \frac{p_2(x_l, x_k, t)}{p_1(x_l, t)\, p_1(y_k, t)} \int_0^L \psi(x_l, y_k, z, t)\, \mathrm{d}z \\ &= \frac{p_2(x_l, y_k, t)}{p_1(x_l, t)\, p_1(y_k, t)} \frac{h}{2} \sum_s \Big[\psi(x_l, y_k, z_{s+1}, t) + \psi(x_l, y_k, z_s, t)\Big] + O(h^2), \end{aligned} \tag{2A.22}$$

where the trapezoidal rule with step h on an equally spaced mesh is used to approximate the integrals. We now apply the method of lines to Equation (2A.20) and obtain the following system of coupled ordinary differential equations,

$$\begin{aligned} \frac{\mathrm{d}p_{ij}^2}{\mathrm{d}t} = \frac{D}{h^2}\Big[p_{i+1,j} - 2p_{ij} + p_{i-1,j} + p_{i,j+1} - 2p_{ij} + p_{i,j-1}\Big] \\ - \frac{1}{2h}\Big[\gamma_{i+1,j} - \gamma_{i-1,j}\Big] + \frac{1}{2h}\Big[\gamma_{i,j+1} - \gamma_{i,j-1}\Big] \\ - (N-2)\frac{1}{2h}\Big[Q_{i+1,j} - Q_{i-1,j}\Big] - (N-2)\frac{1}{2h}\Big[Y_{i,j+1} - Y_{i,j-1}\Big], \end{aligned} \tag{2A.23}$$

where indices i, j denote spatial mesh nodes, and $\gamma_{l,m} = \gamma(x_l, y_m, t)$. Since we consider periodic boundary conditions, this general expression at an interior node can be adapted at the boundaries of the domain. This systems of ordinary differential equations is then solved using a fourth order Runge-Kutta algorithm with constant time steps of duration Δt. The values of spatial discretisation, h, and time step, Δt, are chosen to be sufficiently small so that numerical solutions of Equations (2A.16) using h and Δt are visually indistinguishable from solutions obtained using discretisations $h/2$ and $\Delta t/2$.

Chapter 3

Discrete and continuum approximations for collective cell migration in a scratch assay with cell size dynamics

A paper published in the *Bulletin of Mathematical Biology*.

Matsiaka Oleksii, Penington Catherine, Baker Ruth, Simpson Matthew. Discrete and continuum approximations for collective cell migration in a scratch assay with cell size dynamics *Bulletin of Mathematical Biology*, vol. 80, p. 738-757, 2018.

3.1 Abstract

Scratch assays are routinely used to study the collective spreading of cell populations. In general, the rate at which a population of cells spreads is driven by the combined effects of cell migration and proliferation. To examine the effects of cell migration separately from the effects of cell proliferation, scratch assays are often performed after treating the cells with a drug that inhibits proliferation. Mitomycin-C is a drug that is commonly used to suppress cell proliferation in this context. However, in addition to suppressing cell proliferation, Mitomycin-C also causes cells to change size during the experiment, as each cell in the population approximately doubles in size as a result of treatment. Therefore, to describe a scratch assay that incorporates the effects of cell-to-cell crowding, cell-to-cell adhesion, and dynamic changes in cell size, we present a new stochastic model that incorporates these mechanisms. Our agent-based stochastic model takes the form of a system of Langevin equations that is the system of stochastic

differential equations governing the evolution of the population of agents. We incorporate a time-dependent interaction force that is used to mimic the dynamic increase in size of the agents. To provide a mathematical description of the average behaviour of the stochastic model we present continuum limit descriptions using both a standard mean-field approximation, and a more sophisticated moment dynamics approximation that accounts for the density of agents and density of pairs of agents in the stochastic model. Comparing the accuracy of the two continuum descriptions for a typical scratch assay geometry shows that the incorporation of agent growth in the system is associated with a decrease in accuracy of the standard mean-field description. In contrast, the moment dynamics description provides a more accurate prediction of the evolution of the scratch assay when the increase in size of individual agents is included in the model.

3.2 Introduction

In vitro cell biology assays are used to study the invasive properties of malignant cells, to quantify different mechanisms of wound repair, as well as in the discovery of potential drugs (Riss, 2005; Edmondson et al., 2014; Shah et al., 2016). Typically, cells are placed on a two-dimensional substrate, and are allowed to migrate, proliferate, and interact with each other, as illustrated in Figure 3.1(a)-(b). Different experimental geometries, such as circular barrier assays, are possible, and modern imaging technologies provide means of collecting high resolution images of the cell population as it evolves (Johnston et al., 2015).

An example of a two-dimensional cell biology assay is given in Figure 3.1(a)-(b). This kind of experimental design is routinely referred to as a *scratch assay*. Scratch assays are initiated by uniformly distributing a population of cells on a cell culture plate, which is then incubated for some time to allow cells to attach to the substrate and for the density of the monolayer of cells to increase. After incubation, a sharp-tipped instrument is used to scratch the monolayer to produce an artificial wound (Liang et al., 2007; Jin et al., 2016; Grada et al., 2017). The rate of the recolonisation of the wound space is then observed over time and has been demonstrated to depend on the rate of cell motility, the rate of cell proliferation, and the strength of cell-to-cell interaction forces. It is well known that quantifying the roles of these different mechanisms is challenging, as similar population-level outcomes can arise from different relative contributions of these separate mechanisms (Treloar et al., 2013a). One way of overcoming these issues is to modify the experimental procedure to deliberately separate the effects of cell migration from the effects of cell proliferation, and this approach is routinely used to improve our understanding of the role of cell motility in wound healing and malignant spreading (Glenn et al., 2016; Nyegaard et al., 2016; Grada et al., 2017). The experimental images in Figure 3.1(c)-(d) show a scratch assay that is prepared in exactly the same

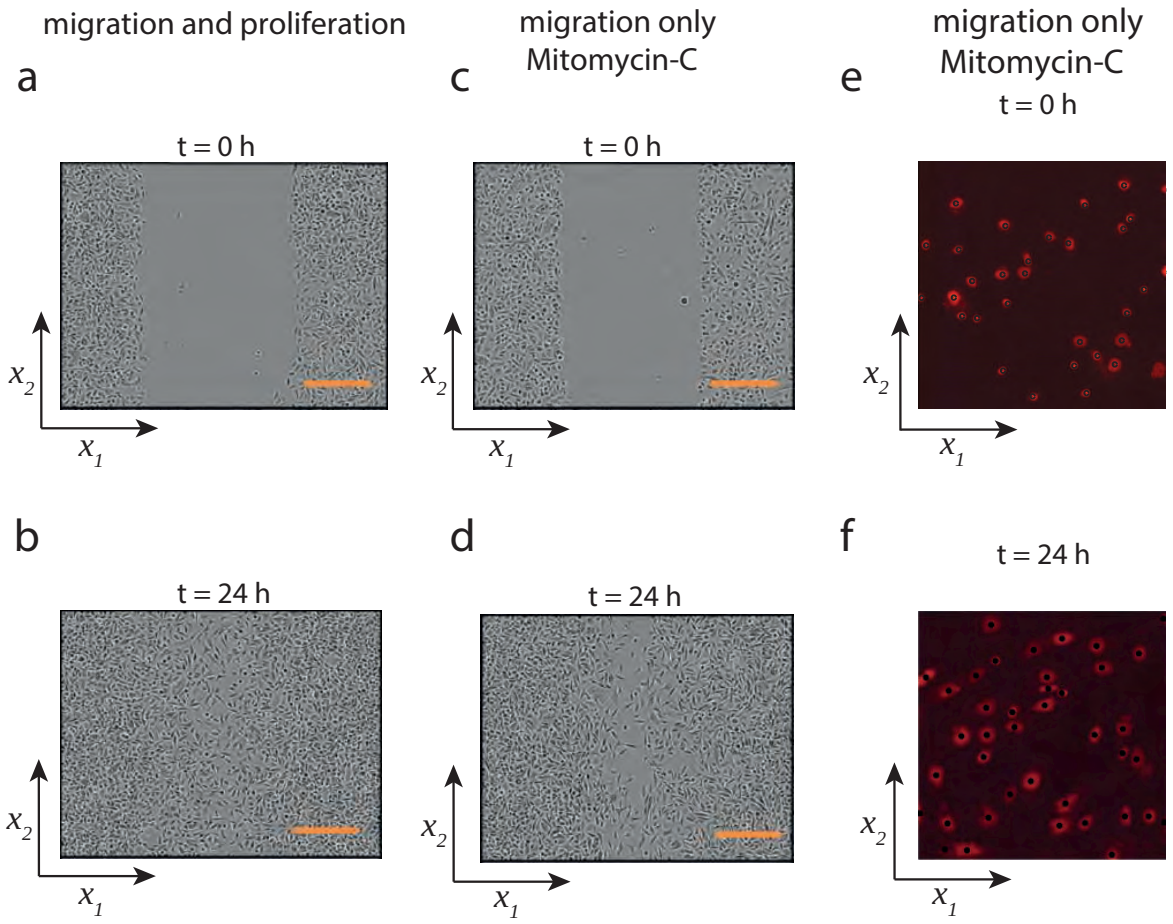


Figure 3.1: (a)-(b) An example of a typical scratch assay, where cells are allowed to close an artificially created gap. In this experiment the cells are PC-3 prostate cancer cells (Kaighn et al., 1979). (c)-(d) Scratch assay with PC-3 cells pretreated with Mitomycin-C to prevent proliferation. In (a)-(d) the scale bars correspond to $300\,\mu$m. (e)-(f) Images showing individual 3T3 fibroblast cells in a circular barrier assay where the cells are treated with Mitomycin-C. In these images a cell nucleus stain is used, and each individual cell is superimposed with a black disk. We denote the two-dimensional coordinates as $\{x_1, x_2\}$, as indicated. Images in (e)-(f) show a square region of length $400\,\mu$m. The images in (a)-(d) are reproduced with permission from Springer (Shah et al. 2016). The images in (e)-(f) are reproduced with permission from The Royal Society (Simpson et al., 2013b).

way as the experimental images in Figure 3.1(a)-(b) except that the cells are treated with a drug to inhibit proliferation (Shah et al., 2016). A visual comparison of the images in Figure 3.1(a)-(b) and Figure 3.1(c)-(d) shows that the combined effects of cell migration and cell proliferation lead to a more rapid wound closure. The experiment in Figure 3.1(c)-(d) involves treating the cells with a drug called Mitomycin-C (Sadeghi et al., 1998). Mitomycin-C is a chemotherapy drug that suppresses mitosis by blocking DNA replication (Sadeghi et al., 1998). While Mitomycin-C is known to prevent cell proliferation without inhibiting cell migration (Simpson et al., 2013b), a consequence of treating cells with Mitomycin-C is that cells increase in size, as if they are about to divide into two daughter cells, but the process of division does not take place. Therefore, cells that are treated with Mitomycin-C do not divide, but instead they approximately

double in size over a period of approximately 24 hours, as illustrated in Figure 3.1(e)-(f), where we show fibroblast cells in a circular barrier assay. The images in Figure 3.1(e)-(f) include a nuclear stain in red, and it is clear that the Mitomycin-C treated cells approximately double in size during the experiment. This change in cell size is typically neglected in mathematical models that describe the collective spreading of cell populations (Simpson et al., 2013b; Jin et al., 2016).

Many experimental images, such as the images in Figure 3.1(c)-(f), demonstrate the potential for significant changes in cell size which may influence the behaviour of the entire cell population since crowding effects are thought to be important in two-dimensional cell biology assays (Simpson et al., 2013b). Many classical continuum models of collective cell behaviour do not incorporate any measure of cell size (Maini et al., 2004; Sherratt and Murray, 1990). To address this limitation, another common approach to model collective cell behaviour is to use lattice-based stochastic models, where individual agents are allowed to move on a discrete lattice (Penington et al., 2011; Markham et al., 2015). Lattice-based models are attractive because they are conceptually straightforward, computationally efficient, and produce time-lapse images that are similar to images obtained from experiments (Simpson et al., 2013b). While lattice-based models typically associate the cell size with the lattice spacing (Simpson et al., 2013b), modelling the collective behaviour of populations of cells involving cell-to-cell crowding effects and dynamic changes in the size of individual cells in this approach is not straightforward. In particular, in a lattice-based model it is difficult to represent the cell size as a continuous function of time (Binder and Simpson, 2016). An alternative approach is to use a lattice-free stochastic model (Codling et al., 2008; Galle et al., 2005; Newman and Grima, 2004). Lattice-free models can be more computationally demanding than lattice-based models when dealing with crowding effects. However, lattice-free models are much more appealing than lattice-based models because agents in the simulation can assume a continuous size, which may be allowed to change dynamically.

While computational implementations of stochastic models are well suited to capturing individual-level details, experimental data is routinely presented in the form of population-level and tissue-level data. Consequently, it is convenient to have access to some continuum approximation to describe the collective behaviour associated with the stochastic simulations. The continuum description often takes the form of a partial differential equation (Penington et al., 2011; Dyson et al., 2012). Most continuum-based models used to describe collective behaviour of cell populations invoke the mean-field approximation (MFA) (Sherratt and Murray, 1990; Maini et al., 2004a; Penington et al., 2011; Simpson et al., 2013b). In effect, the MFA amounts to assuming that the positions of individuals in the population are independent. This assumption is widely invoked, both implicitly (Sherratt and Murray, 1990; Maini et al., 2004a) and explicitly

(Penington et al., 2011; Dyson et al., 2012; Simpson et al., 2013b). However, since spatial structure, such as clustering and patchiness, is often observed experimentally, the MFA is not always appropriate. Clustering and patchiness are observed in a range of natural processes, including cell biology experiments (Steinberg, 1996; Treloar et al., 2013a) and ecology (Levin and Whitfield, 1994), therefore it is also of interest to derive continuum limit descriptions that avoid the MFA, where appropriate. To achieve this, in this chapter we describe a lattice-free model of the collective spreading of a population of cells that incorporates cell motility, dynamic cell size changes, crowding effects, and cell-to-cell adhesion. We derive a continuum description using the standard MFA, as well as introduce an alternative continuum description using a more sophisticated moment dynamics approach (Middleton et al., 2014). Moment dynamics approaches are often used to describe spatially correlated populations in ecological applications and in the study of epidemics (Bolker and Pacala, 1997; Keeling et al., 1997; Sharkey et al., 2006; Sharkey et al., 2015). However, moment dynamics approaches are less common in the study of collective cell behaviour. Moment dynamics approaches can invoke many different approximations to account for spatial correlations (Murrell et al., 2004; House, 2014; Plank and Law, 2015; Binny et al., 2015; Binny et al., 2016a; Binny et al., 2016b), and in this chapter we use the Kirkwood superposition approximation (KSA) which was originally developed to describe the spatial arrangement of molecules in liquids (Kirkwood, 1935; Singer, 2004), and only much later adopted to describe the spatial arrangement of individual cells in the collective cell spreading (Baker and Simpson, 2010; Middleton et al., 2014).

In this chapter we present discrete and continuum descriptions of collective cell behaviour formulated in both one and two dimensions. Two-dimensional models allow us to reproduce the dynamics of experiments such as those depicted in Figure 3.1, and are perfectly suited for visualisation of the experiments. However, as we demonstrate in Chapter 3A, there is little motivation to use two-dimensional descriptions for the scratch assay geometry because the agent density, on average, does not depend on the vertical coordinate.

We denote the position of an arbitrary point in the computational domain by the vector $\vec{u} = \{x_1, x_2\}$. The positions of two other arbitrary points in the domain are given by the vectors $\vec{u}' = \{y_1, y_2\}$ and $\vec{u}'' = \{z_1, z_2\}$, and so on. We utilise the notation x, y, and z to denote the positions of distinct points in the one-dimensional domain when introducing continuum descriptions in one dimension. The position of the ith agent on a two-dimensional domain is $\mathbf{u}^{(i)}$. The position of an agent i in the one-dimensional discrete simulations is given by $x^{(i)}$. This choice of notation allows us to distinguish between the positions of agents in the discrete simulations and the coordinates of fixed points in the continuum description. Furthermore, it is consistent with our previous work which does not involve dynamical changes in agent size (Matsiaka et al., 2017;

Chapter 2).

This chapter is organised as follows. In Section 3.3.1 we describe a stochastic lattice-free model of collective cell migration. The model incorporates cell migration, cell crowding effects and cell-to-cell adhesion, and allows individual cells in the population to change size dynamically. In Section 3.3.2 we introduce two continuum descriptions of the stochastic model: (i) a mean-field based approximation; and (ii) a moment dynamics approximation, based on the KSA. In Section 3.4 we compare the averaged data obtained from repeated simulations of the stochastic model with numerical solutions of both continuum approximations. Finally, in Section 3.5 we summarise our findings and discuss the opportunities for further research.

3.3 Methods

3.3.1 Langevin stochastic model

In this section we describe the lattice-free stochastic model used to simulate the collective behaviour of a population of N agents. Many two-dimensional cell biology experiments, such as the scratch assays depicted in Figure 3.1(a)-(d), can be described using a one-dimensional coordinate system because the density of cells is independent of the vertical coordinate (Johnston et al., 2015). Therefore, we focus our attention and discussion on the one-dimensional discrete model in Chapter 3. Additional results and discussion relating to justifying the use of a one-dimensional model to describe a two-dimensional cell culture experiment is presented in Chapter 3A.

In this chapter we denote the cell diameter using $\delta(t) > 0$, and assume that the dynamic change in cell diameter is logistic,

$$\delta(t) = \frac{2\delta(0)}{1 + \exp(-kt)}, \tag{3.1}$$

where the parameter $k > 0$ describes the growth rate, $\delta(0)$ is the initial cell diameter, and $\lim_{t\to\infty} \delta(t) = 2\delta(0)$. Plots showing typical $\delta(t)$ for different choices of k are given in Figure 3.2. We note that the choice of using a logistic function for $\delta(t)$ is not essential, and all of the analysis presented here can be applied to any other suitable choice of growth model.

We assume that the movement of individual agents in the stochastic model is described by an equation of motion (Newman and Grima, 2004; Middleton et al., 2014). We adopt the Langevin stochastic model where the collective behaviour of the population

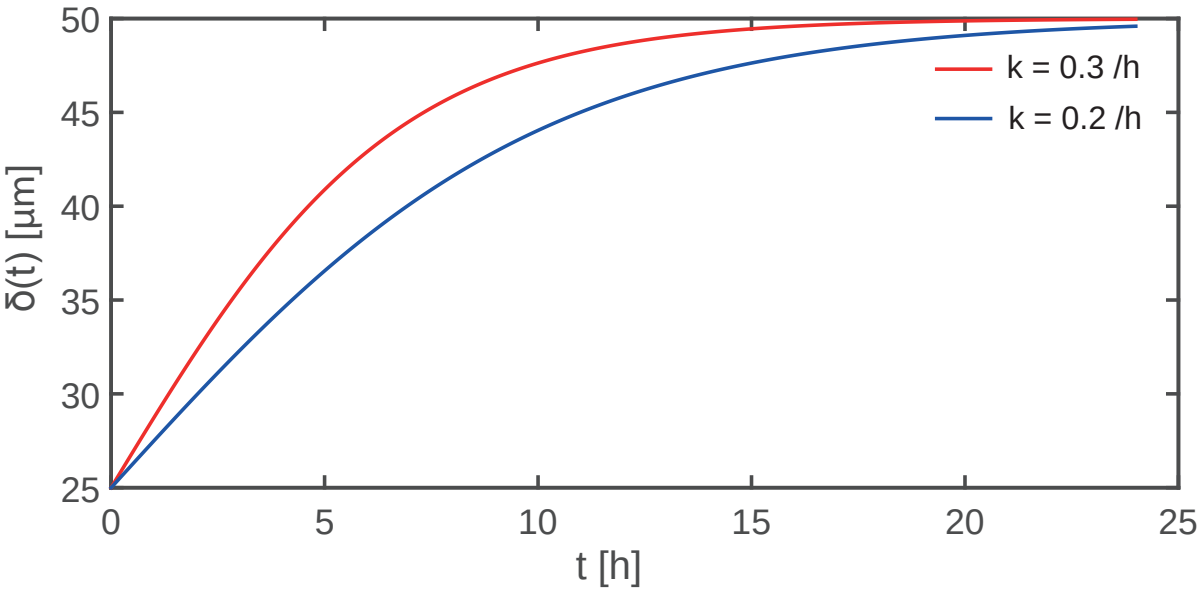


Figure 3.2: Logistic increase in agent diameter, given by Equation (3.1), for $k = 0.3$ /h (red) and $k = 0.2$ /h (blue). In both cases $\delta(0) = 25$ μm and $\lim_{t\to\infty} \delta(t) = 2\delta(0)$.

is described by a system of Langevin equations that can be written as

$$\frac{\mathrm{d}\mathbf{u}^{(i)}}{\mathrm{d}t} = \sum_{j\neq i} \mathbf{F}_{ij} + \boldsymbol{\xi}_i, \qquad i = 1, \ldots, N, \tag{3.2}$$

where $\mathbf{u}^{(i)}$ is the position of the ith agent in a two-dimensional space, $\mathbf{F}_{ij}$ is the interaction force between agent i and agent j, $\boldsymbol{\xi}_i$ is the stochastic force acting on the ith agent, and N is the number of agents in the simulation.

The corresponding one-dimensional model is given by

$$\frac{\mathrm{d}x^{(i)}}{\mathrm{d}t} = \sum_{j\neq i} F_{ij} + \xi_i, \qquad i = 1, \ldots, N. \tag{3.3}$$

Since we consider unbiased movements of isolated individuals, we sample ξ_i from a Gaussian distribution (Newman and Grima, 2004) with variance

$$\mathrm{Var}(\xi_i) = \frac{2D}{\Delta t}, \tag{3.4}$$

where Δt is the value of the time step used to solve the system of Langevin equations in a simulation of the stochastic model. Choosing Δt in this way ensures that the mean squared displacement of an isolated individual agent is independent of the time step chosen to simulate the stochastic model, and the mean squared displacement of an isolated agent is therefore $2Dt$.

The force function, F_{ij}, is chosen to be

$$F_{ij} = f_0 \, Z(r,t) \, \mathrm{sgn}(x^{(i)} - x^{(j)}), \tag{3.5}$$

where f_0 is a constant that describes the strength of the interaction forces, $Z(r,t)$ is the dimensionless function describing how the interaction force depends on the separation of the agents, $r = |x^{(i)} - x^{(j)}|$, t is time, and sgn is the *signum* function (Middleton et al., 2014; Matsiaka et al., 2017; Chapter 2). A schematic showing the arrangement of agents in the model is given in Figure 3.3(a)-(b), where we can see the effects of agent movement and the increase in the size of the agents with time.

We consider two main features of agent-to-agent interactions: (i) a short range repulsion force, which can be thought of as a resistance to deformations and crowding; and, (ii) longer range attraction forces which can be thought of as adhesion between agents. To model these two forces we use a modified Morse potential,

$$Z(r,t) = \begin{cases} 2\Big(\exp(-2a\,[r - \delta(t)]) - \exp(-a\,[r - \delta(t)])\Big), & r < 2\delta(t), \\ 2\Big(\exp(-2a\,[r - \delta(t)]) - \exp(-a\,[r - \delta(t)])\Big)\, g(r,t), & 2\delta(t) \le r \le 3\delta(t), \\ 0, & r > 3\delta(t), \end{cases} \tag{3.6}$$

where $a > 0$ is a parameter that controls the shape of the force function, $\delta(t)$ is the time-dependent agent diameter, given by Equation (3.1) or some other appropriate functional form. Here, the spatial range of interactions is finite, and set to three agent diameters, giving $Z(r,t) = 0$ for $r > 3\delta(t)$. The function $g(r)$ is the Tersoff cut-off function (Tersoff et al., 1987), which is included to capture the finite range of interactions,

$$g(r,t) = \frac{1}{2}\Big(1 - \sin\Big[\frac{\pi(2r - \delta(t))}{2\delta(t)}\Big]\Big). \tag{3.7}$$

Figure 3.3(c) shows a typical interaction function, $Z(r,t)$, over a period of 24 hours. At each instant in time the interaction force function incorporates repulsion at short distances, and attraction at longer distances, up to a finite range of three agent diameters. Since $\delta(t)$ increases with t, the interaction function also changes with time, and we can interpret the change in $Z(r,t)$ with t in Figure 3.3(c) as a result of the increase in agent diameter with time. The choice of the force function in Equation (3.6) is not unique, but rather one of many other possible functional forms that incorporate short range repulsion and longer range attraction such as the Lennard-Jones potential, Hertz potential, or a nonlinear spring model (Byrne and Preziosi, 2003; Jeon et al., 2010; Murray et al., 2012).

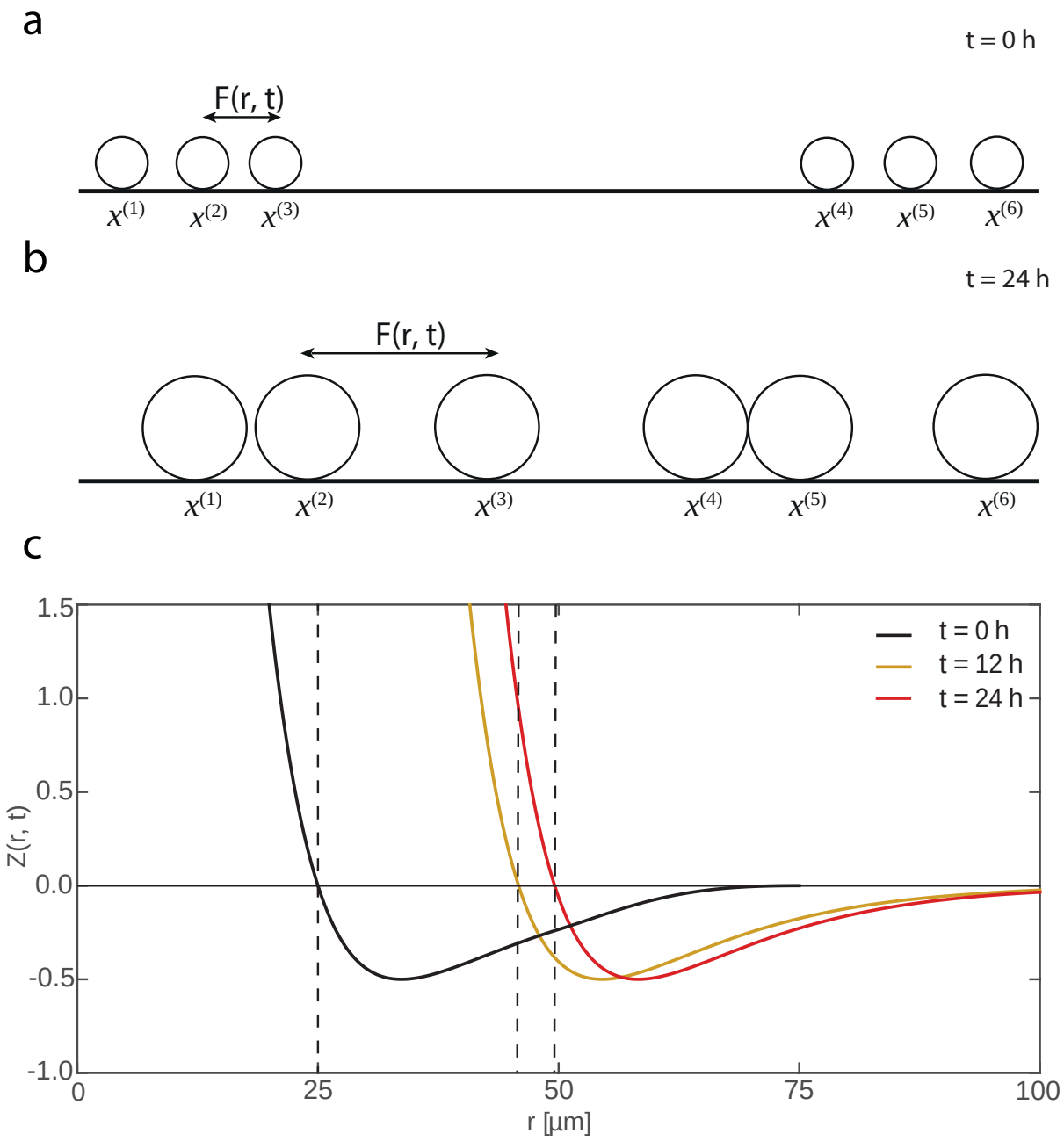


Figure 3.3: (a)-(b) Schematic illustration of individual agent motility and interaction forces in a population of agents where the diameters of individual agents double over an interval of approximately $t = 24\,\mathrm{h}$, as described by Equation (3.1). (c) Dimensionless potential, $Z(r,t)$, at $t = 0, 12$, and $24\,\mathrm{h}$ in black, orange and red, respectively. All plots correspond to $k = 0.2\,/\mathrm{h}$ and $\delta(0) = 25\ \mu$m. The horizontal line at $Z(r,t) = 0$ denotes the change from short range repulsion ($Z(r,t) > 0$) to longer range attraction ($Z(r,t) < 0$). The series of three vertical dashed lines indicate the diameter of the agent at $t = 0, 12$ and 24 hours.

3.3.2 Continuum description

In this section we present two different continuum approximations of the lattice-free stochastic model described in Section 3.3.1. In particular, we present both a standard continuum approximation, based on invoking the MFA, and a moment dynamics continuum approximation, based on invoking the KSA. We note that the process of deriving both continuum approximations has been presented previously in the simpler case where

the agent diameter is a constant (Middleton et al. 2014; Matsiaka et al. 2017; Chapter 2). The focus of the current study is to consider the continuum approximations for the case where the agent diameter increases with time. Therefore, we do not repeat the derivation of the continuum limit descriptions here in Chapter 3. Instead, complete details of the derivations are given in Chapter 3A, and here we focus on reporting the continuum descriptions and examining the accuracy of the continuum descriptions.

Two-dimensional continuum model

Here we first present the two-dimensional continuum limit of the stochastic model, and later consider the one-dimensional analogue of this model. The mean-field continuum description is given by an integro-partial differential equation (IPDE),

$$\frac{\partial p_1(\mathbf{u},t)}{\partial t} = D\Delta p_1(\mathbf{u},t) - (N-1)\nabla(p_1(\mathbf{u},t)\,\mathbf{V}(\mathbf{u},t)), \tag{3.8}$$

where the agent density, $p_1(\mathbf{u},t)$, depends on the position $\mathbf{u} = \{x_1, x_2\}$ and time t, D is the diffusivity, N is the number of agents, operator $\Delta = \partial^2/\partial x_1^2 + \partial^2/\partial x_2^2$, operator $\nabla = \partial/\partial\vec{\mathbf{u}} = \{\partial/\partial x_1, \partial/\partial x_2\}$, and

$$\mathbf{V}(\mathbf{u},t) = \int_\Omega \mathbf{F}(\mathbf{u}-\mathbf{u}',t)\,p_1(\mathbf{u}',t)\,\mathrm{d}\mathbf{u}' \tag{3.9}$$

is the velocity field induced by the agent-to-agent interactions. The force function $\mathbf{F}(\mathbf{u}-\mathbf{u}',t)$ is defined as

$$\mathbf{F}(\mathbf{u}-\mathbf{u}',t) = f_0\,Z(|\mathbf{u}-\mathbf{u}'|,t)\,\frac{\mathbf{u}-\mathbf{u}'}{|\mathbf{u}-\mathbf{u}'|}, \tag{3.10}$$

and depends on the separation distance, $|\mathbf{u}-\mathbf{u}'|$ and time t.

The two-dimensional moment dynamics model, based on the KSA approximation (Singer, 2004; Middleton et al., 2014), can be written as

$$\frac{\partial p_1(\mathbf{u},t)}{\partial t} = D\Delta p_1(\mathbf{u},t) - (N-1)\nabla\Big(\int_\Omega \mathbf{F}(\mathbf{u}-\mathbf{u}',t)\, p_2(\mathbf{u},\mathbf{u}',t)\,\mathrm{d}\mathbf{u}'\Big), \tag{3.11}$$

$$\begin{aligned}\frac{\partial p_2(\mathbf{u},\mathbf{u}',t)}{\partial t} &= D\Delta p_2(\mathbf{u},\mathbf{u}',t)\\ &\quad - \frac{\partial}{\partial \mathbf{u}}\Big(\mathbf{F}(\mathbf{u}-\mathbf{u}',t)\, p_2(\mathbf{u},\mathbf{u}',t)\Big)\\ &\quad - \frac{\partial}{\partial \mathbf{u}'}\Big(\mathbf{F}(\mathbf{u}'-\mathbf{u},t)\, p_2(\mathbf{u},\mathbf{u}',t)\Big)\\ &\quad - (N-2)\,\frac{\partial}{\partial \mathbf{u}}\int_\Omega \mathbf{F}(\mathbf{u}-\mathbf{u}'',t)\,\frac{p_2(\mathbf{u},\mathbf{u}',t)\, p_2(\mathbf{u},\mathbf{u}'',t)\, p_2(\mathbf{u}',\mathbf{u}'',t)}{p_1(\mathbf{u},t)\, p_1(\mathbf{u}',t)\, p_1(\mathbf{u}'',t)}\,\mathrm{d}\mathbf{u}''\\ &\quad - (N-2)\,\frac{\partial}{\partial \mathbf{u}'}\int_\Omega \mathbf{F}(\mathbf{u}'-\mathbf{u}'',t)\,\frac{p_2(\mathbf{u},\mathbf{u}',t)\, p_2(\mathbf{u},\mathbf{u}'',t)\, p_2(\mathbf{u}',\mathbf{u}'',t)}{p_1(\mathbf{u},t)\, p_1(\mathbf{u}',t)\, p_1(\mathbf{u}'',t)}\,\mathrm{d}\mathbf{u}'',\end{aligned} \tag{3.12}$$

where $p_2(\mathbf{u},\mathbf{u}',t)$ is the density-density correlation function that captures correlations in the positions of agents at locations $\mathbf{u}$ and $\mathbf{u}'$, at time t (Middleton et al., 2014), operator Δ is defined as $\Delta = \partial^2/\partial x_1^2 + \partial^2/\partial x_2^2$ in Equation (3.11), and $\Delta = \partial^2/\partial x_1^2 + \partial^2/\partial x_2^2 + \partial^2/\partial x_1^{2\prime} + \partial^2/\partial x_2^{2\prime}$ in Equation (3.12), respectively.

Simplified one-dimensional continuum model

The experimental images depicted in Figure 3.1(a)-(d) demonstrate the evolution of the cell population on a two-dimensional substrate. We note that the cell density in a scratch assay is independent of the vertical coordinate (Figure 3.1(a)-(d)) so that the experiment can be described in terms of a one-dimensional coordinate system. The underlying discrete model for Equation (3.8) and Equations (3.11)-(3.12) is a system of Langevin equations, Equation (3.3), as described in Section 3.3.1. To justify the one-dimensional approach we reproduce the experimental image in Figure 3.1(a) using two-dimensional Langevin model, and compare the density profiles obtained from the two-dimensional model with the results from the one-dimensional stochastic model (Chapter 3A). Results presented in Figure 3A.1 (Chapter 3A) show that the simpler one-dimensional model produces similar results to the two-dimensional model for this special initial condition where the agent density is independent of the vertical coordinate. Motivated by these considerations, we neglect density variations in the vertical direction and write Equation (3.8) and Equations (3.11)-(3.12) in a one-dimensional format. We note that the use of one-dimensional continuum models to understand and interpret multidimensional transport phenomena with appropriate symmetry imposed by the initial conditions and boundary conditions is relatively common in both the mathematical biology literature (Callaghan et al., 2006; Khain et al., 2011; Smith et al., 2017), as well as in other areas of engineering and applied science (Simpson, 2009b).

The mean-field continuum limit of the one-dimensional stochastic model is given by an IDPE that can be written as

$$\frac{\partial p_1(x,t)}{\partial t} = D\Delta p_1(x,t) - (N-1)\nabla(p_1(x,t)\,V(x,t)), \tag{3.13}$$

where the agent density, $p_1(x,t)$, depends on the position x and time t, operator $\Delta = \partial^2/\partial x^2$, operator $\nabla = \partial/\partial x$, and

$$V(x,t) = \int_\Omega F(x-y,t)\,p_1(y,t)\,\mathrm{d}y, \tag{3.14}$$

is the velocity field induced by the interactions between agents. We note that the diffusivity in Equation (3.13) is directly related to the stochastic force, ξ_i, in the stochastic model, Equation (3.3).

As stated previously, continuum limit descriptions based on the MFA amount to assuming that the positions of agents are independent. However, in many practical situations, cells and other living organisms can adhere to each other and form clusters (Steinberg, 1996). In these situations the assumptions underpinning MFA-based continuum models are questionable. To address this limitation, we now make use of a more sophisticated moment dynamics continuum description that accounts for the density of agents and the density of pairs of agents. The moment dynamics model, based on the KSA, can be written as

$$\frac{\partial p_1(x,t)}{\partial t} = D\Delta p_1(x,t) - (N-1)\nabla\Big(\int_\Omega F(x-y,t)\,p_2(x,y,t)\,\mathrm{d}y\Big), \tag{3.15}$$

$$\begin{aligned}\frac{\partial p_2(x,y,t)}{\partial t} &= D\Delta p_2(x,y,t)\\ &\quad - \frac{\partial}{\partial x}\Big(F(x-y,t)\,p_2(x,y,t)\Big)\\ &\quad - \frac{\partial}{\partial y}\Big(F(y-x,t)\,p_2(x,y,t)\Big)\\ &\quad - (N-2)\,\frac{\partial}{\partial x}\int_\Omega F(x-z,t)\,\frac{p_2(x,y,t)\,p_2(x,z,t)\,p_2(y,z,t)}{p_1(x,t)\,p_1(y,t)\,p_1(z,t)}\,\mathrm{d}z\\ &\quad - (N-2)\,\frac{\partial}{\partial y}\int_\Omega F(y-z,t)\,\frac{p_2(x,y,t)\,p_2(x,z,t)\,p_2(y,z,t)}{p_1(x,t)\,p_1(y,t)\,p_1(z,t)}\,\mathrm{d}z,\end{aligned} \tag{3.16}$$

where $p_1(x,t)$ is the average density of agents at location x and time t, and $p_2(x,y,t)$ is the density-density correlation function that captures correlations in the positions of agents at locations x and y at time t, operator Δ is defined as $\Delta = \partial^2/\partial x^2$ in Equation (3.15), and $\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2$ in Equation (3.16), respectively.

3.4 Results and discussion

Here we focus our attention on a typical scratch assay geometry (Figure 3.1(a)-(d)). In this experiment the cell density does not depend, on average, on the vertical position (Figure 3.1(a)-(d)). Therefore, we can approximate this experiment as a one-dimensional problem.

The experimental images in Figure 3.1(a)-(d)) show only a small region of the population, which extends well beyond the vertical boundaries of the image (Johnston et al., 2015). We apply periodic boundary conditions in all simulations (Middleton et al., 2014). The initial condition is given by sampling from a distribution, $\alpha(x)$, that is given by

$$\alpha(x) = \begin{cases} 12.5 \times 10^{-3}, & 0\ \mu\text{m} \leq x < 600\ \mu\text{m}, \\ 0, & 600\ \mu\text{m} \leq x \leq 1400\ \mu\text{m}, \\ 12.5 \times 10^{-3}, & 1400\ \mu\text{m} < x \leq 2000\ \mu\text{m}, \end{cases} \tag{3.17}$$

where the length of the domain, 2000 μm, is a typical width of an experimental image (Jin et al., 2016). To initialise the stochastic simulations we randomly place agents in the two intervals, $0 \leq x \leq 600$ μm, and $1400 \leq x \leq 2000$ μm. This initial distribution of agents in the simulations mimics the initial distribution of cells in the images of scratch assays, given in Figure 3.1(a) and Figure 3.1(c). In all of our results we report the agent density profiles in terms of both the dimensional density of agents, $p_1(x,t)$ [agents/μm], as well as a non-dimensional agent density, $p_1(x,t)/C$, where $C = \mathcal{N}\delta(0)/L$ is the carrying capacity density of agents with diameter $\delta(0)$. Here $\mathcal{N}$ is the maximum number of agents of diameter $\delta(0)$ that can be distributed, without compression, along a line of length L. In Equation (3.17) we choose the maximum density to be 12.5×10^{-3} cells/μm because this corresponds to a non-dimensional density of approximately $p_1(x,t)/C = 0.625$, and this is a typical initial density used in practice (Jin et al 2016; Liang et al. 2007). To initialise our simulations we place agents of diameter $\delta(0)$, at random, until the density of agents in the two intervals, $0 \leq x \leq 600$ μm, and $1400 \leq x \leq 2000$ μm, is $p_1(x,t)/C = 0.625$.

We describe the evolution of the scratch assay in three different ways. First, we perform individual realisations of the stochastic model to produce individual snapshots showing the distribution of agents in each simulation. Second, we perform a large number of identically prepared realisations of the stochastic model, and count the numbers of agents in I equally-spaced intervals across the domain. Averaging the number of agents in each interval allows us to quantify the spatial variation in the average agent density. Finally, we solve both the MFA- and KSA-based continuum models numerically, and compare the solutions of the continuum models with the average data from the suite of

stochastic simulations to examine the accuracy of the two continuum descriptions. To quantify the accuracy of the two continuum descriptions we use

$$\mathrm{E}_{\mathrm{MFA}}(t) = \frac{1}{I}\sum_{i=1}^{I}\Big[p_{1,\mathrm{MFA}}(i,t) - p_{1,\mathrm{discrete}}(i,t)\Big]^2, \tag{3.18}$$

$$\mathrm{E}_{\mathrm{KSA}}(t) = \frac{1}{I}\sum_{i=1}^{I}\Big[p_{1,\mathrm{KSA}}(i,t) - p_{1,\mathrm{discrete}}(i,t)\Big]^2, \tag{3.19}$$

where, the index i denotes spatial node, I is the total number of nodes used to quantify the averaged agent density, $p_{1,\mathrm{MFA}}(i,t)$ is the density of agents predicted by the MFA-based continuum model, $p_{1,\mathrm{KSA}}(i,t)$ is the density of agents predicted by the KSA-based continuum model, and $p_{1,\mathrm{discrete}}(i,t)$ is the density of agents predicted by averaging a suite of identically prepared realizations of the stochastic Langevin model. Here, $\mathrm{E}_{\mathrm{MFA}}(t)$ is a measure of the error associated with the MFA-based continuum description, and $\mathrm{E}_{\mathrm{KSA}}(t)$ is a measure of the error associated with the KSA-based continuum description.

We first examine the most straightforward situation where we consider a scratch assay with a population of cells where the diameter of cells remains constant. Following this preliminary case, we then examine two additional situations where we consider the same scratch assay except that the diameter of individual agents in the populations increases at different rates: (i) relatively slow growth, $k = 0.2$ /h; and (ii) faster growth, $k = 0.3$ /h. In all cases we fix the diffusivity to be a typical value for the PC-3 prostate cancer cell line, $D = 1200\,\mu\mathrm{m}^2/\mathrm{h}$ (Jin et al. 2016).

The stochastic model, Equation (3.3), is solved numerically using a first order explicit Euler method (Press et al. 2007). The number of individual realisations used to construct the density profiles is chosen to be 10^5. This choice produces averaged density data with fluctuations that are two orders of magnitude smaller than the density data. For example, at $t = 0$, the density in the region $x < 600\,\mu$m is approximately 10^{-2} agents/μm, whereas the standard deviation of the agent density is approximately 10^{-4} agents/μm. After performing 10^5 identically prepared simulations of the stochastic model, the spatio-temporal distribution of agent density is estimated by averaging results from identically-prepared realisations. The initial condition for the MFA-based continuum model, Equation (3.13), is $p_1(x,0) = \alpha(x)$, and the solution of the MFA-based continuum model is obtained by solving Equation (3.13) numerically, as described in Chapter 3A. Since we choose the positions of agents in the stochastic simulations to be random at $t = 0$, there are no correlations in the initial distribution of agents. Consequently, the initial condition for the KSA-based continuum model, Equations (3.15)-(3.16), is $p_1(x,0) = \alpha(x)$, and $p_2(x,y,0) = \alpha(x)\alpha(y)$. To predict the evolution of the system using the KSA-based continuum model, numerical solutions of Equations (3.15)-(3.16) are obtained using techniques outlined in Chapter 3A. In all cases, the numerical

solutions of both continuum models are obtained using a sufficiently fine spatial and temporal discretisation that the results are grid independent.

The results in Figures 3.4-3.6 compare solutions of the MFA- and KSA-based continuum models with the averaged results from stochastic simulations in the cases of: (i) no growth, $k = 0$ /h; (ii) relatively slow growth, $k = 0.2$ /h; and (iii) faster growth, $k = 0.3$ /h. To quantify the performance of the MFA and KSA models we compute the time evolution of $E_{MFA}(t)$ and $E_{KSA}(t)$, given by Equations (3.18)–(3.19), respectively. The only difference between the continuum-stochastic comparisons in Figures 3.4-3.6 is the rate of increase of the agent diameter, k. All other parameters in the simulations in Figures 3.4-3.6 are held constant to avoid ambiguity and to highlight the influence of agent growth on the dynamics of the population and the performance of the two different continuum descriptions.

The results in Figure 3.4(a)-(c), Figure 3.5(a)-(c) and Figure 3.6(a)-(c) show stochastic simulations evolving from different realisations of the initial condition, Equation (3.17), to give a spatial distribution of agents after 24 and 48 hours. Note that the distributions of agents in Figure 3.4(a)-(c), Figure 3.5(a)-(c) and Figure 3.6(a)-(c) are given as a series of 100 separate, one-dimensional simulations that are plotted adjacent to each other (Matsiaka et al., 2017; Chapter 2). Presenting the stochastic results in this way is convenient because it highlights the randomness in the stochastic model. In general, we see that over a period of 48 hours the wound, of initial width 800 μm, becomes recolonised by agents and the wound appears to close. Comparing the evolution of the stochastic models in Figure 3.4(a)-(c), Figure 3.5(a)-(c) and Figure 3.6(a)-(c) with the experimental images in Figure 3.1(c)-(d) suggests that this choice of parameters in the stochastic model is reasonable, as the rate of wound closure in the stochastic simulations is similar to the rate of wound closure in the experimental images.

In the case where there is no growth (Figure 3.4), the solution of the MFA-based continuum model matches the averaged agent density profile from the stochastic simulations very well (Figure 3.4(d)). Similarly, comparing the solution of the KSA-based continuum model with the averaged agent density profile (Figure 3.4(e)) reveals an excellent match. In this case, there seems to be little justification for use of the more complicated KSA-based continuum model as the simpler MFA-based model captures the evolution of the averaged agent density extremely well. Furthermore, quantitative comparison of the accuracy of the MFA-based continuum model with the accuracy of the KSA-based continuum model (Figure 3.4(f)) confirms that there is no advantage in using the KSA-based model for this problem where the size of the agents remains fixed.

In contrast, when we consider the situation where agents increase in size, $k > 0$ (Figures 3.5 and 3.6), the improved performance of the KSA-based continuum model becomes clear. Results in Figure 3.5(d) compare the evolution of the MFA-based model and averaged agent density data from the stochastic model, showing that there is a

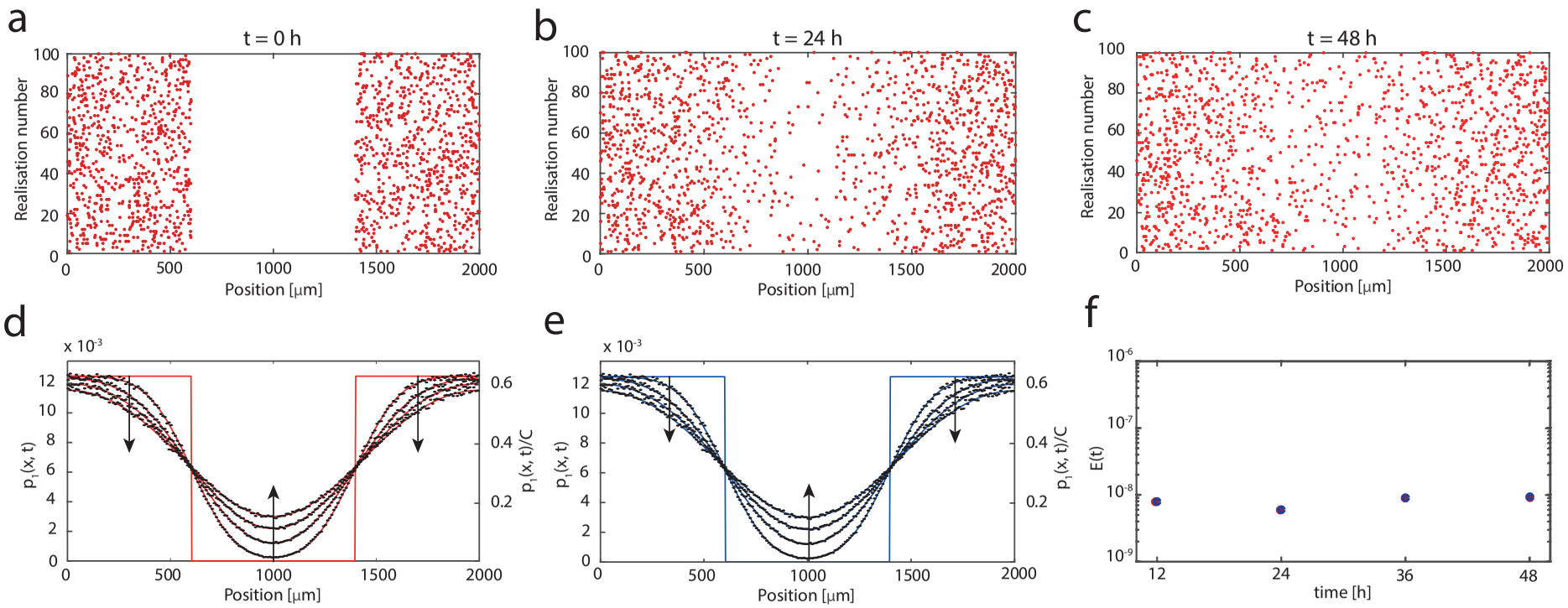


Figure 3.4: Comparison of the averaged agent density data from stochastic simulations and the solutions of the MFA- and KSA-based continuum models for a scratch assay with constant cell size, $\delta(t) \equiv 25\,\mu$m. Snapshots in (a)-(c) show 100 identically prepared realisations of the stochastic model, given by Equation (3.3), at $t = 0, 24$ and 48 h, respectively. The initial density is given by Equation (3.17). Results in (d) and (e) show the averaged agent density profiles using an ensemble of 10^5 identically prepared simulations (black dots) with binsize 10 μm. The averaged results are compared with solutions of the MFA-based continuum model, Equation (3.13) (red lines), and the KSA-based continuum model, Equation (3.16) (blue lines). Profiles are given at $t = 0, 12, 24, 36$, and 48 h, with the arrows indicating the direction of increasing t. Results in (f) show the evolution of the mean squared errors (Equations (3.18) and (3.19)) for the MFA- and KSA-based continuum models. Equation (3.3) is integrated with $\Delta t = 2 \times 10^{-2}$ h, Equation (3.13) is integrated with $\Delta x = 4$ μm and $\Delta t = 1 \times 10^{-3}$ h, and Equation (3.16) is integrated with $\Delta x = \Delta y = 4$ μm, and $\Delta t = 1 \times 10^{-3}$ h. The remaining parameters are $N = 15$, $a = 0.05\,/\mu$m, $f_0 = 0.1$ μm/h.

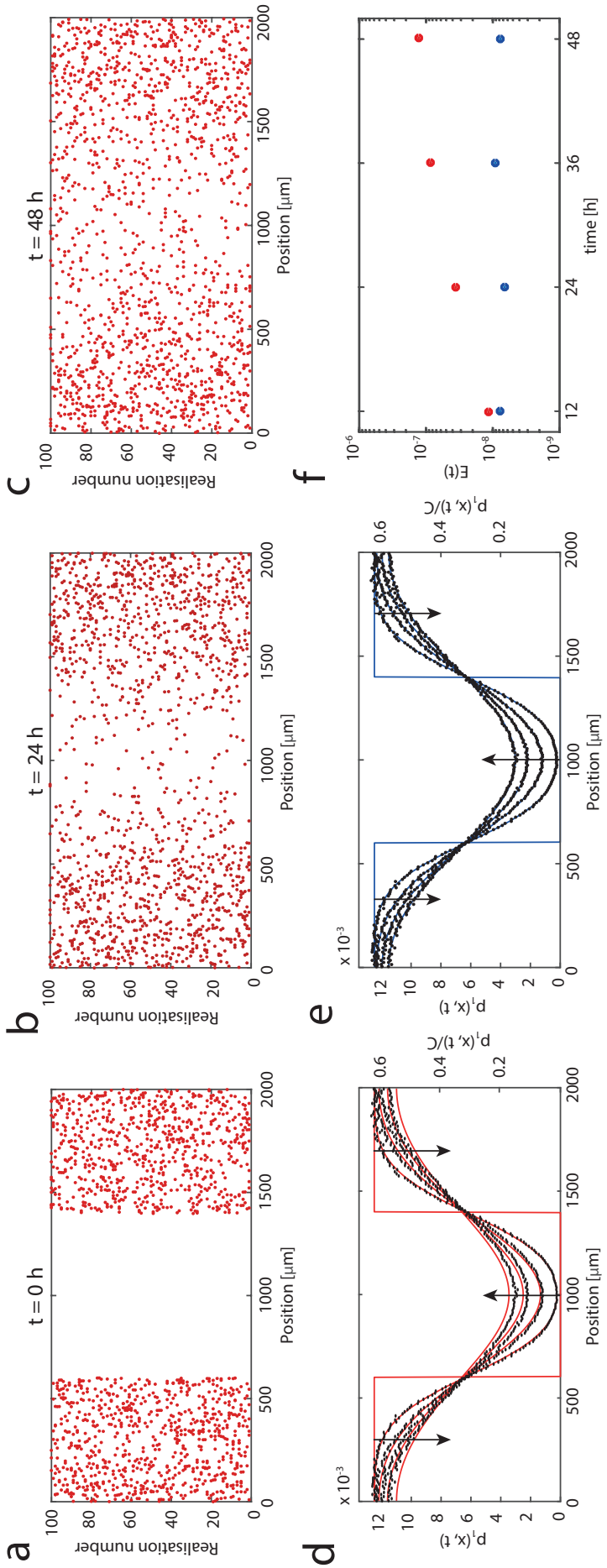


Figure 3.5: Comparison of the averaged agent density data from stochastic simulations and the solutions of the MFA- and KSA-based continuum models for a scratch assay with agents where the diameter grows relatively slowly, with $k = 0.2$ /h. Snapshots in (a)-(c) show 100 identically prepared realisations of the stochastic model, given by Equation (3.3), at $t = 0, 24$ and 48 h, respectively. The initial density is given by Equation (3.17). Results in (d) and (e) show the averaged agent density profiles using an ensemble of 10^5 identically prepared simulations (black dots) with binsize 10 μm. The averaged results are compared with solutions of the MFA-based continuum model, Equation (3.13) (red lines), and the KSA-based continuum model, Equation (3.16) (blue lines). Profiles are given at $t = 0, 12, 24, 36$, and 48 h, with the arrows indicating the direction of increasing t. Results in (f) show the evolution of the mean squared errors (Equations (3.18) and (3.19)) for the MFA- and KSA-based continuum models. Equation (3.3) is integrated with $\Delta t = 2 \times 10^{-2}$ h, Equation (3.13) is integrated with $\Delta x = 4$ μm and $\Delta t = 1 \times 10^{-3}$ h, and Equation (3.16) is integrated with $\Delta x = \Delta y = 4$ μm, and $\Delta t = 1 \times 10^{-3}$ h. The remaining parameters are $N = 15$, $a = 0.05$ /μm, $f_0 = 0.1$ μm/h.

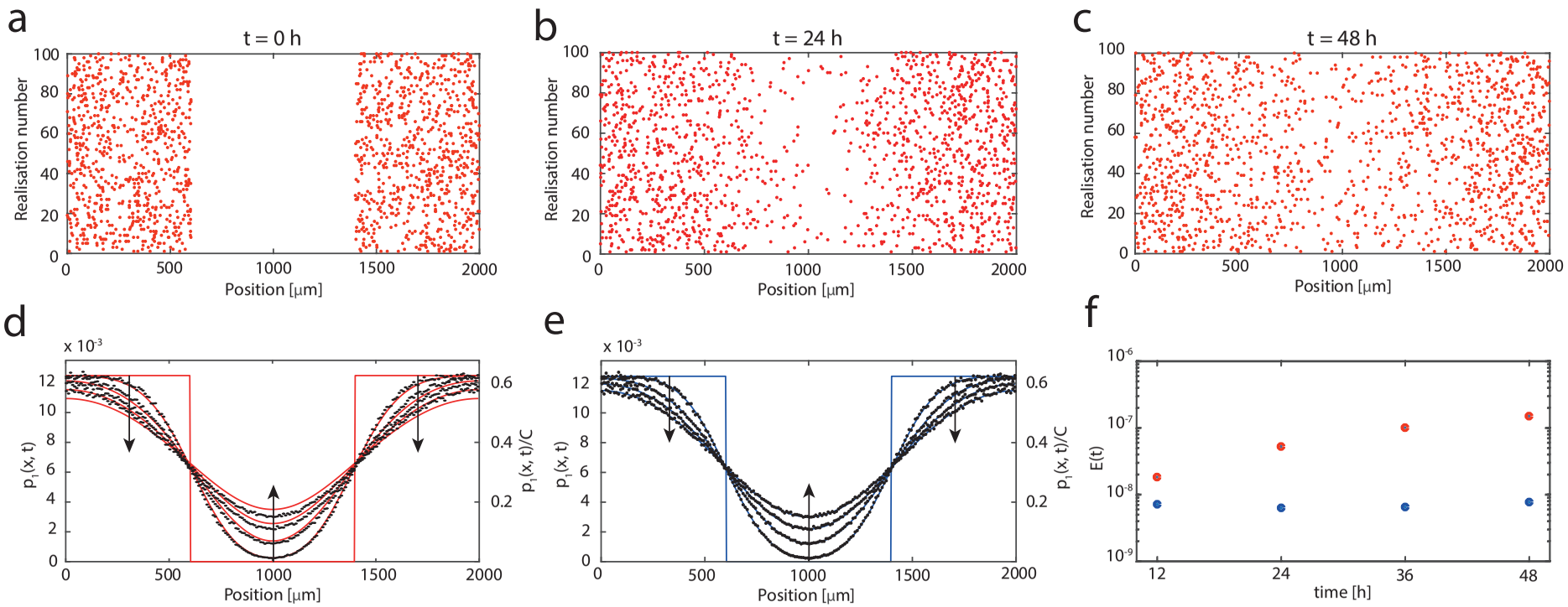


Figure 3.6: Comparison of the averaged agent density data from stochastic simulations and the solutions of the MFA- and KSA-based continuum models for a scratch assay with agents where the diameter grows faster, with $k = 0.3$ /h. Snapshots in (a)-(c) show 100 identically prepared realisations of the stochastic model, given by Equation (3.3), at $t = 0, 24$ and 48 h, respectively. The initial density is given by Equation (3.17). Results in (d) and (e) show the averaged agent density profiles using an ensemble of 10^5 identically prepared simulations (black dots) with binsize 10 μm. The averaged results are compared with solutions of the MFA-based continuum model, Equation (3.13) (red lines), and the KSA-based continuum model, Equation (3.16) (blue lines). Profiles are given at $t = 0, 12, 24, 36,$ and 48 h, with the arrows indicating the direction of increasing t. Results in (f) show the evolution of the mean squared errors (Equations (3.18) and (3.19)) for the MFA- and KSA-based continuum models. Equation (3.3) is integrated with $\Delta t = 2 \times 10^{-2}$ h, Equation (3.13) is integrated with $\Delta x = 4$ μm and $\Delta t = 1 \times 10^{-3}$ h, and Equation (3.16) is integrated with $\Delta x = \Delta y = 4$ μm, and $\Delta t = 1 \times 10^{-3}$ h. The remaining parameters are $N = 15$, $a = 0.05\,/\mu$m, $f_0 = 0.1$ μm/h.

clear and visually discernible difference between two sets of profiles. This difference is quantified in Figure 3.5(f). In contrast, the accuracy of the KSA-based continuum model, shown in Figure 3.5(f), remains excellent. Similar comparisons between the performance of the MFA- and KSA-based continuum models in Figure 3.6 for faster growth confirms the improved accuracy of the KSA-based continuum model in the case when the agents are allowed to increase in size dynamically.

3.5 Conclusions

In this chapter we present stochastic and continuum models of collective cell migration that can be applied to mimic scratch assays. In particular, we pay careful attention to allow for the case where the agents in the stochastic simulation change size dynamically. This feature can be important when we consider scratch assays with cells that are treated with Mitomycin-C to prevent proliferation. The stochastic model we present takes the form of a system of Langevin equations, and this framework can be used to describe the collective behaviour of a population of cells with constant size, or the collective behaviour of a population of cells with variable size. In addition to considering variable cell size, the stochastic framework describes random cell motility, crowding effects via a short range repulsive force, and cell-to-cell adhesive effects via longer range attraction. There is a crucial difference between two key parameter regimes that we consider: constant agent size, and dynamically increasing agent size. In the case when agent size remains constant, the average force acting on each individual agent remains approximately constant in regions of spatially uniform density. In contrast, increasing the size of individual agents leads to increased interactions between agents. Consequently, when we consider cases where the agent size increases dynamically, the MFA continuum description provides poorer match to the averaged discrete results at later times (Figure 3.5(d), Figure 3.6(d)). We note that the growth rate introduced in the stochastic model does not depend on the spatial positions of agents. As a result, it means that certain biological phenomena, such as contact growth inhibition, are not included in the model and local cell density can be over 100% if interaction forces are strong.

In addition to relying on repeated stochastic simulations, we also wish to develop continuum approximations of the stochastic model so that we can predict population-level and tissue-level data. To achieve this we first consider a continuum description based on the usual MFA that neglects correlations in the positions of agents. In this approach the position of any individual agent is treated as being independent of the positions of all other agents. The MFA-based model is relatively fast to simulate, as it takes only a few minutes to produce results depicted in Figure 3.4d, Figure 3.5d, and Figure 3.6d on a single desktop machine. While the MFA leads to a straightforward continuum model, the neglect of spatial correlations suggests that the approach might not always be valid, for example in situations where spatial structure and clustering develops. To overcome these potential limitations we also make use of a more advanced moment dynamics approach using the KSA. The KSA-based continuum description is more complicated to derive and much more numerically intense than the MFA-based approach, but it is attractive since it avoids inaccuracies of the MFA.

Generally, both the MFA- and KSA-based continuum models lead to reasonable pre-

dictions of the averaged stochastic results for the experimentally motivated problems that we describe here. In the case when the agent size remains constant, both the MFA- and KSA-based continuum models lead to an excellent match with the averaged data from the stochastic simulations. In this case, the simpler MFA-based continuum model is preferable to the computationally expensive KSA-based continuum model. However, in cases where agents increase in size, the KSA-based model outperforms the MFA-based model. This is due to the fact that agent growth increases agent-to-agent crowding effects, and these effects are incorporated in a relatively simplistic way in the MFA-based continuum model. Instead of simply concluding that the KSA-based model is always preferable to the MFA-based model, we acknowledge that the increased accuracy of the KSA-based approach comes at the cost of significantly increased computational expense. Specifically, it takes a couple of days to produce the results shown in Figure 3.4(e), Figure 3.5(e), and Figure 3.6(e) on high performance computing facilities without parallelizing techniques (QUT High Performance Computing). Therefore, we take a flexible view and present both continuum models. Furthermore, we acknowledge that the MFA-based approach will be preferable in some circumstances, whereas the KSA-based approach will be preferable in other circumstances.

There are many ways that the work presented here can be extended. For example, all cases presented here involve particular choices of functional forms for $\delta(t)$ and $Z(r,t)$, yet many other choices are possible. Note that the stochastic algorithm described here, and the two continuum approximations are sufficiently flexible that other functional forms for $\delta(t)$ and $Z(r,t)$ can be used directly in these frameworks, if required. We note that, here we consider the change in cell diameter since this is the simplest possible way that we can mimic an increase in cell size. However, alternative approaches are possible, such as modelling dynamic changes in cell volume. Our modelling approach can be used to mimic dynamic changes in volume by assuming that cells are spherical, and expressing the radius as a function of volume. Here we do not pursue this approach as the experimental images in Figure 3.1 provide little information about the three-dimensional shape of the cells, so we feel it is more natural to work with a simpler measure, namely the approximate diameter, $\delta(t)$. Another assumption that we make is that all agents in the population behave identically in that each agent has the same initial size and grows at the same rate. An interesting extension of this work would be to consider a heterogeneous population of cells that is made up of distinct subpopulations. Using the framework presented it would be possible to consider different subpopulations with different initial sizes, and to consider different subpopulations that grow at different rates. This kind of model could be described using a more complicated multi-species framework (Matsiaka et al., 2017; Chapter 2). However, since this is the first time that a model of collective cell migration in a scratch assay that incorporates crowding effects and cell size dynamics has been explored, we leave this extension to the multi-species

case for future consideration.

Chapter 3A

Additional results for Chapter 3

3A.1 Comparing results from the two-dimensional and one-dimensional stochastic models

In this section we present results from solving the Langevin equations (Equation (3.3)) for the scratch assay geometry in two dimensions, and compare them with the corresponding solutions to the Langevin equations in one dimension. Employing the vector notation introduced in Chapter 3 we can write the Langevin equations for the two-dimensional model as follows,

$$\frac{\mathrm{d}\mathbf{u}^{(i)}}{\mathrm{d}t} = \sum_{j \neq i} \mathbf{F}_{ij} + \boldsymbol{\xi}_i, \qquad i = 1, \ldots, N, \tag{3A.1}$$

where $\mathbf{u}^{(i)}$ is the position of the ith agent, $\mathbf{F}_{ij}$ is the interaction force between agent i and agent j, $\boldsymbol{\xi}_i$ is the stochastic force acting on the ith agent, and N is the number of agents in the simulation.

The initial positions of agents in both the two-dimensional and one-dimensional simulations are chosen from the following distribution

$$\alpha(x) = \begin{cases} 12.5 \times 10^{-3}, & 0\ \mu\mathrm{m} \leq x < 600\ \mu\mathrm{m}, \\ 0, & 600\ \mu\mathrm{m} \leq x \leq 1400\ \mu\mathrm{m}, \\ 12.5 \times 10^{-3}, & 1400\ \mu\mathrm{m} < x \leq 2000\ \mu\mathrm{m}. \end{cases} \tag{3A.2}$$

In the case of the two-dimensional simulation, the vertical positions of agents are placed at random to form a homogeneous density distribution. It ensures that there is no density gradient in the vertical direction.

We fix all model parameters to be the same in both the two-dimensional and one-dimensional simulations to avoid any other possible source of variability. The size of agents is fixed to a typical size of skin cell size of $25\,\mu\mathrm{m}$ (Simpson et al., 2013b).

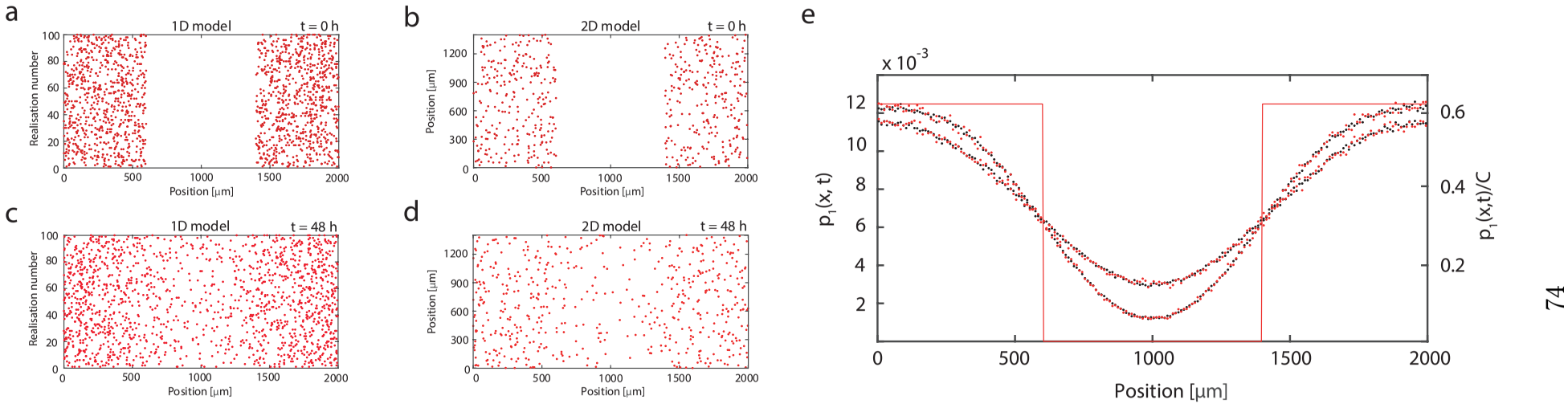


Figure 3A.1: Comparison of the averaged agent density data from the two-dimensional stochastic simulations and its one-dimensional counterpart. Snapshots in (a) and (c) show 100 identically prepared realisations of the stochastic model, given by Equation (3.3), at $t = 0$ and 48 h. Snapshots in (b) and (d) represent one realisation of the stochastic model, given by Equation (3.3), at $t = 0$ and 48 h. Results in (e) show the averaged agent density profiles using an ensemble of 10^5 identically prepared simulations of the one-dimensional model (black dots) with binsize 10 μm. These results compared with the averaged ensemble of 10^3 identically prepared realisations of the two-dimensional model, Equation (3.3), at $t = 24$ and 48 h. Equations (3.3) and Equation (3A.1) are solved with $\Delta t = 2 \times 10^{-2}$ h. The red solid line shows initial conditions given by Equation (3A.2).

We choose the number of agents in each one-dimensional simulation to be 15. The number of agents in each two-dimensional simulation is 534. The number of agents in both simulations is chosen so that the initial nondimensional density outside of the scratched region is $p_1/C = 0.625$, where C is the carrying capacity density of agents with diameter $25\,\mu$m. The size of the domain in the two-dimensional simulations is $2000\,\mu\text{m} \times 1400\,\mu\text{m}$, which is typical for scratch assay experiments (Figure 3.1). Periodic boundary conditions are imposed in both the two-dimensional and one-dimensional simulations.

Results are summarised in Figure 3A.1. To compare the one-dimensional and two-dimensional simulations we average results from the two-dimensional model in the vertical direction to obtain a one-dimensional agent density distribution. We show that averaged two-dimensional results demonstrate very similar population level behaviour compared to simple one-dimensional model provided all model parameters and initial densities are fixed.

3A.2 Derivation of the $p_1(x,t)$ and $p_2(x,y,t)$ governing equations for the one-dimensional model

In this section we derive the equations for the evolution of the density $p_1(x,t)$ and the pair-correlation density $p_2(x,y,t)$ presented in Chapter 3. These equations represent the first two levels of the full hierarchy of equations that incorporates all spatial moments (Middleton et al., 2014; Matsiaka et al., 2017; Chapter 2). To begin, we introduce an agent density function, $\rho(x,t)$. The evolution of the agent density is given by continuity equation (van Kampen, 1976):

$$\frac{\partial \rho(x,t)}{\partial t} = -\sum_{i=1}^{N} \frac{\partial J_i}{\partial x^{(i)}}, \tag{3A.3}$$

where J_i is the component of the total flux of agents associated with the agent i, and N is the total number of agents. If $J_i = \rho\, \mathrm{d}x^{(i)}/\mathrm{d}t$, we have

$$\frac{\partial \rho(x,t)}{\partial t} = -\sum_{i=1}^{N} \frac{\partial}{\partial x^{(i)}} \Big(\sum_{j \neq i} F_{ij} + \xi_i \Big) \rho(x,t), \tag{3A.4}$$

where $\sum_{j\neq i} F_{ij} + \xi_i$ is the right hand side of Equation (3.3) and ξ_i is treated as a fixed parameter (van Kampen, 1976).

Suppose we treat ξ_i as a random variable. We can obtain a solution to Equation

(3A.4) with the initial conditions

$$\rho(x,0) = \prod_{i=1}^{N} \delta(x - x^{(i)}(0)) = \delta^{(N)}(x - x^{(i)}(0)), \tag{3A.5}$$

where δ is the Dirac delta function, and $x^{(i)}(0)$ is the initial position of the ith agent. For any initial density distribution, $\rho(x,0)$, we assume that we can obtain the average over many different realisations of the stochastic force ξ_i, $\langle \rho(x,t) \rangle_\xi$.

Let $P_1^i(x,t)$ be the probability density function (PDF) for an individual agent. Then the one-agent PDF is given by (van Kampen, 1976)

$$P_1^i(x,t) = \langle \langle \rho(x,t) \rangle_\xi \rangle_{\mathrm{IC}}, \tag{3A.6}$$

where the average is taken over different realisations of the initial distribution and the random variable ξ_i. Since the averaged local density can be expressed as

$$\langle \langle \rho(x,t) \rangle_\xi \rangle_{\mathrm{IC}} = \langle \langle \delta^{(N)}(x - x^{(i)}(t)) \rangle_\xi \rangle_{\mathrm{IC}}, \tag{3A.7}$$

we can define the one-agent PDF in the following way,

$$P_1^i(x,t) = \langle \langle \delta^{(N)}(x - x^{(i)}(t)) \rangle_\xi \rangle_{\mathrm{IC}}, \tag{3A.8}$$

where $x^{(i)}(t)$ is the position of the ith agent at time t, as given by Equation (3.3).

Similarly, we define the two-agent PDF as

$$P_2^{ij}(x,y,t) = \langle \langle \delta^{(N)}(x - x^{(i)}(t))\, \delta^{(N)}(y - y^{(j)}(t)) \rangle_\xi \rangle_{\mathrm{IC}}. \tag{3A.9}$$

The evolution of $P_1^i(x,t)$ and $P_2^{ij}(x,y,t)$ are given by (García-Ojalvo and Sancho, 1999)

$$\frac{\partial P_1^i(x,t)}{\partial t} = D \Delta P_1^i(x,t) - \frac{\partial}{\partial x}\left(f_i\, P_1^i(x,t) \right), \tag{3A.10}$$

$$\frac{\partial P_2^{ij}(x,y,t)}{\partial t} = D \Delta P_2^{ij}(x,y,t) - \frac{\partial}{\partial x}\left(f_i\, P_2^{ij}(x,y,t) \right) - \frac{\partial}{\partial y}\left(f_j\, P_2^{ij}(x,y,t) \right), \tag{3A.11}$$

where the total force f_i acting on agent i can be expressed as

$$f_i = \sum_{j \neq i} F_{ij}. \tag{3A.12}$$

Combining Equations (3A.8), (3A.10) and (3A.12), and taking the convolution of the

interaction force and a δ function centred at $y^{(j)}$, we obtain

$$\frac{\partial P_1^i(x,t)}{\partial t} = D\Delta P_1^i(x,t) - \nabla\Big\langle \sum_{j\in\mathcal{L},j\neq i} \int_\Omega F(x^{(i)}-y,t)\,\delta(x-x^{(i)}(t))\,\delta(y-y^{(j)}(t))\,\mathrm{d}y \Big\rangle, \tag{3A.13}$$

where Ω denotes the domain and $\mathcal{L}$ is the set of agents. The second term on the right hand side of Equation (3A.13) is an advection term. Combining Equations (3A.9) and (3A.13), and interchanging summation and integration, we obtain

$$\frac{\partial P_1^i(x,t)}{\partial t} = D\Delta P_1^i(x,t) - \nabla \int_\Omega F(x-y,t) \sum_{j\in\mathcal{L},j\neq i} P_2^{ij}(x,y,t)\,\mathrm{d}y, \tag{3A.14}$$

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 population level dynamics, we define the following quantities,

$$p_1(x,t) = \frac{1}{N}\sum_{i\in\mathcal{L}} P_1^i(x,t), \tag{3A.15}$$

$$p_2(x,y,t) = \frac{1}{N(N-1)}\sum_{i\in\mathcal{L}}\sum_{j\in\mathcal{L},j\neq i} P_2^{ij}(x,y,t), \tag{3A.16}$$

where $p_1(x,t)$ is the normalised one-agent density distribution, and $p_2(x,y,t)$ is the density-density correlation function that captures correlations in agent positions.

To proceed, we sum over the index i in Equation (3A.14) and apply the definitions given in Equations (3A.15)-(3A.16) to obtain

$$\frac{\partial p_1(x,t)}{\partial t} = D\Delta p_1(x,t) - (N-1)\nabla\Big(\int_\Omega F(x-y,t)\,p_2(x,y,t)\,\mathrm{d}y\Big). \tag{3A.17}$$

To derive an evolution equation for $p_2(x,y,t)$ we begin with the two-agent Fokker-Planck equation,

$$\frac{\partial P_2^{ij}(x,y,t)}{\partial t} = D\Delta P_2^{ij}(x,y,t) - \frac{\partial}{\partial x}\left(f_i\,P_2^{ij}(x,y,t)\right) - \frac{\partial}{\partial y}\left(f_j\,P_2^{ij}(x,y,t)\right), \tag{3A.18}$$

where indices i and j denote arbitrary agents in population.

Adopting the interaction force law, Equation (3.10), using the definition of the two-agent PDF, as given by Equation (3A.9), and evaluating the required convolutions,

allows us to rewrite Equation (3A.18) as

$$\begin{aligned}\frac{\partial P_2^{ij}(x,y,t)}{\partial t} &= D\Delta P_2^{ij}(x,y,t)\\ &- \frac{\partial}{\partial x}\left\langle F(x-y,t)\,\delta(x-x^{(i)}(t))\,\delta(y-y^{(j)}(t))\right\rangle\\ &- \frac{\partial}{\partial y}\left\langle F(y-x,t)\,\delta(x-x^{(i)}(t))\,\delta(y-y^{(j)}(t))\right\rangle\\ &- \frac{\partial}{\partial x}\Big\langle \sum_{g\in\mathcal{L},g\neq i,j}\int_\Omega F(x-z,t)\,\delta(x-x^{(i)}(t))\,\delta(y-y^{(j)}(t))\,\delta(z-z^{(g)}(t))\,\mathrm{d}z\Big\rangle\\ &- \frac{\partial}{\partial y}\Big\langle \sum_{g\in\mathcal{L},g\neq i,j}\int_\Omega F(y-z,t)\,\delta(x-x^{(i)}(t))\,\delta(y-y^{(j)}(t))\,\delta(z-z^{(g)}(t))\,\mathrm{d}z\Big\rangle, \end{aligned} \tag{3A.19}$$

where the second and third terms on the right hand side of Equation (3A.19) represent interactions between agents i and j, the fourth and fifth terms on the right hand side of Equation (3A.19) represent interactions between agents i and j and other agents within the population.

The three-agent normalised density functions can be defined as

$$p_3(x,y,z,t) = \frac{1}{N(N-1)(N-2)}\sum_i\sum_{j\neq i}\sum_{g\neq i,j} P_3^{ijg}(x,y,z,t). \tag{3A.20}$$

We therefore require a definition for the three-agent PDF, $P_3^{ijg}(x,y,z,t)$. Similar to Equation (3A.9),

$$P_3^{ijg}(x,y,z,t) = \langle \delta(x-x^{(i)}(t))\,\delta(y-y^{(j)}(t))\,\delta(z-z^{(g)}(t))\rangle. \tag{3A.21}$$

To proceed we divide Equation (3A.19) by $N(N-1)$, and combine Equations (3A.19)-(3A.21), summing over the indices i and j, to obtain an expression for the evolution of $p_2(x,y,t)$

$$\begin{aligned}\frac{\partial p_2(x,y,t)}{\partial t} &= D\Delta p_2(x,y,t)\\ &- \frac{\partial}{\partial x}\Big(F(x-y)\,p_2(x,y,t)\Big)\\ &- \frac{\partial}{\partial y}\Big(F(y-x)\,p_2(x,y,t)\Big)\\ &- (N-2)\,\frac{\partial}{\partial x}\int_\Omega F(x-z,t)\,p_3(x,y,z,t)\,\mathrm{d}z\\ &- (N-2)\,\frac{\partial}{\partial y}\int_\Omega F(y-z,t)\,p_3(x,y,z,t)\,\mathrm{d}z. \end{aligned} \tag{3A.22}$$

This procedure can be repeated to yield the hierarchy of $N-1$ coupled integro partial differential equations and one Fokker-Planck equation. At any arbitrary level $d \in$

$[1, N-1]$ of this hierarchy, the d-density, p_d, depends on the higher order density, p_{d+1}. This means that full hierarchy of equations is intractable for analysis. Consequently, we invoke two approximations to simplify the hierarchy of density equations: (i)the standard mean-field approximation (MFA); and (ii) the Kirkwood superposition approximation (KSA).

The MFA approximates the pair correlation function $p_2(x, y, t)$ in terms of $p_1(x, t)$ and $p_1(y, t)$:

$$p_2(x, y, t) = p_1(x, t)p_1(y, t). \tag{3A.23}$$

This expression implies that the probability of finding an agent at x is statistically independent of the probability of finding an agent at y. The KSA approximates the three-agent normalised density function as the product of two-agent density functions, and can be written as (Singer, 2004; Middleton et al., 2014)

$$p_3(x, y, z, t) = \frac{p_2(x, y, t)\, p_2(x, z, t)\, p_2(y, z, t)}{p_1(x, t)\, p_1(y, t)\, p_1(z, t)}. \tag{3A.24}$$

3A.3 Discretisation scheme for the one-dimensional MFA model

In this section we present the discretisation scheme used to solve Equation (3.13) governing the evolution of the agent density $p_1(x, t)$ in one dimension. The MFA-based continuum equation is qiven by

$$\frac{\partial p_1(x, t)}{\partial t} = D\Delta p_1(x, t) - (N-1)\nabla(p_1(x, t)\, V(x, t)), \tag{3A.25}$$

where

$$V(x, t) = \int_\Omega F(x - y, t)\, p_1(y, t)\, \mathrm{d}y, \tag{3A.26}$$

and Ω is the integration domain.

We introduce the following quantities

$$\beta(x, y, t) = F(x - y, t)\, p_1(y, t) = f_0\, Z(r, t)\, \mathrm{sgn}(x - y)\, p_1(y, t), \tag{3A.27}$$

$$\begin{aligned} I_l &= p_1(x_l, t) \int_\Omega \beta(x_l, y, t)\, \mathrm{d}y \\ &= p_1(x_l, t)\frac{h}{2} \sum_s \Big[\beta(x_l, y_{s+1}, t) + \beta(x_l, y_s, t)\Big] + O(h^2), \end{aligned} \tag{3A.28}$$

where the trapezoidal rule with a stepsize h is used for numerical integration, and the indices l and s denote equally-spaced spatial mesh nodes. Here, the stepsize is chosen to be $h = 4\,\mu$m, while the binsize used in discrete results presented in Chapter 3 is $10\,\mu$m.

Using the definitions presented in Equations (3A.27)-(3A.28), we now apply the method of lines to Equation (3A.25) and obtain the system of ordinary differential equations

$$\frac{\mathrm{d}p_i^1}{\mathrm{d}t} = \frac{D}{h^2}\Big[p_{i+1} - 2p_i + p_{i-1}\Big] - (N-1)\frac{1}{2h}\Big[I_{i+1} - I_{i-1}\Big], \tag{3A.29}$$

where the index i denotes a spatial mesh node. This system of ordinary differential equations is solved using the first order explicit Euler method with a constant time step Δt. This expression is valid for an arbitrary interior node and, since we apply periodic boundary conditions, it can be easily adapted on the boundaries of the domain.

3A.4 Discretisation scheme for the one-dimensional KSA model

We now write down the discretisation scheme used to solve Equation (14) governing the evolution of $p_2(x,y,t)$ in the KSA-based framework. Note that we only solve the equation for $p_2(x,y,t)$ and obtain $p_1(x,t)$ by numerical integration, using

$$p_1(x,t) = \int_\Omega p_2(x,y,t)\mathrm{d}y. \tag{3A.30}$$

The governing equation that we consider is as follows,

$$\begin{aligned}
\frac{\partial p_2(x,y,t)}{\partial t} &= D\Delta p_2(x,y,t) \\
&- f_0\Big(\frac{\partial}{\partial x} - \frac{\partial}{\partial y}\Big)\Big(Z(|x-y|,t)\,\mathrm{sgn}(x-y)\,p_2(x,y,t)\Big) \\
&- f_0(N-2)\frac{\partial}{\partial x}\int_\Omega Z(|x-z|,t)\,\mathrm{sgn}(x-z)\,\frac{p_2(x,y,t)\,p_2(x,z,t)\,p_2(y,z,t)}{p_1(x,t)\,p_1(y,t)\,p_1(z,t)}\,\mathrm{d}z \\
&- f_0(N-2)\frac{\partial}{\partial y}\int_\Omega Z(|y-z|,t)\,\mathrm{sgn}(y-z)\,\frac{p_2(x,y,t)\,p_2(x,z,t)\,p_2(y,z,t)}{p_1(x,t)\,p_1(y,t)\,p_1(z,t)}\,\mathrm{d}z.
\end{aligned} \tag{3A.31}$$

In order to present the numerical method as briefly as possible, we define the following quantities

$$\gamma(x,y,t) = f_0\,Z(|x-y|,t)\,\mathrm{sgn}(x-y)\,p_2(x,y,t), \tag{3A.32}$$

$$\phi(x,y,z,t) = f_0\,Z(|x-z|,t)\,\mathrm{sgn}(x-z)\,\frac{p_2(x,z,t)\,p_2(y,z,t)}{p_1(z,t)}, \tag{3A.33}$$

$$\psi(x,y,z,t) = f_0\,Z(|y-z|,t)\,\mathrm{sgn}(y-z)\,\frac{p_2(x,z,t)\,p_2(y,z,t)}{p_1(z,t)}. \tag{3A.34}$$

Upon substituting Equations (3A.32)-(3A.34) into Equation (3A.31), the evolution equa-

tion for $p_2(x, y, t)$ becomes

$$\begin{aligned}\frac{\partial p_2(x,y,t)}{\partial t} &= D\Delta p_2(x,y,t) - \frac{\partial}{\partial x}\gamma(x,y,t) + \frac{\partial}{\partial y}\gamma(x,y,t) \\ &\quad - (N-2)\frac{\partial}{\partial x}\left[\frac{p_2(x,y,t)}{p_1(x,t)\,p_1(y,t)}\int_\Omega \phi(x,y,z,t)\,\mathrm{d}z\right] \\ &\quad - (N-2)\frac{\partial}{\partial y}\left[\frac{p_2(x,y,t)}{p_1(x,t)\,p_1(y,t)}\int_\Omega \psi(x,y,z,t)\,\mathrm{d}z\right]. \end{aligned} \tag{3A.35}$$

We now introduce the discretised quantities

$$\begin{aligned} Q_{l,k} &= \frac{p_2(x_l,y_k,t)}{p_1(x_l,t)\,p_1(y_k,t)}\int_\Omega \phi(x_l,y_k,z,t)\,\mathrm{d}z \\ &= \frac{p_2(x_l,y_k,t)}{p_1(x_l,t)\,p_1(y_k,t)}\frac{h}{2}\sum_s \Big[\phi(x_l,y_k,z_{s+1},t) + \phi(x_l,y_k,z_s,t)\Big] + O(h^2), \end{aligned} \tag{3A.36}$$

$$\begin{aligned} Y_{l,k} &= \frac{p_2(x_l,x_k,t)}{p_1(x_l,t)\,p_1(y_k,t)}\int_\Omega \psi(x_l,y_k,z,t)\,\mathrm{d}z \\ &= \frac{p_2(x_l,y_k,t)}{p_1(x_l,t)\,p_1(y_k,t)}\frac{h}{2}\sum_s \Big[\psi(x_l,y_k,z_{s+1},t) + \psi(x_l,y_k,z_s,t)\Big] + O(h^2), \end{aligned} \tag{3A.37}$$

where the trapezoidal rule with stepsize h on an equally spaced mesh is used to approximate the integrals. We now apply the method of lines to Equation (3A.35) and obtain the following system of equations

$$\begin{aligned}\frac{\mathrm{d}p_{ij}^2}{\mathrm{d}t} &= \frac{D}{h^2}\Big[p_{i+1,j} - 2p_{ij} + p_{i-1,j} + p_{i,j+1} - 2p_{ij} + p_{i,j-1}\Big] \\ &\quad - \frac{1}{2h}\Big[\gamma_{i+1,j} - \gamma_{i-1,j}\Big] + \frac{1}{2h}\Big[\gamma_{i,j+1} - \gamma_{i,j-1}\Big] \\ &\quad - (N-2)\frac{1}{2h}\Big[Q_{i+1,j} - Q_{i-1,j}\Big] - (N-2)\frac{1}{2h}\Big[Y_{i,j+1} - Y_{i,j-1}\Big], \end{aligned} \tag{3A.38}$$

where indices i, j denote spatial mesh nodes, and $\gamma_{l,m} = \gamma(x_l, y_m, t)$. This expression is valid for an arbitrary interior node and, since we apply periodic boundary conditions, it can be easily adapted on the boundaries of the domain. This systems is then solved using the first order explicit Euler method with a constant time step of duration Δt.

Chapter 4

Mechanistic and experimental models of cell migration reveal the importance of cell-to-cell pushing in cell invasion

A paper published in the *Biomedical Physics & Engineering Express*.

Matsiaka Oleksii, Baker Ruth, Shah Esha, Simpson Matthew. Mechanistic and experimental models of cell migration reveal the importance of intercellular interactions in cell invasion. *Biomedical Physics & Engineering Express*, vol. 5, 045009, 2019.

4.1 Abstract

Moving fronts of cells are essential for development, repair and disease progression. Therefore, understanding and quantifying the details of the mechanisms that drive the movement of cell fronts is of wide interest. Quantitatively identifying the role of intercellular interactions, and in particular the role of cell pushing, remains an open question. In this chapter, we report a combined experimental-modelling approach showing that intercellular interactions contribute significantly to the spatial spreading of a population of cells. We use a novel experimental data set with PC-3 prostate cancer cells that have been pretreated with Mitomycin-C to suppress proliferation. This allows us to experimentally separate the effects of cell migration from cell proliferation, thereby enabling us to focus on the migration process in detail as the population of cells recolonizes an initially-vacant region in a series of two-dimensional experiments. We quantitatively model the experiments using a stochastic modelling framework, based on Langevin dy-

namics, which explicitly incorporates random motility and various intercellular forces including: (i) long range attraction (adhesion); and (ii) finite size effects that drive short range repulsion (pushing). Quantitatively comparing the ability of this model to describe the experimentally observed population-level behaviour provides us with quantitative insight into the roles of random motility and intercellular interactions. To quantify the mechanisms at play, we calibrate the stochastic model to match experimental cell density profiles to obtain estimates of cell diffusivity, D, and the amplitude of intercellular forces, f_0. Our analysis shows that taking a standard modelling approach which ignores intercellular forces provides a poor match to the experimental data whereas incorporating intercellular forces, including short-range pushing and longer range attraction, leads to a faithful representation of the experimental observations. These results demonstrate a significant role of cell pushing during cell front movement and invasion.

4.2 Introduction

Moving cell fronts occur during many physiological processes, such as wound healing, morphogenesis, and malignant invasion (Tao et al., 2007; Friedl et al., 2004; Ramis-Conde et al., 2008; Kabla, 2012). Typically, cell fronts are observed as advancing, sharp boundaries between densely occupied and vacant regions, or as a moving interface between two distinct populations of cells (Hakim and Silberzan, 2017). An example of the first scenario is wound healing, where populations of cells close and recolonize an initially vacant space (Maini et al., 2004b; Jin et al., 2016a), as shown in Figure 4.1. An advancing interface between two populations of cells is often associated with malignant invasion into surrounding tissues (Eves et al., 2003; Lowengrub et al., 2010). Improving our understanding of how cell populations spread can provide important, clinically-relevant information about the nature of moving cell fronts. Historically, moving cell fronts have been studied, both *in vitro* and *in vivo*, to provide both qualitative and quantitative information about the mechanisms that drive front movement. We note that quantifying the precise contributions of various cellular-level mechanisms that lead to population-level front behavior is a nontrivial task that requires the integration of many different types of experimental data (Treloar et al., 2013a). Often it is assumed that the movement of advancing cell fronts is driven by the combined effects of undirected cell migration and carrying capacity-limited cell proliferation (Treloar et al., 2013a). At present, a fundamental question, which remains largely overlooked in the mathematical biology literature, is *what is the role of cell-to-cell pushing and how does it influence population-level front movement?*

Cell motility is a complicated process involving both the interplay and competition between various individual-level mechanisms (Treloar et al., 2013a; Treloar et al., 2014; Hakim et al., 2017). One of the most well-studied individual-level cell motility mech-

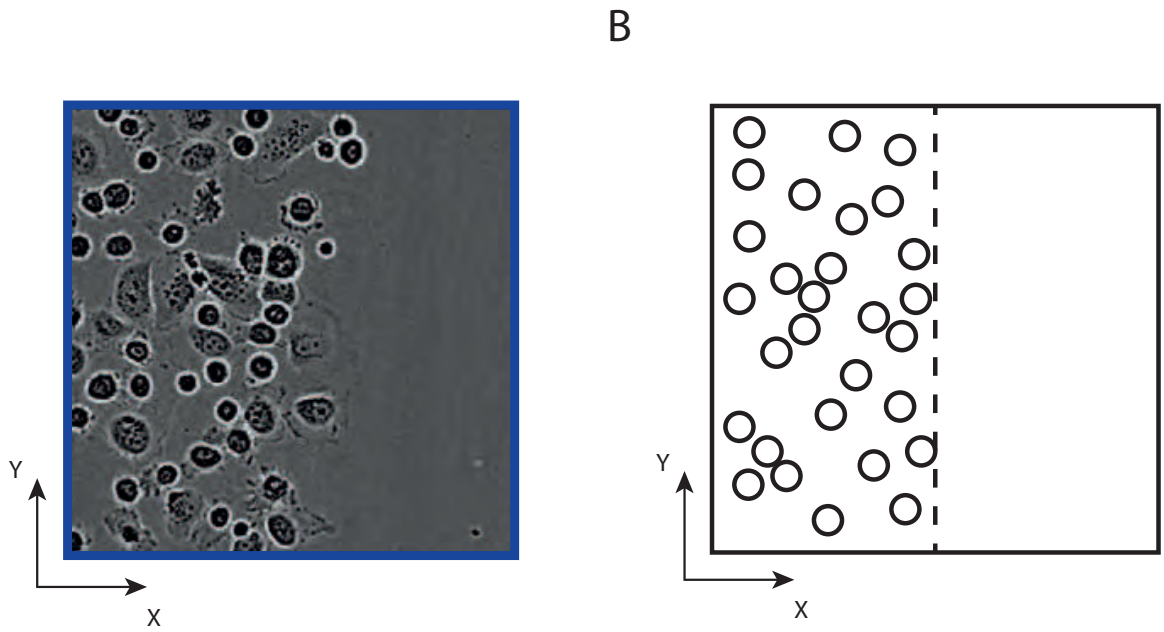


Figure 4.1: What drives the movement of cell fronts? A: Experimental image showing the leading edge of a moving front of PC-3 prostate cancer cells. This front is moving in the positive x direction. B: Schematic of A showing the position of the front (vertical dashed line) with the front moving in the positive x direction. The location of the region in A is the blue rectangle superimposed in Figure 4.1(A).

anisms is lamellipodial cell migration where cells undergo undirected movement due to myosin-powered contractions of the actin network under the cytoplasmic membrane (Nickaeen et al., 2017). Since this process is observed in many cell types, it remains prevalent in many mathematical modelling frameworks. As such, the assumption that cells undergo Brownian motion is often invoked and cells are represented, either implicitly or explicitly, as non-interacting point particles that move according to a white Gaussian process (Codling et al., 2008). While this approach is appealing due to its simplicity, it neglects the effects of intercellular interactions, such as adhesion and finite size (crowding) effects. The neglect of cell-to-cell adhesion can be problematic because it is known that mesenchymal cell types, such as keratinocytes, can be strongly affected by cell-to-cell adhesion during wound healing (Nardini et al., 2016). Furthermore, adherent cells can form clusters that exhibit qualitatively different behaviour from isolated cells (Steinberg, 1996; Nadell et al., 2010; Painter et al., 2010; Carmona-Fontaine et al., 2011). For example, strongly adherent groups of cells can undergo movement mediated by adherens junctions and move as a whole in coordinated process known as *locomotion* (George et al., 2017; Etienne-Manneville, 2014). During this motion cells can change shape because of the imbalance between traction and friction forces induced by leader and follower cells (Trepat, 2009). A few individual cells can heavily influence front expansion in a process known as *trailblazing* and form finger-like protrusions (McLennan, 2015; Köpf and Pismen, 2013). Tightly packed interacting monolayers of cells can exhibit properties similar to the solid state matter with a transition to the fluid-like

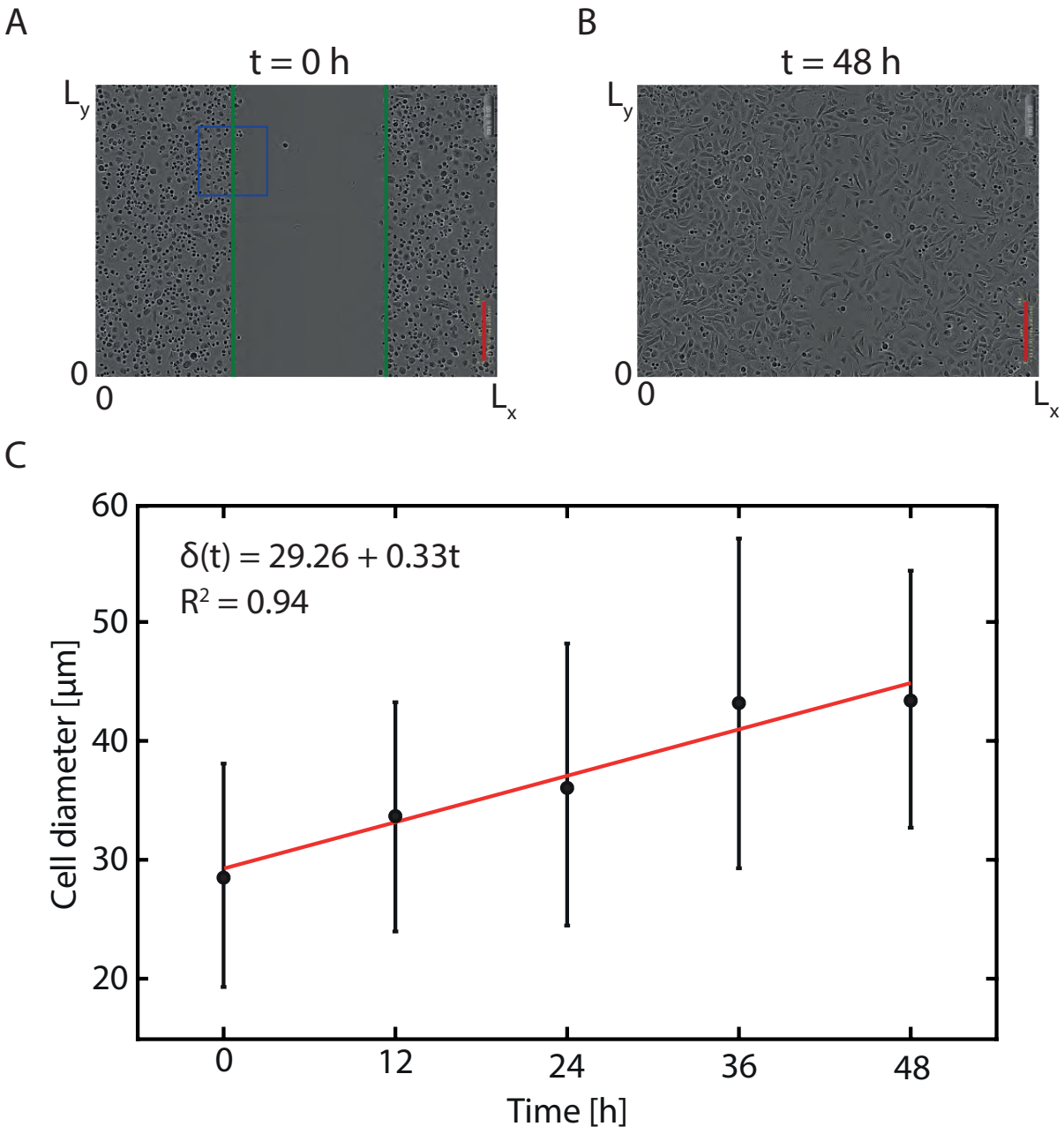


Figure 4.2: Experimental data. A-B: Images from an IncuCyte ZOOM™ assay with Mitomycin-C pretreated PC-3 prostate cancer cells. The scale bar in each image corresponds to 300 μm. The green solid lines show the initial position of the two opposingly directed cell fronts. The blue rectangle denotes the location of the subregion highlighted in Figure 4.1(A). C: Cell diameter data as a function of time, $\delta(t)$, from a sample of 30 randomly chosen cells at each time point. Black dots indicate the sample mean and the error bars denote the sample standard deviation about the sample mean. Red solid line represents the best-fit linear approximation, $\delta(t) = 29.26 + 0.33t$. R^2 is the adjusted coefficient of determination measuring the goodness of fit.

behaviour (Park et al., 2015).

Arguably, some of the most striking examples of the front-like spreading of a cell population occur during embryonic development, such as neural crest cell invasion in the developing gut tissues, which is thought to arise as a consequence of combined undirected Brownian cell motility and carrying capacity limited cell proliferation (Simpson

et al., 2006b; Simpson et al., 2009b; Weijer, 2009). However, previous investigations have made the point that short range cell pushing can also play a role in driving the movement of cell fronts in confined environments, such as living tissues (Hawkins et al., 2009). Henceforth, we hypothesize that cells in a confined space may generate population pressure, driven by finite size effects and local repulsion, which can stimulate spatial expansion of the population.

Perhaps the most popular mathematical framework for modelling the movement of cell fronts involves using reaction-diffusion equations (Fisher, 1937; Sherratt and Murray, 1990; Johnston et al., 2015), including the Fisher–Kolmogorov equation, and generalisations thereof. Previously, the lack of individual-level experimental data meant that classical reaction-diffusion models were a useful way to conceptualize and simulate collective cell behaviour. However, with the increasing availability of individual-level information it is becoming important to develop mathematical models that provide both population-level information and individual-level information. There is a vast number of discrete models that focus on many individual-level features of cell behaviour (Buske et al., 2011; Gardiner et al., 2015; Osborne et al., 2017). The downside of many modelling frameworks presented in the literature is a large number of free parameters which makes parameter estimation problems computationally intractable and simulations hard to interpret. For example, in Buske et al., 2011 the authors develop a comprehensive mathematical model of stem cells dynamics in the intestinal crypt. This model requires the selection of 25 free parameters. Some of these parameters are not directly measurable, and other parameters in the model had not been estimated before this study. Consequently, quantifying the precise contribution of different factors that control cell colony behaviour and dynamics is very challenging. To deal with this question of over parameterization, we deal only with a fundamental model that is characterized by a minimal number of degrees of freedom that we can use experimental data to identify the parameters in the model and hence to quantitatively explore the roles of undirected migration and cell-to-cell pushing in our experiments.

Previously, cell pushing has been incorporated into lattice-based models of cell motility where agents move on a spatial domain that is represented as a regular lattice (Nan et al., 2018; Yates et al., 2015). However, these previous studies about the role of cell pushing are primarily theoretical studies that do not consider calibrating models to quantitatively match any experimental data. Our current work is the first attempt to incorporate short range cell pushing into a more realistic spatially continuous, off-lattice discrete model of cell migration. Importantly, we directly apply the model to quantitatively mimic a novel experimental data set. Figure 4.2 illustrates the IncuCyte ZOOM^{TM} scratch assay experimental protocol that we use in this chapter, at $t = 0$ h and $t = 48$ h. In our experiments we isolate the role of cell migration from the effects of cell proliferation by working with a population of cells that has been pre-treated with a

chemotherapy drug, Mitomycin-C, to block DNA replication and, consequently, suppress proliferation (Sadeghi et al., 1998). This means that the number of cells present in the experimental field of view over the duration of the experiment remains approximately constant. However, a side effect of Mitomycin-C pretreatment is that individual cells increase in size during the experiment as the cells prepare to proceed through the cell cycle but are unable to divide (Simpson et al., 2013). This dynamic increase in cell size, which is typically neglected in previous modelling studies (Simpson et al., 2013), can significantly influence intercellular interactions during the experiment and so we take great care to incorporate these effects into our mathematical model. Our approach for incorporating dynamic cell size effects in the model is justified by also working with a simpler model that neglects dynamic changes in cell size; we find that the simpler, standard model leads to a much poorer match with the experimental data.

This chapter is structured as follows. We begin by describing the IncuCyte ZOOM™ experimental protocol, the experimental data, and the procedure we use to process the experimental images. We then introduce the discrete mathematical model which accounts for random motility and intercellular interactions, including short range pushing and longer range attraction, as well as incorporating a mechanism for describing dynamic cell size changes. We refer to this model as *Model I* since it incorporates all four mechanisms that are thought to be relevant to the experimental system. To quantitatively explore the significance of these various cell-level mechanisms we systematically repeat the model calibration process for a range of simpler, more commonly used models. These simplified models account for: (i) random motility and intercellular forces (*Model II*); (ii) intercellular forces only (*Model III*); and (iii) random motility only (*Model IV*). We discuss the performance of each model when applied to the IncuCyte ZOOM™ data in the Results and Discussion. Finally, in the Conclusions we summarize our findings and discuss alternative applications and extensions of our modelling framework.

4.3 Materials and methods

4.3.1 IncuCyte ZOOM™ experimental data

Monolayer scratch assays are performed using the IncuCyte ZOOM™ system (Essen BioScience, Ann Arbor, MI) as shown in Figure 4.2(A)-(B). All experiments are performed using the PC-3 prostate cancer cell line (Kaighn et al., 1979) from the American Type Culture Collection (ATCC, Manassas, USA). The procedure of growing the cell culture in a flask is outlined by Jin et al., (2016b). After growing, cells are removed from the flask using TrypLE™ (Life Technologies) in phosphate buffered saline, resuspended in medium and seeded at a density of 20,000 cells per well in 96-well ImageLock plates (Essen BioScience, Ann Arbor, MI) as shown in Figure 4.3. The diameter of each

individual well is 9000 μm.

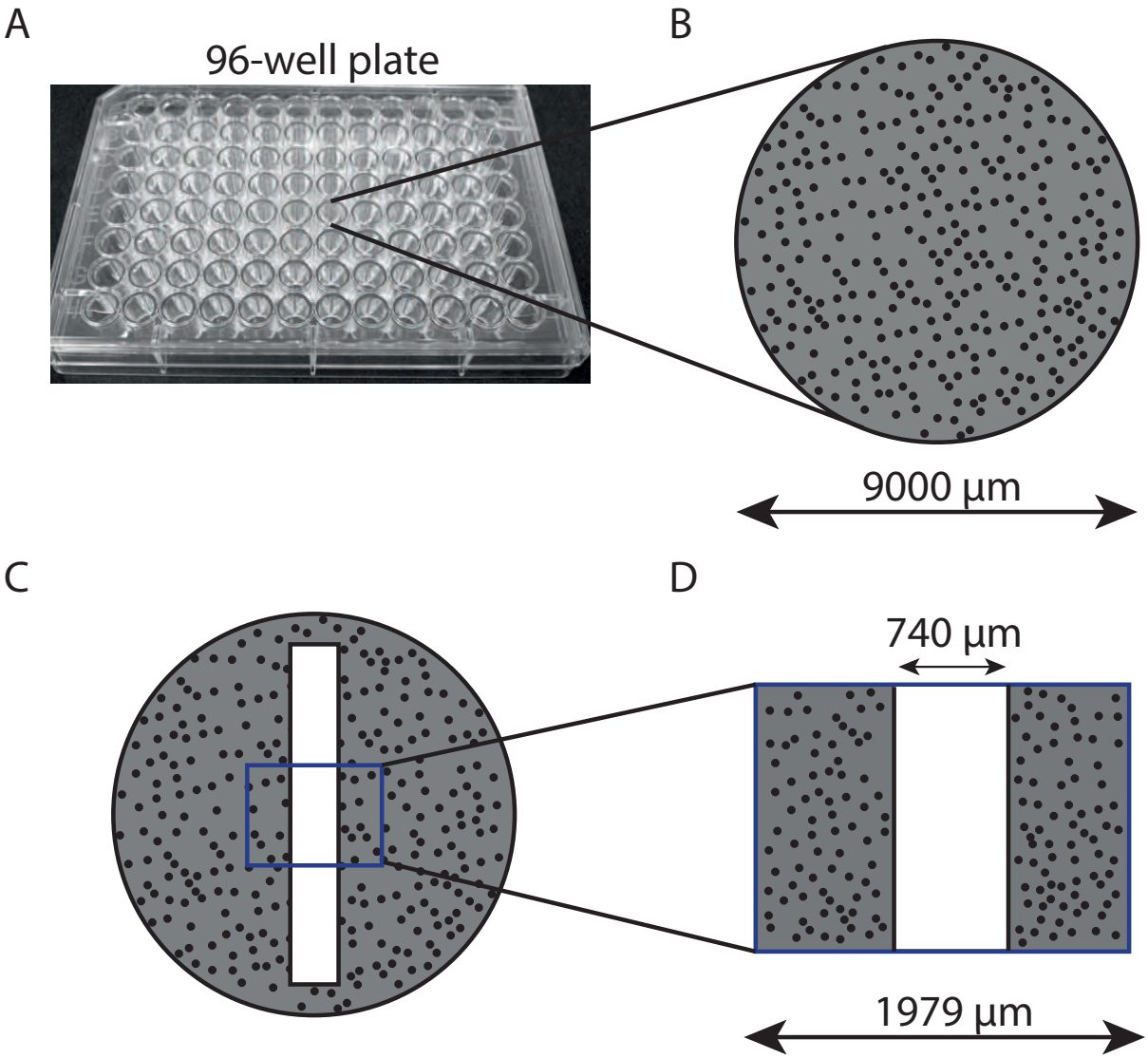


Figure 4.3: Experimental geometry. A: Image of a 96-well culture plate. The diameter of each well is 9000 μm. B: Schematic demonstrating the monolayer of cells (black dots) with approximately constant density. C: Schematic showing an artificial wound (white) in the monolayer of cells. D: Field of view of the experimental images showing that the field of view is much smaller than the extent of the well in the 96-well plate.

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; Kumari et al., 2009; Tlili et al., 2018). A WoundMaker™ (Essen BioScience, Ann Arbor, MI) 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 IncuCyte ZOOM™ and photographed every 2 hours for 48 hours. In total, these experiments are performed in eight of the 96 wells on the 96-well plate. After a preliminary visual inspection of the resulting eight experimental images, we selected four typical wells for analysis. Throughout this chapter we will refer to these four identically prepared experiments as Experiment A, B, C and D, respectively. All

experiments are initiated with a monolayer of cells close to confluence. Even though the cell culture was treated with Mitomycin-C, which ensures the absence of proliferation and approximately constant cell count during an experiment, the monolayer of cells remains close to confluence throughout the duration of the experiment due to increase in the size of individual cells (Figure 4A.1, Chapter 4A).

By the end of the experiment, summarised in Figure 4.2(A)-(B), we see that the Mitomycin-C pretreated cells have approximately doubled in size (Simpson et al., 2013). There is little information about the change in volume of cells since all data is in the form of 2D images. However, we note that, due to effects of Mitomycin-C pretreatment that allows cells to go through cell cycle to the point when they are ready to divide, the observed growth is rather volumetric and reflects actual change in cell size. To quantify this increase in cell size we randomly choose 30 cells from the experimental images at $t = 0, 12, 24, 36$ and 48 h and use these cells to estimate the average diameter as a function of time, $\delta(t)$. To do this we estimate the area on the image occupied by each particular cell, and then convert this estimate of area into an equivalent diameter, $\delta = \sqrt{4A/\pi}$, where A is the area estimate. With 30 estimates of the diameter at $t = 0, 12, 24, 36$ and 48 h, we compute the sample mean and sample standard deviation at each time point and plot the data in Figure 4.2(C). Visually we see that the average diameter appears to increase approximately linearly with time, and so we fit a linear model to the data. The cell diameter data and the linear model are shown in Figure 4.2(C). Note that had our experiments been performed over a longer period of time it would be more appropriate to use a different model, such as the logistic growth function, to model the temporal cell size dynamics.

4.3.2 Image analysis

An example of a raw experimental image and a detailed description of the procedure we use to extract density information from that image are summarized in Figure 4.4. Here, the size of the field of view is $(L_x \times L_y) = (1979\,\mu\text{m} \times 1439\,\mu\text{m})$, as shown in Figure 4.4(A). Throughout this chapter we use data from four identically-prepared experimental replicates of the scratch assay. Experimental images at the beginning of the experiment for each of the four replicates are shown in Figure 4.5. For completeness, the time series of experimental images for all four identically prepared experiments is given in Chapter 4A (Figure 4A.1), giving a total of 20 experimental images.

To process each experimental image we first separate the background of the image from the cells using Ilastik (Sommer et al., 2011). Ilastik is a machine learning tool that enables automatic object identification, and allows us to separate the cells in each image from the background. An example of a grey scale segmented image is shown in Figure 4.4(B). A visual comparison of the raw image in Figure 4.4(A) with the segmented

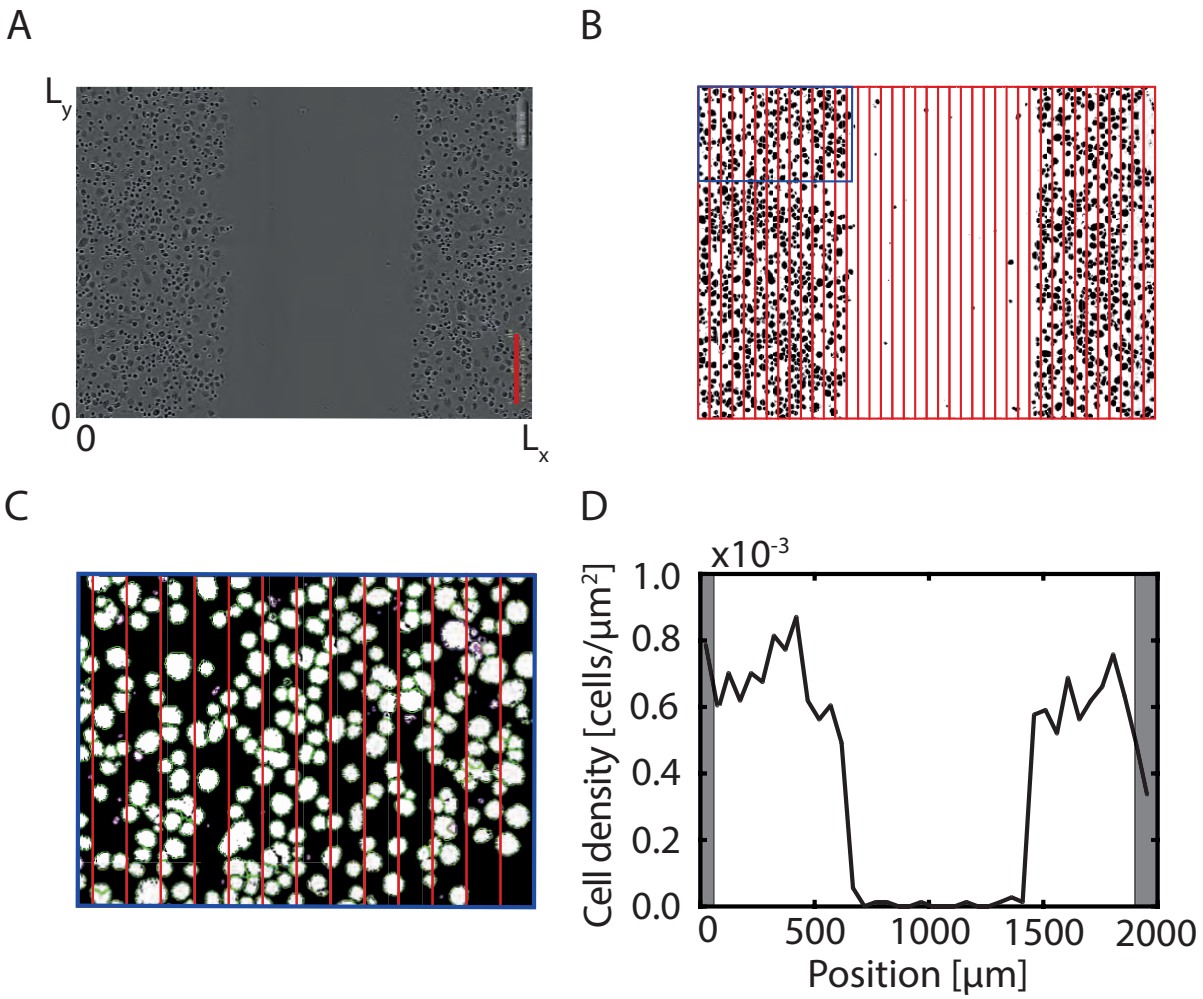


Figure 4.4: Processing of raw experimental images from the IncuCyte ZOOM™ scratch assay. A: Raw experimental image showing the field of view, of dimension $(L_x \times L_y) = (1979\,\mu\text{m} \times 1439\,\mu\text{m})$. The scale bar corresponds to 300 μm. B: Binary image obtained after segmenting the raw image with Ilastik. C: Zoomed-in image showing the region contained within the blue rectangle in B after processing with CellProfiler. Faint green outlines denote individual objects that are identified as cells. D: The one-dimensional cell density profile is estimated by counting the number of objects per equally-spaced column. Shaded regions in D show boundary regions that are neglected owing to the presence of scale bar and time label that are automatically superimposed on the IncuCyte ZOOM™ images.

image in Figure 4.4(B) confirms that the identification of cells from the background in the image is accurate. Since the density of cells in the original images is independent of the vertical coordinate, as shown in Figure 4.5, we divide each image into 40 equally-spaced columns, 49.5 μm wide each. We then use CellProfiler to automatically estimate the number of cells per column (Carpenter et al., 2016). With this data we divide the number of cells per column by the area of each column to give an estimate of the cell density across the horizontal coordinate of the experimental images.

A typical column-averaged cell density profile, shown in Figure 4.4(D), summarizes the spatial variations in density as a function of the horizontal coordinate. We employ this technique to extract density profiles at each time point, $t = 0, 12, 24, 36$, and 48 h, for each experimental replicate. The data presented in Figure 4.4(D) shows the

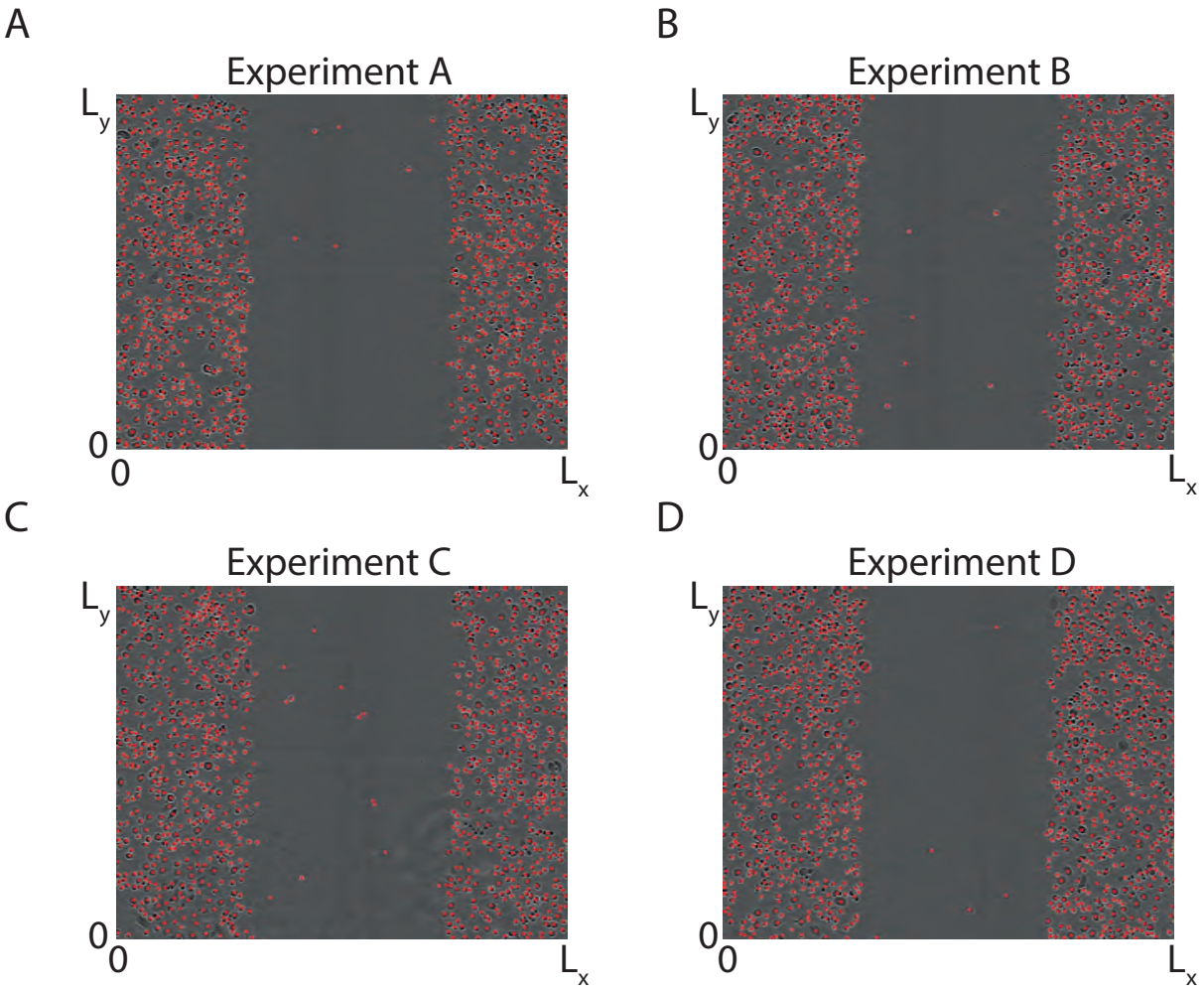


Figure 4.5: Experimental images of four identically prepared IncuCyte ZOOM^{TM} scratch assays just after the scratch has been made, $t = 0$ h. A-D: In each experimental replicate the positions of individual cells are extracted using CellProfiler and highlighted with red dots.

average cell density that we associate with the centre of each column. This particular density profile is relatively noisy because it is associated with the single image in Figure 4.4(A). Before we proceed, we discard density data from the two right-most columns of each image because the time label and scale bar are automatically superimposed on the IncuCyte ZOOM^{TM} images and these objects partially obscure the numbers of cells in these subregions of each images. We also discard the left-most column from each image, which leaves us with a slightly smaller image that is discretized into 37 equally-spaced columns of width 49.5 μm, and a reduced domain width of 1831 μm. Next, to reduce the magnitude of the fluctuations in the cell density profile, we average the density profiles associated with each of the four experimental replicates to obtain a single averaged cell density profile as a function of time, as summarized in Figure 4.6. For completeness, the relatively noisy column-averaged cell density profiles associated with each of the four individual experimental replicates are given in Chapter 4A (Figure 4A.3).

Now that we have quantified our experimental observations in terms of the temporal

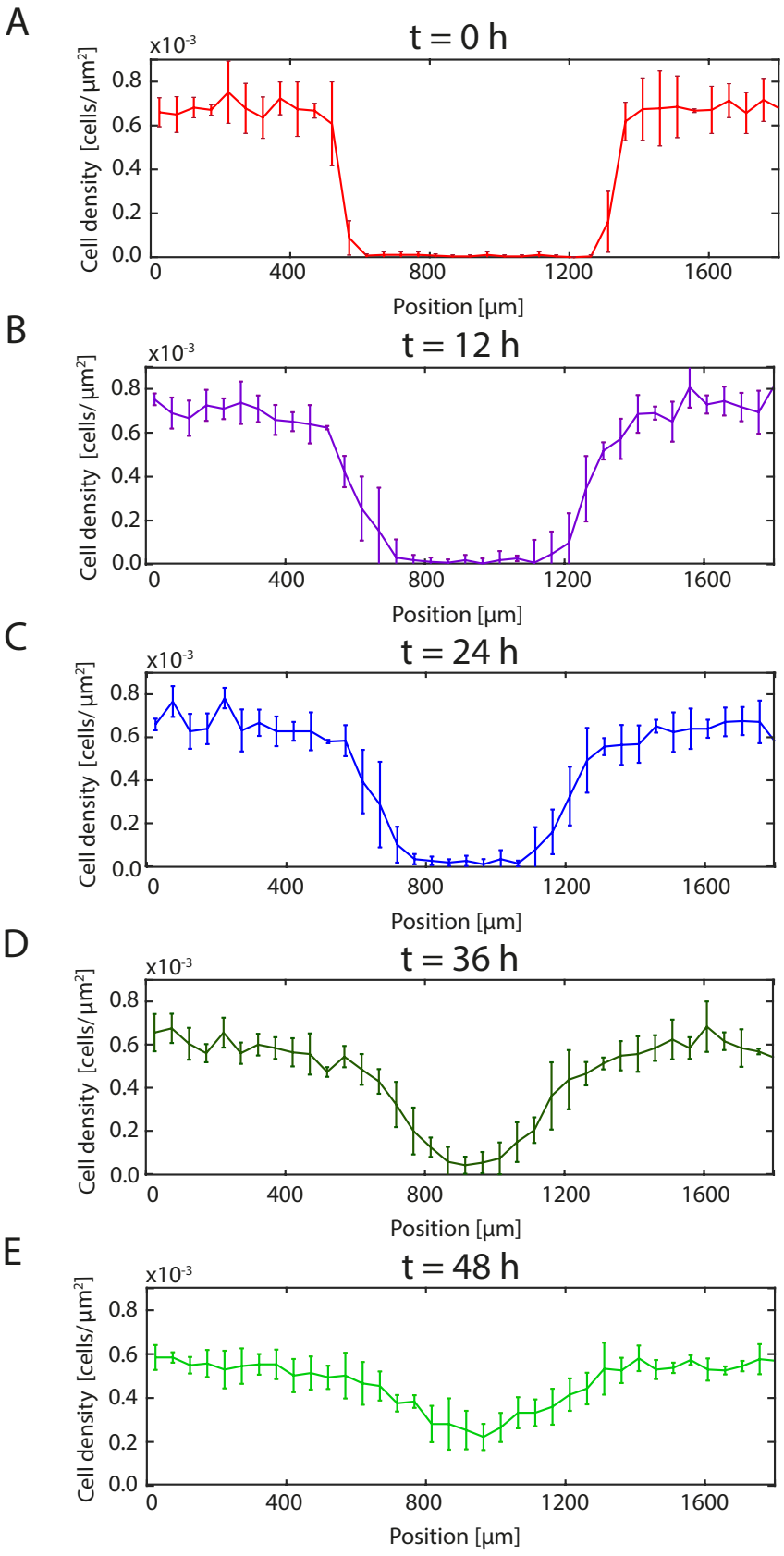


Figure 4.6: Averaged cell density profiles. A-E: Averaged cell density profiles at $t = 0, 12, 24, 36$, and 48 h, as indicated. All profiles report the sample mean density computed using four identically prepared experimental replicates. The error bars denote the sample standard deviation.

variation in the column-averaged cell density profiles, further averaged across four identically prepared experimental replicates, we will now attempt to use a suite of discrete mathematical models to mimic the experimental data set. To provide the most realistic discrete simulations, we always take care to initialize each discrete simulation using the exact same number and locations of cells that are present in the experimental images at $t = 0$ h (Figure 4.5).

4.3.3 Discrete stochastic model

To simulate our experimental data set we use a two-dimensional discrete model of cell motility that incorporates random motility, intercellular interactions including both long range cell-to-cell adhesion and short range cell pushing, as well as capturing dynamic changes in cell size. We refer to this model as *Model I* since it describes the situation where all four processes are acting simultaneously. We choose to work with a discrete modelling framework because discrete individual-based approaches are more natural to compare with experimental images than continuum models (Sepulveda et al., 2013; Chiou et al., 2012). Such discrete individual-based models are used to study a range of cell biology phenomena, including malignant invasion (Zhu et al., 2015), wound healing (De Jesus et al., 2016), self-organization (Osborne et al., 2017), angiogenesis (Peirce, 2008) and embryonic development (Weijer, 2007). Since we work with an off-lattice discrete framework, each agent is allowed to move in any direction on a continuous domain. This off-lattice approach is more realistic than a simpler lattice-based model where the locations of agents are restricted to an artificial lattice structure (Codling et al., 2008; Osborne et al., 2017; Ermentrout and Edelstein-Keshet, 1993; Wynn et al., 2013).

We begin by introducing Model I and then describe three simplifications of this model in which we systematically neglect certain features. To explore the relevance of the mechanisms inherent in these four models we carefully calibrate each model to match the density data summarized in Figure 4.6 and quantitatively compare the results of the calibration procedure.

Model I: random motility, long range cell-to-cell adhesion, short range cell pushing and dynamic changes in cell size

A key novelty of our approach is that we simulate the dynamic change in agent diameter, $\delta(t)$. This approach is very different to standard approaches where agents in discrete models are thought of as having either no size or a constant size. Here, to match our experimental measurements presented in Figure 4.2(C), we assume that $\delta(t)$ increases linearly,

$$\delta(t) = 29.26 + 0.33t, \tag{4.1}$$

where t is time, measured in hours.

We employ a Langevin stochastic framework (Middleton et al., 2014), where a population of N cells is modelled by a system of N first order stochastic differential equations

$$\frac{\mathrm{d}\mathbf{u}^{(i)}}{\mathrm{d}t} = \overbrace{\sum_{j\neq i} \mathbf{F}_{ij}(r,t)}^{\text{directed force}} + \overbrace{\boldsymbol{\xi}_i}^{\text{random noise}}, \qquad i = 1,\ldots,N, \tag{4.2}$$

where $\mathbf{u}^{(i)}$ is the position vector of the ith agent on a two-dimensional domain, $\mathbf{F}_{ij}$ is the deterministic interaction force between agents i and j that are separated by distance r (Mousavi et al., 2014), and $\boldsymbol{\xi}_i$ is a random stochastic force exerted on the ith agent. The stochastic force $\boldsymbol{\xi}_i$ is sampled from the Gaussian distribution (Middleton et al., 2014) with zero mean and variance $2D/\Delta t$, where D is the diffusivity, and Δt is the time step used to numerically solve Equation (4.2). Since the Langevin equation formalism does not include any inertial forces, this framework implicitly neglects agent acceleration. This assumption is reasonable at low Reynolds numbers, and is routinely invoked at cellular length scales (Berg, 1993). The model given by Equation (4.2) does not address possible directional bias near the edge of a wound resulting from the chemical signalling responses to wounding (Nikolic et al., 2006). Including directional bias and chemotaxis will result in additional free parameters in the system (Johnston et al., 2015), that we have little prior knowledge about, which, in turn, will make the parameter estimation problem significantly harder. Increasing number of parameters in the model requires increased complexity of the experimental data to avoid ambiguities in the parameter estimation (Warne et al., 2019). As such, we focus on describing a cell population using the fundamental model where the cell migration is mediated by a directional movement due to cell-to-cell adhesive and repulsive forces, $\mathbf{F}_{ij}$, and the undirected stochastic force, $\boldsymbol{\xi}_i$.

The details of the deterministic interaction force, $\mathbf{F}_{ij}$, can be chosen to incorporate a range of relevant phenomena such as long range attraction and short range repulsion (Matsiaka et al., 2018; Chapter 3), as illustrated in Figure 4.7. In this chapter we specify that the interaction force, $\mathbf{F}_{ij}$, depends upon the distance between agents, r, and time, t, and is given by

$$\mathbf{F}_{ij}(r,t) = f_0\, Z(r,t)\, \frac{\mathbf{u}^{(i)} - \mathbf{u}^{(j)}}{|\mathbf{u}^{(i)} - \mathbf{u}^{(j)}|}, \tag{4.3}$$

where f_0 is dimensional magnitude of the interaction force, $Z(r,t)$ is a dimensionless function describing how the interaction force depends on the separation between the agents, $r = |\mathbf{u}^{(i)} - \mathbf{u}^{(j)}|$. As with all models of the form of Equation (4.2), the dimensional magnitude of the interaction force, f_0, has dimension of velocity. Previously, the interaction forces between cells in the epithelial monolayer have been inferred using a discrete framework by Chiou et al., (2012), where the authors explicitly demonstrate

the significance of population pressure and mechanically driven population spreading (Ranft et al., 2010). The key novelty in the expression for the force function, Equation (4.3), is the time dependent cell-to-cell interaction potential, $Z(r,t)$, Figure (4.7). This approach allows us to explicitly parameterise the dynamic change in cell size and simulate mechanically driven cell pushing, which is in contrast to many models of cell motility and adhesion present in the literature where cells have constant size (Huang et al., 2005). The cell-to-cell pushing is expressed by a combination of cell-to-cell repulsive forces incorporated in the force function, Equation (4.3), dynamical change in cell size, and cell random motility. Here, we address basic phenomenological principles of cell-to-cell adhesion without considering biochemical pathways that facilitate adhesion, such as E-cadherin protein junctions (Cai et al., 2014; Peglion et al., 2014).

In this chapter we use the dimensionless function $Z(r,t)$ to incorporate three main features of cell-to-cell interactions: (i) short range repulsion forces which mimic cell pushing; (ii) longer range attraction forces which mimic cell-to-cell adhesion; and (iii) dynamic changes in agent size. The short range repulsion forces can be interpreted as cell resistance to deformation, which leads to crowding and volume exclusion effects (Bruna and Chapman, 2012). In contrast, the longer range attraction forces mimic intercellular attraction. These cell-to-cell attraction forces are thought to be a predominant factor in the cell-to-cell adhesion (De Palo et al., 2017). To incorporate these effects we use a modified Morse potential (Matsiaka et al., 2017; Chapter 2),

$$Z(r,t) = \begin{cases} 2\Big(\exp(-2a\,[r-\delta(t)]) - \exp(-a\,[r-\delta(t)])\Big), & r < 2\delta(t), \\ 2\Big(\exp(-2a\,[r-\delta(t)]) - \exp(-a\,[r-\delta(t)])\Big)\, g(r,t), & 2\delta(t) \leq r \leq 3\delta(t), \\ 0, & r > 3\delta(t), \end{cases} \tag{4.4}$$

where $\delta(t)$ is the time-dependent agent diameter, $a > 0$ is a parameter that controls the shape of the force function, and $g(r,t) = \Big(1 - \sin\Big[(2\pi r - \pi\delta(t))/2\delta(t)\Big]\Big)/2$ is the Tersoff cut-off function (Tersoff, 1988) which ensures a finite range of interactions. The cell-to-cell interaction range is finite, and set to three agent diameters (Middleton et al., 2014). As such, the interaction force is zero for separation distances of greater than three agent diameters. For all results presented here we set $a = 0.08$ which implies that the agents are relatively rigid and highly unlikely to overlap (Matsiaka et al., 2017; Chapter 2).

Schematics showing the key features of $Z(r,t)$ at $t = 0$ h and $t = 48$ h are shown in Figure 4.7(C-D). Here we see that at short separation distances we have strong positive $Z(r,t)$, which captures short range repulsion and pushing, owing to finite size effects. Over longer separation distances we have smaller negative $Z(r,t)$ which models attraction, such as adhesion. Finally, over sufficiently large distances we have no interactions as $Z(r,t) = 0$. To capture the effects of the increase in cell size, all length scales in Figure 4.7(C-D) are given in terms of the average cell diameter, $\delta(t)$, which can vary

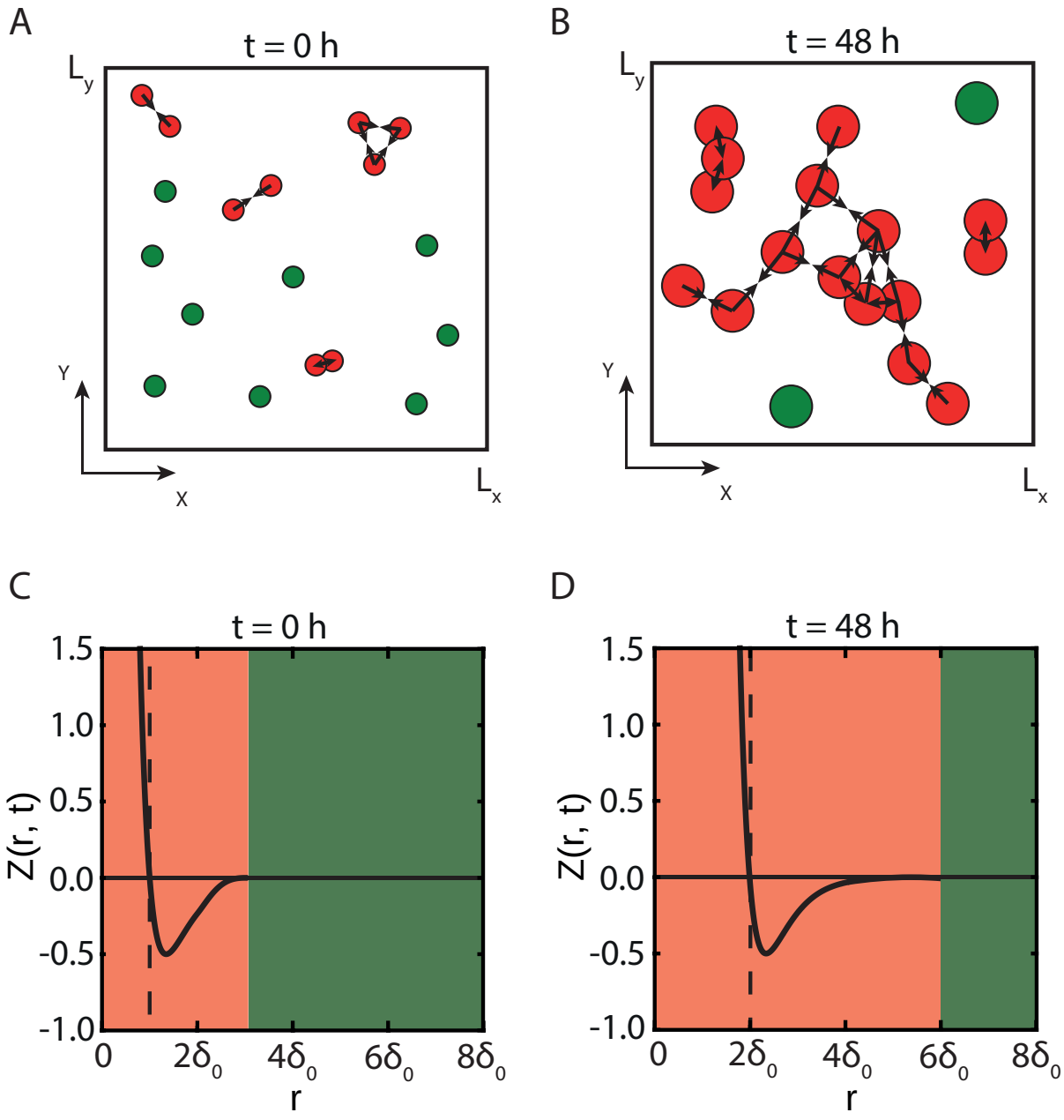


Figure 4.7: Schematic showing how short and long range forces are introduced in the model through the dimensionless force function, $Z(r,t)$. A-B: Green circles denote isolated agents that are unaffected by cell-to-cell interactions and, as a consequence, undergo random migration. Red circles indicate agents that are sufficiently close to other agents that they interact with them. Comparing the schematics in A and B shows that the increase in agent size leads to additional agent-to-agent interactions because agent growth reduces the distance between agents. The arrows in A and B indicate the direction of the deterministic forces $\mathbf{F}_{ij}$. C-D: Dimensionless force function $Z(r,t)$ for $t = 0$ h and $t = 48$ h. δ_0 is the agent size at the start of the experiment, $t = 0$ h. The red shaded regions indicate a sufficiently small distance between agents, $r \leq 3\delta(t)$, where agent-to-agent interactions are present. The green shaded regions indicate a sufficiently large distance between agents, $r > 3\delta(t)$, so that there is no interaction.

with time. In this chapter we use a linear function for $\delta(t)$ because this matches our experimental observations, however other functional forms for $\delta(t)$ are possible. In all simulations we apply the Langevin model on a domain of size $1831.5\,\mu\text{m} \times 1439\,\mu\text{m}$,

which is the size of the experimental field of view after the boundary columns have been neglected.

Reduced models

- Model II: constant cell size ($\delta(t) = 29.26\,\mu$m) with cell diffusion and intercellular forces

- Model III: variable cell size with intercellular forces but no cell diffusion, $D = 0$ and $f_0 > 0$

- Model IV: variable cell size with diffusion and without intercellular forces, $f_0 = 0$ and $D > 0$

To simulate Models I–IV we must apply appropriate boundary conditions to reflect the conditions relevant to the experiment. To determine these boundary conditions, we note that the experimental images, shown in Figure 4.2(A)-(B), correspond to a field of view that is much smaller than the spatial extent of the experiment. For example, the width of the field of view in Figure 4.2(A)-(B) is 1979 μm, which is much smaller than the diameter of the well in the 96-well plate (9000 μm), as shown in Figure 4.3. The schematic in Figure 4.3 is important because it emphasizes that the images from this kind of assay only show a small proportion of the population of cells present in the well. In particular, it is important to remember that the boundaries around the field of view are not physical boundaries since the spatially uniform population of cells extends far beyond the boundaries around the field of view. This means that whenever the density is below confluence, cells will migrate, in each direction, across the boundaries of the field of view. However, since the population of cells is placed uniformly into each well of the 96-well plate, the net flux of cells across the boundaries of the field of view will be approximately zero owing to symmetry. To justify zero net flux boundary conditions we also count a number of cells in each experimental replica at $t = 0, 12, 24, 36$, and 48 hours (Figure 4A.2, Chapter 4A). These additional results demonstrate that cell counts remain approximately constant with average fluctuations of 8.1% of cell count at $t = 0$ h across four experiments. Similar experimental set up without Mitomycin-C pretreatment would have resulted in about 100 % increase in cell count due to proliferation. Therefore, we impose zero net flux boundary conditions around all boundaries of the field of view (Simpson et al., 2018). We implement these boundary conditions by simply aborting any potential movement event that would take a particular agent across one of the boundaries.

4.4 Results and Discussion

To quantitatively compare the suitability of Models I–IV to describe our experimental data set, we calibrate each model to provide the best match to the experimental density data. For Model I our aim is to estimate the two model parameters, (D, f_0), that lead to the best match with the experimental measurements. To facilitate this we introduce a measure of the discrepancy between the experimental data and the model solution,

$$E(D, f_0) = \frac{1}{148} \sum_{j=1}^{4} \sum_{i=1}^{37} \left[p_{\text{data}}(x_i, t_j) - p_{\text{model}}(x_i, t_j) \right]^2, \tag{4.5}$$

where $E(D, f_0)$ measures the discrepancy between the experimental cell density, $p_{\text{data}}(x_i, t_j)$, and the density predicted by the model, $p_{\text{model}}(x_i, t_j)$, for given values of D and f_0. The index i denotes column number along the x coordinate, and index j denotes time so that $j = 1, 2, 3$ and 4 correspond to the experimental time points $t = 12, 24, 36,$ and 48 h, respectively. The experimental density estimates, $p_{\text{data}}(x_i, t)$, correspond to the averaged experimental density, where the average is taken across all four identically-prepared experimental replicates. We find that it is necessary to estimate $E(D, f_0)$ using averaged experimental data, rather than working with the four experimental data sets separately since the fluctuations in the data from the individual replicates lead to sufficiently large fluctuations in our estimates of $E(D, f_0)$. We will refer to the function $E(D, f_0)$ as an *error surface*. We will visualize the surface and seek to find values of D and f_0 that minimize $E(D, f_0)$, and we denote these estimates as $\bar{D}$ and $\bar{f}_0$, respectively. All results, for Models I–IV, are obtained by numerically integrating the governing equations in two-dimensional space using a forward Euler method with a constant time step of duration $\delta t = 0.02$ h. This choice of time step is sufficiently small that our results are grid-independent. We then construct one-dimensional density distributions from the model output, $p_{\text{model}}(x_i, t_j)$, using the same procedure that is used to convert the distribution of cells in the two-dimensional experimental images into one-dimensional density profiles.

Since we have access to four identically prepared initial conditions for our experimental data set, each time we attempt to match the models with the averaged experimental data we repeat the process four times using the four different choices of initial conditions associated with experimental data sets A, B, C and D. This approach means that we can estimate and visualize the error surface four times for each particular model. A mathematical model that is compatible with the data ought to lead to estimates of D and f_0 that are consistent across the four initial conditions, and so we will examine Models I–IV to see whether they are capable of providing parameter estimates that are consistent across the four different initial conditions.

Previous modelling studies based on using reaction-diffusion equations indicate that

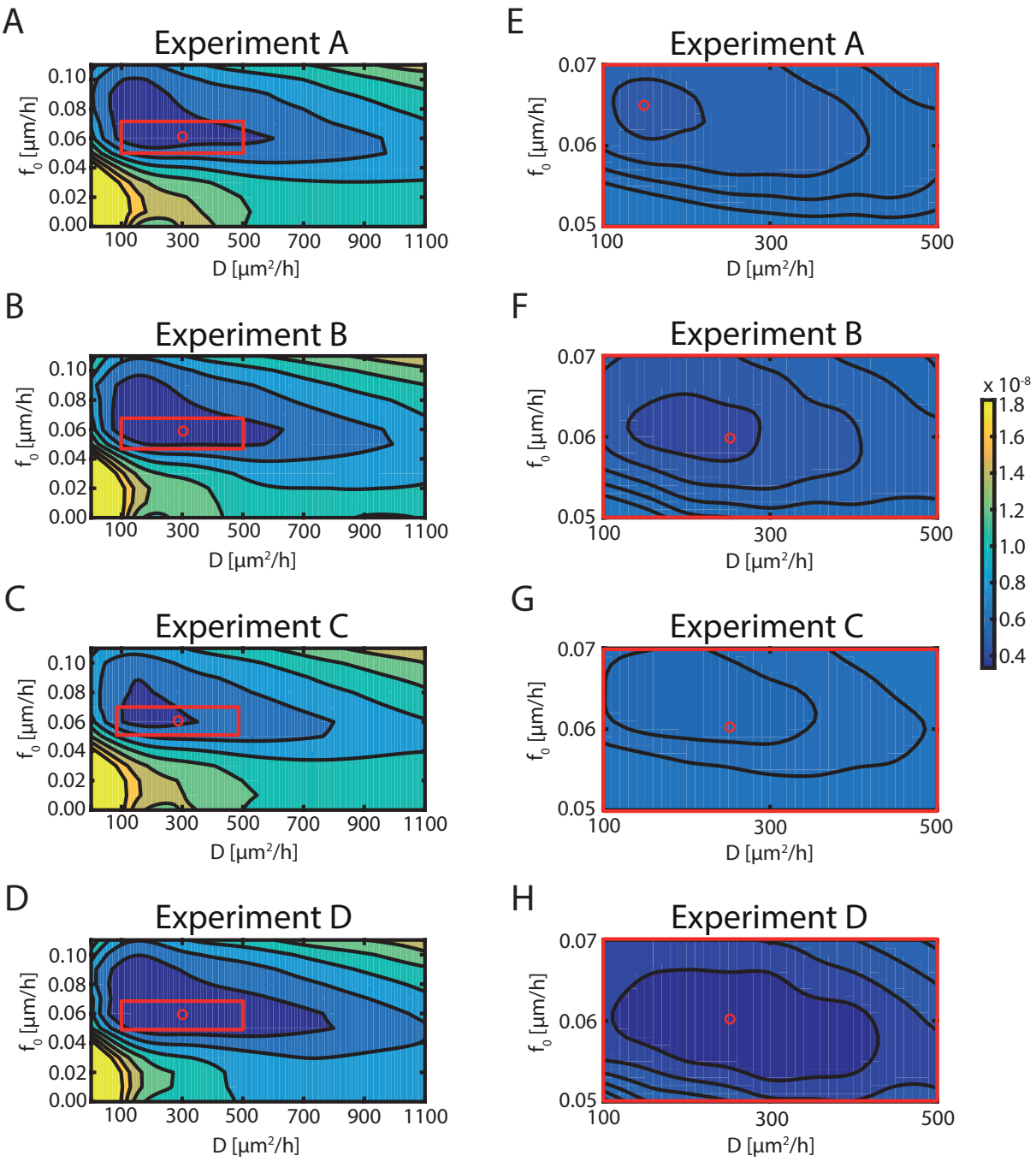


Figure 4.8: $E(D, f_0)$ for Model I. A-D: Error surface contours, $E(D, f_0)$, for four identically-prepared IncuCyte ZOOM^{TM} scratch assays. Each surface contour plot is constructed using seven values of the diffusivity in $0 \leq D \leq 1100$ μm^2/h, and 12 equally-spaced values of the force amplitude in $0 \leq f_0 \leq 0.11$ μm/h. E-H: Refined error surface contours obtained by estimating $E(D, f_0)$ on a refined discretisation of the parameter space within the red rectangles in A-D. The values of $E(D, f_0)$ are shown on the color bar. The location of the best-fit estimate $(\bar{D}, \bar{f}_0)$ is shown as a red circle in each subfigure.

estimates of diffusivity for PC-3 prostate cancer cells varies significantly, from about 300 μm^2/h for low density conditions to approximately 1000 μm^2/h for high density conditions (Jin et al., 2016a). Consequently, we focus our search for parameter estimates in the interval $0 \leq D \leq 1100$ μm^2/h. In contrast, there are no previous estimates of f_0 for the PC-3 cell line. Since we have little initial guidance about an appropriate choice

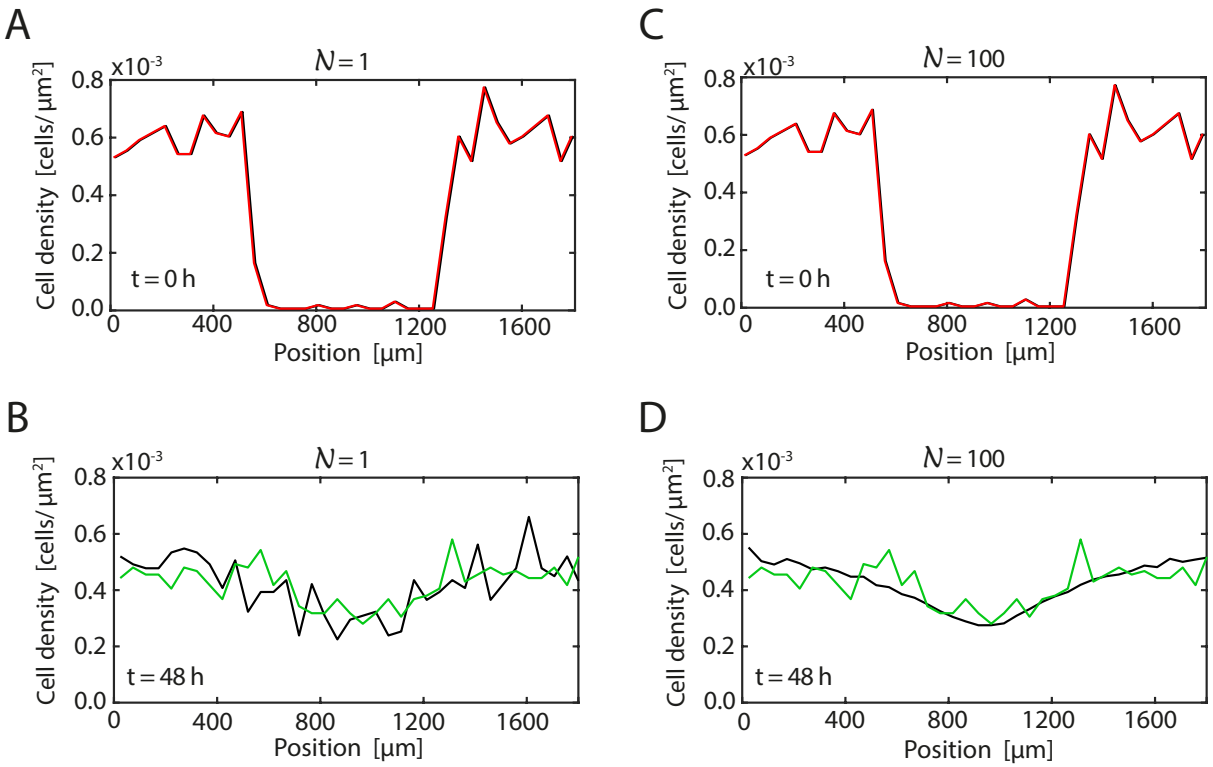


Figure 4.9: Comparison of the calibrated solutions of Model I and experimental profiles from experiment D. Results in A-B compare the experimental data (red, green) and Model I solution (black) at $t = 0$ and $t = 48$ h, respectively, for $\mathcal{N} = 1$ model realization. Results in C-D compare the experimental data (red, green) and Model I solution (black) at $t = 0$ and $t = 48$ h, respectively, averaged over $\mathcal{N} = 100$ identically prepared realizations of the model. The best-fit parameter estimates are $(\bar{D}, \bar{f}_0) = (250, 0.06)$ from Figure 4.8(H).

of f_0, we first conducted a series of preliminary simulations (not shown) to determine an acceptable range of f_0 for each model. This exercise suggests that acceptable ranges are approximately $0 \leq f_0 \leq 0.11$ μm/h for Model I; $0 \leq f_0 \leq 0.6$ μm/h for Model II; and $0 \leq f_0 \leq 0.11$ μm/h for Model III.

To estimate $\bar{D}$ and $\bar{f}_0$ for Model I we estimate discrete values of $E(D, f_0)$ by using a series of numerical solutions of Equation (3A.1) over many (D, f_0) pairs. For each parameter pair we generate an ensemble of 100 identically prepared realizations of the stochastic model and then estimate $E(D, f_0)$ by averaging the density data from the 100 identically prepared realizations. Results in Figure 4.8(A)-(D) show the error surfaces constructed using seven values of diffusivity, $D = 0, 100, 300, 500, 700, 900, 1100$ μm^2/h and 12 equally-spaced values of the force amplitude in the interval $0 \leq f_0 \leq 0.11$ μm/h. All contour lines in Figure 4.8 are obtained using the Matlab spline interpolation function `griddedInterpolant`. Since the initial positions of agents in the discrete simulations exactly correspond to the positions of cells in experimental images, given in Figure 4.5, the magnitude of the fluctuations in the experimental data are consistent with the magnitude of the fluctuations from one realization of the stochastic model. In general, the magnitude of the fluctuations in the stochastic model decreases as the

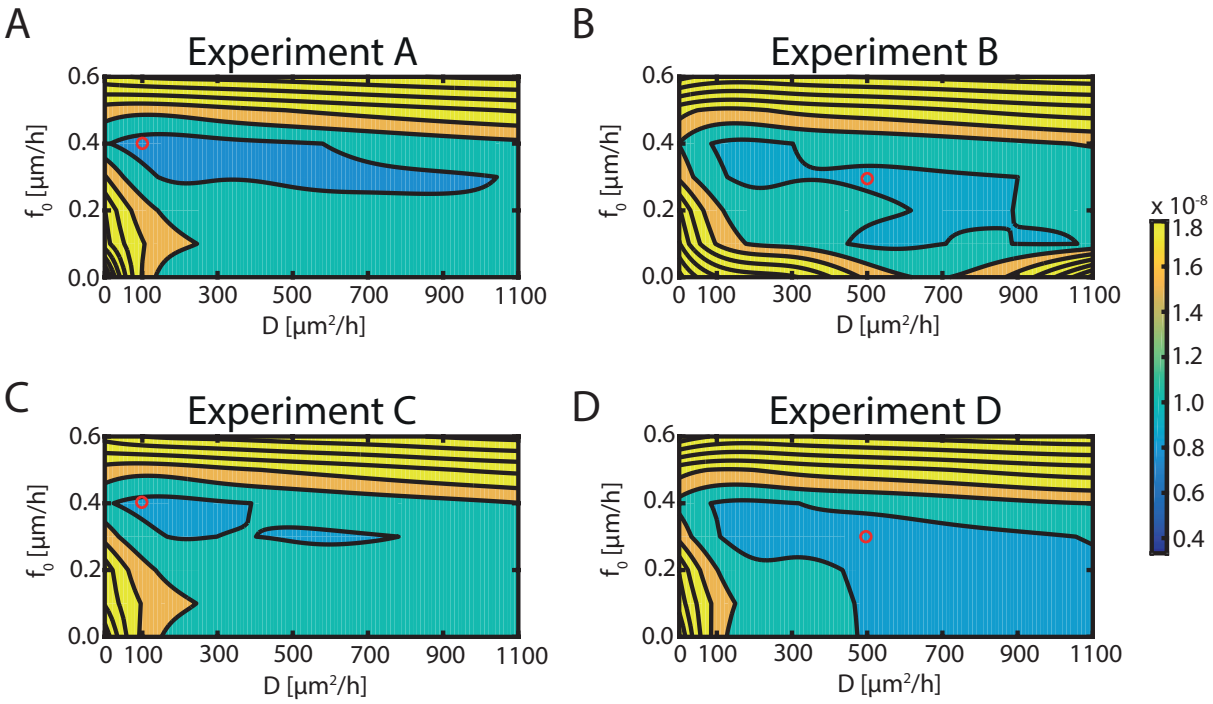


Figure 4.10: $E(D, f_0)$ for Model II. A-D: Error surface contours, $E(D, f_0)$, for four identically-prepared IncuCyte ZOOM^{TM} scratch assays. Each surface contour plot is constructed using seven values of diffusivity in $0 \leq D \leq 1100 \mu m^2/h$ and seven equally-spaced values of the force amplitude in $0 \leq f_0 \leq 0.6$ μm/h. The values of $E(D, f_0)$ are shown on the color bar. In each case the location of the best-fit estimate $(\bar{D}, \bar{f}_0)$ is shown as red circle.

number of realisations increases, $\sim 1/\sqrt{N}$. Therefore, our choice of using $N = 100$ realizations leads to averaged discrete density profiles with fluctuations that are approximately one order of magnitude smaller than fluctuations in the experimental data.

Preliminary estimates of $\bar{D}$ and $\bar{f}_0$ are obtained by evaluating the error surface across a relatively coarse discretisation of the parameter space. These estimates are shown in Figure 4.8(A)-(D) as red circles. We then refine our estimates of $\bar{D}$ and $\bar{f}_0$ by considering a refined discretisation of a subregion surrounding each red circle in Figure 4.8(A)-(D). These subregions are shown as red rectangles in Figure 4.8(A)-(D). Refined plots of the error surface in Figure 4.8(E)-(H) are obtained by calculating $E(D, f_0)$ across five equally-spaced values of D in $100 \leq D \leq 500$ $\mu m^2/h$ and five equally-spaced values of f_0 in $0.05 \leq f_0 \leq 0.07$ μm/h. Each individual plot of $E(D, f_0)$ in Figure 4.8(E)-(H) shows that we have a well-defined minmum from which we can estimate $\bar{D}$ and $\bar{f}_0$. Comparing data in Figure 4.8(E)-(H) shows that our estimates of $\bar{D}$ and $\bar{f}_0$ are consistent between the four experimental replicates. In fact, three of the four refined plots give remarkably consistent estimates of $(\bar{D}, \bar{f}_0) = (250, 0.06)$. Only one of the experimental replicates, shown in Figure 4.8(E), gives slightly different parameter estimates, $(\bar{D}, \bar{f}_0) = (150, 0.065)$.

To visualize the quality of match between the experimental data and discrete profiles predicted by the stochastic model we use the best-fit parameter estimates $(\bar{D}, \bar{f}_0) = (250,$

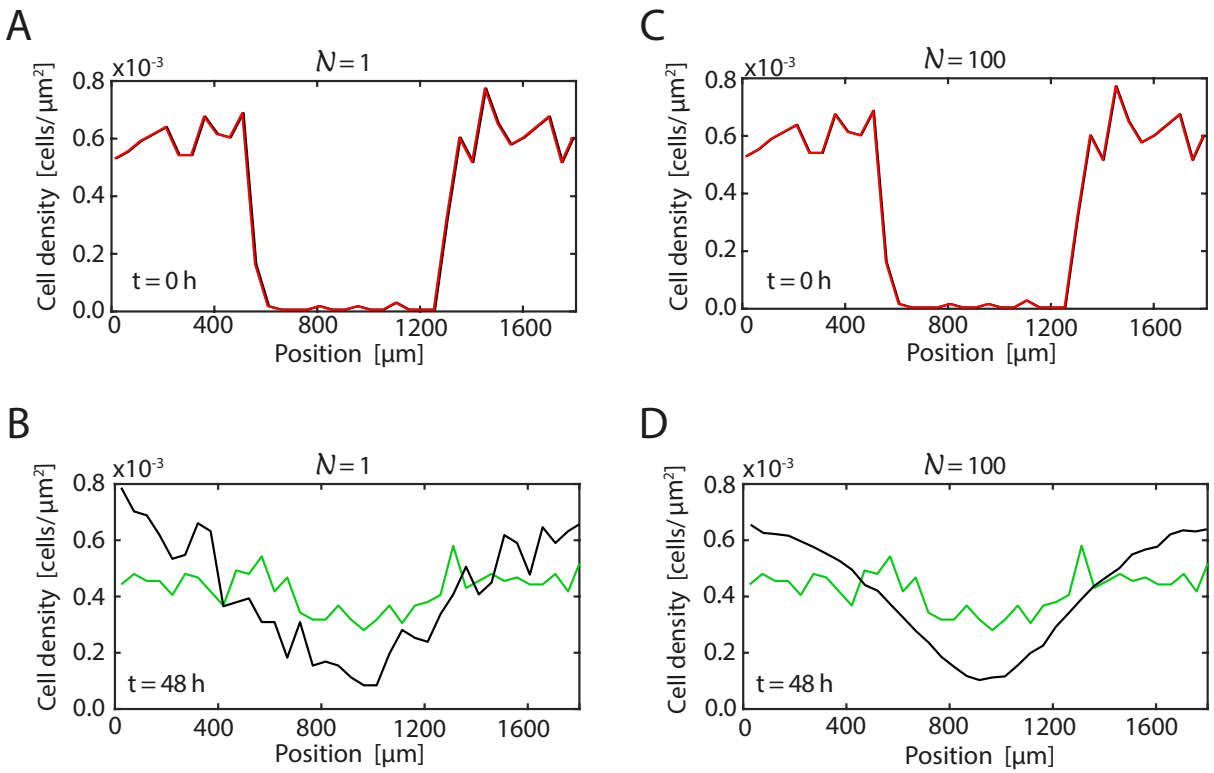


Figure 4.11: Comparison of the calibrated solutions of Model II and experimental profiles from experiment D. Results in A-B compare the experimental data (red, green) and Model II solution (black) at $t = 0$ and $t = 48$ h, respectively, for $\mathcal{N} = 1$ model realization. Results in C-D compare the experimental data (red, green) and Model II solution (black) at $t = 0$ and $t = 48$ h, respectively, averaged over $\mathcal{N} = 100$ identically prepared realizations of the model. The best-fit parameter estimates are $(\bar{D}, \bar{f}_0) = (500, 0.3)$ from Figure 4.10(D).

0.06) for experiment D. First, we solve Model I with $(\bar{D}, \bar{f}_0) = (250, 0.06)$ and calculate the density profiles from the simulations as before. Second, we superimpose density profiles from the discrete simulations with the experimental density distributions, as shown in Figure 4.9. The choice of initial conditions in the stochastic model guarantees that we have an exact match between the experimental density profile and the simulation density profiles at $t = 0$ h. However, since we are dealing with stochastic experimental data and a stochastic mathematical model we do not expect there to be an exact match at later times. Results in Figure 4.9(A)-(B) show that we have a reasonable match between the calibrated simulation results and the experimental density data for a single realization of the stochastic model. Similarly, results in Figure 4.9(C)-(D) show that we also have a good match between the simulation results and the experimental density data over 100 identically prepared realizations of the stochastic model where we see that the magnitude of the fluctuations in the averaged stochastic data are reduced.

We now turn our attention to calibrating Model II to match the experimental data. To achieve this we repeat the exact same calibration process except that we implement Model II with constant agent size. Comparing results in Figure 4.9(A)-(D) and Figure 4.10(A)-(D) shows that the best fit parameters in Model II lead to a larger value of

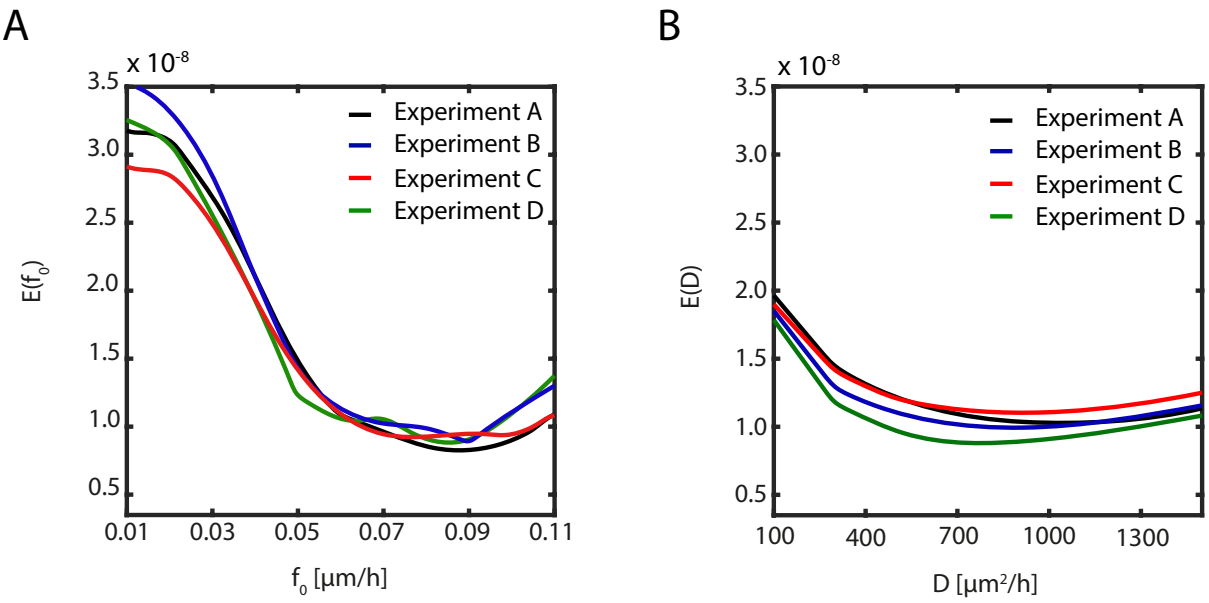


Figure 4.12: $E(f_0)$ and $E(D)$ for Models III and IV. A: $E(f_0)$ for Model III constructed for each experimental replicate with 11 equally-spaced values of f_0 in $0.01 \leq f_0 \leq 0.11$ μm/h. B: $E(D)$ for Model IV constructed for each experimental replicate with eight equally-spaced values of D in $100 \leq D \leq 1500$ μm^2/h.

$E(D, f_0)$. Furthermore, comparing estimates of $\bar{D}$ and $\bar{f}_0$ for Model II between the four identically prepared experimental data sets shows that we have a much higher degree of variability between our parameter estimates for Model II than we did for Model I. Overall, these results suggest that Model I is more consistent with our experimental data than Model II, and so we do not proceed with any further refinement of our parameter estimates for Model II. This result shows that the traditional approach of neglecting the dynamical changes in cell size has clear impact on the ability of the model to describe the behaviour of the entire cell population.

To visualize the quality of match between the experimental data and best-fit density profiles predicted by Model II we use the best-fit parameter estimates $(\bar{D}, \bar{f}_0) = (500, 0.3)$ for experimental replicate D. Again, we solve Model II with these parameter estimates and then estimate the density profiles from those simulations. Results in Figure 4.11 show the density profiles from the discrete simulations superimposed on the corresponding experimental density distributions. Visually, we see that the quality of match in Figure 4.11(B) and Figure 4.11(D) is notably poorer than the quality of match in Figure 4.9(B) and Figure 4.9(D). This visual difference is consistent with the quantitative differences in $E(D, f_0)$ in Figure 4.8 and Figure 4.10.

Finally, we calibrate Models III and IV to match the experimental data. Note that Model III neglects the role of random motility so our parameter estimation involves just one parameter, f_0. Similarly, Model IV neglects the role of intercellular forces so our parameter estimation involves just one parameter, D. Following a now familiar procedure, we compute measures of discrepancy, $E(f_0)$ and $E(D)$, for Models III and

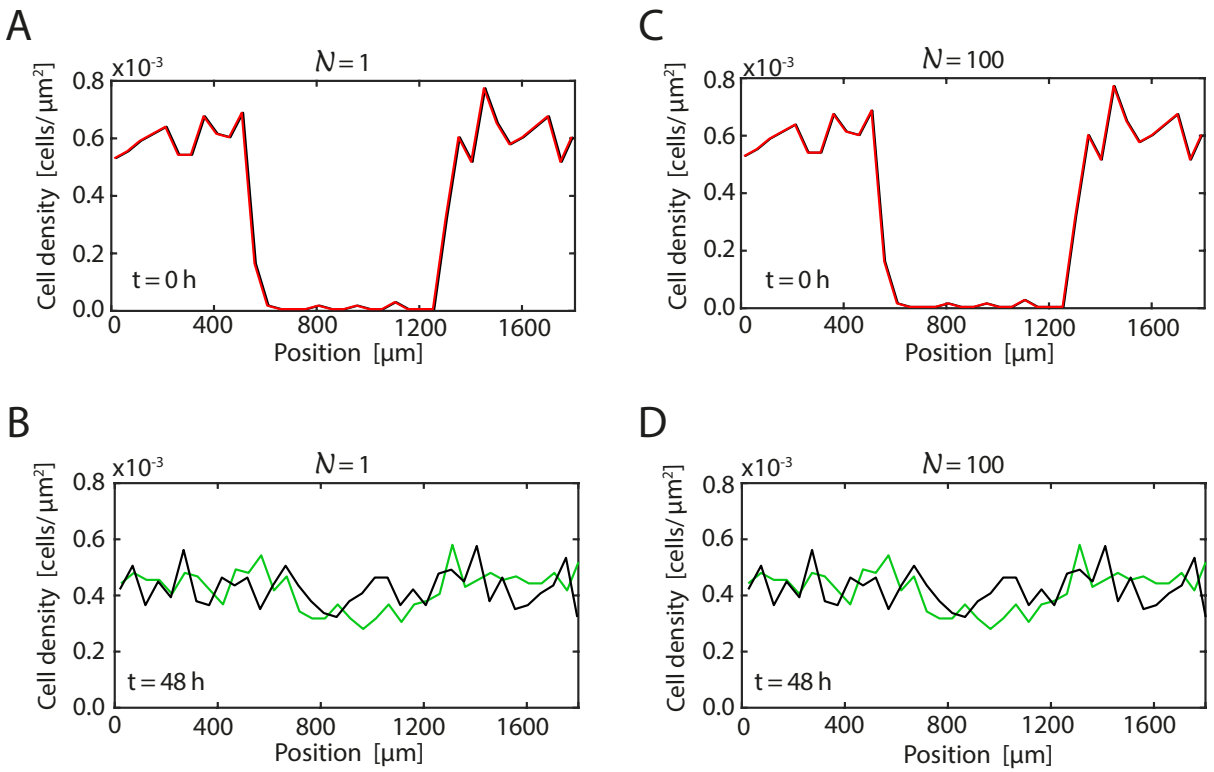


Figure 4.13: Comparison of the calibrated solutions of Model III and experimental profiles from experiment D. Results in A-B compare the experimental data (red, green) and Model III solution (black) at $t = 0$ and $t = 48$ h, respectively, for $\mathcal{N} = 1$ model realization. Results in C-D compare the experimental data (red, green) and Model II solution (black) at $t = 0$ and $t = 48$ h, respectively, averaged over $\mathcal{N} = 100$ identically prepared realizations of the model. The best-fit parameter estimate is $\bar{f}_0 = 0.08$ μm/h from Figure 4.12(A).

IV, respectively. Results in Figure 4.12(A) show a well-defined minimum for each of the four experimental replicates for Model III, however the best-fit estimates are in the range $0.07 \leq f_0 \leq 0.1$ μm/h, which is approximately double the estimate we identified previously for Model I. In contrast, results in Figure 4.12(B) show that we have relatively poorly-defined minimum for all four experimental replicates for Model IV. In this case the best-fit estimates are in the range $700 \leq D \leq 1000$ μm$^2/h$ which is approximately four times greater than the estimates we obtained for Model I.

Results in Figure 4.13 compare the discrete density profiles obtained using Model III parameterized with the best-fit estimate $f_0 = 0.09$ μm/h and the experimental density profiles from experiment D. We note that Model III is deterministic and produces the same density distribution regardless of the number of model realizations. The quality of match between calibrated Model III and the experimental data for experimental replicate D is reasonable, however the value of $E(\bar{f}_0)$ is greater than the value of $E(\bar{D}, \bar{f}_0)$ for Model I, thereby indicating that Model I produces an improved match to the experimental data. Results in Figure 4.14 compares discrete density profiles obtained using Model IV parameterized with the best-fit estimate $D = 700$ μm^2/h where we see no improvement in the quality of match between the calibrated mathematical model and the experimental

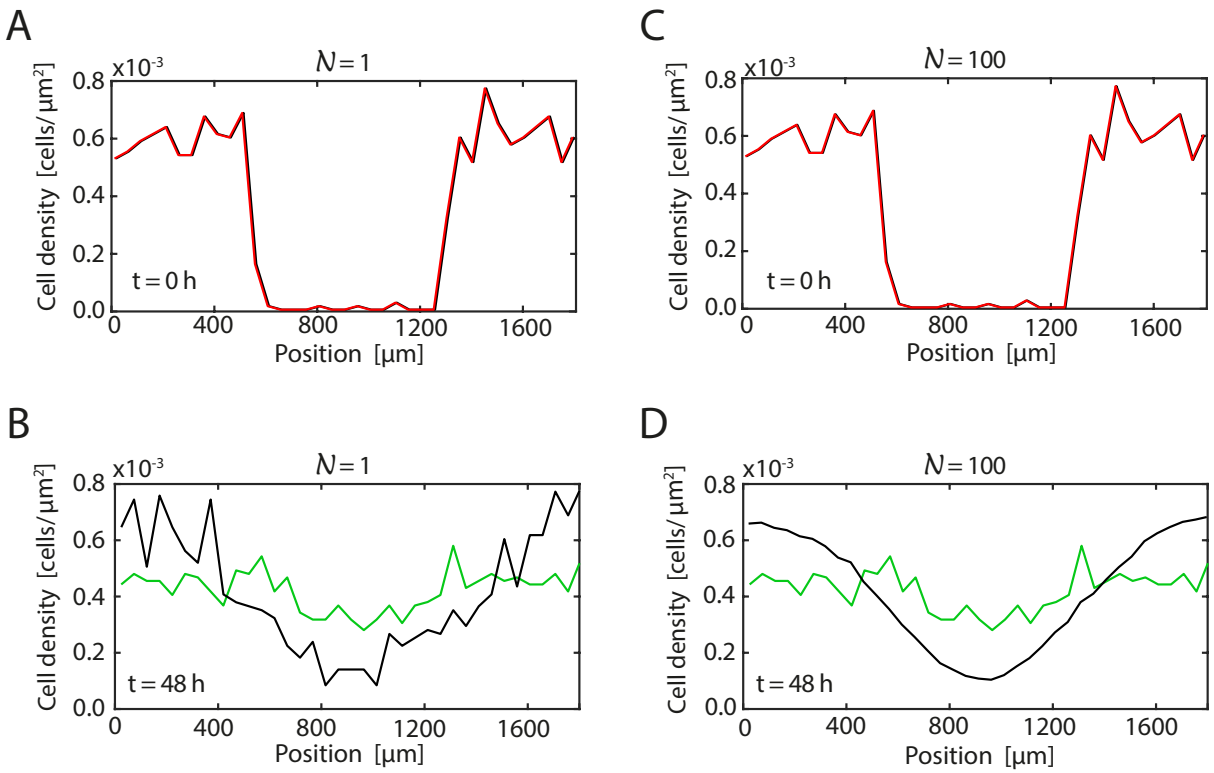


Figure 4.14: Comparison of the calibrated solutions of Model IV and experimental profiles from experiment D. Results in A-B compare the experimental data (red, green) and Model IV solution (black) at $t = 0$ and $t = 48$ h, respectively, for $\mathcal{N} = 1$ model realization. Results in C-D compare the experimental data (red, green) and Model II solution (black) at $t = 0$ and $t = 48$ h, respectively, averaged over $\mathcal{N} = 100$ identically prepared realizations of the model. The best-fit parameter estimates is $\bar{D} = 700$ μm^2/h from Figure 4.12(B).

data relative to Model 1.

All results presented in this section of Chapter 4 focus on comparing the quality of match between Models I–IV and the experimental data using experimental replicate D. Similar comparisons between the best-fit solution of Models I–IV and experimental data from experimental replicates A, B and C are given in Chapter 4A (Figures 4A.4-4A.15). These additional comparisons are consistent with the comparisons made here in Chapter 4.

4.5 Conclusions

In this chapter we use a combined experimental-mathematical modelling approach to quantitatively explore the contribution of cell-to-cell pushing and random motility in driving the movement of cell fronts. We perform a series of IncuCyte ZOOM™ scratch assay experiments in which cells are pretreated with the chemotherapy drug, Mitomycin-C. This approach is useful because Mitomycin-C suppresses proliferation, thereby allowing us to focus on the role of cell migration in the experiments.

We quantitatively assess the role of cell-to-cell interactions, including short range

pushing and longer range adhesion, by calibrating an off-lattice discrete stochastic model to match our experimental data set. The mathematical model that we use accounts for random cell motility, long range cell-to-cell attraction (adhesion), short range cell-to-cell repulsion (pushing) and dynamic cell size changes. We refer to this model as *Model I*. To explore the significance of these various cell-level mechanisms we systematically repeat the model calibration process for a range of simpler, more commonly used models. These simplified models account for: (i) random cell motility, long range cell-to-cell attraction (adhesion) and short range cell-to-cell repulsion (pushing) without any dynamical changes in cell size (*Model II*); (ii) long range cell-to-cell attraction (adhesion) and short range cell-to-cell repulsion (pushing) (*Model III*); and (iii) random cell motility only (*Model IV*).

The novelty in our work resides in the use of the time-dependent cell-to-cell interaction function, Equation (4.3), that allows us to model the dynamical increase in the cell size during the experiments. In contrast, in most of the existing literature it is commonly assumed that cell size remains constant over the time scale of the experiments. We focus on examining the role of the cell-to-cell pushing as a result of combined effects of cell-to-cell repulsion, motility, and dynamic increases in cell size. The role of cell-to-cell pushing is not normally discussed when individual-based models are applied to the experimental data. Here, we use experimental data with cells that grow in size as a result of Mitomycin-C treatment to suppress proliferation. As such, the effects of pushing are much more pronounced in our experiments, as opposed to other experiments where the dynamic change in cell size is not as important.

The phenomenological principles of cell-to-cell interactions modelled in our work are applicable to a range of cell types. However, different cell types can have different balances of mechanisms. For example, epithelial skin cells are normally tightly packed in thin sheets with strong cell-to-cell bonds, while mesenchymal stem cells are highly motile. Consequently, applying our model to a specific cell type would require specific adjustments. Also, we do not consider biochemical pathways of cell-to-cell adhesion and motility, such as E-Cadherin junctions, trailblazing and leader-follower interactions. The leader-follower interactions have a potential to severely influence estimates of the leading edge position, which is normally used to evaluate the rate of spatial spreading of the cell colony. Incorporating these effects is theoretically possible, however, they are not always pronounced in the experimental results and are case specific.

The crucial advantage of our framework is that even in the most advanced *Model I* cell motility is parametrised by only two free parameters. This is important because we are able to explore parameter space reasonably efficiently. In contrast, many discrete models presented in the literature are not as much plausible for calibrating experimental data while addressing same aspects of individual-level behaviour. We note that there is still a certain degree of freedom associated with the choice of cell-to-cell interaction

function. However, our choice is fairly typical and is suitable for our purposes.

Comparing the calibration of these four models to our experimental data provides insight into which model provides the most faithful representation of the experimental observation. Comparing estimates of $E(\bar{D}, \bar{f}_0)$ between the four models shows that Model I provides the best match to the experimental data. This result suggests that properly accounting for random motility, intercellular forces, including short range pushing and longer range attraction, as well as dynamic changes in cell size, are important for this fairly typical experimental protocol. In contrast, calibrating Models II-IV to the data always provides estimates of model parameters that give the best match to the experimental data, but this does not mean that these simpler frameworks are the best model of the underlying biological processes. This is an important result because often in the mathematical biology literature a single type of model will be used to mimic an experiment, without investigation of the more important question of whether that model provides a reasonable description of the underlying biological processes. Here, by systematically comparing the performance of four different, but related, mathematical models to our novel experimental data set, we provide insight into the underlying biological mechanisms in a way that is not possible when working with a single mathematical model in isolation. This approach of calibrating multiple competing mathematical models to match a single data set is a useful way to provide insight into the underlying biological processes, as well as providing insight into the important question of model selection (Jin et al., 2016a; Ciupe et al., 2006).

There are many ways in which our study could be extended to provide additional insight. A key simplifying assumption that we make in our modelling of the experiments is that we give all agents in the simulations the same size at the beginning of the simulation, $\delta(0) = \delta_0$. While our estimates of this initial size are based on experimental estimates given in Figure 4.2(C), our approach is to represent the distribution of observed cell sizes using a sample mean. A close examination of the experimental images in Figure 4.5 shows that there is considerable variability in the distribution of cell sizes at the beginning of the experiment. This variability is captured in the error bars in Figure 4.2(C) but neglected in our analysis. Therefore, a reasonable extension of the current analysis would be to incorporate this initial variability into the stochastic models with a view to understanding how this initial variability influences the population-level motion of the cell fronts. In our work we do not analyse cell-to-cell correlations or spatial structure (Binder and Simpson, 2013). Previously, velocity-velocity correlations have been used to study cell swarming and mechanical waves during tissue expansion and similar analysis could be applied to the experimental data used in our work (Tlili et al., 2018; Serra-Picamal et al., 2018; Zaritsky et al., 2014). We leave these extensions for future consideration.

Chapter 4A

Additional results for Chapter 4

4A.1 Images of IncuCyte ZOOM™ scratch assay experiments A, B, C, and D at times t = 0, 12, 24, 36, and 48 h.

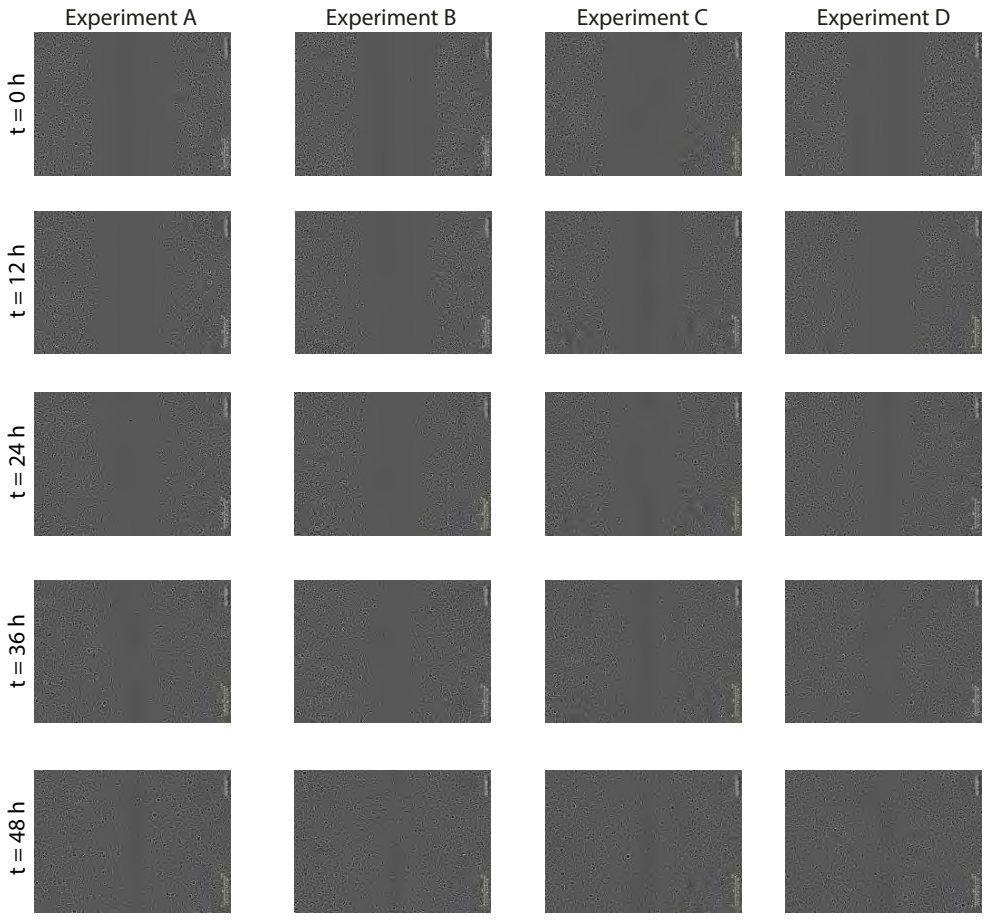


Figure 4A.1: Images of IncuCyte ZOOM™ scratch assay experiments A, B, C, and D at times t = 0, 12, 24, 36, and 48 h.

4A.2 Cell counts for experiments A, B, C, and D at times t = 0, 12, 24, 36, and 48 h.

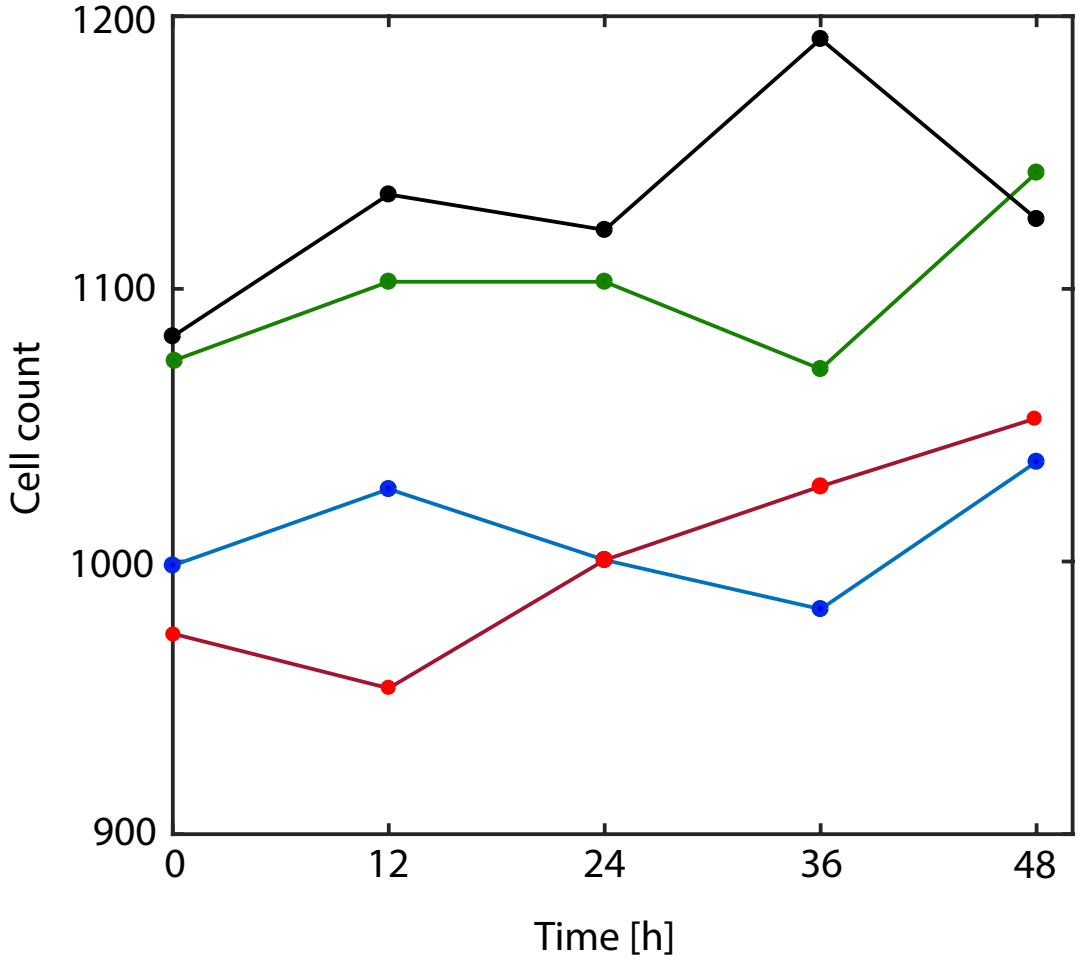


Figure 4A.2: Cell counts for experiment A (blue solid), experiment B (green solid), experiment C (red solid), and experiment D (black solid) at times t = 0, 12, 24, 36, and 48 h.

4A.3 Cell density distributions for experiments A, B, C, and D at times t = 0, 12, 24, 36, and 48 h.

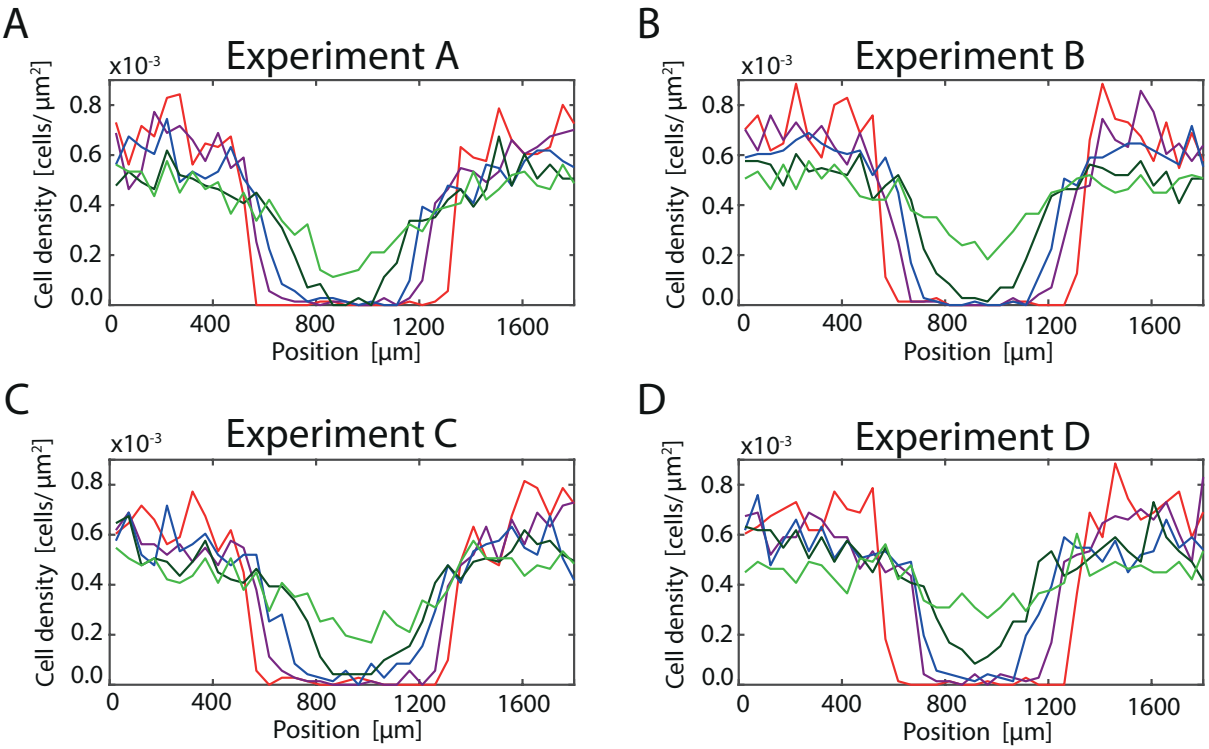


Figure 4A.3: Cell density distributions for experiments A, B, C, and D at times t = 0, 12, 24, 36, and 48 h.

4A.4 Additional results for Models I-IV

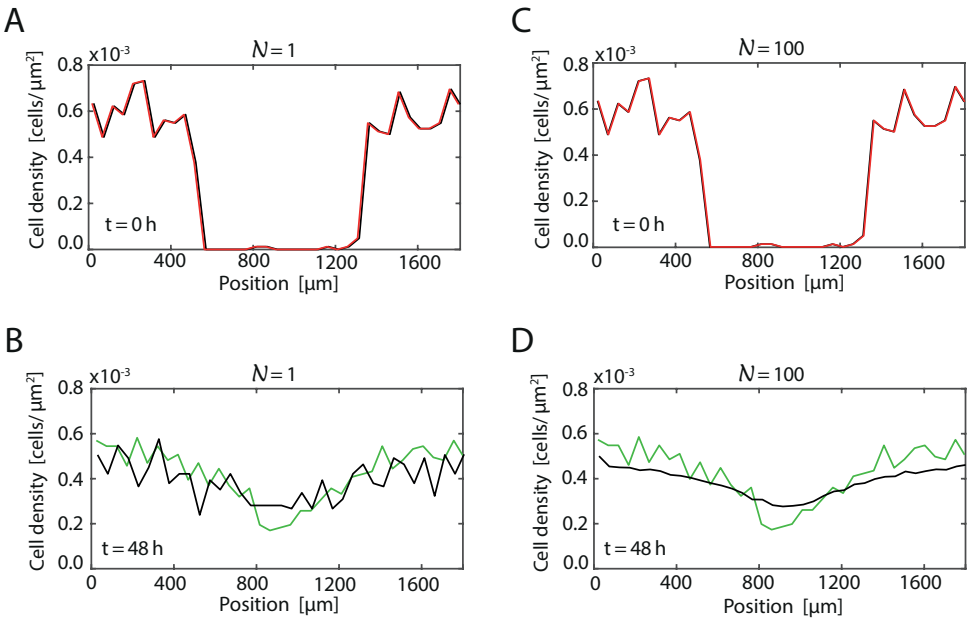


Figure 4A.4: Comparisons of the calibrated solutions of Model I and density profiles obtained from experiment A. Black solid lines denote solutions of the stochastic model constructed from one single realisation, subfigures A-B, and constructed from the averaged solutions of 100 identically prepared realisations of the stochastic model, subfigures C-D, respectively. Discrete profiles are constructed using best-fit estimates $\bar{D} = 150\,\mu\text{m}^2/\text{h}$ and $\bar{f}_0 = 0.065\,\mu\text{m}/\text{h}$. Red and green solid lines represent cell density distributions in experiment A for the time points $t = 0$ and 48 hours.

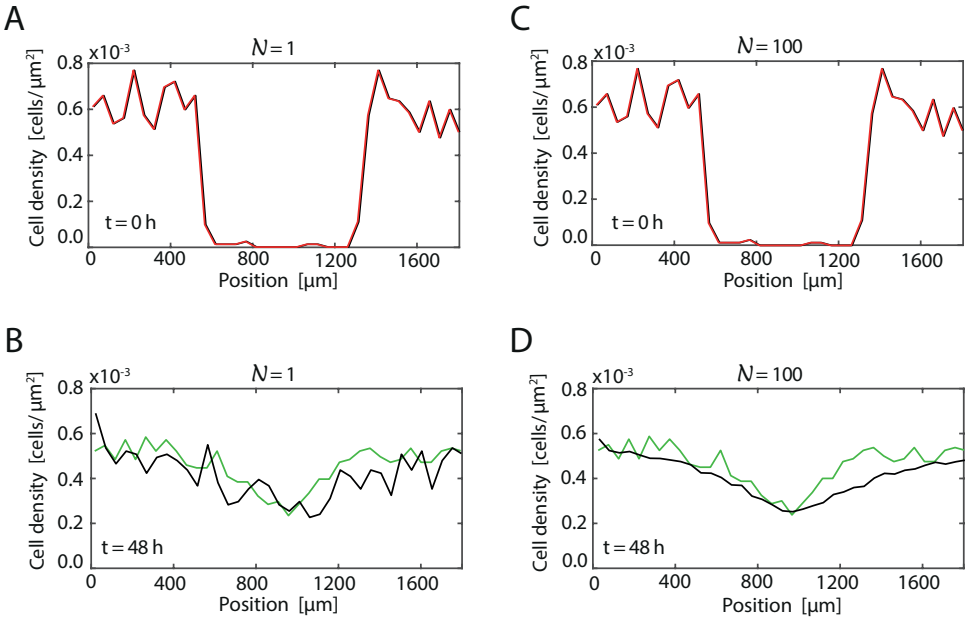


Figure 4A.5: Comparisons of the calibrated solutions of Model I and density profiles obtained from experiment B. Black solid lines denote solutions of the stochastic model constructed from one single realisation, subfigures A-B, and constructed from the averaged solutions of 100 identically prepared realisations of the stochastic model, subfigures C-D, respectively. Discrete profiles are constructed using best-fit estimates $\bar{D} = 250\,\mu\text{m}^2/\text{h}$ and $\bar{f}_0 = 0.06\,\mu\text{m}/\text{h}$. Red and green solid lines represent cell density distributions in experiment B for the time points $t = 0$ and 48 hours.

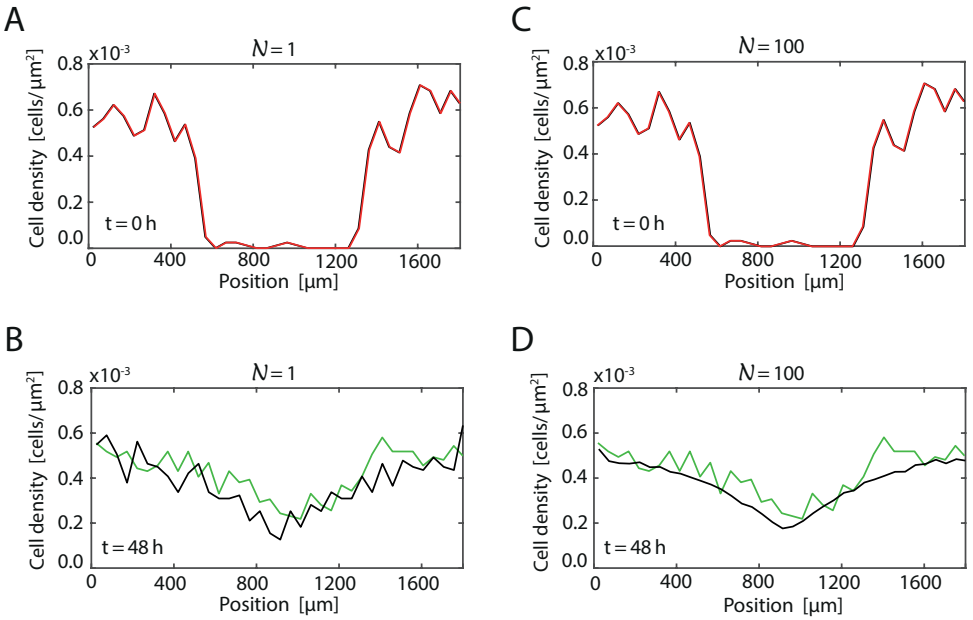


Figure 4A.6: Comparisons of the calibrated solutions of Model I and density profiles obtained from experiment C. Black solid lines denote solutions of the stochastic model constructed from one single realisation, subfigures A-B, and constructed from the averaged solutions of 100 identically prepared realisations of the stochastic model, subfigures C-D, respectively. Discrete profiles are constructed using best-fit estimates $\bar{D} = 250\,\mu\text{m}^2/\text{h}$ and $\bar{f}_0 = 0.06\,\mu\text{m}/\text{h}$. Red and green solid lines represent cell density distributions in experiment C for the time points $t = 0$ and 48 hours.

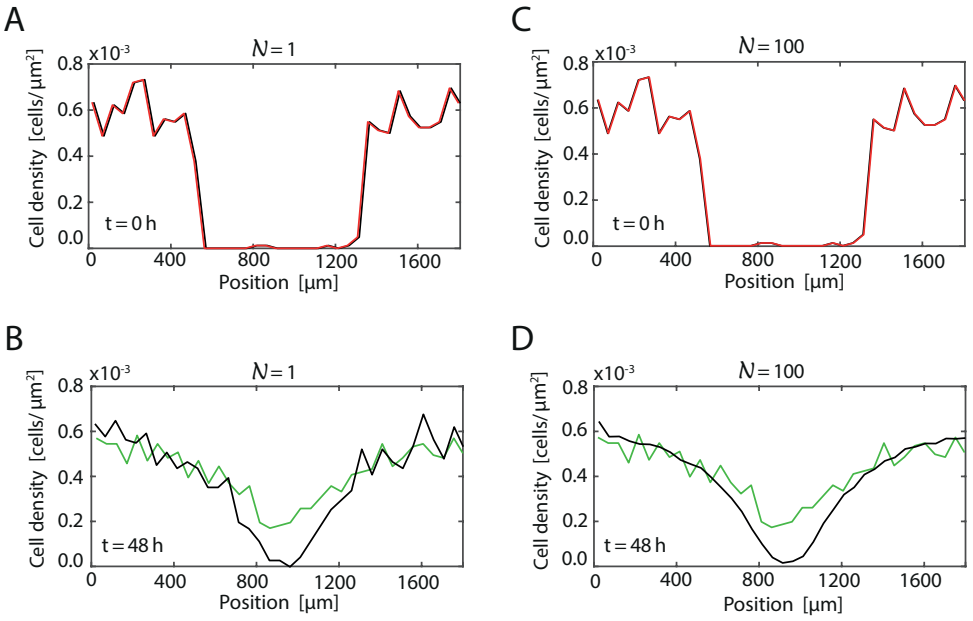


Figure 4A.7: Comparisons of the calibrated solutions of Model II and density profiles obtained from experiment A. Black solid lines denote solutions of the stochastic model constructed from one single realisation, subfigures A-B, and constructed from the averaged solutions of 100 identically prepared realisations of the stochastic model, subfigures C-D, respectively. Discrete profiles are constructed using best-fit estimates $\bar{D} = 150\,\mu\text{m}^2/\text{h}$ and $\bar{f}_0 = 0.4\,\mu\text{m}/\text{h}$. Red and green solid lines represent cell density distributions in experiment A for the time points $t = 0$ and 48 hours.

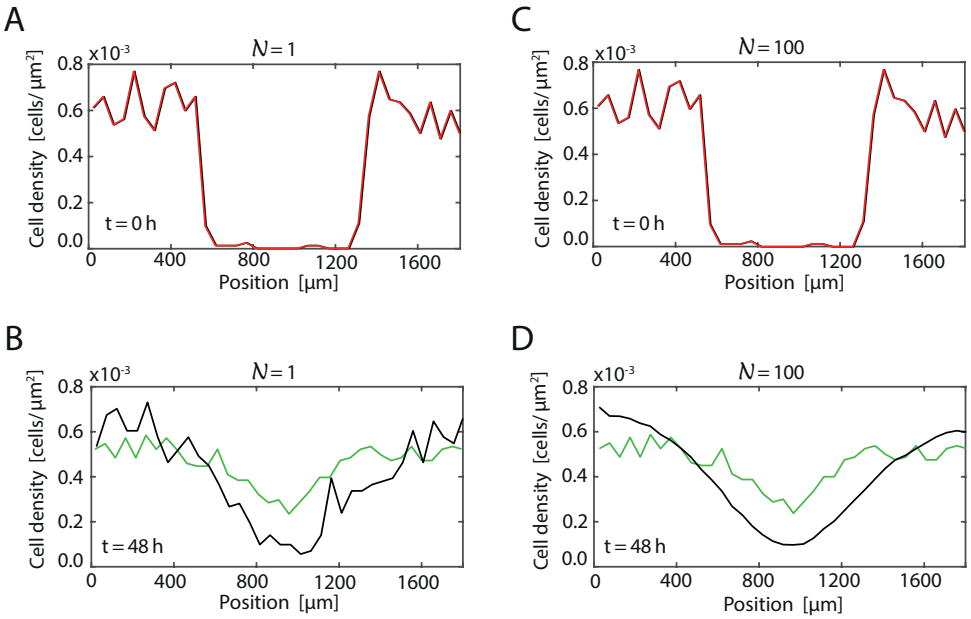


Figure 4A.8: Comparisons of the calibrated solutions of Model II and density profiles obtained from experiment B. Black solid lines denote solutions of the stochastic model constructed from one single realisation, subfigures A-B, and constructed from the averaged solutions of 100 identically prepared realisations of the stochastic model, subfigures C-D, respectively. Discrete profiles are constructed using best-fit estimates $\bar{D} = 500\,\mu\text{m}^2/\text{h}$ and $\bar{f}_0 = 0.3\,\mu\text{m}/\text{h}$. Red and green solid lines represent cell density distributions in experiment B for the time points $t = 0$ and 48 hours.

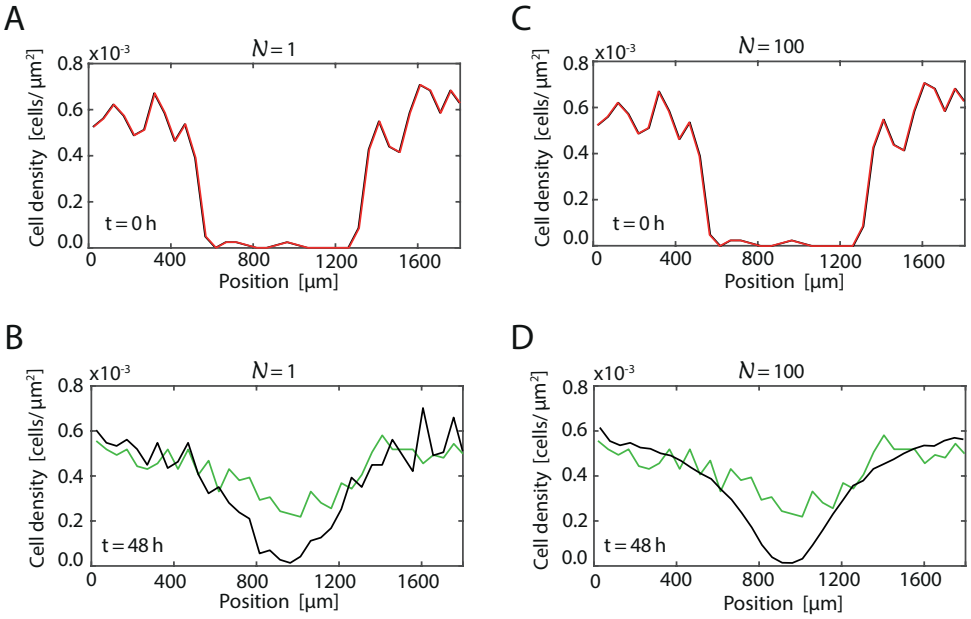


Figure 4A.9: Comparisons of the calibrated solutions of Model II and density profiles obtained from experiment C. Black solid lines denote solutions of the stochastic model constructed from one single realisation, subfigures A-B, and constructed from the averaged solutions of 100 identically prepared realisations of the stochastic model, subfigures C-D, respectively. Discrete profiles are constructed using best-fit estimates $\bar{D} = 150\,\mu\text{m}^2/\text{h}$ and $\bar{f}_0 = 0.4\,\mu\text{m}/\text{h}$. Red and green solid lines represent cell density distributions in experiment C for the time points $t = 0$ and 48 hours.

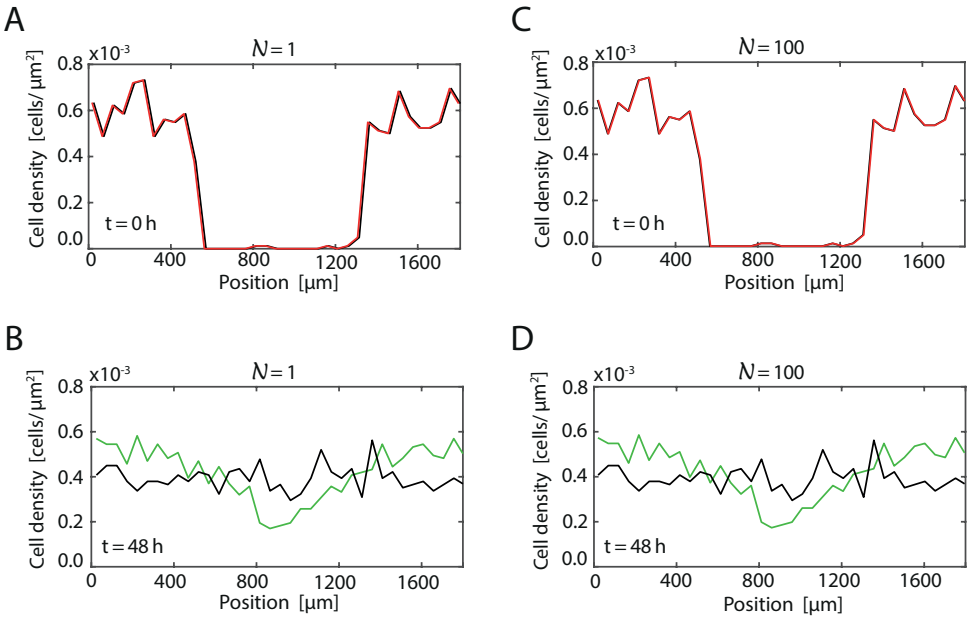


Figure 4A.10: Comparisons of the calibrated solutions of Model III and density profiles obtained from experiment A. Black solid lines denote solutions of the stochastic model constructed from one single realisation, subfigures A-B, and constructed from the averaged solutions of 100 identically prepared realisations of the stochastic model, subfigures C-D, respectively. Discrete profiles are constructed using best-fit estimate $\bar{f}_0 = 0.09\,\mu$m/h. Red and green solid lines represent cell density distributions in experiment A for the time points $t = 0$ and 48 hours.

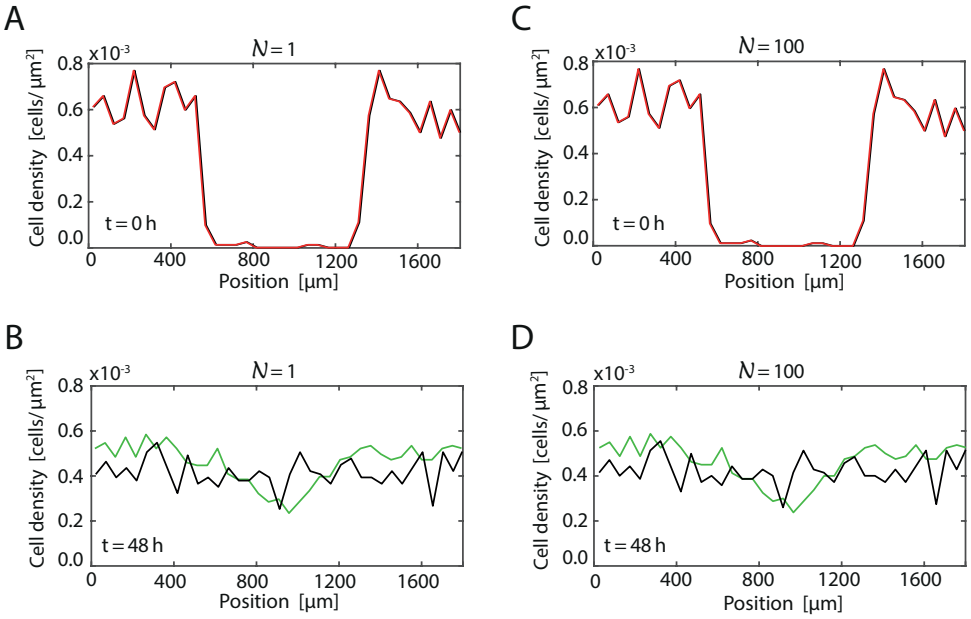


Figure 4A.11: Comparisons of the calibrated solutions of Model III and density profiles obtained from experiment B. Black solid lines denote solutions of the stochastic model constructed from one single realisation, subfigures A-B, and constructed from the averaged solutions of 100 identically prepared realisations of the stochastic model, subfigures C-D, respectively. Discrete profiles are constructed using best-fit estimate $\bar{f}_0 = 0.09\,\mu$m/h. Red and green solid lines represent cell density distributions in experiment B for the time points $t = 0$ and 48 hours.

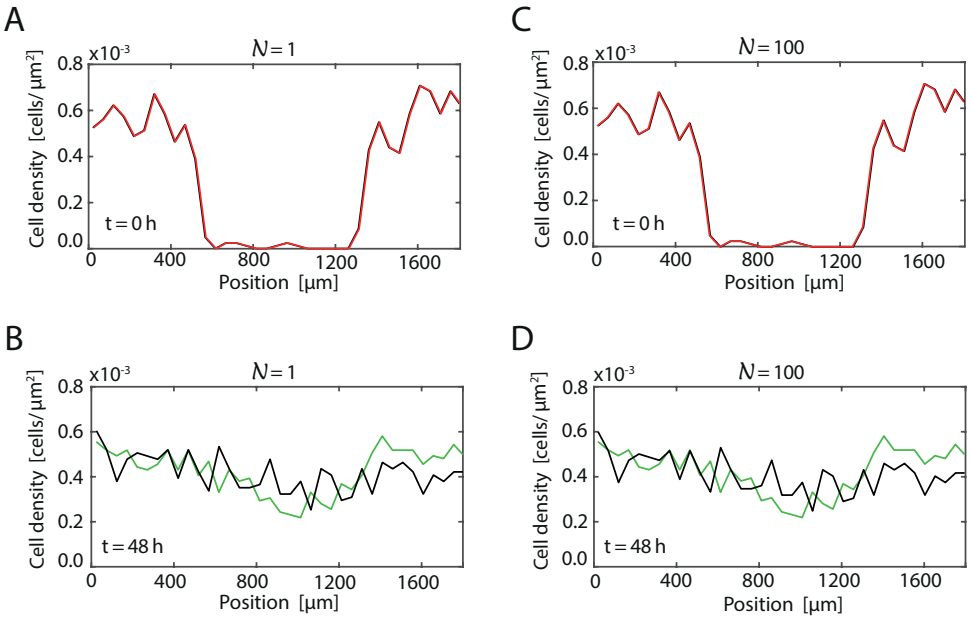


Figure 4A.12: Comparisons of the calibrated solutions of Model III and density profiles obtained from experiment C. Black solid lines denote solutions of the stochastic model constructed from one single realisation, subfigures A-B, and constructed from the averaged solutions of 100 identically prepared realisations of the stochastic model, subfigures C-D, respectively. Discrete profiles are constructed using best-fit estimate $\bar{f}_0 = 0.08\,\mu$m/h. Red and green solid lines represent cell density distributions in experiment C for the time points $t = 0$ and 48 hours.

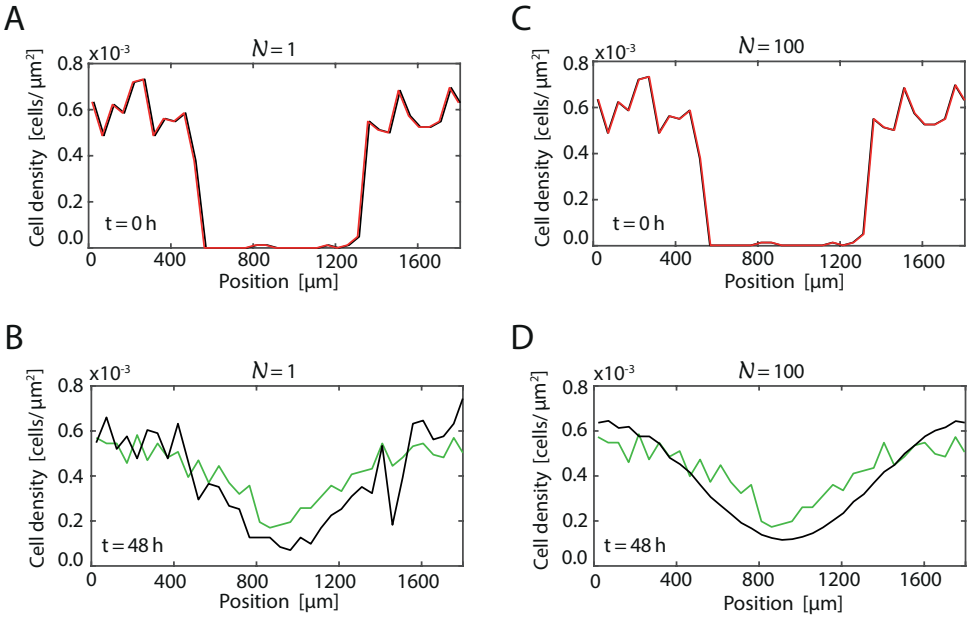


Figure 4A.13: Comparisons of the calibrated solutions of Model IV and density profiles obtained from experiment A. Black solid lines denote solutions of the stochastic model constructed from one single realisation, subfigures A-B, and constructed from the averaged solutions of 100 identically prepared realisations of the stochastic model, subfigures C-D, respectively. Discrete profiles are constructed using best-fit estimate $\bar{D} = 900\,\mu$m^2/h. Red and green solid lines represent cell density distributions in experiment A for the time points $t = 0$ and 48 hours.

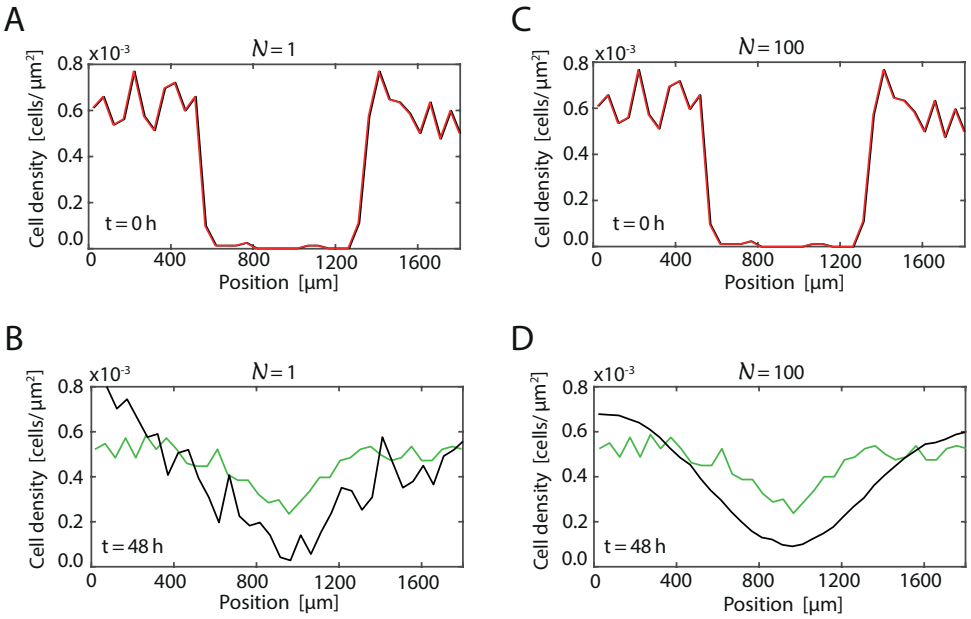


Figure 4A.14: Comparisons of the calibrated solutions of Model IV and density profiles obtained from experiment B. Black solid lines denote solutions of the stochastic model constructed from one single realisation, subfigures A-B, and constructed from the averaged solutions of 100 identically prepared realisations of the stochastic model, subfigures C-D, respectively. Discrete profiles are constructed using best-fit estimate $\bar{D} = 900\,\mu\text{m}^2/\text{h}$. Red and green solid lines represent cell density distributions in experiment B for the time points $t = 0$ and 48 hours.

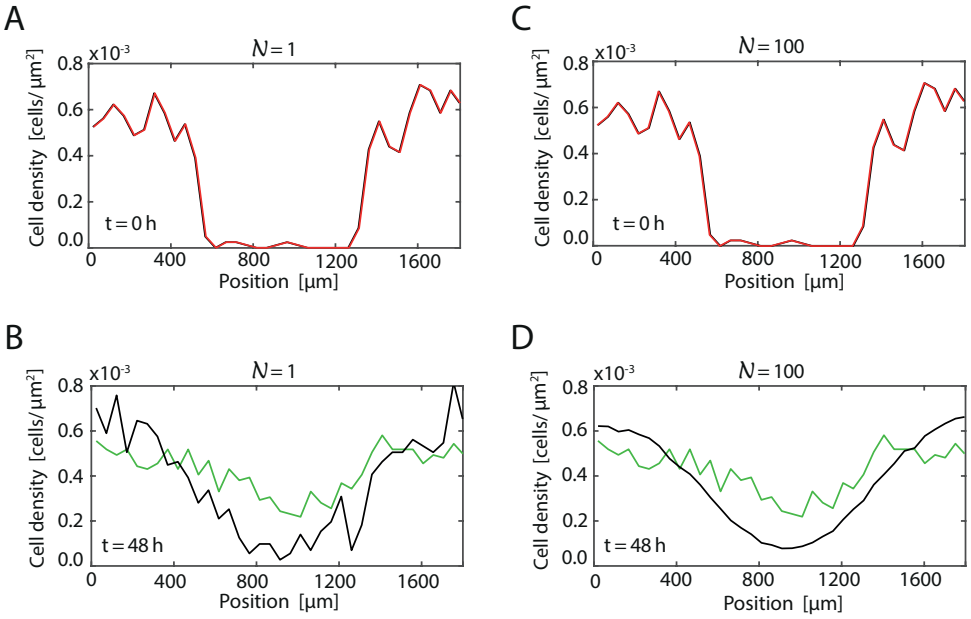


Figure 4A.15: Comparisons of the calibrated solutions of Model IV and density profiles obtained from experiment C. Black solid lines denote solutions of the stochastic model constructed from one single realisation, subfigures A-B, and constructed from the averaged solutions of 100 identically prepared realisations of the stochastic model, subfigures C-D, respectively. Discrete profiles are constructed using best-fit estimate $\bar{D} = 900\,\mu\text{m}^2/\text{h}$. Red and green solid lines represent cell density distributions in experiment C for the time points $t = 0$ and 48 hours.

Chapter 5

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

A paper published in the *Journal of Theoretical Biology*.

Matsiaka Oleksii, Baker Ruth, Simpson Matthew. Continuum descriptions of spatial spreading for heterogeneous cell populations: theory and experiment. *Journal of Theoretical Biology*, vol. 482, 109997, 2019.

5.1 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 chapter, 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 monotonically decreasing and uniform distributions 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 find that the simpler single-species homogeneous model performs relatively poorly when applied to non-monotonic and monotonically increasing distributions of cell sizes. Additional results for heterogeneity in parameters describing both undirected and directed cell migration are also considered, and we find that similar results apply.

5.2 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; Vo et al., 2015). 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).

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; Menon et al., 2018). Figure

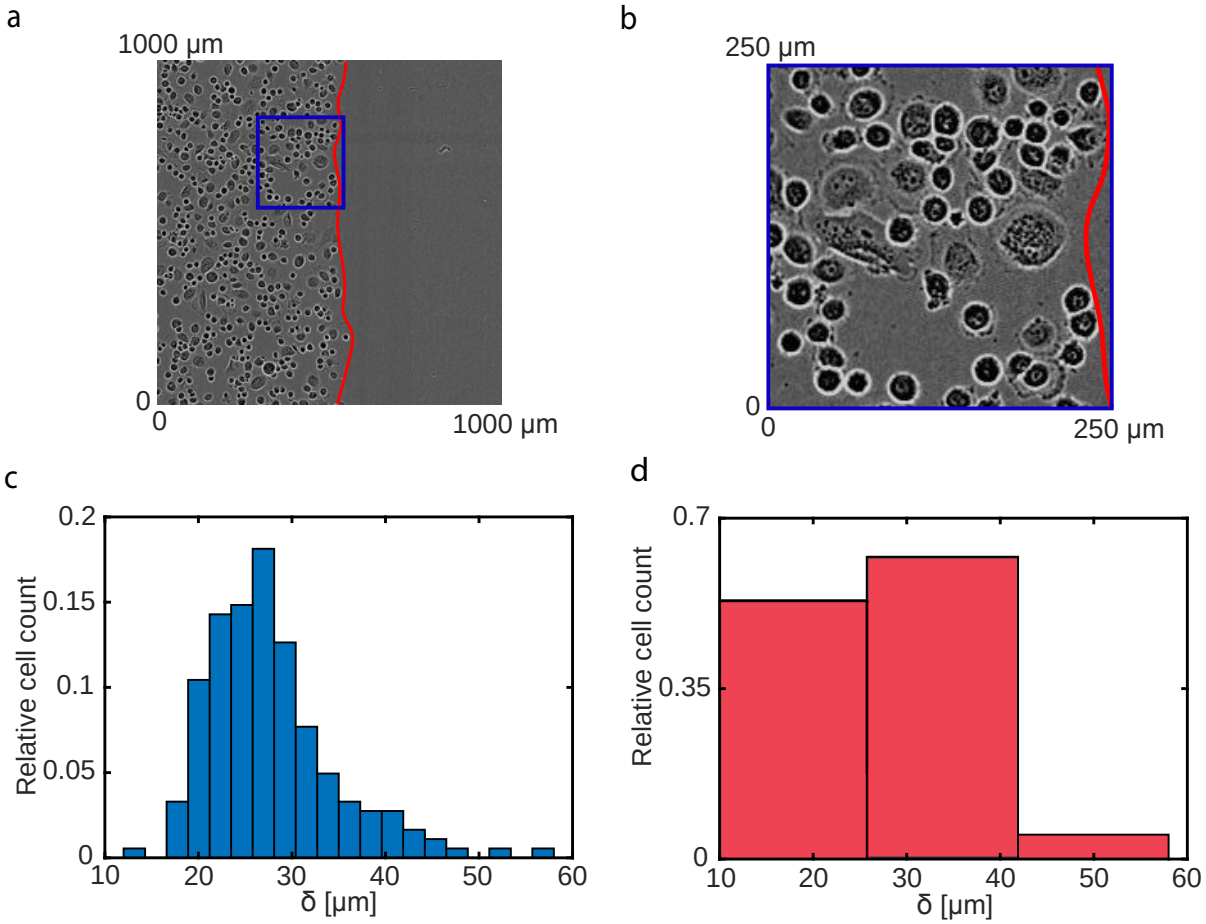


Figure 5.1: Heterogeneity in a population of PC-3 prostate cancer cells (Kaighn et al., 1979). (a) Experimental image of an advancing cell population and corresponding cell size distribution. The red solid line denotes position of the leading edge. (b) Detailed image of the subregion denoted in the blue rectangle in Figure 5.1(a). (c) Cell size distribution with a bin size of $2.3\,\mu$m. 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 $15\,\mu$m. The histogram in Figure 5.1(d) is constructed using the same sample of 184 cells.

5.1(a)-(b) shows a typical experiment where we can clearly visually observe cells of different sizes. The measured cell size distribution in Figure 5.1(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; Chapter 2). Previous work has also attempted to estimate parameters in heterogeneous models that describe glioblastoma progression (Rutter et al., 2018). 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 chapter 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 5.1, we use the multi-species continuum approach introduced by Matsiaka et al. (2017) (Chapter 2). 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 5.1(d). Throughout this chapter we consider four distinct distributions of cell sizes: (i) monotonically decreasing (Set Ia); (ii) uniform (Set Ib); (iii) non-monotonic (Set Ic); and (iv) monotonically increasing (Set Id). The monotonically decreasing distribution, as shown in Figure 5.3(a), is a fairly accurate approximation of the experimentally observed cell size distribution in Figure 5.1(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 chapter is organised in the following way. In Section 5.3 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 5.4 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 5.5 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 presented in Chapter 5A explore the role of: (i) heterogeneity in undirected (diffusive) migration, Set II; and (ii) heterogeneity in directed (adhesion/cell-to-cell contacts) migration, Set III. Finally, in Section 5.6 we summarise our result and propose potential extensions.

5.3 Experimental data

Monolayer scratch assays are performed using the IncuCyte ZOOM™ 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™, Manassas, USA). After growing, cells are removed from the flask using TrypLE™ (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™ (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™ 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 $t = 0\,\text{h}$, shown in Figure 5.1, 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 5.1(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 $2.3\,\mu$m, is shown in Figure 5.1(c). The bin width $2.3\,\mu$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 5.1(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 5.1(d) demonstrates the histogram of cell sizes constructed using the same sample of cells with a larger bin size width of $15\,\mu$m. Here, we have three subpopulations that capture the key trends in the heterogeneity in Figure 5.1(c) without needing to deal with 17 distinct subpopulations.

In this chapter we use experimental data to extract the cell size distribution at $t = 0\,\text{h}$ and use this data to generate the initial conditions in the three-species heterogeneous model (Set Ia, Figure 5.3). An interesting side effect of Mitomycin-C pretreatment is that cells increase in size abnormally fast compared to similar experiments without pretreatment. As a result of pretreatment, the cell size distribution changes significantly with time, which, in turn, represents an additional degree of freedom in the problem. To keep our work tractable, we consider the most fundamental problem where we treat the cell size distribution as being constant through time, and we leave an extension to the case where the cell size distribution varies with time for future analysis.

5.4 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 5.3 (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

$$\frac{\mathrm{d}\vec{x}_i}{\mathrm{d}t} = \sum_{i \neq j} \vec{F}_{ij} + \vec{\xi}_i, \tag{5.1}$$

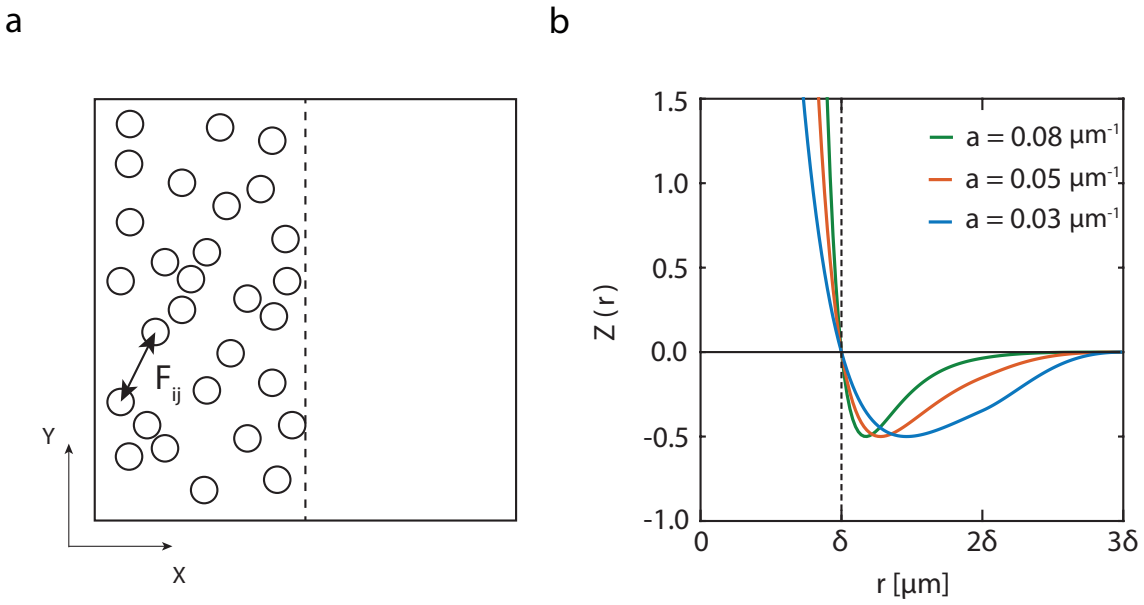


Figure 5.2: (a) An idealisation of the front-like distribution of cells in the experimental design shown in Figure 5.1(a). Here all cells are of constant size. F_{ij} is the interaction force between cell i and cell j. The vertical dashed line represents the approximate leading edge of the population. (b) A typical cell-to-cell interaction force function in the form of the modified Morse potential, $Z(r)$, (Equation (3.6)) 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 $r < \delta$).

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, 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; Chapter 3). 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, as such, are able to retain certain features of a discrete model (Middleton

et al., 2014; O'Dea and King, 2012). In this chapter 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; Chapter 2). 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 5.1, 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; Chapter 3).

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

$$\frac{\partial p^{(1)}(x,t)}{\partial t} = D_1 \Delta p^{(1)}(x,t) + \nabla(p^{(1)}(x,t)\, V^{(1,1)}(x,t)) - \sum_{m=1}^{3} n_m \nabla(p^{(1)}(x,t)\, V^{(1,m)}(x,t)), \tag{5.2}$$

$$\frac{\partial p^{(2)}(x,t)}{\partial t} = D_2 \Delta p^{(2)}(x,t) + \nabla(p^{(2)}(x,t)\, V^{(2,2)}(x,t)) - \sum_{m=1}^{3} n_m \nabla(p^{(2)}(x,t)\, V^{(2,m)}(x,t)), \tag{5.3}$$

$$\frac{\partial p^{(3)}(x,t)}{\partial t} = D_3 \Delta p^{(3)}(x,t) + \nabla(p^{(3)}(x,t)\, V^{(3,3)}(x,t)) - \sum_{m=1}^{3} n_m \nabla(p^{(3)}(x,t)\, V^{(3,m)}(x,t)), \tag{5.4}$$

$$V^{(l,m)}(x,t) = \int_{\Omega} F^{(l,m)}(x-y)\, p^{(m)}(y,t)\, \mathrm{d}y, \tag{5.5}$$

where $p^{(1)}(x,t)$, $p^{(2)}(x,t)$, and $p^{(3)}(x,t)$ are the cell densities associated with each subpopulation and depend on position x and time t. In this heterogeneous model, D_1, D_2, and D_3 are diffusivities of subpopulations 1, 2, and 3, n_1, n_2 and n_3 are the numbers of cells in each subpopulation, operator Δ is defined as $\Delta = \partial^2/\partial x^2$, operator $\nabla = \partial/\partial x$, and $V^{(l,m)}(x,t)$ is the velocity field of subpopulation l induced by subpopulation m (Matsiaka

et al., 2017; Chapter 2). The diffusivity constants parameterise the undirected migration of each subpopulation and the velocity fields describe the directed migration of each subpopulation that is driven by a combination of cell-to-cell adhesion and crowding effects. Overall, the parameter regime used here to simulate three-species populations of cells is consistent with the parameter regime for which mean field model have previously demonstrated to be in a good agreement with the individual-based model (Chapter 2 and Chapter 3).

The interaction force between subpopulations l and i that describes directed migration is given by

$$F^{(l,i)}(x-y) = f_0^{(i)} \mathcal{Z}_i(r)\,\mathrm{sgn}(x-y), \tag{5.6}$$

where $f_0^{(i)}$ is the dimensional amplitude of the interaction force acting on subpopulation i, $\mathcal{Z}_i(r)$ is a dimensionless 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; Painter et al., 2010). A number of different phenomenological laws, $\mathcal{Z}_i(r)$, 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

$$\mathcal{Z}(r) = \begin{cases} 2\Big(\exp[-2a\,(r-\delta)] - \exp[-a\,(r-\delta)]\Big), & r < 2\delta, \\ 2\Big(\exp[-2a\,(r-\delta)] - \exp[-a\,(r-\delta)]\Big)\,g(r), & 2\delta \le r \le 3\delta, \\ 0, & r > 3\delta, \end{cases} \tag{5.7}$$

where a is the parameter that controls the shape of the force function, δ_i is the cell size in the subpopulation i, $i = 1, 2, 3$, and $r = |x-y|$. We fix the value of the shape parameter at $a = 0.08\,\mu\mathrm{m}^{-1}$ (Matsiaka et al., 2017; Chapter 2). The function $g_i(r) = \Big(1 - \sin\Big[(2\pi r - \pi\delta_i)/2\delta_i\Big]\Big)/2$ 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 (5.7) for different values of the parameter a is shown in Figure 5.2(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 the heterogeneous three-species model are: (i) the cell sizes, δ_1, δ_2 and δ_3; (ii) the cell diffusivities, D_1, D_2 and D_3; and (iii) the amplitudes of interaction forces, $f_0^{(1)}$, $f_0^{(2)}$ and $f_0^{(3)}$. In this chapter 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 relatively simple homogeneous, single-species models. Since our experimental data in Figure 5.1 allows us to explicitly characterise the heterogeneity

in cell size, all results in Chapter 5 focus on cell size. Additional results in Chapter 5A focus on heterogeneity in diffusivity and amplitude of interaction forces to provide additional insight into the role of heterogeneity in these kinds of experiments.

We define the total density of the heterogeneous population as

$$\mathcal{P}(x,t) = \sum_{i=1}^{3} \left[p^{(i)}(x,t) \right], \tag{5.8}$$

where $p^{(i)}(x,t)$ is the cell density of subpopulation $i = 1,2,3$ predicted by Equations (5.2)-(5.4), and $\mathcal{P}(x,t)$ is the total cell density. It is important to interpret the solutions of Equations (5.2)-(5.4) 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 (5.2)-(5.4), to obtain a single-species homogeneous model in the form,

$$\frac{\partial P(x,t)}{\partial t} = \bar{D}\Delta P(x,t) - (N-1)\nabla\Big(P(x,t)\,V(x,t)\Big), \tag{5.9}$$

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 $\delta_i = \bar{\delta}$, $D_i = \bar{D}$, and $f_i^{(i)} = \bar{f}_0$ for $i = 1,2,3$. The key differences between the homogeneous single-species model, Equation (5.9), and the three-species heterogeneous model, Equations (5.2)-(5.4) are: (i) the three-species heterogeneous model incorporates 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 5.1(a). As such, we adopt an initial cell distribution in the form of the one-dimensional step function,

$$P(x,0) = \mathcal{P}(x,0) = \begin{cases} 23.9 \times 10^{-3}\ \text{cells}/\mu\text{m}, & 0\,\mu\text{m} < x < 1000\,\mu\text{m}, \\ 0\ \text{cells}/\mu\text{m}, & 1000\,\mu\text{m} < x < 2000\,\mu\text{m}, \end{cases} \tag{5.10}$$

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 Figures 5.3(a)-5.6(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; Chapter 2). We note that the boundary of the experimental image in Figure 5.1(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 $x = 0\,\mu\text{m}$ and $x = 2000\,\mu\text{m}$.

All continuum results for single-species homogeneous and three-species heterogeneous models, given by Equation (5.9) and Equations (5.2)-(5.4), respectively, are solved numerically using the method of lines with $\Delta x = 4\,\mu\text{m}$ and $\Delta t = 0.005\,\text{h}$ on $0\,\mu\text{m} < x < 2000\,\mu\text{m}$ domain (Matsiaka et al., 2017; Chapter 2). The spatial and temporal discretisations are chosen sufficiently fine to produce grid independent results. The detailed discretisation scheme used in this chapter is presented in the Chapter 5A.

5.5 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, $\bar{\delta}$, while keeping $\bar{D}$ and $\bar{f}_0$ fixed at $\bar{D} = 250\,\mu\text{m}^2/\text{h}$ and $\bar{f}_0 = 1.0\,\mu\text{m/h}$. The values of D_i and $f_0^{(i)}$ in the heterogeneous three-species model are fixed at $D_i = 250\,\mu\text{m}^2/\text{h}$ and $f_0^{(i)} = 1.0\,\mu\text{m/h}$ for $i = 1, 2, 3$. These values of diffusivity and amplitude of cell-to-cell interaction forces are based on detailed experimental measurements reported previously (Matsiaka et al., 2019; Chapter 4).

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; Kollimada et al., 2016; 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, $E(\bar{\delta})$, associated

with the position of the leading edge,

$$E(\bar{\delta}) = \frac{1}{\alpha} \sum_{j} \left[\mathcal{S}(t_j) - S(t_j) \right]^2, \tag{5.11}$$

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 $\alpha = 49$ is the number of discrete time points we use to compute $E(\bar{\delta})$. 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). An alternative approach is to use an error measure based on the discrepancy between cell density profiles. At first, this approach of using the entire cell density profile might be thought to be preferable to working with leading edge data since density profiles incorporate much more detailed spatial information than just using the position of the leading edge. However, extracting the density data from experiments is much more tedious because it often involves manual cell counting in regions where cell densities are high and this is both difficult to reproduce and very time consuming (Treloar et al. 2014). Therefore, to keep our work as practical as possible, here we present only results with an error measure solely based on the leading edge data. Additional result that measure the discrepancy between the models using the entire density information are presented in Chapter 5A (Figure 5A.1 and Figure 5A.2), and we find that this more complicated approach gives very similar results to the leading edge data. Therefore, in this chapter, we focus on the using leasing edge data.

The experimental distribution of cell sizes in Figure 5.1(d) provides insights into potential choices of the cell size distribution in Equations (5.2)-(5.4). Here we define three subpopulations based on the equivalent cell size: small ($\delta_1 = 18$ μm), medium ($\delta_2 = 34$ μm), and large cells ($\delta_3 = 50$ 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 5.3). After considering the experimentally-motivated monotonically decreasing distribution, we then systematically explore: (i) uniform (Set Ib, Figure 5.4), (ii) non-monotonic (Set Ic, Figure 5.5), and (iii) monotonically increasing distributions (Set Id, Figure 5.6).

Figure 5.3(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, $E(\bar{\delta})$, demonstrates a clear minimum which ensures the unique choice of a best-fit cell size, $\bar{\delta}$. Results in Figure 5.3(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 5.3(d) (with additional details at the leading edge shown in the magnified

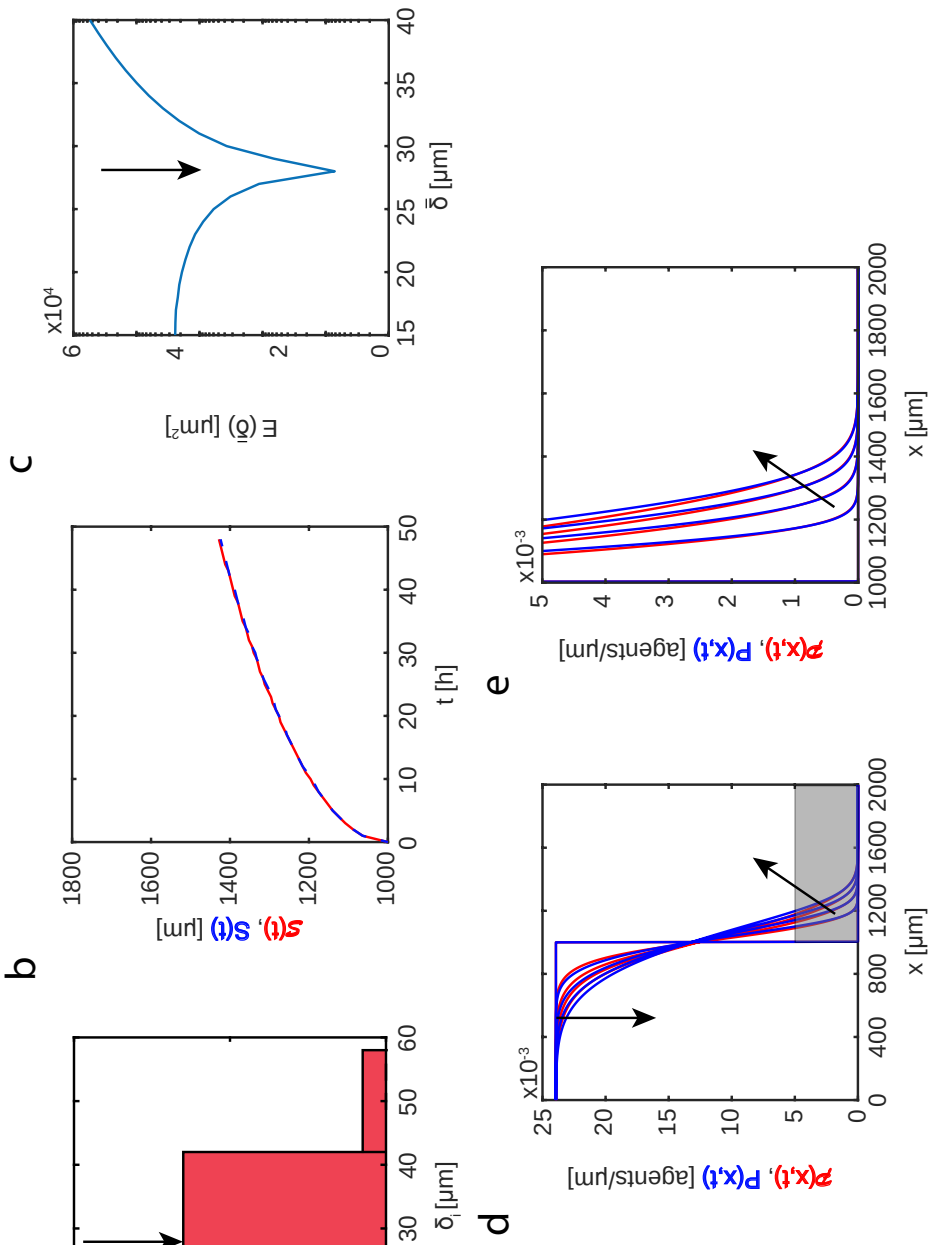


Figure 5.3: Set Ia. Heterogeneity in cell sizes: monotonically decreasing distribution. (a) Cell size distribution adopted in the three-species heterogeneous model, Equations (5.2)-(5.4). Here the proportions of cells of different sizes are set to: (i) $n_1/N = 0.472$; (ii) $n_2/N = 0.472$; (iii) $n_3/N = 0.056$. (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, $E(\bar{\delta})$, between the position of the leading edge given by the three-species heterogeneous model and the position predicted by the single-species homogeneous model as a function of cell size, $\bar{\delta}$. The black arrow denotes the best-fit value of cell size, $\bar{\delta} = 28\,\mu$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).

region in Figure 5.3(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 5.3 - 5.6 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 5.3(b)-5.6(b)). In contrast, the quality of match between 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 motivated distribution in Figure 5.3(a) (Set Ia) leads to a remarkably good match between the three-species heterogeneous model and the single-species homogeneous model. Similarly, the uniform distribution shown in Figure 5.4(a) (Set Ib) also leads to a reasonably good quality of match between two different models. In contrast, the density profiles associated with the non-monotonic cell size distribution (Figure 5.5, Set Ic) and monotonically increasing cell size distribution (Figure 5.6, Set Id) 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, $\bar{\delta}$, 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 5.4, Set Ib), $\bar{\delta} = 36\,\mu\text{m}$, is quite close to the weighted average value of $34\,\mu\text{m}$ for the distribution in Figure 5.4(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 (Set Ia) and monotonically increasing (Set Id) cell size distributions, shown in Figure 5.3 and Figure 5.6, respectively. In contrast, the best-fit value of the cell size for the non-monotonic distribution (Set Ic), $\bar{\delta} = 40\,\mu\text{m}$, differs significantly from the weighted average of $34\,\mu\text{m}$. Therefore, these results suggest that great care ought to be exercised when taking a distribution of parameter values and attempting to select the most appropriate single representative value of that parameter.

In addition to the results in Figures 5.3 - 5.6 exploring the role of heterogeneity in cell size, we present an additional suite of results, called Set II and Set III, that systemat-

ically explore the role of heterogeneity in the diffusivity and interaction strength. These additional results are presented in Chapter 5A. 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.

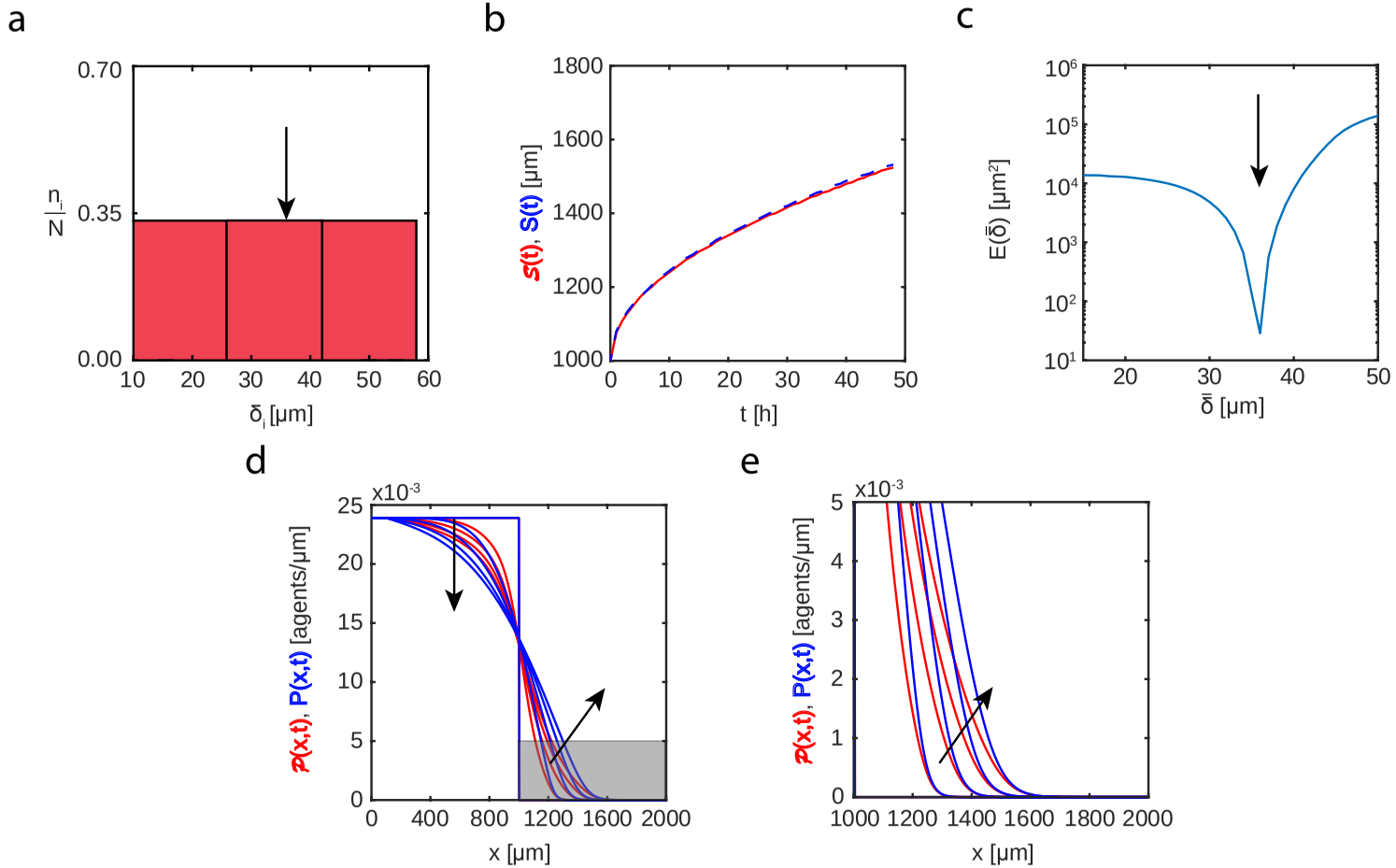


Figure 5.4: Set Ib. Heterogeneity in cell sizes: uniform distribution. (a) Cell size distribution adopted in the three-species heterogeneous model, Equations (5.2)-(5.4). Here the proportions of cells of different sizes are set to: (i) $n_1/N = 0.33(3)$; (ii) $n_2/N = 0.33(3)$; (iii) $n_3/N = 0.33(3)$. (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, $E(\bar{\delta})$, between the position of the leading edge given by the three-species heterogeneous model and the position predicted by the single-species homogeneous model as a function of cell size, $\bar{\delta}$. The black arrow denotes the best-fit value of cell size, $\bar{\delta} = 36\,\mu$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 $\bar{\delta}$, $P(x,t)$ (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).

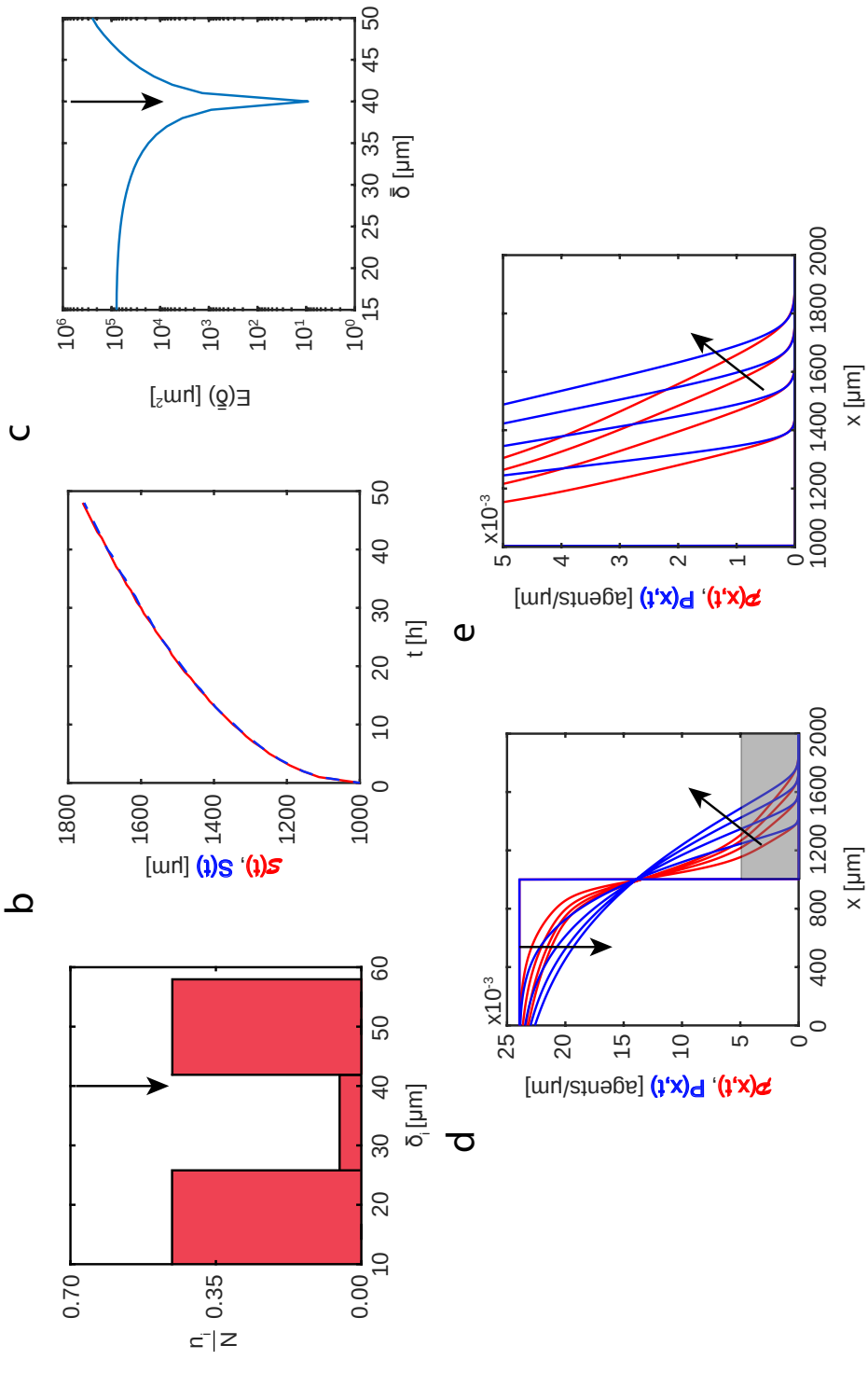


Figure 5.5: Set Ic. Heterogeneity in cell sizes: non-monotonic distribution. (a) Cell size distribution adopted in the three-species heterogeneous model, Equations (5.2)-(5.4). Here the proportions of cells of different sizes are set to: (i) $n_1/N = 0.472$; (ii) $n_2/N = 0.056$; (iii) $n_3/N = 0.472$. (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, $E(\bar{\delta})$, between the position of the leading edge given by the three-species heterogeneous model and the position predicted by the single-species homogeneous model as a function of cell size, $\bar{\delta}$. The black arrow denotes the best-fit value of cell size, $\bar{\delta} = 40\,\mu$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 $\bar{\delta}$, $P(x,t)$ (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).

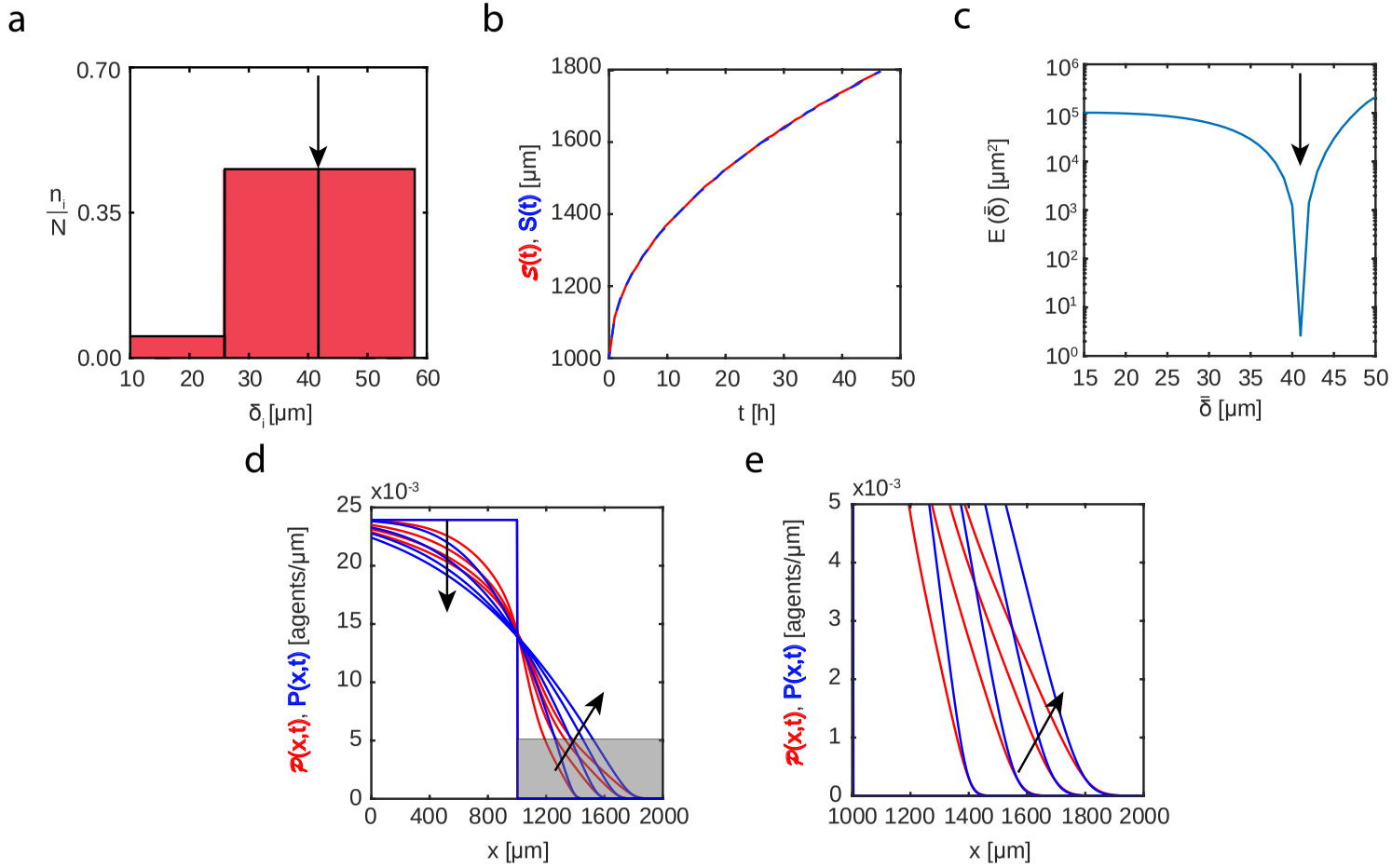


Figure 5.6: Set Id. Heterogeneity in cell sizes: monotonically increasing distribution. (a) Cell size distribution adopted in the three-species heterogeneous model, Equations (5.2)-(5.4). Here the proportions of cells of different sizes are set to: (i) $n_1/N = 0.056$; (ii) $n_2/N = 0.472$; (iii) $n_3/N = 0.472$. (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, $E(\bar{\delta})$, between the position of the leading edge given by the three-species heterogeneous model and the position predicted by the single-species homogeneous model as a function of cell size, $\bar{\delta}$. The black arrow denotes the best-fit value of cell size, $\bar{\delta} = 41\,\mu$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 $\bar{\delta}$, $P(x,t)$ (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).

5.6 Conclusions

In this chapter, 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 5.1. 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 different 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 5.3(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 5.1. Similarly, the homogeneous distribution, Figure 5.4, shows that single-species homogeneous model is able to accurately replicate the three-species heterogeneous model results. This is an expected result because in this special case the cells of each subpopulation are the same size. In contrast, the single-species homogeneous model does not perform so well when applied to both non-monotonic and monotonically increasing distributions in Figures 5.5-5.6, respectively. Additionally we explore potential heterogeneity in diffusivity and amplitude of the cell-to-cell interactions (Chapter 5A). 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 chapter 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 chapter could be extended by considering

additional subpopulations. Another simplification that we invoke is to assume that the measured heterogeneity remains constant for the duration of the experiment. Future studies could address the significantly more complicated question of allowing the distributions to evolve in time on the same time scale as the experiment to see if it is still possible to use a simpler homogeneous model in this more complicated scenario. Another avenue for further exploration would be to consider heterogeneity in more than one parameter at a time, whereas in this chapter 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.

Chapter 5A

Additional results for Chapter 5

5A.1 Heterogeneity in the cell size assessed using density profile data

In this section we repeat the analysis contained in Figure 5.3 and Figure 5.5 (Chapter 5), except that we use a different measure of the discrepancy, that is instead of using the leading edge data, we use a measure that is based on the entire cell density profile. In this analysis we keep D_i and $f_0^{(i)}$ fixed at $D_i = 250\,\mu\text{m}^2/\text{h}$ and $f_0^{(i)} = 1.0\,\mu\text{m/h}$ for $i = 1, 2, 3$, respectively (Set I, Chapter 5). The main difference is that here we define the error measure, $\mathbb{E}(\bar{\delta})$, as the mean square difference between the density profiles given by the three-species heterogeneous model and profiles predicted the single-species homogeneous model,

$$\mathbb{E}(\bar{\delta}) = \frac{1}{\alpha \mathcal{N}} \sum_i \sum_j \left[\mathcal{P}(x_i, t_j) - P(x_i, t_j)\right]^2, \tag{5A.1}$$

where $\mathcal{P}(x_i, t_j)$ is the total agent density given by the three-species heterogeneous model, $P(x_i, t_j)$ is the agent density predicted by the single-species homogeneous model, $\alpha = 49$ is the number of discrete time points that we used to compute $\mathbb{E}(\delta)$, and $\mathcal{N} = 500$ is the number of discrete spatial points in the discretisation scheme.

Results in Figures 5A.1-5A.2 are analogous to those results in Figure 5.3 and Figure 5.5 (Chapter 5). The best-fit value of the cell size, $\bar{\delta} = 26\,\mu\text{m}$, is relatively close to a best-fit estimate of $\bar{\delta} = 28\,\mu\text{m}$ reported in Chapter 5 using the error measure, $E(\bar{\delta})$, based on the position of the leading edge (compare Figure 5A.1(b) and Figure 5.3(c) in Chapter 5). Here we find that the single species homogeneous model can be used to accurately describe the three-species heterogeneous data for both error measures, and this is obvious when we visually compare Figure 5A.1(c) and Figure 5.3(d) in Chapter 5. In contrast, the density profiles associated with the best-fit homogeneous model in Figure 5A.2(c) do not provide an accurate match the three-species heterogeneous data

generated. This means that the heterogeneous density profiles associated with the non-monotonic distribution cannot be faithfully replicated using a simpler homogeneous model. We note that, for all cases presented, the discrepancy between the three-species and single-species density profiles significantly increases with time, and this might be an important factor to take into account when choosing appropriate model. In summary, the conclusions we draw in Chapter 5 about the suitability of the homogeneous model to accurately describe results generated using the heterogeneous model are the same regardless of whether we use leading edge data or the entire cell density profile to measure the discrepancy between the two modelling frameworks.

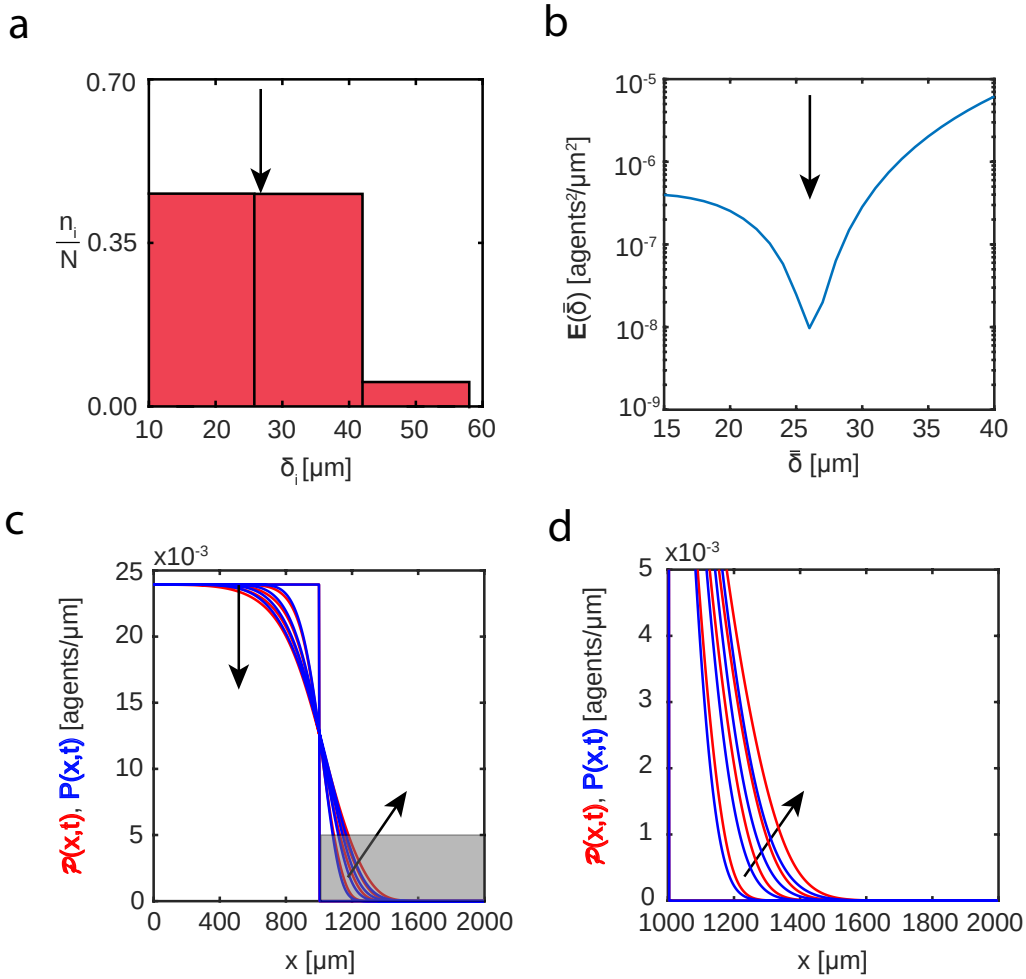


Figure 5A.1: Heterogeneity in cell sizes: monotonically decreasing distribution. (a) Cell size distribution adopted in the three-species heterogeneous model, Equations (5.2)-(5.4), (Chapter 5). Here the proportions of cells of different sizes are set to: (i) $n_1/N = 0.472$; (ii) $n_2/N = 0.472$; (iii) $n_3/N = 0.056$. (b) Error measure, $\mathbb{E}(\bar{\delta})$, between the cell density profiles, $\mathcal{P}(x,t)$, given by the three-species heterogeneous model and density profiles, $P(x,t)$, predicted by the single-species homogeneous model as a function of cell size, $\bar{\delta}$. The black arrow denotes the best-fit value of cell size, $\bar{\delta} = 26\,\mu$m. (c)-(d) 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 $\bar{\delta}$, $P(x,t)$ (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 (d) show a close-up comparison right near the leading edge, denoted by the gray shaded region in (c).

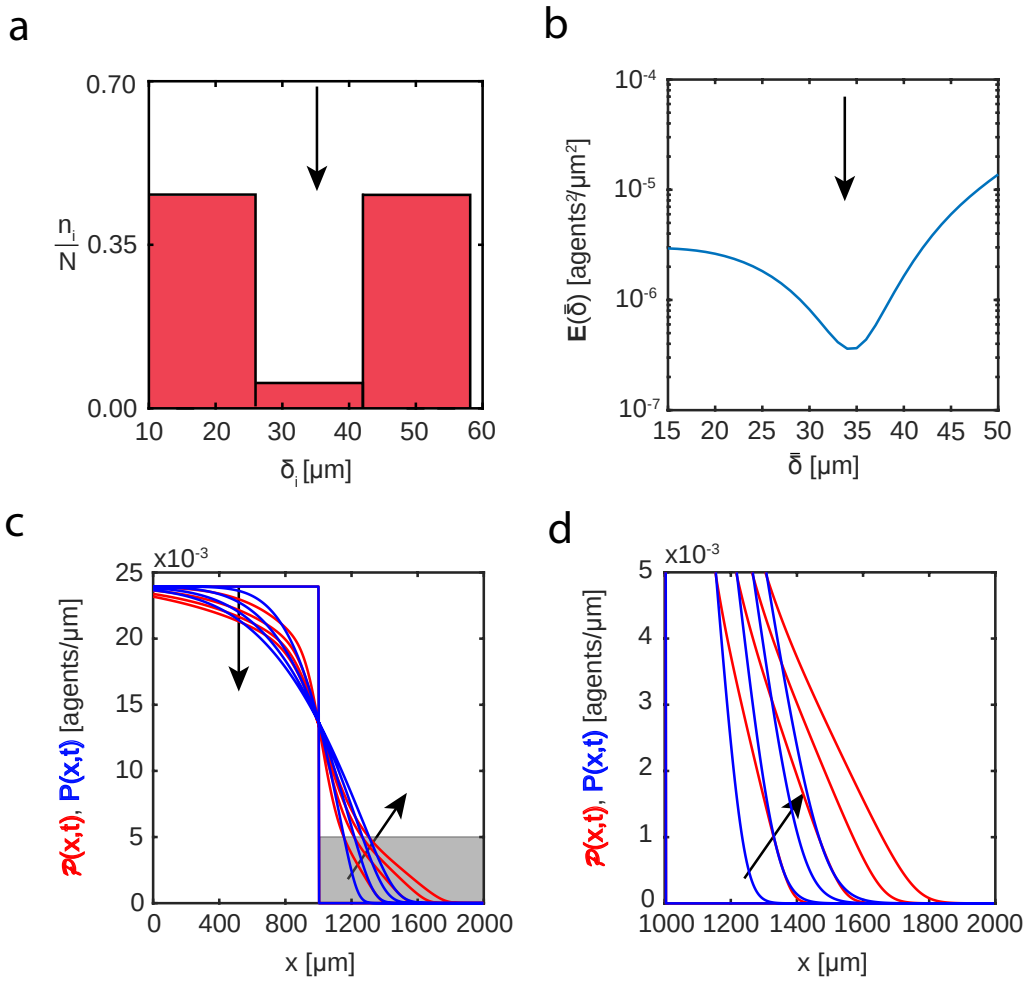


Figure 5A.2: Heterogeneity in cell sizes: non-monotonic distribution. (a) Cell size distribution adopted in the three-species heterogeneous model, Equations (5.2)-(5.4), (Chapter 5). Here the proportions of cells of different sizes are set to: (i) $n_1/N = 0.472$; (ii) $n_2/N = 0.056$; (iii) $n_3/N = 0.472$. (b) Error measure, $\mathbb{E}(\bar{\delta})$, between the cell density profiles, $\mathcal{P}(x,t)$, given by the three-species heterogeneous model and density profiles, $P(x,t)$, predicted by the single-species homogeneous model as a function of cell size, $\bar{\delta}$. The black arrow denotes the best-fit value of cell size, $\bar{\delta} = 34\,\mu$m. (c)-(d) 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 $\bar{\delta}$, $P(x,t)$ (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 (d) show a close-up comparison right near the leading edge, denoted by the gray shaded region in (c).

5A.2 Heterogeneity in the interaction force, f_0

In this data set we explore the heterogeneity in the interaction forces where we fix values of the diffusivity, $D_i = 250\,\mu\text{m}^2/\text{h}$, and the cell size, $\delta_i = 34\,\mu\text{m}$ for $i = 1, 2, 3$. We note that these estimates are typical parameter values for PC-3 cells (Matsiaka et al., 2019; Chapter 4). To analyse performance of the single-species homogeneous model (Equation (5.9)) applied to data generated by the three-species heterogeneous model (Equations (5.2)-(5.4)) we consider four interaction force distributions: (i) uniform distribution, Figure 5A.3(a), (ii) monotonically decreasing distribution, Figure 5A.4(a), (iii) non-monotonic distribution, Figure 5A.5(a), and (iv) monotonically increasing distribution, Figure 5A.6(a). For all cases presented we are able to predict a position of the leading edge as well as accurately describe cell density profiles given by the three-species heterogeneous model.

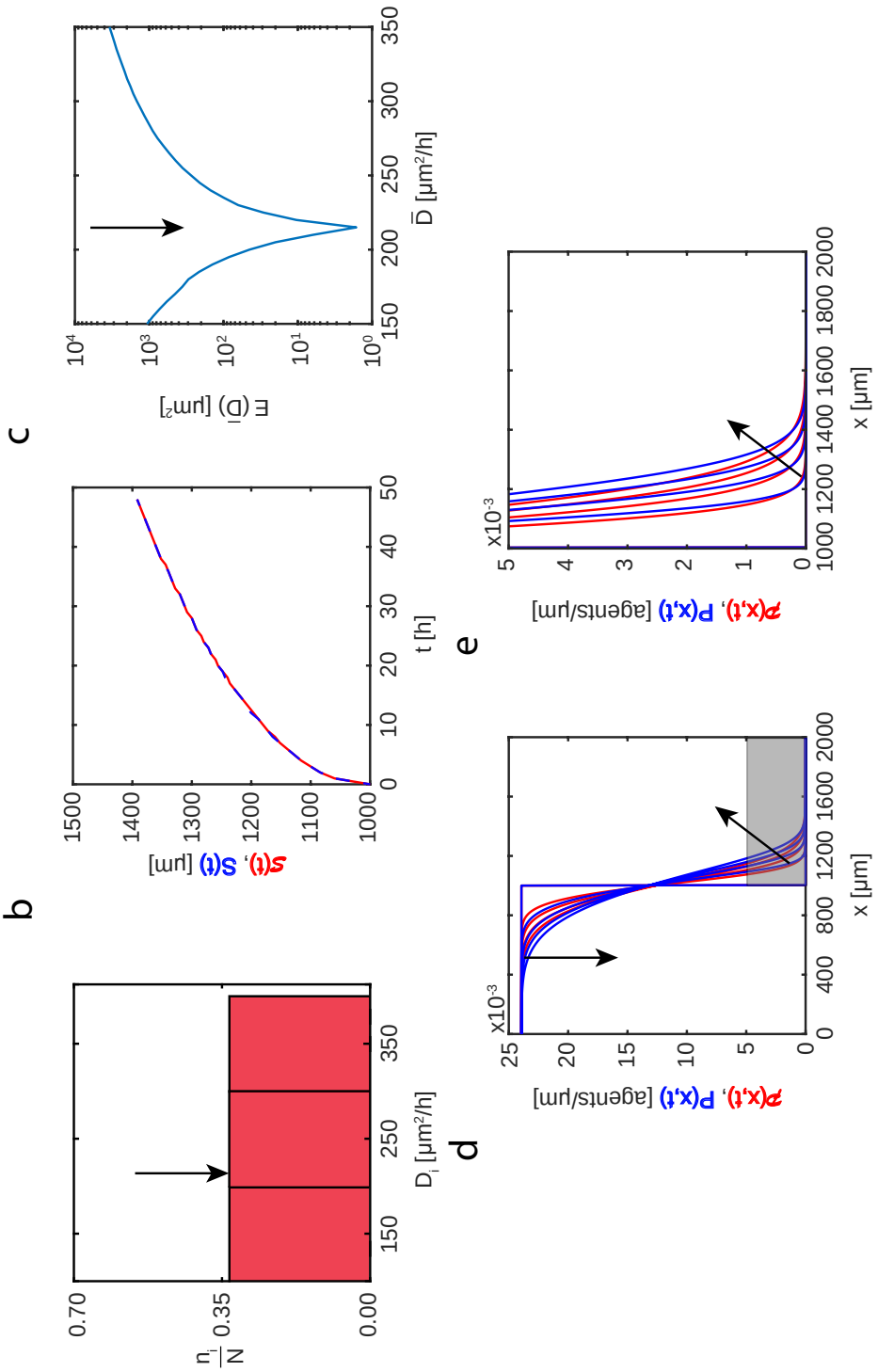


Figure 5A.3: Set IIa. Heterogeneity in interaction forces: uniform distribution. (a): Interaction force distribution adopted in the three-species heterogeneous model, Equations (5.2)-(5.4). Here the proportions of cells of different sizes are set to: (i) $n_1/N = 0.333$; (ii) $n_2/N = 0.333$; (iii) $n_3/N = 0.333$. (b): Leading edge predicted by the three-species heterogeneous model, $S(t)$ (solid red), and the best-fit approximation given by the single-species homogeneous model, $S(t)$ (blue dashed). (c): Error measure, $E(\bar{f}_0)$, between the position of the leading edge given by the three-species heterogeneous model and the position predicted by the single-species homogeneous model as a function of amplitude of the interaction force, $\bar{f}_0$. (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 $\bar{f}_0$, $P(x,t)$ (solid blue). The black arrow denotes the best-fit value of interaction force, $\bar{f}_0 = 0.055\,\mu m/h$. 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).

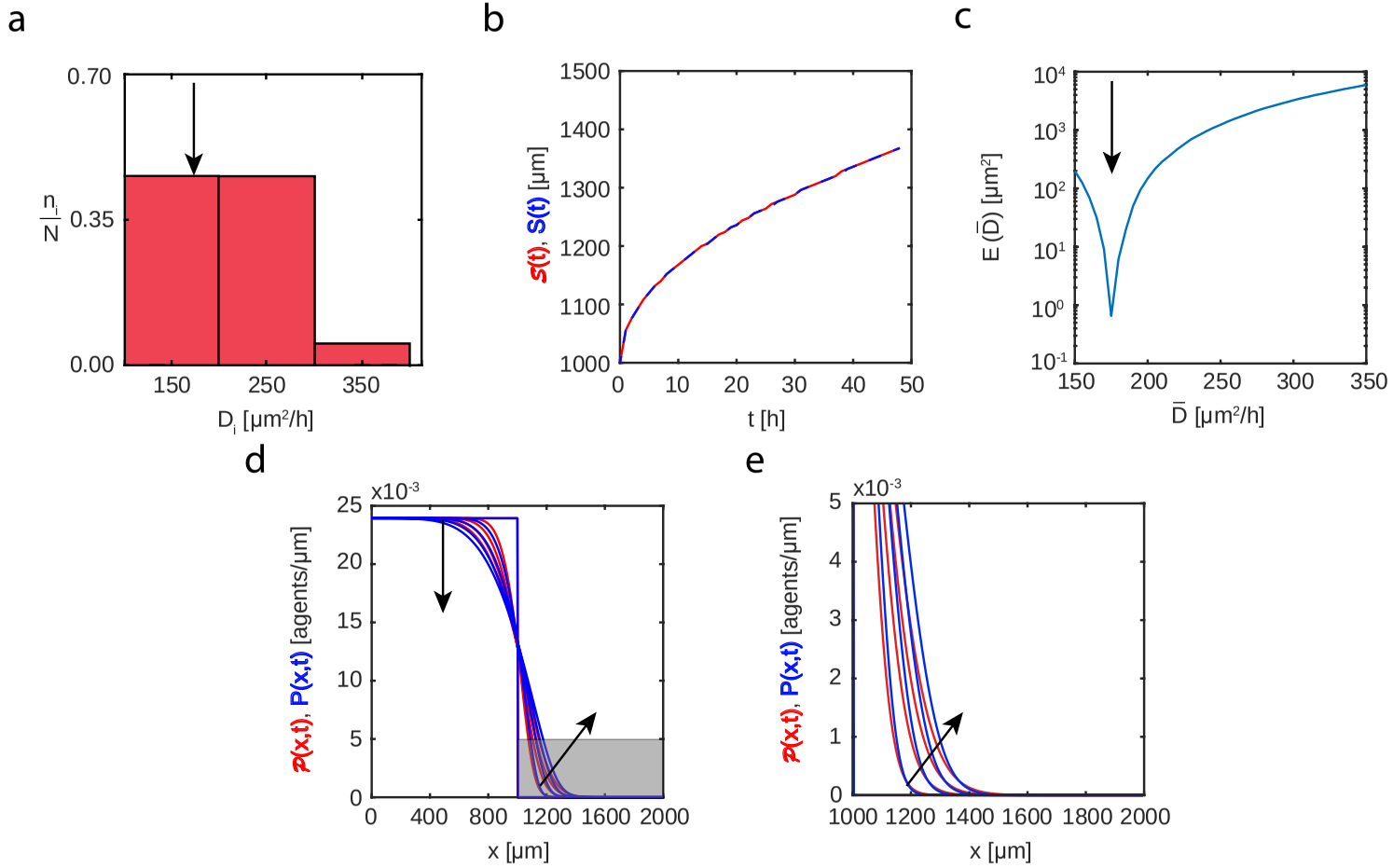


Figure 5A.4: Set IIb. Heterogeneity in interaction forces: monotonically decreasing distribution. (a): Interaction force distribution adopted in the three-species heterogeneous model, Equations (5.2)-(5.4).Here the proportions of cells of different sizes are set to: (i) $n_1/N = 0.472$; (ii) $n_2/N = 0.472$; (iii) $n_3/N = 0.056$. (b): Leading edge 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, $E(\bar{f}_0)$, between the position of the leading edge given by the three-species heterogeneous model and the position predicted by the single-species homogeneous model as a function of amplitude of the interaction force, $\bar{f}_0$. (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 $\bar{f}_0$, $P(x,t)$ (solid blue). The black arrow denotes the best-fit value of interaction force, $\bar{f}_0 = 0.06\,\mu$m/h. 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).

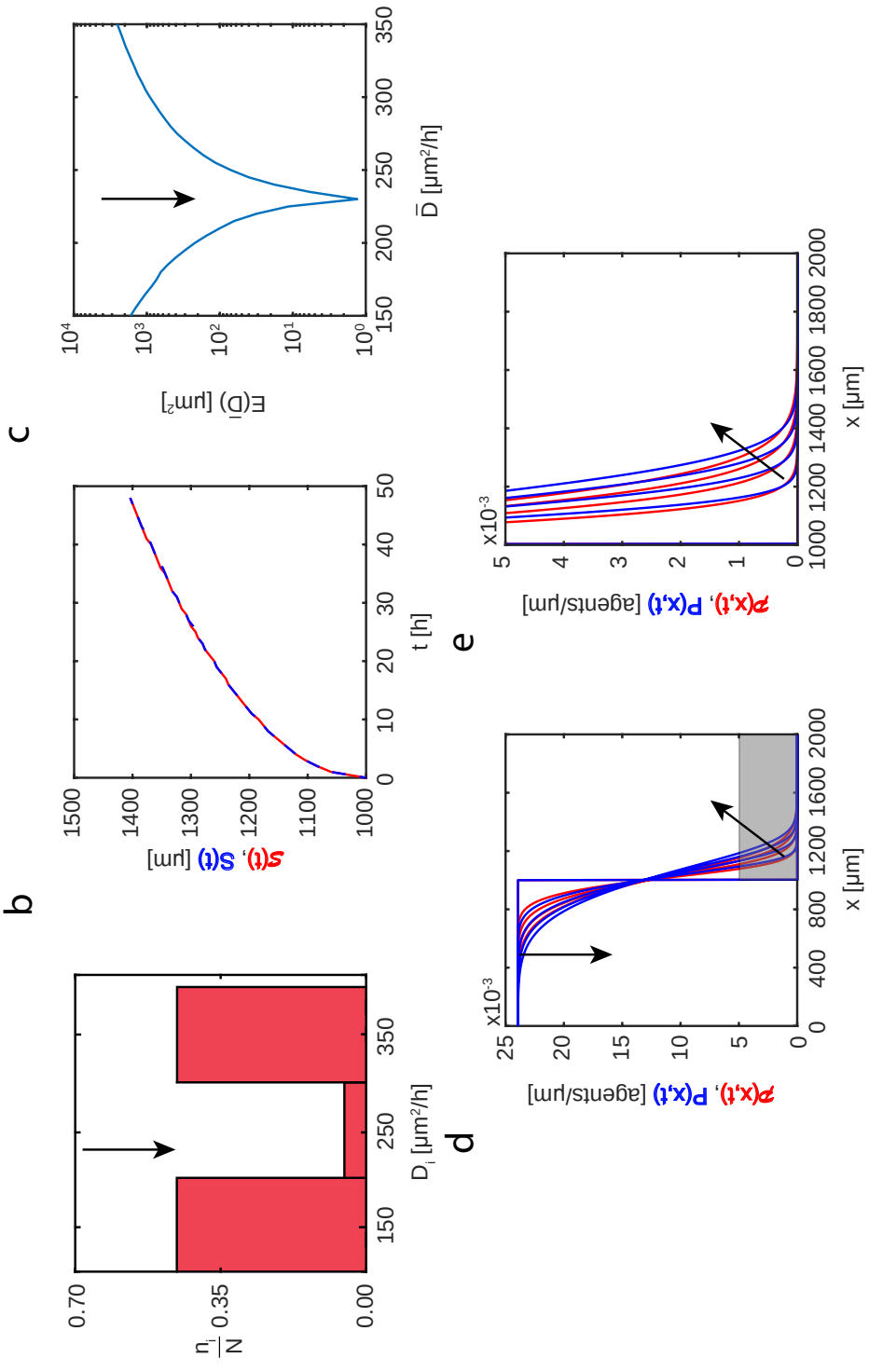


Figure 5A.5: Set IIc. Heterogeneity in interaction forces: non-monotonic distribution. (a): Interaction force distribution adopted in the three-species heterogeneous model, Equations (5.2)-(5.4). Here the proportions of cells of different sizes are set to: (i) $n_1/N = 0.472$; (ii) $n_2/N = 0.056$; (iii) $n_3/N = 0.472$. (b): Leading edge predicted by the three-species heterogeneous model, $S(t)$ (solid red), and the best-fit approximation given by the single-species homogeneous model, $S(t)$ (blue dashed). (c): Error measure, $E(\bar{f}_0)$, between the position of the leading edge given by the three-species heterogeneous model and the position predicted by the single-species homogeneous model as a function of amplitude of the interaction force, $\bar{f}_0$. (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 $\bar{f}_0$, $P(x,t)$ (solid blue). The black arrow denotes the best-fit value of interaction force, $\bar{f}_0 = 0.065\,\mu m/h$. 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).

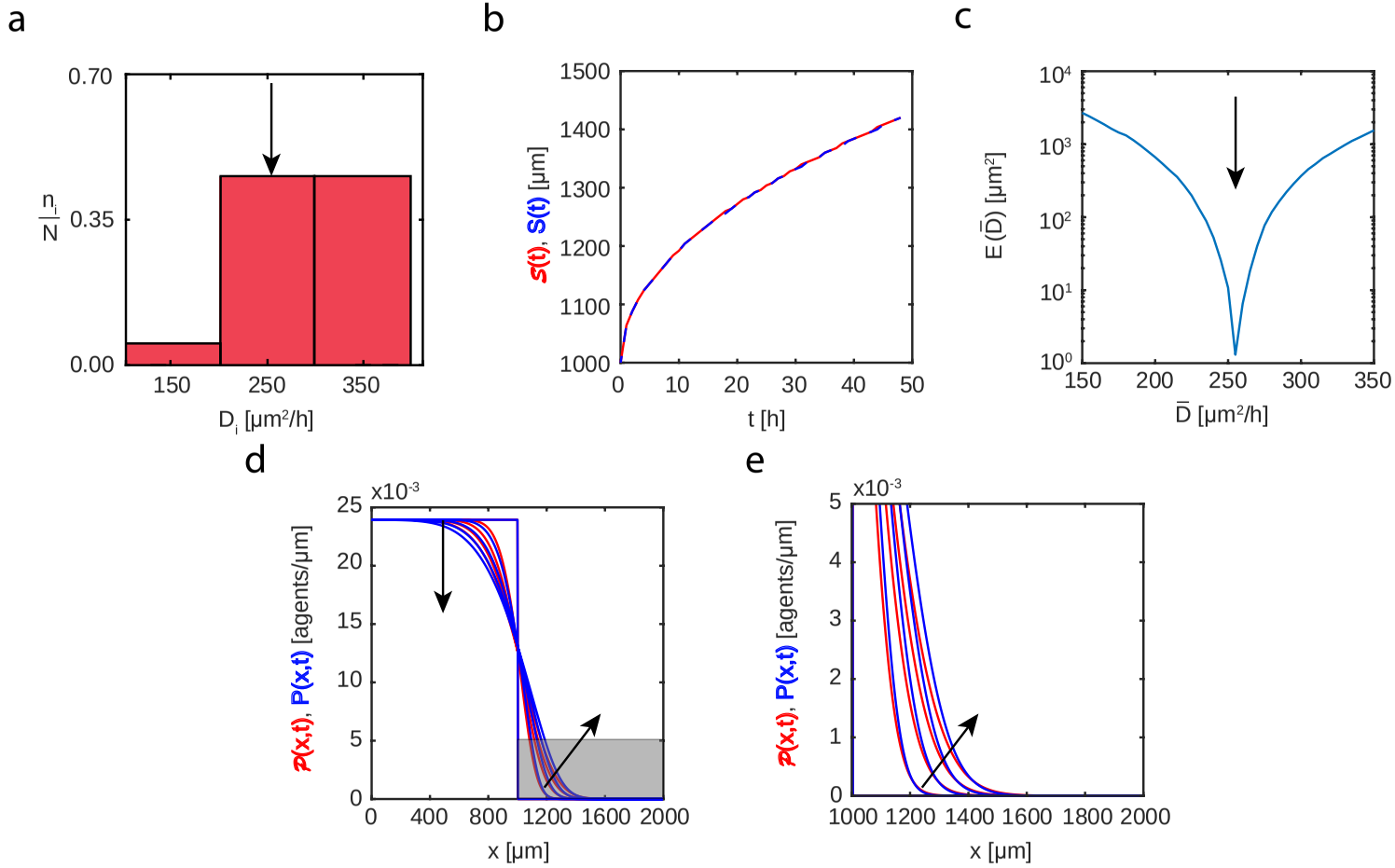


Figure 5A.6: Set IId. Heterogeneity in interaction forces: monotonically increasing distribution. (a): Interaction force distribution adopted in the three-species heterogeneous model, Equations (5.2)-(5.4). Here the proportions of cells of different sizes are set to: (i) $n_1/N = 0.056$; (ii) $n_2/N = 0.472$; (iii) $n_3/N = 0.472$. (b): Leading edge 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, $E(\bar{f}_0)$, between the position of the leading edge given by the three-species heterogeneous model and the position predicted by the single-species homogeneous model as a function of amplitude of the interaction force, $\bar{f}_0$. (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 $\bar{f}_0$, $P(x,t)$ (solid blue). The black arrow denotes the best-fit value of interaction force, $\bar{f}_0 = 0.07\,\mu$m/h. 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).

5A.3 Heterogeneity in the diffusivity, D

In this data set we explore the heterogeneity in the diffusivity where we fix values of the amplitude of the interaction force, $f_0^{(i)} = 0.05\,\mu$m/h, and the cell size, $\delta_i = 34\,\mu$m for $i = 1, 2, 3$. We note that these estimates are typical parameter values for PC-3 cells (Matsiaka et al., 2019; Chapter 4). To analyse performance of the single-species homogeneous model (Equation (5.9)) applied to data generated by the three-species heterogeneous model (Equations (5.2)-(5.4)) we consider four different diffusivity distributions: (i) uniform distribution, Figure 5A.7(a), (ii) monotonically decreasing distribution, Figure 5A.8(a), (iii) non-monotonic distribution, 5A.9(a), and (iv) monotonically increasing distribution, Figure 5A.10(a). For all cases presented we are able to predict a position of the leading edge as well as accurately describe cell density profiles given by the three-species heterogeneous model.

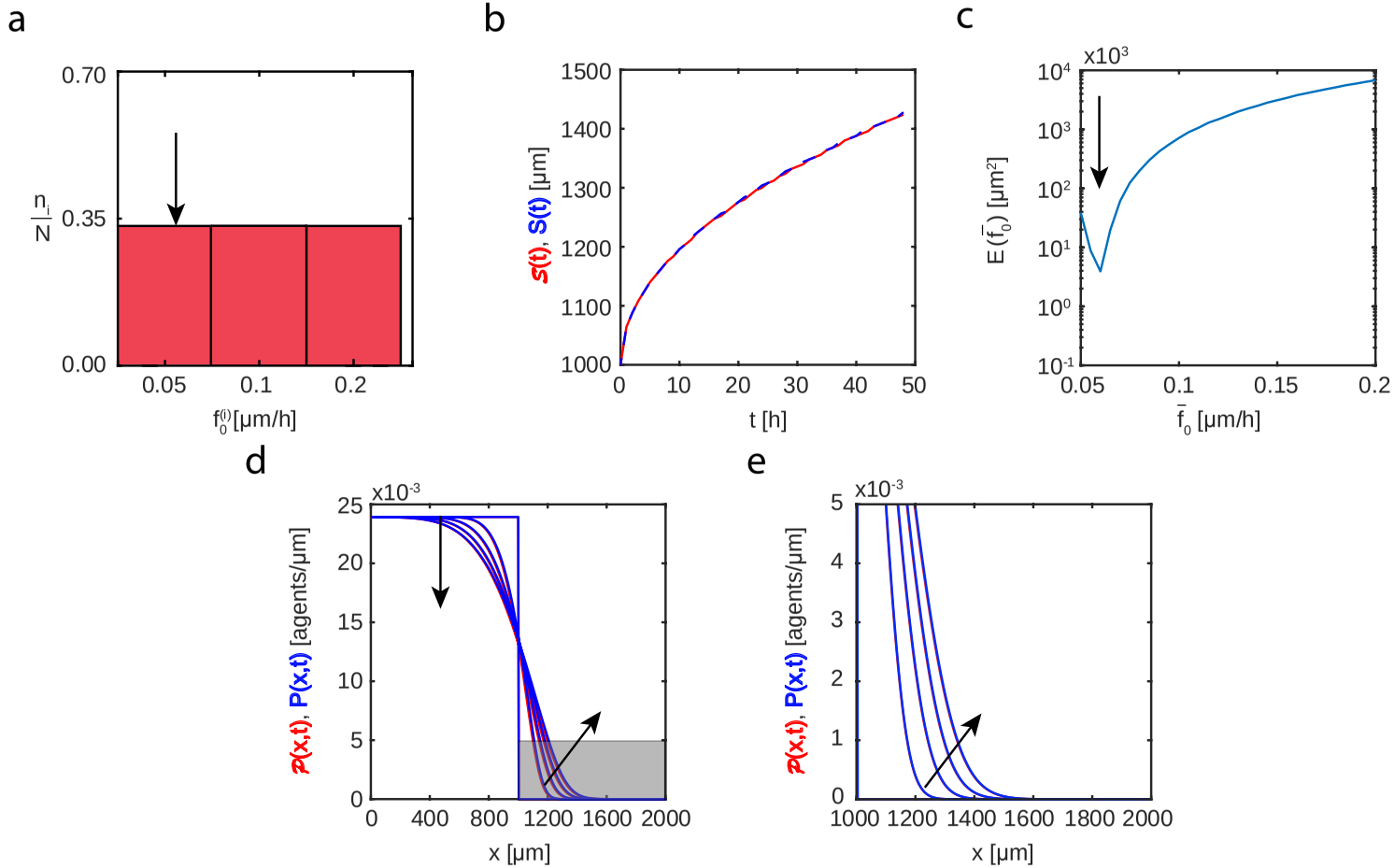


Figure 5A.7: Set IIIa. Heterogeneity in diffusivity: uniform distribution. (a): Diffusivity distribution adopted in the three-species heterogeneous model, Equations (5.2)-(5.4). Here the proportions of cells of different sizes are set to: (i) $n_1/N = 0.333$; (ii) $n_2/N = 0.333$; (iii) $n_3/N = 0.333$. (b): Leading edge 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, $E(\bar{D})$, between the position of the leading edge given by the three-species heterogeneous model and the position predicted by the single-species homogeneous model as a function of diffusivity, $\bar{D}$. The black arrow denotes the best-fit value of diffusivity, $\bar{D} = 215\,\mu\text{m}^2/\text{h}$. (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 $\bar{D}$, $P(x,t)$ (solid blue). The continuum results for both models are presented at $t = 0, 12, 24, 36$, and $48\,\text{h}$. 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).

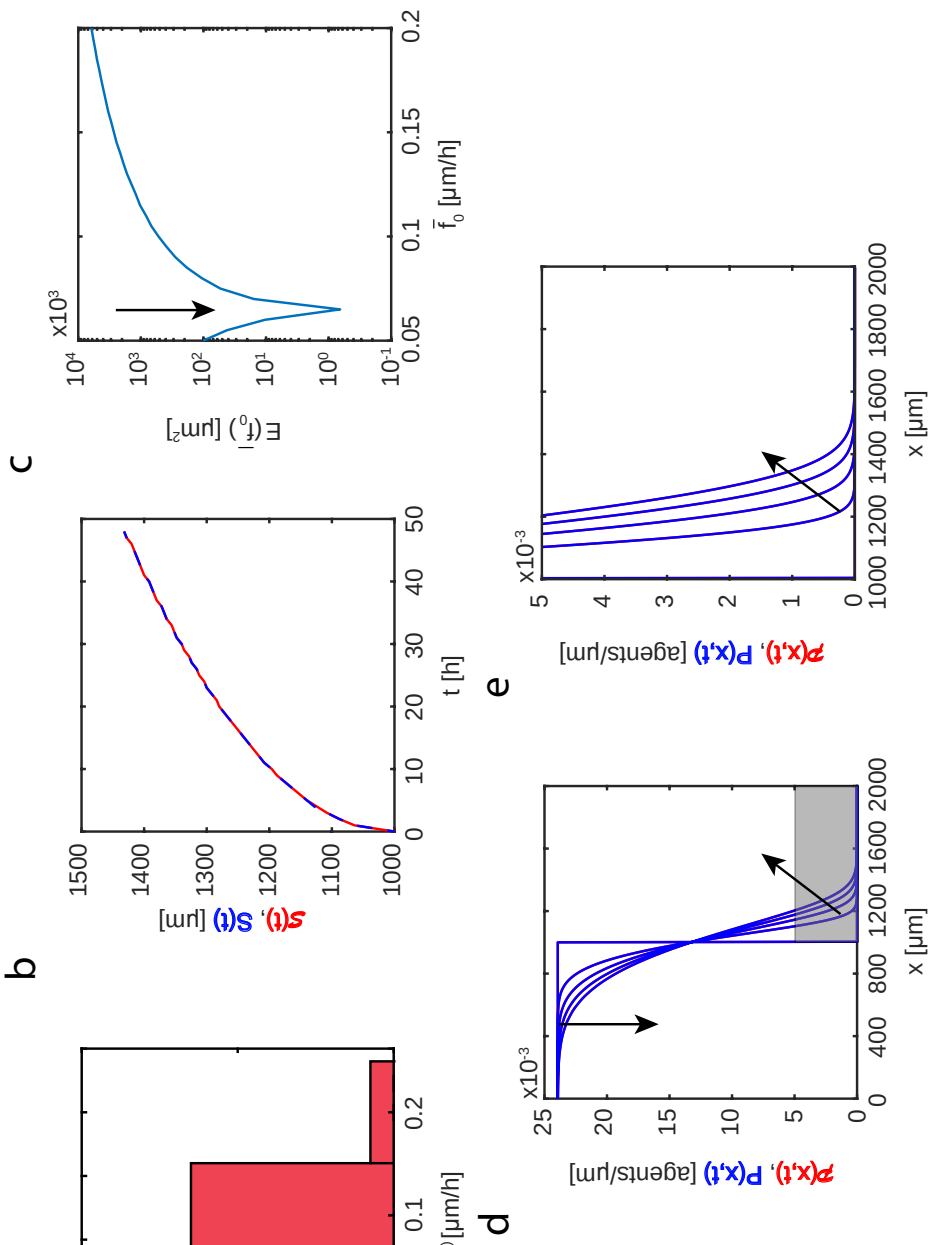


Figure 5A.8: Set IIIb. Heterogeneity in diffusivity: monotonically decreasing distribution. (a): Diffusivity distribution adopted in the three-species heterogeneous model, Equations (5.2)-(5.4). Here the proportions of cells of different sizes are set to: (i) $n_1/N = 0.472$; (ii) $n_2/N = 0.472$; (iii) $n_3/N = 0.056$. (b): Leading edge predicted by the three-species heterogeneous model, $S(t)$ (solid red), and the best-fit approximation given by the single-species homogeneous model, $S(t)$ (blue dashed). (c): Error measure, $E(\bar{D})$, between the position of the leading edge given by the three-species heterogeneous model and the position predicted by the single-species homogeneous model as a function of diffusivity, $\bar{D}$. The black arrow denotes the best-fit value of diffusivity, $\bar{D} = 175\,\mu\text{m}^2/\text{h}$. (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 $\bar{D}$, $P(x,t)$ (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).

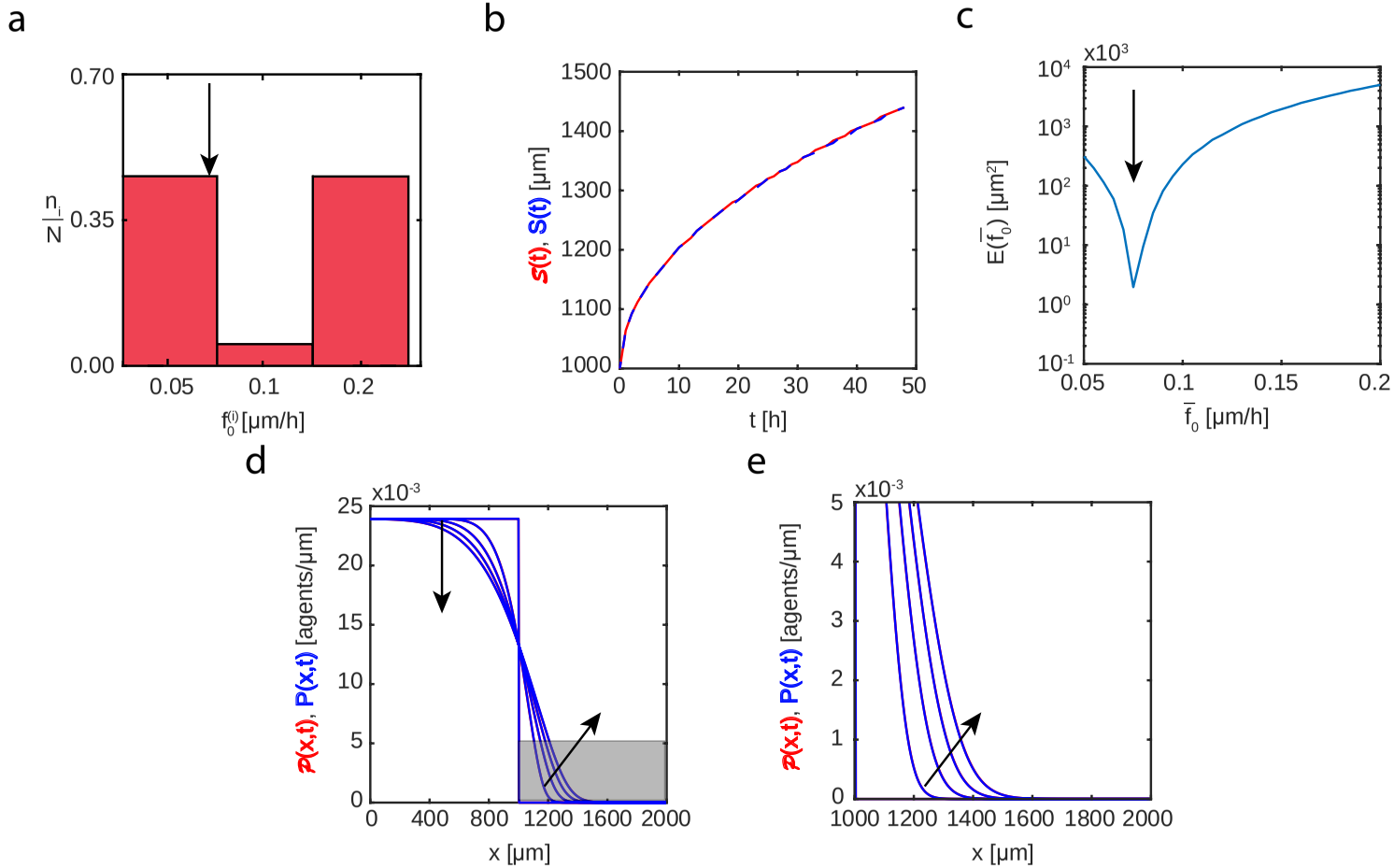


Figure 5A.9: Set IIIc. Heterogeneity in diffusivity: non-monotonic distribution. (a): Diffusivity distribution adopted in the three-species heterogeneous model, Equations (5.2)-(5.4). Here the proportions of cells of different sizes are set to: (i) $n_1/N = 0.472$; (ii) $n_2/N = 0.056$; (iii) $n_3/N = 0.472$. (b): Leading edge 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, $E(\bar{D})$, between the position of the leading edge given by the three-species heterogeneous model and the position predicted by the single-species homogeneous model as a function of diffusivity, $\bar{D}$. The black arrow denotes the best-fit value of diffusivity, $\bar{D} = 230\mu\text{m}^2/\text{h}$. (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 $\bar{D}$, $P(x,t)$ (solid blue). The continuum results for both models are presented at $t = 0, 12, 24, 36$, and $48\,\text{h}$. 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).

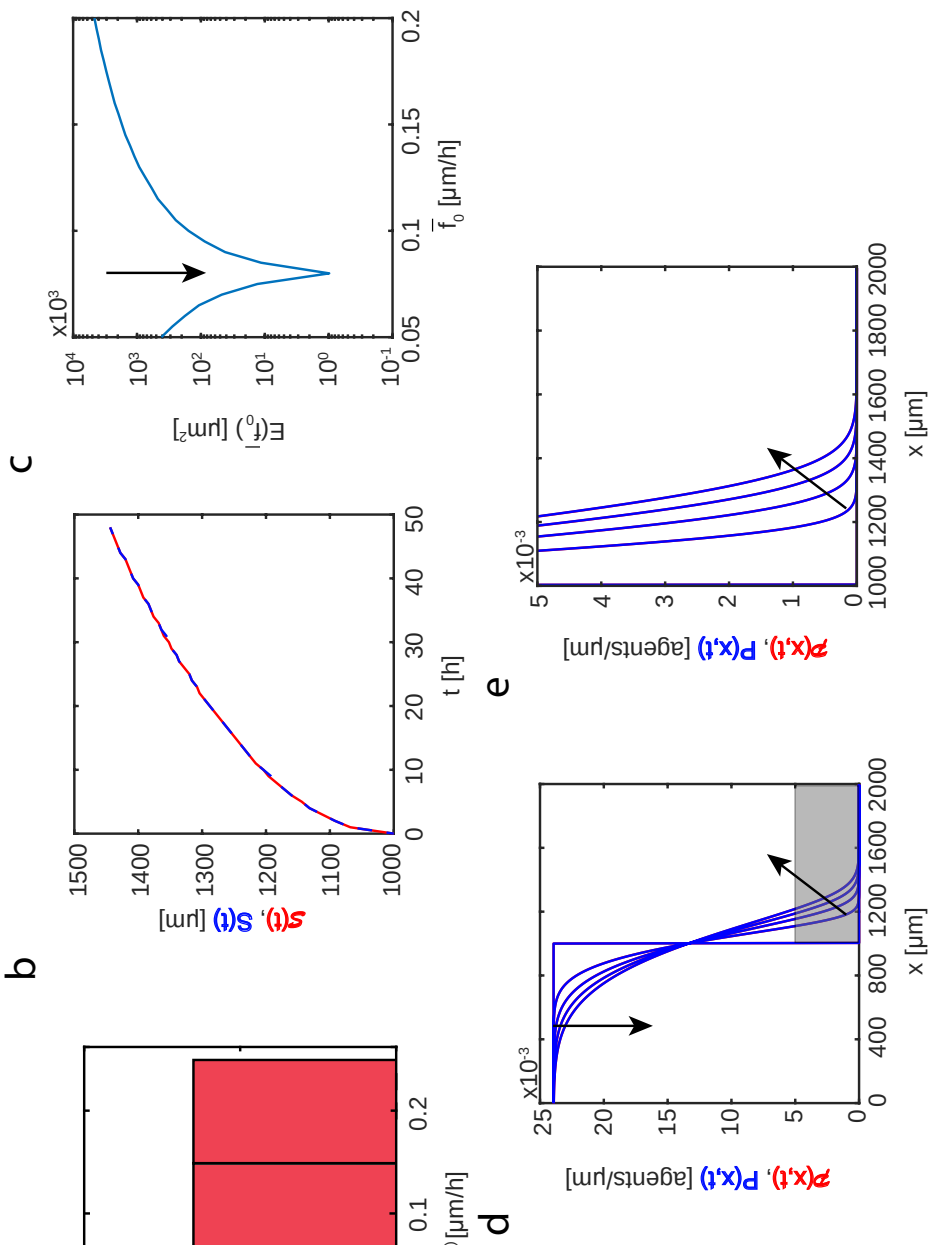


Figure 5A.10: Set IIId. Heterogeneity in diffusivity: monotonically increasing distribution. (a): Diffusivity distribution adopted in the three-species heterogeneous model, Equations (5.2)-(5.4). Here the proportions of cells of different sizes are set to: (i) $n_1/N = 0.056$; (ii) $n_2/N = 0.472$; (iii) $n_3/N = 0.472$. (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, $E(\bar{D})$, between the position of the leading edge given by the three-species heterogeneous model and the position predicted by the single-species homogeneous model as a function of diffusivity, $\bar{D}$. The black arrow denotes the best-fit value of diffusivity, $\bar{D} = 255\,\mu\text{m}^2/\text{h}$. (d)-(e): Cell density profiles as 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 $\bar{D}$, $P(x,t)$ (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).

5A.4 Discretisation scheme for the single-species homogeneous model and heterogeneous three-species model

In this section we present the discretisation scheme used to obtain the numerical solution of the single-species homogeneous model in the mean-field framework. In summary, the governing equation that we consider is as follows,

$$\frac{\partial P(x,t)}{\partial t} = D\Delta P(x,t) - (N-1)\nabla(P(x,t)\,V(x,t)), \tag{5A.2}$$

where

$$V(x,t) = \int F(x-y)\,P(y,t)\,\mathrm{d}y \tag{5A.3}$$

is the velocity filed induced by intercellular interaction forces, and $N = 36$ is the total number of cells in the simulations.

To present the numerical scheme as succinctly as possible, we define

$$\sigma(x,y,t) = F(x-y)\,P(y,t), \tag{5A.4}$$

$$\begin{aligned} I_s &= P(x_s,t)\int \sigma(x_s,y)\,\mathrm{d}y \\ &= P(x_s,t)\frac{h}{2}\sum_i \Big[\sigma(x_s,y_{i+1}) + \sigma(x_s,y_i)\Big] + O(h^2), \end{aligned} \tag{5A.5}$$

where the rectangle rule with step h is used for numerical integration, and indices s and i denote the equally-spaced spatial mesh nodes.

Using the definitions in Equations (5A.4)-(5A.5), we apply the method of lines to Equation (5A.2) and obtain the following system of coupled ordinary differential equations,

$$\frac{\mathrm{d}P_i}{\mathrm{d}t} = \frac{D}{h^2}\Big[P_{i+1} - 2P_i + P_{i-1}\Big] - (N-1)\frac{1}{2h}\Big[I_{i+1} - I_{i-1}\Big], \tag{5A.6}$$

where index i denotes a spatial mesh node. This systems of ordinary differential equations is solved using an explicit forward Euler algorithm with constant time steps of duration Δt.

Similarly, the three-species model is given by three coupled integro-PDEs in the following form,

$$\frac{\partial p^{(1)}}{\partial t} = D_1 \Delta p^{(1)} - (n_1 - 1)\nabla(p^{(1)} V^{(11)}) - n_2 \nabla(p^{(1)} V^{(12)}) - n_3 \nabla(p^{(1)} V^{(13)}), \tag{5A.7}$$

$$\frac{\partial p^{(2)}}{\partial t} = D_2 \Delta p^{(2)} - (n_2 - 1)\nabla(p^{(2)} V^{(22)}) - n_1 \nabla(p^{(2)} V^{(21)}) - n_3 \nabla(p^{(2)} V^{(23)}), \tag{5A.8}$$

$$\frac{\partial p^{(3)}}{\partial t} = D_3 \Delta p^{(3)} - (n_3 - 1)\nabla(p^{(3)} V^{(33)}) - n_1 \nabla(p^{(3)} V^{(31)}) - n_2 \nabla(p^{(3)} V^{(32)}), \tag{5A.9}$$

$$V^{(lk)} = \int_\Omega F^{(lk)}(x - y)\, p^{(k)}(y, t)\, \mathrm{d}y. \tag{5A.10}$$

We define

$$\sigma^{lk}(x, y, t) = F^{(lk)}(x - y)\, p^{(k)}(y, t), \tag{5A.11}$$

$$\begin{aligned} I_s^{lk} &= p^{(l)}(x_s, t) \int \sigma^{lk}(x_s, y)\, \mathrm{d}y \\ &= p^{(l)}(x_s, t) \frac{h}{2} \sum_i \Big[\sigma^{lk}(x_s, y_{i+1}) + \sigma^{lk}(x_s, y_i)\Big] + O(h^2), \end{aligned} \tag{5A.12}$$

where $k = 1, 2, 3$ is the subpopulation index, indices i and s denote the equally-spaced spatial mesh nodes, and h is spatial discretisation step.

Using the definitions in Equations (5A.11)-(5A.12), we apply the method of lines to Equations (5A.7)-(5A.9) and obtain the following system of coupled ordinary differential equations,

$$\frac{\mathrm{d}p_i^{(k)}}{\mathrm{d}t} = \frac{D_k}{h^2}\Big[P_{i+1}^{(k)} - 2P_i^{(k)} + P_{i-1}^{(k)}\Big] - (n_k - 1)\frac{1}{2h}\Big[I_{i+1}^{kk} - I_{i-1}^{kk}\Big] - \sum_{l \neq k} n_l \frac{1}{2h}\Big[I_{i+1}^{lk} - I_{i-1}^{lk}\Big], \tag{5A.13}$$

where $k = 1, 2, 3$ is the subpopulation index.

Chapter 6

Conclusions

In this chapter we summarise the results of our work and propose possible extensions.

6.1 Summary of the research

Our study uses combined experimental and modelling approach in the attempt to describe and quantify previously overlooked cell motility mechanisms caused by dynamical changes in cell size and cell-to-cell mechanical interactions. Additionally we explore the role of cell heterogeneity in the model choice for description of complex multi-species populations. In our work we use experimental data in the form of two-dimensional scratch assays containing malignant PC-3 prostate cancer cells. These cells had been pretreated before the experiment with the chemotherapy drug Mytomocyn-C that blocks proliferation and leads to abnormally fast increase in cell size, as shown in Chapter 4. The increase in cell size allows us to explore the role of cell-to-cell pushing and mechanical interactions in collective cell migration and cell front expansion. We develop a multi-species discrete and continuum approach that is able to parametrise and describe heterogeneous cell populations.

The primary objectives of this thesis are to:

1. Develop novel discrete and continuum models of cell motility and adhesion that are able to describe multi-species populations of cells.

2. Extend models developed in Objective 1 to allow for dynamical cell size.

3. Validate new models with a novel experimental data that shows significant increase in cell sizes and quantify the role of cell-to-cell pushing.

4. Explore the ability of a single-species model to describe heterogeneous cell populations.

Chapter 2 and Chapter 3 constitute model development part of the thesis. In Chapter 2 we develop discrete and continuum multi-species models of cell motility and adhesion and test different approximations in order to simplify the continuum model. The continuum model takes a form of the hierarchy of integro-partial differential equations (IPDEs) that requires certain approximations to make it computationally feasible. We consider two different continuum approximations: (i) mean-field approximation, and (ii) moment dynamics approximation in the form of Kirkwood Superposition Approximation (KSA). The mean-field approximation neglects correlations in the cell-to-cell positions and is commonly invoked to describe cell populations. In contrast, the moment dynamics approximation is much more sophisticated and is able to capture correlations in cell positions resulting from cell crowding and volume exclusion effects (Binny et al., 2015). Moment dynamics models are widely represented in mathematical ecology and epidemiology literature for their ability to capture spatio-temporal correlations and clustering in populations (Bolker and Pacala, 1997; Bolker et al., 2003; Keeling, 1997). We explore the accuracy of both approximations as applied to novel multi-species models of cell migration and adhesion in Chapter 2 and investigate different parameter regimes. In Chapter 3 we introduce dynamical cell size into models presented in Chapter 2.

Currently, mean-field approximation remains popular assumption in many models in mathematical biology (Capasso and Frandoli, 2019). The major downside of the mean-field assumption resides in neglecting correlations in cell-to-cell positions that can be especially pronounced in crowded environments and clusters. Recently, there have been attempts to partially incorporate the effects of cell-to-cell correlations into the modified mean-field models of cell migration (Baker and Simpson, 2010; Markham et al., 2013). However, the only reliable method to model correlations is to explicitly include cell-to-cell correlation function in the model. On the other side, moment dynamics models tend to be computationally demanding with increasing number of cells and strength of the cell-to-cell interactions and, as such, are much less plausible to describe larger populations. In this thesis, we compare the mean-field and more sophisticated KSA-based modelling frameworks and demonstrate that the strength of cell-to-cell interactions, which intuitively translates to the adhesion strength on the population scale, is a paramount factor determining the suitability of each approximation. We demonstrate that when the cell-to-cell interactions are strong enough, mean-field approximation provides inadequate description of the discrete stochastic data. Therefore, great care should be taken when applying mean-field model to describe and parametrise populations of cells with notable adhesion. This result is especially important in the context of populations of cells that increase in size (Chapter 3) because increasing size of each individual implies more frequent interaction which, consequently, induce correlations and volume exclusion effects.

Chapter 4 and Chapter 5 present model validation part of the thesis. In Chapter 4

we use a novel experimental data in conjunction with a discrete modelling framework to provide insights into the role of cell-to-cell pushing. Specifically, we use the range of discrete models: (i) Model I, which includes cell diffusion, cell-to-cell interaction forces, and dynamical cell size; (ii) Model II: constant cell size with cell diffusion and cell-to-cell interaction forces; (iii) Model III: dynamical cell size with no diffusion but cell-to-cell interaction forces; (iv) Model IV: diffusion only. Here Model I is the most complete model which includes all cell motility mechanisms we are interested in. A parameter estimation technique is applied in the attempt to match experimentally measured cell density profiles with density profiles predicted by Models I-IV. These results are presented in Chapter 4 and demonstrate that models that assume the constant cell size (Model II), or no cell size at all (Model IV) always underperform in comparison to the Model I and Model III and suggest that the mechanical cell-to-cell pushing plays important role in the spreading of the cell colony. We note that the cell-to-cell pushing is strikingly pronounced in our experiments since cells increase in size at an increased rate. However, our results do not imply that cell pushing is exclusive to cells that increase in size. In a crowded environments where volume exclusion effects play important role and collisions are frequent cell pushing may contribute significantly to a collective behaviour of the entire population.

In Chapter 5 we use the multi-species continuum model developed in Chapter 2 to explore the ability of a single-species model describe behaviour of the heterogeneous cell population. We explore the heterogeneity in cell size, δ, diffusivity, D, (undirected motility), and amplitude of cell-to-cell interactions, f_0, (directed motility and adhesion). For each of the aforementioned parameters we consider four different distributions: homogeneous, monotonically increasing, monotonically decreasing, and non-monotonic distribution. Our results in Chapter 5 show that for certain types of heterogeneity, such as monotonically decreasing heterogeneity in cell sizes, single-species model can accurately describe behaviour of the heterogeneous multi-species population. This is an important result, because single-species models tend to be computationally more efficient than their multi-species counterparts. On the other hand, non-homogeneous distribution in cell sizes demonstrates a poor match between density profiles generated by the multi-species heterogeneous model and profiles predicted by the single-species model. Therefore, the choice between single-species and multi-species models should be made after carefully assessing the degree of heterogeneity in a population. We acknowledge the fact, that some forms of heterogeneity might be fairly straightforward to identify and quantify as demonstrated in Chapter 5 for the heterogeneity in cell sizes. However, heterogeneity in interaction forces is much more challenging and there is no established procedure to measure cell-to-cell interaction forces efficiently.

6.2 Future work

In this section we present four most broad pathways for potential extensions to our discrete and continuum modelling frameworks and summarise them in Figure 6.1.

- Pathway **I**: Introducing proliferation (mitosis) in the single species discrete and continuum models of cell motility and adhesion. Green disks denote agents that increase in size until they are ready to divide (12 h) and then split into two daughter cells (24 h). The time scale is chosen to be arbitrary for illustrative purposes.

  Discrete and continuum models presented in Chapter 2 and Chapter 3 are developed assuming constant number of cells, that is neglecting proliferation. This assumption is justified when we work with the experiments where proliferation has been inhibited, such as presented in Chapter 4. Approximately constant cell count can be also justified when applying models to the experimental data with cells that have doubling time much longer than duration of the experiments. Although it is relatively straightforward to extend discrete model to include proliferation, it is non-trivial to extend IPDE-based models to include proliferation. Including proliferation will require different derivation procedure due to some underlying assumptions based around van Kampen lemma used in Chapter 2A.

- Pathway **II**: Two-species heterogeneous model of cell motility and adhesion with proliferation. Here, all model parameters can potentially vary between populations including proliferation rate and rate of change of the cell size. Similarly to scenario presented in Pathway **I**, cells increase in size until they are about to divide (12 h) and then split into two daughter cells (24 h). Green and blue cells can be either different cell types, or cells of different sizes going through different stages of their cell cycle.

  In the context of this thesis, all parameter estimation procedures presented in Chapter 4 were performed assuming that all cells have the same size at the beginning of the experiment. Although the cell size heterogeneity is major topic of this thesis, Chapter 4 is focused on the effects of the dynamical cell size on the movement of cell fronts and does not address how the differences in cell sizes could possibly affect our estimates of the model parameters. Simulations in Chapter 4 assume that all cells increase in size at the same rate. As such, it is of interest to explore how including the heterogeneity in a cell growth rate can extend our understanding of the cell front expansion and collective cell migration in general. Quantifying the innate differences in rates of increasing of the cell sizes might require sophisticated cell tracking procedures.

- Pathway **III**: Multi-species heterogeneous model of cell motility, adhesion, and proliferation with *continuous* heterogeneity in all parameters. In this model, every single cell can be assigned unique values of model parameters, or entire population split into arbitrary number of functional groups, types, and sizes. Green disks of various shades of green denote cells of different sizes and types.

  The model of the cell-to-cell interaction forces can be extended to explicitly incorporate the dynamics of E-cadherin protein bonds responsible for adhesion on a cellular level. This way adhesive forces between cells can be modelled to depend on the cell-to-cell contact area rather then the intercellular distance. Throughout this thesis we assume that the friction forces resulting from movement of a cell in a viscous environment do not depend on its size (Equation 2.2). Taking into account increased friction as the cell size increases with time is a potential extension to improve realism of the discrete model.

- Pathway **IV**: *In silico* reconstruction of living tissue environment including cell motility, adhesion, proliferation, extracellular matrix, and biochemical clues. The green disks of various shades of green denote cells of different sizes and types. The orange solid lines denote extracellular matrix. The blue zigzag arrows show the direction of chemotactic gradient that only small light-green cells are responsive to.

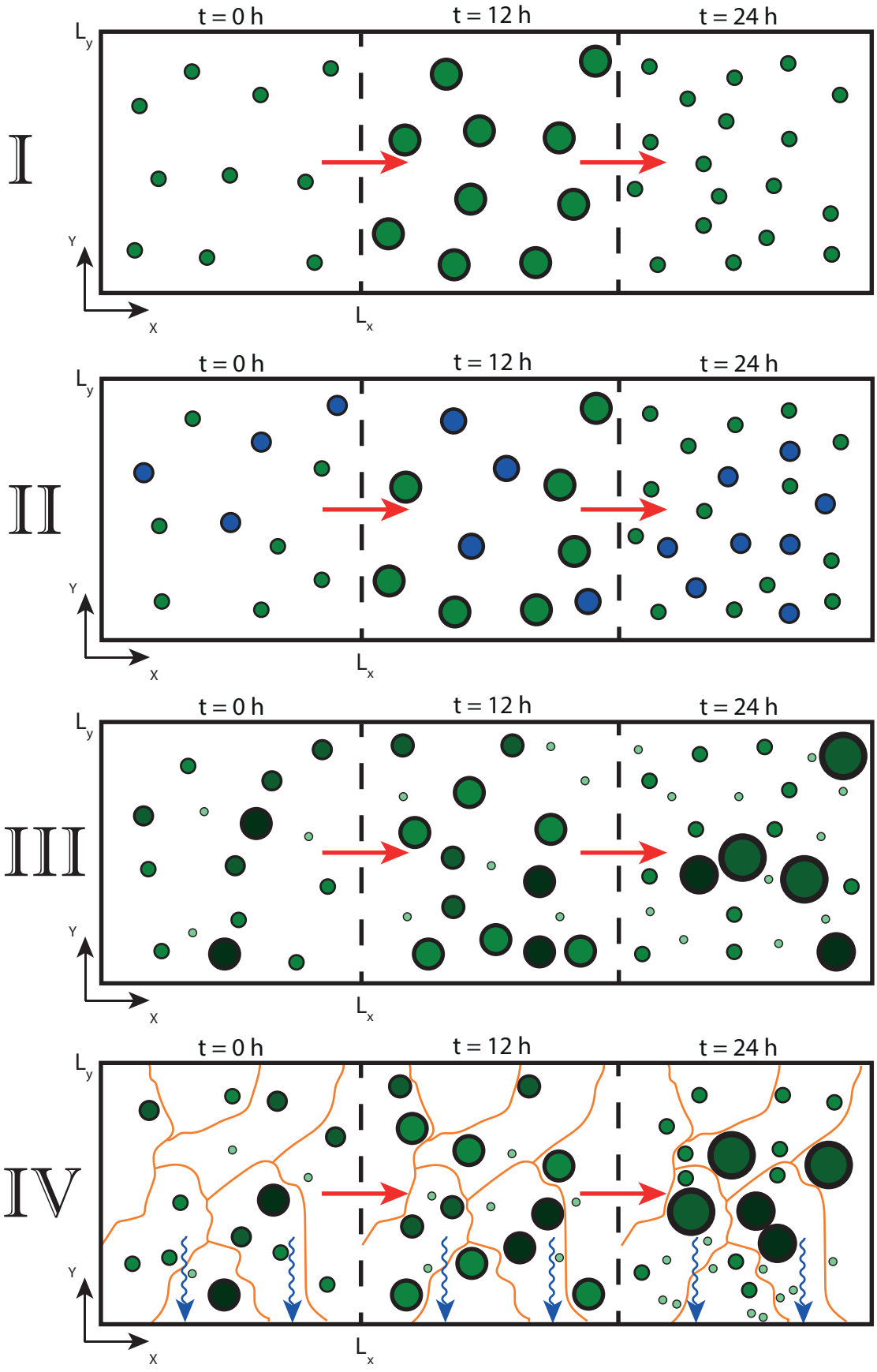


Figure 6.1: Pathways for potential extensions: (I) Single-species model of cell motility, adhesion and proliferation; (II) Two-species heterogeneous model of cell motility, adhesion, and proliferation; (III) Multi-species heterogeneous model of cell motility, adhesion, and proliferation with continuous heterogeneity in parameters; (IV) *In silico* reconstruction of living tissue environment.

Literature cited

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

Amodeo, A.A., Skotheim, J.M. Cell-Size Control. *Cold Spring Harbor perspectives in biology*, vol. 8(4), a019083, 2016.

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

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

Berg, H.C. *Random walks in biology*. Princeton University Press, 1993.

Binder, B.J., Simpson, M.J. Quantifying spatial structure in experimental observations and agent-based simulations using pair-correlation functions. *Physical Review E*, vol. 88, 022705, 2013.

Binder, B.J., Simpson, M.J. Cell density and cell size dynamics during *in vitro* tissue growth experiments: Implications for mathematical models of collective cell behaviour. *Applied Mathematical Modelling*, vol. 40, pp. 3438-3446, 2016.

Binny, R.N., Plank, M.J., James, A. Spatial moment dynamics for collective cell movement incorporating a neighbour-dependent directional bias. *Journal of the Royal Society Interface*, vol. 12, 20150228, 2015.

Binny, R.N., Haridas, P., James, A., Law, R., Simpson, M.J., Plank, M.J. Spatial structure arising from neighbour dependent bias in collective cell movement. *PeerJ*, vol. 4, e1689, 2016a.

Binny, R.N., James, A., Plank, M.J. Collective cell behaviour with neighbour-dependent proliferation, death and directional bias. *Bulletin of Mathematical Biology*, vol. 78(11), pp. 2277-2301, 2016b.

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

Bocci, F., Gearhart-Serna, L., Boareto, M., Ribeiro, M., Ben-Jacob, E., Devi, G.R., Levine, H., Onuchic, J.N., Jolly, M.K. Toward understanding cancer stem cell heterogeneity in the tumor microenvironment *Proceedings of the National Academy of Sciences*, vol. 116 (1), pp. 148-157, 2019.

Bolker, B.M., Pacala, S.W. Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. *Theoretical Population Biology*, vol. 52, pp. 179-197, 1997.

Bolker, B.M., Pacala S.W., Neuhauser, C. Spatial dynamics in model plant communities:What do we really know? *The American Naturalist*, vol. 162, pp. 135-148, 2003.

Bruna, M., Chapman, S.J. Excluded-volume effects in the diffusion of hard spheres. *Physical Review E*, vol. 85, 011103, 2012.

Buske, P., Galle, J., Barker, N., Aust, G., Clevers, H., Loeffler, M. A comprehensive model of spatio-temporal stem cell and tissue organisation in the intestinal crypts. *PLoS Computational Biology*, vol. 7(1), e1001045, 2011.

Byrne, H., Preziosi, L. Modelling solid tumour growth using the theory of mixtures. *Mathematical Medicine and Biology*, vol. 20, pp. 341-366, 2003.

Cai, D., Chen, S.C., Prasad, M., He, L., Wang, X,, Choesmel-Cadamuro, V., Sawyer, J.K., Danuser, G., Montell, D.J. Mechanical feedback through E-cadherin promotes direction sensing during collective cell migration. *Cell*, vol. 157(5), pp. 1146-1159, 2014.

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

Capasso, V., Flandoli, F. On the mean field approximation of a stochastic model of tumour-induced angiogenesis. *European Journal of Applied Mathematics*, vol. 30, pp. 619-658, 2019.

Carmona-Fontaine C., Theveneau, E., Tzekou, A., Tada, M., Woods, M., Page, K.M., Parsons, M., Lambris, J.D., Mayor, R. Complement fragment C3a controls mutual cell attraction during collective cell migration. *Developmental Cell*, vol. 21(6), pp. 1026-1037, 2011.

Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J., Golland, P., Sabatini, D.M. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biology*, vol. 7(10), R100, 2016.

Chiou, K.K., Hufnagel, L., Shraiman, B.I. Mechanical stress inference for two dimensional cell arrays. *PLoS Computational Biology*, vol. 8(5), e1002512, 2012.

Ciupe, M.S., Bivort, B.L., Bortz, D.M., Nelson, pp.W. Estimating kinetic parameters from HIV primary infection data through the eyes of three different mathematical models. *Mathematical Biosciences*, vol. 200, pp. 1-17, 2006.

Codling, E.A., Plank, M.J., Benhamou, S. Random walk models in biology. *Journal of the Royal Society Interface*, vol. 5, pp. 813-834, 2008.

De Jesus, A.M., Aghvami, M., Sander, E.A. A combined in vitro imaging and multiscale modeling system for studying the role of cell matrix interactions in cutaneous wound healing. *PLoS ONE*, vol. 11(2), e0148254, 2016.

Delgado-SanMartin, J.A., Hare, J.I., Davies, E.J., Yates, J.W.T. Multiscalar cellular automaton simulates in-vivo tumour-stroma patterns calibrated from in-vitro assay data. *BMC Medical Informatics and Decision Making*, vol. 17, 88, 2017.

De Leso, M.L., Pei, J.V. An accurate and cost-effective alternative method for measuring cell migration with the circular wound closure assay. *Bioscience Reports*, vol. 38(5), BSR20180698, 2018.

De Palo, G., Yi, D., Endres, R.G. A critical-like collective state leads to longrange cell communication in *Dictyostelium discoideum* aggregation. *PLoS Biology*, vol. 15(4),

e1002602, 2017

Drasdo, D., Kree, R., McCaskill, J.S. Monte-Carlo approach to tissue-cell populations. *Physical Review E*, vol. 52, pp. 6635-6657, 1995.

Drasdo, D., Höhme, S. A single-cell-based model of tumor growth in vitro: monolayers and spheroids. *Physical Biology*, vol. 2(3), pp. 133-147, 2005.

Dyson, L., Maini, P.K., Baker, R.E. Macroscopic limits of individual-based models for motile cell populations with volume exclusion. *Physical Review E*, vol. 86, 031903, 2012.

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

Ermentrout, G.B., Edelstein-Keshet, L. Cellular automata approaches to biological modeling. *Journal of Theoretical Biology*, vol. 160(1), pp. 97-133, 1993.

Essen BioScience: IncuCyte ZOOM live cell imaging. Available from: http://www.essenbioscience.com/essen-products/incucyte/ (Accessed: April 2019).

Etienne-Manneville, S. Neighborly relations during collective migration. *Current Opinion in Cell Biology*, vol. 30, pp. 51-59, 2014.

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

Field, T.R., Tough, R.J.A. Coupled dynamics of populations supported by discrete sites and their continuum limit. *Proceedings of the Royal Society A*, vol. 466, pp. 2561-2586, 2010.

Fisher, R.A. The wave of advance of advantageous genes. *Annals of Eugenics*, vol. 7, pp. 355-69, 1937.

Frasca, M., Sharkey, K.J. Discrete-time moment closure models for epidemic spreading in populations of interacting individuals. *Journal of Theoretical Biology*, vol. 399, pp.

13-21, 2016.

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

Friedl, P., Hegerfeldt, Y., Tusch, M. Collective cell migration in morphogenesis and cancer. *The International Journal of Developmental Biology*, vol. 48(5-6), pp. 441-449, 2004.

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

García-Ojalvo, J., Sancho, J.M. *Noise in spatially extended systems*. Springer New York, 1999.

Gardiner, B.S., Wong, K.K.L, Joldes, G.R., Rich, A.J., Tan, C.W., Burgess, A.W., Smith, D.W. Discrete element framework for modelling extracellular matrix, deformable cells and subcellular components. *PLoS Computational Biology*, vol. 11(10), e1004544, 2015.

George, M., Bullo, F., Campas, O. Connecting individual to collective cell migration. *Scientific Reports*, vol. 264(3), pp. 1057-1067, 2017.

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

Glenn, H.L., Messner, J., Meldrum, D.R. A simple non-perturbing cell migration assay insensitive to proliferation effects. *Scientific Reports*, vol. 6, 31694, 2016.

Goers, L., Freemont, P., Polizzi, K.M. Co-culture systems and technologies: taking synthetic biology to the next level. *Journal of the Royal Society Interface*, vol. 11, 20140065, 2014.

Grada, A., Otero-Vinas, M., Prieto-Castrillo, F., Obagi, Z., Falanga, V. Research techniques made simple: analysis of collective cell migration using the wound healing assay. *Journal of Investigative Dermatology*, vol. 137, pp. e11-e16, 2017.

Griffith, L.G., Swartz, M.A. Capturing complex 3D tissue physiology in vitro. *Nature Reviews Molecular Cell Biology*, vol. 7(3), pp. 211-224, 2006.

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

Hakim, V., Silberzan, P. Collective cell migration: a physics perspective. *Reports on Progress in Physics*, vol. 80(7), 076601, 2017.

Hastings, R.J., Franks, L.M. Cellular heterogeneity in a tissue culture cell line derived from a human bladder carcinoma. *British journal of cancer*, vol. 47(2), pp. 233-244, 1983.

Hawkins, R.J., Piel, M., Faure-Andre, G., Lennon-Dumenil, A.M., Joanny, J.F., Prost, J., Voituriez, R. Pushing off the walls: A mechanism of cell motility in confinement. *Physical Review Letters*, vol. 102(5), 058103, 2009.

Holle, A.W., Young, J.L., Spatz, J.P. *In vitro* cancer cell-ECM interactions inform *in vivo* cancer treatment. *Advanced Drug Delivery Reviews*, vol. 97, pp. 270-279, 2016.

House, T. Algebraic moment closure for population dynamics on discrete structures. *Bulletin of Mathematical Biology*, vol. 77, pp. 646-659, 2014.

Huang, S., Brangwynne, C.P., Parker, K.K., Ingber, D.E. Symmetry-breaking in mammalian cell cohort migration during tissue pattern formation: Role of random-walk persistence. *Cell Motility and the Cytoskeleton*, vol. 61, pp. 201-213, 2005.

Jeon, J., Quaranta, V., Cummings, P.T. An off-lattice hybrid discrete-continuum model of tumor growth and invasion. *Biophysical Journal*, vol. 98(1), pp. 37-47, 2010.

Jin, W., Shah, E.T., Penington, C.J., McCue, S.W., Chopin, L.K., Simpson, M.J. Reproducibility of scratch assays is affected by the initial degree of confluence: Experiments, modelling and model selection. *Journal of Theoretical Biology*, vol. 390, pp. 136-145, 2016a.

Jin, W., Penington, C.J., McCue, S.W., Simpson, M.J. Stochastic simulation tools and continuum models for describing two-dimensional collective cell spreading with universal growth functions. *Physical Biology*, vol. 13(5), 056003, 2016b.

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

Johnston, S.T., Simpson, M.J., Plank, M.J. Lattice-free descriptions of collective motion with crowding and adhesion. *Physical Review E*, vol. 88, 062720, 2013.

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

Johnston, S.T., Shah, E.T., Chopin, L.K., McElwain, D.L.S., Simpson, M.J. Estimating cell diffusivity and cell proliferation rate by interpreting IncuCyte ZOOM™ assay data using the Fisher–Kolmogorov model. *BMC Systems Biology*, vol. 9, 38, 2015.

Kabla, A. Collective cell migration: leadership, invasion and segregation. *Journal of the Royal Society Interface*, vol. 9, pp. 3268-3278, 2012.

Kaighn, M.E., Narayan, K.S., Ohnuki, Y., Lechner, J.F., Jones, L.W. Establishment and characterization of a human prostatic carcinoma cell line (PC-3). *Investigative Urology*, vol. 17, pp. 16-23, 1979.

Keeling, M.J., Rand, D.A., Morris, A.J. Correlation models for childhood epidemics. *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 264, pp. 1149-1156, 1997.

Keller, E.F., Segel, L.A. Traveling bands of chemotactic bacteria: a theoretical analysis. *Journal of Theoretical Biology*, vol. 30, pp. 235-248, 1971.

Khain, E., Sander, L.M., Schneider-Mizell, C.M. The role of cell-cell adhesion in wound healing. *Journal of Statistical Physics*, vol. 128, pp. 209-218, 2006.

Khain, E., Katakowski, M., Hopkins, S., Szalad, A., Zheng, X., Jiang, F., Chopp, M. Collective behavior of brain tumor cells: The role of hypoxia. *Physical Review E*, vol. 83, 031920, 2011.

Kirkwood, J.G. Statistical mechanics of fluid mixtures. *Journal of Chemical Physics*, vol. 3(5), pp. 300-313, 1935.

Köpf, M.H., Pismen, L.M. A continuum model of epithelial spreading. *Soft Matter*, vol. 9, pp. 3727-3734, 2013.

Kumari, R., Sharma, A., Ajay, A.K., Bhat, M.K. Mitomycin C induces bystander killing in homogeneous and heterogeneous hepatoma cellular models. *Molecular Cancer*, vol. 8, 87, 2009.

Levin, S.A., Whitfield, M. Patchiness in marine and terrestrial systems: from individuals to populations. *Philosophical Transactions: Biological Sciences*, vol. 343, pp. 99-103, 1994.

Liang, C.C., Park, A.Y., Guan, J.L. *In vitro* scratch assay: a convenient and inexpensive method for analysis of cell migration *in vitro*. *Nature Protocols*, vol. 2, pp. 329-333, 2007.

Lowengrub, J.S., Frieboes, H.B., Jin, F., Chuang, Y.-L., Li, X., Macklin, P., Wise, S.M., Cristini, V. Nonlinear modelling of cancer: bridging the gap between cells and tumours. *Nonlinearity*, vol. 23(1), R1R9, 2010.

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

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

Markham, D.C., Simpson, M.J., Maini, P.K., Gaffney E.A, Baker, R.E. Incorporating spatial correlations into multispecies mean-field models, *Physical Review E*, vol. 88, 052713, 2013.

Markham, D.C., Simpson, M.J., Baker, R.E. Choosing an appropriate modelling framework for analysing multispecies co-culture cell biology experiments. *Bulletin of Mathematical Biology*, vol. 77, pp. 713-734, 2015.

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

Matsiaka, O.M., Penington, C.J., Baker, R.E., Simpson, M.J. Discrete and continuum

approximations for collective cell migration in a scratch assay with cell size dynamics. *Bulletin of Mathematical Biology*, vol. 80, pp. 738-757, 2018.

Matsiaka, O.M., Baker, R.E., Shah, E.T., Simpson, M.J. Mechanistic and experimental models of cell migration reveal the importance of intercellular interactions in cell invasion. *Biomedical Physics and Engineering Express*, vol. 5, 045009, 2009.

McLennan, R., Schumacher, L.J., Morrison, J.A., Teddy, J.M., Ridenour, D.A., Box, A.C., Semerad, C.L., Li, H., McDowell, W., Kay, D., Maini, P.K., Baker, R.E., Kulesa, P.M. Neural crest migration is driven by a few trailblazer cells with a unique molecular signature narrowly confined to the invasive front. *Development*, vol. 142, pp. 1-12, 2015.

Menon, T., Kumar, R., Nair, S. Cell size heterogeneity early in development is required for collective cell migration during gastrulation in zebrafish. BioRxiv doi:10.1101/481325, 2018.

Middleton, A.M., Fleck, C., Grima, R. A continuum approximation to an off-lattice individual-cell based model of cell migration and adhesion. *Journal of Theoretical Biology*, vol. 359, pp. 220-232, 2014.

Mousavi, S.J., Doblare, M., Doweidar, M.H. Computational modelling of multi-cell migration in a multi-signalling substrate. *Physical Biology*, vol. 11(2), 026002, 2014.

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

Murray, P.J., Edwards, C.M., Tindall, M.J., Maini, P.K. Classifying general nonlinear force laws in cell-based models via the continuum limit. Physical Review E, vol. 85, 021921, 2012.

Murrell, D.J., Dieckmann, U., Law, R. On moment closures for population dynamics in continuous space. *Journal of Theoretical Biology*, vol. 229, pp. 421-432, 2004.

Nadell, C.D., Foster, K.R., Xavier, J.B. Emergence of spatial structure in cell groups and the evolution of cooperation. *PLoS Computational Biology*, vol. 6(3), e1000716, 2010.

Nan, P., Walsh, D.M., Landman, K.A., Hughes, B.D. Distinguishing cell shoving mechanisms. *PLoS ONE* , vol. 13(3), e0193975, 2018.

Nardini, J.T., Chapnick, D.A., Liu, X., Bortz, D.M. Modeling keratinocyte wound healing dynamics: Cellcell adhesion promotes sustained collective migration. *Journal of Theoretical Biology*, vol. 400, pp. 103-117, 2016.

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

Nickaeen, M., Novak, I.L., Pulford, S., Rumack, A., Brandon, J., Slepchenko, B.M., Mogilner, A. A free-boundary model of a motile cell explains turning behavior *PLoS Computational Biology*, vol. 13(11), e1005862, 2017.

Nikolic, D.L., Boettiger, A.N., Bar-Sagi, D., Carbeck, J.D., Shvartsman, S.Y. Role of boundary conditions in an experimental model of epithelial wound healing. *American Journal of Physiology –Cell Physiology*, vol. 291(1), pp. C68-75, 2006.

Nyegaard, S., Christensen, B., Rasmussen, J.T. An optimized method for accurate quantification of cell migration using human small intestine cells. *Metabolic Engineering Communications,* vol. 3, pp. 76-83, 2016.

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

Painter, K.J., Sherratt, J.A. Modelling the movement of interacting cell populations. *Journal of Theoretical Biology*, vol. 225, pp. 327-339, 2003.

Painter, K.J., Armstrong, N.J., Sherratt, J.A. The impact of adhesion on cellular invasion processes in cancer and development. *Journal of Theoretical Biology*, vol. 264(3), pp.1057-1067, 2010.

Park, J.-A., Atia, L., Mitchel, J.A., Fredberg, J.J,, Butler, J.pp. Collective migration and cell jamming in asthma, cancer and development. *Journal of Cell Science*, vol. 129(18), pp. 3375-3383, 2016.

Peglion, F., Llense, F., Etienne-Manneville, S. Adherens junction treadmilling during collective migration. *Nature Cell Biology*, vol. 16(7), pp. 639-651, 2014.

Peirce, S.M. Computational and mathematical modeling of angiogenesis. *Microcirculation*, vol. 15(8), pp. 739-751, 2008.

Penington, C.J., Hughes, B.D., Landman, K.A. Building macroscale models from microscale probabilistic models: A general probabilistic approach for nonlinear diffusion and multispecies phenomena. *Physical Review E*, vol. 84, 041120, 2011.

Pillay, S., Byrne, H.M., Maini, pp.K. The impact of exclusion processes on angiogenesis models. *Journal of Mathematical Biology*. vol. 77(6-7), pp. 1721-1759, 2018.

Plank, M.J., Law, R. Spatial point processes and moment dynamics in the life sciences: a parsimonious derivation and some extensions. *Bulletin of Mathematical Biology*, vol. 77, pp. 586-613, 2015.

Progatzky, F., Dallman, M.J., Lo Celso, C. From seeing to believing: labelling strategies for *in vivo* cell-tracking experiments. *Interface Focus*, vol. 3(3), 20130001, 2013.

Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T. *Numerical recipes: The art of scientific computing (3rd ed.)*. Cambridge University Press, 2007.

Ramis-Conde, I., Chaplain, M.A.J., Anderson, A.R.A. Mathematical modelling of cancer cell invasion of tissue. *Mathematical and Computer Modelling*, vol. 47(5-6), pp. 533-545, 2008.

Ranft, J., Basan, M., Elgeti, J., Joanny, J.-F., Prost, J., Jülicher, F. Fluidization of tissues by cell division and apoptosis. *PNAS*, vol. 107(49), pp. 20863-20868, 2010.

Regnault, C., Dheeman, D.S., Hochstetter, A. Microfluidic Devices for Drug Assays. *High-throughput*, vol. 7(2), 18, 2018.

Riss, T. Selecting cell-based assays for drug discovery screening. *Cell Notes*, vol. 13, pp. 16-21, 2005.

Rognoni, E., Watt, F.M. Skin cell heterogeneity in development, wound healing, and cancer. *Trends in cell biology*, vol. 28(9), pp. 709-722, 2018.

Rosso, F., Giordano, A., Barbarisi, M., Barbarisi, A. From cellECM interactions to tissue engineering. *Journal of cellular physiology*, vol. 199, pp. 174-180, 2004.

Rutter, E.M., Banks, H.T., Flores, K.B. Estimating intratumoral heterogeneity from spatiotemporal data. *Journal of Mathematical Biology*, vol. 77, pp. 1999-2022, 2018.

Sadeghi, H.M., Seitz, B., Hayashi, S., LaBree, L., McDonnell, P.J. In vitro effects of Mitomycin-C on human keratocytes. *Journal of Refractive Surgery*, vol. 14, pp. 534-540, 1998.

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

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

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

Sepulveda, N., Petitjean, L., Cochet, O., Grasland-Mongrain, E., Silberzan, P., Hakim, V. Collective cell motion in an epithelial sheet can be quantitatively described by a stochastic interacting particle model. *PLoS Computational Biology*, vol. 9(3), e1002944, 2013.

Serra-Picamal, X., Conte V., Vincent, R., Anon, E., Tambe, D.T., Bazellieres, E., Butler, J.P., Fredberg, J.J., Trepat, X. Mechanical waves during tissue expansion. *Nature Physics*, vol. 8, pp. 628-634, 2012.

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

Sharkey, K.J., Fernandez, C., Morgan, K.L., Peeler, E., Thrush, M., Turnbull, J.F., Bowers, R.G. Pair-level approximations to the spatio-temporal dynamics of epidemics on asymmetric contact networks. *Journal of Mathematical Biology*, vol. 53, pp. 61-85, 2006.

Sharkey, K.J., Kiss, I.Z., Wilkinson, R.R., Simon, P.L. Exact equations for SIR epidemics on tree graphs. *Bulletin of Mathemtical Biology*, vol. 77, pp. 614-645, 2015.

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

Sherratt, J.A., Murray, J.D. Models of epidermal wound healing. *Proceedings of the Royal Society of London B*, vol. 241, pp. 29-36, 1990.

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

Simpson, M.J., Landman, K.A., Hughes, B.D., Newgreen, D.F. Looking inside an invasion wave of cells using continuum models: Proliferation is the key. *Journal of Theoretical Biology*, vol. 243, pp. 343-360, 2006a.

Simpson, M.J., Landman, K.A., Newgreen, D.F. 2006 Chemotactic and diffusive migration on a nonuniformly growing domain: numerical algorithm development and applications. *Journal of Computational and Applied Mathematics*, vol. 192, pp. 282-300, 2006b.

Simpson, M.J., Zhang, D.C., Landman, K.A., Newgreen, D.F. Cell proliferation drives neural crest cell invasion of the intestine. *Developmental Biology*, vol. 302, pp. 553-568, 2007.

Simpson, M.J., Landman, K.A., Hughes, B.D. Multi-species simple exclusion processes. *Physica A: Statistical Mechanics and its Applications*, vol. 388, pp. 399-406, 2009a.

Simpson, M.J. Depth-averaging errors in reactive transport modelling. *Water Resources Research*, vol. 45, W02505, 2009b.

Simpson, M.J., Binder, B.J., Haridas, P., Wood, B.K., Treloar, K.K., McElwain, D.L.S., Baker, R.E. Experimental and modelling investigation of monolayer development with clustering. *Bulletin of Mathematical Biology*, vol. 75, pp. 871-889, 2013a.

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

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

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

Singer, A. Maximum entropy formulation of the Kirkwood superposition approximation. *The Journal of Chemical Physics*, vol. 121, pp. 3657-3666, 2004.

Smith, S., Cianci, C., Grima, R. Macromolecular crowding directs the motion of small molecules inside cells. *Journal of the Royal Society Interface*, vol. 14, 20170047, 2017.

Sommer, C., Strähle, C., Köthe, U., Hamprecht, F.A. Ilastik: Interactive learning and segmentation toolkit. In: *Biomedical imaging: From nano to macro. Proceeding of Eighth IEEE International Symposium on Biomedical Imaging*, 2011.

Srinivas, S., Rodriguez, T., Clements, M., Smith, J.C., Beddington, R.S.pp. Active cell migration drives the unilateral movements of the anterior visceral endoderm. *Development*, vol. 131(5), pp. 1157, 2004.

Steinberg, M.S. Adhesion in development: An historical review. *Developmental Biology*, vol. 180, pp. 377-388, 1996.

Szabó, A., Merks, R.M. Cellular potts modeling of tumor growth, tumor invasion, and tumor evolution. *Frontiers in Oncology*, vol. 3, 87, 2013.

Tao, H., Berno, A.J., Cox, D.R, Frazer, K.A. In vitro human keratinocyte migration rates are associated with SNPs in the KRT1 Interval. *PLoS ONE*, vol. 2(8), e697, 2007.

Tersoff, J. New empirical approach for the structure and energy of covalent systems. *Physical Review B*, vol. 37(12), pp. 6991-7000, 1987.

Tlili, S., Gauquelin, E., Li, B., Cardoso, O., Ladoux, B., Delanoë-Ayari, H., Graner, F. Collective cell migration without proliferation: density determines cell velocity and wave velocity. *Royal Society Open Science*, vol. 5, 172421, 2018.

Treloar, K.K, Simpson, M.J., Haridas, P., Manton, K.J., Leavesley, D.I., McElwain, D.L.S., Baker, R.E. Multiple types of data are required to identify the mechanisms influencing the spatial expansion of melanoma cell colonies. *BMC Systems Biology*, vol. 7, 137, 2013a.

Treloar, K.K., Simpson, M.J. Sensitivity of edge detection methods for quantifying cell migration assays. *PLoS ONE*, vol. 8(6), e67389, 2013b.

Treloar, K.K., Simpson, M.J., McElwain, D.L.S., Baker, R.E. Are in vitro estimates of cell diffusivity and cell proliferation rate sensitive to assay geometry? *Journal of Theoretical Biology*, vol. 356, pp. 71-84, 2014.

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

Trepat, X., Wasserman, M.R., Angelini, T.E., Millet, E., Weitz, D.A., Butler, J.P., Fredberg, J.J. Physical forces during collective cell migration. *Nature Physics*, vol. 5, pp. 426-430, 2009.

Trinh, A.L., Chen, H., Chen, Y., Hu, Y., Li, Z., Siegel, E.R., Linskey, M.E., Wang, pp.H., Digman, M.A., Zhou, Y.H. Tracking functional tumor cell subpopulations of malignant glioma by phasor fluorescence lifetime imaging microscopy of NADH. *Cancers*, vol. 9(12), E168, 2017.

Van Kampen, N.G. Stochastic differential equations. *Physics Reports*, vol. 24c(3), pp. 171-228, 1976.

Warne, D.J., Baker, R.E., Simpson, M.J. Using experimental data and information criteria to guide model selection for reactiondiffusion problems in mathematical biology. *Bulletin of Mathematical Biology*, vol. 81(6), pp.1760-1804, 2019.

Weigelt, B., Ghajar, C.M., Bissell, M.J. The need for complex 3D culture models to unravel novel pathways and identify accurate biomarkers in breast cancer. *Advanced Drug Delivery Reviews*, vol. 69-70, pp. 42-51, 2014.

Weijer, C.J. Collective cell migration in development. *Journal of Cell Science*, vol. 122(18), pp. 3215-3223, 2009.

Weinberg, R.A. *The biology of cancer. Second edition*. Garland Science, 2009.

Wise, S.M., Lowengrub, J.S., Frieboes, H.B., Cristini, V. Three-dimensional multi-species nonlinear tumor growth–I Model and numerical method. *Journal of Theoretical Biology*, vol. 253(3), pp. 524-543, 2008.

Wynn, M.L., Rupp, P., Trainor, P.A., Schnell, S., Kulesa, P.M. Follow-the-leader cell migration requires biased cell-cell contact and local microenvironmental signals. *Physical Biology*, vol. 10(3), 035003, 2013.

Yates, C., Parker, A., Baker, R.E. Incorporating pushing in exclusion-process models of cell migration. *Physical Review E*, vol. 91, 052711, 2015.

Zaritsky, A., Kaplan, D., Hecht, I., Natan S., Wolf, L., Gov, N.S., Ben-Jacob, E., Tsarfaty, I. Propagating waves of directionality and coordination orchestrate collective cell migration. *PLoS Computational Biology*, vol. 10(7), e1003747, 2014.

Zhu, J., Liang, L., Jiao, Y., Liu, L., US-China Physical Sciences-Oncology Alliance. Enhanced invasion of metastatic cancer cells via extracellular matrix interface. *PLoS ONE*, vol. 10(2), e0118058, 2015.