Mathematical models for quantifying collective cell behaviour

Stuart T. Johnston
BMath, BAppSc (Hons) (Maths)

under the supervision of

Prof. Matthew Simpson
and
Prof. Scott McCue

Mathematical Sciences
Science and Engineering Faculty
Queensland University of Technology

2017

Submitted in fulfillment of the requirements of the degree of
Doctor of Philosophy

Abstract

Collective behaviour is critical to a variety of biological and ecological processes, including tumour invasion, wound healing and the spread of invasive species. Mathematical modelling techniques provide an opportunity for obtaining insight into the underlying mechanisms governing collective behaviour. Furthermore, mathematical models can be interfaced with experimental data to enhance quantitative information obtained from experiments. The aim of this thesis is two-fold. First, to investigate the application of mathematical models to experimental data to obtain robust estimates of the parameters governing collective behaviour. Second, to develop novel continuum mathematical descriptions of individual-based models of birth, death and movement.

For the first part of this thesis, I begin by examining the application of a well-known mathematical model of cell motility and cell proliferation, the Fisher-Kolmogorov model, to IncuCyte ZOOM$^{\text{TM}}$ assay data for a prostate cancer (PC-3) cell population. Standard techniques used to interpret IncuCyte ZOOM$^{\text{TM}}$ assay data do not provide insight about the individual contributions of cell motility and cell proliferation to the overall migration of the cell population. I find that by combining experimental measurements of the evolution of the position of the leading edge of the cell population and the evolution of the cell density away from the leading edge, I am able to obtain unique estimates of the cell diffusivity, cell proliferation rate and cell carrying capacity density. Furthermore, I am able to quantify how these parameters are influenced by the presence of varying concentrations of epidermal growth factor.

Next, I investigate whether the position of the leading edge of a cell population in a scratch assay can be combined with an appropriate mathematical model to provide unique estimates of the cell motility rate and the cell proliferation rate. Leading edge data is a commonly-reported experimental measurement from scratch assays, which are typically interpreted in a qualitative fashion or with a quantitative technique that does not isolate the individual roles of motility and proliferation. I implement a lattice-based random walk model of motility and proliferation, mimic the geometry of a scratch assay, and combine this mathematical model with experimental leading edge data through an automated edge detection algorithm. I find that, provided the difference in the time scales of cell proliferation and cell motility is accounted for, this technique produces unique estimates of the cell diffusivity and cell proliferation rate.

I then consider an approximate Bayesian computation (ABC) parameter recovery approach and examine which experimental measurements from a scratch assay provide the most information about the cell diffusivity and cell proliferation rate parameters. ABC techniques provide parameter distributions rather than point-estimates and hence contain quantitative information about the uncertainty associated with the parameter estimates. I find that an ABC approach provides parameter estimates that are consistent with previously-obtained parameter estimates for 3T3 fibroblasts. Furthermore, I find that measuring the number of cells present within the scratch assay and the pair correlation function provides sufficient information to obtain robust parameter estimates.

For the final piece of work associated with the first part of this thesis, I extend my ABC technique to examine the influence of experimental design choices on parameter estimates. There is significant variation in the experimental design of scratch assays presented in the literature. Specifically, the number and timing of experimental observations captured, as well as whether the scratch assay contains one front or two opposingly-directed fronts, varies throughout the literature. These design choices are not discussed and the influence of the choices has not been quantified. I find that capturing two experimental observations, after the initial observation, is sufficient to obtain robust parameter estimates. Additionally, I find that a scratch assay with a single front contains less uncertainty about the recovered parameters, compared to a scratch assay with two opposingly-directed fronts.

These first four pieces of work highlight the importance of selecting an appropriate parameter recovery framework, which consists of a mathematical model, an experimental measurement or summary statistic, and a parameter calibration technique. Throughout these pieces of work I demonstrate that making naïve framework choices does not necessarily elucidate the individual roles of cell motility and cell proliferation in the collective spreading of a cell population. Furthermore, even if the roles of cell motility and cell proliferation are isolated, mathematical modelling provides insight into how to refine the framework choices made such that the information obtained about cell motility and cell proliferation is maximised. In combination, these four pieces of work provide guidance on potential model and parameter recovery technique choices for given types of experimental data.

For the second part of this thesis, I first examine a corrected mean-field description of a lattice-based model of two interacting subpopulations, where the agents in each subpopulation undergo birth, death and movement events. Standard mean-field descriptions for such models have been presented previously but provide inaccurate approximations of the average discrete behaviour outside limited parameter regimes. Corrected mean-field descriptions for related models have provided more accurate approximations than the corresponding mean-field descriptions. My investigation provides guidance about the choice of mean-field description required in a suite of parameter regimes for a number of case studies inspired by cell biology experiments.

Next, I present a novel interpretation of a lattice-based random walk model of birth, death and adhesive movement. Instead of considering the average occupancy of each lattice site, as is typical for both mean-field and corrected mean-field descriptions, I consider the lattice as a combination of groups of contiguous occupied sites and contiguous vacant sites. I find that modelling the dynamics of these groups, referred to as chains and gaps, respectively, provides a more accurate description of the average discrete behaviour than either mean-field or corrected mean-field descriptions. Furthermore, this novel description contains a measure of the spatial clustering present in the discrete model.

I then extend the random walk model of birth, death and movement to incorporate competitive and/or co-operative mechanisms, where there is a negative or positive benefit associated with belonging to the bulk population. Such mechanisms arise throughout cell biology and ecology. I derive the mean-field description of this random walk and analyse the corresponding long-time predictions in a suite of parameter regimes. I find that the inclusion of competitive and/or co-operative mechanisms results in a variety of interesting behaviour, depending on the relative rates of birth, death and movement. For example, I observe the Allee effect, where growth is negative below a threshold density, as well as travelling wave solutions that contain sharp fronts and shocks.

The final study in the second part of this thesis examines an extension of the chain-and-gap description that can be used to model processes containing moving fronts, as well as competitive and/or co-operative mechanisms. As such mechanisms give rise to more interesting behaviour than standard models of birth, death and movement, the question of whether the population persists or becomes extinct is critical. I find that the extended chain-and-gap description accurately predicts both the long-time persistence or extinction of the population in all parameter regimes, as well as providing an accurate estimate of the rate of invasion of the population. Furthermore, the description provides details of the spatial clustering within moving fronts.

These second four pieces of work provide detailed information about the continuum approximations required to accurately reflect the average behaviour of discrete individual-based models of collective behaviour. The type of continuum approximation depends on the mechanisms present in the discrete model, as well as the relative rates at which these mechanisms occur. In particular, if the combination of mechanisms and mechanism rates results in significant spatial clustering due to nearest-neighbour effects, typical approximations are inaccurate. As the chain-and-gap description accurately describes these cases, the combination of these four pieces of work provides a framework for informing the choice of continuum approximation for a particular discrete model.

In conclusion, I detail the contributions of each of the eight publications to the overall literature and provide discussion regarding potential avenues for future work related to collective behaviour. Overall, the work contained in this thesis provides a framework for modelling collective behaviour with random walks. In particular, I detail techniques for implementing an appropriate continuum description of random walks with a suite of mechanisms and parameter regimes, and methods for applying these random walk models of collective behaviour to relevant experimental data to obtain meaningful quantitative information.

Declaration

I hereby declare that this submission is my own work and to the best of my knowledge it contains no material previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any other degree or diploma at QUT or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by colleagues, with whom I have worked at QUT or elsewhere, during my candidature, is fully acknowledged.

I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.

Signature ____________________ Date ____________________

Acknowledgements

First and foremost, I would like to thank my supervisor, Prof. Matthew Simpson, for his guidance and advice throughout my studies. Without his tireless efforts and dedication this thesis would not have been possible. He made every effort to meet whenever was necessary and provided significant research opportunities to enhance my training. I am deeply grateful for his supervision.

Second, to all of the researchers I worked collaboratively with during my doctoral training: E/Prof. Sean Mcelwain, Prof. Ruth Baker, A/Prof. Joshua Ross, Dr Ben Binder, Prof. Lisa Chopin, Ms. Parvathi Haridas and Ms. Esha Shah. I would like to thank each one of these researchers for allowing me to work with them, and for the unique insight resulting from each collaboration. Particular thanks must be given to E/Prof. Sean McElwain for sharing his wealth of knowledge in applied mathematics.

I am grateful for the funding provided by the Australian Government, through the Australian Postgraduate Award, the Wound Management Innovation Co-operative Research Centre, and the Queensland University of Technology, through the Vice-Chancellor's Top-Up Scholarship. Without this support this thesis would not exist.

Finally, I am thankful for the advice and support provided by my family, friends and fellow doctoral students. The fruitful discussions and coffee breaks provided an enjoyable change of pace from research. I feel fortunate to have so many generous and supportive people in my life. This work is for all of you.

List of abbreviations

- **ABC** - Approximate Bayesian computation.
- **ABC-MCMC** - Approximate Bayesian computation Markov chain Monte Carlo.
- **ABDM** - Adhesive birth-death-movement.
- **CI** - Credible interval.
- **C&G** - Chain-and-gap.
- **CMF** - Corrected mean-field.
- **EGF** - Epidermal growth factor.
- **KSA** - Kirkwood superposition approximation.
- **MD** - Moment dynamics.
- **MF** - Mean-field.
- **ODE** - Ordinary differential equation.
- **PDE** - Partial differential equation.

Contents

Chapter 1

Introduction

1.1 Overview

Collective cell behaviour is a key feature underpinning biological processes such as development, malignant spreading and wound healing [50, 78]. The migration of a cell population determines, for example, the rate of spread of malignant tumours or the time taken for a wound to heal [50, 78]. Factors that influence the migration of a cell population are therefore a key focus of drug development [65, 86]. Unfortunately it is typically too experimentally complex to perform investigations into these factors *in vivo* and subsequently *in vitro* experiments are used to observe collective cell behaviour [11, 44, 58, 64, 82, 89].

Scratch assays are a class of *in vitro* cell biology assays that are commonly performed to examine the collective behaviour of a cell population during the closure of a model wound [44, 58, 64, 82, 89]. In a scratch assay, cells are placed on a dish and are allowed to grow until the cell population forms a confluent monolayer [44]. An artificial wound is induced by removing a portion of the cell monolayer, typically via a sharp-tipped instrument such as a pipette. The cells migrate into the vacant space introduced by the creation of the wound, and experimental images of this migration are captured. Eventually, the cell population reforms a confluent monolayer through a combination of motility and proliferation [73]. There is significant variation in the experimental set-up for scratch assays. For example, the wound can be induced such that there is a single cell front [10, 47] or two opposingly-directed cell fronts [1, 44, 59]. The number of experimental images captured, as well as the duration of the experiment, can vary significantly between experiments [1, 46, 47, 58, 69, 82]. For example, some investigations report a large number of experimental observations [10, 46, 47], whereas others report two or three observations [1, 58, 69, 82]. There is significant benefit in reducing the number of experimental observations, as it is both time-consuming and expensive to capture and interpret the experimental images.

Typically, scratch assays are interpreted in a qualitative fashion, where a scratch assay is performed with some kind of chemical treatment and compared to a scratch assay performed under control conditions [41, 58, 82]. For example, Teppo *et al.* [82] demonstrates the influence of hypoxia on collective cell behaviour by qualitatively comparing the rate of spread for two cell populations, where one cell population is under hypoxic conditions. However, no quantitative insight is offered into the underlying cellular mechanisms that are affected by hypoxic conditions [82]. It is therefore instructive to consider approaches that provide quantitative estimates of the parameters governing cellular mechanisms, as the influence of a particular chemical treatment or experimental design choice can be quantified.

One approach for obtaining quantitative parameter estimates from scratch assays is to employ a mathematical model [4, 46, 47, 63, 73, 85]. Maini *et al.* [46, 47] consider the Fisher-Kolmogorov model [19, 40],

$$\frac{\partial C(x,t)}{\partial t} = D\frac{\partial^2 C(x,t)}{\partial x^2} + \lambda C(x,t)\left(1 - \frac{C(x,t)}{K}\right),$$

where $C(x,t)$ is the cell density, λ is the cell proliferation rate, D is the cell diffusivity and K is the cell carrying capacity density, and match the long-time travelling wave speed arising from the model, $v = 2\sqrt{\lambda D}$, to the position of the leading edge of the cell population in a scratch assay using a least-squares fit. Unfortunately this method does not produce unique estimates of D and λ, as the product of D and λ is not unique. Furthermore, this method requires experimental data that is obtained over a sufficiently long period, which is not practical given the typical approach of an experimental time of 24 h or less [11, 58, 69]. Other methods are able to isolate estimates of D and λ but require experimental measurements that are either more difficult to measure than the position of the leading edge, or are more time-consuming to interpret. For example, Arciero *et al.* [4] propose a continuum model that is parameterised using the experimental cell density as a function of distance, as well as the position of leading edge. Determining the cell density as a function of distance requires counting the number of cells throughout the experimental domain, which is time-consuming and hence is not suited for high-throughput assays.

While different experimental measurements have varying levels of time and cost, it is unclear how the amount of information gained about the cell parameters through the parameterisation of an appropriate mathematical model corresponds to the cost. Ideally, an experimental measurement that can be measured automatically, such as the position of the leading edge of the cell population [84, 88] or the area of the induced wound [60], could be employed to obtain all of the relevant information about the cell parameters of interest. Experimental measurements that can obtained automatically are cheap and are suited to experiments that are high-throughput or require many experimental replicates. However, quantifying the information gained about the parameter estimates requires statistical techniques as performing experimental replicates only provides information about the variability present in the experiment rather than the uncertainty associated with the parameter recovery technique [48, 56, 83].

There are many open questions in the literature regarding the influence of the experimental design, the information present in standard experimental measurements, approaches that combine mathematical techniques and experimental data to obtain quantitative parameter estimates, and techniques that provide a measure of the uncertainty associated with the parameter recovery method. Therefore, there is significant opportunity in this area of research to develop novel techniques and obtain quantitative insight into the cellular processes governing collective cell migration.

More generally, throughout the physical and life sciences, lattice-based random walks are commonly employed to model collective behaviour [13, 16, 17, 31, 33, 37, 51, 72]. For example, both Deroulers *et al.* [16] and Khain *et al.* [37] use a lattice-based random walk to describe the migration of glioma cells. In the context of molecular aggregation, Mackie *et al.* [45] describe the formation of micelles through a random walk process. The ability to define relevant individual mechanisms in a random walk process, such as motility, proliferation, death and agent-agent adhesion, makes random walks a flexible modelling tool [72]. The stochastic nature of the random walk processes means that it can be computationally intensive to determine the average behaviour of the random walk, particularly for processes that are spatially heterogeneous. Furthermore, it is difficult to obtain insight into the average behaviour through analysis of the random walk.

Due to these shortfalls, deterministic continuum approximations of the average random walk behaviour have been proposed [6, 16, 21, 37, 51, 70, 72, 73]. The most common continuum description is a mean-field description, which involves an assumption that the spatial correlations between lattice sites are negligible [16, 37, 72]. Mean-field descriptions of random walk processes have been used to model glioma cell migration [16, 37] and the migration of neural crest cells [72]. However, the assumption that the spatial correlations are negligible is only valid in an extremely limited set of parameter regimes [6, 35, 70]. If the rate of proliferation or death is not sufficiently small compared to the rate of motility, or if there is agent-agent adhesion or repulsion that is not sufficiently weak, the spatial correlations become significant and hence the mean-field assumption is invalid [6, 35, 70].

Approaches that explicitly describe the spatial correlations between lattice sites have been proposed and are known as corrected mean-field descriptions [6, 35, 49, 70]. Such descriptions model the occupancy of lattice sites in terms of the occupancy of lattice pairs which, in turn, are modelled in terms of the occupancy of lattice triplets, and so on [6, 35, 49, 70]. It is analytically intractable to deal with the resulting infinite system of equations and hence, in practice, an appropriate moment closure technique is implemented to close the system of equations [6, 35, 42, 70, 75]. Specifically, these techniques are used to approximate the occupancy of lattice triplets in terms of the occupancy of lattice pairs [75]. Corrected mean-field descriptions provide a more accurate description of the average random walk behaviour, compared to mean-field descriptions [6, 35, 49, 70]. For example, corrected mean-field descriptions have been applied to birth-death-movement random walks for both the spatially homogeneous case and the spatially heterogeneous case, as well adhesive movement random walks and dispersal models [6, 35, 42, 49, 70]. Note that corrected mean-field descriptions are also referred to as moment dynamics descriptions, particularly in the ecological literature [42], and as such we use both terms interchangeably.

Again, there are a significant number of open questions in this area of research. For example, as corrected mean-field descriptions are less analytically tractable and more computationally intensive than mean-field descriptions, it would be of interest to determine the scenarios where the mean-field approximation is valid, and where the corrected mean-field approximation is required. Furthermore, there are a significant number of mechanisms that have yet to be examined with non-standard mean-field models. Hence there is potential to develop novel continuum models of collective behaviour that will provide insight into a variety of processes.

1.2 Research questions

Here we detail the research questions that motivate each component of this thesis. We briefly discuss the state of the literature before the investigation was performed, and how the investigation extends upon the collective knowledge contained within the literature. For a more in-depth discussion of the relevant literature for each research question, see the introduction in each chapter associated with a research question. The eight research questions proposed, and subsequently answered, in this thesis are as follows:

(1) Can the Fisher-Kolmogorov model be applied to IncuCyte ZOOM$^{\text{TM}}$ assay to obtain reliable estimates of cell carrying capacity density, cell diffusivity and cell proliferation rate?

Typically, IncuCyte ZOOM$^{\text{TM}}$ assays are interpreted in terms of the percentage of unoccupied space present in the assay, compared to the unoccupied space introduced into the cell monolayer at the beginning of the experiment [30, 60, 61]. Unfortunately, this measurement does not provide quantitative insight into the cellular mechanisms driving the spreading of the cell population into the unoccupied space, which is a key component of tumour spreading and tissue repair [10, 46, 47, 85]. Furthermore, it is difficult to robustly determine any potential change in the cellular mechanisms if a chemical treatment is applied to the cell population, which is critical to drug design [65, 87].

The Fisher-Kolmogorov model has been applied to models of collective cell behaviour previously [19, 40, 46, 47, 52, 62, 63, 66]. For example, Maini *et al.* [46, 47] obtain an estimate of cell diffusivity through a comparison of the position of the leading edge of the cell population and the long-time travelling wave speed arising from the Fisher-Kolmogorov model [52]. However, this approach does not produce unique estimates for the three parameters present in the model: cell diffusivity, cell proliferation rate and cell carrying capacity density [46, 47]. It is unclear which experimental measurements are required to obtain robust and unique estimates of these parameters from IncuCyte ZOOM$^{\mathrm{TM}}$ assays. Furthermore, it is unclear whether applying the Fisher-Kolmogorov model to a IncuCyte ZOOM$^{\mathrm{TM}}$ assay under different experimental conditions can provide quantitative estimates of the change in the cellular mechanisms due to, for example, the application of a putative chemical treatment. We propose a methodical approach that estimates each parameter in the model using a different experimental measurement. Specifically, we propose obtaining the cell carrying capacity density from late-time observations away from the leading edge of the cell population, the cell proliferation rate by measuring the evolution of the cell density away from the leading edge of the population, and the cell diffusivity by tracking the time evolution of the leading edge of the population.

(2) Does the position of the leading edge of a cell population in a scratch assay contain sufficient information to obtain robust estimates of cell proliferation rate and cell diffusivity?

Scratch assays are employed ubiquitously to observe the movement of cell fronts, which form a critical component of malignant spreading and tissue repair [58, 64, 82, 89]. These experiments are typically performed over a period of time that is less than 24 h, due to the cost associated with imaging experiments over a long period of time and the need to replenish nutrients if the experiment is sufficiently long [11, 58, 69]. The standard approach for scratch assays is to capture experimental images at a number of time points and measure the position of the leading edge using an image analysis tool such as ImageJ or Matlab's Image Processing Toolbox [84, 88]. However, the position of the leading edge, in isolation, does not provide insight into the cellular mechanisms driving the movement of the cell front.

One approach to obtain estimates of the processes affecting cell migration is to use a mathematical model. However, previous mathematical models either rely on expensive and time-consuming measurements [9, 63] or assume that the experimental time is sufficiently long to justify that the speed of the cell front has approached the long-time asymptotic wave speed arising from a reaction-diffusion equation [46, 47]. We propose a stochastic random walk model of cell movement and proliferation that mimics the geometry of the scratch assay and hence allows for the measurement of the position of the leading edge using the same image analysis techniques as the experiment. Furthermore, we propose dividing the experimental measurements into two intervals to obtain unique estimates of the cell diffusivity and cell proliferation rate. Cell proliferation occurs over a significantly longer time scale than cell motility [38, 47], which implies that at short time the movement of the cell front is driven entirely by cell motility. Hence we use short-time leading edge data to estimate the cell diffusivity and use the remaining data to obtain an estimate of the cell proliferation rate.

(3) Is approximate Bayesian computation a valid technique to extract information about the cell proliferation rate and the cell diffusivity from scratch assays; and, if so, which summary statistic provides the most information about the cell proliferation rate and the cell diffusivity?

Obtaining reliable parameter estimates from *in vitro* cell biology experiments, such as scratch assays, is critical to quantifying the influence of putative treatments of diseases such as chronic wounds and malignant tumour spreading [20, 27, 39, 86]. Unfortunately, the majority of scratch assays are interpreted qualitatively [41, 58]. Quantitative techniques have been applied to scratch assays but only provide point estimates of the parameters [79, 80]. Hence the uncertainty in the parameter estimates are associated with the number of experimental replicates performed, rather than the mathematical technique used to obtain the parameter estimates.

Approximate Bayesian computation (ABC) provides a parameter distribution that includes information about the uncertainty present in the parameter recovery process [48, 56, 83]. However, the use of ABC techniques for spatiotemporal processes is a relatively recent development and ABC has never been applied to scratch assays previously. We propose the use of a random walk model of cell motility and cell proliferation with geometry that mimics a scratch assay. ABC algorithms require a summary statistic, that is, an experimental measurement, to compare mathematical models with experimental data [48]. It is unclear which summary statistic provide sufficient information to obtain robust estimates of the cell motility and cell proliferation rate. The ideal choice of summary statistic is the location of each cell. However, due to the number of cells this summary statistic is too highly dimensional to be a computationally tractable summary statistic. The pair correlation function incorporates information about the distance between pairs of cells, as well as the number of cells, and may provide a suitable choice of summary statistic [8]. Hence an investigation into the choice of summary statistic and the corresponding information obtained about the cell motility and cell proliferation rate will provide insight.

(4) How does the choice of experimental design influence the amount of information obtained from a scratch assay about the cell proliferation rate and the cell diffusivity?

As performing and interpreting experiments is both time-consuming and expensive, there is a significant benefit in determining experimental designs that reduce the cost of the experiment while preserving sufficient information about the cell parameters. A combination of approximate Bayesian computation and an appropriate summary statistic provides robust estimates of cell motility and the cell proliferation rate [36]. However, it is unclear whether the experimental design choices can influence the parameter estimates obtained, or the amount of information obtained about the cell parameters. For example, it is unknown whether the same parameter estimates are obtained from a scratch assay with a single cell front or two opposingly-directed cell fronts results. Furthermore, even if the same parameter estimates are obtained, it is unclear whether one experimental design results in less uncertainty about the parameters, compared to a different experimental design. We propose an investigation into parameter estimates, and their corresponding uncertainty, for a suite of design choices to determine the experimental design that reduces experimental cost while still producing robust estimates of the cell parameters.

(5) When are mean-field descriptions of interacting subpopulations inadequate and, subsequently, when are corrected mean-field descriptions of interacting cell populations required?

Interacting subpopulations are present in many biological and ecological processes, such as the invasion of tumour cells through the stroma or the spread of an invasive species via predation of a local species [7, 15, 24, 32, 43, 55, 76]. The standard approach is to consider a mean-field partial different equation description of such processes [23, 54, 67, 71, 74, 77]. However, mean-field models are either phenomenologically-derived and hence do not consider individual-level interactions or, alternatively, rely on an implicit assumption that spatial correlations and local clustering do not have any impact on the population [29, 73]. In certain situations, the mean-field assumption is valid [70]. However, it is unclear in which situations, with respect to the relative ratios of birth, death and movement rates of the two subpopulations, the assumption is valid.

To address this, we propose a corrected mean-field description [6, 35, 70] that incorporates explicit information about the spatial correlations, and is derived from an individual-based model of two interacting subpopulations. Through comparisons of the corrected mean-field description, the standard mean-field description and the underlying discrete process we can obtain insight into the parameter regimes where a corrected mean-field description is required to accurately approximate the average discrete behaviour, and where the standard mean-field description is sufficient.

(6) Can lattice-based birth-death-movement processes be accurately described by modelling the dynamics of groups of contiguous occupied lattice sites and contiguous unoccupied lattice sites?

Lattice-based birth-death-movement random walk processes are widely used to describe biological, ecological and physical phenomena [13, 16, 17, 31, 33]. However, due to the stochastic nature of the random walk, it is computationally intensive to determine the average behaviour of the random walk and it is difficult to obtain insight into the relevant phenomenon through analysis of the random walk. To address these issues, deterministic continuum descriptions of random walk processes have been proposed [3, 18]. However, there are shortfalls associated with previously-proposed continuum descriptions [6, 35, 70]. Specifically, standard mean-field descriptions are only valid for processes that contain weak spatial correlations [6, 70], and corrected mean-field descriptions are not valid if the spatial correlations are significant [6, 70].

It is unclear whether the random walk process can be interpreted through the dynamics of a summary statistic, excluding the lattice occupancy. We propose interpreting the random walk through a model of the dynamics of groups of contiguous occupied lattices sites and contiguous unoccupied lattice sites. This approach avoids making assumptions about the spatial correlations in the system, as the dynamics of all possible lengths of contiguous occupied sites and contiguous unoccupied sites are considered. It is instructive to examine whether this interpretation provides a valid approximation of the average random walk behaviour, particularly in parameter regimes where previous deterministic continuum descriptions are invalid.

(7) How does introducing competitive/co-operative mechanisms to lattice-based birth-death-movement processes change the corresponding mean-field description?

Many biological and ecological processes exhibit either co-operative or competitive behaviour, where there is either a positive or negative benefit associated with being part of, or separate from, the bulk population [5, 14, 28, 57, 81]. For example, if an individual is separate from the bulk population, the birth rate for that individual can be reduced due to the difficulty associated with finding a mate [14, 81]. Conversely, an individual that is separate from the bulk population may be less susceptible to eradication efforts if an isolated individual is less likely to be detected, compared to the bulk population [28, 81].

Introducing birth, death and movement parameters that are dependent on the number of occupied nearest-neighbour sites in a lattice-based random walk process allows for modelling of both competitive and co-operative mechanisms. Without this occupancy dependence, competitive mechanisms are restricted to logistic growth, where the birth rate decreases with density. It is unclear how introducing these new mechanisms influences the mean-field description of the population. It is therefore instructive to investigate the form of the mean-field description, and the corresponding predictions of the population behaviour obtained from the mean-field description in a suite of parameter regimes.

(8) **Can modelling the dynamics of groups of contiguous occupied lattice sites and contiguous unoccupied lattice sites provide a reliable prediction of the persistence or extinction of an invasive population with competitive/co-operative mechanisms?**

Moving fronts occur ubiquitously in biological and ecological processes, such as glioma cells moving through the brain to form a glioblastoma and the spread of an invasive species [2, 12, 22, 25, 26, 50, 53, 55, 57, 68, 80]. Obtaining reliable predictions of the direction and speed of the moving front is critical for informing biological and ecological control measures. Lattice-based random walk processes provide a flexible framework for modelling moving fronts [16, 21, 33, 37, 72], and allow for the implementation of birth, death and movement parameters that depend on the size of the group of contiguous occupied sites that an agent belongs to. Deterministic continuum descriptions have been proposed for certain cases, specifically, the case where the parameters are independent of group size and where the parameters depend on whether an agent has zero, or at least one nearest-neighbour agent [34, 72].

However, these continuum descriptions rely on the spatial correlations in the system being sufficiently weak to obtain accurate estimates of the front velocity [70]. Hence, in many parameter regimes, previous continuum descriptions provide an inaccurate estimate of the front velocity. Describing lattice-based random walks in terms of the dynamics of groups of contiguous occupied sites and contiguous unoccupied sites provides an accurate approximation of the average random walk behaviour in all cases for a process that is, on average, spatially uniform. It is unknown whether a similar description provides an accurate approximation of the average random walk behaviour for processes containing a moving front. Hence it is instructive to examine the predictions about the front velocity, as well as the long-time persistence or extinction of the population, compared to the average random walk behaviour.

The work contained in this thesis involves investigating and subsequently addressing these unanswered questions.

1.3 Aims and outcomes of this thesis

The main aims of this thesis are two-fold. First, to use a combination of mathematical modelling, experimental techniques and parameter recovery approaches to obtain quantitative insight into the cellular mechanisms governing collective cell behaviour. Second, to use mathematical modelling techniques to develop novel continuum descriptions of collective behaviour from individual-based models of motility, birth and death.

Each of these main aims is compromised of four sub-aims. The first aim consists of the following four sub-aims:

- Determine the requisite experimental measurements and procedure for applying the Fisher-Kolmogorov model to relevant experimental data such that reliable estimates of the cell carrying capacity, cell diffusivity and cell proliferation rate can be obtained from an IncuCyte ZOOM$^{\mathrm{TM}}$ assay.
- Analyse the information contained in the position of the leading edge of the cell population in a scratch assay to determine whether robust estimates of the cell diffusivity and cell proliferation rate can be obtained.
- Investigate whether approximate Bayesian computation is a valid technique for estimating cell parameters from experimental data obtained from a scratch assay and, if so, determine which experimental measurements contain sufficient information for robust parameter estimates.
- Quantify the change in parameter estimates, and uncertainty associated with the parameter estimates, due to experimental design choices in a scratch assay using approximate Bayesian computation.

The second aim consists of the following four sub-aims:

- Develop a corrected mean-field description for multiple interacting subpopulations to obtain insight into the parameter regimes where corrected mean-field descriptions are required to accurately approximate lattice-based random walk models of multiple interacting subpopulations.
- Determine whether modelling the dynamics of groups of contiguous occupied lattice sites and contiguous unoccupied lattice sites provides a valid approximation of the average behaviour of birth-death-movement processes, in particular for parameter regimes where mean-field and corrected mean-field descriptions are inaccurate.
- Investigate how introducing competitive and co-operative mechanisms into lattice-based birth-death-movement processes changes the standard mean-field description of the process, and the implications on long-time behaviour for the population.
- Develop a model for the dynamics of groups of contiguous occupied lattice sites and contiguous unoccupied lattice sites for an invasive population and compare the predictions for the long-time behaviour of the population with predictions obtained from previously-proposed continuum descriptions.

This thesis is prepared and presented by published papers and contains eight peer-reviewed papers that have been published in international journals. In all eight journal articles the candidate is the first author and hence contributed significantly to all articles. The work presented here satisfies the requirements of three published and/or submitted papers for the Queensland University Technology to award a thesis by published papers.

This thesis consists of the following eight publications:

- **Johnston, S. T.**, Shah, E. T., Chopin, L. K., McElwain, D. L. S. and Simpson, M. J. [2015], 'Estimating cell diffusivity and cell proliferation rate by interpreting

IncuCyte ZOOM$^{\text{TM}}$ assay data using the Fisher-Kolmogorov model', *BMC Systems Biology* **9**(1), 38. (Chapter 2.)

- **Johnston, S. T.**, Simpson, M. J. and McElwain, D. L. S. [2014], 'How much information can be obtained from tracking the position of the leading edge in a scratch assay?', *Journal of the Royal Society Interface* **11**(97), 20140325. (Chapter 3.)
- **Johnston, S. T.**, Simpson, M. J., McElwain, D. L. S., Binder, B. J. and Ross, J. V. [2014], 'Interpreting scratch assays using pair density dynamics and approximate Bayesian computation', *Open Biology* **4**(9), 140097. (Chapter 4.)
- **Johnston, S. T.**, Ross, J. V., Binder, B. J., McElwain, D. L. S., Haridas, P. and Simpson, M. J. [2016], 'Quantifying the effect of experimental design choices for *in vitro* scratch assays', *Journal of Theoretical Biology* **400**, 19-31. (Chapter 5.)
- **Johnston, S. T.**, Simpson, M. J. and Baker, R. E. [2015], 'Modelling the movement of interacting cell populations: A moment dynamics approach', *Journal of Theoretical Biology* **370**, 81-92. (Chapter 6.)
- **Johnston, S. T.**, Baker, R. E. and Simpson, M. J. [2016], 'Filling the gaps: A robust description of adhesive birth-death-movement processes', *Physical Review E* **93**(4), 042413. (Chapter 7.)
- **Johnston, S. T.**, Baker, R. E., McElwain, D. L. S. and Simpson, M. J. [2017], 'Co-operation, competition and crowding: A discrete framework linking Allee kinetics, nonlinear diffusion, shocks and sharp-fronted travelling waves', *Scientific Reports* **7**, 421134. (Chapter 8.)
- **Johnston, S. T.**, Baker, R. E. and Simpson, M. J. [2017], 'A new and accurate continuum description of moving fronts', *New Journal of Physics* **19**(3), 033010. (Chapter 9.)

As each chapter of this thesis contains a previously-published paper, barring the introductory and concluding chapters, this thesis is designed such that each chapter can be read independently. The formatting from the published papers has been standardised such that each chapter is visually consistent. However, the content of the published papers has been reproduced without change and hence overlaps in ideas for models, experiments and literature evaluations may occur. Each chapter consists of the following sections: an introductory section, containing a critical evaluation of the relevant literature; a models section, containing an explanation of the models and experiments considered in the chapter; a results section, detailing the outcomes obtained from the relevant experimental and modelling approaches, and; a concluding section, discussing the results in the context of the literature and providing potential future directions for research.

1.4 Structure of this thesis

This thesis is comprised of four major areas: an introductory chapter, four chapters relating to original research focused on the application of mathematical models to experimental data, four chapters relating to original research focused on developing novel continuum models of collective cell behaviour, and a concluding chapter.

In this introductory chapter we summarise and review the relevant literature relating to mathematical models of collective cell behaviour, techniques for the application of such mathematical models to experimental data and continuum descriptions of lattice-based random walk models.

1.5 Statement of joint authorship

Here we detail the contribution of the PhD candidate and the co-authors for each of the eight articles comprising this thesis. All co-authors have consented to the inclusion of each relevant article. To provide further detail about the levels of contribution we highlight in italics the co-author that contributed most to a particular area of contribution.

Chapter 2: Estimating cell diffusivity and cell proliferation rate by interpreting IncuCyte ZOOM$^{\mathrm{TM}}$ assay data using the Fisher-Kolmogorov model

For this chapter, the associated published reference is:

Johnston, S. T., Shah, E. T., Chopin, L. K., McElwain, D. L. S. and Simpson, M. J. [2015], 'Estimating cell diffusivity and cell proliferation rate by interpreting IncuCyte ZOOM$^{\mathrm{TM}}$ assay data using the Fisher-Kolmogorov model', *BMC Systems Biology* **9**(1), 38.

Statement of contribution:

- **Johnston, S. T. (Candidate)** conceived the study, designed the experiments, *performed the experiments*, *analysed the data*, *wrote the manuscript*, and *edited and approved the final manuscript*.
- Shah, E. T. conceived the study, designed the experiments, *performed the experiments*, and edited and approved the final manuscript.
- Chopin, L. K. conceived the study, *designed the experiments*, and edited and approved the final manuscript.
- McElwain, D. L. S. conceived the study, designed the experiments, analysed the data, and edited and approved the final manuscript.
- Simpson, M. J. *conceived the study*, designed the experiments, analysed the data, wrote the manuscript, edited and approved the final manuscript, and *acted as corresponding author*.

Chapter 3: How much information can be obtained from tracking the position of the leading edge in a scratch assay?

For this chapter, the associated published reference is:

Johnston, S. T., Simpson, M. J. and McElwain, D. L. S. [2014], 'How much information can be obtained from tracking the position of the leading edge in a scratch assay?', *Journal of the Royal Society Interface* **11**(97), 20140325.

Statement of contribution:

- **Johnston, S. T. (Candidate)** conceived the study, designed the experiments, *performed the experiments*, *analysed the data*, *wrote the manuscript*, and edited and approved the final manuscript.
- Simpson, M. J. *conceived the study*, *designed the experiments*, analysed the data, wrote the manuscript, *edited and approved the final manuscript*, and *acted as corresponding author.*
- McElwain, D. L. S. conceived the study, designed the experiments, analysed the data, and edited and approved the final manuscript.

Chapter 4: Interpreting scratch assays using pair density dynamics and approximate Bayesian computation

For this chapter, the associated published reference is:

Johnston, S. T., Simpson, M. J., McElwain, D. L. S., Binder, B. J. and Ross, J. V. [2014], 'Interpreting scratch assays using pair density dynamics and approximate Bayesian computation', *Open Biology* **4**(9), 140097.

Statement of contribution:

- **Johnston, S. T. (Candidate)** conceived the study, designed the experiments, *performed the experiments*, *analysed the data*, *wrote the manuscript*, and *edited and approved the final manuscript.*
- Simpson, M. J. conceived the study, *designed the experiments*, analysed the data, wrote the manuscript, edited and approved the final manuscript, and *acted as corresponding author.*
- McElwain, D. L. S. conceived the study, designed the experiments, and edited and approved the final manuscript.
- Binder, B. J. conceived the study, designed the experiments, and edited and approved the final manuscript.
- Ross, J. V. *conceived the study*, designed the experiments, analysed the data, and edited and approved the final manuscript.

Chapter 5: Quantifying the effect of experimental design choices for *in vitro* scratch assays

For this chapter, the associated published reference is:

Johnston, S. T., Ross, J. V., Binder, B. J., McElwain, D. L. S., Haridas, P. and Simpson, M. J. [2016], 'Quantifying the effect of experimental design choices for *in vitro* scratch assays', *Journal of Theoretical Biology* **400**, 19-31.

Statement of contribution:

- **Johnston, S. T. (Candidate)** conceived the study, designed the experiments, *performed the experiments*, *analysed the data*, *wrote the manuscript*, and *edited and approved the final manuscript.*
- Ross, J. V. *conceived the study*, designed the experiments, analysed the data, and edited and approved the final manuscript.
- Binder, B. J. conceived the study, designed the experiments, and edited and approved the final manuscript.
- McElwain, D. L. S. conceived the study, designed the experiments, and edited and approved the final manuscript.
- Haridas, P. conceived the study, designed the experiments, *performed the experiments*, and edited and approved the final manuscript.
- Simpson, M. J. conceived the study, *designed the experiments*, analysed the data, wrote the manuscript, edited and approved the final manuscript, and *acted as corresponding author.*

Chapter 6: Modelling the movement of interacting cell populations: A moment dynamics approach

For this chapter, the associated published reference is:

Johnston, S. T., Simpson, M. J. and Baker, R. E. [2015], 'Modelling the movement of interacting cell populations: A moment dynamics approach', *Journal of Theoretical Biology* **370**, 81-92.

Statement of contribution:

- **Johnston, S. T. (Candidate)** conceived the study, designed the experiments, *performed the experiments*, *analysed the data*, *wrote the manuscript*, *edited and approved the final manuscript*, and *acted as corresponding author.*
- Simpson, M. J. *conceived the study*, *designed the experiments*, analysed the data, wrote the manuscript, and edited and approved the final manuscript.
- Baker, R. E. conceived the study, designed the experiments, and edited and approved the final manuscript.

Chapter 7: Filling the gaps: A robust description of adhesive birth-death-movement processes

For this chapter, the associated published reference is:

Johnston, S. T., Baker, R. E. and Simpson, M. J. [2016], ‘Filling the gaps: A robust description of adhesive birth-death-movement processes’, *Physical Review E* **93**(4), 042413.

Statement of contribution:

- **Johnston, S. T. (Candidate)** *conceived the study*, *designed the experiments*, *performed the experiments*, *analysed the data*, *wrote the manuscript*, edited and approved the final manuscript, and *acted as corresponding author.*
- Baker, R. E. conceived the study, and edited and approved the final manuscript.
- Simpson, M. J. conceived the study, analysed the data, wrote the manuscript, and *edited and approved the final manuscript.*

Chapter 8: Co-operation, competition and crowding: A discrete framework linking Allee kinetics, nonlinear diffusion, shocks and sharp-fronted travelling waves

For this chapter, the associated published reference is:

Johnston, S. T., Baker, R. E., McElwain, D. L. S. and Simpson, M. J. [2017], ‘Co-operation, competition and crowding: A discrete framework linking Allee kinetics, nonlinear diffusion, shocks and sharp-fronted travelling waves’, *Scientific Reports* **7**, 421134.

Statement of contribution:

- **Johnston, S. T. (Candidate)** *conceived the study*, *designed the experiments*, *performed the experiments*, *analysed the data*, *wrote the manuscript*, and *edited and approved the final manuscript.*
- Baker, R. E. conceived the study, designed the experiments, and edited and approved the final manuscript.
- McElwain, D. L. S. conceived the study, designed the experiments, analysed the data, and edited and approved the final manuscript.
- Simpson, M. J. conceived the study, designed the experiments, analysed the data, wrote the manuscript, edited and approved the final manuscript, and *acted as corresponding author.*

Chapter 9: A new and accurate continuum description of moving fronts

For this chapter, the associated published reference is:

Johnston, S. T., Baker, R. E. and Simpson, M. J. [2017], ‘A new and accurate continuum description of moving fronts’, *New Journal of Physics* **19**(3), 033010.

Statement of contribution:

- **Johnston, S. T. (Candidate)** *conceived the study*, *designed the experiments*, *performed the experiments*, *analysed the data*, *wrote the manuscript*, and *edited and approved the final manuscript*.
- Baker, R. E. edited and approved the final manuscript.
- Simpson, M. J. analysed the data, wrote the manuscript, edited and approved the final manuscript, and *acted as corresponding author*.

Bibliography

[1] Aichele, K., Bubel, M., Deubel, G., Pohlemann, T. and Oberringer, M. [2013], 'Bromelain down-regulates myofibroblast differentiation in an in vitro wound healing assay', *Naunyn-Schmiedeberg's Archives of Pharmacology* **386**, 853–763.

[2] Alvord Jr, E. C. and Shaw, C. M. [1991], *Neoplasms affecting the nervous system of the elderly*, Oxford University Press.

[3] Anguige, K. and Schmeiser, C. [2009], 'A one-dimensional model of cell diffusion and aggregation, incorporating volume filling and cell-to-cell adhesion', *Journal of Mathematical Biology* **58**(3), 395–427.

[4] Arciero, J. C., Mi, Q., Branca, M. F., Hackam, D. J. and Swigon, D. [2011], 'Continuum model of collective cell migration in wound healing and colony expansion', *Biophysical Journal* **100**(3), 535–543.

[5] Axelrod, R., Axelrod, D. E. and Pienta, K. J. [2006], 'Evolution of cooperation among tumor cells', *Proceedings of the National Academy of Sciences* **103**(36), 13474–13479.

[6] Baker, R. E. and Simpson, M. J. [2010], 'Correcting mean-field approximations for birth-death-movement processes', *Physical Review E* **82**(4), 041905.

[7] Bhowmick, N. A. and Moses, H. L. [2005], 'Tumor–stroma interactions', *Current Opinion in Genetics & Development* **15**(1), 97–101.

[8] Binder, B. J. and Simpson, M. J. [2013], 'Quantifying spatial structure in experimental observations and agent-based simulations using pair-correlation functions', *Physical Review E* **88**(2), 022705.

[9] Bloomfield, J.-M., Sherratt, J.-A., Painter, K.-J. and Landini, G. [2010], 'Cellular automata and integrodifferential equation models for cell renewal in mosaic tissues', *Journal of the Royal Society Interface* **7**(52), 1525–1535.

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

[11] Chigurupati, S., Mughal, M. R., Okun, E., Das, S., Kumar, A., McCaffery, M., Seal, S. and Mattson, M. P. [2012], 'Effects of cerium oxide nanoparticles on the growth of keratinocytes, fibroblasts and vascular endothelial cells in cutaneous wound healing', *Biomaterials* **34**, 2194–2201.

[12] Clavero, M. and García-Berthou, E. [2005], 'Invasive species are a leading cause of animal extinctions', *Trends in Ecology and Evolution* **20**(3), 110–110.

[13] Codling, E. A., Plank, M. J. and Benhamou, S. [2008], 'Random walk models in biology', *Journal of the Royal Society Interface* **5**, 813–834.

[14] Courchamp, F., Clutton-Brock, T. and Grenfell, B. [1999], ‘Inverse density dependence and the Allee effect’, *Trends in Ecology and Evolution* **14**(10), 405–410.

[15] De Wever, O. and Mareel, M. [2003], ‘Role of tissue stroma in cancer cell invasion’, *The Journal of Pathology* **200**(4), 429–447.

[16] Deroulers, C., Aubert, M., Badoual, M. and Grammaticos, B. [2009], ‘Modeling tumor cell migration: from microscopic to macroscopic models’, *Physical Review E* **79**(3), 031917.

[17] Derrida, B., Evans, M., Hakim, V. and Pasquier, V. [1993], ‘Exact solution of a 1D asymmetric exclusion model using a matrix formulation’, *Journal of Physics A* **26**(7), 1493.

[18] Fernando, A. E., Landman, K. A. and Simpson, M. J. [2010], ‘Nonlinear diffusion and exclusion processes with contact interactions’, *Physical Review E* **81**(1), 011903.

[19] Fisher, R. A. [1937], ‘The wave of advance of advantageous genes’, *Annals of Eugenics* **7**(4), 355–369.

[20] Fishman, D. A., Liu, Y., Ellerbroek, S. M. and Stack, M. S. [2001], ‘Lysophosphatidic acid promotes matrix metalloproteinase (MMP) activation and MMP-dependent invasion in ovarian cancer cells’, *Cancer Research* **61**(7), 3194–3199.

[21] Fort, J. and Pujol, T. [2007], ‘Time-delayed fronts from biased random walks’, *New Journal of Physics* **9**(7), 234.

[22] Gabriely, G., Wurdinger, T., Kesari, S., Esau, C. C., Burchard, J., Linsley, P. S. and Krichevsky, A. M. [2008], ‘MicroRNA 21 promotes glioma invasion by targeting matrix metalloproteinase regulators’, *Molecular and Cellular Biology* **28**(17), 5369–5380.

[23] Gatenby, R. A. and Gawlinski, E. T. [1996], ‘A reaction-diffusion model of cancer invasion’, *Cancer Research* **56**(24), 5745–5753.

[24] Gatenby, R. A., Gawlinski, E. T., Gmitro, A. F., Kaylor, B. and Gillies, R. J. [2006], ‘Acid-mediated tumor invasion: a multidisciplinary study’, *Cancer Research* **66**(10), 5216–5223.

[25] Gerlee, P. and Nelander, S. [2012], ‘The impact of phenotypic switching on glioblastoma growth and invasion’, *PLOS Computational Biology* **8**(6), e1002556.

[26] Gerlee, P. and Nelander, S. [2016], ‘Travelling wave analysis of a mathematical model of glioblastoma growth’, *Mathematical Biosciences* **276**, 75–81.

[27] Gingis-Velitski, S., Loven, D., Benayoun, L., Munster, M., Bril, R., Voloshin, T., Alishekevitz, D., Bertolini, F. and Shaked, Y. [2011], ‘Host response to short-term, single-agent chemotherapy induces matrix metalloproteinase-9 expression and accelerates metastasis in mice’, *Cancer Research* **71**(22), 6986–6996.

[28] Godin, J. G. J. [1986], ‘Antipredator function of shoaling in teleost fishes: a selective review’, *Le Naturaliste Canadien* **113**, 241–250.

[29] Grima, R. [2008], ‘Multiscale modeling of biological pattern formation’, *Current Topics in Developmental Biology* **81**, 435–460.

[30] Gujral, T. S., Chan, M., Peshkin, L., Sorger, P. K., Kirschner, M. W. and MacBeath, G. [2014], ‘A noncanonical frizzled2 pathway regulates epithelial-mesenchymal transition and metastasis’, *Cell* **159**, 844–856.

[31] Hackett-Jones, E. J., Landman, K. A. and Fellner, K. [2012], 'Aggregation patterns from nonlocal interactions: Discrete stochastic and continuum modeling', *Physical Review E* **85**(4), 041912.

[32] Hastings, A., Cuddington, K., Davies, K. F., Dugaw, C. J., Elmendorf, S., Freestone, A., Harrison, S., Holland, M., Lambrinos, J., Malvadkar, U. et al. [2005], 'The spatial spread of invasions: new developments in theory and evidence', *Ecology Letters* **8**(1), 91–101.

[33] Illien, P., Bénichou, O., Mejía-Monasterio, C., Oshanin, G. and Voituriez, R. [2013], 'Active transport in dense diffusive single-file systems', *Physical Review Letters* **111**(3), 038102.

[34] Johnston, S. T., Baker, R. E. and Simpson, M. J. [2016], 'Filling the gaps: A robust description of adhesive birth-death-movement processes', *Physical Review E* **93**(4), 042413.

[35] Johnston, S. T., Simpson, M. J. and Baker, R. E. [2012], 'Mean-field descriptions of collective migration with strong adhesion', *Physical Review E* **85**(5), 051922.

[36] Johnston, S. T., Simpson, M. J., McElwain, D. L. S., Binder, B. J. and Ross, J. V. [2014], 'Interpreting scratch assays using pair density dynamics and approximate Bayesian computation', *Open Biology* **4**, 140097.

[37] Khain, E., Katakowski, M., Charteris, N., Jiang, F. and Chopp, M. [2012], 'Migration of adhesive glioma cells: Front propagation and fingering', *Physical Review E* **86**(1), 011904.

[38] Khain, E., Katakowski, M., Hopkins, S., Szalad, A., Zheng, X., Jiang, F. and Chopp, M. [2011], 'Collective behavior of brain tumor cells: the role of hypoxia', *Physical Review E* **83**(3), 031920.

[39] Knecht, D. A., LaFleur, R. A., Kahsai, A. W., Argueta, C. E., Beshir, A. B. and Fenteany, G. [2010], 'Cucurbitacin I inhibits cell motility by indirectly interfering with actin dynamics', *PLOS ONE* **5**(11), e14039.

[40] Kolmogorov, A., Petrovsky, A. and Piscounov, N. [1937], 'Etude de lequation de la diffusion avec croissance de la quantite de matiere et son application a unprobleme biologique.', *Moscow University Mathematics Bulletin* **1**, 1–25.

[41] Landesberg, R., Cozin, M., Cremers, S., Woo, V., Kousteni, S., Sinha, S., Garrett-Sinha, L. and Raghavan, S. [2008], 'Inhibition of oral mucosal cell wound healing by bisphosphonates', *Journal of Oral and Maxillofacial Surgery* **66**(5), 839–847.

[42] Law, R. and Dieckmann, U. [2000], 'A dynamical system for neighborhoods in plant communities', *Ecology* **81**(8), 2137–2148.

[43] Li, G., Satyamoorthy, K., Meier, F., Berking, C., Bogenrieder, T. and Herlyn, M. [2003], 'Function and regulation of melanoma–stromal fibroblast interactions: when seeds meet soil', *Oncogene* **22**(20), 3162–3171.

[44] Liang, C.-C., Park, A. Y. and Guan, J.-L. [2007], 'In vitro scratch assay: a convenient and inexpensive method for analysis of cell migration in vitro', *Nature Protocols* **2**(2), 329–333.

[45] Mackie, A. D., Panagiotopoulos, A. Z. and Szleifer, I. [1997], 'Aggregation behavior of a lattice model for amphiphiles', *Langmuir* **13**(19), 5022–5031.

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

[47] Maini, P. K., McElwain, D.-L. S. and Leavesley, D. I. [2004*b*], 'Travelling waves in a wound healing assay', *Applied Mathematics Letters* **17**(5), 575–580.

[48] Marjoram, P., Molitor, J., Plagnol, V. and Tavaré, S. [2003], 'Markov chain Monte Carlo without likelihoods', *Proceedings of the National Academy of Sciences* **100**(26), 15324–15328.

[49] Markham, D. C., Simpson, M. J., Maini, P. K., Gaffney, E. A. and Baker, R. E. [2013], 'Incorporating spatial correlations into multispecies mean-field models', *Physical Review E* **88**(5), 052713.

[50] Martin, P. [1997], 'Wound healing–aiming for perfect skin regeneration', *Science* **276**(5309), 75–81.

[51] McKane, A. J. and Newman, T. J. [2004], 'Stochastic models in population biology and their deterministic analogs', *Physical Review E* **70**(4), 041902.

[52] Murray, J. D. [2002], *Mathematical Biology I: An Introduction*, Springer, New York.

[53] Nnetu, K. D., Knorr, M., Käs, J. and Zink, M. [2012], 'The impact of jamming on boundaries of collectively moving weak-interacting cells', *New Journal of Physics* **14**(11), 115012.

[54] Painter, K. J. and Sherratt, J. A. [2003], 'Modelling the movement of interacting cell populations', *Journal of Theoretical Biology* **225**(3), 327–339.

[55] Phillips, B. L., Brown, G. P., Greenlees, M., Webb, J. K. and Shine, R. [2007], 'Rapid expansion of the cane toad (bufo marinus) invasion front in tropical australia', *Austral Ecology* **32**(2), 169–176.

[56] Plagnol, V. and Tavaré, S. [2004], Approximate Bayesian computation and MCMC, *in* 'Monte Carlo and Quasi-Monte Carlo Methods 2002', Springer, pp. 99–113.

[57] Poujade, M., Grasland-Mongrain, E., Hertzog, A., Jouanneau, J., Chavrier, P., Ladoux, B., Buguin, A. and Silberzan, P. [2007], 'Collective migration of an epithelial monolayer in response to a model wound', *Proceedings of the National Academy of Sciences* **104**(41), 15988–15993.

[58] Rausch, V., Liu, L., Apel, A., Rettig, T., Gladkich, J., Labsch, S., Kallifatidis, G., Kaczorowski, A., Groth, A., Gross, W. et al. [2012], 'Autophagy mediates survival of pancreatic tumour-initiating cells in a hypoxic microenvironment', *The Journal of Pathology* **227**(3), 325–335.

[59] Riahi, R., Yang, Y., Zhang, D. D. and Wong, P. K. [2012], 'Advances in wound-healing assays for probing collective cell migration', *Journal of Laboratory Automation* **17**(1), 59–65.

[60] Salomon, C., Kobayashi, M., Ashman, K., Sobrevia, L., Mitchell, M. D. and Rice, G. E. [2013], 'Hypoxia-induced changes in the bioactivity of cytotrophoblast-derived exosomes', *PLOS ONE* **8(11)**, e79636.

[61] Salomon, C., Torres, M. J., Kobayashi, M., Scholz-Romero, K., Sobrevia, L., Dobierzewska, A., Illanes, S. E., Mitchell, M. D. and Rice, G. E. [2014], 'A gestational

profile of placental exosomes in maternal plasma and their effects on endothelial cell migration', *PLOS ONE* **9(6)**, e98667.

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

[63] Sengers, B. G., Please, C. P. and Oreffo, R. O.-C. [2007], 'Experimental characterization and computational modelling of two-dimensional cell spreading for skeletal regeneration', *Journal of the Royal Society Interface* **4**(17), 1107–1117.

[64] Sharili, A.-S., Allen, S., Smith, K., Price, J. and McGonnell, I. M. [2013], 'Snail2 promotes osteosarcoma cell motility through remodelling of the actin cytoskeleton and regulates tumor development', *Cancer Letters* **333**, 170–179.

[65] Shatkin, A.-J., Reich, E., Franklin, R.-M. and Tatum, E.-L. [1962], 'Effect of mitomycin-c on mammalian cells in culture', *Biochimica et Biophysica Acta* **55**(3), 277–289.

[66] Sheardown, H. and Cheng, Y.-L. [1996], 'Mechanisms of corneal epithelial wound healing', *Chemical Engineering Science* **51**(19), 4517–4529.

[67] Sherratt, J. A. [2000], 'Wavefront propagation in a competition equation with a new motility term modelling contact inhibition between cell populations', *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* **456**(2002), 2365–2386.

[68] Sherratt, J. A., Lewis, M. A. and Fowler, A. C. [1995], 'Ecological chaos in the wake of invasion.', *Proceedings of the National Academy of Sciences* **92**(7), 2524–2528.

[69] Shibata, A., Tanabe, E., Inoue, S., Kitayoshi, M., Okimoto, S., Hirane, M., Araki, M., Fukushima, N. and Tsujiuchi, T. [2013], 'Hydrogen peroxide stimulates cell motile activity through LPA receptor-3 in liver epithelial WB-F344 cells', *Biochemical and Biophysical Research Communications* **433**, 317–321.

[70] Simpson, M. J. and Baker, R. E. [2011], 'Corrected mean-field models for spatially dependent advection-diffusion-reaction phenomena', *Physical Review E* **83**, 051922.

[71] Simpson, M. J., Landman, K. A. and Bhaganagarapu, K. [2007], 'Coalescence of interacting cell populations', *Journal of Theoretical Biology* **247**(3), 525–543.

[72] Simpson, M. J., Landman, K. A. and Hughes, B. D. [2010], 'Cell invasion with proliferation mechanisms motivated by time-lapse data', *Physica A: Statistical Mechanics and its Applications* **389**(18), 3779–3790.

[73] Simpson, M. J., Treloar, K. K., Binder, B. J., Haridas, P., Manton, K. J., Leavesley, D. I., McElwain, D.-L. S. and Baker, R. E. [2013], 'Quantifying the roles of cell motility and cell proliferation in a circular barrier assay', *Journal of the Royal Society Interface* **10**(82), 20130007.

[74] Simpson, M. J., Zhang, D. C., Mariani, M., Landman, K. A. and Newgreen, D. F. [2007], 'Cell proliferation drives neural crest cell invasion of the intestine', *Developmental Biology* **302**(2), 553–568.

[75] Singer, A. [2004], 'Maximum entropy formulation of the Kirkwood superposition approximation', *The Journal of Chemical Physics* **121**(8), 3657–3666.

[76] Skellam, J. [1951], 'Random dispersal in theoretical populations', *Biometrika* pp. 196–218.

[77] Smallbone, K., Gavaghan, D. J., Gatenby, R. A. and Maini, P. K. [2005], 'The role of acidity in solid tumour growth and invasion', *Journal of Theoretical Biology* **235**(4), 476–484.

[78] Stetler-Stevenson, W. G., Aznavoorian, S. and Liotta, L. A. [1993], 'Tumor cell interactions with the extracellular matrix during invasion and metastasis', *Annual Review of Cell Biology* **9**(1), 541–573.

[79] Swanson, K. R. [2008], 'Quantifying glioma cell growth and invasion in vitro', *Mathematical and Computer Modelling* **47**(5), 638–648.

[80] Swanson, K. R., Bridge, C., Murray, J.-D. and Alvord Jr, E. C. [2003], 'Virtual and real brain tumors: using mathematical modeling to quantify glioma growth and invasion', *Journal of the Neurological Sciences* **216**(1), 1–10.

[81] Taylor, C. M. and Hastings, A. [2005], 'Allee effects in biological invasions', *Ecology Letters* **8**(8), 895–908.

[82] Teppo, S., Sundquist, E., Vered, M., Holappa, H., Parkkisenniemi, J., Rinaldi, T., Lehenkari, P., Grenman, R., Dayan, D., Risteli, J. et al. [2012], 'The hypoxic tumor microenvironment regulates invasion of aggressive oral carcinoma cells', *Experimental Cell Research* **319**, 376–389.

[83] Toni, T. and Stumpf, M. P. [2010], 'Simulation-based model selection for dynamical systems in systems and population biology', *Bioinformatics* **26**(1), 104–110.

[84] Treloar, K. K. and Simpson, M. J. [2013], 'Sensitivity of edge detection methods for quantifying cell migration assays', *PLOS ONE* **8**(6), e67389.

[85] Tremel, A., Cai, A., Tirtaatmadja, N., Hughes, B.-D., Stevens, G.-W., Landman, K.-A. and O'Connor, A.-J. [2009], 'Cell migration and proliferation during monolayer formation and wound healing', *Chemical Engineering Science* **64**(2), 247–253.

[86] Upton, Z., Wallace, H. J., Shooter, G. K., van Lonkhuyzen, D. R., Yeoh-Ellerton, S., Rayment, E. A., Fleming, J. M., Broszczak, D., Queen, D., Sibbald, R. G. et al. [2011], 'Human pilot studies reveal the potential of a vitronectin: growth factor complex as a treatment for chronic wounds', *International Wound Journal* **8**(5), 522–532.

[87] Weissberg, J. B., Son, Y. H., Papac, R. J., Sasaki, C., Fischer, D. B., Lawrence, R., Rockwell, S., Sartorelli, A. C. and Fischer, J. J. [1989], 'Randomized clinical trial of mitomycin-c as an adjunct to radiotherapy in head and neck cancer', *International Journal of Radiation Oncology* **17**(1), 3–9.

[88] Zaritsky, A., Manor, N., Wolf, L., Ben-Jacob, E. and Tsarfaty, I. [2013], 'Benchmark for multi-cellular segmentation of bright field microscopy images', *BMC Bioinformatics* **14**(1), 319.

[89] Zhao, X., Li, D.-C., Zhu, X.-G., Gan, W.-J., Li, Z., Xiong, F., Zhang, Z.-X., Zhang, G.-B., Zhang, X.-G. and Zhao, H. [2013], 'B7-H3 overexpression in pancreatic cancer promotes tumor progression', *International Journal of Molecular Medicine* **31**(2), 283–291.

Part I

Application of mathematical models to experimental data

Chapter 2

Estimating cell diffusivity and cell proliferation rate by interpreting IncuCyte ZOOM$^{\mathrm{TM}}$ assay data using the Fisher-Kolmogorov model

A paper published in *BMC Systems Biology.*

Johnston, S. T., Shah, E. T., Chopin, L. K., McElwain, D. L. S. and Simpson, M. J. [2015], 'Estimating cell diffusivity and cell proliferation rate by interpreting IncuCyte ZOOM$^{\mathrm{TM}}$ assay data using the Fisher-Kolmogorov model', *BMC Systems Biology* **9**(1), 38.

Abstract

Standard methods for quantifying IncuCyte ZOOM$^{\mathrm{TM}}$ assays involve measurements that quantify how rapidly the initially-vacant area becomes re-colonised with cells as a function of time. Unfortunately, these measurements give no insight into the details of the cellular-level mechanisms acting to close the initially-vacant area. We provide an alternative method enabling us to quantify the role of cell motility and cell proliferation separately. To achieve this we calibrate standard data available from IncuCyte ZOOM$^{\mathrm{TM}}$ images to the solution of the Fisher-Kolmogorov model. The Fisher-Kolmogorov model is a reaction-diffusion equation that has been used to describe collective cell spreading driven by cell migration, characterised by a cell diffusivity, D, and carrying capacity limited proliferation with proliferation rate, λ, and carrying capacity density, K. By analysing temporal changes in cell density in several subregions located well-behind the initial position of the leading edge we estimate λ and K. Given these estimates, we then apply automatic leading edge detection algorithms to the images produced by the IncuCyte ZOOM$^{\mathrm{TM}}$ assay and match this data with a numerical solution of the Fisher-Kolmogorov equation to provide an estimate of D. We demonstrate this method by applying it to interpret a suite of IncuCyte ZOOM$^{\mathrm{TM}}$ assays using PC-3 prostate cancer cells and obtain estimates of D, λ and K. Comparing estimates of D, λ and K for a control assay with estimates of D, λ and K for assays where epidermal growth factor (EGF) is applied in varying concentrations

confirms that EGF enhances the rate of scratch closure and that this stimulation is driven by an increase in D and λ, whereas K is relatively unaffected by EGF. Our approach for estimating D, λ and K from an IncuCyte ZOOM$^{\text{TM}}$ assay provides more detail about cellular-level behaviour than standard methods for analysing these assays. In particular, our approach can be used to quantify the balance of cell migration and cell proliferation and, as we demonstrate, allow us to quantify how the addition of growth factors affects these processes individually.

2.1 Background

Scratch assays are commonly used to quantify the potential for collective cell spreading by taking a spatially uniform population of cells on a two-dimensional substrate, creating an artificial scratch in the monolayer, and then making observations about the rate at which the remaining population spreads into the vacant region [2, 4, 15, 16, 19–21, 24, 25, 46]. Scratch assays are routinely used since they are technically straightforward, fast and inexpensive [23]. Data obtained from scratch assays can be used to improve our understanding of drug design, malignant spreading and tissue repair [23].

A key limitation of scratch assays is the question of whether they are reproducible since the scratch can be made with various types of instruments and varying degrees of pressure, and the assay can be performed on several different types of substrates. All of these variables have the potential to affect the results of the scratch assay. Inspired by these limitations, new platforms to perform scratch assays, such as the IncuCyte$^{\text{TM}}$ and IncuCyte ZOOM$^{\text{TM}}$ real time live cell imaging assays have been developed [7]. IncuCyte ZOOM$^{\text{TM}}$ assays have the advantage that the scratch is reproducibly created with a mechanical tool and live images are obtained without the need to interrupt the experiment for imaging purposes [7].

Typical approaches to quantify IncuCyte ZOOM$^{\text{TM}}$ assay data involve making use of automated features that allow the user to quantify the proportion of the initially-scratched area that becomes re-colonised by cells as a function of time. As the assay proceeds and the cell population spreads into the initially-vacant area, the proportion of the area covered by cells increases with time. Typically, this data is presented as a plot of *relative wound density* as a function of time [9, 30, 31]. The relative wound density is a ratio of the occupied area of the initially-scratched area to the total area of the scratch [7]. To illustrate this typical approach we present a series of images from an IncuCyte ZOOM$^{\text{TM}}$ assay with PC-3 cells [17] in Figure 2.1. PC-3 cells are a prostate cancer cell line with high metastatic potential [14, 17]. The experimental image in Figure 2.1(a) shows the initial scratch, and the subsequent re-colonisation of the initially-vacant area is shown in Figures 2.1(b)–(d). The data in Figure 2.1(e) demonstrates the temporal variation in the relative wound density, which is automatically calculated by the IncuCyte ZOOM$^{\text{TM}}$ system [7]. While this kind of standard approach for quantifying IncuCyte ZOOM$^{\text{TM}}$ assays can provide useful information about how quickly a particular cell population is able to re-colonise the initially-vacant area, it does not distinguish between the relative roles of various cellular functions. The collective spreading of a population of cells is driven by both cell motility and cell proliferation [2, 24, 25, 44, 46]. However, traditional data extracted from IncuCyte ZOOM$^{\text{TM}}$ assays does not give us any indication of the relative roles of cell motility and cell proliferation. This additional information could be important in terms of understanding how a particular growth factor or a potential drug treatment affects collective spreading since it is possible that the addition of a growth factor or drug treatment could affect: (i) cell motility alone, (ii) cell proliferation alone, or (iii) both cell motility and cell proliferation, simultaneously.

In this methodology article we describe an alternative method for interpreting IncuCyte ZOOM$^{\text{TM}}$ assay data using a continuum mathematical model. Our approach allows us to quantify the rate of cell migration in terms of an undirected cell diffusivity, D, and the rate of cell proliferation in terms of the proliferation rate, λ, and carrying capacity density, K. Applying this approach to a suite of IncuCyte ZOOM$^{\text{TM}}$ assay data using PC-3 prostate cancer cells allows us to obtain estimates of D, K and λ for these cells. Under control conditions our method gives $D \approx 1.32 \times 10^2\ \mu\text{m}^2/\text{h}$, $K \approx 1.13 \times 10^{-3}\,\text{cells}/\mu\text{m}^2$ and $\lambda \approx 5.07 \times 10^{-2}\,/\text{h}$, which corresponds to a cell doubling time of approximately 14 h. We provide additional datasets where all experiments are repeated with varying concentrations of human epidermal growth factor (EGF) [12, 13], which leads to enhanced collective spreading. Applying our technique to this additional data indicates that the EGF acts affects both D and λ, but not K. In particular, our results suggest that D increases monotonically with EGF concentration whereas we observe a nonmonotonic relationship between λ and EGF concentration, with a maximum proliferation rate when the assays are treated with 50 ng/mL EGF. Although the techniques described here have been used previously to calibrate mathematical models to experimental data from circular barrier assays [41, 44, 45], this is the first time that IncuCyte ZOOM$^{\text{TM}}$ assay data has been used to calibrate these parameters, and the first time that this process has been used to quantify how estimates of D, λ and K depend on the concentration of EGF in an IncuCyte ZOOM$^{\text{TM}}$ assay.

2.2 Methods

2.2.1 IncuCyte ZOOM$^{\text{TM}}$ assay

We perform a monolayer scratch assay using the IncuCyte ZOOM$^{\text{TM}}$ live cell imaging system (Essen BioScience, MI USA). This system measures scratch closure in real time and automatically calculates the relative wound density within the initially-vacant area at each time point. The relative wound density is the ratio of the occupied area to the total area of the initial scratched region. All experiments are performed using the PC-3 prostate cancer cell line [17], which is obtained from the American Type Culture Collection (ATCC, Manassas, USA). Cells are routinely propagated in RPMI 1640 medium (Life Technologies, Australia) in 10% foetal calf serum (Sigma-Aldrich, Australia), with 100 u/mL penicillin, 100 μg/mL streptomycin (Life Technologies), in plastic flasks (Corning

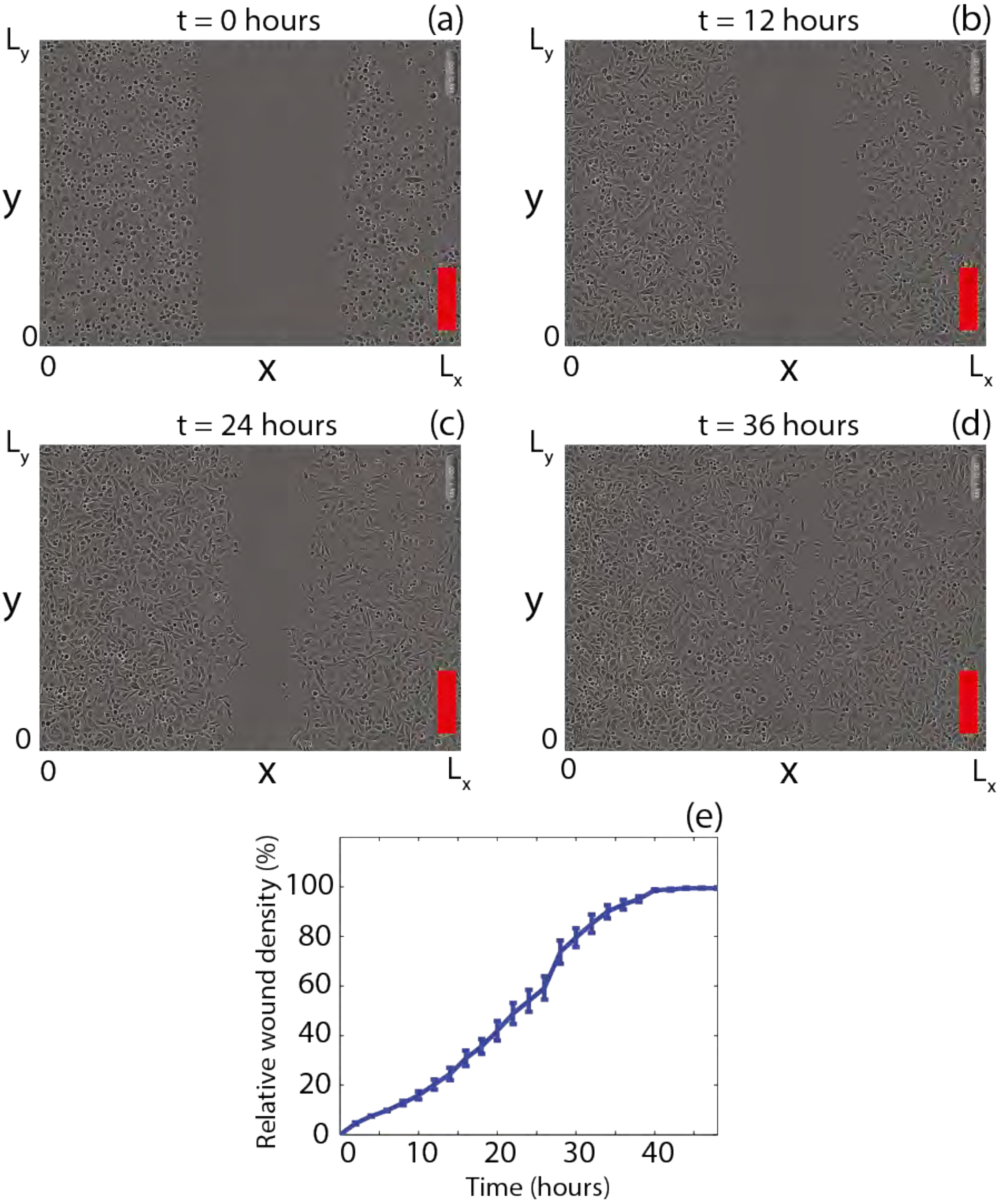

Figure 2.1: Images from the control IncuCyte ZOOM$^{\text{TM}}$ assay with PC-3 cells showing, (a) the initial position of the scratch, and the subsequent collective cell spreading after 12, 24 and 36 h in (b)–(d), respectively. Scale bar corresponds to 300 μm. The results in (e) show the standard way of presenting IncuCyte ZOOM$^{\text{TM}}$ assay data for these experiments as the relative wound density as a function of time. Here we present the average relative wound density from $n = 3$ identically prepared experimental replicates. The error bars in (e) indicate one standard deviation about the mean.

Life Sciences, Asia Pacific) in 5% CO_2 and air in a Panasonic incubator (VWR International) at 37°C. Cells are regularly screened for *Mycoplasma* (ATCC). Cells are removed from the monolayer using TrypLE$^{\text{TM}}$ (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). After seeding, cells are grown overnight to form a spatially uniform monolayer. We use a WoundMaker$^{\text{TM}}$ (Essen BioScience) to create uniform, reproducible scratches in all the wells of a 96-well plate. After creating the scratch, the medium is aspirated and the wells are washed twice with fresh medium to remove any cells from the scratched area. Following the washes, for the control assay, 100 μL of fresh medium is added to each well. We also perform a series of experiments where, following the washes, fresh medium containing different concentrations of EGF (Life Technologies) is added to the wells. The concentrations of EGF we use are: 25, 50, 75, 100 and 125 ng/mL. We will refer to these assays as EGF-25, EGF-50, EGF-75, EGF-100 and EGF-125, respectively. Once the fresh medium is added, the plate is placed into the IncuCyte ZOOM$^{\text{TM}}$ apparatus and images of the collective cell spreading are recorded every two hours for a total duration of 46 hours. For the control assay and each different EGF concentration we perform three identically prepared experimental replicates ($n = 3$).

2.2.2 Image analysis

We use Matlab's Image Processing Toolbox [26, 43] to estimate the position of the leading edge of the spreading cell population in the IncuCyte ZOOM$^{\text{TM}}$ images. The experimental image is imported and converted to greyscale using the `imread` and `rgb2gray` commands, respectively. We detect edges in the images using `edge` with the Canny method [3] and automatically-selected threshold values. Detected edges outside of these threshold values are ignored. Remaining edges are dilated using the `imdilate` command and a structuring element, defined using `strel`, with a circular element of size 15. Using the `bwareaopen` command with a component size of 10,000 pixels, we remove any remaining vacant spaces in the image while preserving the vacant scratch. Edge dilation is reversed using the `imerode` command with the same structuring element defined previously to erode the image. Finally, edges within the image are smoothed using `medfilt2` and the area of the remaining vacant space, $A(t)$, representing the vacant area, is estimated using the `regionprops` command.

We calculate the position of the leading edge, which we define to be the distance between the centre of the experimental domain and the position of the leading edge using

$$L_E(t) = \frac{L_x L_y - A(t)}{2L_y}, \tag{2.1}$$

where L_x is the horizontal width of the image and L_y is the vertical height of the image. For all experiments we have $L_x = 1970$ μm and $L_y = 1430$ μm. Equation (2.1) allows us to examine the time evolution of the scratched area in terms of $L_E(t)$, which is the half-width of the scratch (Figure 2.4(a)).

2.2.3 Mathematical model

We interpret our experimental results using the Fisher-Kolmogorov equation [8, 22, 28], which is a continuum reaction-diffusion model describing the spatiotemporal evolution of cell density in a population of cells where cell migration is driven by random (undirected) cell motility and cell proliferation is driven by carrying capacity limited logistic growth. The Fisher-Kolmogorov equation, and extensions of the Fisher-Kolmogorov equation, have been previously applied to *in vitro* [32, 35, 36] and *in vivo* [1, 29] data describing collective cell spreading in a range of contexts including wound healing [10, 37], tissue repair [2, 46] and malignant spreading [4, 19, 20, 34, 42].

Since our scratch assay takes place on a two-dimensional substrate (Figures 2.1(a)–(c)), we start with the two-dimensional analogue of the Fisher-Kolmogorov equation in Cartesian coordinates

$$\frac{\partial \overline{C}(x,y,t)}{\partial t} = D\left(\frac{\partial^2 \overline{C}(x,y,t)}{\partial x^2} + \frac{\partial^2 \overline{C}(x,y,t)}{\partial y^2}\right) + \lambda \overline{C}(x,y,t)\left(1 - \frac{\overline{C}(x,y,t)}{K}\right), \tag{2.2}$$

where $\overline{C}(x,y,t)$ [cells/μm^2] is the cell density, or average number of cells per unit area, at location (x,y) and time t. For our experiments we have $0 \leq x \leq L_x$ and $0 \leq y \leq L_y$. There are three parameters in the Fisher-Kolmogorov equation: (i) the cell diffusivity, D [μm^2/h], (ii) the cell proliferation rate, λ [/h], and (iii) the carrying capacity density, K [cells/μm^2]. The proliferation rate, λ, is related to the cell doubling time $t_d = \log_{\mathrm{e}}(2)/\lambda$. We note that we make the standard assumption that D, λ and K are constants [2, 24, 25, 35, 37, 46]

Since the initial cell monolayer is spatially uniform and the initial scratch is made perpendicular to the x-direction (Figures 2.1(a)), we can simplify the mathematical model by averaging in the y-direction [4, 19, 20]. To do this we average the two-dimensional cell density

$$C(x,t) = \frac{1}{L_y}\int_0^{L_y} \overline{C}(x,y,t)\ \mathrm{d}y, \tag{2.3}$$

which allows us to write Equation (2.2) as a one-dimensional partial differential equation

$$\frac{\partial C(x,t)}{\partial t} = D\frac{\partial^2 C(x,t)}{\partial x^2} + \lambda C(x,t)\left(1 - \frac{C(x,t)}{K}\right). \tag{2.4}$$

In general, approximating a two-dimensional nonlinear partial differential equation, such as Equation (2.2), by an averaged one-dimensional nonlinear partial differential equation, such as Equation (2.4), can introduce an averaging error. However, for initial conditions such as ours where the initial density is independent of the vertical location, this error vanishes, and a detailed analysis of this error is presented elsewhere [38, 39]. The initial condition for Equation (2.4) is given by the width of the scratch (Figure 2.1(a))

$$C(x,0) = \begin{cases} C_0 & 0 \leq x < 985 - L_E(0)\ \mu\text{m}\,, \\ 0 & 985 - L_E(0) \leq x < 985 + L_E(0)\ \mu\text{m}\,, \\ C_0 & 985 + L_E(0) \leq x \leq 1970\ \mu\text{m} \end{cases} \tag{2.5}$$

where C_0 is the initial density of cells in the monolayer and $L_E(0)$ is the initial position of the leading edge. Since we use a WoundMaker$^{\text{TM}}$ tool to create uniform scratches in all experimental replicates, the initial condition, given by Equation (2.5), applies to all experiments and cannot be varied.

The physical distribution of cells in each experiment extends well-beyond the $L_x \times L_y$ rectangular region imaged by the IncuCyte ZOOM$^{\text{TM}}$ apparatus. Therefore, since the cells are spatially uniform except for the scratched region, there will be no net flux of cells across the vertical boundaries along the lines $x = 0$ and $x = L_x$. We model this by using zero-flux boundary conditions

$$\begin{aligned} \frac{\partial C(x,t)}{\partial x} &= 0 \quad \text{at} \quad x = 0, \\ \frac{\partial C(x,t)}{\partial x} &= 0 \quad \text{at} \quad x = L_x. \end{aligned} \tag{2.6}$$

These boundary conditions do not imply that cells are stationary at $x = 0$ and $x = L_x$. Instead, these boundary conditions imply that the cell density profile is spatially uniform, $\partial C(x,t)/\partial x = 0$, so that there is no net flux of cells across the vertical boundaries at $x = 0$ and $x = L_x$.

We solve Equation (2.4) using a finite difference numerical method [27]. The spatial domain, $0 \leq x \leq L_x$, is discretised uniformly with grid spacing δx, and the spatial derivatives are approximated using a central-difference approximation [27]. This leads to a system of coupled nonlinear ordinary differential equations that are integrated through time using a backward-Euler approximation with constant time steps of duration δt [27]. The resulting system of coupled nonlinear algebraic equations are linearised using Picard (fixed-point) iteration, with absolute convergence tolerance ϵ [47]. The associated tridiagonal system of linear equations is solved using the Thomas algorithm [27]. For all results presented here we always chose δx, δt and ϵ so that our numerical algorithm produces grid-independent results.

We also apply Equation (2.4) to some simplified situations where we focus on the time evolution of the cell density in small subregions, located well-behind the initial position of the scratch, where the cell density is approximately spatially uniform. This implies that $C(x,t) \approx C(t)$ within these subregions [41, 44, 45]. Since the cell density is approximately spatially uniform we have $\partial C(x,t)/\partial x = 0$, and the first term on the right of Equation

(2.4) vanishes and, subsequently in these subregions, the partial differential equation simplifies to the logistic equation,

$$\frac{\mathrm{d}C(t)}{\mathrm{d}t} = \lambda C(t)\left(1 - \frac{C(t)}{K}\right), \tag{2.7}$$

whose solution is given by

$$C(t) = \frac{KC(0)}{C(0) - \mathrm{e}^{-\lambda t}\left(C(0) - K\right)}, \tag{2.8}$$

where $C(0) = C_0$ is the initial density at $t = 0$. The simplification of approximating Equation (2.4) by Equation (2.7) in subregions located well-behind the leading edge where the cell density is spatially uniform does not imply that cells are stationary in these subregions. Instead, Equation (2.7) represents the situation where there is no gradient in cell density and cells are free to move amongst the extracellular space within these subregions. The key advantage of applying this approximation is that cell motion in these spatially uniform subregions does not contribute to any temporal changes in cell density. Instead, when the cell density is spatially uniform, any temporal change in cell density is solely associated with the proliferation term in Equation (2.4) [41, 44, 45].

2.2.4 Parameter estimation

We estimate the three parameters in the Fisher-Kolmogorov model using a sequential approach. First, using cell counting, we estimate the parameters governing cell proliferation: K and λ. Second, using data describing the temporal changes in the position of the leading edge, we estimate the cell diffusivity, D. Although it is possible to use a different approach, based on a multivariate regression technique to estimate D, λ and K simultaneously, we prefer to estimate these parameters sequentially. Estimating the three parameters sequentially, emphasises the differences in the interpretation of these parameters, as well as emphasising the differences in the mechanisms of cell proliferation and cell motility. If, instead, a multivariate approach is used to estimate the three parameters simultaneously, we anticipate that the interpretation of the mechanisms associated with these parameters might not be obvious as it is in our approach.

Carrying capacity density

To estimate K we focus on experimental images from the latter part of the experiment, $t = 46$ h, where the cell population has grown to confluence. We identify three smaller subregions, located well-behind the initial leading edge, and count the number of cells within each subregion, N. To quantify the variability in our estimate we analyse three different subregions in each image and count N in each replicate subregion. Using this data we estimate the average carrying capacity density as

$$K = \frac{\langle N \rangle}{A_{SR}} \tag{2.9}$$

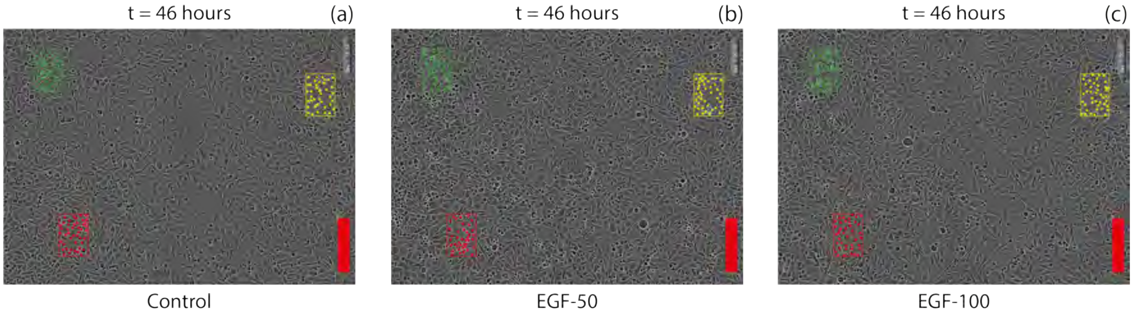

Figure 2.2: Final time experimental images (t = 46 hours) for three IncuCyte ZOOM$^{\text{TM}}$ assays for (a) Control, (b) EGF-50, and (c) EGF-100. The three coloured boxes indicate the location of the three subregions used to estimate K and λ. Each coloured square within the subregions indicates the centre of an individual cell in the cell counting step. Scale bar corresponds to 300 μm.

where $\langle N \rangle$ is the average number of cells within the subregion of area $A_{SR} = 3.789 \times 10^4$ μm^2. To examine whether EGF has any impact on the carrying capacity density we estimate K for the control assay and for each experiment treated with a different EGF concentration. Figure 2.2 shows IncuCyte ZOOM$^{\text{TM}}$ images at $t = 46$ h with the location of three subregions superimposed. The images in Figures 2.2(a)–(c) show the control, EGF-50 and EGF-100 assays, respectively. We note that the location of all three subregions in each image is well-behind the initial position of the scratch (Figure 2.1(a)) so that after $t = 46$ h the local density of cells within each subregion has grown to confluence. To quantify the variability in our estimate of K, we calculate the sample standard deviation for each EGF concentration, and report results as a mean value for K, with the variation in our estimate given by plus or minus one standard deviation about the mean. Results are summarised in Table 2.1.

Proliferation rate

The logistic equation, given by Equation (2.7), describes the time evolution of the cell density where there is, on average, no spatial variation in cell density. To apply the logistic equation to our data we analyse three subregions within each IncuCyte ZOOM$^{\text{TM}}$ image at several time points during the assay. Counting the total number of cells in each subregion and dividing by the area of the subregion gives an estimate of the local cell density in that subregion. In all cases the subregions contain approximately 20-30 cells at $t = 0$ h. Repeating this procedure for three different subregions, at fixed locations, for each experimental replicate, at five different time points, allows us to calculate the average cell density as a function of time, $C(t)$. With this data, together with our previous estimates of K, we find the value of λ in Equation (2.8) that matches our $C(t)$ data across several time points. For consistency, when we estimate λ we always analyse the same three subregions that we used previously to estimate K. The location of these three subregions is shown in Figure 2.3. We estimate the initial cell density, $C(0)$, from the first image taken immediately after the scratch is made at $t = 0$ h. Images of the assay in Figures

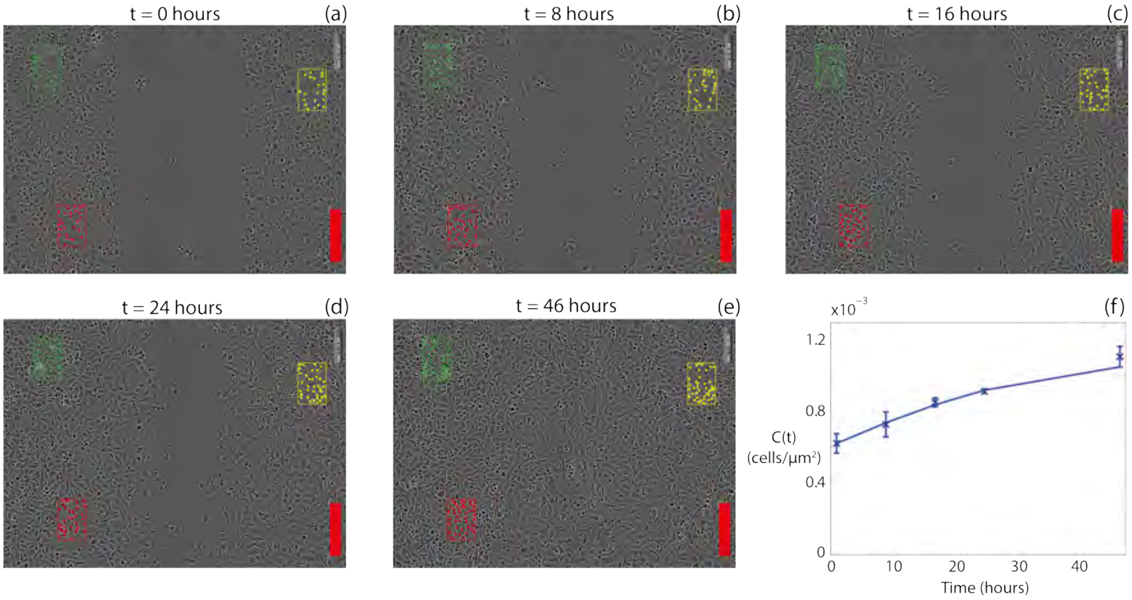

Figure 2.3: (a)-(e) Time evolution of an EGF-75 IncuCyte ZOOM$^{\text{TM}}$ assay. Images taken after (a) 0, (b) 8, (c) 16, (d) 24, and (e) 46 h after the scratch was performed. The three coloured boxes indicate the location of the three subregions used to calculate K and λ. Each coloured square within the subregions indicates the centre of an individual cell in the cell counting step. Scale bar corresponds to 300 μm. (f) Comparison of the average experimental cell density $C(t)$ (crosses) and the logistic growth curve using our estimates of K and λ (solid).

2.3(a)–(e) correspond to 0, 8, 16, 24 and 46 h, respectively. The location of each subregion is chosen to be well-behind the initial position of the leading edge of the population so that the cell density is approximately spatially uniform locally within each subregion. In each subregion $C(t)$ increases with time, and we attribute this increase to cell proliferation. The data in Figure 2.3(f) shows the time evolution of the average cell density, $C(t)$, calculated by averaging the three estimates of cell density from each subregion, at each time point. Using our previous estimate of K, we estimate λ by matching the solution of Equation (2.7) with the observed $C(t)$ data.

For each EGF concentration we have three sets of data describing the temporal variation in average cell density per experimental replicate. For each set of time series data we use Matlab's `lsqnonlin` function, a nonlinear least squares minimisation routine [5], to estimate λ. To quantify the average proliferation rate we average the three estimates of λ from each experimental replicate. A comparison of the resulting logistic growth curve using our average estimate of K and λ with the observed $C(t)$ data is given in Figure 2.3(f) for the EGF-25 experiment, indicating that the solution of Equation (2.7) matches the data reasonably well. To quantify the variability in λ we report our average estimate of λ and the variation as the average plus or minus one sample standard deviation. We also report the average and variability in the C_0 values used to obtain estimates of λ. Results are summarised in Table 2.1.

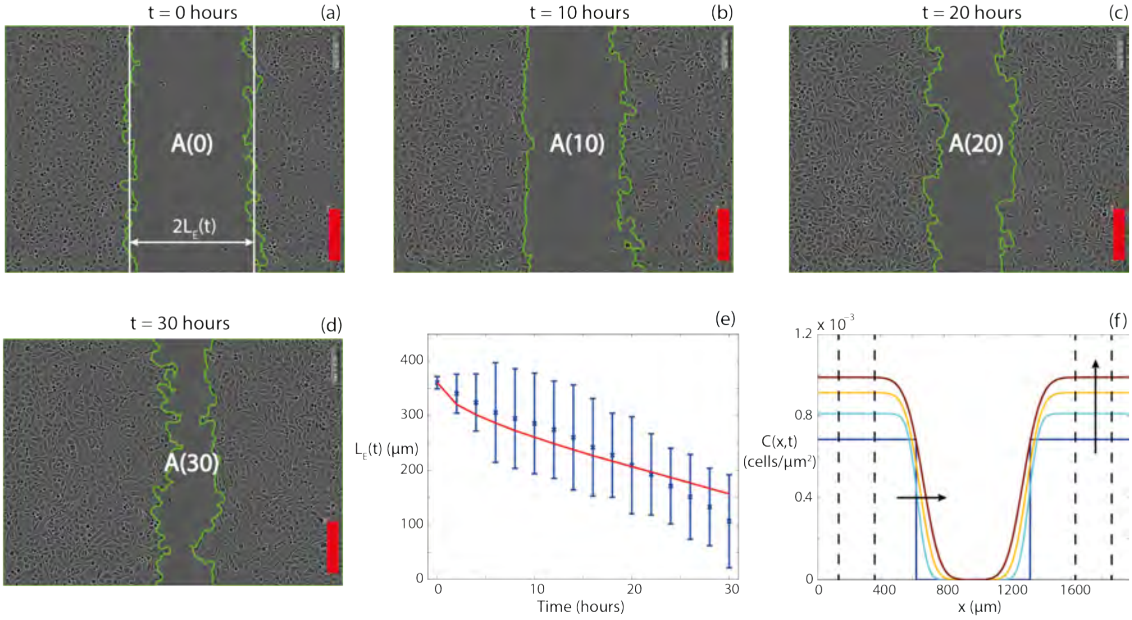

Figure 2.4: (a)-(d) Indicate the area of remaining vacant space, $A(t)$, as determined by the edge detection algorithm at (a) 0, (b) 10, (c) 20, and (d) 30 h for the control assay. The position of the detected leading edge is given in green. The straight vertical lines superimposed on (a) (white) indicate the average width of the scratch, $2L_E(t)$. Scale bar corresponds to 300μm. (e) Average $L_E(t)$ data estimated from the control assay experimental images (blue). The error bars correspond to one standard deviation about the mean. Numerical $L_E(t)$ data (red), corresponding to the numerical solution of Equation (4) using the relevant estimates of D, λ and K (Table 2.1). (f) Evolution of $C(x,t)$ profiles at $t = 0,\ 10,\ 20,\ 30$ h corresponding to the numerical solution of Equation (2.4) using the relevant estimates of D, λ and K (Table 2.1). Arrows indicate the direction of increasing time. Numerical solutions of Equation (2.4) correspond to $\delta x = 1$ μm, $\delta t = 0.1$ h and $\epsilon = 1 \times 10^{-6}$. The vertical lines show the locations of the subregions where the estimates of λ and K were obtained.

Cell diffusivity

Several different approaches have been used in previous studies to estimate D from *in vitro* assays describing collective cell spreading processes. For example, Treloar *et al.* [45] estimate D in a circular barrier assay by applying two different methods to the same data set. First, they estimate D using a cell labelling and cell counting technique to provide an estimate of the detailed cell density profiles near the leading edge of the spreading population. Treloar *et al.* [45] calibrate the solution of a partial differential equation to that data to give an estimate of D which matches the position and shape of the spreading cell density profile. Second, using the same data set, Treloar *et al.* [45] use automated leading edge image analysis [26, 43] to quantify temporal changes in the position of the leading edge of the spreading cell density profile without counting individual cells. Treloar *et al.* [45] calibrate their model to this leading edge data to obtain a second estimate of D. Given that the two approaches implemented by Treloar *et al.* [45] produce similar estimates of D, here we choose to estimate D using leading edge data since this is the most straightforward approach which avoids the need for labelling and counting individual cells within the spreading population.

Figures 2.4(a)-(d) show IncuCyte ZOOM$^{\text{TM}}$ images at 0, 10, 20 and 30 h, respectively. The position of the two detected leading edges of the spreading population is superimposed on each image. A visual comparison of how the position of the detected leading edges changes with time suggests that the initially-vacant region closes symmetrically with time. The edge detection results allow us to calculate the area of the vacant region, $A(t)$, and with this information Equation (2.1) allows us to estimate the half-width, $L_E(t)$, which is decreasing function of time. Previously, Treloar *et al.* [43, 45] showed that the location of the automatically detected leading edge corresponds to a cell density of approximately 2% of the carrying capacity density. Therefore, we conclude that the Fisher-Kolmogorov model adequately describes this experimental system as these kinds of standard data can be compared to the model output, and this approach provides parameter estimates consistent with those obtained from other parameter recovery techniques [45].

Given our previous estimates of K and λ, and assuming that the position of the detected leading edge corresponds to the location where the density is 2% of the carrying capacity, we use Matlab's `lsqnonlin` function to find an estimate of D that minimises the difference between the observed time series of $L_E(t)$ and the time series of $L_E(t)$ data from the numerical solution of Equation (2.4). We present an example of the match between the experimental measurements of $L_E(t)$ and numerical prediction of $L_E(t)$ in Figure 4(c). For all time points, the numerical estimate of $L_E(t)$ is within one sample standard deviation of the average of the experimental measurements. Given our estimates of D, λ and K, we can use our numerical solution of Equation (2.4) to explore how $C(x,t)$ varies across the entire width of the domain, for the duration of the assay, as illustrated in Figure 2.4(f). These profiles show that the cell density remains approximately spatially uniform well-behind the initial location of the scratch. In fact, we have indicated the position of the location of the various subregions used to estimate K and λ on the profiles in Figure 4(f), and we see that the predicted cell density profile is spatially uniform, $\partial C(x,t)/\partial x = 0$, at these locations for the duration of the assay, which is visually consistent with the subregions presented in Figure 2.2. The cell density in the three subregions well-behind the initial location of the scratch increases with time owing to cell proliferation. These profiles also show the cell density front near the location of the scratch moves inward to close the initially-scratched region with time.

Using our approach we calculate an average value of D by estimating the cell diffusivity for each experimental replicate and then averaging the results. We note that Treloar *et al.* found that varying the Matlab edge detection threshold parameters led to a small variation in the position of the detected leading edge corresponding to a cell density in the range of approximately 1–5% of the carrying capacity density [43]. To quantify the variability in our estimate of D we repeated the edge detection by assuming that the position of the detected leading edge corresponds to both 1% and 5% of the carrying capacity density. Results are summarised in Table 2.1. We note that the maximum value of D is obtained by assuming that the detected leading edge corresponds to 5% of the carrying capacity since this upper bound implies additional spreading. Conversely, the minimum value of D is obtained by assuming that the detected leading edge corresponds to 1% of the carrying capacity since this lower bound implies less spreading.

Experiment	K (cells/μm^2)	λ (/h)
Control	1.13×10^{-3} ($1.11 \times 10^{-3} - 1.14 \times 10^{-3}$)	5.07×10^{-2} ($4.12 \times 10^{-2} - 6.03 \times 10^{-2}$)
EGF-25	1.04×10^{-3} ($1.01 \times 10^{-3} - 1.07 \times 10^{-3}$)	5.59×10^{-2} ($4.40 \times 10^{-2} - 6.79 \times 10^{-2}$)
EGF-50	1.12×10^{-3} ($1.11 \times 10^{-3} - 1.13 \times 10^{-3}$)	6.94×10^{-2} ($5.60 \times 10^{-2} - 8.27 \times 10^{-2}$)
EGF-75	1.12×10^{-3} ($1.11 \times 10^{-3} - 1.13 \times 10^{-3}$)	5.74×10^{-2} ($5.44 \times 10^{-2} - 6.05 \times 10^{-2}$)
EGF-100	1.16×10^{-3} ($1.11 \times 10^{-3} - 1.20 \times 10^{-3}$)	6.13×10^{-2} ($5.17 \times 10^{-2} - 7.08 \times 10^{-2}$)
EGF-125	1.11×10^{-3} ($1.09 \times 10^{-3} - 1.12 \times 10^{-3}$)	5.48×10^{-2} ($5.35 \times 10^{-2} - 5.62 \times 10^{-2}$)

Experiment	D (μm^2/h)	C_0 (cells/μm^2)
Control	1.32×10^{2} ($1.05 \times 10^{2} - 1.98 \times 10^{2}$)	6.84×10^{-4} ($5.77 \times 10^{-4} - 7.91 \times 10^{-4}$)
EGF-25	1.59×10^{2} ($1.27 \times 10^{2} - 2.38 \times 10^{2}$)	6.42×10^{-4} ($5.01 \times 10^{-4} - 7.83 \times 10^{-4}$)
EGF-50	1.53×10^{2} ($1.21 \times 10^{2} - 2.18 \times 10^{2}$)	7.79×10^{-4} ($6.32 \times 10^{-4} - 9.26 \times 10^{-4}$)
EGF-75	1.64×10^{2} ($1.30 \times 10^{2} - 2.44 \times 10^{2}$)	6.81×10^{-4} ($5.75 \times 10^{-4} - 7.87 \times 10^{-4}$)
EGF-100	2.06×10^{2} ($1.65 \times 10^{2} - 3.12 \times 10^{2}$)	7.12×10^{-4} ($6.10 \times 10^{-4} - 8.15 \times 10^{-4}$)
EGF-125	2.40×10^{2} ($1.90 \times 10^{2} - 3.51 \times 10^{2}$)	7.68×10^{-4} ($6.55 \times 10^{-4} - 7.68 \times 10^{-4}$)

Table 2.1: Estimated K, λ, D and C_0 values for PC-3 cells for different EGF concentrations. Results are reported as a mean, with the estimate of the variability given in the parenthesis.

2.3 Results

By applying the parameter estimation procedures described in Section 2.4 we obtain estimates of K, λ and D, as well an estimates of the variability in these values. These results are summarised in Table 2.1. Comparing our estimates of K, λ and D for each assay with a different concentration of EGF provides us with information about how EGF affects cell proliferation and cell motility for the PC-3 cell line. Data in Table 2.1 indicates that K varies by no more than approximately 8% between the experiments with different EGF concentrations relative to the control experiment. In comparison, our estimates of λ and D vary by approximately 37% and 82%, respectively, between the experiments with different EGF concentrations, relative to the control experiment. These results imply that EGF affects both the rate of cell motility and the rate of cell proliferation; however, EGF appears to have a smaller influence on the carrying capacity density, suggesting that it has minimal impact on the physical shape and maximum packing density of PC-3 cells.

Results in Table 2.1 suggest that we observe an increase in the rate of proliferation with small concentrations of EGF. However, there appears to be a reduction in the rate of cell proliferation at larger concentration of EGF, implying that there is an optimal stimulation of proliferation at an EGF concentration of approximately 50 ng/mL. This kind of nonmonotonic response to EGF has been observed in previous experiments involving both PC-3 [14] and other cell types [33]. However, unlike these previous studies [14, 33], our approach allows us to estimate how EGF affects both cell migration and cell proliferation separately.

Results in Table 2.1 suggest that EGF significantly enhances cell motility, D. However, unlike the response in λ, our results suggest that the diffusivity of PC-3 cells appears to be an monotonically increasing function of EGF for the concentrations of EGF that we consider.

2.4 Discussion and conclusions

In this work we provide an alternative method for analysing IncuCyte ZOOM$^{\text{TM}}$ assays [7]. The traditional approach for analysing IncuCyte ZOOM$^{\text{TM}}$ assays is to report the temporal variation in the relative wound density (Figure 2.1), which is the ratio of the occupied area to the initially-vacant area of the scratch [9, 30, 31]. While this data allows us to quantify the rate of collective cell spreading, it does not provide any quantitative insight into the relative roles of different mechanisms that drive collective cell spreading. As an alternative, we present a method which allows us to analyse standard images from IncuCyte ZOOM$^{\text{TM}}$ assays by interpreting the results quantitatively using the Fisher-Kolmogorov equation [8, 22]. Our approach provides a quantitative measure of the relative roles of cell migration and cell proliferation by estimating the carrying capacity density, K, the cell proliferation rate λ, and the cell diffusivity, D.

To estimate K we focus on images from the latter part of the IncuCyte ZOOM$^{\text{TM}}$ assay, $t = 46$ h, by which time the cell monolayer has approximately grown to confluence. We count the number of cells in several subregions located well-behind the initial position of the scratch. Dividing the cell counts by the area of the subregion gives us an estimate of the carrying capacity density, K. To examine the influence of our choice of the area of the subregion, we also examine the sensitivity of our estimate of K for the control assay to variations in the area of the subregion. For example, with $A_{SR} = 3.789 \times 10^4$ μm^2, we obtain $K = 1.13 \times 10^{-3} \pm 0.01 \times 10^{-3}$ cells/μm^2 for the control assay. Repeating the procedure and doubling A_{SR} gives an estimate of K which is within 2% of the original estimate. Therefore, our estimate of K is practically insensitive to the size of the subregion.

To estimate λ we count the numbers of cells in several subregions, located well-behind the initial position of the scratch, at several time points during the assay. This allows us to quantify how the cell density behind the initial scratch increases with time to reach carrying capacity density. Using this information we calibrate the solution of the logistic equation, using our previous estimate of K, to that data, which provides an estimate of λ. To estimate λ we focus on relatively early time data, $t = 0, 8, 16, 24$ and 46 h, since we anticipate that most of the proliferation activity occurs before the cell population reaches confluence. Using this approach for the control assay we obtain $\lambda = 5.07 \times 10^{-2} \pm 0.96 \times 10^{-2}$ /h. To examine the influence of our choice of time points we re-estimate λ using data at $t = 0, 8, 16, 24, 30$ and 46 h, giving $\lambda = 5.53 \times 10^{-2}$ /h, which is well-within the variability of the original estimate. Since the process of identifying and counting cells in various subregions is the most time-consuming aspect of our method together with the fact that including additional data at these intermediate times does not significantly alter our estimates of λ, we conclude that our choice of focusing on relatively early-time observations is adequate to provide estimates of λ.

Given our estimates of λ and K, we then estimate D by solving the Fisher-Kolmogorov equation numerically and finding a value of D which provides the best match between the position of the leading edge observed in the experiments and the position of the leading edge predicted by the numerical solution of the Fisher-Kolmogorov equation. Using this approach, for our control assays, we estimate $D \approx 1.32 \times 10^2$ μm^2/h, $\lambda \approx 5.07 \times 10^{-2}$/h and $K \approx 1.13 \times 10^{-3}$ cells/μm^2 for the PC-3 prostate cancer cell line [17]. Since typical values of D reported in the literature vary in the range of 10^1 - 10^3 μm^2/h [2, 35, 41, 45, 46], our estimate of D seems reasonable since it is well within previously reported values for different cell lines.

In addition to analysing the control assay, we also estimate D, λ and K for a suite of assays where the cell culture medium is supplemented with different concentrations of EGF (25, 50, 75, 100 and 125 ng/mL) [12, 13]. Using our approach we estimate D, λ and K for each EGF concentration, which provides insight into how EGF affects the rate of cell migration, the rate of cell proliferation and the carrying capacity density for PC-3 cells in these assays. In summary, we find there is no consistent trend in our estimates of K with the different EGF treatments. The maximum variability in our estimate of K between different EGF concentrations is approximately 8%, indicating that the carrying capacity density is relatively unaffected by EGF. In contrast, we find that our estimates of D and λ are both sensitive to EGF. The maximum variability in D and λ amongst different EGF treatments is approximately 82% and 37%, respectively. Therefore, our analysis suggests that EGF affects both cell motility and cell proliferation, with the impact on cell motility being more pronounced than the impact on cell proliferation. Interestingly our results suggests that we have a monotonic increase of D with EGF concentration whereas we have a nonmonotonic relationship between λ and EGF concentration. We observe a maximum stimulation of proliferation at an EGF concentration of 50 ng/mL.

Similar to other applications of the Fisher-Kolmogorov equation [2, 24, 25, 35–37, 41, 46], we have made the standard assumption that the parameters in each experiment, D, λ and K, are constants that do not vary with position, time or cell density. Recently, there has been considerable interest in the theoretical physics and applied mathematics literature regarding the analysis of extensions of the Fisher-Kolmogorov equation where D and λ vary with position, time or cell density [6, 11]. Although these extensions are mathematically interesting, we have not attempted to apply such an extension here since the precise form of the putative spatial or temporal dependence is unknown, and at this stage, we anticipate that more detailed experimental data would be required to calibrate these more detailed mathematical models. We leave this extension as a potential topic for future analysis.

The question of whether there is any role for chemotaxis in this particular IncuCyte ZOOM$^{\text{TM}}$ assay has not been addressed in this work. Since we have been able to obtain reasonable estimates for D, λ and K by calibrating the solution of the Fisher-Kolmogorov equation to our experimental data it is not obvious that we need to consider applying a more complicated model incorporating chemotactic cell migration at this time. However, it is possible that the cells produce a chemical signal, $G(x,t)$, which as a result of diffusion and decay, could lead to the formation of a chemical gradient that stimulates additional directed cell motion [18]. An extension of the Fisher-Kolmogorov model which incorporates these effects can be written as [4, 18, 28, 40]

$$\begin{aligned}\frac{\partial C(x,t)}{\partial t} &= D\frac{\partial^2 C(x,t)}{\partial x^2} - \chi\frac{\partial}{\partial x}\left(C(x,t)\frac{\partial G(x,t)}{\partial x}\right)\\ &\quad + \lambda C(x,t)\left(1-\frac{C(x,t)}{K}\right), \qquad (2.10)\\ \frac{\partial G(x,t)}{\partial t} &= D_g\frac{\partial^2 G(x,t)}{\partial x^2} + k_1 C(x,t) - k_2 G(x,t), \qquad (2.11)\end{aligned}$$

where χ is the chemotactic sensitivity coefficient, D_g is the diffusivity of the chemotactic chemical, k_1 is the rate at which cells produce the chemotactic chemical, and k_2 is the rate at which the chemotactic chemical undergoes decay. This model can be used to simulate chemoattraction by setting $\chi > 0$ or chemorepulsion by setting $\chi < 0$ [18, 28, 40]. Comparing this chemotactic extension of the Fisher-Kolmogorov equation with the standard model, Equation (2.4), indicates that there are an additional four parameters to estimate in order to apply the chemotaxis model: χ, D_g, k_1 and k_2. Given that standard applications of the IncuCyte ZOOM$^{\mathrm{TM}}$ assay do not attempt to make any measurement of the presence of any putative chemotactic factor, $G(x,t)$ [9, 30, 31], nor have we made any measurements of χ, D_g, k_1 or k_2, we do not attempt to calibrate this more complicated chemotaxis model to our IncuCyte ZOOM$^{\mathrm{TM}}$ assay data. Instead, we suggest that if this kind of chemotaxis model were to be applied to an IncuCyte ZOOM$^{\mathrm{TM}}$ assay data set, additional experimental measurements of these kinds of details are warranted.

Availability of supporting data

The data set supporting our results is included within the article and the supplementary material document.

Abbreviations

EGF - Epidermal growth factor.

Competing interests

The authors declare that they have no competing interests.

Author's contributions

STJ, ETS, LKC, DLSM and MJS conceived the study and designed the experiments. STJ and ETS performed the experiments. STJ, DLSM and MJS analysed the data. STJ and MJS wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgments

We appreciate experimental assistance and advice from Dr Penny Jeffery. This work is supported by the Australian Research Council (DP140100249, FT130100148), the National Health and Research Council (Australia), the Movember Foundation and the Prostate Cancer Foundation of Australia through a Movember Revolutionary Team Award. We also thank the three referees who made valuable suggestions that improved this work.

Bibliography

[1] Baldock, A. L., Ahn, S., Rockne, R., Johnston, S., Neal, M., Corwin, D., Clark-Swanson, K., Sterin, G., Trister, A. D., Malone, H., Ebiana, V., Sonaben, A. M., Mrugala, M., Rockhill, J. K., Silbergeld, D. L., Lai, A., Cloughesy, T., McKhann,

G. M., Bruce, J. N., Rostomily, R. C., Canoll, P. and Swanson, K. R. [2014], 'Patient-specific metrics of invasiveness reveal significant prognostic benefit of resection in a predictable subset of gliomas', *PLOS ONE* **9(10)**, e99057.

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

[3] Canny, J. [1986], 'A computational approach to edge detection', *IEEE Transactions on Pattern Analysis and Machine Intelligence* (6), 679–698.

[4] Charteris, N. and Khain, E. [2014], 'Modeling chemotaxis of adhesive cells: stochastic lattice approach and continuum description', *New Journal of Physics* **16**(2), 025002.

[5] Coleman, T. F. and Li, Y. [1996], 'An interior trust region approach for nonlinear minimization subject to bounds', *SIAM Journal on Optimization* **6**(2), 418–445.

[6] Curtis, C. W. and Bortz, D. M. [2012], 'Propagation of fronts in the Fisher-Kolmogorov equation with spatially varying diffusion', *Physical Review E* **86**, 066108.

[7] EssenBioScience [Accessed: January 2015], 'Incucyte zoom live cell imaging', http://www.essenbioscience.com/essen-products/incucyte/.

[8] Fisher, R. A. [1937], 'The wave of advance of advantageous genes', *Annals of Eugenics* **7**(4), 355–369.

[9] Gujral, T. S., Chan, M., Peshkin, L., Sorger, P. K., Kirschner, M. W. and MacBeath, G. [2014], 'A noncanonical frizzled2 pathway regulates epithelial-mesenchymal transition and metastasis', *Cell* **159**, 844–856.

[10] Habbal, A., Barelli, H. and Malandin, G. [2014], 'Assessing the ability of the 2D FisherKPP equation to model cell-sheet wound closure', *Mathematical Biosciences* **252**, 45–59.

[11] Hammond, J. F. and Bortz, D. M. [2011], 'Analytical solutions to Fishers equation with time-variable coefficients', *Applied Mathematics and Computation* **218**, 2497–2508.

[12] Harris, R. C., Chung, E. and Coffey, R. J. [2003], 'EGF receptor ligands', *Experimental Cell Research* **284**, 2–13.

[13] Herbst, R. S. [2004], 'Review of epidermal growth factor receptor biology', *International Journal of Radiation Oncology Biology Physics* **59**, 21–26.

[14] Jarrard, D. F., Blitz, B. F., Smith, R. C., Patai, B. L. and Rukstalis, D. B. [1994], 'Effect of epidermal growth factor on prostate cancer cell line PC3 growth and invasion', *Prostate* **24**, 46–53.

[15] Johnston, S. T., Simpson, M. J. and McElwain, D. L. S. [2014], 'How much information can be obtained from tracking the position of the leading edge in a scratch assay?', *Journal of the Royal Society Interface* **11**(97), 20140325.

[16] Johnston, S. T., Simpson, M. J., McElwain, D. L. S., Binder, B. J. and Ross, J. V. [2014], 'Interpreting scratch assays using pair density dynamics and approximate Bayesian computation', *Open Biology* **4**, 140097.

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

[18] Keller, E. F. and Segel, L. A. [1971], 'Model for chemotaxis', *Journal of Theoretical Biology* **30**(2), 225–234.

[19] Khain, E., Katakowski, M., Charteris, N., Jiang, F. and Chopp, M. [2012], 'Migration of adhesive glioma cells: Front propagation and fingering', *Physical Review E* **86**(1), 011904.

[20] Khain, E., Katakowski, M., Hopkins, S., Szalad, A., Zheng, X., Jiang, F. and Chopp, M. [2011], 'Collective behavior of brain tumor cells: the role of hypoxia', *Physical Review E* **83**(3), 031920.

[21] Kharait, S., Hautaniemi, S., Wu, S., Iwabu, A., Lauffenburger, D. A. and Wells, A. [2007], 'Decision tree modeling predicts effects of inhibiting contractility signaling on cell motility', *BMC Systems Biolgy* **1**(1), 9.

[22] Kolmogorov, A., Petrovsky, A. and Piscounov, N. [1937], 'Etude de lequation de la diffusion avec croissance de la quantite de matiere et son application a unprobleme biologique.', *Moscow University Mathematics Bulletin* **1**, 1–25.

[23] Liang, C.-C., Park, A. Y. and Guan, J.-L. [2007], 'In vitro scratch assay: a convenient and inexpensive method for analysis of cell migration in vitro', *Nature Protocols* **2**(2), 329–333.

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

[25] Maini, P. K., McElwain, D. L. S. and Leavesley, D. I. [2004*b*], 'Travelling waves in a wound healing assay', *Applied Mathematics Letters* **17**(5), 575–580.

[26] Mathworks [Accessed: January 2015], 'Image processing toolbox user guide r2013b', http://www.mathworks.com.au/help/images/index.html.

[27] Morton, K. W. and Mayers, D. F. [2005], *Numerical solution of partial differential equations*, Cambridge University Press, Cambridge.

[28] Murray, J. D. [n.d.], *Mathematical Biology I: An Introduction.*

[29] Rockne, R. C., Trister, A. D., Jacobs, J., Hawkins-Daarud, A. J., Neal, M. L., Hendrickson, K., Mrugala, M. M., Rockhill, J. K., Kinahan, P., Krohn, K. K. and Swanson, K. R. [2014], 'A patient-specific computational model of hypoxia-modulated radiation resistance in glioblastoma using 18F-FMISO-PET', *Journal of the Royal Society Interface* **12**, 20141174.

[30] Salomon, C., Kobayashi, M., Ashman, K., Sobrevia, L., Mitchell, M. D. and Rice, G. E. [2013], 'Hypoxia-induced changes in the bioactivity of cytotrophoblast-derived exosomes', *PLOS ONE* **8(11)**, e79636.

[31] Salomon, C., Torres, M. J., Kobayashi, M., Scholz-Romero, K., Sobrevia, L., Dobierzewska, A., Illanes, S. E., Mitchell, M. D. and Rice, G. E. [2014], 'A gestational profile of placental exosomes in maternal plasma and their effects on endothelial cell migration', *PLOS ONE* **9(6)**, e98667.

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

[33] Schultz, G., Chegini, N., Grant, M., Khaw, P. and MacKay, S. [1992], 'Effects of growth factors on corneal wound healing', *Acta Ophthalmologica* **70**(S202), 60–66.

[34] Scott, J. G., Basanta, D., Anderson, A. R. A. and Gerlee, P. [2013], 'A mathematical model of tumour self-seeding reveals secondary metastatic deposits as drivers of primary tumour growth', *Journal of the Royal Society Interface* **10**, 20130011.

[35] Sengers, B. G., Please, C. P. and Oreffo, R. O. [2007], 'Experimental characterization and computational modelling of two-dimensional cell spreading for skeletal regeneration', *Journal of the Royal Society Interface* **4**(17), 1107–1117.

[36] Sheardown, H. and Cheng, Y.-L. [1996], 'Mechanisms of corneal epithelial wound healing', *Chemical Engineering Science* **51**(19), 4517–4529.

[37] Sherratt, J. A. and Murray, J. D. [1990], 'Models of epidermal wound healing', *Proceedings of the Royal Society London, Series B* **241**, 29–36.

[38] Simpson, M. J. [2009], 'Depth-averaging errors in reactive transport modeling', *Water Resources Research* **45**, W02505.

[39] Simpson, M. J., Landman, K. A. and Hughes, B. D. [2009], 'Multi-species simple exclusion processes', *Physica A: Statistical Mechanics and its Applications* **388**(4), 399–406.

[40] Simpson, M. J., Landman, K. A., Hughes, B. D. and Newgreen, D. F. [2006], 'Looking inside an invasion wave of cells using continuum models: proliferation is the key', *Journal of Theoretical Biology* **243**(3), 343–360.

[41] Simpson, M. J., Treloar, K. K., Binder, B. J., Haridas, P., Manton, K. J., Leavesley, D. I., McElwain, D. L. S. and Baker, R. E. [2013], 'Quantifying the roles of cell motility and cell proliferation in a circular barrier assay', *Journal of the Royal Society Interface* **10**(82), 20130007.

[42] Swanson, K. R., Bridge, C., Murray, J. D. and Alvord Jr, E. C. [2003], 'Virtual and real brain tumors: using mathematical modeling to quantify glioma growth and invasion', *Journal of the Neurological Sciences* **216**(1), 1–10.

[43] Treloar, K. K. and Simpson, M. J. [2013], 'Sensitivity of edge detection methods for quantifying cell migration assays', *PLOS ONE* **8**(6), e67389.

[44] Treloar, K. K., Simpson, M. J., Haridas, P., Manton, K. J., Leavesley, D. I., McElwain, D. L. S. and Baker, R. E. [2013], 'Multiple types of data are required to identify the mechanisms influencing the spatial expansion of melanoma cell colonies', *BMC Systems Biology* **7**(1), 137.

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

[46] Tremel, A., Cai, A., Tirtaatmadja, N., Hughes, B. D., Stevens, G. W., Landman, K. A. and O'Connor, A. J. [2009], 'Cell migration and proliferation during monolayer formation and wound healing', *Chemical Engineering Science* **64**(2), 247–253.

[47] Zheng, C. and Bennett, G. D. [2002], *Applied Contaminant Transport Modeling*, John Wiley and Sons, New York.

2.5 Supplementary material

2.5.1 Carrying capacity density

The carrying capacity density is calculated by counting the number of cells within each subregion at the final experimental time point ($t = 46$ hours) and dividing by the area of the subregion. For a schematic, see Figure 2.2 (main document). Each subregion is the same size and has dimension 160.32 μm × 236.36 μm, resulting in an area of 3.79×10^4 μm^2.

	Region			
Replicate	Red	Green	Yellow	Mean
1	1.19×10^{-3}	1.19×10^{-3}	1.03×10^{-3}	1.13×10^{-3}
2	1.13×10^{-3}	1.13×10^{-3}	1.13×10^{-3}	1.13×10^{-3}
3	1.08×10^{-3}	1.21×10^{-3}	1.03×10^{-3}	1.11×10^{-3}
Standard Deviation		1.52×10^{-5}	Mean	1.13×10^{-3}

Table 2.2: Carrying capacity density (cells/μm^2) for the control assay.

	Region			
Replicate	Red	Green	Yellow	Mean
1	1.08×10^{-3}	1.11×10^{-3}	1.00×10^{-3}	1.06×10^{-3}
2	1.06×10^{-3}	1.08×10^{-3}	1.00×10^{-3}	1.05×10^{-3}
3	1.03×10^{-3}	1.03×10^{-3}	0.98×10^{-3}	1.01×10^{-3}
Standard Deviation		2.69×10^{-5}	Mean	1.04×10^{-3}

Table 2.3: Carrying capacity density (cells/μm^2) for the EGF-25 assay.

	Region			
Replicate	Red	Green	Yellow	Mean
1	1.06×10^{-3}	1.19×10^{-3}	1.11×10^{-3}	1.12×10^{-3}
2	1.11×10^{-3}	1.21×10^{-3}	1.08×10^{-3}	1.13×10^{-3}
3	1.16×10^{-3}	1.13×10^{-3}	1.06×10^{-3}	1.12×10^{-3}
Standard Deviation		1.01×10^{-5}	Mean	1.12×10^{-3}

Table 2.4: Carrying capacity density (cells/μm^2) for the EGF-50 assay.

Region				
Replicate	Red	Green	Yellow	Mean
1	1.16×10^{-3}	1.13×10^{-3}	1.03×10^{-3}	1.11×10^{-3}
2	1.24×10^{-3}	1.13×10^{-3}	1.03×10^{-3}	1.13×10^{-3}
3	1.19×10^{-3}	1.11×10^{-3}	1.06×10^{-3}	1.12×10^{-3}
Standard Deviation		1.34×10^{-5}	Mean	1.12×10^{-3}

Table 2.5: Carrying capacity density (cells/μm^2) for the EGF-75 assay.

Region				
Replicate	Red	Green	Yellow	Mean
1	1.08×10^{-3}	1.13×10^{-3}	1.11×10^{-3}	1.11×10^{-3}
2	1.27×10^{-3}	1.16×10^{-3}	1.16×10^{-3}	1.20×10^{-3}
3	1.16×10^{-3}	1.27×10^{-3}	1.08×10^{-3}	1.17×10^{-3}
Standard Deviation		4.51×10^{-5}	Mean	1.16×10^{-3}

Table 2.6: Carrying capacity density (cells/μm^2) for the EGF-100 assay.

Region				
Replicate	Red	Green	Yellow	Mean
1	1.16×10^{-3}	1.13×10^{-3}	1.06×10^{-3}	1.12×10^{-3}
2	1.16×10^{-3}	1.13×10^{-3}	1.03×10^{-3}	1.11×10^{-3}
3	1.13×10^{-3}	1.06×10^{-3}	1.08×10^{-3}	1.09×10^{-3}
Standard Deviation		1.34×10^{-5}	Mean	1.11×10^{-3}

Table 2.7: Carrying capacity density (cells/μm^2) for the EGF-125 assay.

2.5.2 Proliferation rate

The cell density is calculated by counting the number of cells within each subregion at five experimental time points ($t = 0,\ 8,\ 16,\ 24$ and 46 hours) and dividing by the area of the subregion. For a schematic, see Figure 2.3 (main document). Each subregion is the same size and has dimension 160.32 μm × 236.36 μm, resulting in an area of 3.79×10^4 μm^2. We use this cell density to obtain the proliferation rate from the logistic equation. For a full explanation, see the main document.

	Region			
Time (hours)	Red	Green	Yellow	Mean
0	0.79×10^{-3}	0.69×10^{-3}	0.53×10^{-3}	0.67×10^{-3}
8	0.98×10^{-3}	0.87×10^{-3}	0.61×10^{-3}	0.82×10^{-3}
16	1.03×10^{-3}	0.92×10^{-3}	0.69×10^{-3}	0.88×10^{-3}
24	1.11×10^{-3}	0.98×10^{-3}	0.84×10^{-3}	0.98×10^{-3}
46	1.19×10^{-3}	1.19×10^{-3}	1.03×10^{-3}	1.13×10^{-3}

Table 2.8: Cell densities (cells/μm^2) for the control assay, replicate one.

	Region			
Time (hours)	Red	Green	Yellow	Mean
0	0.71×10^{-3}	0.71×10^{-3}	0.48×10^{-3}	0.63×10^{-3}
8	0.77×10^{-3}	0.84×10^{-3}	0.58×10^{-3}	0.73×10^{-3}
16	0.84×10^{-3}	0.90×10^{-3}	0.66×10^{-3}	0.80×10^{-3}
24	0.98×10^{-3}	0.95×10^{-3}	0.82×10^{-3}	0.91×10^{-3}
46	1.13×10^{-3}	1.13×10^{-3}	1.13×10^{-3}	1.13×10^{-3}

Table 2.9: Cell densities (cells/μm^2) for the control assay, replicate two.

	Region			
Time (hours)	Red	Green	Yellow	Mean
0	0.74×10^{-3}	0.77×10^{-3}	0.74×10^{-3}	0.75×10^{-3}
8	0.84×10^{-3}	0.82×10^{-3}	0.84×10^{-3}	0.84×10^{-3}
16	0.92×10^{-3}	0.90×10^{-3}	0.87×10^{-3}	0.90×10^{-3}
24	0.98×10^{-3}	0.95×10^{-3}	0.95×10^{-3}	0.96×10^{-3}
46	1.08×10^{-3}	1.21×10^{-3}	1.03×10^{-3}	1.11×10^{-3}

Table 2.10: Cell densities (cells/μm^2) for the control assay, replicate three.

	Region			
Time (hours)	Red	Green	Yellow	Mean
0	0.77×10^{-3}	0.79×10^{-3}	0.42×10^{-3}	0.66×10^{-3}
8	0.82×10^{-3}	0.84×10^{-3}	0.50×10^{-3}	0.72×10^{-3}
16	0.84×10^{-3}	0.90×10^{-3}	0.66×10^{-3}	0.80×10^{-3}
24	0.98×10^{-3}	0.98×10^{-3}	0.77×10^{-3}	0.91×10^{-3}
46	1.08×10^{-3}	1.11×10^{-3}	1.00×10^{-3}	1.06×10^{-3}

Table 2.11: Cell densities (cells/μm^2) for the EGF-25 assay, replicate one.

	Region			
Time (hours)	Red	Green	Yellow	Mean
0	0.66×10^{-3}	0.77×10^{-3}	0.63×10^{-3}	0.69×10^{-3}
8	0.77×10^{-3}	0.87×10^{-3}	0.66×10^{-3}	0.77×10^{-3}
16	0.82×10^{-3}	1.00×10^{-3}	0.95×10^{-3}	0.92×10^{-3}
24	0.87×10^{-3}	1.00×10^{-3}	0.95×10^{-3}	0.94×10^{-3}
46	1.06×10^{-3}	1.08×10^{-3}	1.00×10^{-3}	1.05×10^{-3}

Table 2.12: Cell densities (cells/μm^2) for the EGF-25 assay, replicate two.

	Region			
Time (hours)	Red	Green	Yellow	Mean
0	0.45×10^{-3}	0.74×10^{-3}	0.55×10^{-3}	0.58×10^{-3}
8	0.53×10^{-3}	0.82×10^{-3}	0.69×10^{-3}	0.68×10^{-3}
16	0.71×10^{-3}	0.92×10^{-3}	0.71×10^{-3}	0.78×10^{-3}
24	0.82×10^{-3}	0.92×10^{-3}	0.77×10^{-3}	0.84×10^{-3}
46	1.03×10^{-3}	1.03×10^{-3}	0.98×10^{-3}	1.01×10^{-3}

Table 2.13: Cell densities (cells/μm^2) for the EGF-25 assay, replicate three.

	Region			
Time (hours)	Red	Green	Yellow	Mean
0	0.71×10^{-3}	0.98×10^{-3}	0.87×10^{-3}	0.85×10^{-3}
8	0.79×10^{-3}	0.98×10^{-3}	0.90×10^{-3}	0.89×10^{-3}
16	0.95×10^{-3}	1.11×10^{-3}	1.00×10^{-3}	1.02×10^{-3}
24	1.00×10^{-3}	1.13×10^{-3}	1.06×10^{-3}	1.06×10^{-3}
46	1.06×10^{-3}	1.19×10^{-3}	1.11×10^{-3}	1.12×10^{-3}

Table 2.14: Cell densities (cells/μm^2) for the EGF-50 assay, replicate one.

	Region			
Time (hours)	Red	Green	Yellow	Mean
0	0.63×10^{-3}	0.63×10^{-3}	0.74×10^{-3}	0.67×10^{-3}
8	0.82×10^{-3}	0.84×10^{-3}	0.87×10^{-3}	0.84×10^{-3}
16	0.98×10^{-3}	1.00×10^{-3}	0.87×10^{-3}	0.96×10^{-3}
24	1.03×10^{-3}	1.08×10^{-3}	0.98×10^{-3}	1.03×10^{-3}
46	1.11×10^{-3}	1.21×10^{-3}	1.08×10^{-3}	1.13×10^{-3}

Table 2.15: Cell densities (cells/μm^2) for the EGF-50 assay, replicate two.

	Region			
Time (hours)	Red	Green	Yellow	Mean
0	1.00×10^{-3}	0.82×10^{-3}	0.63×10^{-3}	0.82×10^{-3}
8	1.00×10^{-3}	1.00×10^{-3}	0.79×10^{-3}	0.93×10^{-3}
16	1.06×10^{-3}	1.06×10^{-3}	0.84×10^{-3}	0.99×10^{-3}
24	1.11×10^{-3}	1.08×10^{-3}	0.90×10^{-3}	1.03×10^{-3}
46	1.16×10^{-3}	1.13×10^{-3}	1.06×10^{-3}	1.12×10^{-3}

Table 2.16: Cell densities (cells/μm^2) for the EGF-50 assay, replicate three.

	Region			
Time (hours)	Red	Green	Yellow	Mean
0	0.69×10^{-3}	0.63×10^{-3}	0.55×10^{-3}	0.62×10^{-3}
8	0.79×10^{-3}	0.77×10^{-3}	0.63×10^{-3}	0.73×10^{-3}
16	0.87×10^{-3}	0.87×10^{-3}	0.82×10^{-3}	0.85×10^{-3}
24	0.90×10^{-3}	0.92×10^{-3}	0.92×10^{-3}	0.91×10^{-3}
46	1.16×10^{-3}	1.13×10^{-3}	1.03×10^{-3}	1.11×10^{-3}

Table 2.17: Cell densities (cells/μm^2) for the EGF-75 assay, replicate one.

	Region			
Time (hours)	Red	Green	Yellow	Mean
0	0.61×10^{-3}	0.69×10^{-3}	0.69×10^{-3}	0.66×10^{-3}
8	0.77×10^{-3}	0.79×10^{-3}	0.79×10^{-3}	0.78×10^{-3}
16	0.87×10^{-3}	0.95×10^{-3}	0.84×10^{-3}	0.89×10^{-3}
24	0.95×10^{-3}	0.98×10^{-3}	0.90×10^{-3}	0.94×10^{-3}
46	1.24×10^{-3}	1.13×10^{-3}	1.03×10^{-3}	1.13×10^{-3}

Table 2.18: Cell densities (cells/μm^2) for the EGF-75 assay, replicate two.

	Region			
Time (hours)	Red	Green	Yellow	Mean
0	0.87×10^{-3}	0.82×10^{-3}	0.58×10^{-3}	0.76×10^{-3}
8	0.90×10^{-3}	0.90×10^{-3}	0.71×10^{-3}	0.84×10^{-3}
16	1.08×10^{-3}	0.95×10^{-3}	0.77×10^{-3}	0.93×10^{-3}
24	1.19×10^{-3}	0.98×10^{-3}	0.95×10^{-3}	1.04×10^{-3}
46	1.19×10^{-3}	1.11×10^{-3}	1.06×10^{-3}	1.12×10^{-3}

Table 2.19: Cell densities (cells/μm^2) for the EGF-75 assay, replicate three.

	Region			
Time (hours)	Red	Green	Yellow	Mean
0	0.66×10^{-3}	0.79×10^{-3}	0.63×10^{-3}	0.69×10^{-3}
8	0.82×10^{-3}	0.95×10^{-3}	0.74×10^{-3}	0.84×10^{-3}
16	0.87×10^{-3}	1.00×10^{-3}	0.84×10^{-3}	0.91×10^{-3}
24	0.92×10^{-3}	1.06×10^{-3}	0.90×10^{-3}	0.96×10^{-3}
46	1.08×10^{-3}	1.13×10^{-3}	1.11×10^{-3}	1.11×10^{-3}

Table 2.20: Cell densities (cells/μm^2) for the EGF-100 assay, replicate one.

	Region			
Time (hours)	Red	Green	Yellow	Mean
0	0.77×10^{-3}	0.71×10^{-3}	0.74×10^{-3}	0.74×10^{-3}
8	0.92×10^{-3}	0.84×10^{-3}	0.84×10^{-3}	0.87×10^{-3}
16	0.98×10^{-3}	0.95×10^{-3}	0.87×10^{-3}	0.93×10^{-3}
24	1.00×10^{-3}	1.08×10^{-3}	0.98×10^{-3}	1.02×10^{-3}
46	1.27×10^{-3}	1.16×10^{-3}	1.16×10^{-3}	1.20×10^{-3}

Table 2.21: Cell densities (cells/μm^2) for the EGF-100 assay, replicate two.

	Region			
Time (hours)	Red	Green	Yellow	Mean
0	0.84×10^{-3}	0.77×10^{-3}	0.50×10^{-3}	0.70×10^{-3}
8	1.00×10^{-3}	0.90×10^{-3}	0.74×10^{-3}	0.88×10^{-3}
16	1.00×10^{-3}	0.98×10^{-3}	0.92×10^{-3}	0.97×10^{-3}
24	1.00×10^{-3}	1.00×10^{-3}	1.03×10^{-3}	1.01×10^{-3}
46	1.16×10^{-3}	1.27×10^{-3}	1.08×10^{-3}	1.17×10^{-3}

Table 2.22: Cell densities (cells/μm^2) for the EGF-100 assay, replicate three.

	Region			
Time (hours)	Red	Green	Yellow	Mean
0	0.98×10^{-3}	0.74×10^{-3}	0.74×10^{-3}	0.82×10^{-3}
8	1.00×10^{-3}	0.87×10^{-3}	0.84×10^{-3}	0.91×10^{-3}
16	1.03×10^{-3}	0.98×10^{-3}	0.90×10^{-3}	0.97×10^{-3}
24	1.11×10^{-3}	1.00×10^{-3}	0.92×10^{-3}	1.01×10^{-3}
46	1.16×10^{-3}	1.13×10^{-3}	1.06×10^{-3}	1.12×10^{-3}

Table 2.23: Cell densities (cells/μm^2) for the EGF-125 assay, replicate one.

	Region			
Time (hours)	Red	Green	Yellow	Mean
0	0.74×10^{-3}	0.84×10^{-3}	0.63×10^{-3}	0.74×10^{-3}
8	0.79×10^{-3}	0.90×10^{-3}	0.74×10^{-3}	0.81×10^{-3}
16	0.92×10^{-3}	0.98×10^{-3}	0.79×10^{-3}	0.90×10^{-3}
24	1.06×10^{-3}	1.06×10^{-3}	0.90×10^{-3}	1.00×10^{-3}
46	1.16×10^{-3}	1.13×10^{-3}	1.03×10^{-3}	1.11×10^{-3}

Table 2.24: Cell densities (cells/μm^2) for the EGF-125 assay, replicate two.

	Region			
Time (hours)	Red	Green	Yellow	Mean
0	0.87×10^{-3}	0.74×10^{-3}	0.63×10^{-3}	0.75×10^{-3}
8	0.92×10^{-3}	0.92×10^{-3}	0.77×10^{-3}	0.87×10^{-3}
16	0.95×10^{-3}	0.95×10^{-3}	0.82×10^{-3}	0.91×10^{-3}
24	1.06×10^{-3}	1.00×10^{-3}	0.87×10^{-3}	0.98×10^{-3}
46	1.13×10^{-3}	1.06×10^{-3}	1.08×10^{-3}	1.09×10^{-3}

Table 2.25: Cell densities (cells/μm^2) for the EGF-125 assay, replicate three.

2.5.3 Leading edge position

The leading edge position is obtained from the edge detection algorithm described in the main document. Here we report the averaged position obtained from the three replicates for each EGF concentration, as well as the standard deviation.

Time (hours)	0	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30
$\langle L_E \rangle$ (μm)	361	341	325	306	295	286	275	260	242	228	209	193	171	151	133	107
St. Dev. (μm)	11	36	52	91	91	92	89	96	89	77	89	75	69	78	71	85

Table 2.26: Leading edge position (μm) for the control assay.

Time (hours)	0	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30
$\langle L_E \rangle$ (μm)	346	330	301	291	270	260	232	221	203	183	166	153	136	123	105	84
St. Dev. (μm)	42	59	73	78	96	97	127	124	136	159	163	155	164	161	166	163

Table 2.27: Leading edge position (μm) for the EGF-25 assay.

Time (hours)	0	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30
$\langle L_E \rangle$ (μm)	318	299	289	276	267	251	227	206	187	168	133	104	84	64	52	39
St. Dev. (μm)	44	46	34	42	34	38	52	69	74	87	122	133	135	128	97	82

Table 2.28: Leading edge position (μm) for the EGF-50 assay.

Time (hours)	0	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30
$\langle L_E \rangle$ (μm)	363	350	336	320	310	305	279	252	228	210	184	161	140	118	90	72
St. Dev. (μm)	41	44	59	69	73	60	86	76	86	75	56	62	51	59	74	69

Table 2.29: Leading edge position (μm) for the EGF-75 assay.

Time (hours)	0	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30
$\langle L_E \rangle$ (μm)	363	339	326	308	292	284	255	223	198	185	151	135	110	90	68	41
St. Dev. (μm)	37	31	54	44	60	65	79	122	113	126	153	136	146	131	123	95

Table 2.30: Leading edge position (μm) for the EGF-100 assay.

Time (hours)	0	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30
$\langle L_E \rangle$ (μm)	359	341	325	296	283	251	231	208	176	158	133	113	93	66	58	38
St. Dev. (μm)	24	28	20	34	55	65	71	80	49	53	65	56	51	20	21	31

Table 2.31: Leading edge position (μm) for the EGF-125 assay.

Chapter 3

How much information can be obtained from tracking the position of the leading edge in a scratch assay?

A paper published in the *Journal of the Royal Society Interface*.

Johnston, S. T., Simpson, M. J. and McElwain, D. L. S. [2014], 'How much information can be obtained from tracking the position of the leading edge in a scratch assay?', *Journal of the Royal Society Interface* **11**(97), 20140325.

Abstract

Moving cell fronts are an essential feature of wound healing, development and disease. The rate at which a cell front moves is driven, in part, by the cell motility, quantified in terms of the cell diffusivity D, and the cell proliferation rate λ. Scratch assays are a commonly-reported procedure used to investigate the motion of cell fronts where an initial cell monolayer is scratched and the motion of the front is monitored over a short period of time, often less than 24 hours. The simplest way of quantifying a scratch assay is to monitor the progression of the leading edge. Leading edge data is very convenient since, unlike other methods, it is nondestructive and does not require labeling, tracking or counting individual cells amongst the population. In this work we study short time leading edge data in a scratch assay using a discrete mathematical model and automated image analysis with the aim of investigating whether such data allows us to reliably identify D and λ. Using a naïve calibration approach where we simply scan the relevant region of the (D, λ) parameter space, we show that there are many choices of D and λ for which our model produces indistinguishable short time leading edge data. Therefore, without due care, it is impossible to estimate D and λ from this kind of data. To address this, we present a modified approach accounting for the fact that cell motility occurs over a much shorter time scale than proliferation. Using this information we divide the duration of the experiment into two periods, and we estimate D using data from the first period, while we estimate λ using data from the second period. We confirm the accuracy of our approach using *in silico* data and a new set of *in vitro* data, which shows that our method

recovers estimates of D and λ that are consistent with previously-reported values except that that our approach is fast, inexpensive, nondestructive and avoids the need for cell labeling and cell counting.

3.1 Introduction

Moving cell fronts are key features of tissue repair [15] and tumour spreading [28]. The rate at which the front of a population of cells moves is influenced by the rate at which individual cells within the population migrate and proliferate [27]. Random, undirected cell migration is typically quantified in terms of the cell diffusivity D, while cell proliferation is quantified in terms of the proliferation rate λ. Developing methods to estimate D and λ from experimental observations is important so that we can assess the effectiveness of intervention strategies which often aim at influencing either D or λ [22, 36]. For example, drugs such as Mitomycin-C, which inhibit proliferation [22], are used to reduce tumour spreading [36], whereas steroid treatment, which stimulates cell migration [11], is often studied with the aim of enhancing wound healing.

Scratch assays [19, 21, 31, 38], also known as scrape or wound healing assays [31, 38], are routinely used to investigate the motion of cell fronts by creating a scratch in a cell monolayer and observing the motion of the cell front. Images of the front are captured over a period of time that is typically less than 24 hours [5, 19, 24]. Short time scale experimental data is very common since it avoids the need for replenishing the nutrients in the assay. There are various ways that data from a scratch assay are reported and analysed. The most common method is to present a qualitative, visual comparison between a control assay and another assay where some treatment has been applied. This kind of data is often presented without any attempt to estimate D or λ. For example, Teppo [31] presented scratch assay data showing that hypoxia increased the rate at which the fronts of cancer cells moved, but they did not determine how the hypoxic conditions affected D and/or λ.

Another approach to analyse scratch assays is to use a mathematical model, such as the Fisher-Kolmogorov equation [9] or an extension of this reaction-diffusion equation [16, 23, 29, 30] (Supplementary material). Some previous studies have focused on matching the experimental front speed with the long time asymptotic travelling wave speed of the Fisher-Kolmogorov equation, $c = \sqrt{4\lambda D}$ [7, 13, 14]. Unfortunately, this approach is of little practical use for most experiments which are conducted over short time scales where no such travelling wave forms [5, 19, 24]. Another way of analysing scratch assays is to generate cell density profiles which can be matched to numerical solutions of a reaction-diffusion equation [20, 27, 35]. Unfortunately, this approach is expensive and time consuming since it requires some kind of direct or indirect cell counting technique to construct the density profiles. Other mathematical models have been used to interpret scratch assays, such as mechanistic [2] and biased continuum models [12]. However, the experimental procedures required to parameterise these models are time consuming since they involved individual cell counting [2] or individual cell tracking [12].

The simplest and most cost effective measurement that can be made to characterise a scratch assay is to record the location of the cell front as a function of time [1, 5, 21]. The widespread availability of automatic edge detection algorithms [33, 37] means that it is straightforward to obtain this information. Given that most scratch assays are conducted for short time periods, here we seek to determine whether it is possible to reliably estimate D and λ from short time leading edge data alone without constructing cell density profiles [20, 27, 35]. To explore this question, we use automatic edge detection algorithms to analyse a discrete model of collective cell spreading driven by cell migration and cell proliferation [26]. While such models have been used to analyse various types of *in vitro* assays previously [3, 20], these studies have not focused on short time leading edge data. Our work shows that great care must be taken when interpreting short time leading edge data since the most straightforward model calibration approach indicates that there are many choices of D and λ which lead to indistinguishable leading edge data. To overcome this we develop a novel method by dividing the leading edge time series data into two intervals allowing us to estimate D from the first time interval, and then we separately estimate λ using the second time interval. We test the method using both *in silico* and *in vitro* data showing that we recover estimates of D and λ that are consistent with previously-reported results obtained using far more complicated experimental procedures.

This manuscript is organised in the following way. In Section 3.2 we describe a discrete model for simulating the motion of cell fronts. Section 3.2 describes the image analysis and experimental procedure. Data in Section 3.3 shows that a straightforward model calibration procedure implies that there are many choices of D and λ that match short time leading edge data. As a result, we also describe, in Section 3.3, a modified method that leads to unique estimates of D and λ, and we validate our results using both *in silico* and *in vitro* data. Finally, in Section 3.4, we discuss our results and outline options for extending the work.

3.2 Methods

3.2.1 Experimental method

The experimental method has been presented previously [27]. Briefly, murine fibroblast 3T3 cells [32] were grown in T175 cm^2 tissue culture flasks and one μL of cell suspension was placed into the well, with diameter 15.6 mm, of a tissue culture plate. The tissue culture plate was incubated at 37 °C and 5% CO$_2$ until the population became confluent. A scratch was made in the monolayer using a P1000 pipette tip (Lab Advantage, Australia). Images were recorded at $t = 0, 3, 6, 9, 12$ and 24 h, and a schematic illustration of the assay at $t = 0$ is given in Figure 3.1(a).

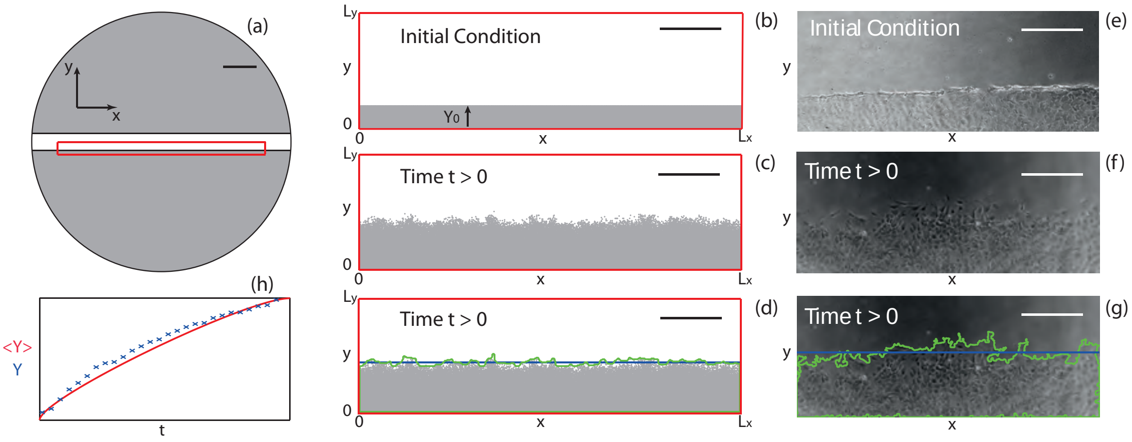

Figure 3.1: Schematic illustrating the initial scratch in the cell monolayer, simulation and leading edge data. (a) Monolayer of cells (grey) immediately after the scratch (white) has been made. The red rectangle indicates the spatial region which we simulate. (b) The initial confluent cell monolayer (grey) has height Y_0 and the width L_x, corresponding to the width of the red rectangle in (a). The height of the domain, L_y is chosen to be sufficiently large that the agents in the simulation never touch this boundary within the 24 hour period of the simulation. (c) Simulation after time t. (d) The simulation results are analysed using the image analysis tools to detect the leading edge (green) which is used to estimate the average position of the leading edge (blue). Scale bars in (a)-(d) are 2 mm. (e) Typical experimental image immediately after a scratch has been made. To illustrate the edge detection algorithm we show in (f) an experimental image and in (g) the detected leading edge (green) and the average position of the front (blue). Scale bars in (e)-(g) are 400 μm. (h) Typical temporal evolution of the position of the leading edge for experimental data (blue crosses) and averaged simulation data (red).

3.2.2 Mathematical model

We consider a lattice-based random walk model on a two-dimensional square lattice, with lattice spacing Δ [6, 26]. Each site may be occupied by, at most, one agent, and each simulation contains a total of $Z(t)$ agents which have the ability to move and proliferate, with probability $P_m \in [0, 1]$ and $P_p \in [0, 1]$, respectively, during each time step of duration τ. We make the standard assumption that P_m and P_p are constants, which are related to D and λ by

$$D = \frac{P_m \Delta^2}{4\tau}, \qquad \lambda = \frac{P_p}{\tau}, \tag{3.1}$$

which means that we can view the parameters (P_m, P_p) as being interchangeable with (D, λ). During each time step $Z(t)$ agents are chosen, at random, one at a time, and given the opportunity to move [26]. An agent at (x, y) will attempt to step to $(x \pm \Delta, y)$ or $(x, y \pm \Delta)$, with the target site chosen with equal probability. After $Z(t)$ potential motility events have been attempted, an additional $Z(t)$ agents are selected, at random, one at a time, and given the opportunity to proliferate. A proliferative agent at (x, y) will attempt to place a daughter agent at $(x \pm \Delta, y)$ or $(x, y \pm \Delta)$, with the target site chosen with equal probability. Potential motility and proliferation events will only succeed if the target site is vacant, otherwise the event is aborted. Implicitly, this means that individual agents in crowded regions will be relatively immobile and unable to proliferate, whereas uncrowded agents will behave differently and will have a greater opportunity to move and proliferate. The continuum-limit description of this model is a generalisation of the Fisher-Kolmogorov equation in two-dimensions [26] (Supplementary material). This description is valid only when the ratio P_p/P_m is sufficiently small [25, 26].

We apply this model to mimic the geometry of the scratch assay. In all results we set $\Delta = 25$ μm, corresponding to a typical cell diameter [27, 34]. The simulation domain, shown in Figure 3.1(b), is $0 \leq x \leq L_x$, $0 \leq y \leq L_y$. We choose $L_x = 12.5$ mm so that our domain captures almost the entire population within the well without directly simulating the curved boundaries. Although it is possible to simulate such curved geometries [27, 34], we neglect these details here since our experimental data, described in Section 3.3.3, focuses on several rectangular subregions within the well, away from the circular boundary. We choose $L_y = 3.75$ mm which is sufficient to ensure that agents in simulations never reach the boundary, $y = 3.75$ mm, during the 24 hour simulation period. Symmetry conditions are applied along the lines $x = 0$, $x = L_x$, $y = 0$ and $y = L_y$. To match our experimental conditions, agents are initially placed on the lattice so that the region $y < Y_0$ is confluent. All simulation data is presented for a particular choice of τ and we re-simulated all results with smaller values of τ to ensure our results are insensitive to τ.

The results in this work could have been generated using a lattice-free model [17, 18]. Instead, we chose a lattice-based model since lattice-free models with crowding effects are far more computationally expensive [17, 18]. Furthermore, our recent work showed that lattice-based and lattice-free models produce equivalent data at the leading edge [17, 18] which means that there is no advantage in using a lattice-free model here if we are focusing on leading edge data.

3.2.3 Image analysis

We use Matlab's Image Processing Toolbox to estimate the position of the leading edge from the experimental and modelling images. The experimental image is imported and converted to greyscale using `imread` and `rgb2gray`, respectively. The simulation data is converted from a matrix representing occupied and vacant sites into a greyscale image using `mat2gray`. Henceforth, the procedure for analysing the experimental images and simulation data is identical. Edges are detected using `edge` with the Canny method [4] and a threshold between 0.04 and 0.1. Detected edges weaker than the threshold are ignored. Remaining edges are dilated, using `imdilate`, by a stretching element, defined using `strel`, with a square element of size seven. Any remaining vacant spaces are filled, using `imfill`, after which the dilation was reversed by eroding the image with the stretching element, defined previously, using `imerode`. The edges within the image were smoothed using `medfilt2`, and the area enclosed by the leading edge estimated using `regionprops`. For illustrative purposes, this algorithm was applied to the simulation data in Figure 3.1(c) and the detected edge is superimposed in Figure 3.1(d). To estimate the vertical position of the leading edge, Y, we use

$$Y = \frac{A}{L_x}, \tag{3.2}$$

where A is the area enclosed by the detected leading edge. The average position of the leading edge, Y, is superimposed in Figure 3.1(d). To estimate how Y changes with time we repeat the process at many time points and subtract the initial position to give a measure of the net displacement of the leading edge as a function of time. Schematic results in Figure 3.1(h) indicate how the net displacement of the leading edge evolves with time for a representative set of experimental and averaged simulation data. We acknowledge that the edge detection could have been performed with ImageJ rather than Matlab. For this work we chose to use Matlab since our previous comparison of Matlab and ImageJ edge detection algorithms showed that Matlab allows greater flexibility in the choice and control of threshold and dilation parameters [33].

3.3 Results

In this work we will generate, and refer to, two distinct types of data: *experimental data* and *averaged simulation data.* The differences between these types of data can be described as follows,

1. Experimental data: Describes the position of the leading edge as a function of time obtained from a single experiment. Furthermore, we consider two different ways of generating experimental data:
   - *In vitro* experimental data: Corresponds to data from experimental images, and
   - *In silico* experimental data: Corresponds to data from discrete simulation images.
2. Averaged simulation data: Describes the average position of the leading edge, where the average has been constructed using data from many identically-prepared realisations of the discrete model, that is, simulations performed with an identical algorithm, initial condition and parameters.

We construct the averaged simulation data using

$$\langle Y^i \rangle = \frac{1}{M} \sum_{n=1}^{M} Y_n^i, \tag{3.3}$$

where Y_n^i is the position of the leading edge, at time step i, in the n^{th} identically-prepared realisation and M is the total number of identically-prepared realisations. To measure the differences between different sets of experimental data and averaged simulation data, we define

$$E = \frac{1}{NY_{\max}} \sqrt{\sum_{i=1}^{N} \left(\langle Y^i \rangle - Y^i\right)^2}, \tag{3.4}$$

where $\langle Y^i \rangle$ is the position of the leading edge, at time step i, using averaged simulation data, Y^i is the position of the leading edge, at time step i, using experimental data, N is the number of time steps and $Y_{\max}$ is the maximum value of Y^i, for $i = 1, 2, 3, \ldots, N$.

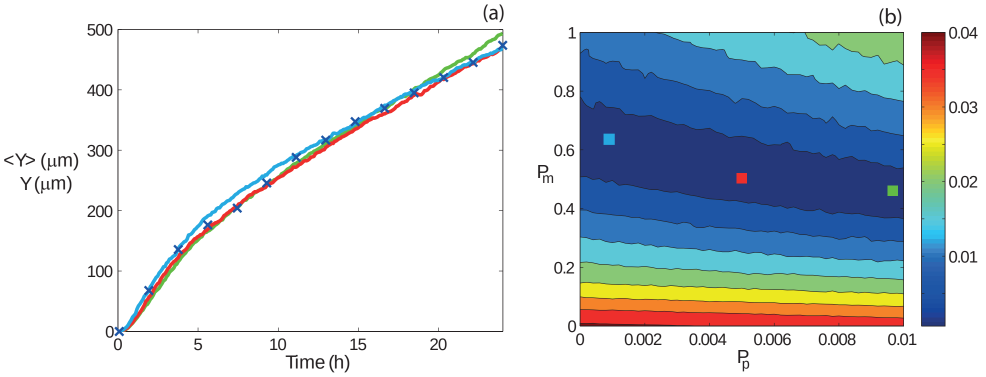

Figure 3.2: Comparison of *in silico* experimental data and averaged simulation data. (a) Leading edge *in silico* experimental data corresponds to $(P_m, P_p) = (0.5, 5 \times 10^{-3})$ (blue crosses). Data is presented at every 20th time step. (b) Contour plot of E (Equation (3.4)) measuring the difference between the *in silico* experimental data and averaged simulation data within the region $P_m \in [0, 1], P_p \in [0, 0.01]$. Simulation parameters are $M = 10, Y_0 = 750$ μm, $\Delta = 25$ μm and $\tau = 0.09191$ h, with a final time of 24 h. The contour plot of E was generated by considering 2601 different parameter combinations; 51 equally-spaced values of P_m, and 51 equally-spaced values of P_p. The light blue, green and red coloured squares in (b) correspond to three different parameter combinations: $(P_m, P_p) = (0.6, 1.4 \times 10^{-3})$, $(0.46, 8.8 \times 10^{-3})$ and $(0.5, 5 \times 10^{-3})$, respectively. Averaged simulation data from these three different parameter combinations are superimposed in (a), showing that all three parameter combinations lead to indistinguishable short time leading edge data. All averaged simulation data are insensitive to τ.

3.3.1 Naïve parameter recovery

To explore whether it is possible to reliably estimate D and λ from short time leading edge data we first analyse a representative set of *in silico* experimental data corresponding to $(P_m, P_p) = (0.5, 5 \times 10^{-3})$, with $\Delta = 25$ μm, which is reported in Figure 3.2(a). We note that it is difficult to draw specific conclusions by a simple visual inspection of this data set, with the exception that it appears that the front speed is not constant over this time interval. To analyse this data we generate a suite of averaged simulation data, sampling 2601 equally-spaced parameter combinations within the region $P_m \in [0, 1], P_p \in [0, 0.01]$, and we present a contour plot of E, given by Equation (3.4), in Figure 3.2(b). We expect that if there is a unique choice of D and λ that matches the data in Figure 3.2(a), we would see a unique minimum on the E surface. Instead, we observe a relatively large, flat region, within which E takes on small, indistinguishable, values. This region extends right across this portion of the parameter space, indicating that there are many combinations of D and λ which match the experimental data equally well. To demonstrate that our observations for this parameter set hold more generally, we repeated the process focusing on *in silico* experimental data with a higher proliferation rate and found similar results (Supplementary material).

To demonstrate the redundancy in the short time leading edge data we choose three different combinations of (P_m, P_p), highlighted in Figure 3.2(b), and we superimpose the corresponding averaged leading edge data on the experimental data in Figure 3.2(a). Comparing these data sets confirms that there are several parameter combinations which give indistinguishable short time leading edge data. Furthermore, we found that any parameter combination within the dark blue region in Figure 3.2(b) also gives averaged simulation data that matches the experimental data (not shown). These results indicate that short time leading edge information should be treated with care since a standard model calibration procedure may not provide useful information.

3.3.2 Parameter recovery accounting for the separation of time scales

Our results in Section 3.3.1 imply that additional information needs to be incorporated into our parameter estimation procedure if we are to infer useful information from short time leading edge data. Here we make use of the fact that there is a large separation of time scales between cell proliferation processes and cell motility processes. Typical estimates of the cell doubling time are approximately 15-30 hours [14, 27] whereas the time scale of cell motility events is approximately 10-20 minutes [10]. This separation of time scales implies that the first part of the leading edge time series data will be dominated by the influence of cell motility and we can make use of this information by dividing our time series data into two intervals: (i) $t < T$, and (ii) $t > T$, where T is a time interval during which the motion of the leading edge is dominated by cell motility. Intuitively, we expect that T ought to be chosen to be much less than the cell doubling time, and we will discuss this choice in Section 3.3.3.

To make use of this separation of time scales we estimate P_m and P_p iteratively:

Step 1 Estimate P_m by considering experimental data for $t < T$, we set $P_p = 0$ and systematically vary P_m so that our averaged simulation data matches the experimental data,

Step 2 Estimate P_p by considering experimental data for $t > T$, we set P_m to be the value found previously, and we systematically vary P_p so that our averaged simulation data matches the experimental data.

Step 3 Re-estimate P_m by considering experimental data for $t < T$, we set P_p to be the value found in step 2, and we systematically vary P_m so that our averaged simulation data matches the experimental data. Repeat steps 2 and 3 until both P_m and P_p converge.

We now apply this method to *in silico* experimental data and then examine *in vitro* data in Section 3.3.3. Figure 3.3(a) shows same *in silico* experimental data presented previously in Figure 3.2(a). The results from estimating P_m using the iterative procedure are given in Figure 3.3(b) and show that by choosing $T = 3$ h and focusing on the interval $P_m \in [0, 1]$ we observe a relatively well-defined minimum in the plot of E indicating that we have $P_m \approx 0.48$. The results from estimating P_p are given in Figure 3.3(c)

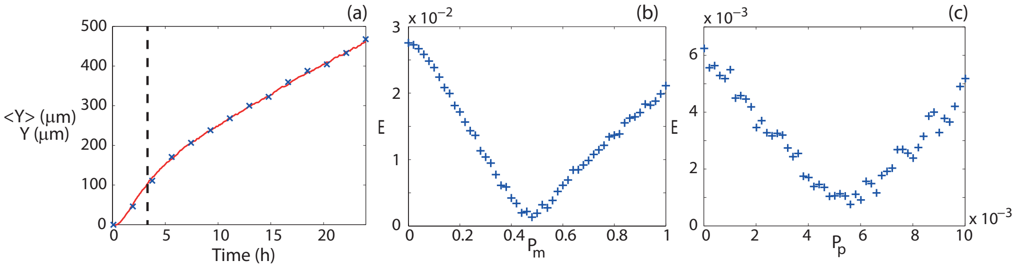

Figure 3.3: Parameter recovery for *in silico* experimental data using the iterative separation of time scales approach. (a) *In silico* experimental data, at every 20th time step, with $(P_m, P_p) = (0.5, 5 \times 10^{-3})$ (blue crosses). The vertical line represents $T = 3$ h. (b) Observing the minimum value of E, (Equation (3.4)), measuring the difference between the *in silico* experimental data and averaged simulation data for $t < 3$ h suggests that P_m is approximately 0.48. The averaged simulation data corresponds to 51 equally-spaced values of P_m in the interval $P_m \in [0, 1]$, and $P_p = 0$. (c) E, (Equation (3.4)), measuring the difference between the *in silico* experimental data and averaged simulation data for $3 < t < 24$ h. The averaged simulation data corresponds to 51 equally-spaced values of P_p in the interval $P_p \in [0, 0.01]$, and $P_m = 0.48$, giving $P_p \approx 5.6 \times 10^{-3}$. Simulation data was generated with $M = 10, Y_0 = 750$ μm, $\Delta = 25$ μm and $\tau = 0.09191$ h, with a final time of 24 h. P_m and P_p required two iterations to converge. All averaged simulation data are insensitive to τ.

and show that with $T = 3$ h and $P_m = 0.48$, we observe a well-defined minimum in E indicating that we have $P_p \approx 5.6 \times 10^{-3}$. We note that it took two iterations for P_m and P_p to converge. These parameter estimates are a great improvement on the results in Section 3.3.1 where we found it was impossible to distinguish between many different parameter combinations. We note that our parameter estimates do not precisely coincide with the expected values of $P_m = 0.50$ and $P_p = 5 \times 10^{-3}$, and there are two potential explanations for this. First, our *in silico* experimental data corresponds to one realisation of the discrete model which might not be representative of the expected behaviour we would observe when considering many identically prepared realisations. Secondly, when we generated the averaged simulation data, we only used a modest number of simulations, $M = 10$, and we expect that our estimate could be improved by increasing M. To further illustrate the robustness of our approach we also applied it to data generated using several different parameter combinations, including some for larger proliferation rates, and we found that this method also gave reliable parameter estimates for these additional cases (Supplementary material).

Once we have obtained estimates of P_m and P_p it is possible to re-examine the suitability of our choice of T. Our estimate of P_p indicates that the average time taken for an isolated agent to undergo a proliferation event is approximately 18 hours, whereas our estimate of P_m indicates that the average time taken for an isolated agent to undergo a motility event is approximately 30 minutes. These time scales give a physical explanation for why our choice of $T = 3$ h is sufficient, since agents have plenty of opportunity to undergo motility events during the first 3 hours of the simulation whereas there is hardly any opportunity for proliferation to occur during this interval. To further demonstrate the robustness of our results, we repeated the process of estimating P_p and P_m using the data in Figure 3.3(a) and found that we obtained excellent estimates of the parameters regardless of whether we chose $T = 2, 3$ or 4 hours.

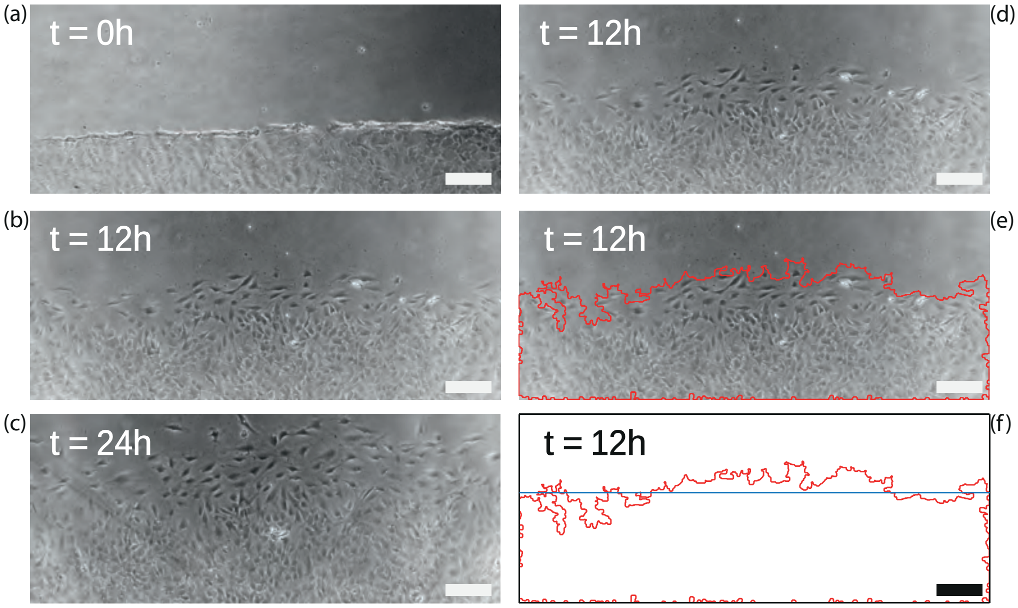

Figure 3.4: Time evolution of a scratch assay with 3T3 fibroblasts cells. Experimental images are shown at:(a) 0, (b) 12 and (c) 24 h. To illustrate the application of the edge detection algorithm we show in (d) the image at 12 h and in (e) we superimpose the detected leading edge (red). (f) Shows the detected leading edge (red) and the average position of the front (blue). Scale bar corresponds to 200 μm.

3.3.3 *In vitro* data

We obtained *in vitro* experimental data for a scratch assay using 3T3 fibroblast cells as described in Section 3.2.1. At each time point, we took four different images, at different spatial locations, in the scratch assay. The field of view in each image is approximately 2 mm wide and 0.8 mm high. The spatial location of the four sets of images are approximately evenly-spaced along the edge of the scratch within the red rectangle in Figure 3.1(a). One set of such images, at $t = 0, 12$ and 24 h, is presented in Figure 3.4(a)-(c). An example of the results from the edge detection algorithm, applied to the image at $t = 12$ h, is illustrated in Figure 3.4(d)-(f). Results summarising the average position of the leading edge as a function of time are given in Figure 3.5(a), and the original data set from the four sets of images at all time points is given in the Supplementary material.

We first apply the naïve parameter recovery method, described in Section 3.3.1, where we consider the difference between our *in vitro* experimental data and averaged simulation data using the entire time series of leading edge data. The averaged simulation data was generated using 2601 equally-spaced parameter combinations within the region $P_m \in [0, 1]$, $P_p \in [0, 0.01]$. Results in Figure 3.5(b) show a contour plot of E, defined by Equation (3.4), which confirms that there is a large region within the parameter space for which the short time leading edge data is indistinguishable. To confirm that multiple parameter combinations match the *in vitro* experimental data, we consider three distinct parameter pairs, highlighted in Figure 3.5(b), and superimpose the corresponding averaged simulation data in Figure 3.5(a).

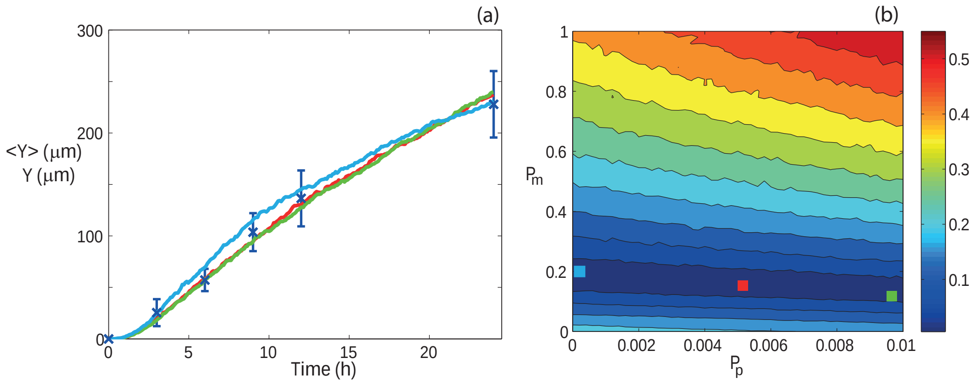

Figure 3.5: Comparison of *in vitro* experimental data and averaged simulation data. (a) Leading edge *in vitro* data (blue crosses) is presented with the error bars indicating one standard deviation from the mean. (b) Contour plot of E (Equation (3.4)) measuring the difference between the *in vitro* data and the averaged simulation data in the region $P_m \in [0, 1], P_p \in [0, 0.01]$. Simulation parameters are $M = 10, Y_0 = 750$ μm, $\Delta = 25$ μm and $\tau = 0.09191$ h, with a final time of 24 h. The contour plot of E was generated by considering 2601 different parameter combinations; 51 equally-spaced values of P_m and 51 equally-spaced values of P_p. The red, green and light blue coloured squares in (b) correspond to three different parameter combinations: $(P_m, P_p) = (0.16, 5.4 \times 10^{-3})$, $(0.14, 9.6 \times 10^{-3})$ and $(0.2, 2 \times 10^{-4})$, respectively. Averaged simulation data from these three different combinations are superimposed in (a), showing that all three parameter combinations lead to indistinguishable short time leading edge data. All averaged simulation data are insensitive to τ.

We now apply the approach described in Section 3.3.2 to our *in vitro* data choosing $T = 6$ h. Results in Figure 3.6(a) show the averaged experimental data. A plot of E, given in Figure 3.6(b), constructed using 51 equally-spaced values of P_m in the interval $P_m \in [0, 1]$, with $P_p = 0$, indicates that the optimal value of P_m is approximately 0.18. A plot of E, given in Figure 3.6(c), constructed using 51 equally-spaced values of P_p in the interval $P_p \in [0, 0.01]$, with $P_m = 0.18$, indicates that the proliferation parameter lies within the subinterval $P_p \in [0, 5 \times 10^{-3}]$. We now refine our parameter estimates by repeating the process and increasing the number of realisations used to generate the averaged simulation data from $M = 10$ to $M = 50$. Furthermore, we now focus our attention on the subintervals $P_m \in [0, 0.5]$ and $P_p \in [0, 5 \times 10^{-3}]$, highlighted by the rectangles superimposed on Figure 3.6(b)-(c). By repeating the parameter estimation process, we obtained the refined results shown in Figure 3.6(e)-(f), indicating that the optimal parameter pair is $(P_m, P_p) = (0.17, 2.7 \times 10^{-3})$, or $(D, \lambda) \approx (300\ \mu\text{m}^2/\text{h}, 0.03\ /\text{h})$. To quantify the uncertainty in our estimates, we repeated the same process using the mean experimental data plus or minus one sample standard deviation of the experimental data. This gave $P_m = 0.17$ $(0.14 - 0.20)$ and $P_p = 2.7 \times 10^{-3}$ $(1.6 \times 10^{-3} - 3.5 \times 10^{-3})$, where the ranges in the parenthesis indicate an estimate of the uncertainty. Our estimates of P_m and P_p were obtained using just one iteration.

To explore whether our results are sensitive to our choice of T, we repeated the process using $T = 9$ h and found that this also gave $(P_m, P_p) = (0.17, 2.7 \times 10^{-3})$, or $(D, \lambda) \approx (300\ \mu\text{m}^2/\text{h}, 0.03\ /\text{h})$, indicating that our results are relatively insensitive to T. The reason for this insensitivity can be explained by considering the time scales implied by our parameter estimates. Our estimate of λ indicates that the average time required for

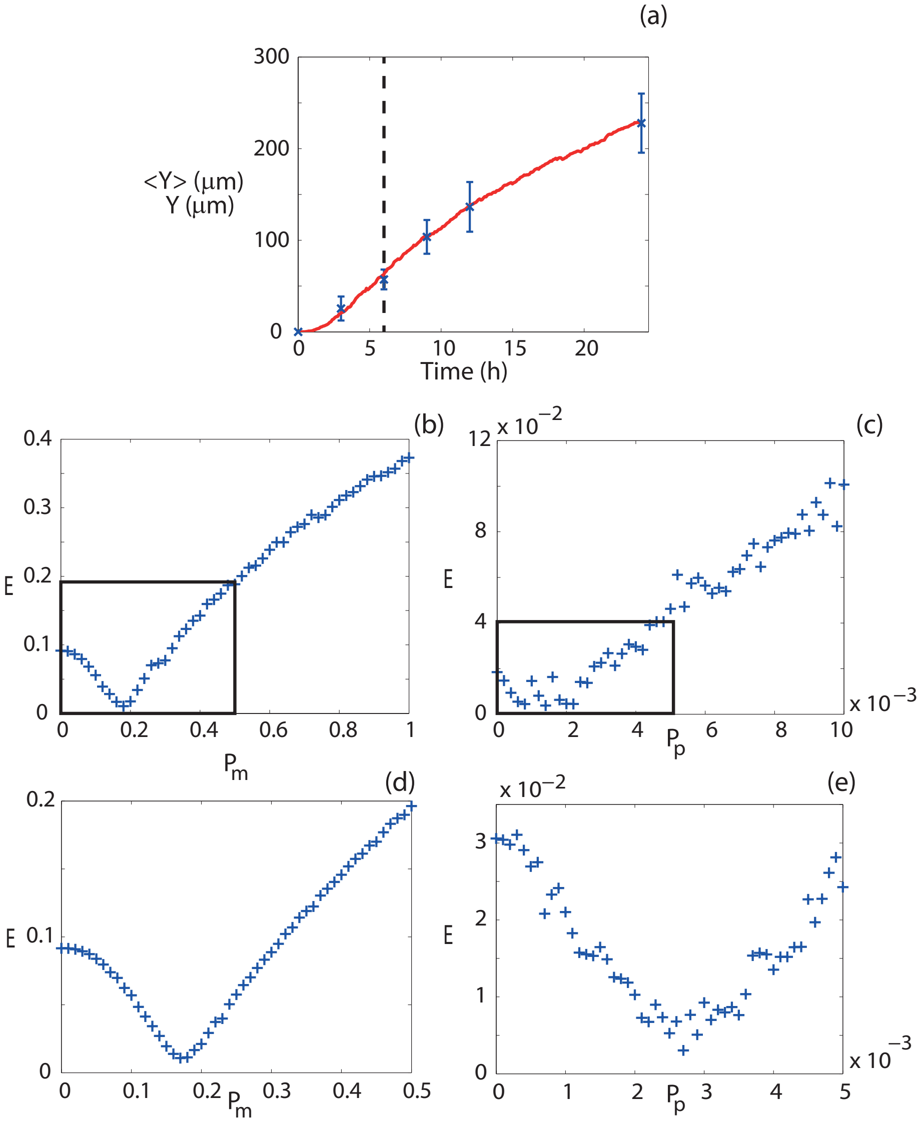

Figure 3.6: Parameter recovery for *in vitro* data using the separation of time scales approach. (a) *in vitro* data (blue crosses) showing the average position of the leading edge with the error bars denoting one standard deviation from the mean (See Supplementary material for the original data set). The vertical line indicates $T = 6$ h. (b) Plot of E (Equation (3.4)) measuring the difference between the *in vitro* experimental data and averaged simulation data for $t < T$ with $P_p = 0$. The plot of E was generated by considering 51 equally-spaced values of P_m within the interval $P_m \in [0, 1]$, and indicates that P_m is approximately 0.18. (c) Plot of E measuring the difference between the *in vitro* experimental data and the averaged simulation data for $t > T$ with $P_m = 0.18$. The plot of E, generated using 51 equally-spaced values of P_p within the interval $P_p \in [0, 0.01]$, indicates that the true value of P_p lies in the subinterval $P_p \in [0, 5 \times 10^{-3}]$. (d) Plot of E measuring the difference between the *in vitro* experimental data and the averaged simulation data for $t < T$ with $P_p = 0$. The plot of E, generated using 51 equally-spaced values of P_m within the subinterval $P_m \in [0, 0.5]$, indicates that an improved estimate is $P_m \approx 0.17$. (e) Plot of E measuring the difference between the *in vitro* experimental data and the averaged simulation data for $t > T$ with $P_m = 0.17$. The plot of E, generated using 51 equally-spaced values of P_p within the subinterval $P_p \in [0, 5 \times 10^{-3}]$, indicates that the true value of P_p is approximately 2.7×10^{-3}. Averaged simulation data showing the position of the leading edge with $P_m = 0.17$ and $P_p = 2.7 \times 10^{-3}$ are superimposed on the *in vitro* experimental data in (a). All simulation data corresponds to $Y_0 = 750$ μm, $\Delta = 25$ μm and $\tau = 0.09191$ h. Results in (b)-(c) correspond to $M = 10$, while results in (d)-(e) correspond to $M = 50$. Our estimates of P_m and P_p required one iteration to converge. All averaged simulation data are insensitive to τ.

an isolated cell to proliferate is approximately 34 hours. In comparison, our estimate of D indicates that the average time taken for an isolated cell to undergo a motility event is approximately 30 minutes. This indicates that that either $T = 6$ or 9 h are appropriate since either of these time scales are sufficiently small relative to the proliferation time scale as well as being sufficiently large compared to the motility time scale. Averaged simulation data produced using $(P_m, P_p) = (0.17, 2.7 \times 10^{-3})$ are superimposed in Figure 3.6(a), confirming that the simulated leading edge data matches the measurements. We note that our parameter estimates are consistent with Tremel's [35] previously-reported estimates. However, we would also like to point out that our estimates of D and λ were obtained simply and inexpensively, using only short time leading edge data, whereas Tremel's results were obtained by constructing cell density profiles and tracking individual cells, both of which are time consuming and expensive.

3.4 Discussion and conclusions

Moving cell fronts [19, 21, 31, 38] play a key role in development, disease and tissue repair. The rate at which the cell front moves depends both on the motility and proliferation of individual cells within the population. Mathematical models can be used to interpret scratch assays, with some previous studies focusing exclusively on matching experimental estimates of the front speed with the long time asymptotic wave speed of the travelling wave solution of the Fisher-Kolmogorov equation, $c = \sqrt{4\lambda D}$ [16]. This approach suffers from two limitations. First, travelling wave solutions require a large amount of time to develop whereas most scratch assays are performed for short time intervals. Second, even if large time experimental data is available, this approach determines the product, λD, and not the values of λ and D separately [13, 14]. Other methods for interpreting scratch assays have involved calibrating the numerical solution of a reaction-diffusion equation to observed cell density profiles [20, 27, 35] to provide estimates of λ and D. Unfortunately, this approach is expensive and time consuming since it requires either a direct or indirect method for counting individual cells to construct the cell density profiles.

In this work we describe a different approach for analysing scratch assays relying only on determining short time leading edge data. Our method can be implemented either for new experimental images, or retrospectively, using previously published images. The simplicity of our approach derives from the fact that we do not require any analysis or counting of individual cells. Using a discrete model of cell motility and cell proliferation, we show that care ought to be exercised when analysing short time leading edge data since a straightforward model calibration procedure, whereby we match the entire time history of the position of the leading edge, reveals that there are many parameter combinations for which the short time leading edge data from the model are equivalent. To overcome this we make use of the fact that cell migration takes place on a short time scale compared to cell proliferation, and we introduce a new iterative method where we analyse the leading edge time series data in two steps. First, we analyse the interval $t < T$, setting $P_p = 0$ in the model, to provide an estimate of P_m. Second, we analyse the time interval $t > T$, using our previously-determined estimate of P_m, to provide an estimate of P_p. These two steps can be applied iteratively until our estimates converge to within some tolerance. Our approach relies on estimating some time, T, which is sufficiently large compared to the time scale of cell migration, yet is sufficiently small compared to the time scale of proliferation. We confirm our approach using both *in silico* and *in vitro* data, and we note that our estimates of D and λ for the *in vitro* data are consistent with previously-published values for the same cell line in a similar experiment [35].

As we demonstrate, once the data has been analysed to produce an estimate of D and λ, we can test the sensitivity of our estimates to our choice of T. For our *in silico* data we found that we obtained similar results regardless of whether we chose $T = 2, 3$ or 4 h. Similarly, for our *in vitro* data, where we had less experimental data points from which to choose T, we found that we obtained the same values for D and λ regardless of whether we chose $T = 6$ or 9 h.

Our parameter estimates for the *in vitro* data indicate that care should be taken when interpreting leading edge data with the long time asymptotic wave speed expression for the Fisher-Kolmogorov equation [7, 13, 14]. Our parameter estimates for the *in vitro* data correspond to $(D, \lambda) \approx (300\ \mu\text{m}^2/\text{h}, 0.03/\text{h})$. While it is possible to use these parameters to estimate the speed of the travelling wave solution of the Fisher-Kolmogorov equation [7, 13, 14], this result is valid only in the long time limit, $t \to \infty$. Since our experimental results have been reported over a time interval which is less than the doubling time, we expect that it is inappropriate to use such a result since there has been insufficient time for the travelling wave to form. Indeed, comparing the slope of the data in Figure 3.5(a) with $c = \sqrt{4\lambda D}$, evaluated using our parameter estimates, confirms that these approaches give different estimates of the front speed.

Our approach of combining simulation data with automated leading edge analysis can be extended in several ways. One important point, not considered here, is that certain cells, such as melanoma [34] and glioma cells [10], exhibit significant cell-to-cell adhesion. To incorporate cell-to-cell adhesion we could consider a different discrete model with an additional parameter controlling the adhesion strength [10]. Under these conditions it would be interesting to explore whether the three parameters governing cell migration, cell proliferation and the strength of adhesion could be uniquely determined by short time leading edge data. Alternatively, we could apply our model to scratch assays performed on different substrates [13, 14] to analyse the effect of cell-to-substrate adhesion. Another approach may be to apply our model to narrow wounds, where cell proliferation is negligible or absent [8]. A further application of our model would be to analyse a series of scratch assays where we considered some control assay relative to a set of other assays where a chemical inhibitor or promotor has been applied. Our approach could be used to determine precisely how D and/or λ varies as a function of the concentration of the chemical, and therefore play a role in the design of intervention strategies aimed at manipulating the movement of cell fronts.

Acknowledgements. We thank the CRC for Wound Management Innovation, the Australian Research Council (FT130100148) and Ms Parvathi Haridas.

Bibliography

[1] Aichele, K., Bubel, M., Deubel, G., Pohlemann, T. and Oberringer, M. [2013], 'Bromelain down-regulates myofibroblast differentiation in an in vitro wound healing assay', *Naunyn-Schmiedeberg's Archives of Pharmacology* **386**, 853–763.

[2] Arciero, J. C., Mi, Q., Branca, M. F., Hackam, D. J. and Swigon, D. [2011], 'Continuum model of collective cell migration in wound healing and colony expansion', *Biophysical Journal* **100**(3), 535–543.

[3] Bloomfield, J.-M., Sherratt, J.-A., Painter, K.-J. and Landini, G. [2010], 'Cellular automata and integrodifferential equation models for cell renewal in mosaic tissues', *Journal of the Royal Society Interface* **7**(52), 1525–1535.

[4] Canny, J. [1986], 'A computational approach to edge detection', *IEEE Transactions on Pattern Analysis and Machine Intelligence* (6), 679–698.

[5] Chigurupati, S., Mughal, M. R., Okun, E., Das, S., Kumar, A., McCaffery, M., Seal, S. and Mattson, M. P. [2012], 'Effects of cerium oxide nanoparticles on the growth of keratinocytes, fibroblasts and vascular endothelial cells in cutaneous wound healing', *Biomaterials* **34**, 2194–2201.

[6] Codling, E. A., Plank, M. J. and Benhamou, S. [2008], 'Random walk models in biology', *Journal of the Royal Society Interface* **5**, 813–834.

[7] Doran, M. R., Mills, R. J., Parker, A. J., Landman, K. A. and Cooper-White, J. J. [2009], 'A cell migration device that maintains a defined surface with no cellular damage during wound edge generation', *Lab on a Chip* **9**(16), 2364–2369.

[8] Farooqui, R. and Fenteany, G. [2005], 'Multiple rows of cells behind an epithelial wound edge extend cryptic lamellipodia to collectively drive cell-sheet movement', *Journal of Cell Science* **118**(1), 51–63.

[9] Fisher, R. A. [1937], 'The wave of advance of advantageous genes', *Annals of Eugenics* **7**(4), 355–369.

[10] Khain, E., Katakowski, M., Hopkins, S., Szalad, A., Zheng, X., Jiang, F. and Chopp, M. [2011], 'Collective behavior of brain tumor cells: the role of hypoxia', *Physical Review E* **83**(3), 031920.

[11] Knecht, D. A., LaFleur, R. A., Kahsai, A. W., Argueta, C. E., Beshir, A. B. and Fenteany, G. [2010], 'Cucurbitacin i inhibits cell motility by indirectly interfering with actin dynamics', *PLOS ONE* **5**(11), e14039.

[12] Lee, P. and Wolgemuth, C. W. [2011], 'Crawling cells can close wounds without purse strings or signaling', *PLoS Computational Biology* **7**(3), e1002007.

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

[14] Maini, P. K., McElwain, D.-L. S. and Leavesley, D. I. [2004*b*], 'Travelling waves in a wound healing assay', *Applied Mathematics Letters* **17**(5), 575–580.

[15] Martin, P. [1997], 'Wound healing–aiming for perfect skin regeneration', *Science* **276**(5309), 75–81.

[16] Murray, J. D. [2002], *Mathematical Biology I: An Introduction*, Vol. 2, Springer.

[17] Plank, M. J. and Simpson, M. J. [2012], 'Models of collective cell behaviour with crowding effects: comparing lattice-based and lattice-free approaches', *Journal of the Royal Society Interface* **9**(76), 2983–2996.

[18] Plank, M. J. and Simpson, M. J. [2013], 'Lattice-free models of cell invasion: Discrete simulations and travelling waves', *Bulletin of Mathematical Biology* **75**(11), 2150–2166.

[19] Rausch, V., Liu, L., Apel, A., Rettig, T., Gladkich, J., Labsch, S., Kallifatidis, G., Kaczorowski, A., Groth, A., Gross, W. et al. [2012], 'Autophagy mediates survival of pancreatic tumour-initiating cells in a hypoxic microenvironment', *The Journal of Pathology* **227**(3), 325–335.

[20] Sengers, B. G., Please, C. P. and Oreffo, R. O.-C. [2007], 'Experimental characterization and computational modelling of two-dimensional cell spreading for skeletal regeneration', *Journal of the Royal Society Interface* **4**(17), 1107–1117.

[21] Sharili, A.-S., Allen, S., Smith, K., Price, J. and McGonnell, I. M. [2013], 'Snail2 promotes osteosarcoma cell motility through remodelling of the actin cytoskeleton and regulates tumor development', *Cancer Letters* **333**, 170–179.

[22] Shatkin, A.-J., Reich, E., Franklin, R.-M. and Tatum, E.-L. [1962], 'Effect of mitomycin-c on mammalian cells in culture', *Biochimica et Biophysica Acta* **55**(3), 277–289.

[23] Sherratt, J. A. and Murray, J. D. [1990], 'Models of epidermal wound healing', *Proceedings of the Royal Society of London Series B, Biological Sciences* **241**, 29–36.

[24] Shibata, A., Tanabe, E., Inoue, S., Kitayoshi, M., Okimoto, S., Hirane, M., Araki, M., Fukushima, N. and Tsujiuchi, T. [2013], 'Hydrogen peroxide stimulates cell motile activity through lpa receptor-3 in liver epithelial wb-f344 cells', *Biochemical and Biophysical Research Communications* **433**, 317–321.

[25] Simpson, M. J. and Baker, R. E. [2011], 'Corrected mean-field models for spatially dependent advection-diffusion-reaction phenomena', *Physical Review E* **83**, 051922.

[26] Simpson, M. J., Landman, K. A. and Hughes, B. D. [2010], 'Cell invasion with proliferation mechanisms motivated by time-lapse data', *Physica A: Statistical Mechanics and its Applications* **389**(18), 3779–3790.

[27] Simpson, M. J., Treloar, K. K., Binder, B. J., Haridas, P., Manton, K. J., Leavesley, D. I., McElwain, D.-L. S. and Baker, R. E. [2013], 'Quantifying the roles of cell motility and cell proliferation in a circular barrier assay', *Journal of the Royal Society Interface* **10**(82), 20130007.

[28] Stetler-Stevenson, W. G., Aznavoorian, S. and Liotta, L. A. [1993], 'Tumor cell interactions with the extracellular matrix during invasion and metastasis', *Annual Review of Cell Biology* **9**(1), 541–573.

[29] Swanson, K. R. [2008], 'Quantifying glioma cell growth and invasion in vitro', *Mathematical and Computer Modelling* **47**(5), 638–648.

[30] Swanson, K. R., Bridge, C., Murray, J.-D. and Alvord Jr, E. C. [2003], 'Virtual and real brain tumors: using mathematical modeling to quantify glioma growth and invasion', *Journal of the Neurological Sciences* **216**(1), 1–10.

[31] Teppo, S., Sundquist, E., Vered, M., Holappa, H., Parkkisenniemi, J., Rinaldi, T., Lehenkari, P., Grenman, R., Dayan, D., Risteli, J. et al. [2012], 'The hypoxic tumor microenvironment regulates invasion of aggressive oral carcinoma cells', *Experimental Cell Research* **319**, 376–389.

[32] Todaro, G. J. and Green, H. [1963], 'Quantitative studies of the growth of mouse embryo cells in culture and their development into established lines', *The Journal of Cell Biology* **17**(2), 299–313.

[33] Treloar, K. K. and Simpson, M. J. [2013], 'Sensitivity of edge detection methods for quantifying cell migration assays', *PLOS ONE* **8**(6), e67389.

[34] Treloar, K. K., Simpson, M. J., Haridas, P., Manton, K. J., Leavesley, D. I., McElwain, D.-L. S. and Baker, R. E. [2013], 'Multiple types of data are required to identify the mechanisms influencing the spatial expansion of melanoma cell colonies', *BMC Systems Biology* **7**(1), 137.

[35] Tremel, A., Cai, A., Tirtaatmadja, N., Hughes, B.-D., Stevens, G.-W., Landman, K.-A. and O'Connor, A.-J. [2009], 'Cell migration and proliferation during monolayer formation and wound healing', *Chemical Engineering Science* **64**(2), 247–253.

[36] Weissberg, J. B., Son, Y. H., Papac, R. J., Sasaki, C., Fischer, D. B., Lawrence, R., Rockwell, S., Sartorelli, A. C. and Fischer, J. J. [1989], 'Randomized clinical trial of mitomycin-c as an adjunct to radiotherapy in head and neck cancer', *International Journal of Radiation Oncology* **17**(1), 3–9.

[37] Zaritsky, A., Manor, N., Wolf, L., Ben-Jacob, E. and Tsarfaty, I. [2013], 'Benchmark for multi-cellular segmentation of bright field microscopy images', *BMC Bioinformatics* **14**(1), 319.

[38] Zhao, X., Li, D.-C., Zhu, X.-G., Gan, W.-J., Li, Z., Xiong, F., Zhang, Z.-X., Zhang, G.-B., Zhang, X.-G. and Zhao, H. [2013], 'B7-h3 overexpression in pancreatic cancer promotes tumor progression', *International Journal of Molecular Medicine* **31**(2), 283–291.

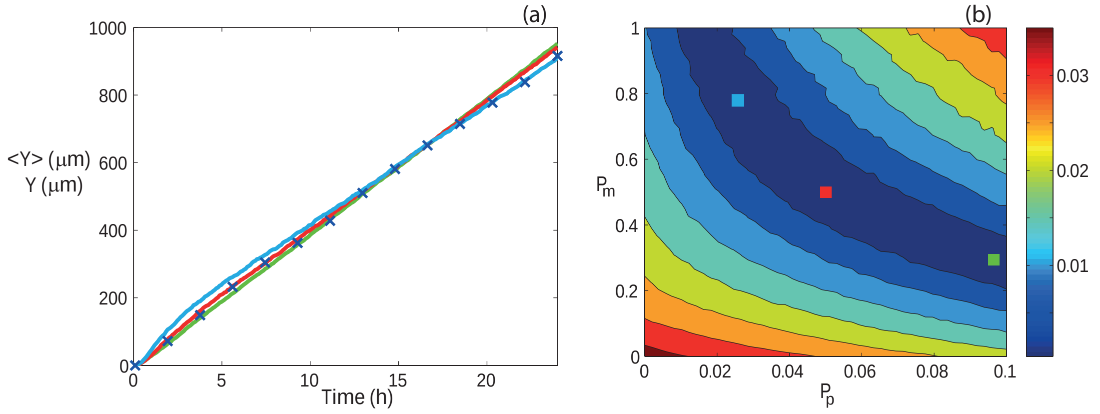

Figure 3.7: Comparison of *in silico* experimental data and averaged simulation data. (a) Leading edge *in silico* experimental data (blue crosses) corresponds to $(P_m, P_p) = (0.5, 0.02)$ (blue crosses). Data is presented every 20th time step. (b) Contour plot of E (Equation (4), main paper) measuring the difference between the *in silico* data and averaged simulation data within the region $P_m \in [0,1], P_p \in [0, 0.1]$. Simulation parameters are $M = 10, Y_0 = 750$ μm, $\Delta = 25$ μm and $\tau = 0.09191$ h, with a final time of 24 h. The contour plot of E was generated by considering 2601 different parameter combinations; 51 equally-spaced values of P_m and 51 equally-spaced values of P_p. The red, green and light blue coloured squares in (b) correspond to three different parameter combinations: $(P_m, P_p) = (0.5, 0.05), (0.32, 0.096)$ and $(0.78, 0.028)$, respectively. Averaged simulation data from these three different parameter combinations are superimposed in (a), showing that all three parameter combinations lead to indistinguishable short time leading edge data. All averaged simulation data are insensitive to τ.

3.5 Supplementary material

3.5.1 Reaction-diffusion model

The continuum-limit description of the discrete model described in Section 3.2.2 is a two-dimensional generalization of the Fisher-Kolmogorov equation [13],

$$\frac{\partial u}{\partial t} = D\nabla^2 u + \lambda u(1-u),$$

where $0 \leq u \leq 1$ is a scaled density, D is the diffusivity and λ is the proliferation rate. This model, applied in one-dimensional geometry, where $\nabla^2 u = \partial^2 u/\partial x^2$, supports travelling wave solutions in the long time limit, $t \to \infty$ [14]. For initial conditions with compact support the long time travelling wave speed is $c = \sqrt{4D\lambda}$ [14].

3.5.2 Naïve parameter recovery with high proliferation rate

We now consider results that are equivalent to those presented in Section 3.3.1 (main paper) except that we consider a higher proliferation rate with parameters $(P_m, P_p) = (0.8, 0.02)$. The experimental data is given in Figure 3.7(a) and the corresponding contour plot of E, generated using Equation (4) (main paper), is given in Figure 3.7(b) indicates that there is no single well-defined minimum on this surface. Instead, we observer a large dark blue region, extending right across the parameter space, from which any combination of D and λ produce indistinguishable short time leading edge data. We demonstrate this

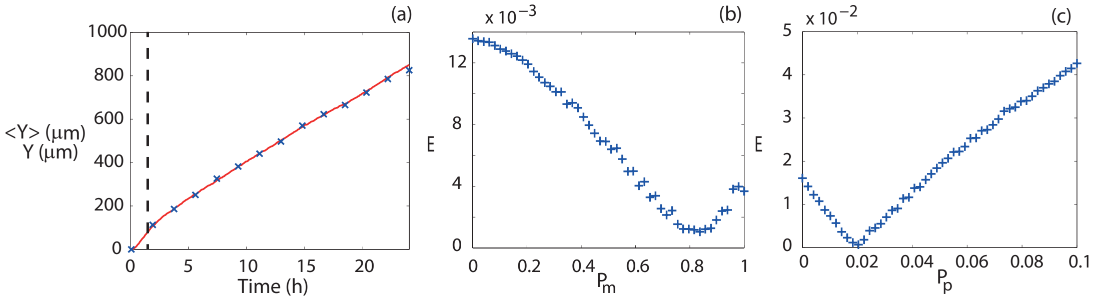

Figure 3.8: Recovery of parameter values from *in silico* experimental data using a separation of timescales. (a) *In silico* experimental data (blue crosses) with $(P_m, P_p) = (0.8, 0.02)$. The vertical line denotes $T = 1.5$ h. (b) Shows a plot of E (Equation 4, main paper) measuring the difference between the *in silico* experimental data and averaged simulation data for $t < T$ with $P_p = 0$. The plot of E was generated by considering 51 equally-spaced values of P_m within the interval $P_m \in [0, 1]$, and indicates that P_m is approximately 0.84. (c) Plot of E measuring the difference between the *in silico* experimental data and averaged simulation data for $t > T$ with $P_m = 0.84$. The plot of E, generated using 51 equally-spaced values of P_p within the interval $P_p \in [0, 0.1]$, indicates that P_p is approximately 0.022. All simulations were performed with $M = 10, Y_0 = 750$ μm, $\Delta = 25$ μm and $\tau = 0.09191$ h, with a final time of 24h. P_m and P_p took three iterations to converge. All simulation results are independent of the duration of the time step.

by choosing three different parameter combinations, highlighted in Figure 3.7(b), and give the corresponding averaged simulation data in Figure 3.7(a).

3.5.3 Parameter recovery accounting for the separation of time scales with high proliferation rate

We now apply the technique described in Section 3.3.2 (main paper) to the *in silico* experimental data from Figure 3.7(a) with $T = 1.5$ h, as shown in Figure 3.8(a). We consider the leading edge data from the first 1.5 h of the experiment and scan the parameter space in the interval $P_m \in [0, 1]$. Results in Figure 3.8(b) indicate that we have $P_m \approx 0.84$. We then consider the leading edge data in the interval $1.5 < t < 24$ h and scan the parameter space in the interval $P_p \in [0, 0.1]$. Results in Figure 3.8c indicate that we have $P_p \approx 0.022$, which provides an excellent match to the real parameters. Our procedure to estimate P_m and P_p required three iterations to converge.

Experimental data

The experimental data set describing $n = 4$ images from the scratch assay, performed according to the methods outlined in Section 3.2.1, are presented in Table 3.1.

Table 3.1: Data describing the mean, $\langle Y \rangle$, and standard deviation, σ, of the leading edge position from four images of the scratch assay. Data corresponds to $t = 0, 3, 6, 9, 12$ and 24 h, as indicated.

	$t = 0$ h	$t = 3$ h	$t = 6$ h	$t = 9$ h	$t = 12$ h	$t = 24$ h
Y_1 (μm)	0	39.99	46.38	109.53	161.82	183.84
Y_2 (μm)	0	13.87	52.46	82.18	130.29	230.53
Y_3 (μm)	0	14.85	71.74	97.34	125.65	261.14
Y_4 (μm)	0	33.04	58.27	125.60	155.06	235.92
$\langle Y \rangle$ (μm)	0	25.44	57.21	103.66	136.45	227.86
σ (μm)	0	13.11	31.03	18.42	27.12	32.34

Chapter 4

Interpreting scratch assays using pair density dynamics and approximate Bayesian computation

A paper published in *Open Biology.*

Johnston, S. T., Simpson, M. J., McElwain, D. L. S., Binder, B. J. and Ross, J. V. [2014], 'Interpreting scratch assays using pair density dynamics and approximate Bayesian computation', *Open Biology* **4**(9), 140097.

Abstract

Quantifying the impact of biochemical compounds on collective cell spreading is an essential element of drug design, with various applications including developing treatments for chronic wounds and cancer. Scratch assays are a technically simple and inexpensive method used to study collective cell spreading; however, most previous interpretations of scratch assays are qualitative and do not provide estimates of the cell diffusivity, D, or the cell proliferation rate, λ. Estimating D and λ is important for investigating the efficacy of a potential treatment and provides insight into the mechanism through which the potential treatment acts. While a few methods for estimating D and λ have been proposed, these previous methods lead to point estimates of D and λ and provide no insight into the uncertainty in these estimates. Here we compare various types of information that can be extracted from images of a scratch assay and quantify D and λ using discrete computational simulations and approximate Bayesian computation. We show that it is possible to robustly recover estimates of D and λ from synthetic data as well as a new set of experimental data. For the first time our approach also provides a method to estimate the uncertainty in our estimates of D and λ. We anticipate that our approach can be generalised to deal with more realistic experimental scenarios in which we are interested in estimating D and λ, as well as additional relevant parameters such the strength of cell–to–cell adhesion or the strength of cell–to–substrate adhesion.

4.1 Introduction

Scratch assays, otherwise known as scrape or wound healing assays [18, 27], are a common experimental method used to study collective cell spreading. Cells are grown to confluence on a culture plate, after which an artificial gap is created in the monolayer with a fine-tipped instrument [18]. Microscopic images of the cell front moving into the vacated area are captured over approximately 12-24 hours [6, 21, 26, 30]. Scratch assays are often used to evaluate the impact of biochemical compounds on cell migration and proliferation [10, 12, 16, 46]. For example, scratch assays have been used to study wound healing treatments [16, 17], compounds that promote metastasis [10] and chemotherapeutic drugs [12]. Unfortunately, the majority of these evaluations are qualitative [17, 26] or focus on measurements that do not distinguish between the roles of cell diffusivity and cell proliferation [10, 12, 16, 19, 29]. Quantitative comparisons between control assays and assays where a treatment has been applied are critical to providing information about the efficacy of a treatment. There is, therefore, considerable interest in the development of robust approaches that recover estimates of the cell diffusivity D, and cell proliferation rate λ, as these parameters provide important information about the effectiveness, and the mechanism of action, of a putative treatment.

Previous quantitative interpretations of scratch assays have considered a variety of experimental measurements, including counting cell numbers to construct detailed cell density profiles [5, 28, 33, 45], estimating the position of the leading edge of the spreading population [9, 13, 19] and recording detailed individual cell trajectories [5, 45]. In some cases these measurements have been compared with the results of a mathematical model to produce point estimates of D and λ [36, 37]. Presently, it is unclear whether some of these experimental measurements lead to improved estimates of D and λ relative to other experimental measurements, and it remains unclear whether an optimal experimental measurement from a scratch assay can be identified. To the best of our knowledge, pair density information and pair correlation functions [2, 43] have not been previously considered as a means of estimating D and λ from a scratch assay. Unlike previous quantitative interpretations, the data required to calculate the pair correlation function from a scratch assay is straightforward to obtain since it can be calculated simply by inspecting images of the assay at several time points without detailed cell labelling techniques or real time tracking of individual cells. Calculating the pair correlation function from experimental images incorporates information about both the counts of pair distances and the number of cells in the image. This kind of information can also be easily extracted from a discrete, individual-based random walk simulations incorporating random cell movement, governed by the cell diffusivity D, and cell proliferation, governed by the cell proliferation rate λ.

Typically, D and λ are estimated by minimising a measure of the difference between some experimental measure and a prediction of a mathematical model, giving rise to point-estimates of D and λ [13, 19, 37, 45]. However, any information about the uncertainty of the recovered parameters is ignored by this standard approach. Understanding and quantifying the uncertainty in our estimates is important since previously reported estimates of D vary widely [37] and so it is insightful to employ parameter estimation techniques that provide more information than traditional approaches. Approximate Bayesian computation (ABC) generates a parameter distribution that contains this information and hence provides more insight to the recovered parameters [20, 23, 40]. The use of ABC algorithms in spatiotemporal problems is relatively novel and has not been considered in the context of a scratch assay.

Therefore, in this work we focus on a relatively straightforward experimental system by working with 3T3 fibroblast cells which are widely assumed to undergo migration and proliferation without significant cell–to–cell or cell–to–substrate adhesion effects [5, 33, 45]. This simplification allows us to focus on the estimation of two parameters, D and λ. Of course, if the technique described in this work were to be applied to other cell types where other mechanisms, such as cell–to–cell adhesion, cell–to–substrate adhesion or other mechanical effects were present, a more detailed random walk framework with additional parameters would be required. For example, Khain and coworkers [15] describe such an extension whereby individual motility events in the random walk simulation are affected by adhesion, and this is incorporated into the computer simulations through the use of an additional parameter. Other extensions are also possible, such as the incorporation of mechanical forces [1, 35, 38]. While the present work does not incorporate these additional details, we anticipate that the general framework presented here for the simpler random walk simulations with just two parameters could be extended to deal with these further details in future applications.

Here we interpret new experimental images from a scratch assay using discrete random walk simulations, pair correlation functions and ABC. In Section 4.2 we describe the experimental procedure, present a random walk simulation framework that approximates the behaviour of cells in a scratch assay [31] and describe the process of comparing the simulation predictions with experimental data. We note that the random walk model is applied by performing repeated stochastic computational simulations and, henceforth, we refer to our random walk model as a computational simulation. In Section 4.3 we present the results from an ABC algorithm, applied to synthetically generated data, and compare our ability to estimate D and λ using various pieces of information from the images of the synthetic scratch assay. We show that combining estimates of the pair correlation function and the number of cells in the image allows us to robustly estimate D and λ. Applying the same technique to new experimental data we recover estimates of D and λ that are well-defined and consistent with previous point-estimates [33]; however, we also present information about the uncertainty in our parameter estimates which has not been presented previously. In Section 4.4, we discuss our results and suggest directions for future study.

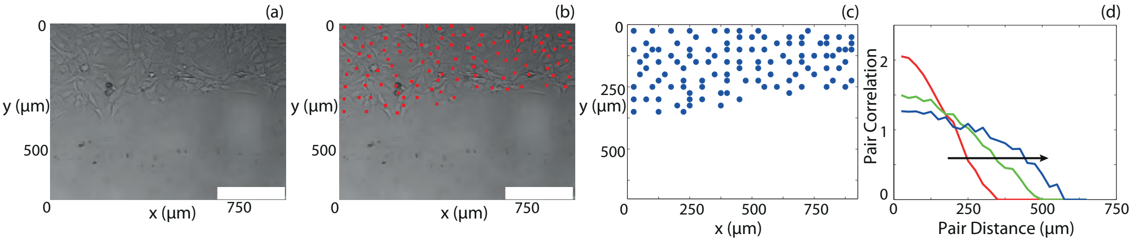

Figure 4.1: (a) Typical experimental image obtained from a scratch assay performed using 3T3 fibroblast cells. (b) Identification of the position of the cells. Scale bar corresponds to 250 μm. (c) Position of cells mapped to a square lattice, where the lattice size is equal to the cell diameter, $\Delta = 25$ μm. (d) Pair correlation function $q(i)$, obtained from experimental images at time $t = 4$ h (red), $t = 8$ h (green), $t = 12$ h (blue). Arrow indicates direction of increasing time. See Section 4.2.3 and Supplementary material for details about the calculation and interpretation of $q(i)$.

4.2 Methods

4.2.1 Experimental method

The details of the experimental method have been presented previously [42]. Briefly, murine fibroblast 3T3 cells [39] were grown in T175 cm^2 tissue culture flasks. One μL of cell suspension was carefully inserted into the well of a tissue culture plate to ensure that cells were approximately evenly distributed. The tissue culture plate was placed in a humidified incubator at 37°C and 5% CO$_2$ until the population became confluent. A scratch was made in the population using a P1000 pipette tip (Lab Advantage, Australia). Images of the spreading population were recorded using a Leica AF6000 automated microscope every five minutes for 24 hours.

4.2.2 Computational simulation

We consider a discrete random walk incorporating motility and proliferation mechanisms on a two-dimensional square lattice with lattice spacing Δ, where each lattice site may be occupied by, at most, one agent [8, 31]. At time t the lattice contains $N(t)$ agents, which have the ability to move and proliferate, with probability $P_m \in [0, 1]$ and $P_p \in [0, 1]$, respectively, during each timestep of fixed duration τ. Invoking the standard assumption that P_m and P_p are constant, the parameters in the computational simulation are related to D and λ by [31]

$$D = \frac{P_m \Delta^2}{4\tau}, \quad \lambda = \frac{P_p}{\tau}. \tag{4.1}$$

Using these relationships we can treat the parameters in the simulation, P_m and P_p, as interchangeable with D and λ, respectively.

During each timestep $N(t)$ agents are selected with replacement, at random, one at a time [7], and given a chance to move [31]. Once selected, an agent at (x, y) randomly chooses, and attempts to move to either $(x \pm \Delta, y)$ or $(x, y \pm \Delta)$. Potential motility events are successful provided that the target site is vacant; otherwise the event is aborted. After $N(t)$ motile events have been attempted, an additional $N(t)$ agents are selected with replacement, at random, one at a time, and are given the opportunity to proliferate. A proliferative agent at (x, y) attempts to place a daughter agent at $(x \pm \Delta, y)$ or $(x, y \pm \Delta)$. Attempted proliferation events can only be successful if the target site is unoccupied; otherwise the event is aborted. We note that our random walk simulation is an idealisation in which it is always theoretically possible that an agent in the simulations could occasionally proliferate twice in quick succession and we note that this is not biologically relevant. However, for parameter values relevant to our, and many other, biological systems, this feature is expected to have minimal impact. To see this, we note that the average time between motility events for an isolated agent is τ/P_m whereas the average time between proliferation events for an isolated agent is τ/P_p. Therefore, for our simulations to be biologically realistic we expect that the quantity τ/P_m to be in the order of 10–20 minutes [15] whereas the quantity τ/P_p to be in the order of approximately 24-48 hours [13]. We will make a comment on these details in Section 4.3.2 when we interpret our results.

We choose the geometry of our simulation to mimic the scratch assay presented in Figures 4.1(a)-(b). The average cell diameter is approximately 25 μm [33], giving $\Delta = 25$ μm. The simulation domain, an X by Y lattice, presented in Figure 4.1(c), corresponds with the size of the experimental images. The image in Figure 4.1(a) is approximately 900 μm wide and 675 μm high, corresponding to $X = 36$ and $Y = 27$. We apply symmetry (zero flux) boundary conditions along the boundaries at $x = 0$, $x = X\Delta$, $y = 0$ and $y = Y\Delta$. To initiate the computational simulation we place $N(0)$ agents, at random, ensuring that no two agents occupy the same site, in the region for $y \leq Y_0\Delta$. We estimate $N(0)$ by counting the number of cells present at $t = 0$ in the experimental images. We note that $N(t)$ depends on time, t, but we refer to this quantity as N from this point for notational convenience.

4.2.3 Pair correlation functions

There is a significant amount of information available in an experimental image of a scratch assay. For example, cell density profiles [33, 45], individual cell trajectories [45] and the position of the leading edge of the spreading cell front [13, 19] have all been estimated from experimental images and used to provide point estimates of D and λ. Here, we consider estimating the pair correlation function [2] as an experimental measurement, henceforth referred to as a summary statistic. Summary statistics are lower-dimensional summaries of data that provide tractable comparisons between sets of data [20]. Since summary statistics merely summarise a data set, it is important to examine whether a particular summary statistic is informative, that is, a statistic that contains sufficient information to distinguish the model output obtained from different parameters.

To calculate the pair correlation function we consider a data set corresponding to a square lattice of dimensions X by Y, where each lattice site can be occupied by, at most, one agent. Each lattice site has an index, (x, y), where $1 \leq x \leq X$, $1 \leq y \leq Y$. All occupied lattice sites at time t, (x_j, y_j), are uniquely indexed by $j = 1, \ldots, N$. The number of occupied lattice pairs for each pair distance $i = 1, \ldots, Y-1$ is then given by

$$c(i) = \sum_{k=1}^{N} \sum_{m=k+1}^{N} 1_{x_k = x_m} 1_{|y_k - y_m| = i}, \; i = 1, \ldots, Y-1, \tag{4.2}$$

where 1_a is the indicator function, which is equal to one if a is true and is equal to zero otherwise. We have oriented our lattice such that the x direction is parallel to the direction of the initial scratch and the y direction is perpendicular to the direction of the initial scratch. Previous analysis [2] indicates that there is more information in the y direction for this kind of scratch assay and so we focus on counting the pairs of agents in the y direction from this point onward.

Binder and Simpson [2] demonstrated that it was possible to normalise Equation (4.2) to produce a pair correlation function which accounted for volume exclusion and crowding effects, and here we use the same approach. Binder and Simpson's [2] normalisation term describes the expected number of pairs of occupied lattice sites, for each pair distance i, in a randomly distributed population without any spatial correlation. The normalisation term is given by

$$\hat{c}(i) = X^2(Y-i)\rho\overline{\rho}, \text{ for } i = 1, 2, \ldots, Y-1, \tag{4.3}$$

where $\rho = N/(XY)$, $\overline{\rho} = (N-1)/(XY-1)$ and N is the number of occupied lattice sites. Therefore, the pair correlation function is given by

$$q(i) = \frac{c(i)}{\hat{c}(i)}. \tag{4.4}$$

We note that $q(i)$ is a nondimensional measure of the abundance of pairs of objects relative to a uniformly distributed population whereas $c(i)$ is a dimensional measure of the number of pairs. Intuitively, we expect that measurements relating to pair density information could provide important information about the rates of cell motility and cell proliferation since proliferation events produce pairs separated by a short distance whereas motility events act to increase the distance between cell pairs. However, without any quantitative comparisons, it is unclear whether there is any advantage in using $q(i)$ or $c(i)$ to recover estimates of D and λ.

To compare our experimental data with the predictions from our computational simulation we map the positions of cells in the experimental images onto the same lattice used in the simulation (Supplementary material). We then calculate the pair correlation function, $q(i)$, for both the experimental images and the images produced by the computational simulation, using the method outlined by Binder and Simpson [2] . Values of $q(i)$ greater than unity indicate that the distribution of cells or agents is such that we are more likely to find a pair of cells or agents separated by a distance i than in a spatially uniform distribution. Similarly, values of $q(i)$ less than unity indicate that the distribution of cells or agents is such that we are less likely to find a pair of cells or agents separated by a distance i than in a spatially uniform distribution. If we find that $q(i)$ is approximately unity for all pair distances, the domain is populated uniformly at random [2]. Calculating $q(i)$ requires information about the counts of pair distances $c(i)$, and the number of cells or agents N. To calculate the pair correlation function we normalise the pair distance counts by the density, which depends on N. Therefore, we expect some information regarding N will be lost when considering the pair correlation function only as a summary statistic. However, since the pair correlation function has been used previously to analyse *in vitro* cell biology assays [2, 43], it is relevant for us to examine whether there is sufficient information in the pair correlation function to robustly recover estimates of D and λ using an ABC framework. For completeness, in Section 4.3.1, we also examine and compare results generated by considering other potential summary statistics to ensure that we use the most appropriate information in our parameter estimation.

The process of mapping cells from a continuous image onto a lattice can involve some discretisation error, which we investigated, in detail, recently [43]. This previous study explored the impact of using different sized lattices to discretise similar experimental images and we computed the pair correlation function for various sized lattices, and this showed that the pair correlation function was insensitive to the size of the lattice provided that the lattice spacing was at least as small as the average cell diameter [43]

4.2.4 Approximate Bayesian computation

ABC is a useful method for computing posterior distributions of unknown model parameters in situations where the likelihood function is intractable [34]. ABC algorithms consider parameter values that generate model predictions that attempt to match observed experimental data [20, 23, 34, 40, 41]. To approximate the posterior distribution, $f(\theta|\beta)$, we consider a prior distribution, $\pi(\theta)$, and a simulation that provides a summary statistic based on a parameter set, θ. We note that β represents the experimental data and define $S(\beta)$ as the corresponding summary statistic. Making no assumptions, we consider a uniform prior, $P_m \in [0, 1]$, $P_p \in [0, 1]$, to generate parameter values and corresponding simulations. Given that the timescale of cell proliferation is typically much larger than the timescale of cell motility, $P_p/P_m \ll 1$ [31], we anticipate that a significant region of the parameter space will not produce realisations that match the experimental data. To reduce the computation time we therefore implement the ABC Markov Chain Monte Carlo algorithm (ABC-MCMC) [20, 34], an ABC algorithm that evolves based on previously successful parameter values (Supplementary material).

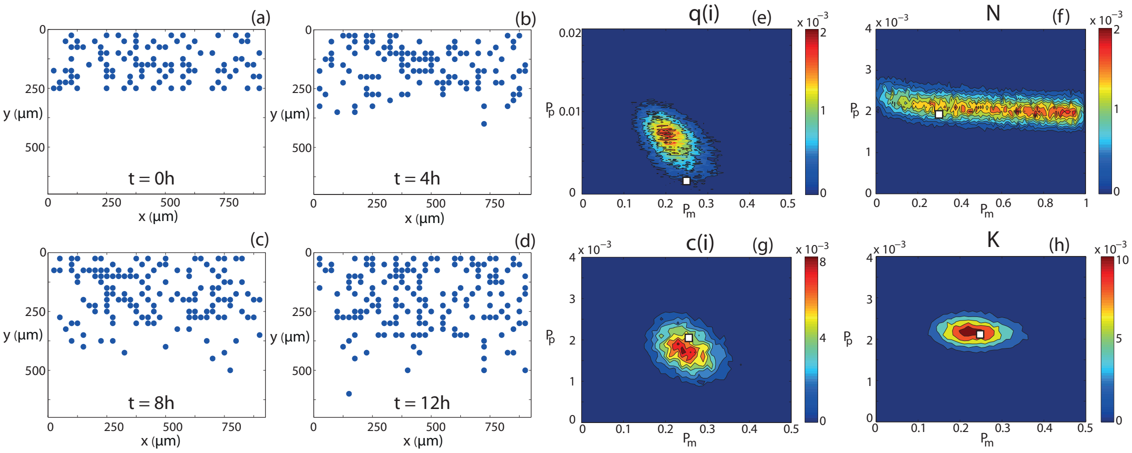

Figure 4.2: (a)-(d): Typical evolution of the discrete simulation described in Section 4.2.2, generated with $(P_m, P_p) = (0.25, 2 \times 10^{-3})$, presented at time (a) $t = 0$ h, (b) $t = 4$ h, (c) $t = 8$ h, (d) $t = 12$ h. (e)-(h): Calculated posterior distribution obtained from a summary statistic generated with $(P_m, P_p) = (0.25, 2\times10^{-3})$ using (e) the pair correlation function, $q(i)$, (f) the number of cells, N, (g) the counts of pair distances, $c(i)$, (h) K, the combination of $q(i)$ and N as the summary statistic. The details of the process to obtain the distributions are given in the Supplementary material. In brief, ϵ is the maximum difference between the summary statistics for θ to be accepted, Γ is the width of the distribution of potential step sizes in the ABC-MCMC algorithm and M is the total number of steps attempted. Parameter values used were (e) $\epsilon = 0.012$, (f) $\epsilon = 0.01$, (g) $\epsilon = 0.06$, (h) $\epsilon = 0.012$. ϵ values were chosen such that the posterior distribution did not significantly change if ϵ was reduced (Supplementary material). For all simulations $\tau = 1/24$ h, $\Gamma = (10^{-1}, 10^{-3})$, $M = 10^6$, $N(0) = 100$, $Y_0 = 10$. Red areas indicate high relative frequency while blue areas indicate low relative frequency. All simulation data are insensitive to τ. The white squares represent the parameter values used to generate the synthetic data.

4.3 Results

4.3.1 Synthetic data

To examine the robustness of our method and the validity of using the pair correlation function as a summary statistic we first attempt to recover parameter values from data that was generated synthetically. We choose a biologically relevant parameter set $(P_m, P_p) = (0.25, 2 \times 10^{-3})$, which with $\Delta = 25$ μm and $\tau = 1/24$ h, corresponds to $(D, \lambda) = (937.5\ \mu\text{m}^2/\text{h}, 4.8 \times 10^{-2}\ /\text{h})$. We perform a single realisation of the simulation with these parameters and calculate the relevant summary statistics at time $t = 4, 8, 12$ h. The average distance between the summary statistic for the synthetic data and the simulation prediction at $t = 4, 8, 12$ h is calculated and compared to either accept or reject potential parameter values to estimate the posterior distribution (Supplementary material). The evolution of the computational simulation is presented in Figures 4.2(a)-(d). We apply the ABC-MCMC algorithm (Supplementary material) and present the resulting posterior distribution for the pair correlation function $q(i)$, in Figure 4.2(e). If the pair correlation function were a close-to-sufficient summary statistic we would observe a well-defined posterior distribution centred at $(P_m, P_p) = (0.25, 2 \times 10^{-3})$, with a narrow spread about the mean in the distributions of P_m and P_p. Instead, we observe that the posterior distribution is centred at $(P_m, P_p) \approx (0.22, 6.7 \times 10^{-3})$ with significant spread.

These observations suggest that additional information ought to be incorporated into our ABC algorithm. We note that identically prepared simulations using the same values of Δ and τ, but different values of P_m and P_p, can occasionally produce similar or indistinguishable summary statistics. This is due to the fact that our random walk computer simulations are stochastic. For this reason we feel it is more appropriate to interpret our experimental results using a probabilistic ABC approach, leading to a distribution of D and λ, rather than using a more traditional approach and arriving at point estimates of D and λ.

We now attempt to refine our estimates of P_m and P_p by examining the posterior distributions obtained by considering the number of cells N, and the pre-normalised counts of the pair distances $c(i)$, as summary statistics in Figures 4.2(f) and 4.2(g), respectively. Intuitively, we expect that N may provide some information about P_p but less information about P_m. Indeed, the posterior distribution in Figure 4.2(f) suggests that all values in the range $P_m \in [0, 1]$ are potentially acceptable and there is little correlation between P_m and P_p. The counts of pair distances correspond to the pre-normalised pair correlation functions. Since information regarding P_m and P_p may be lost in the normalisation which converts $c(i)$ into $q(i)$, we anticipate that $c(i)$ could be a more relevant summary statistic than $q(i)$. We observe in Figure 4.2(g) that $c(i)$ is indeed an excellent summary statistic as the calculated posterior distribution is centred on $(P_m, P_p) \approx (0.25, 2 \times 10^{-3})$.

The final summary statistic we consider is the average of $q(i)$ and N, $K = \{d[q(i)] + d[N]\}/2$, where $d[L]$ is a measure of the difference between two data sets using an arbitrary summary statistic L (Supplementary material). Therefore, the summary statistic, K, includes information about the number of cells or agents which is lost due to the formulation of the pair correlation function, $q(i)$. We present the posterior distribution calculated using K in Figure 4.2(h) and we find that the distribution is similarly centred on $(P_m, P_p) \approx (0.25, 2 \times 10^{-3})$. Due to the explicit inclusion of N in the summary statistic, we observe a reduction in the spread of the posterior distribution in the P_p direction, compared to Figure 4.2(g), while maintaining a similar spread in the P_m direction. While in theory we could continue to incorporate additional information in our summary statistic to obtain a further refined posterior distribution, there is an important computational trade off between the more complicated summary statistic and the improvement in the posterior distribution [3].

Since the $c(i)$ and K summary statistics both lead to reasonable posterior distributions we now compare them by repeating the ABC-MCMC algorithm on ten sets of identically prepared simulation data, that is, simulation data generated using the same parameter values, initial and boundary conditions, and investigate the average of the ten resulting posterior distributions, presented in Figure 4.3. We observe that the distribution in Figure 4.3(a) using K as the summary statistic has a significantly smaller spread than in Figure 4.3(b) which used $c(i)$ as the summary statistic. To quantitatively compare the posterior distributions we calculate the Kullback-Leibler divergence [4], which is a measure of the

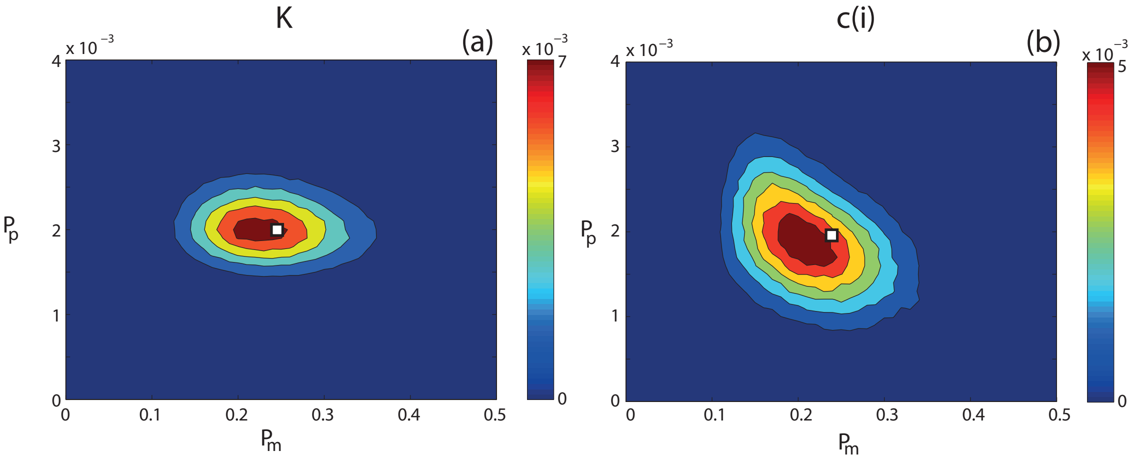

Figure 4.3: Averaged posterior distributions calculated using summary statistics obtained simulations as prepared in Figure 2 for (a) K, a combination of the pair correlation function $q(i)$, and the number of cells N, (b) the counts of pair distances $c(i)$, as a summary statistic. The details of the process to obtain the distributions are given in the Supplementary material. Parameter values used were (a) $\epsilon = 0.012$, (b) $\epsilon = 0.06$. For all simulations $\tau = 1/24$ h, $\Gamma = (10^{-1}, 10^{-3})$, $M = 10^6$, $N(0) = 100$, $Y_0 = 10$. The white squares represent the parameter values used to generate the synthetic data. For further details about the parameters, see Figure 2.

information gained from moving from the prior to the posterior distribution, and is defined as

$$D_{KL}(f|\pi) = \sum_j f(\theta_j|\beta) \ln\left(\frac{f(\theta_j|\beta)}{\pi(\theta_j)}\right), \tag{4.5}$$

where the index j accounts for all possible discretised parameter pairs. A larger $D_{KL}(f|\pi)$ value implies that more information is obtained when moving from the prior to the posterior distribution [4]. We discretise our posterior distribution onto a lattice with 10^2 equally-spaced values of P_m in the interval $P_m \in [0,1]$ and 10^4 equally-spaced values of P_p in the interval $P_p \in [0,1]$ and count the number of successful observations for each parameter combination, and use this information to calculate $D_{KL}(f|\pi)$. We find that $D_{KL}(f|\pi)$ is higher for the posterior distribution calculated using K rather than $c(i)$, $D_{KL}(f|\pi) = 7.98$ and $D_{KL}(f|\pi) = 7.61$, respectively. For perspective, the $D_{KL}(f|\pi)$ values for $q(i)$ and N, individually, were 6.32 and 6.93, respectively. Therefore, a difference in $D_{KL}(f|\pi)$ of approximately 0.3 is relevant. Guided by this information we will interpret our experimental data using K as the summary statistic.

4.3.2 Experimental data

We obtained experimental data from a scratch assay, calculated $q(i)$ and $c(i)$, and counted N, at $t = 4, 8, 12$ h. The position of the cells in the experiments was mapped to a square lattice with $\Delta = 25$ μm (Supplementary material). Figures 4.4(a)-(i) illustrate the process of mapping the cell positions to the lattice. We applied the ABC-MCMC algorithm (Supplementary material) to the experimental data using K as a summary statistic and the average distance between the summary statistic for the experimental data and the simulation prediction at $t = 4, 8, 12$ h to either reject or accept potential parameter

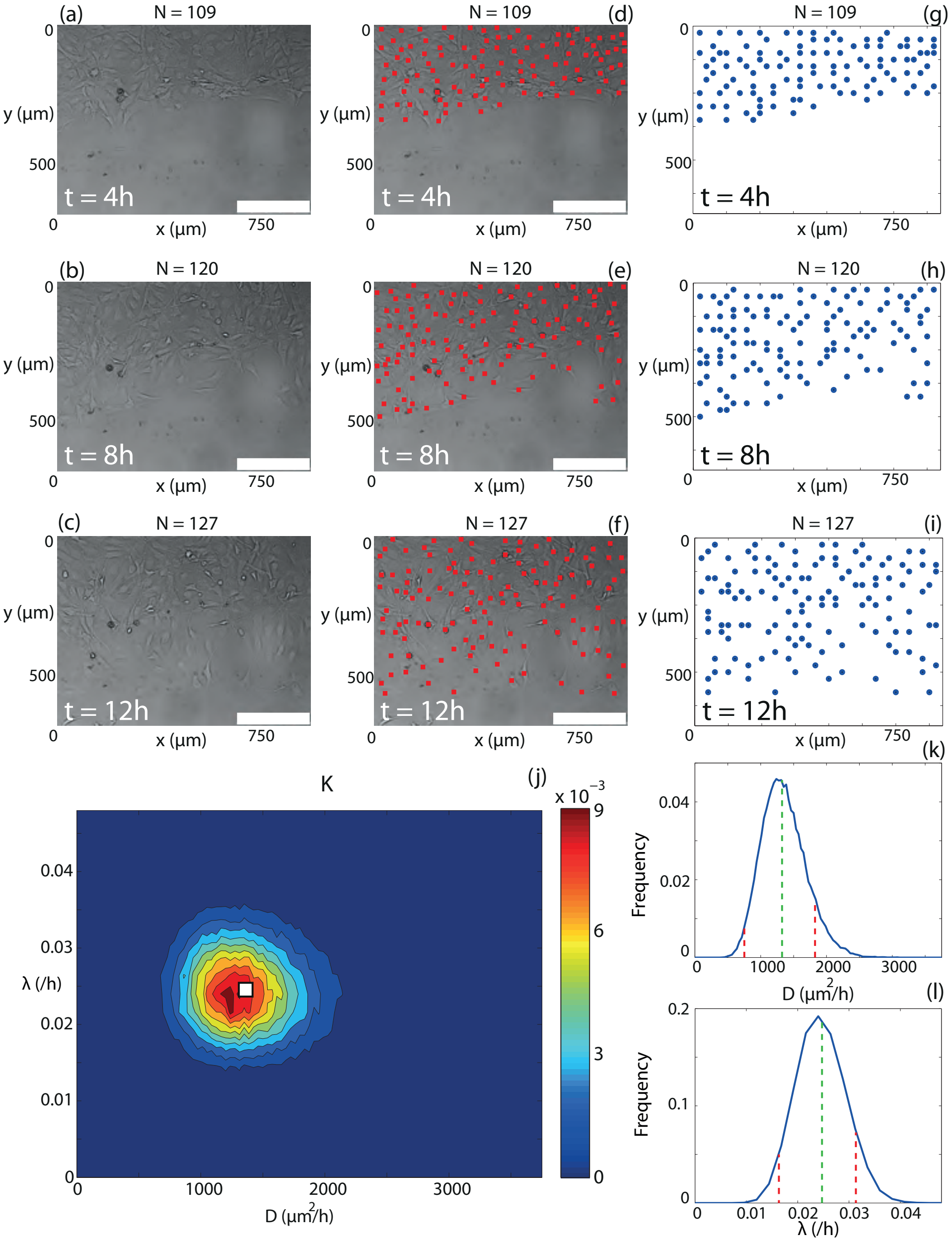

Figure 4.4: Typical experimental images of a scratch assay at (a) $t = 4$ h, (b) $t = 8$ h, (c) $t = 12$ h. Experimental images of a scratch assay with the position of cells indicated (red squares) at (d) $t = 4$ h, (e) $t = 8$ h, (f) $t = 12$ h. The position of cells when mapped onto a lattice at (g) $t = 4$ h, (h) $t = 8$ h, (i) $t = 12$ h. (j) Posterior distribution calculated using K, which is the average of $q(i)$ and N, as a summary statistic and $\epsilon = 0.015$. The details of the process to obtain the distribution are given in the Supplementary material. In brief, ϵ is the maximum difference between the summary statistics for θ to be accepted, Γ is the width of the distribution of potential step sizes in the ABC-MCMC algorithm and M is the total number of steps attempted. For all simulations $\tau = 1/24$ h, $\Gamma = (10^{-1}, 10^{-3})$, $M = 10^6$, $N(0) = 102$, $Y_0 = 10$. Red areas indicate high relative frequency while blue areas indicate low relative frequency. ϵ was chosen such that the posterior distribution did not significantly change if ϵ was reduced (Supplementary material). The white square represents the mean parameter values, $(D, \lambda) \approx (1350\ \mu\text{m}^2/\text{h}, 2.5 \times 10^{-2}\ /\text{h})$. All simulation data are insensitive to τ. (k) Posterior distribution of λ obtained by averaging over D with the mean, $\lambda = 2.5 \times 10^{-2}$ /h, (dashed green) and 90% credible interval, $(1.7 \times 10^{-2} - 3.1 \times 10^{-2})$ /h (dashed red) superimposed. (l) Posterior distribution of D obtained by averaging over λ with the mean, $D = 1350\ \mu\text{m}^2/\text{h}$, (dashed green) and 90% credible interval, $(675 - 1800)\ \mu\text{m}^2/\text{h}$ (dashed red), superimposed.

values in the estimation of the posterior distribution, given in Figure 4.4(j). We observe that the resulting bivariate posterior distribution is well-defined and contains a relatively narrow spread in both the D and λ directions. To provide further insight we estimate the corresponding univariate distributions of D and λ, presented in Figures 4.4(k)-(l), by averaging the posterior distribution in each of the λ and D directions, respectively. Since the univariate posterior distributions do not appear to be significantly skewed, we choose to report the mean of the univariate posterior distributions as our estimate of D and λ. To provide quantitative insight into the uncertainty in our estimates we calculate the 90% credible interval by finding the interval, symmetric about the mode, containing approximately 90% of the total area under the univariate distribution. The mean and 90% credible intervals are $D \approx 1350$ $(675 - 1800)$ μm^2/h and $\lambda \approx 2.5 \times 10^{-2}$ $(1.7 \times 10^{-2} - 3.1 \times 10^{-2})$ /h. We note that our estimates of D and λ are consistent (within a factor of two) with previously reported point estimates [33]. However, unlike previous point-estimates of D and λ, our approach provides a well-defined quantitative estimate of the uncertainty present in the parameter recovery. Furthermore, our approach does not require overly complicated and time-consuming experimental procedures such as tracking individual cells [5, 45], constructing cell density profiles [45] or performing multiple sets of assays in which proliferation is artificially suppressed [33]. We note that other measures of uncertainty could be implemented, such as the minimal credible interval, but as the marginal distributions are approximately normal the credible interval is relatively insensitive to the choice of uncertainty measure.

For our parameter estimates ($D \approx 1350$ μm^2/h and $\lambda \approx 2.5 \times 10^{-2}$/h) with $\Delta = 25$ μm and $\tau = 1/24$ h, Equation (4.1) gives $P_m = 0.36$ and $P_p = 0.00104$. Therefore, the average time between motility events for an isolated agent, τ/P_m, is approximately 7 minutes whereas the average time between proliferation events for an isolated agent, τ/P_p, is approximately 40 hours. These quantities are biologically realistic since the timescale of cell motility is much shorter than the timescale of cell proliferation, and these quantities are consistent with previous estimates of the timescale of cell motility [15] and previous estimates of the timescale of cell proliferation [13]. While it is possible to impose additional conditions on our computational simulations, such as explicitly ensuring that no two proliferation events ever occur in rapid succession, we have avoided introducing such details to ensure that our computational simulations are consistent with previously-reported algorithms [31]

4.4 Discussion and conclusions

Scratch assays are a technically simple and inexpensive method used to observe spreading cell fronts [18] which can be thought of as a simple representation of wound healing, malignant spreading or certain developmental processes. The impact of biochemical compounds on cell diffusivity and cell proliferation, vital to cancer and wound healing research, can, in principle, be measured using a scratch assay [10, 12, 16, 46]. However, the majority of these previous studies have reported qualitative data [17, 26] which cannot separately identify D and λ or the impact of the potential treatment on D and λ [10, 12, 16, 19]. While mathematical models have been used to obtain separate point estimates of D and λ [13, 28, 36, 45], these previous studies have neglected to consider the uncertainty present in the parameter recovery process.

The work presented here addresses two common limitations of previous interpretations of scratch assays. First, our method provides quantitative estimates of D and λ by comparing images from a scratch assay with predictions from a lattice-based computational simulation of cell migration and proliferation. Second, to compare the experimental images with the simulation we implement an ABC-MCMC algorithm with an appropriate summary statistic to approximate the posterior distribution of D and λ. The posterior distribution contains vital information about the uncertainty and variability of the recovered parameters, information that is not present in previous interpretations of scratch assays. Using an ABC technique which quantifies this uncertainty will be useful for investigating the efficacy of putative drug treatments, which could be relevant for studying both wound healing [46] and cancer [12]. For example, a traditional approach of estimating D and λ could be used to provide point estimates of D and λ under control conditions and comparing these to separate point estimates of D and λ for an experiment in which the drug has been applied. Alternatively, our approach could be used to produce a distribution of D and λ under control conditions and comparing these distributions to those obtained by analysing a set of equivalent experiments where the drug was applied. Comparing distributions of D and λ provides additional information that is not possible when comparing point estimates. For example, it allows us to assess our confidence in stating that one treatment is better than another. Furthermore, it will assist us in determining an appropriate number of experimental comparisons to ensure reliable assessment.

Our approach for estimating D and λ from a scratch assay provides more comprehensive information than a traditional method, which typically produces point estimates of D and λ only. However, one of the limitations of our approach is that it relies upon obtaining highly-resolved images of the scratch assay such that the location of each cell in the image can be determined, and acknowledge that this could be non-trivial in some situations. Although we have achieved this here using non-labelled cells, another approach might be to use some kind of nuclear stain to help identify the location of individual cells in the population [33].

To investigate the validity of applying ABC to spatiotemporal experiments, such as scratch assays, we initially attempted to recover estimates of P_m and P_p from synthetic data generated using our computational simulation with pre-specified values of P_m and P_p. By comparing different summary statistics we found that using a weighted average of the pair correlation function $q(i)$, and the number of cells or agents present N, provided a simple yet insightful summary statistic. After confirming the validity of our approach using synthetic data, we applied the same approach to a new experimental data set. Our posterior distribution of D and λ allowed us to estimate $(D, \lambda) \approx (1350\ \mu\text{m}^2/\text{h}, 2.5 \times 10^{-2}\ /\text{h})$, which are consistent with previously-reported estimates [33]. However, unlike previous point estimates, we also obtained information about the uncertainty present in the parameter recovery. The posterior distribution allowed us to estimate credible intervals for both $D = (675 - 1800)$ μm^2/h and $\lambda = (1.7 \times 10^{-2} - 3.1 \times 10^{-2})$ /h very simply using a single experimental data set.

Our approach of interpreting scratch assays using ABC together with a combination of the pair correlation function and the number of cells present in the experimental images can be extended in several ways. For example, in this work, we have only considered experimental data where cell–to–cell adhesion is negligible [39]. However, many cell types, such as glioma [15] and melanoma cells [44] exhibit significant cell-to-cell adhesion. An extension of the computational simulation framework presented here, such as the one presented by Khain *et al.* [15], could be employed to analyse scratch assays conducted with adhesive cells. Khain's random walk model includes an additional dimensionless parameter, $\hat{q} \in [0, 1]$, describing the strength of cell–to–cell adhesion and it would be interesting to investigate whether there is sufficient information present in images from a scratch assay using adhesive cells to robustly recover estimates of Khain's three model parameters, D, λ and $\hat{q}$. Furthermore, other types mathematical models could be considered, either with more detailed descriptions of cell migration and proliferation [38], other more detailed mechanisms of cell–to–cell interaction [22, 25] or different types of mechanical interactions [35]. However, since the application of ABC techniques to interpret scratch assay data has not been previously attempted, the present study focused on a relatively straightforward experimental system which could be interpreted with a model relying on just two parameters. Of course, further extensions are possible and these include applying three–dimensional random walk simulations to describe three–dimensional assays [11, 32, 47]. Alternatively, we could investigate the influence of the assumption of memoryless proliferation, particularly for applications where a large proliferation rate is relevant. Another possible extension of our present study is to explore the limitations of using a lattice–based random walk model. This could be achieved by repeating the ABC analysis using a lattice–free random walk [14, 24] and comparing the estimates of D and λ in the lattice–based and lattice–free frameworks. While this comparison is, in principle, possible to carry out, we note that ABC techniques rely on repeated simulations of the random walk and we note that lattice–free models of collective cell behaviour with crowding effects are significantly more computationally demanding than lattice–based models. Therefore, we leave the extension of applying ABC techniques to a lattice–free model for future analysis.

Acknowledgements. We thank the CRC for Wound Management Innovation and the Australian Research Council (FT130100148). We appreciate the experimental assistance provided by Ms Parvathi Haridas. We thank the two anonymous referees for their helpful comments.

Availability of supporting data. The data set supporting the results of this article is included within the article.

Bibliography

[1] Arciero, J. C., Mi, Q., Branca, M. F., Hackam, D. J. and Swigon, D. [2011], ‘Continuum model of collective cell migration in wound healing and colony expansion’, *Biophysical Journal* **100**(3), 535–543.

[2] Binder, B. J. and Simpson, M. J. [2013], ‘Quantifying spatial structure in experimental observations and agent-based simulations using pair-correlation functions’, *Physical Review E* **88**(2), 022705.

[3] Blum, M. G., Nunes, M., Prangle, D., Sisson, S. et al. [2013], ‘A comparative review of dimension reduction methods in approximate Bayesian computation’, *Statistical Science* **28**(2), 189–208.

[4] Burnham, K. P. and Anderson, D. R. [2002], *Model selection and multimodel inference: a practical information-theoretic approach*, Springer.

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

[6] Chigurupati, S., Mughal, M. R., Okun, E., Das, S., Kumar, A., McCaffery, M., Seal, S. and Mattson, M. P. [2012], ‘Effects of cerium oxide nanoparticles on the growth of keratinocytes, fibroblasts and vascular endothelial cells in cutaneous wound healing’, *Biomaterials* **34**, 2194–2201.

[7] Chowdhury, D., Schadschneider, A. and Nishinari, K. [2005], ‘Physics of transport and traffic phenomena in biology: from molecular motors and cells to organisms’, *Physics of Life Reviews* **2**(4), 318–352.

[8] Codling, E. A., Plank, M. J. and Benhamou, S. [2008], ‘Random walk models in biology’, *Journal of the Royal Society Interface* **5**(25), 813–834.

[9] Doran, M. R., Mills, R. J., Parker, A. J., Landman, K. A. and Cooper-White, J. J. [2009], ‘A cell migration device that maintains a defined surface with no cellular damage during wound edge generation’, *Lab on a Chip* **9**(16), 2364–2369.

[10] Fishman, D. A., Liu, Y., Ellerbroek, S. M. and Stack, M. S. [2001], ‘Lysophosphatidic acid promotes matrix metalloproteinase (MMP) activation and MMP-dependent invasion in ovarian cancer cells’, *Cancer Research* **61**(7), 3194–3199.

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

[12] Gingis-Velitski, S., Loven, D., Benayoun, L., Munster, M., Bril, R., Voloshin, T., Alishekevitz, D., Bertolini, F. and Shaked, Y. [2011], ‘Host response to short-term, single-agent chemotherapy induces matrix metalloproteinase-9 expression and accelerates metastasis in mice’, *Cancer Research* **71**(22), 6986–6996.

[13] Johnston, S. T., Simpson, M. J. and McElwain, D. L. S. [2014], ‘How much information can be obtained from tracking the position of the leading edge in a scratch assay?’, *Journal of The Royal Society Interface* **11**(97), 20140325.

[14] Johnston, S. T., Simpson, M. J. and Plank, M. J. [2013], ‘Lattice-free descriptions of collective motion with crowding and adhesion’, *Physical Review E* **88**(6), 062720.

[15] Khain, E., Katakowski, M., Hopkins, S., Szalad, A., Zheng, X., Jiang, F. and Chopp, M. [2011], 'Collective behavior of brain tumor cells: the role of hypoxia', *Physical Review E* **83**(3), 031920.

[16] Knecht, D. A., LaFleur, R. A., Kahsai, A. W., Argueta, C. E., Beshir, A. B. and Fenteany, G. [2010], 'Cucurbitacin I inhibits cell motility by indirectly interfering with actin dynamics', *PLOS ONE* **5**(11), e14039.

[17] Landesberg, R., Cozin, M., Cremers, S., Woo, V., Kousteni, S., Sinha, S., Garrett-Sinha, L. and Raghavan, S. [2008], 'Inhibition of oral mucosal cell wound healing by bisphosphonates', *Journal of Oral and Maxillofacial Surgery* **66**(5), 839–847.

[18] Liang, C.-C., Park, A. Y. and Guan, J.-L. [2007], 'In vitro scratch assay: a convenient and inexpensive method for analysis of cell migration in vitro', *Nature Protocols* **2**(2), 329–333.

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

[20] Marjoram, P., Molitor, J., Plagnol, V. and Tavaré, S. [2003], 'Markov chain Monte Carlo without likelihoods', *Proceedings of the National Academy of Sciences* **100**(26), 15324–15328.

[21] Meier, F. M., Frommer, K. W., Peters, M. A., Brentano, F., Lefèvre, S., Schröder, D., Kyburz, D., Steinmeyer, J., Rehart, S., Gay, S. et al. [2012], 'Visfatin/pre-b-cell colony-enhancing factor (PBEF), a proinflammatory and cell motility-changing factor in rheumatoid arthritis', *Journal of Biological Chemistry* **287**(34), 28378–28385.

[22] Nnetu, K. D., Knorr, M., Käs, J. and Zink, M. [2012], 'The impact of jamming on boundaries of collectively moving weak-interacting cells', *New Journal of Physics* **14**(11), 115012.

[23] Plagnol, V. and Tavaré, S. [2004], Approximate Bayesian computation and MCMC, *in* 'Monte Carlo and Quasi-Monte Carlo Methods 2002', Springer, pp. 99–113.

[24] Plank, M. J. and Simpson, M. J. [2012], 'Models of collective cell behaviour with crowding effects: comparing lattice-based and lattice-free approaches', *Journal of The Royal Society Interface* **9**(76), 2983–2996.

[25] Poujade, M., Grasland-Mongrain, E., Hertzog, A., Jouanneau, J., Chavrier, P., Ladoux, B., Buguin, A. and Silberzan, P. [2007], 'Collective migration of an epithelial monolayer in response to a model wound', *Proceedings of the National Academy of Sciences* **104**(41), 15988–15993.

[26] Rausch, V., Liu, L., Apel, A., Rettig, T., Gladkich, J., Labsch, S., Kallifatidis, G., Kaczorowski, A., Groth, A., Gross, W. et al. [2012], 'Autophagy mediates survival of pancreatic tumour-initiating cells in a hypoxic microenvironment', *The Journal of Pathology* **227**(3), 325–335.

[27] Riahi, R., Yang, Y., Zhang, D. D. and Wong, P. K. [2012], 'Advances in wound-healing assays for probing collective cell migration', *Journal of Laboratory Automation* **17**(1), 59–65.

[28] Sengers, B. G., Please, C. P. and Oreffo, R. O. [2007], 'Experimental characterization and computational modelling of two-dimensional cell spreading for skeletal regeneration', *Journal of the Royal Society Interface* **4**(17), 1107–1117.

[29] Sherratt, J. A. and Murray, J. [1990], 'Models of epidermal wound healing', *Proceedings of the Royal Society of London. Series B: Biological Sciences* **241**(1300), 29–36.

[30] Shibata, A., Tanabe, E., Inoue, S., Kitayoshi, M., Okimoto, S., Hirane, M., Araki, M., Fukushima, N. and Tsujiuchi, T. [2013], 'Hydrogen peroxide stimulates cell motile activity through LPA receptor-3 in liver epithelial WB-F344 cells', *Biochemical and Biophysical Research Communications* **433**, 317–321.

[31] Simpson, M. J., Landman, K. A. and Hughes, B. D. [2010], 'Cell invasion with proliferation mechanisms motivated by time-lapse data', *Physica A: Statistical Mechanics and its Applications* **389**(18), 3779–3790.

[32] Simpson, M. J., Towne, C., McElwain, D. L. S. and Upton, Z. [2010], 'Migration of breast cancer cells: Understanding the roles of volume exclusion and cell-to-cell adhesion', *Physical Review E* **82**(041901).

[33] Simpson, M. J., Treloar, K. K., Binder, B. J., Haridas, P., Manton, K. J., Leavesley, D. I., McElwain, D. L. S. and Baker, R. E. [2013], 'Quantifying the roles of cell motility and cell proliferation in a circular barrier assay', *Journal of The Royal Society Interface* **10**(82), 20130007.

[34] Sisson, S. A., Fan, Y. and Tanaka, M. M. [2007], 'Sequential monte carlo without likelihoods', *Proceedings of the National Academy of Sciences* **104**(6), 1760–1765.

[35] Smith, A. M., Baker, R. E., Kay, D. and Maini, P. K. [2012], 'Incorporating chemical signalling factors into cell-based models of growing epithelial tissues', *Journal of Mathematical Biology* **65**(3), 441–463.

[36] Swanson, K. R. [2008], 'Quantifying glioma cell growth and invasion in vitro', *Mathematical and Computer Modelling* **47**(5), 638–648.

[37] Swanson, K. R., Bridge, C., Murray, J. and Alvord Jr, E. C. [2003], 'Virtual and real brain tumors: using mathematical modeling to quantify glioma growth and invasion', *Journal of the Neurological Sciences* **216**(1), 1–10.

[38] Tambe, D. T., Hardin, C. C., Angelini, T. E., Rajendran, K., Park, C. Y., Serra-Picamal, X., Zhou, E. H., Zaman, M. H., Butler, J. P., Weitz, D. A. et al. [2011], 'Collective cell guidance by cooperative intercellular forces', *Nature Materials* **10**(6), 469–475.

[39] Todaro, G. J. and Green, H. [1963], 'Quantitative studies of the growth of mouse embryo cells in culture and their development into established lines', *The Journal of Cell Biology* **17**(2), 299–313.

[40] Toni, T. and Stumpf, M. P. [2010], 'Simulation-based model selection for dynamical systems in systems and population biology', *Bioinformatics* **26**(1), 104–110.

[41] Toni, T., Welch, D., Strelkowa, N., Ipsen, A. and Stumpf, M. P. [2009], 'Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems', *Journal of the Royal Society Interface* **6**(31), 187–202.

[42] Treloar, K. K. and Simpson, M. J. [2013], 'Sensitivity of edge detection methods for quantifying cell migration assays', *PLOS ONE* **8**(6), e67389.

[43] Treloar, K. K., Simpson, M. J., Binder, B. J., McElwain, D. L. S. and Baker, R. E. [2014], 'Assessing the role of spatial correlations during collective cell spreading', *Scientific Reports* **4**(5713).

[44] Treloar, K. K., Simpson, M. J., Haridas, P., Manton, K. J., Leavesley, D. I., McElwain, D. L. S. and Baker, R. E. [2013], 'Multiple types of data are required to identify the mechanisms influencing the spatial expansion of melanoma cell colonies', *BMC Systems Biology* **7**(1), 137.

[45] Tremel, A., Cai, A., Tirtaatmadja, N., Hughes, B., Stevens, G., Landman, K. and O'Connor, A. [2009], 'Cell migration and proliferation during monolayer formation and wound healing', *Chemical Engineering Science* **64**(2), 247–253.

[46] Upton, Z., Wallace, H. J., Shooter, G. K., van Lonkhuyzen, D. R., Yeoh-Ellerton, S., Rayment, E. A., Fleming, J. M., Broszczak, D., Queen, D., Sibbald, R. G. et al. [2011], 'Human pilot studies reveal the potential of a vitronectin: growth factor complex as a treatment for chronic wounds', *International Wound Journal* **8**(5), 522–532.

[47] Zaman, M. H., Trapani, L. M., Sieminski, A. L., MacKellar, D., Gong, H., Kamm, R. D., Wells, A., Lauffenburger, D. A. and Matsudaira, P. [2006], 'Migration of tumor cells in 3d matrices is governed by matrix stiffness along with cell-matrix adhesion and proteolysis', *Proceedings of the National Academy of Sciences* **103**(29), 10889–10894.

4.5 Supplementary material

4.5.1 Lattice mapping

To map the positions of cells from the experimental image, where cell position is a continuous variable, to a discrete lattice we first calculate the position of each cell. We then define a mapping from cell position (x_c, y_c) to lattice site (x_L, y_L) through the relationship

$$x_L = \left\lceil \frac{x_c}{\Delta} \right\rceil, \qquad y_L = \left\lceil \frac{y_c}{\Delta} \right\rceil, \tag{4.6}$$

where $\lceil x \rceil$ denotes the ceiling function.

4.5.2 ABC algorithm

Marjoram *et al.* [24] provide a full description of this approach and here we only give a brief outline of the algorithm used in our study.

R1 If at θ step to θ' according to a transition kernel $w(\theta \to \theta')$.
R2 Simulate β' from the model using θ' and calculate the summary statistic $S(\beta')$.
R3 Calculate the distance $||S(\beta) - S(\beta')||$, using an appropriate distance measure.
R4 If $||S(\beta) - S(\beta')|| > \epsilon$ reject θ' and return to **R1**.
R5 Calculate

$$h = \min\left(1, \frac{\pi(\theta')w(\theta' \to \theta)}{\pi(\theta)w(\theta \to \theta')}\right).$$

R6 Accept θ' with probability h.
R7 Return to **R1** until M steps have been attempted.

Initially, we sample θ randomly from the prior distribution, until the corresponding summary statistic is sufficiently close to the experimental summary statistic. We define the transition kernel that proposes θ' values as a bivariate uniform distribution, so that $\theta' \in \theta \pm \Gamma$, where Γ defines the width of the uniform distribution. The transition kernel ensures that $P_m \in [0,1]$, $P_p \in [0,1]$ by truncating the bivariate uniform distribution at the boundaries of the parameter space, if necessary. To measure the differences between two summary statistics we define

$$d\left[S(\beta)\right] = ||S(\beta) - S(\beta')|| = \frac{1}{Q}\left[\sum_{i=1}^{Q}\left(\frac{S(\beta)_i - S(\beta')_i}{1 + S(\beta)_i}\right)^2\right]^{\frac{1}{2}}, \tag{4.7}$$

where $S(\beta)_i$ is the i^{th} data point in $S(\beta)$ and Q is the number of data points in $S(\beta)$. We note that we take the average $d\left[S(\beta)\right]$ value when there are summary statistics taken at multiple time points.

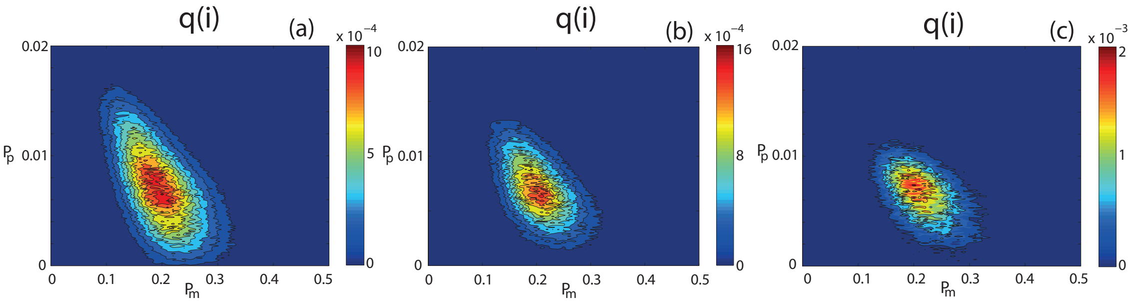

Figure 4.5: Convergence of the posterior distribution for a single synthetic data set for the summary statistic consisting of the pair correlation function $q(i)$. Synthetic data was generated with $P_m = 0.25$, $P_p = 2\times 10^{-3}$. The maximum distance between summary statistics for θ to be accepted was (a) $\epsilon = 0.0135$, (b) $\epsilon = 0.012$, (c) $\epsilon = 0.011$. Red indicates high relative frequency while blue indicates low relative frequency.

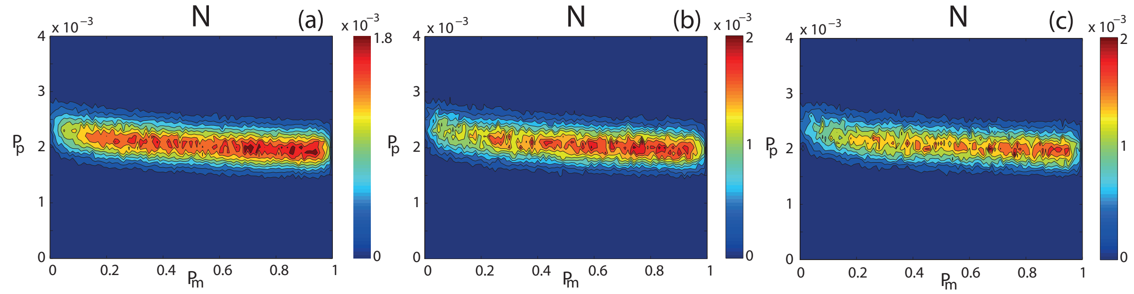

Figure 4.6: Convergence of the posterior distribution for a single synthetic data set for the summary statistic consisting of the number of cells N. Synthetic data was generated with $P_m = 0.25$, $P_p = 2\times 10^{-3}$. The maximum distance between summary statistics for θ to be accepted was (a) $\epsilon = 0.012$, (b) $\epsilon = 0.012$, (c) $\epsilon = 0.01$. Red indicates high relative frequency while blue indicates low relative frequency.

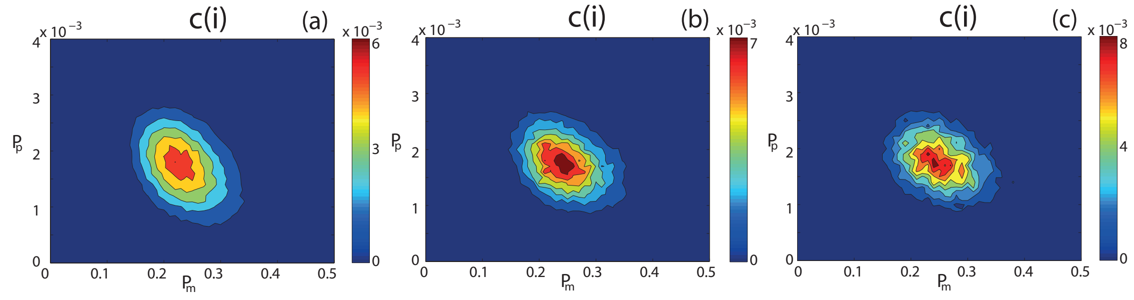

Figure 4.7: Convergence of the posterior distribution for a single synthetic data set for the summary statistic consisting of the counts of pair distances $c(i)$. Synthetic data was generated with $P_m = 0.25$, $P_p = 2\times 10^{-3}$. The maximum distance between summary statistics for θ to be accepted was (a) $\epsilon = 0.075$, (b) $\epsilon = 0.065$, (c) $\epsilon = 0.06$. Red indicates high relative frequency while blue indicates low relative frequency.

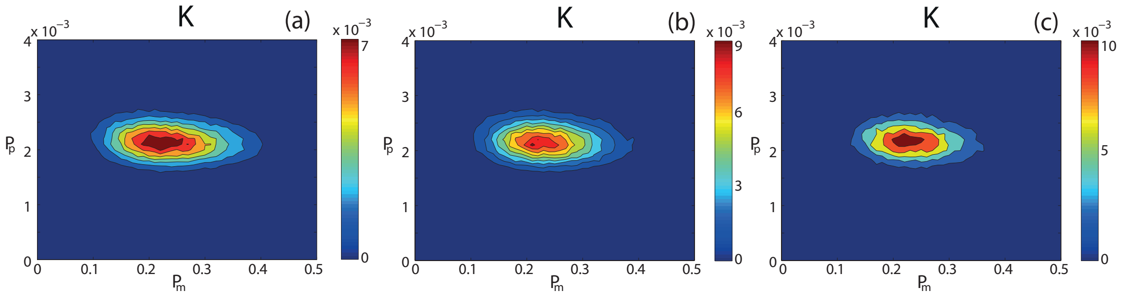

Figure 4.8: Convergence of the posterior distribution for a single synthetic data set for the K summary statistic, where $K = \{d[q(i)] + d[N]\}/2$. Synthetic data was generated with $P_m = 0.25$, $P_p = 2 \times 10^{-3}$. The maximum distance between summary statistics for θ to be accepted was (a) $\epsilon = 0.015$, (b) $\epsilon = 0.0135$, (c) $\epsilon = 0.012$. Red indicates high relative frequency while blue indicates low relative frequency.

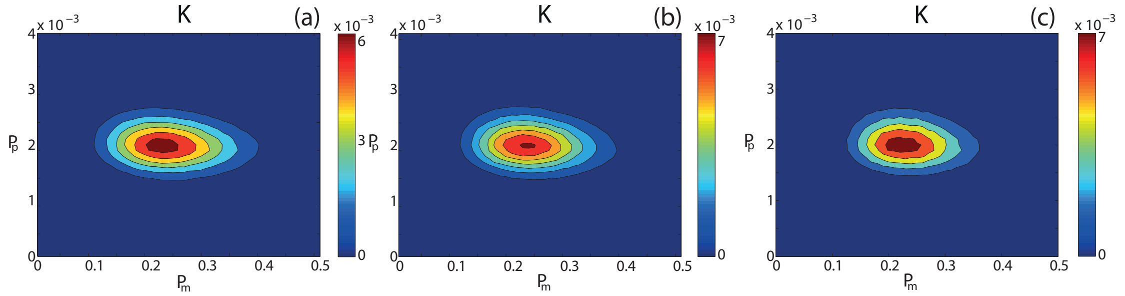

Figure 4.9: Convergence of the averaged posterior distribution for ten identically prepared synthetic data sets for the K summary statistic, where $K = \{d[q(i)] + d[N]\}/2$. Synthetic data was generated with $P_m = 0.25$, $P_p = 2 \times 10^{-3}$. The maximum distance between summary statistics for θ to be accepted was (a) $\epsilon = 0.015$, (b) $\epsilon = 0.0135$, (c) $\epsilon = 0.012$. Red indicates high relative frequency while blue indicates low relative frequency.

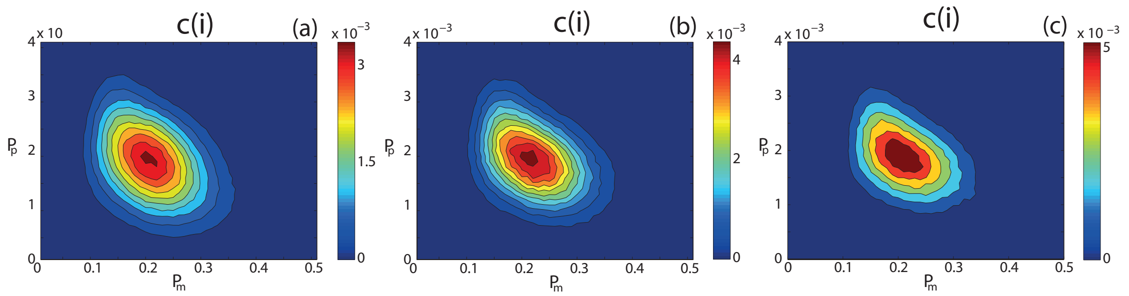

Figure 4.10: Convergence of the averaged posterior distribution for ten identically prepared synthetic data sets for the summary statistic consisting of the counts of pair distances $c(i)$. Synthetic data was generated with $P_m = 0.25$, $P_p = 2 \times 10^{-3}$. The maximum distance between summary statistics for θ to be accepted was (a) $\epsilon = 0.075$, (b) $\epsilon = 0.065$, (c) $\epsilon = 0.06$. Red indicates high relative frequency while blue indicates low relative frequency.

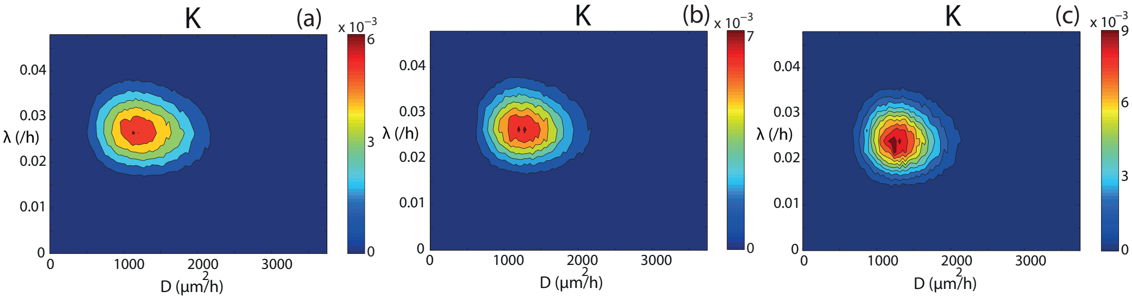

Figure 4.11: Convergence of the posterior distribution for a single experimental data set generated from a scratch assay using the K summary statistic, where $K = \{d[q(i)] + d[N]\}/2$. Synthetic data was generated with $P_m = 0.25$, $P_p = 2 \times 10^{-3}$. The maximum distance between summary statistics for θ to be accepted was (a) $\epsilon = 0.02$, (b) $\epsilon = 0.0175$, (c) $\epsilon = 0.015$. Red indicates high relative frequency while blue indicates low relative frequency.

4.5.3 Distribution convergence

To examine whether the posterior distribution generated from our ABC algorithm approximates $f(\theta|\beta)$, we consider posterior distributions generated with different ϵ values. The posterior distributions are calculated using the same data sets and identically prepared simulations. If the change in the posterior distribution between ϵ values is insignificant then the estimated posterior distribution provides a close approximation to $f(\theta|\beta)$ [23]. In Figures 4.5(a)-(c) we present results using three different ϵ values that demonstrate that the posterior distribution approximately converges for $\epsilon = 0.012$. We observe that the distribution is centred at approximately the same position, with regard to P_m and P_p, and that the spread of the distribution in the P_m and P_p directions are consistent between Figures 4.5(b)-(c). We repeat this process in Figures 4.6-4.11 for all distributions presented in this work and demonstrate the values of ϵ chosen are appropriate.

Chapter 5

Quantifying the effect of experimental design choices for *in vitro* scratch assays

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

Johnston, S. T., Ross, J. V., Binder, B. J., McElwain, D. L. S., Haridas, P. and Simpson, M. J. [2016], 'Quantifying the effect of experimental design choices for *in vitro* scratch assays', *Journal of Theoretical Biology* **400**, 19-31.

Abstract

Scratch assays are often used to investigate potential drug treatments for chronic wounds and cancer. Interpreting these experiments with a mathematical model allows us to estimate the cell diffusivity, D, and the cell proliferation rate, λ. However, the influence of the experimental design on the estimates of D and λ is unclear. Here we apply an approximate Bayesian computation (ABC) parameter inference method, which produces a posterior distribution of D and λ, to new sets of synthetic data, generated from an idealised mathematical model, and experimental data for a non-adhesive mesenchymal population of fibroblast cells. The posterior distribution allows us to quantify the amount of information obtained about D and λ. We investigate two types of scratch assay, as well as varying the number and timing of the experimental observations captured. Our results show that a scrape assay, involving one cell front, provides more precise estimates of D and λ, and is more computationally efficient to interpret than a wound assay, with two opposingly-directed cell fronts. We find that recording two observations, after making the initial observation, is sufficient to estimate D and λ, and that the final observation time should correspond to the time taken for the cell front to move across the field of view. These results provide guidance for estimating D and λ, while simultaneously minimising the time and cost associated with performing and interpreting the experiment.

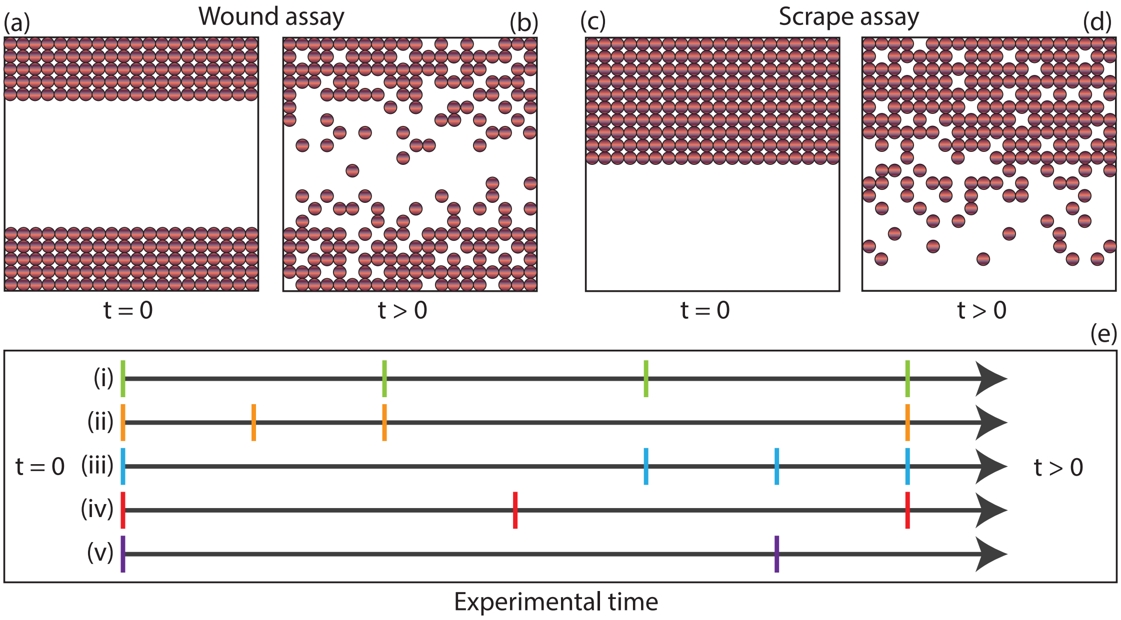

Figure 5.1: Schematic representation of scratch assay design parameters. (a)-(b) Schematic wound assay. (a) Immediately after the scratch, and (b) as the two fronts move and the wound begins to close. (c)-(d) Schematic scrape assay. (c) Immediately after the scratch, and (d) as the single cell front moves. (e) Potential combinations of the timing of experimental observations.

5.1 Introduction

Scratch assays are commonly used to observe the migration and proliferation of cells [11, 19–21, 27, 41]. Cells are placed on a culture dish and incubated, eventually forming a confluent monolayer [21]. An artificial wound is created and images of the resulting collective cell spreading, driven by combined cell migration and proliferation, are captured over 12-24 hours [1, 5, 25, 26, 30]. The number of experimental images captured and reported varies considerably. For example, some results that are reported include just one or two observations [1, 26, 30], whereas others report many more [5, 14, 25]. The influence of the number of observations on the experimental findings has not been quantified.

The majority of scratch assays fall into one of two categories. The first, which we refer to as a *wound assay*, involves the creation of a thin wound [21, 27], which produces two opposingly directed cell fronts that eventually merge [1], as shown in the schematic in Figures 5.1(a)-(b). The second, which we refer to as a *scrape assay*, involves a monolayer that has been more extensively scratched so there is only one cell front [5, 23], as shown in the schematic in Figures 5.1(c)-(d). Again, the influence of the assay choice on the experimental findings has not been quantified.

Previously, scratch assays have been used to investigate the influence of various chemical compounds and potential drug treatments on the rates of cell motility and proliferation [11, 19, 20, 41]. For example, wound healing can be stimulated by steroid treatment [41], while cell proliferation can be inhibited by chemotherapeutic drugs [29]. Developing methods that allow us to robustly quantify the rates of cell migration and proliferation is therefore critical to drug design. However, the majority of scratch assay results are interpreted qualitatively [20, 26] or use simple quantitative measures that do not distinguish between the roles of cell motility and proliferation [11, 19]. In contrast, mathematical models that explicitly distinguish between the roles of cell diffusivity, D, and cell proliferation, λ, for scratch assays have also been presented [5, 12, 14–17, 28, 40]. Previous approaches using both deterministic models [14, 15] and stochastic models [16] lead to point-estimates of D and λ. More recently, approximate Bayesian computation (ABC) has been used to provide estimates of D and λ, together with a measure of the uncertainty associated with these estimates [17]. While ABC methods are computationally demanding compared to deterministic data calibration techniques, such as the Levenberg-Marquadt algorithm [15], the advantage of ABC is that additional information regarding the uncertainty of the model parameter estimates is obtained. Furthermore, prior knowledge about the system can be incorporated in a principled and systematic way, such that knowledge can be accumulated as more data is available. The ABC method of Johnston and coworkers [17] has been used to obtain parameter estimates for 3T3 fibroblasts but did not consider the influence of design parameters on the model parameter estimates.

In this work we refer to two different categories of parameters:

- *Model parameters*, which govern the cell diffusivity, D, and cell proliferation rate, λ, and
- *Design parameters*, which distinguish between different experimental designs of a scratch assay, such as shown in Figure 5.1.

It is instructive to consider how the posterior distribution of D and λ from the ABC algorithm is influenced by the choice of design parameters. The information gained in moving from a prior to a posterior distribution can be quantified using the Kullback-Leibler divergence, D_{KL} [4]. Numerical approximations of posterior distributions are calculated using ABC [24] and have been applied to parameter inference in dynamical systems [37] and spatio-temporal models [17, 42]. Bayesian experimental design is concerned with determining the optimal experimental design that maximises D_{KL}, by adjusting the design parameters [6]. Our mathematical framework, which mimics the random motility and proliferation of mesenchymal (non-adhesive) cells in a scratch assay, contains a significant number of design parameters [17, 32]. However, we require that the design parameters can be adjusted experimentally, as well as in the mathematical model, otherwise the results may have limited practical relevance. The simplest design parameter to alter is the number and timing of the experimental observations. An example of different combinations of experimental observations is presented in Figure 5.1(e). The combination of observation times in (i) contains four equally-spaced observations, whereas in (ii) and (iii) the observations are clustered at the start and end of the experiment, respectively. Currently, it is unclear which combination of observation times provides the most information about the model parameters, D and λ, and Bayesian experimental design allows us to quantify these

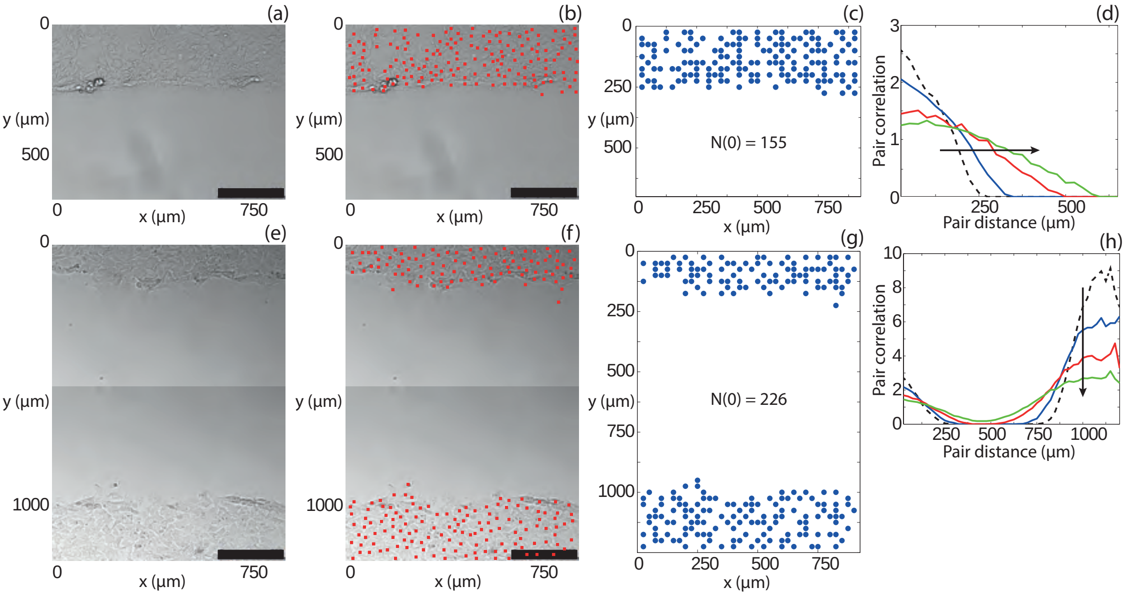

Figure 5.2: Interpretation of experimental images for (a)-(d) the scrape assay and (e)-(h) the wound assay. (a),(e) Experimental images. (b),(f) Identification of the position of cells. The red markers denote the cell centres, and the markers are deliberately chosen to be smaller than the size of the individual cells so that individual cells are not obscured. The scale bar corresponds to 300 μm. (c),(g) Cell positions mapped to a square lattice with $\Delta = 25$ μm. (d),(h) Temporal evolution of the pair correlation function. The pair correlation function is presented at 0 h (dashed black), 4 h (blue), 8 h(red) and 12 h (green). The arrows indicate the direction of increasing time.

differences. Therefore, we may be able to identify the optimal number of experimental images required and consequently avoid collecting additional unnecessary observations, reducing the cost associated with performing and interpreting the experiments.

Here we implement the ABC parameter inference method of Johnston and coworkers [17]. In Section 5.2 we provide a brief description of the experimental procedure and the mathematical model. The influence of the assay choice on the model parameter estimates, using both synthetic and experimental data sets, is examined in Section 5.3. We also investigate the influence of the number and timing of observations on the model parameter estimates, using both synthetic and experimental data sets. In Section 5.4 we discuss the implications of our results, and in Section 5.5, we make recommendations about the experimental design, and suggest options for future work.

5.2 Methods

5.2.1 Experimental method

The experimental method has been presented previously [34]. Briefly, murine fibroblast 3T3 cells [36] are grown in T175 cm^2 tissue culture flasks and one μL of cell suspension is placed into the well of a tissue culture plate. The tissue culture plate is incubated at 37 °C in 5% CO$_2$ overnight so that the cell population settles, attaches to the substrate, and the density begins to increase as cell proliferation takes place. A scratch is made in the monolayer using a P100 pipette tip for the wound assay, and a P1000 pipette tip for the scrape assay. Microscopic images are captured every five minutes for 24 hours after the initial scratch is made using a Leica AF6000 automated microscope. Experimental images of the scrape and wound assays are given in Figures 5.2(a) and (e), respectively. We note that the cell populations in Figures 5.2(a) and (e) appear to be confluent away from the scratched region. However, 3T3 fibroblasts are known to grow to reach significantly higher densities than in Figures 5.2(a) and (e) [34]. This implies that the cell density will increase well beyond the density in Figures 5.2(a) and (e) provided that sufficient time is allowed for more cell proliferation to take place.

5.2.2 Mathematical model

Individual-based random walk models with crowding effects are widely used to mimic the behaviour of cells in biological systems [9, 16, 17, 22, 32]. We consider a two-dimensional random walk on a square lattice, with lattice spacing Δ [7, 32], where each site may be occupied by, at most, one agent. Agents have the potential to move and proliferate with constant probability $P_m \in [0,1]$ and $P_p \in [0,1]$, respectively, during each timestep of fixed duration τ. These probabilities are related to the diffusivity and proliferation rate by [32]

$$D = \frac{P_m \Delta^2}{4\tau}, \qquad \lambda = \frac{P_p}{\tau}. \tag{5.1}$$

Using these relationships we consider the parameters in the discrete model, (P_m, P_p, Δ, τ), as being interchangeable with the model parameters, D and λ [16, 17]. We treat Δ and τ as constants: Δ is measurable [34] and τ is a constant that is chosen to be sufficiently small that the results are independent of the time step [38]. For all simulations $\tau = 1/24$ h. Although our model is an exclusion process, whereby individual agents are subject to crowding effects, previous analysis has shown that the two-dimensional spreading of a population of these agents due to random motility with crowding is described by a linear diffusion mechanism with $D = P_m \Delta^2 / 4\tau$ [31].

In any simulation we have $N(t)$ agents, and during each timestep, $N(t)$ agents are selected sequentially at random, with replacement, and given the opportunity to move with probability P_m. If an agent undergoing a movement event is currently at (x, y) it attempts to move to $(x \pm \Delta, y)$ or $(x, y \pm \Delta)$, without bias. After $N(t)$ agents have attempted to move, an additional $N(t)$ agents are selected sequentially at random, with replacement, and given the opportunity to proliferate with probability P_p. An agent undergoing a proliferation event at (x, y) will attempt to place a daughter agent at $(x \pm \Delta, y)$ or $(x, y \pm \Delta)$, without bias. Potential motility or proliferation events only succeed if the target site is vacant [32].

The geometry of a scratch assay is approximated with a lattice of height $Y\Delta$ and width $X\Delta$. The field of view of our microscope and average cell size provide a natural choice for Δ, Y and X. The average cell diameter of 3T3 fibroblasts has been measured previously, giving $\Delta = 25$ μm [34]. For our synthetic data set we consider an idealised case where X and Y are the same for both assays. For our experimental data set, the experimental image of a scrape assay in Figure 5.2(a) gives $Y = 27$ and $X = 36$, while the experimental image of a wound assay in Figure 5.2(e) gives $Y = 49$ and $X = 36$. Since the approximately spatially uniform population extends well beyond the boundaries of the field of view, there will be no net flux of cells across these boundaries [15]. Hence we apply no net flux (symmetry) boundary conditions along the lines $y = 0,\ y = Y\Delta,\ x = 0,\ x = X\Delta$ [15–17]. To model a scrape assay, we randomly place $N(0)$ agents in the region $y \leq Y_0\Delta$ such that each lattice site is occupied by, at most, one agent. We note that Y_0 is the height of the initially-occupied region in a scrape assay. Our wound assay contains a total of $N(0)$ agents in two regions, $y \leq Y_{0,1}\Delta$ and $y > (Y - Y_{0,2})\Delta$, where $Y_{0,1}$ and $Y_{0,2}$ are the height of the initially occupied regions at $y = 0$ and $y = Y\Delta$, respectively. To calculate $N(0)$ we count the number of cells in the relevant experimental observation at $t = 0$ h.

In the mathematical model we can alter X, Y, $N(0)$, the height of the initially-confluent monolayer and the number and timing of observations captured. However, since we aim to mimic an experiment, we recognise that some design parameters are constrained by the experimental procedure. For example, X and Y are defined by the field of view of the microscope. Therefore, without additional equipment, we cannot alter this feature of the experimental domain. In addition, $N(0)$ is determined by the cell density at confluence, which is cell-type specific and cannot be easily varied. The height of the initially confluent monolayer, either Y_0 or $Y_{0,1}$ and $Y_{0,2}$, depends on the instrument used to perform the scratch and is difficult to reproduce reliably, and is therefore inappropriate to consider as a design parameter. We therefore focus on varying the number and timing of experimental observations captured. A single observation must be captured to determine the state of the system at $t = 0$ h and, consequently, we can only vary the number of observations captured after the initial observation. Therefore, the number of experimental observations captured refers to the number of observations captured after the initial observation.

5.2.3 Data interpretation

Both the synthetic data set, produced by our mathematical model, and the experimental data set contain the spatial positions of $N(t)$ cells in each observation. We note that the cell positions in the experimental data set are obtained manually while the cell positions in the synthetic data set are obtained automatically. However, it is computationally intractable to compare the data sets using the spatial position of all cells. Johnston and coworkers [17] found that considering the number of cells, $N(t)$, and the pair correlation function, $q(i)$ [3], provides a summary statistic that contains a large percentage of the complete information. A summary statistic provides a lower-dimensional summary of a data set that allows tractable comparisons between data sets [10, 24]. Full details of the pair correlation function are presented elsewhere [3], and are summarised here. Briefly, we count the number of pairs of cells, separated by a distance of $i\Delta$ μm, in the y-direction and define this as $c(i)$, which is the counts of pair distances. We note that we could repeat

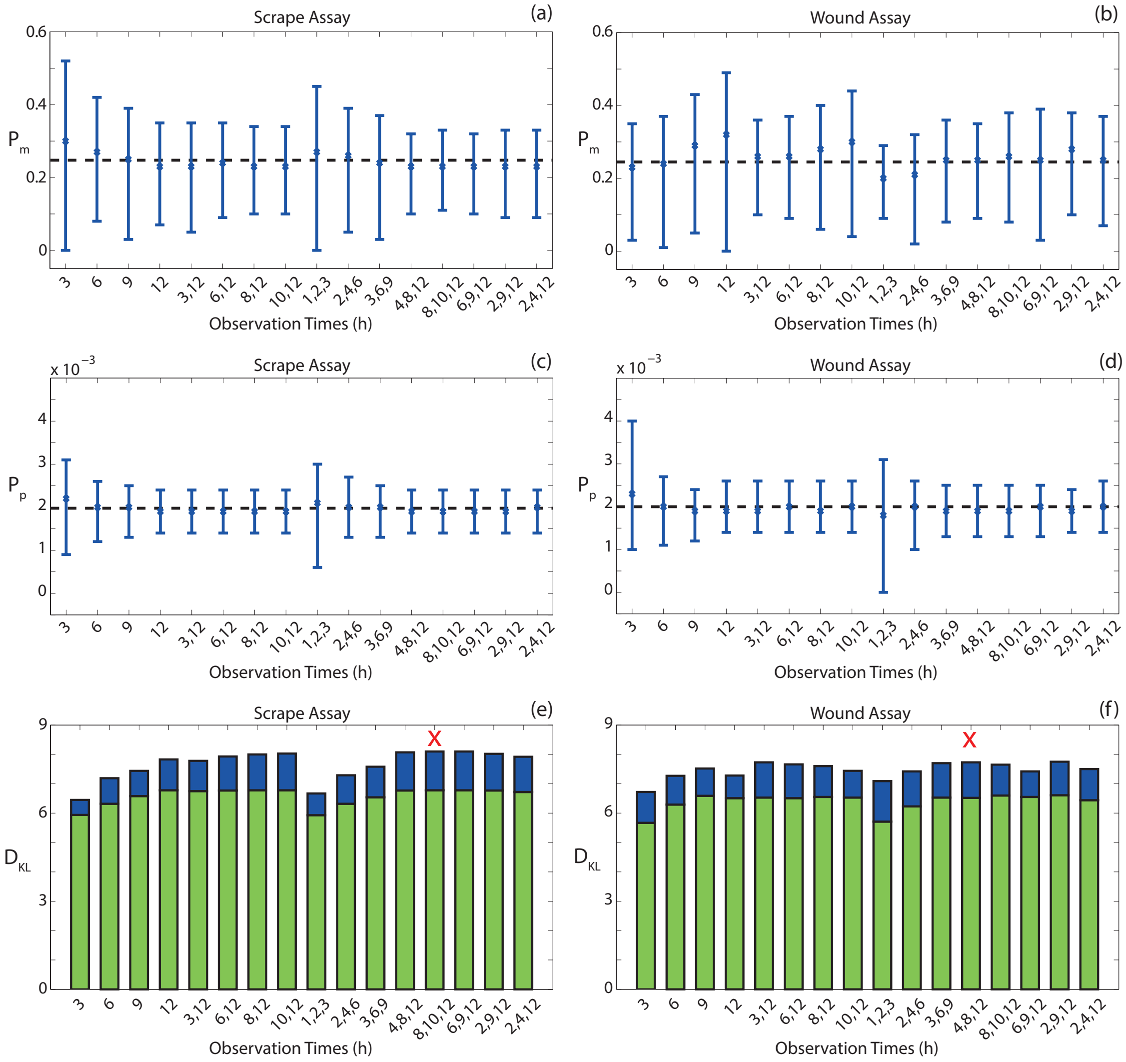

Figure 5.3: Mean model parameter estimates with 90% credible intervals, for a synthetically generated data set interpreted with a suite of combinations of observation times. (a)-(b) Mean estimates of P_m. (c)-(d) Mean estimates of P_p. The dashed line corresponds to the model parameter values used to obtain the synthetic data set. The blue crosses correspond to the mean values. (e)-(f) Mean D_{KL} values. The blue bar corresponds to the information gained about P_m and the green bar corresponds to information gained about P_p. The red crosses denote the combination of observation times that results in the highest D_{KL} value. Data was generated using $(P_m, P_p) = (0.25,\ 2 \times 10^{-3})$. For all simulations $\tau = 1/24$ h, $N(0) = 100$, $M = 10^6$, $X = 36$, $Y = 27$, $\Delta = 25$ μm. For scrape assays, $Y_0 = 10$. For wound assays, $Y_{0,1} = Y_{0,2} = 5$.

this for the x-direction, however since the initial distribution of cells in the x-direction is, on average, uniform, and remains uniform throughout the experiment [3], we choose to focus on the counts of pair distances in the y direction only. The counts of pair distances are normalised to obtain the pair correlation function

$$q(i) = \frac{c(i)}{X^2(Y-i)\rho\bar{\rho}}, \tag{5.2}$$

where $\rho = N(t)/(XY)$ is the mean density and $\bar{\rho} = (N(t) - 1)/(XY - 1)$. The process of obtaining $q(i)$ from the experimental images is shown in Figure 5.2. To do this we identify the position of the centre of each cell in Figures 5.2(a) and (e) and mark their position with a red square in Figures 5.2(b) and (f), respectively. The red markers are deliberately chosen to be smaller than the cell diameter so that superimposing the markers on the experimental images does not obscure the view of the cells. We map the cell positions onto the lattice described in Section 5.2.2. We present the resulting lattice for the cell positions identified in Figures 5.2(b) and (f) in Figures 5.2(c) and (g), respectively. The corresponding pair correlation functions are presented in Figures 5.2(d) and (h), respectively, at 0, 4, 8 and 12 h.

5.2.4 Approximate Bayesian computation

To investigate which values of D and λ can generate summary statistics that are consistent with our synthetic or experimental data set we apply an ABC algorithm [2, 17, 24, 35, 37]. The ABC algorithm involves performing M identically-prepared stochastic realisations of the mathematical framework described in Section 5.2.2 and uses a combination of $N(t)$ and $q(i)$ as a summary statistic to determine an approximate posterior distribution of D and λ [17]. We note that we consider the uniform prior distribution for all data sets. Further details are given in the Supplementary material.

5.3 Results

5.3.1 Assay choice: synthetic data set

In the experimental literature there has been no explicit discussion of whether the choice of performing a scrape or wound assay influences our ability to estimate D and λ. Therefore, we attempt to recover estimates of D and λ from synthetic data sets generated from the mathematical framework, for both assays. We consider biologically relevant model parameters $(P_m, P_p) = (0.25, 2\times10^{-3})$, where cell proliferation occurs over a significantly longer timescale than cell motility [34]. We perform ten identically-prepared realisations of the mathematical model for both assays. For each realisation and assay we calculate $q(i)$ and $N(t)$ at $t = 0, 1, 2, ..., 12$ h [17]. We note that, in our synthetic experiments, $t = 12$ h approximately corresponds to the observed time taken for a cell front to move across a typical experimental field of view. For each realisation, we apply the ABC algorithm with the appropriate initial condition, and calculate the approximate posterior distribution of D and λ. For each assay, we then average the posterior distributions [16]. To measure the amount of information gained from the ABC algorithm we discretise the posterior distribution into 10^4 equally-spaced values of P_p, for $P_p \in [0, 1]$, and 10^2 equally-spaced values of P_m, for $P_m \in [0, 1]$. Since proliferation takes place on a longer timescale than motility [17, 39], we require a finer resolution for P_p than P_m. We then calculate the Kullback-Leibler divergence [4]

$$D_{KL}(f|\pi) = \sum_{j=1}^{10^6} f(\underline{\theta}_j|\beta)\ln\left(\frac{f(\underline{\theta}_j|\beta)}{\pi(\underline{\theta}_j)}\right), \tag{5.3}$$

Observation Times (h)	3	6	9	12
Scrape Assay P_m 90% CI	0.52	**0.34**	**0.36**	**0.28**
Wound Assay P_m 90% CI	**0.32**	0.36	0.38	0.49
Scrape Assay P_p 90% CI (10^{-3})	**2.2**	**1.4**	**1.2**	**1.0**
Wound Assay P_p 90% CI (10^{-3})	3.0	1.6	**1.2**	1.2

Observation Times (h)	3,12	6,12	8,12	10,12
Scrape Assay P_m 90% CI	0.30	**0.26**	**0.24**	**0.24**
Wound Assay P_m 90% CI	**0.26**	0.28	0.34	0.40
Scrape Assay P_p 90% CI (10^{-3})	**1.0**	**1.0**	**1.0**	**1.0**
Wound Assay P_p 90% CI (10^{-3})	1.2	1.2	1.2	1.2

Observation Times (h)	1,2,3	2,4,6	3, 6,9	4,8,12
Scrape Assay P_m 90% CI	0.45	0.34	0.34	**0.22**
Wound Assay P_m 90% CI	**0.20**	**0.30**	**0.28**	0.26
Scrape Assay P_p 90% CI (10^{-3})	**2.4**	**1.4**	**1.2**	**1.0**
Wound Assay P_p 90% CI (10^{-3})	3.1	1.6	**1.2**	1.2

Observation Times (h)	8,10,12	6,9,12	2,9,12	2,4,12
Scrape Assay P_m 90% CI	**0.22**	**0.22**	**0.24**	**0.24**
Wound Assay P_m 90% CI	0.30	0.36	0.28	0.30
Scrape Assay P_p 90% CI (10^{-3})	**1.0**	**1.0**	**1.0**	**1.0**
Wound Assay P_p 90% CI (10^{-3})	1.2	1.2	**1.0**	1.2

Table 5.1: Comparison between the width of the 90% credible interval (CI), symmetric around the mode, for (i) P_m in the scrape assay; (ii) P_m in the wound assay; (iii) P_p in the scrape assay, and (iv) P_p in the wound assay for a suite of combinations of observation times. The assay type with the narrower credible interval for each combination of observation times and parameter is presented in bold.

where f is the posterior distribution, β is a data set, π is the uniform prior distribution, $\underline{\theta}$ is a model parameter pair and the index j accounts for all possible parameter pairs. We also calculate the marginal distributions for P_m and P_p by averaging the posterior distribution across P_p and P_m, respectively, and evaluate the corresponding value of D_{KL}. The mean and 90% credible interval, symmetrical around the mode, for P_m and P_p, are then estimated from the marginal distributions.

The posterior distributions of D and λ are calculated, for both assays, using a suite of combinations of observation times. Typically, scratch assays are interpreted with, at most, two observation times [1, 26, 30]. Therefore, we restrict our investigation to, at most, three observation times after the initial scratch is made.

The mean model parameter estimates and 90% credible intervals for both assays are given in Figures 5.3(a)-(d), indicating that the majority of combinations of observation times provide estimates of P_m and P_p that match those used to generate the synthetic data set for both assays. However, the credible intervals indicate that there is greater uncertainty in the estimates of P_m for the majority of combinations of observation times in the wound assay. We highlight the assay type with a narrower credible interval, corresponding to less uncertainty, for each combination of observation times in Figure 5.3 in Table 5.1. We attribute the greater uncertainty about P_m in the wound assay to the fact that the two cell fronts begin to interact toward the end of the experiment, which results in pair correlation functions that can be replicated with different values of P_m.

Interestingly, the credible intervals for P_p are similar for both the wound and scrape assay, suggesting that the difference in assay geometry does not influence the uncertainty associated with P_p. It is possible that the uncertainty associated with P_p is proportional to $N(0)$, which is the same for both synthetic assays. To explore this possibility, we repeat the synthetic scrape assay with $N(0) = 50$ and find that, for all combinations of observation times, there is more uncertainty in the recovered parameters compared to the synthetic scrape assay with $N(0) = 100$ (Tables 5.1-5.2, Supplementary material). To examine the amount of information gained about the model parameters, we present D_{KL} for the suite of combinations of observation times, for both assay types, in Figures 5.3(e)-(f). Our results indicate that, predominantly, the scrape assay gives higher D_{KL} values when compared to the same combination of observation times in the wound assay. We note that there are cases where the D_{KL} value is higher for the wound assay. However, this occurs for combinations of observations times where the final observation is made at short time and hence these do not correspond to the designs that provide the most information. We denote the combination of observation times that provides the most information both the scrape and wound assay with a red cross in Figures 5.3(e)-(f), respectively. The most informative design in the scrape assay results in a D_{KL} value of 8.11, compared to 7.90 in the wound assay. For both assay types, the most informative design includes a final observation time at $t = 12$h, which corresponds to the final observation captured. We note that the Kullback-Leibler divergence is logarithmic, meaning that a large difference in uncertainty can correspond to a small difference in D_{KL}. To illustrate this, consider the D_{KL} values and uncertainty associated with P_p in the scrape assay at 1, 2, 3 h and 2, 4, 6 h. The D_{KL} values are 5.93 and 6.32, respectively, while the width of the 90% credible intervals are 2.4×10^{-3} and 1.4×10^{-3}, respectively. In this case a 6% increase in D_{KL} corresponds to a 42% reduction in the credible interval. Tabular form of the data in Figure 5.3 is provided in the Supplementary material.

5.3.2 Assay choice: experimental data set

We now repeat the process in Section 5.3.1 for an experimental data set obtained using the procedure described in Section 5.2.1. We note that the wound assay was performed over 24 h whereas the scrape assay was performed over 12 h. These timescales were chosen since they indicate the amount of time required for the cell front(s) in the two different assays to move across their respective field of view. Without *a priori* information about D, it is difficult to choose an appropriate final time. Therefore, we capture experimental observations relatively frequently until the cell front has moved across the field of view. After the final observation has been captured, we select the relevant earlier

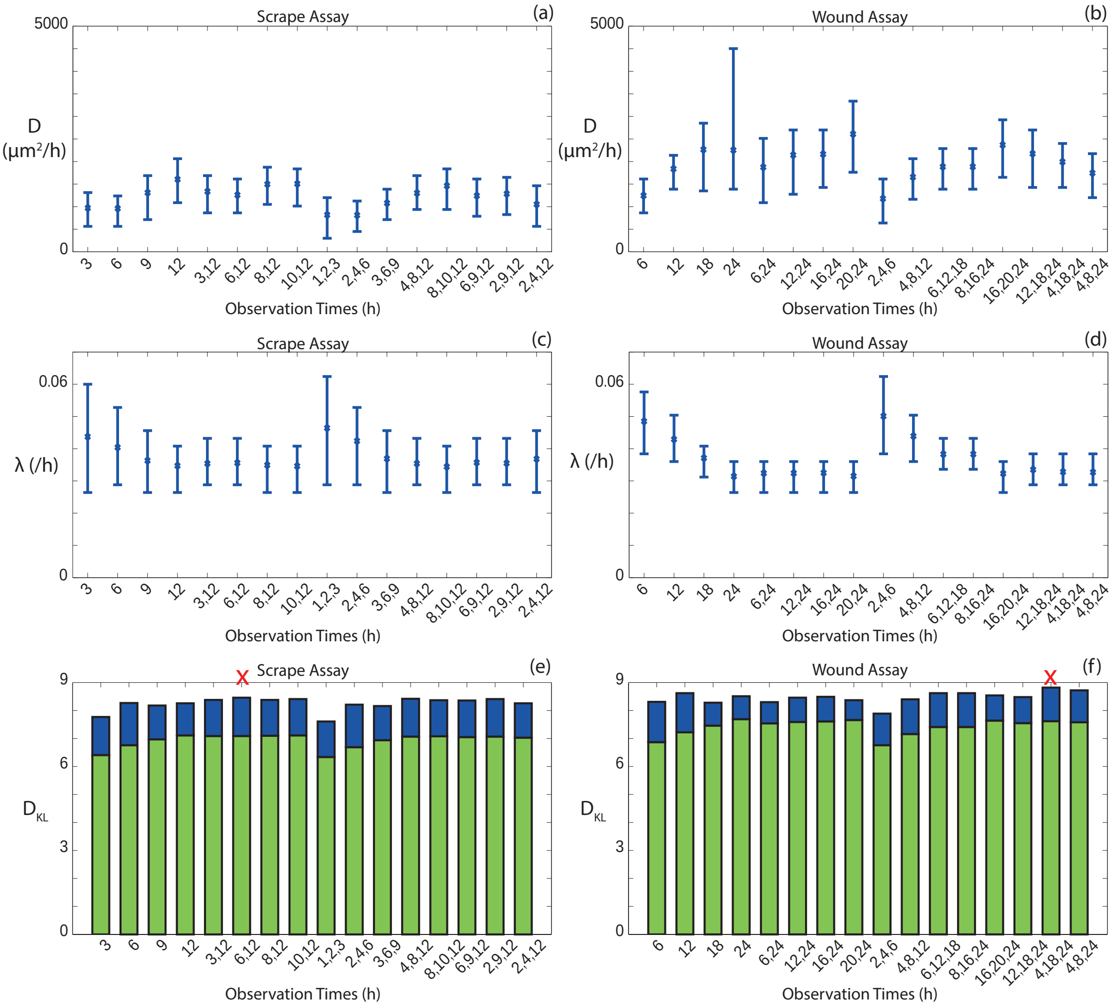

Figure 5.4: Mean model parameter estimates with 90% credible intervals, for an experimentally generated data set interpreted with a suite of combinations of observation times. (a)-(b) Mean estimates of D, given by Equation (5.3). (c)-(d) Mean estimates of λ. The blue crosses correspond to the mean values. (e)-(f) Mean D_{KL} values, given by Equation (5.3). The blue bar corresponds to the information gained about D and the green bar corresponds to information gained about λ. The red crosses denote the combination of observation times that results in the highest D_{KL} value. For all simulations $\tau = 1/24$ h, $M = 10^6$, $X = 36$, $Y = 27$, $\Delta = 25$ μm. For scrape assays, $N(0) = 155$, $Y_0 = 10$. For wound assays, $N(0) = 226$, $Y_{0,1} = 7$, $Y_{0,2} = 9$.

observations to interpret our experimental results. While it is possible to capture experimental observations after the front(s) have moved across the field of view, it is unlikely that there is any additional information to be gained. Since the wound and scrape assays are different experiments, and hence are captured over different time periods, we compare combinations of observation times relative to the final time of the experiment, as this systematically utilises all relevant data for each experiment. For example, 6 and 12 h in the scrape assay are compared to 12 and 24 h in the wound assay. It would be computationally infeasible to explore all possible observation times and, as such, we attempt to provide a fair comparison by selecting observation times in a sensible manner.

We present mean estimates of D and λ, with 90% credible intervals, symmetric about the mode, for both assays in Figures 5.4(a)-(d). We observe that the qualitative trend for the mean model parameter estimates with regard to the combination of observation times is consistent for both assays. This suggests that comparing combinations of observation times after the same proportion of the final observation time is reasonable.

Unlike the synthetic data set, we observe that the scrape and wound assay provide different estimates of the cell diffusivity, D. Given that estimates of D for 3T3 fibroblasts are reported to be in the range 30-3000 μm^2/h [5, 17, 39, 40], our observed variation is not necessarily surprising. However, we note that our estimates of D are consistent with previous results obtained by interpreting experiments using 3T3 cells with an ABC algorithm [17]. Comparing the credible intervals for the two assays we observe that, again, there is more uncertainty associated with D for the majority of combinations of observation times in the wound assay.

Our estimates of the cell proliferation rate, λ, from the two assays are similar. This is consistent with the synthetic data set, suggesting that estimates of λ are relatively insensitive to whether we consider a scrape or wound assay. However, the 90% credible intervals for λ are always smaller for the wound assay. The reduction in uncertainty may be attributed to the larger number of initial cells in the wound assay, as we observed similar credible intervals in the synthetic data set where $N(0)$ is the same for both assays. Evaluating the D_{KL} value for each combination of observation times and both assays, presented in Figures 5.4(e)-(f), respectively, suggests that more information is obtained from the wound assay. However, decomposing the D_{KL} value into the D_{KL} values for the marginal distributions for D and λ, which correspond to the blue and green bars in Figures 5.4(e)-(f), respectively, suggests that while more information is obtained about λ in the wound assay, the scrape assay provides more information about D. The most informative combination of observation times, denoted by the red crosses in Figures 5.4(e)-(f), provides more information in the wound assay than in the scrape assay. Again, however, this additional information is associated with λ. Furthermore, similar to the results obtained from the synthetic data set, the most informative combination of observations for both assays include the final experimental observation. Tabular form of the data in Figure 5.4 is provided in the Supplementary material.

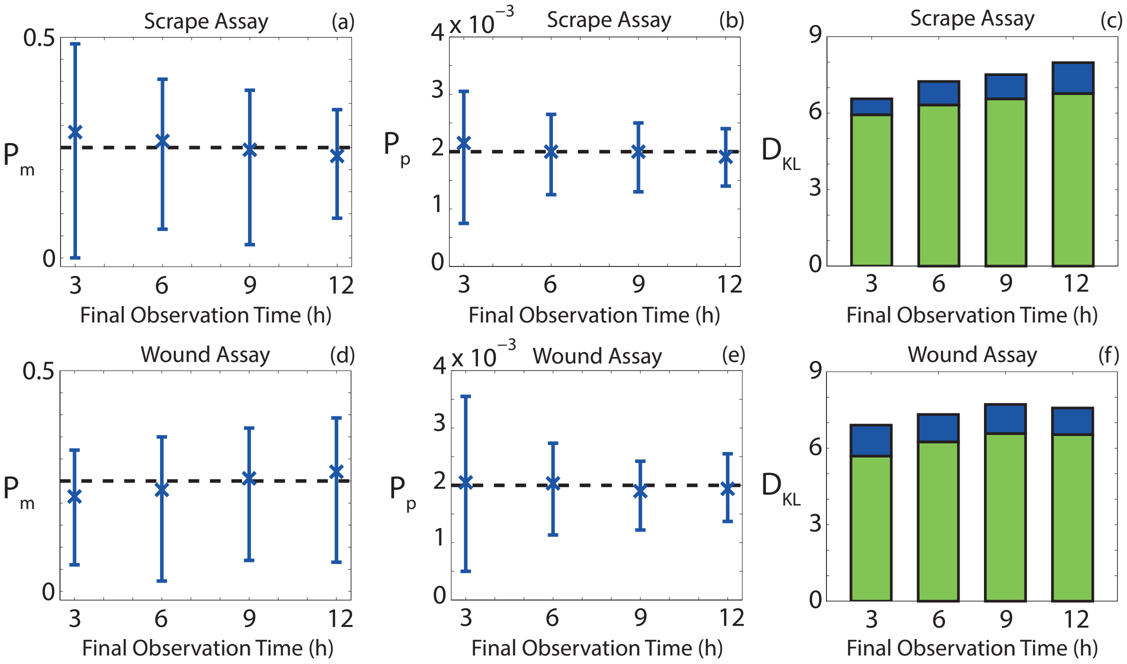

Figure 5.5: Results obtained from the parameter inference approach applied to the synthetic data set generated with $(P_m, P_p) = (0.25, 2 \times 10^{-3})$. (a)-(c) Mean estimates of P_m and P_p with corresponding 90% credible intervals, and the mean D_{KL} values, respectively, for the scrape assay for a final observation time of 3 h, 6 h, 9 h and 12 h. (d)-(f) Mean estimates of P_m and P_p with corresponding 90% credible intervals, and the mean D_{KL} values, respectively, for the wound assay for a final observation time of 3 h, 6 h, 9 h and 12 h. The mean D_{KL} values were calculated using Equation (5.3). The green bar corresponds to the D_{KL} value for P_p and the blue bar corresponds to the D_{KL} value for P_m. For all simulations $\tau = 1/24$ h, $N(0) = 100$, $M = 10^6$, $X = 36$, $Y = 27$, $\Delta = 25$ μm. For scrape assays, $Y_0 = 10$. For wound assays, $Y_{0,1} = Y_{0,2} = 5$. The crosses correspond to the mean values.

Unfortunately, the additional information obtained from the wound assay is associated with two important limitations. First, the interpretation of the experimental images is significantly more time-consuming due to higher $N(t)$ values. Second, the time required to perform the ABC algorithm for the experimental wound assay is significantly longer than for the experimental scrape assay, due to the increase in both the final time and initial number of cells. Running in parallel on ten cores (2.66 GHz Intel Xeon E5-2670), the ABC algorithm for the wound assay required approximately one week of computation time, whereas the scrape assay required approximately one day of computation time.

5.3.3 Choice of observation times: synthetic data set

Typically, in the experimental literature there is no explicit discussion about the choice of the duration of the experiment [1, 17, 19, 26]. Therefore, it is instructive to compare estimates of D and λ, and the corresponding D_{KL} values, for different final time points using synthetic data. Using the approximate posterior distributions generated in Section 5.3 we calculate the mean estimates of P_m and P_p, and the corresponding mean 90% credible intervals, symmetric around the mode, for a final observation time of 3, 6, 9 and 12 h. Results are presented in Figures 5.5(a)-(b) and Figures 5.5(d)-(e), for the

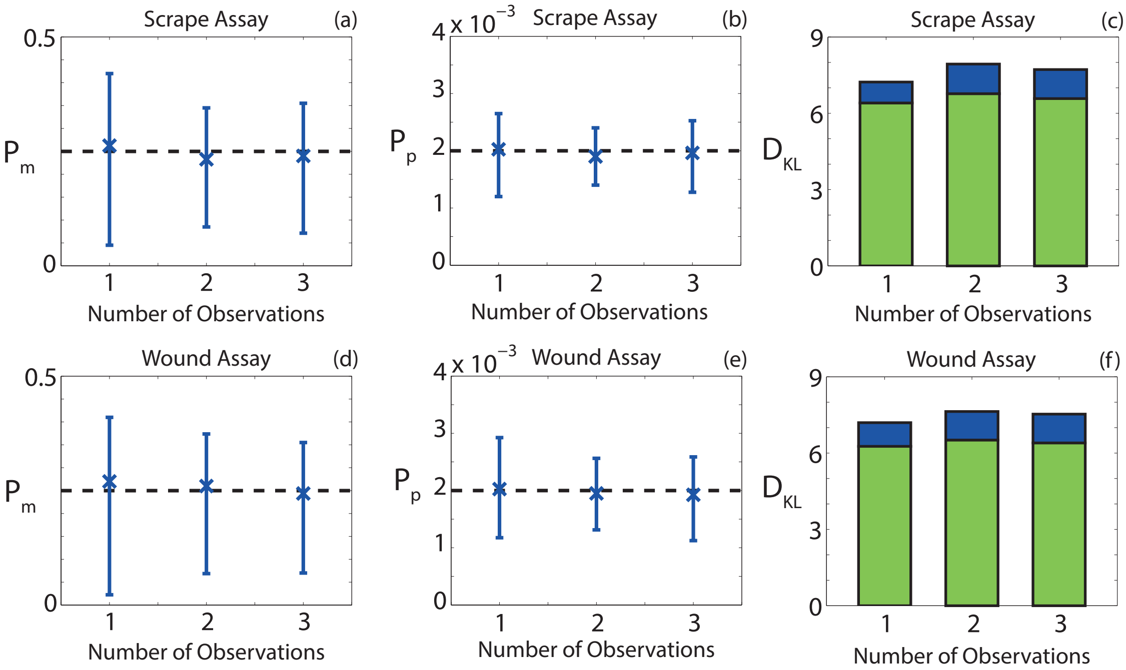

Figure 5.6: Results obtained from the parameter inference approach applied to the synthetic data set generated with $(P_m, P_p) = (0.25, 2 \times 10^{-3})$. (a)-(c) Mean estimates of P_m and P_p with corresponding 90% credible intervals, and the mean D_{KL} values, respectively, for the scrape assay for one, two and three observations. (d)-(f) Mean estimates of P_m and P_p with corresponding 90% credible intervals, and the mean D_{KL} values, respectively, for the wound assay for one, two and three observations. The mean D_{KL} values were calculated using Equation (5.3). The green bar corresponds to the D_{KL} value for P_p and the blue bar corresponds to the D_{KL} value for P_m. For all simulations $\tau = 1/24$ h, $N(0) = 100$, $M = 10^6$, $X = 36$, $Y = 27$, $\Delta = 25$ μm. For scrape assays, $Y_0 = 10$. For wound assays, $Y_{0,1} = Y_{0,2} = 5$. The crosses correspond to the mean values.

scratch assay and the wound assay, respectively. For both assays we see that the mean parameter estimates depend on the choice of final observation time. To determine which final observation time provides the most information about D and λ we calculate the mean D_{KL} value, for both P_m and P_p, for each final observation time. The mean D_{KL} values are presented in Figures 5.5(c) and (f) for the scrape assay and wound assay, respectively. We note that the green bar reflects the information gained about P_p while the blue bar reflects the information gained about P_m. For the scrape assay, we observe that a final observation time of 12 h provides the most information about both P_m and P_p and note that 12 h corresponds to the approximate time taken for the cell front to move across the experimental field of view. This result is intuitive, as we are able to observe cell migration and proliferation until the cell front crosses, and leaves, the experimental field of view. For the wound assay, we observe that the information gained peaks at a final observation time of 9 h. Again, this result is intuitive as the two opposingly-directed cell fronts begin to interact towards the end of the experiment, and hence information is lost beyond that time. This suggests that interpreting experimental observations in a wound assay after the two cell fronts merge does not provide additional information about D and λ.

Currently, there is no explicit discussion in the experimental literature about the choice of the number of observations. Typically there is, at most, two observations made after the initial observation [1, 26, 30], which is due to the cost associated with capturing and interpreting experimental observations. Therefore, we restrict our analysis to consider, at most, three observations after the initial observation. We present the mean parameter estimates, and the mean of the corresponding 90% credible intervals, symmetric around the mode, that we obtain when we analyse one, two and three observations in Figures 5.6(a)-(b) and Figures 5.6(d)-(e), for the scrape and wound assays, respectively. While our estimates of P_m and P_p are sensitive to the number of observations, they are less sensitive compared to the choice of final observation time. To determine whether there is any benefit in capturing and interpreting additional experimental observations, we calculate the mean D_{KL} value for P_m and P_p, for one, two and three observations. To obtain the mean D_{KL} value for one observation we consider the mean of four D_{KL} values. Each of these D_{KL} values is obtained using a combination of observation times that contains a single observation, that is, either 3 h, 6 h, 9 h or 12 h. We follow a similar process to obtain the mean D_{KL} values for two and three observations. For two and three observations we calculate the mean of the D_{KL} values obtained from all combinations of observation times that contain two and three observations, respectively. The mean D_{KL} values are given in Figures 5.6(c) and (f). In both assays, on average, there is a slight increase in the amount of information gained when we make two observations relative to when we make one observation. However, there is a much smaller amount of information gained when we make three observations, compared to two observations. The third observation does not provide a significant amount of additional information, even if we only compare combinations of observations that include $t = 12$ h. With this additional restriction, in the scrape assay, we obtain mean D_{KL} values of 7.95 and 8.05 for two and three observations, respectively. In the wound assay, we obtain mean D_{KL} values of 7.64 for both two and three observations. The lack of additional information obtained from interpreting three observations compared to two observations implies that there is likely to be further diminishing returns and we therefore recommend that making two observations is sufficient.

5.3.4 Choice of observation times: experimental data set

We repeat the process described in Section 5.3.3 for our experimental data set and calculate the mean model parameter estimates, and corresponding mean 90% credible intervals, for different final observation times. Again, we note that the wound assay was performed over 24 h and that we compare observation times after the same proportion of time has elapsed relative to the final time. For example, we compare an observation at 6 h in the scrape assay with an observation at 12 h in the wound assay. We present the mean model parameter estimates and 90% credible intervals in Figures 5.7(a)-(b) and Figures 5.7(d)-(e), for the scrape and wound assays, respectively. Similar to the synthetic results, we observe that the estimates of D are more sensitive to the choice of final observation time in the wound assay. Additionally, estimates of λ are more sensitive to the choice of

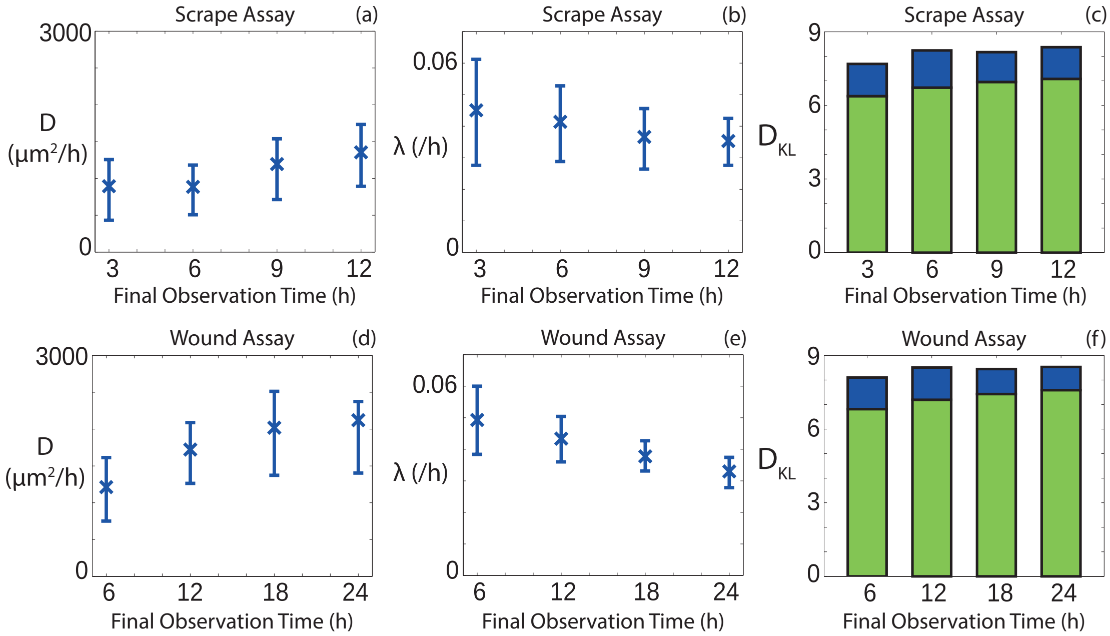

Figure 5.7: Results obtained from the parameter inference approach applied to the experimental data set. (a)-(c) Mean estimates of P_m and P_p with corresponding 90% credible intervals, and the mean D_{KL} values, respectively, for the scrape assay for a final observation time of 3 h, 6 h, 9 h and 12 h. (d)-(f) Mean estimates of P_m and P_p with corresponding 90% credible intervals, and the mean D_{KL} values, respectively, for the wound assay for a final observation time of 3 h, 6 h, 9 h and 12 h. The mean D_{KL} values were calculated using Equation (5.3). The green bar corresponds to the D_{KL} value for λ and the blue bar corresponds to the D_{KL} value for D. For all simulations $\tau = 1/24$ h, $M = 10^6$, $X = 36$, $Y = 49$, $\Delta = 25$ μm. For scrape assays, $N(0) = 155$, $Y_0 = 10$. For wound assays, $N(0) = 226$, $Y_{0,1} = 7, Y_{0,2} = 9$.

final observation time in the wound assay, a trend that is not observed in the synthetic data set. Interestingly, while the estimates of λ are more sensitive in the wound assay, the D_{KL} values corresponding to λ, represented by the green bar in Figures 5.7(c) and (f), are higher for the wound than the scrape assay. We attribute this to the additional number of cells present in the wound assay, which may influence the information gained but not the relationship between λ and the final observation time. The D_{KL} value for the scrape assay, given in Figure 5.7(c), increases with the final observation time. Conversely, the D_{KL} value for the wound assay, given in Figure 5.7(f), does not increase significantly after a final observation time of 12 h. We observe that the composition of the D_{KL} value for the wound assay changes with the final observation time; as the final observation time increases, the D_{KL} value for D decreases and the D_{KL} value for λ increases, while the sum of the two D_{KL} values is approximately constant. This result is consistent with the synthetic wound assay, where we observed that information about D is lost when the two cell fronts began to interact.

We now examine the influence of the number of observations on our estimates of D and λ and the corresponding D_{KL} values. The mean model parameter estimates for all combinations of observation times, for one, two and three observations are presented, with the corresponding mean 90% credible interval for D and λ, in Figures 5.8(a)-(b) and Figures

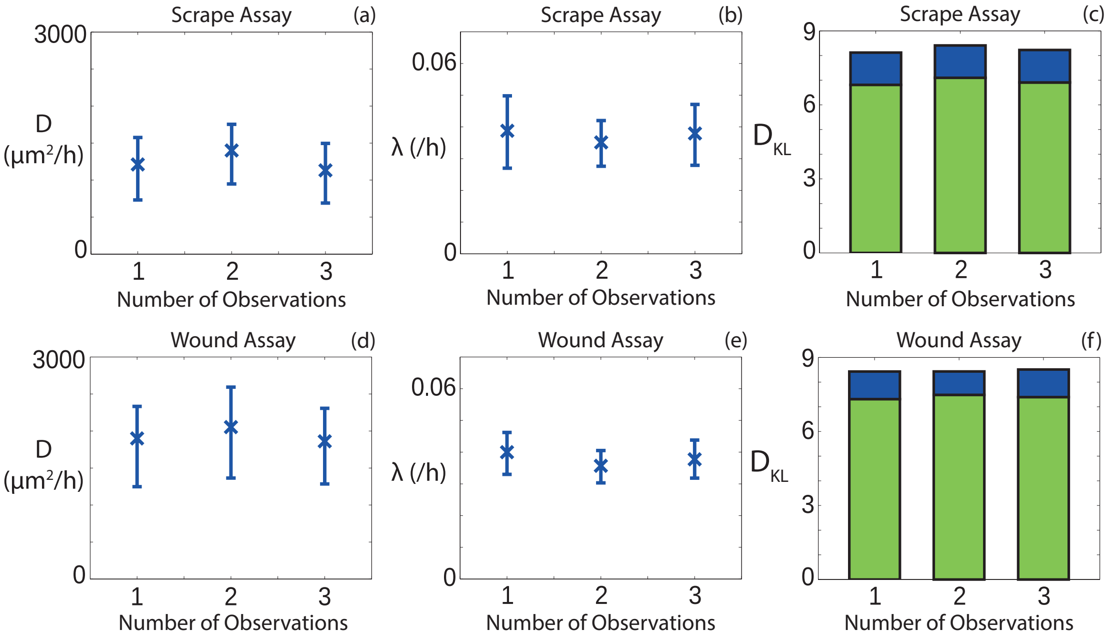

Figure 5.8: Results obtained from the parameter inference approach applied to the experimental data set. (a)-(c) Mean estimates of P_m and P_p with corresponding 90% credible intervals, and the mean D_{KL} values, respectively, for the scrape assay for one, two and three observations. (d)-(f) Mean estimates of P_m and P_p with corresponding 90% credible intervals, and the mean D_{KL} values, respectively, for the wound assay for one, two and three observations. The mean D_{KL} values were calculated using Equation (5.3). The green bar corresponds to the D_{KL} value for λ and the blue bar corresponds to the D_{KL} value for D. For all simulations $\tau = 1/24$ h, $M = 10^6$, $X = 36$, $Y = 49$, $\Delta = 25$ μm. For scrape assays, $N(0) = 155$, $Y_0 = 10$. For wound assays, $N(0) = 226$, $Y_{0,1} = 7, Y_{0,2} = 9$.

5.8(d)-(e) for the scrape and wound assays, respectively. For both model parameters and assays, the parameter estimates are less sensitive to the number of observations than to the final observation time, which is consistent with the synthetic results. It is instructive to consider whether the number of observations influences D_{KL}, to provide guidance about the number of experimental observations that ought to be captured. We present the D_{KL} values for one, two and three observations for the scrape and wound assays in Figures 5.8(c) and (f), respectively. Similar to the synthetic results for the data set in Section 5.3.3, there is no significant increase in information gained between two and three observations, implying that capturing and interpreting additional experimental observations is unnecessary. Given that the interpretation of experimental observations is both time-consuming and expensive, this result provides useful guidance about the number of experimental observations required to estimate D and λ.

For all data sets we observe that it is possible to obtain consistent D_{KL} values with different combinations of observation times, suggesting that various experimental designs are equally informative. However, it is also important to consider the ratio of the D_{KL} values associated with the marginal distributions of P_m and P_p. To illustrate this, we present the approximate posterior distributions of D and λ for the experimental wound assay data set using the combinations of observation times at 6, 12, 18 h, and 12, 18, 24 h

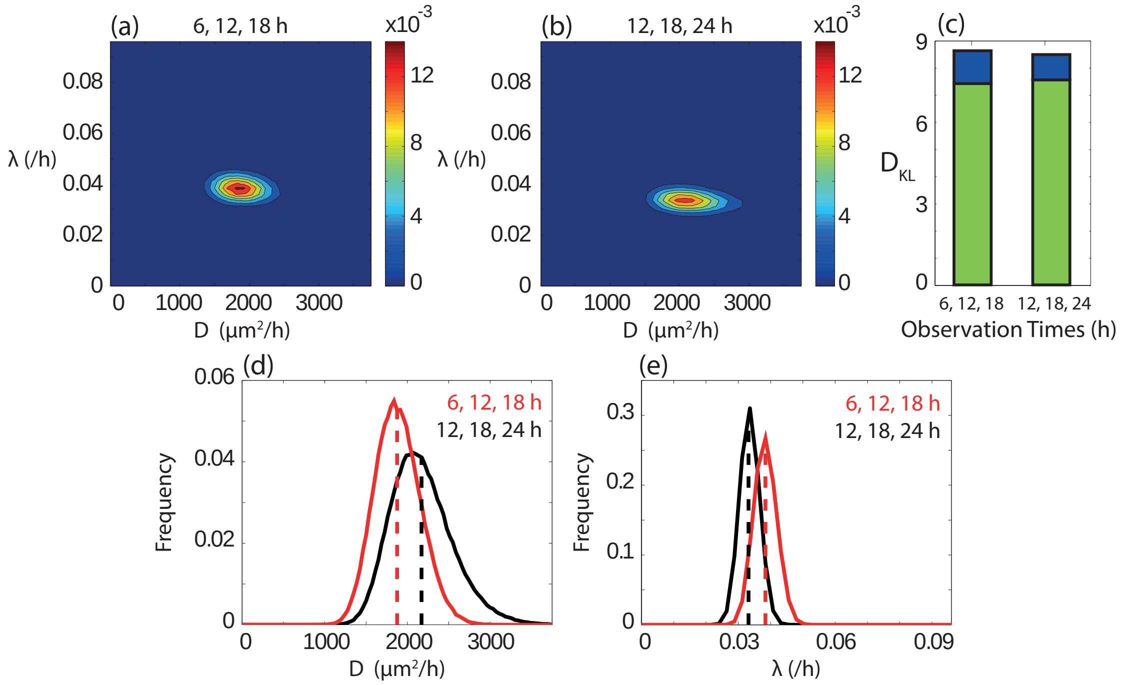

Figure 5.9: Approximate posterior distributions corresponding to different combinations of observation times for the wound assay experimental data set. (a)-(b) Approximate posterior distribution using observations at (a) 6, 12 and 18 h and (b) 12, 18 and 24 h. (c) Mean D_{KL} values for 6, 12 and 18, and for 12, 18 and 24 h. The mean D_{KL} values were calculated using Equation (5.3). The green bar corresponds to the D_{KL} value for λ and the blue bar corresponds to the D_{KL} value for D. (d)-(e) Corresponding marginal distributions for (d) D and (e) λ. Red curves are obtained from observations at 6, 12 and 18 h, while the black curves are obtained from observations at 12, 18 and 24 h. Dashed lines correspond to mean parameter estimates. For all simulations $\tau = 1/24$ h, $M = 10^6$, $X = 36$, $Y = 49$, $N(0) = 226$, $Y_{0,1} = 7, Y_{0,2} = 9$, $\Delta = 25$ μm.

in Figures 5.9(a)-(b), respectively. The D_{KL} values, presented in Figure 5.9(c), for the two posterior distributions are similar: 8.63 and 8.50, respectively, suggesting that both are approximately equally informative. However, if we consider the marginal distributions, we find that the 6, 12, 18 h design results in approximately 30% extra information about D compared to the 12, 18, 24 h design, while retaining 97% of the information about λ. The marginal distributions in Figures 5.9(d)-(e) illustrate the relative information gain for the two designs. We observe that the marginal distribution for D varies significantly, in particular the width of support of the distribution is very different, while the marginal distributions for λ are relatively consistent. The change in distribution for D suggests that the ratio between marginal D_{KL} values can be used to distinguish between posterior distributions that have similar D_{KL} values.

5.4 Summary of results and recommendations

Our results provide guidance about experimental design choices for scratch assays. For both synthetic and experimental data sets we observe that the scrape assay provides estimates of D and λ that are less sensitive to the choice of observation times. For the synthetic data set, where D and λ are known in advance, the scrape assay provides more robust estimates of D and λ. With regard to the information gained about D and λ, defined using the Kullback-Leibler divergence, the scrape assay is superior for the synthetic data set. Conversely, for the experimental data set, the wound assay provides more information about the model parameters than the scrape assay. However, if we consider the D_{KL} values for the marginal distribution of D and λ, we see that the additional information gained in the wound assay is primarily associated with λ. This result is intuitive for the experimental data set, as initially there are significantly more cells in the wound assay than in the scrape assay, in contrast with the synthetic data set where the initial number of cells is the same for both assays. We found that the increase in the initial number of cells results in an increase in computational time of approximately one order of magnitude between the scrape assay and the wound assay. The majority of combinations of observation times for the scrape assay lead to more information about D than the corresponding combinations of observation times for the wound assay. As the D_{KL} values associated with the marginal distribution of D are significantly lower than those associated with the marginal distribution of λ, an increase in information about D is more significant than an equivalent absolute increase in information about λ. Therefore, since the scrape assay provides more robust estimates of D and λ, is less sensitive to the choice of combination of observation times, and is more computationally efficient, we recommend that scrape assays, instead of wound assays, ought to be used to estimate D and λ. We note that these recommendations are based upon experimental observations for a mesenchymal (non-adhesive) cell population and that our recommendations may not be valid for cell populations that involve significant cell-to-cell adhesion.

The uncertainty in our estimates of λ is sensitive to the data obtained at the final observation time. It is intuitive to consider the estimate of λ obtained from a combination of experimental observations that include the latest final observation time, as it is difficult to characterise λ for an experimental time that is significantly less than the cell doubling time, which is approximately 24 h [16]. For the scrape assay, using both experimental and synthetic data sets, we observe that using a final observation time corresponding to time taken for the cell front to cross the experimental field of view provides the most information about D and λ. The choice of the number of observations captured and interpreted affects both the cost and amount of time required to analyse an experiment. Therefore, there is a considerable advantage in minimising the number of observations. We note that the main cost of our approach is associated with the interpretation, rather than the capture, of the experimental images. Therefore, if there is no prior estimate of D, which provides guidance about the final experimental time, images can be captured until the cell front has crossed the experimental field of view. The choice of which images to interpret can then be made after the experiment is performed and the final observation time is determined. For both assays and data sets, we observe that there is, at best, a small increase in information gained by interpreting three experimental observations compared to two. To minimise the expense associated with interpreting experiments, we recommend that two experimental observations are captured, after the initial observation. The time when the first observation is captured, after the initial observation, does not significantly influence the information obtained, compared to the final observation time. As such, we do not provide a recommendation for the time when the first observation should be captured.

5.5 Discussion and conclusions

Scratch assays are widely used to observe collective cell spreading and to examine the influence of potential drugs on the rates of cell motility and proliferation [11, 19, 20, 41]. However, the experimental design of scratch assays reported in the literature varies considerably [1, 5, 16, 17, 25, 26, 30]. The number of experimental observations, the timing at which the observations are captured, and the type of scratch assay are all technically straightforward to vary but there is no explicit discussion in the literature about the influence of these design choices. To the best of our knowledge, there is no biological justification associated with the choice of assay type. Instead, this choice appears to be made according to personal preference. Mathematical models that can be used to estimate D and λ have been presented previously, but these previous applications have not considered the influence of varying experimental designs [5, 16, 17, 28, 40]. To address this limitation, we quantify the information gained about the D and λ, depending on whether a scrape or wound assay is performed. We also investigate the amount of information gained depending on how many experimental observations are captured, and the timing of these observations. To interpret the experimental observations we use an ABC method which provides an approximate bivariate posterior distribution of D and λ [17]. This allows us to quantify the amount of information gained about the model parameters, D and λ, thereby providing guidance about the influence of the experimental design. By identifying experimental designs that allow for more robust parameter estimation, we can make objective recommendations about the number of experimental observations required and, subsequently, reduce the time and cost requirements.

We find that a scrape assay, with just one cell front, provides more robust estimates of the random motility, D, and the cell proliferation rate, λ, is more computationally efficient, requires less time to locate cell positions, and provides more information about D, without sacrificing too much information about λ, compared to the wound assay. Therefore, if the aim is to estimate D and λ, we recommend that scrape assays, and not wound assays, be performed. We find that most information about D and λ is obtained when the final observation time corresponds to the amount of time taken for the front to migrate across the experimental field of view. We also find there is limited benefit to capturing three observations, compared to two.

The work presented here could be extended in several ways. We note that our experimental analysis is relevant for a mesenchymal cell population and that the influence of cell-to-cell or cell-to-substrate adhesion in epithelial cell populations may affect our recommendations. In our mathematical model, we make the standard assumption that both D and λ are constant [16, 17, 33, 34], However, from the results presented in Figure

5.7, we observe that D appears to increase with time. It is possible that the cells are disrupted by the initial scratch or the addition of fresh medium immediately after the scratch is performed. We do not make any suggestion about the putative form of any model parameters that may depend on time or local cell density. It would be instructive to consider an extension to our discrete mathematical model with parameters that are not constant [8, 13] and, subsequently, to investigate whether ABC can result in robust model parameter estimates for a more complicated mathematical model. Such a model may require a different choice of summary statistic and we could employ a semi-automatic approach to determine an appropriate summary statistic [10]. A similar investigation of potential summary statistics for a model that includes parameters that are sensitive to chemical gradients could be performed. However, introducing chemotaxis into the mathematical model increases both the complexity of the model and the number of unknown model parameters. For example, a chemotaxis model would include the diffusivity of the attractant, the production rate of the attractant, the decay rate of the attractant, as well as the parameters governing the chemotactic sensitivity function [18]. We therefore leave this for future work. Alternatively, we could apply the approach outlined in this work to different types of *in vitro* experiments, such as two-dimensional barrier assays or three-dimensional spheroid assays, to quantify the impact of experimental design choices. We make two approximations to compare the cell positions in the experimental images to the mathematical model. First, we assume that cells are incompressible, uniformly-sized disks and, second, we map the cell positions on to a regular lattice with lattice spacing equivalent to the average cell diameter [34]. It would be instructive to relax these assumptions by considering cells that are able to deform or by considering a lattice-free mathematical model. However, both of these approaches would significantly increase the complexity of the mathematical model. Subsequently, the time required to perform the ABC algorithm would become intractable and, as such, we leave this extension for a future study.

Acknowledgements: This work is supported by the Cooperative Research Centre for Wound Management Innovation and the Australian Research Council (FT130100148, FT130100254). We appreciate the support provided by the High Performance Computing Centre at QUT. We thank the three referees for their helpful comments.

Bibliography

[1] Aichele, K., Bubel, M., Deubel, G., Pohlemann, T. and Oberringer, M. [2013], 'Bromelain down-regulates myofibroblast differentiation in an in vitro wound healing assay', *Naunyn-Schmiedeberg's Archives of Pharmacology* **386**, 853–763.

[2] Beaumont, M. A., Zhang, W. and Balding, D. J. [2002], 'Approximate Bayesian computation in population genetics', *Genetics* **162**(4), 2025–2035.

[3] Binder, B. J. and Simpson, M. J. [2013], 'Quantifying spatial structure in experimental observations and agent-based simulations using pair-correlation functions', *Physical Review E* **88**(2), 022705.

[4] Burnham, K. P. and Anderson, D. R. [2002], *Model selection and multimodel inference: a practical information-theoretic approach*, Springer.

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

[6] Chaloner, K. and Verdinelli, I. [1995], 'Bayesian experimental design: A review', *Statistical Science* pp. 273–304.

[7] Codling, E. A., Plank, M. J. and Benhamou, S. [2008], 'Random walk models in biology', *Journal of the Royal Society Interface* **5**(25), 813–834.

[8] Curtis, C. W. and Bortz, D. M. [2012], 'Propagation of fronts in the Fisher-Kolmogorov equation with spatially varying diffusion', *Physical Review E* **86**(6), 066108.

[9] Ermentrout, G. B. and Edelstein-Keshet, L. [1993], 'Cellular automata approaches to biological modeling', *Journal of Theoretical Biology* **160**(1), 97–133.

[10] Fearnhead, P. and Prangle, D. [2012], 'Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation', *Journal of the Royal Statistical Society: Series B* **74**(3), 419–474.

[11] Fishman, D. A., Liu, Y., Ellerbroek, S. M. and Stack, M. S. [2001], 'Lysophosphatidic acid promotes matrix metalloproteinase (MMP) activation and MMP-dependent invasion in ovarian cancer cells', *Cancer Research* **61**(7), 3194–3199.

[12] Habbal, A., Barelli, H. and Malandain, G. [2014], 'Assessing the ability of the 2D Fisher–KPP equation to model cell-sheet wound closure', *Mathematical Biosciences* **252**, 45–59.

[13] Hammond, J. F. and Bortz, D. M. [2011], 'Analytical solutions to Fishers equation with time-variable coefficients', *Applied Mathematics and Computation* **218**(6), 2497–2508.

[14] Jin, W., Shah, E. T., Penington, C. J., McCue, S. W., Chopin, L. K. and Simpson, M. J. [2016], 'Reproducibility of scratch assays is affected by the initial degree of confluence: Experiments, modelling and model selection', *Journal of Theoretical Biology* **390**, 136–145.

[15] Johnston, S. T., Shah, E. T., Chopin, L. K., McElwain, D. S. and Simpson, M. J. [2015], 'Estimating cell diffusivity and cell proliferation rate by interpreting IncuCyte ZOOM assay data using the Fisher-Kolmogorov model', *BMC Systems Biology* **9**(1), 38.

[16] Johnston, S. T., Simpson, M. J. and McElwain, D. L. S. [2014], 'How much information can be obtained from tracking the position of the leading edge in a scratch assay?', *Journal of the Royal Society Interface* **11**(97), 20140325.

[17] Johnston, S. T., Simpson, M. J., McElwain, D. L. S., Binder, B. J. and Ross, J. V. [2014], 'Interpreting scratch assays using pair density dynamics and approximate Bayesian computation', *Open Biology* **4**(9), 140097.

[18] Keller, E. F. and Segel, L. A. [1971], 'Model for chemotaxis', *Journal of Theoretical Biology* **30**(2), 225–234.

[19] Knecht, D. A., LaFleur, R. A., Kahsai, A. W., Argueta, C. E., Beshir, A. B. and Fenteany, G. [2010], 'Cucurbitacin I inhibits cell motility by indirectly interfering with actin dynamics', *PLOS ONE* **5**(11), e14039.

[20] Landesberg, R., Cozin, M., Cremers, S., Woo, V., Kousteni, S., Sinha, S., Garrett-Sinha, L. and Raghavan, S. [2008], 'Inhibition of oral mucosal cell wound healing by bisphosphonates', *Journal of Oral and Maxillofacial Surgery* **66**(5), 839–847.

[21] Liang, C.-C., Park, A. Y. and Guan, J.-L. [2007], 'In vitro scratch assay: a convenient and inexpensive method for analysis of cell migration in vitro', *Nature Protocols* **2**(2), 329–333.

[22] Lukeman, R., Li, Y.-X. and Edelstein-Keshet, L. [2010], 'Inferring individual rules from collective behavior', *Proceedings of the National Academy of Sciences* **107**(28), 12576–12580.

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

[24] Marjoram, P., Molitor, J., Plagnol, V. and Tavaré, S. [2003], 'Markov chain Monte Carlo without likelihoods', *Proceedings of the National Academy of Sciences* **100**(26), 15324–15328.

[25] Meier, F. M., Frommer, K. W., Peters, M. A., Brentano, F., Lefèvre, S., Schröder, D., Kyburz, D., Steinmeyer, J., Rehart, S., Gay, S. et al. [2012], 'Visfatin/pre-b-cell colony-enhancing factor (PBEF), a proinflammatory and cell motility-changing factor in rheumatoid arthritis', *Journal of Biological Chemistry* **287**(34), 28378–28385.

[26] Rausch, V., Liu, L., Apel, A., Rettig, T., Gladkich, J., Labsch, S., Kallifatidis, G., Kaczorowski, A., Groth, A., Gross, W. et al. [2012], 'Autophagy mediates survival of pancreatic tumour-initiating cells in a hypoxic microenvironment', *The Journal of Pathology* **227**(3), 325–335.

[27] Riahi, R., Yang, Y., Zhang, D. D. and Wong, P. K. [2012], 'Advances in wound-healing assays for probing collective cell migration', *Journal of Laboratory Automation* **17**(1), 59–65.

[28] Sengers, B. G., Please, C. P. and Oreffo, R. O. [2007], 'Experimental characterization and computational modelling of two-dimensional cell spreading for skeletal regeneration', *Journal of the Royal Society Interface* **4**(17), 1107–1117.

[29] Shatkin, A., Reich, E., Franklin, R. and Tatum, E.-L. [1962], 'Effect of mitomycin-c on mammalian cells in culture', *Biochimica et Biophysica Acta* **55**(3), 277–289.

[30] Shibata, A., Tanabe, E., Inoue, S., Kitayoshi, M., Okimoto, S., Hirane, M., Araki, M., Fukushima, N. and Tsujiuchi, T. [2013], 'Hydrogen peroxide stimulates cell motile activity through lpa receptor-3 in liver epithelial wb-f344 cells', *Biochemical and Biophysical Research Communications* **433**, 317–321.

[31] Simpson, M. J., Landman, K. A. and Hughes, B. D. [2009], 'Multi-species simple exclusion processes', *Physica A: Statistical Mechanics and its Applications* **388**(4), 399–406.

[32] Simpson, M. J., Landman, K. A. and Hughes, B. D. [2010], 'Cell invasion with proliferation mechanisms motivated by time-lapse data', *Physica A: Statistical Mechanics and its Applications* **389**(18), 3779–3790.

[33] Simpson, M. J., Landman, K. A., Hughes, B. D. and Fernando, A. E. [2010], 'A model for mesoscale patterns in motile populations', *Physica A: Statistical Mechanics and its Applications* **389**(7), 1412–1424.

[34] Simpson, M. J., Treloar, K. K., Binder, B. J., Haridas, P., Manton, K. J., Leavesley, D. I., McElwain, D. L. S. and Baker, R. E. [2013], 'Quantifying the roles of cell motility and cell proliferation in a circular barrier assay', *Journal of the Royal Society Interface* **10**(82), 20130007.

[35] Sisson, S. A., Fan, Y. and Tanaka, M. M. [2007], 'Sequential Monte Carlo without likelihoods', *Proceedings of the National Academy of Sciences* **104**(6), 1760–1765.

[36] Todaro, G. J. and Green, H. [1963], 'Quantitative studies of the growth of mouse embryo cells in culture and their development into established lines', *The Journal of Cell Biology* **17**(2), 299–313.

[37] Toni, T., Welch, D., Strelkowa, N., Ipsen, A. and Stumpf, M. P. [2009], 'Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems', *Journal of the Royal Society Interface* **6**(31), 187–202.

[38] Treloar, K. K., Simpson, M. J., Haridas, P., Manton, K. J., Leavesley, D. I., McElwain, D. L. S. and Baker, R. E. [2013], 'Multiple types of data are required to identify the mechanisms influencing the spatial expansion of melanoma cell colonies', *BMC Systems Biology* **7**(1), 137.

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

[40] Tremel, A., Cai, A., Tirtaatmadja, N., Hughes, B. D., Stevens, G. W., Landman, K. A. and O'Connor, A. J. [2009], 'Cell migration and proliferation during monolayer formation and wound healing', *Chemical Engineering Science* **64**(2), 247–253.

[41] Upton, Z., Wallace, H. J., Shooter, G. K., van Lonkhuyzen, D. R., Yeoh-Ellerton, S., Rayment, E. A., Fleming, J. M., Broszczak, D., Queen, D., Sibbald, R. G. et al. [2011], 'Human pilot studies reveal the potential of a vitronectin: growth factor complex as a treatment for chronic wounds', *International Wound Journal* **8**(5), 522–532.

[42] Vo, B. N., Drovandi, C. C., Pettitt, A. N. and Simpson, M. J. [2015], 'Quantifying uncertainty in parameter estimates for stochastic models of collective cell spreading using approximate Bayesian computation', *Mathematical Biosciences* **263**, 133–142.

5.6 Supplementary material

5.6.1 Approximate Bayesian computation

To investigate which values of D and λ can generate summary statistics that are consistent with those resulting from our synthetic or experimental data set we consider an ABC algorithm [2, 16, 22, 32, 34]. The ABC algorithm performs stochastic identically-prepared realisations of the mathematical framework described in Section 5.2.2 of the main document and uses a combination of $N(t)$ and $q(i)$ as a summary statistic to determine whether a particular model parameter pair, $\underline{\theta}$, generates simulation data that matches the synthetic or experimental data set to a specified tolerance, ϵ [16]. We define a measure of the difference between the summary statistics of two data sets, β and β',

$$||S(\beta) - S(\beta')|| = \frac{1}{Q}\left[\sum_{i=1}^{Q}\left(\frac{S(\beta)_i - S(\beta')_i}{1 + S(\beta)_i}\right)^2\right]^{\frac{1}{2}}, \tag{5.4}$$

where $S(z)$ is an arbitrary summary statistic, $S(\beta)_i$ is the i^{th} element in $S(\beta)$ and Q is the number of elements in $S(\beta)$. We define an overall measure of the difference between the two data sets as

$$L(\beta,\beta') = \frac{1}{2M}\sum_{j=1}^{M}\Big[||S_{\text{pcf}}(\beta) - S_{\text{pcf}}(\beta')||_j\Big] + \frac{1}{2M}\sum_{j=1}^{M}\Big[||S_{\text{N}}(\beta) - S_{\text{N}}(\beta')||_j\Big], \tag{5.5}$$

where M is the number of experimental observations, S_{pcf} is the summary statistic for the pair correlation function, S_{N} is the summary statistic for the number of cells and the subscript j corresponds to an individual experimental observation. For a model parameter pair to be accepted we require that $L(\beta,\beta') < \epsilon$. To reduce computation time we consider an ABC Markov Chain Monte Carlo (ABC-MCMC) algorithm [16, 22, 32], which proposes model parameter values based on previously successful model parameter values. Making minimal assumptions, we initially sample $\underline{\theta}$ from a bivariate uniform distribution, $P_m \in [0,1]$, $P_p \in [0,1]$, until we obtain the first model parameter pair that results in $L(\beta,\beta') < \epsilon$. The ABC-MCMC algorithm then uses a transition kernel $\underline{\Gamma}$ to propose $\underline{\theta}'$ values, where $\underline{\theta}'$ is uniformly sampled from $\underline{\theta} \pm \underline{\Gamma}$ and $\underline{\Gamma} = (10^{-1}, 10^{-3})$. We note that $\underline{\Gamma}$ is truncated, if necessary, such that $P_m \in [0,1]$, $P_p \in [0,1]$. Successful $\underline{\theta}$ values are recorded and are used to generate an approximation to the posterior distribution. For full details of the ABC-MCMC algorithm see [16, 22].

5.6.2 Data

We present results obtained from the ABC algorithm applied to: (i) synthetic scrape assay data with $N(0) = 100$; (ii) synthetic scrape assay data with $N(0) = 50$; (iii) synthetic wound assay data; (iv) experimental scrape assay data, and (v) experimental wound assay data. Each table contains the D_{KL} value for the bivariate posterior distribution of P_m and

P_p, or D and λ, for the synthetic and experimental data sets, respectively. Additionally, the tables contain the mean and 90% credible interval for the recovered estimates of P_m and P_p, or D and λ, for the synthetic and experimental data sets, respectively. Finally, the tables contain the D_{KL} values associated with the univariate distribution for both P_m and P_p, or D and λ, for the synthetic and experimental data sets, respectively.

M	Observation Times (h)	D_{KL}	P_m (90% CI)	$D_{KL}(P_m)$	P_p (90% CI) (10^{-3})	$D_{KL}(P_p)$
1	3	6.47	0.30 (0-0.52)	0.51	2.2 (0.9-3.1)	5.94
1	6	7.21	0.27 (0.08-0.42)	0.87	2.0 (1.2-2.6)	6.32
1	9	7.47	0.25 (0.03-0.39)	0.86	2.0 (1.3-2.5)	6.58
1	12	7.85	0.23 (0.07-0.35)	1.05	1.9 (1.4-2.4)	**6.78**
2	3, 12	7.79	0.23 (0.05-0.35)	1.03	1.9 (1.4-2.4)	6.75
2	6, 12	7.95	0.24 (0.09-0.35)	1.16	1.9 (1.4-2.4)	6.77
2	8, 12	8.00	0.23 (0.10-0.34)	1.22	1.9 (1.4-2.4)	**6.78**
2	10, 12	8.05	0.23 (0.10-0.34)	1.25	1.9 (1.4-2.4)	**6.78**
3	1, 2, 3	6.69	0.27 (0-0.45)	0.74	2.1 (0.6-3.0)	5.93
3	2, 4, 6	7.33	0.26 (0.05-0.39)	0.97	2.0 (1.3-2.7)	6.32
3	3, 6, 9	7.61	0.24 (0.03-0.37)	1.04	2.0 (1.3-2.5)	6.54
3	4, 8, 12	8.08	0.23 (0.10-0.32)	1.30	1.9 (1.4-2.4)	6.77
3	8, 10, 12	**8.11**	0.23 (0.11-0.33)	**1.32**	1.9 (1.4-2.4)	**6.78**
3	6, 9, 12	8.10	0.23 (0.10-0.32)	**1.32**	1.9 (1.4-2.4)	**6.78**
3	2, 9, 12	8.02	0.23 (0.09-0.33)	1.25	1.9 (1.4-2.4)	6.77
3	2, 4, 12	7.93	0.23 (0.09-0.33)	1.20	2.0 (1.4-2.4)	6.72

Table 5.2: Comparison of parameter recovery for ten averaged sets of synthetic scrape assay data with $N(0) = 100$. Synthetic data was identically prepared with $P_m = 0.25$ and $P_p = 2 \times 10^{-3}$.

M	Observation Times (h)	D_{KL}	P_m (90% CI)	$D_{KL}(P_m)$	P_p (90% CI) (10^{-3})	$D_{KL}(P_p)$
1	3	6.46	0.26 (0.02-0.46)	0.62	2.0 (0.7-3.5)	5.67
1	6	6.85	0.25 (0.01-0.41)	0.75	1.8 (0.7-2.7)	6.06
1	9	7.22	0.23 (0.00-0.36)	0.98	1.8 (0.6-2.6)	6.21
1	12	7.45	0.23 (0.04-0.36)	0.96	1.8 (0.8-2.4)	6.42
2	3, 12	7.54	0.22 (0.07-0.35)	1.08	1.8 (1.0-2.4)	6.42
2	6, 12	7.71	0.21 (0.06-0.32)	1.14	1.7 (0.9-2.3)	6.51
2	8, 12	7.56	0.21 (0.02-0.34)	1.06	1.8 (0.8-2.4)	6.45
2	10, 12	7.60	0.22 (0.00-0.34)	1.10	1.8 (1.0-2.4)	6.47
3	1, 2, 3	6.71	0.23 (0.00-0.40)	0.73	1.4 (0.0-2.6)	5.93
3	2, 4, 6	6.97	0.22 (0.00-0.35)	0.94	1.8 (0.1-2.7)	5.97
3	3, 6, 9	7.31	0.22 (0.03-0.35)	1.02	1.7 (0.6-2.4)	6.26
3	4, 8, 12	7.66	0.20 (0.10-0.31)	1.16	1.8 (1.0-2.4)	6.45
3	8, 10, 12	**7.76**	0.20 (0.10-0.31)	1.17	1.7 (1.0-2.2)	**6.53**
3	6, 9, 12	7.69	0.21 (0.04-0.32)	1.17	1.8 (0.9-2.3)	6.47
3	2, 9, 12	7.59	0.22 (0.05-0.33)	1.13	1.8 (0.8-2.4)	6.42
3	2, 4, 12	7.66	0.22 (0.07-0.33)	**1.20**	1.8 (1.0-2.4)	6.42

Table 5.3: Comparison of parameter recovery for ten averaged sets of synthetic scrape assay data with $N(0) = 50$. Synthetic data was identically prepared with $P_m = 0.25$ and $P_p = 2 \times 10^{-3}$.

M	Observation Times (h)	D_{KL}	P_m (90% CI)	$D_{KL}(P_m)$	P_p (90% CI) (10^{-3})	$D_{KL}(P_p)$
1	3	6.75	0.23 (0.03-0.35)	1.05	2.3 (1.0-4.0)	5.67
1	6	7.34	0.24 (0.01-0.37)	0.98	2.0 (1.1-2.7)	6.29
1	9	7.53	0.29 (0.05-0.43)	0.93	1.9 (1.2-2.4)	6.59
1	12	7.30	0.32 (0-0.49)	0.77	1.9 (1.4-2.6)	6.51
2	3, 6	7.32	0.24 (0.04-0.36)	1.06	2.1 (1.3-2.9)	6.22
2	2, 9	7.82	0.25 (0.06-0.36)	1.18	1.9 (1.2-2.4)	6.60
2	4, 9	**7.90**	0.24 (0.10-0.34)	1.28	1.9 (1.2-2.4)	6.59
2	6, 9	7.75	0.25 (0.06-0.36)	1.15	1.9 (1.2-2.4)	6.57
2	3, 12	7.75	0.26 (0.10-0.36)	1.20	1.9 (1.4-2.6)	6.53
2	6, 12	7.71	0.26 (0.09-0.37)	1.15	2.0 (1.4-2.6)	6.51
2	8, 12	7.66	0.28 (0.06-0.40)	1.05	1.9 (1.4-2.6)	6.55
2	10, 12	7.45	0.30 (0.04-0.44)	0.91	2.0 (1.4-2.6)	6.53
3	1, 2, 3	7.13	0.20 (0.09-0.29)	**1.38**	1.8 (0-3.1)	5.71
3	2, 4, 6	7.55	0.21 (0.02-0.32)	1.19	2.0 (1.0-2.6)	6.23
3	3, 6, 9	7.72	0.25 (0.08-0.36)	1.17	1.9 (1.3-2.5)	6.53
3	4, 8, 12	7.80	0.25 (0.09-0.35)	1.21	1.9 (1.3-2.5)	6.52
3	8, 10, 12	7.66	0.26 (0.08-0.38)	1.05	1.9 (1.3-2.5)	6.60
3	6, 9, 12	7.44	0.25 (0.03-0.39)	0.87	2.0 (1.3-2.5)	6.55
3	2, 9, 12	7.77	0.28 (0.10-0.38)	1.14	1.9 (1.4-2.4)	**6.61**
3	2, 4, 12	7.52	0.25 (0.07-0.37)	1.06	2.0 (1.4-2.6)	6.44

Table 5.4: Comparison of parameter recovery for ten averaged sets of synthetic wound assay data. Synthetic data was identically prepared with $P_m = 0.25$ and $P_p = 2 \times 10^{-3}$.

M	Observation Times (h)	D_{KL}	D (90% CI) (μm^2/h)	$D_{KL}(D)$	λ (90% CI) (10^{-2}/h)	$D_{KL}(\lambda)$
1	3	7.78	970 (563-1313)	1.36	4.37 (2.64-6.00)	6.41
1	6	8.27	960 (563-1238)	1.51	4.04 (2.88-5.28)	6.76
1	9	8.18	1310 (713-1688)	1.21	3.63 (2.64-4.56)	6.97
1	12	8.27	1605 (1088-2063)	1.15	3.47 (2.64-4.08)	**7.11**
2	3, 12	8.39	1337 (863-1688)	1.29	3.54 (2.88-4.32)	7.09
2	6, 12	**8.47**	1258 (863-1613)	1.37	3.56 (2.88-4.32)	7.09
2	8, 12	8.38	1498 (1050-1875)	1.28	3.49 (2.64-4.08)	7.10
2	10, 12	8.41	1507 (1013-1838)	1.30	3.46 (2.64-4.08)	**7.11**
3	1, 2, 3	7.61	818 (300-1200)	1.27	4.64(2.88-6.24)	6.34
3	2, 4, 6	8.22	811 (450-1125)	**1.52**	4.24 (2.88-5.28)	6.69
3	3, 6, 9	8.45	1079 (713-1388)	1.22	3.69(2.64-4.56)	6.94
3	4, 8, 12	8.43	1301 (938-1688)	1.35	3.54 (2.88-4.32)	7.07
3	8, 10, 12	8.37	1463 (938-1838)	1.29	3.44 (2.64-4.08)	7.08
3	6, 9, 12	8.37	1243 (788-1613)	1.31	3.57 (2.88-4.32)	7.05
3	2, 9, 12	8.42	1286 (825-1650)	1.34	3.55 (2.88-4.32)	7.07
3	2, 4, 12	8.27	1051 (563-1463)	1.23	3.68 (2.64-4.56)	7.03

Table 5.5: Comparison of parameter recovery for experimental scrape assay data.

M	Observation Times (h)	D_{KL}	D (90% CI) (μm^2/h)	$D_{KL}(D)$	λ (90% CI) (10^{-2}/h)	$D_{KL}(\lambda)$
1	6	8.31	1245 (863-1613)	**1.44**	4.85 (3.84-5.76)	6.87
1	12	8.63	1837 (1388-2138)	1.40	4.29 (3.60-5.04)	7.22
1	18	8.30	2265 (1350-2850)	0.82	3.71 (3.12-4.08)	7.46
1	24	8.53	2253 (1388-2.813)	0.82	3.14 (2.64-3.60)	**7.69**
2	6, 12	8.51	1676 (1238-2063)	1.30	4.33 (3.60-5.04)	7.20
2	4, 18	8.48	1881 (1313-2363)	1.03	3.78 (3.36-4.32)	7.42
2	8, 18	8.54	1940 (1350-2400)	1.09	3.78 (3.36-4.32)	7.43
2	12, 18	8.38	2126 (1463-2663)	0.95	3.78 (3.36-4.32)	7.42
2	6, 24	8.33	1876 (1088-2513)	0.76	3.24 (2.64-3.60)	7.54
2	12, 24	8.47	2145 (1275-2700)	0.87	3.24 (2.64-3.60)	7.59
2	16, 24	8.51	2163 (1425-2700)	0.88	3.25 (2.64-3.60)	7.61
2	20, 24	8.40	2609 (1763-3338)	0.71	3.15 (2.64-3.60)	7.66
3	2, 4, 6	7.90	1179 (638-1613)	1.13	5.01 (3.84-6.24)	6.76
3	4, 8, 12	8.41	1655 (1163-2063)	1.24	4.39 (3.60-5.04)	7.16
3	6, 12, 18	8.63	1884 (1388-2288)	1.21	3.83 (3.36-4.32)	7.41
3	8, 16, 24	8.63	1884 (1388-2288)	1.21	3.83 (3.36-4.32)	7.41
3	16, 20, 24	8.55	2367 (1650-2925)	0.90	3.23 (2.64-3.60)	7.64
3	12, 18, 24	8.50	2175 (1425-2700)	0.93	3.35 (2.88-3.84)	7.55
3	4, 18, 24	**8.83**	1994 (1425-2400)	1.20	3.28 (2.88-3.84)	7.62
3	4, 8, 24	8.74	1745 (1200-2175)	1.14	3.27 (2.88-3.84)	7.58

Table 5.6: Comparison of parameter recovery for experimental wound assay data.

Part II

Novel continuum models of collective behaviour

Chapter 6

Modelling the movement of interacting cell populations: a moment dynamics approach

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

Johnston, S. T., Simpson, M. J. and Baker, R. E. [2015], 'Modelling the movement of interacting cell populations: A moment dynamics approach', *Journal of Theoretical Biology* **370**, 81-92.

Abstract

Mathematical models describing the movement of multiple interacting subpopulations are relevant to many biological and ecological processes. Standard mean-field partial differential equation descriptions of these processes suffer from the limitation that they implicitly neglect to incorporate the impact of spatial correlations and clustering. To overcome this, we derive a moment dynamics description of a discrete stochastic process which describes the spreading of distinct interacting subpopulations. In particular, we motivate our model by mimicking the geometry of two typical cell biology experiments. Comparing the performance of the moment dynamics model with a traditional mean-field model confirms that the moment dynamics approach always outperforms the traditional mean-field approach. To provide more general insight we summarise the performance of the moment dynamics model and the traditional mean-field model over a wide range of parameter regimes. These results help distinguish between those situations where spatial correlation effects are sufficiently strong, such that a moment dynamics model is required, from other situations where spatial correlation effects are sufficiently weak, such that a traditional mean-field model is adequate.

6.1 Introduction

Biological and ecological processes often involve moving fronts of interacting subpopulations. For example, in a biological setting, malignant spreading occurs when tumour cells interact with, and move through, the stroma [3, 7, 10, 18]. In an ecological setting, the spreading of an invasive species involves moving fronts, that, in some cases, is coupled with a retreating front of that species' prey [13, 24, 38].

Figure 1 shows images of two different types of cell biology experiments involving moving fronts of interacting subpopulations. Figure 1(a)-(c) shows images of a co-culture scratch assay [22]. This assay is constructed such that initially we have two subpopulations present in a certain region of the domain that is adjacent to a vacant region. As time proceeds, the two subpopulations spread into the vacant space. The image in Figure 1(c) indicates that one of the subpopulations is clustered, whereas the other subpopulation is more evenly distributed. The image in Figure 1(d) shows a subpopulation of initially confined melanoma cells that are spreading into a surrounding subpopulation of fibroblast cells [18]. These images demonstrate that collective cell spreading processes can involve moving fronts of interacting subpopulations. Given the importance of collective cell spreading processes to a range of biological applications, including wound healing and malignant spreading, it is relevant for us to develop robust mathematical and computational tools that can accurately describe the motion of these kinds of multispecies moving front problems.

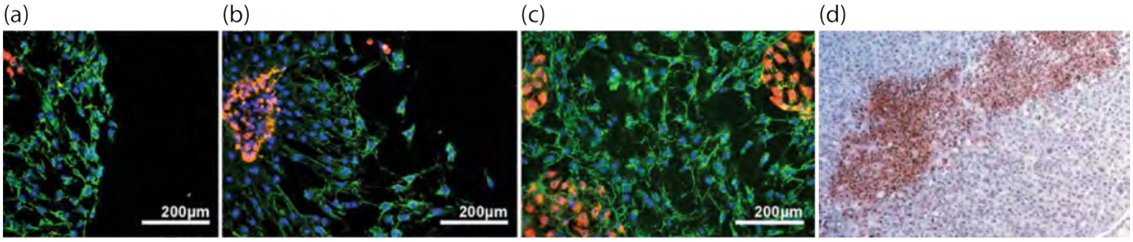

Figure 6.1: (a)-(c) Co-culture scratch assay containing human dermal microvascular endothelial cells (red) and human dermal fibroblasts (green) at (a) 0 hours, (b) 24 hours and (c) 48 hours. Adapted from [22]. (d) Human fibroblasts (blue) and TGF-β1 transduced 451Lu melanoma cells (brown), 19 days after subcutaneous injection into immunodeficient mice. Adapted from [18].

Previous mathematical modelling of problems involving moving fronts of multiple interacting subpopulations have typically involved studying systems of reaction-diffusion partial differential equations (PDEs) [9, 23, 27, 31, 35, 39]. For example, Sherratt [27] considers a two-species model of tumour growth. In this model, the movement of the tumour cell subpopulation, $v(x,t)$, is inhibited by the stroma subpopulation, $u(x,t)$. Cell proliferation is also influenced by crowding, since the rate of proliferation is a decreasing function of the total cell density, $u(x,t) + v(x,t)$ [27]. More generally, Painter and Sherratt [23] suggest that the motion of interacting cell subpopulations depends on the gradient of each particular species' density, as well as the gradient of the total cell density. Focusing specifically on tumour invasion, Gatenby and Gawlinski [9] propose a three-species model, where the density of normal tissue decreases due to an excess concentration of H^+ ions. Smallbone *et al.* [39] extend the Gatenby and Gawlinski three-species model by including a necrotic core within the tumour, which is more consistent with biological observations. However, while these models provide valuable insight into the interaction of multiple cell subpopulations, they are limited in two ways. First, each of these PDE models relies on invoking a mean-field assumption. That is, these models implicitly assume that individuals in an underlying stochastic process interact at a rate that is proportional to the average density [12]. This assumption amounts to the neglect of any spatial structure present in the subpopulations [17]. Second, these PDE models describe population-level behaviour, and do not explicitly consider individual-level information that could be relevant when dealing with certain types of experimental data [34].

Instead of working directly with PDEs, mean-field descriptions of collective cell behaviour have been derived from discrete individual-level models [4, 6, 8, 15, 32, 33]. These discrete models, which can also incorporate crowding [5], can be identified with corresponding mean-field continuum PDE models that aim to describe the average behaviour of the underlying stochastic process. Using this kind of approach gives us access to both discrete individual-level information as well as continuum population-level information. For example, to model the migration of adhesive glioma cells, Khain *et al.* [15] derive a mean-field PDE description of a discrete process which incorporates cell motility, cell-to-cell adhesion and cell proliferation. However, while the relationship between the averaged discrete data and the solution of the corresponding mean-field PDE description is useful in certain circumstances, it is well-known that the assumptions invoked when deriving mean-field PDE descriptions are inappropriate in certain parameter regimes, due to spatial correlations between the occupancy of lattice sites [1, 14, 29]. The impact of spatial correlation is relevant when we consider patchy or clustered distributions of cells, such as in Figures 1(b)-(c). Baker and Simpson [1] partly address this issue by developing a moment dynamics model that approximately incorporates the effect of spatial correlation. Markham *et al.* [19] extend this work, but focus on problems where the initial distribution of cells is spatially uniform, meaning that the modelling and computational tools developed by Markham *et al.* [19] are not suitable for studying the motion of moving fronts of various interacting subpopulations.

In this work we consider a discrete lattice-based model for describing the motion of a population of cells where the total population is composed of distinct, interacting subpopulations. To understand how our work builds on previous methods of analysis, we derive a standard mean-field description of the discrete model and demonstrate that, in certain parameter regimes, the mean-field model does not describe the averaged discrete behaviour. By considering the dynamics of the occupancy of lattice pairs, we derive one- and two-dimensional moment dynamics descriptions that incorporate an approximate description of the spatial correlation present in the system. Motivated by the geometry of the two typical cell biology experiments in Figure 1, we apply our model to two case studies. The first case study is relevant to co-culture scratch assays and the second case study is relevant to the invasion of one subpopulation into another subpopulation, thereby mimicking tumour invasion processes. Through these case studies we demonstrate that our moment dynamics model provides a significantly more accurate description of the averaged discrete model behaviour. Finally, we discuss our results and outline directions for future work.

6.2 Methods

6.2.1 Discrete model

We consider a lattice-based random walk model where each lattice site may be occupied by, at most, one agent [5]. The model is presented for situations where there are two subpopulations, denoted by superscripts G and B, and we note that the framework could be extended to include a larger number of subpopulations if required. The superscripts G and B correspond to the colour scheme in our figures where results relating to the G subpopulation are given in green and results relating to the B subpopulation are given in blue. The discrete process takes place on a one-dimensional lattice, with lattice spacing Δ, where each site is indexed $i \in [1, X]$. Agents on the lattice undergo movement, proliferation and death events at rates P_m^G, P_p^G, P_d^G and P_m^B, P_p^B, P_d^B per unit time, for subpopulations G and B, respectively. During a potential motility event, an agent at site i attempts to move to site $i \pm 1$, with the target site chosen with equal probability. This potential event will be successful only if the target site is vacant. A proliferative agent at site i attempts to place a daughter agent at site $i \pm 1$, with the target site chosen with equal probability. This event will only be successful if the target site is vacant. Agent death occurs by simply removing an agent from the lattice. For all results presented in this work, we apply periodic boundary conditions. However, in practice, we only consider initial conditions and timescales such that the effects of the boundary conditions at $i = 1$ and $i = X$ are unimportant.

For the two-dimensional discrete model, we define a square two-dimensional lattice, with lattice spacing Δ, where each lattice site is indexed (i, j), where $i \in [1, X]$ and $j \in [1, Y]$. A motile agent at (i, j) will attempt to step to site $(i \pm 1, j)$ or $(i, j \pm 1)$, with the target site chosen with equal probability. Similarly, a proliferative agent at (i, j) will attempt to deposit a daughter agent at site $(i \pm 1, j)$ or $(i, j \pm 1)$, with the target site chosen with equal probability. Since the model is an exclusion process, any potential motility or proliferation event that would place an agent on an occupied site is aborted. Agent death occurs by removing an agent from the lattice. While we do not explicitly consider extending this model to a three-dimensional lattice, it is straightforward to perform discrete simulations on a three dimensional lattice [1].

We use the Gillespie algorithm to generate sample paths from the discrete model [11]. An individual realisation of the Gillespie algorithm results in the binary lattice occupancy, $\underline{C}_i^k$, at each site i. To obtain averaged density information we perform M identically-prepared realisations of the discrete algorithm and calculate the average lattice occupancy $C_i^k = \langle \underline{C}_i^k \rangle$, which represents the probability that lattice site i is occupied by an agent of subpopulation $k \in \{G, B\}$.

6.2.2 One-dimensional mean-field approximation

To derive a mean-field description of the discrete model we consider a discrete conservation statement describing the rate of change of the occupancy status of site i. Accounting for all possible motility, proliferation and death events we obtain

$$\frac{\mathrm{d}C_i^k}{\mathrm{d}t} = \frac{P_m^k}{2}\left[C_{i-1}^k\Phi_i + C_{i+1}^k\Phi_i - C_i^k\Phi_{i-1} - C_i^k\Phi_{i+1}\right] + \frac{P_p^k}{2}\left[C_{i-1}^k\Phi_i + C_{i+1}^k\Phi_i\right] - P_d^k C_i^k, \tag{6.1}$$

for subpopulation $k \in \{G, B\}$, where $\Phi_i = 1 - \sum_K C_i^K$ is the probability that site i is vacant. Since we interpret the product of site occupation probabilities in Equation (6.1) as a net transition probability [14], we explicitly assume that the occupancy status of lattice sites are independent, which is equivalent to neglecting the correlations in occupancy between lattice sites. Extending this kind of mean-field conservation statement to apply to our two-dimensional discrete model is straightforward, and the details are given in the supplementary material document. Standard mean-field descriptions of our discrete model, given by Equation (6.1), can be re-written as a PDE description. To see this we expand the $C_{i\pm1}^k$ terms in Equation (6.1) in a Taylor series about site i, neglecting terms of $\mathcal{O}(\Delta^3)$ and smaller. After identifying C_i^k with a continuous function $C^k(x,t)$, we can re-write the resulting expression as a reaction-diffusion PDE for $C^k(x,t)$ [30].

6.2.3 One-dimensional moment dynamics approximation

Instead of treating products of site occupation probabilities as independent quantities, we now consider the time evolution of the relevant n-point distribution functions, $\rho^{(n)}$ [1]. The one-point distribution function is given by $\rho^{(1)}(\sigma_i)$, where σ_i denotes the state of site i and can be interpreted as the probability that site i is in state $\sigma \in \{0, A^G, A^B\}$. We note that the possible states of site i are: (i) A_i^G, which indicates that site i is occupied by an agent from subpopulation G, (ii) A_i^B, which indicates that site i is occupied by an agent from subpopulation B, and (iii) 0_i, which indicates that site i is vacant [1]. The evolution of the one-point distribution function for subpopulation k can be described by accounting for all possible motility, proliferation and death events,

$$\begin{aligned}\frac{\mathrm{d}\rho^{(1)}(A_i^k)}{\mathrm{d}t} =& \frac{P_m^k}{2}\left[\rho^{(2)}(A_{i-1}^k, 0_i) + \rho^{(2)}(A_{i+1}^k, 0_i) - \rho^{(2)}(A_i^k, 0_{i-1}) - \rho^{(2)}(A_i^k, 0_{i+1})\right] + \\ & \frac{P_p^k}{2}\left[\rho^{(2)}(A_{i-1}^k, 0_i) + \rho^{(2)}(A_{i+1}^k, 0_i)\right] - P_d^k \rho^{(1)}(A_i^k).\end{aligned} \tag{6.2}$$

The evolution of the one-point distribution functions depends on the two-point distribution functions, which, in this case, means that the evolution of the occupancy status of individual lattice sites depends on the occupancy of nearest-neighbour lattice pairs. For example, the average occupancy of site i increases due to the probability that site i is unoccupied and site $i-1$ is occupied by subpopulation k. We denote this probability, without the assumption that the occupancies of sites i and $i-1$ are uncorrelated, by

$\rho^{(2)}(A^k_{i-1}, 0_i)$. To measure the correlation between lattice sites i and m, separated by distance $r\Delta = (m-i)\Delta$, we use the correlation function [1]

$$F_i^{a,b}(r\Delta) = \frac{\rho^{(2)}(\sigma_i, \sigma_m)}{\rho^{(1)}(\sigma_i)\rho^{(1)}(\sigma_m)}, \tag{6.3}$$

where a denotes the state of site i and b denotes the state of site m. We note that $F_i^{a,b}(r\Delta)$ depends on time. However, for notational convenience, we do not explicitly include this dependence in our notation. Employing the relationship [1]

$$\rho^{(1)}(\sigma_i) = \sum_{\sigma_m} \rho^{(2)}(\sigma_i, \sigma_m), \tag{6.4}$$

we rewrite Equation (6.2) in terms of the correlation functions. Here, for the specific case where we consider two subpopulations, G and B, we obtain

$$\begin{aligned}\frac{\mathrm{d}C_i^G}{\mathrm{d}t} =& \frac{P_m^G}{2}\Bigg[C_{i-1}^G + C_{i+1}^G - 2C_i^G + C_i^B\left\{2C_i^G - C_{i-1}^G F_{i-1}^{G,B}(\Delta) - C_{i+1}^G F_i^{B,G}(\Delta)\right\} \\ &- C_i^G\left\{2C_i^B - C_{i-1}^B F_{i-1}^{B,G}(\Delta) - C_{i+1}^B F_i^{G,B}(\Delta)\right\}\Bigg] \\ &+ \frac{P_p^G}{2}\Bigg[C_{i-1}^G\left\{1 - C_i^G F_{i-1}^{G,G}(\Delta) - C_i^B F_{i-1}^{G,B}(\Delta)\right\} \\ &+ C_{i+1}^G\left\{1 - C_i^G F_i^{G,G}(\Delta) - C_i^B F_i^{B,G}(\Delta)\right\}\Bigg] - P_d^G C_i^G.\end{aligned} \tag{6.5}$$

Note that if the lattice sites are uncorrelated and hence $F_i^{a,b}(r\Delta) \equiv 1$, Equation (6.5) is equivalent to Equation (6.1). This simplification emphasises that the key difference between the moment dynamics description and the standard mean-field description is in the way that the two approaches deal with the role of spatial correlation effects. We also note interchanging G and B in Equation (6.5) allows us to write down a similar expression for $\mathrm{d}C_i^B/\mathrm{d}t$.

To solve Equation (6.5) and the corresponding expression for $\mathrm{d}C_i^B/\mathrm{d}t$, we must develop a model for the evolution of $F_i^{G,G}(\Delta)$, $F_i^{B,B}(\Delta)$, $F_i^{G,B}(\Delta)$ and $F_i^{B,G}(\Delta)$. To achieve this we consider the evolution of the relevant two-point distribution functions by considering how potential motility, proliferation and death events alter each two-point distribution function. Here we present details for the lattice pair $(i,\ i+1)$, where both sites are occupied by subpopulation G. The evolution of the corresponding two-point distribution function is given by

$$\begin{aligned}\frac{\mathrm{d}\rho^{(2)}(A_i^G, A_{i+1}^G)}{\mathrm{d}t} =& \frac{P_m^G}{2}\Big[\rho^{(3)}(A_{i-1}^G, 0_i, A_{i+1}^G) + \rho^{(3)}(A_i^G, 0_{i+1}, A_{i+2}^G) - \rho^{(3)}(0_{i-1}, A_i^G, A_{i+1}^G) \\ &- \rho^{(3)}(A_i^G, A_{i+1}^G, 0_{i+2})\Big] + \frac{P_p^G}{2}\Big[\rho^{(3)}(A_{i-1}^G, 0_i, A_{i+1}^G) \\ &+ \rho^{(3)}(A_i^G, 0_{i+1}, A_{i+2}^G)\Big] - 2P_d^G\rho^{(2)}(A_i^G, A_{i+1}^G).\end{aligned} \tag{6.6}$$

In general, the evolution of the n-point distribution function depends on the $(n+1)$-point distribution function. This results in an system of equations, the size of which is equivalent to the number of lattice sites, that describe the evolution of the n-point distribution functions. The large number of lattice sites makes this system of equations algebraically intractable, so to make progress we truncate the system using a moment closure approximation [1]. While several different types of moment closure approximations are available in the literature [17], our previous experience with these kinds of models indicates that the Kirkwood superposition approximation (KSA) [36] is a good option. Further, it can be shown that the KSA provides a near-maximal entropic value [20, 36]. Therefore, we apply the KSA

$$\rho^{(3)}(\sigma_i,\sigma_j,\sigma_k) = \frac{\rho^{(2)}(\sigma_i,\sigma_j)\rho^{(2)}(\sigma_i,\sigma_k)\rho^{(2)}(\sigma_j,\sigma_k)}{\rho^{(1)}(\sigma_i)\rho^{(1)}(\sigma_j)\rho^{(1)}(\sigma_k)}, \tag{6.7}$$

to re-write the three-point distribution functions in Equation (6.6) in terms of two-point distribution functions. After using the KSA, we rewrite Equation (6.6) in terms of the correlation functions to obtain

$$\begin{aligned}
\frac{\mathrm{d}F_i^{G,G}(\Delta)}{\mathrm{d}t} = {} & -F_i^{G,G}(\Delta)\left[\frac{1}{C_i^G}\frac{\mathrm{d}C_i^G}{\mathrm{d}t} + \frac{1}{C_{i+1}^G}\frac{\mathrm{d}C_{i+1}^G}{\mathrm{d}t}\right] + \frac{P_m^G}{2}\left[\frac{C_{i-1}^G}{C_i^G}F_{i-1}^{G,G}(2\Delta)\right. \qquad (6.8)\\
& +\frac{C_{i+2}^G}{C_{i+1}^G}F_i^{G,G}(2\Delta) - 2F_L^{G,G}(\Delta) + F_i^{G,G}(\Delta)\Big\{C_{i-1}^B F_{i-1}^{B,G}(\Delta)F_{i-1}^{B,G}(2\Delta)\\
& +C_{i+2}^B F_i^{G,B}(2\Delta)F_{i+1}^{G,B}(\Delta)\Big\} - \frac{C_{i+1}^B C_{i+2}^G}{C_{i+1}^G}F_i^{G,B}(\Delta)F_i^{G,G}(2\Delta)F_{i+1}^{B,G}(\Delta)\\
& \left.-\frac{C_{i-1}^G C_i^B}{C_i^G}F_{i-1}^{G,B}(\Delta)F_{i-1}^{G,G}(2\Delta)F_i^{B,G}(\Delta)\right] + \frac{P_p^G}{2}\left[\frac{1}{C_i^G} + \frac{1}{C_{i+1}^G} - 2F_i^{G,G}(\Delta)\right.\\
& -\frac{C_i^B}{C_i^G}F_i^{B,G}(\Delta) - \frac{C_{i+1}^B}{C_{i+1}^G}F_i^{G,B}(\Delta) + \frac{C_{i-1}^G F_{i-1}^{G,G}(2\Delta)}{C_i^G\left(1 - C_i^G - C_i^B\right)}\times\\
& \Big\{1 - C_i^G F_{i-1}^{G,G}(\Delta) - C_i^B F_{i-1}^{G,B}(\Delta)\Big\}\Big\{1 - C_i^G F_i^{G,G}(\Delta) - C_i^B F_i^{B,G}(\Delta)\Big\}\\
& +\frac{C_{i+2}^G F_i^{G,G}(2\Delta)}{C_{i+1}^G\left(1 - C_{i+1}^G - C_{i+1}^B\right)}\times\\
& \left.\Big\{1 - C_{i+1}^G F_i^{G,G}(\Delta) - C_{i+1}^B F_i^{G,B}(\Delta)\Big\}\Big\{1 - C_{i+1}^G F_{i+1}^{G,G}(\Delta) - C_{i+1}^B F_{i+1}^{B,G}(\Delta)\Big\}\right]\\
& - 2P_d^G F_i^{G,G}(\Delta). \qquad (6.9)
\end{aligned}$$

We observe that the right-hand side of Equation (6.9) is undefined where either $C_i^G = 0$ or $C_i^G + C_i^B = 1$ and we discuss the subsequent method of solution for the system of correlation functions in the supplementary material document.

Equation (6.9) shows that the evolution of nearest-neighbour correlation functions, $F_i^{G,G}(\Delta)$, depends on the next nearest-neighbour correlation function at $r\Delta = 2\Delta$. Therefore, to make progress we must derive expressions for non-nearest-neighbour correlation functions. To do this we consider the evolution of the correlation function for an arbitrary lattice pair, separated by distance $r\Delta$, and the equations governing the evolution of the correlation function for $r\Delta > \Delta$ that are provided in the supplementary material document. For ease of computation we assume that we have some maximum correlation distance for which, when $r\Delta > r_{\max}\Delta$, we have $F_i^{a,b}(r\Delta) \equiv 1$ [1]. This means that the occupancy status of lattice sites that are sufficiently far apart are uncorrelated. For all one-dimensional results presented in this document we set $r_{\max} = 100$ whereas for all two-dimensional results we set $r_{\max} = 5$, and we find that the results of our moment dynamics model are insensitive to further increases in $r_{\max}$. The complete system of governing equations for the one- and two-dimensional correlation functions are given in the supplementary material document.

6.3 Results

To investigate how the moment dynamics model performs relative to the traditional mean-field model, described by Equation (6.1), we now consider two case studies motivated by the experiments illustrated in Figure 1. To compare the performance of the mean-field and moment dynamics models, we calculate

$$E = \frac{1}{X}\left[\sum_{i=1}^{X}\left(\hat{C}_i^k - C_i^k\right)^2\right]^{\frac{1}{2}}, \tag{6.10}$$

where X is the number of lattice sites, $\hat{C}_i^k$ is the average density of subpopulation k calculated using a large number of identically-prepared realisations of the discrete model and C_i^k is the associated solution of the relevant continuum model. In particular, the discrepancy between the averaged discrete results and the traditional mean-field model is denoted E_{MF}, whereas the discrepancy between the averaged discrete results and the moment dynamics model is denoted E_{MD}. In all cases we solve the governing system of coupled ordinary differential equations using Matlab's `ode45` function, which implements an adaptive fourth order Runge-Kutta method [26].

6.3.1 Case study 1: Co-culture scratch assay

One-dimensional co-culture scratch assay

Co-culture scratch assays involve growing two cell cultures on a culture plate, performing a scratch to reveal a vacant region and observing how the population of cells then spreads in to the initially-vacant region [22, 43]. While the scratch assay shown in Figure 1(a)-(c) focuses on spreading in one direction, we consider an initial condition which leads to spreading in two directions,

$$C_i^G(0) = C_i^B(0) = \begin{cases} \epsilon & 1 \leq i < i_1, \\ C_0 & i_1 \leq i < i_2, \\ \epsilon & i_2 \leq i \leq X, \end{cases} \tag{6.11}$$

where $\epsilon \ll 1$ to allow for the possibility of some material remaining after the scratch has been made. This initial condition corresponds to both subpopulations being placed, evenly distributed, at the same density, in the region $i_1 < i < i_2$. Since the cells are located at random we have $F_i^{a,b}(r) \equiv 1$ at $t = 0$.

Representative snapshots from the discrete model at $t = 0$, $t = 100$ and $t = 200$ are presented in Figures 6.2(a)-(c), respectively. While the discrete model is one-dimensional, we show 20 identically-prepared realisations of the model adjacent to each other in Figures 6.2(a)-(c). Reporting the results of the stochastic model in this way gives us a visual indication of the degree of stochasticity in the model. Comparing the spatial distributions of agents at $t = 100$ and $t = 200$ indicates that the more motile blue subpopulation spreads further from the initial condition than the less motile green subpopulation.

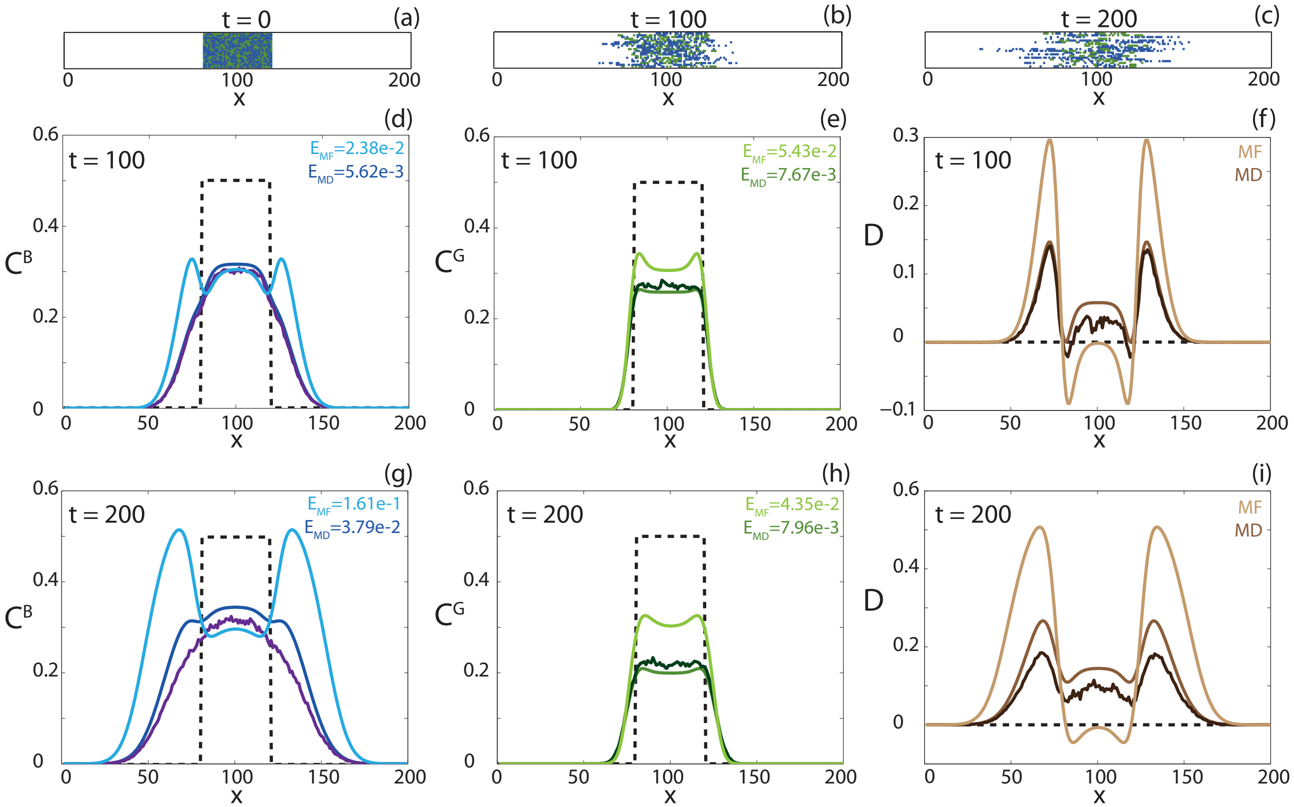

Figure 6.2: One-dimensional model of a co-culture scratch assay. (a)-(c) Snapshots of 20 identically prepared realisations of the discrete model at (a) $t = 0$, (b) $t = 100$ and (c) $t = 200$. (d),(g) Comparison of the averaged discrete model (purple), traditional mean-field solution (light blue) and moment dynamics solution (blue) for cell subpopulation B at (d) $t = 100$ and (g) $t = 200$. (e),(h) Comparison of the averaged discrete model (dark green), traditional mean-field solution (light green) and moment dynamics solution (green) for cell subpopulation G at (e) $t = 100$ and (h) $t = 200$. (f),(i) Comparison of the averaged data from the discrete model (dark brown), traditional mean-field solution (light brown) and moment dynamics solution (brown) describing the difference in density, $D = C^B - C^G$, at (f) $t = 100$ and (i) $t = 200$. Parameters are $P_m^G = 0.1$, $P_m^B = 1$, $P_p^G = P_p^B = 0.05$, $P_d^G = P_d^B = 0.02$, $r_{\max} = 100$, $C_0 = 0.5$, $\epsilon = 10^{-8}$, $i_1 = 81$, $i_2 = 121$, $X = 200$, $\Delta = 1$. Averaged data from the discrete model corresponds to $M = 10^4$ identically prepared realisations. In (d)–(i) the dashed lines corresponds to initial condition, and the discrepancy between the averaged discrete density data and the solution of the traditional mean-field and the moment dynamics models, E_{MF} and E_{MD}, respectively, are given.

The corresponding averaged density profiles, obtained by considering a large number of identically-prepared realisations from the discrete model, are superimposed on the relevant solutions of the mean-field and moment dynamics model for both subpopulations in Figures 6.2(d)-(e), respectively, at $t = 100$. We immediately observe that the traditional mean-field model predicts qualitatively different behaviour to the averaged discrete model. To demonstrate this we plot the difference between the density of the two subpopulations, $D_i = C_i^B - C_i^G$, in Figure 6.2(f). For the averaged discrete density data D_i is predominantly non-negative, whereas the traditional mean-field approach predicts that $D_i < 0$ for a significant portion of the domain. In contrast, the moment dynamics model predicts the same qualitative behaviour as the averaged discrete model. The moment dynamics model provides a closer match to the averaged discrete data ($E_{\mathrm{MD}} = 5.62 \times 10^{-3}$ for subpopulation B and 7.67×10^{-3} for subpopulation G) than the traditional mean-field approach ($E_{\mathrm{MF}} = 2.38 \times 10^{-2}$ for subpopulation B and 5.43×10^{-2} for subpopulation G). An equivalent comparison between the averaged discrete data and the solutions of the traditional mean-field and moment dynamics models at $t = 200$ is given in Figures 6.2(g)-(i). Again, we observe that the traditional mean-field model predicts qualitatively different behaviour to the averaged discrete data, whereas the moment dynamics model provides a reasonable description of the averaged discrete data.

Since the key difference between the derivation of the mean-field model and the moment dynamics model is in the neglect of correlation effects, it is instructive to examine the magnitude of these differences. We can explore these differences since our numerical solution of the moment dynamics model produces estimates of $F_i^{G,G}(r\Delta)$, $F_i^{B,B}(r\Delta)$, $F_i^{G,B}(r\Delta)$ and $F_i^{B,G}(r\Delta)$ for $\Delta \leq r\Delta \leq r_{\max}\Delta$. Solution profiles showing $F_i^{G,G}(r\Delta)$, $F_i^{B,B}(r\Delta)$, $F_i^{G,B}(r\Delta)$ and $F_i^{B,G}(r\Delta)$ are given in the supplementary material document. Given that the mean-field model implicitly assumes that $F_i^{a,b}(r\Delta) \equiv 1$ and that our solution profiles for $F_i^{G,G}(r\Delta)$, $F_i^{B,B}(r\Delta)$, $F_i^{G,B}(r\Delta)$ and $F_i^{B,G}(r\Delta)$ indicate that the correlation function is, at times, up to five orders of magnitude greater than unity, it is not surprising that the traditional mean-field model performs relatively poorly in this case.

The results in Figure 6.2 correspond to one particular choice of the initial cell density in the scratch assay, and we now examine the sensitivity of the performance of the traditional mean-field model relative to the moment dynamics model by decreasing C_0, the initial density of the cell monolayer. We are interested to examine this sensitivity to initial density since previous studies have identified the initial density as playing a key role in the performance of these kinds of models [1, 19]. Results in Figure 6.3 are similar to those in Figure 6.2 except that we consider a much lower initial density of cells by setting $C_0 = 0.1$ in Equation (6.11). In general, we observe that the blue subpopulation in the discrete model moves further away from the initial condition with time than the green subpopulation, as shown in in Figures 6.3(a)-(c). Similar to the results in Figure 6.2, the results in Figures 6.3(d)-(i) show that the traditional mean-field model predicts qualitatively different behaviour than the averaged discrete density data in certain regions of the domain, while the moment dynamics model accurately captures the qualitative trends observed in the averaged discrete data. The details of the correlation functions for this problem are given in the supplementary material document.

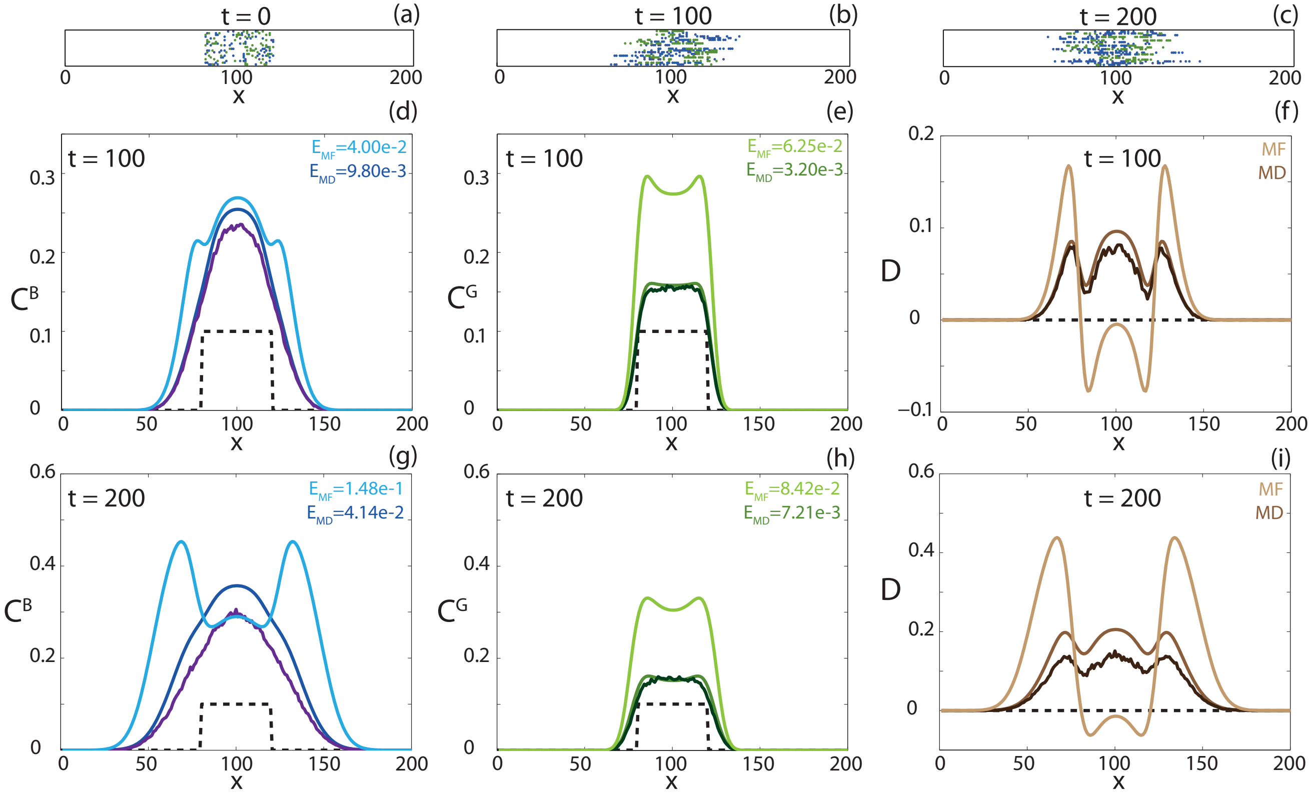

Figure 6.3: One-dimensional model of a co-culture scratch assay. (a)-(c) Snapshots of twenty identically prepared realisations of the discrete model at (a) $t = 0$, (b) $t = 100$ and (c) $t = 200$. (d),(g) Comparison of the averaged discrete model (purple), traditional mean-field solution (light blue) and moment dynamics solution (blue) for subpopulation B at (d) $t = 100$ and (g) $t = 200$. (e),(h) Comparison of the averaged discrete model (dark green), traditional mean-field solution (light green) and moment dynamics solution (green) for subpopulation G at (e) $t = 100$ and (h) $t = 200$. (f),(i) Comparison of the averaged data from the discrete model (dark brown), traditional mean-field solution (light brown) and moment dynamics solution (brown) describing the difference in density, $D = C^B - C^G$, at (f) $t = 100$ and (i) $t = 200$. Parameters are $P_m^G = 0.1$, $P_m^B = 1$, $P_p^G = P_p^B = 0.05$, $P_d^G = P_d^B = 0.02$, $r_{\max} = 100$, $C_0 = 0.1$, $\epsilon = 10^{-8}$, $i_1 = 81$, $i_2 = 121$, $X = 200$, $\Delta = 1$. Averaged data from the discrete model corresponds to $M = 10^4$ identically prepared realisations. In (d)–(i) the dashed lines corresponds to initial condition, and the discrepancy between the averaged discrete density data and the solution of the traditional mean-field and the moment dynamics models, E_{MF} and E_{MD}, respectively, are given.

P_m^B/P_m^G	P_p^B/P_p^G	P_d^B/P_d^G	P_p^B/P_m^B	P_p^G/P_m^G	P_d^B/P_p^B	P_d^G/P_p^G	Mean-field	Corrected mean-field
Large	Intermediate	Intermediate	Intermediate	Intermediate	Intermediate	Intermediate	**X**	✓
Intermediate	Intermediate	Intermediate	Intermediate	Intermediate	Intermediate	Large	**X***	✓
Large	Large	Intermediate	Small	Small	Intermediate	Large	✓	✓
Intermediate	Intermediate	Large	Intermediate	Intermediate	Intermediate	Small	**X***	✓
Intermediate	Intermediate	Intermediate	Small	Small	Large	Large	✓	✓
Intermediate	Intermediate	Intermediate	Large	Large	Intermediate	Intermedate	**X**	✓
Large	Intermediate	Intermediate	Small	Large	Intermediate	Intermediate	**X***	✓

Table 6.1: Parameter ratios and the validity of both the mean-field and moment dynamics models for describing the averaged discrete model for those parameter ratios and the cell co-culture scratch assay initial condition. Large indicates 10^1 or higher, intermediate indicates $5 \times 10^{-2} - 5 \times 10^0$, small indicates less than 10^{-2}. **X** denotes a model that is inappropriate for the corresponding parameter ratio while **X*** denotes a model that provides an accurate prediction for one subpopulation, but not both. The tick symbol denotes a model that provides a prediction that matches the averaged discrete model well.

To further investigate the performance of the moment dynamics model we now summarise results for a wider range of parameter combinations. Since the moment dynamics model requires additional effort to derive and solve compared to the traditional mean-field description, it is of interest to use our model to identify which particular parameter regimes require the application of a moment dynamics model, and which particular parameter regimes can be studied using the simpler traditional mean-field approach. Results in Table 6.1 describe the performance of the moment dynamics and traditional mean-field models for the same problem we considered in Figure 6.2. Using criteria based on Equation (6.10), we conclude that the moment dynamics model outperforms the traditional mean-field description across a large range of parameter combinations. In particular, we observe that the traditional mean-field model fails to describe the average behaviour of the discrete model whenever proliferation is significant, that is, where the proliferation rate is not significantly smaller than the motility rate. We observe that if, for both subpopulations, P_p^k is small compared to P_m^k, then the mean-field model describes the averaged discrete model well for both subpopulations. While the mean-field model is appropriate in certain parameter regimes, the moment dynamics model always provides an improved match to the averaged discrete density data.

Two-dimensional co-culture stencil assay

We now present results for a two-dimensional extension of the model considered in Section 6.3.1. While we motivated the geometry of our simulations in Section 6.3.1 by considering a scratch assay, we note that there are several other types of *in vitro* assays, such as barrier assays [34] or stencil assays [16, 25], that involve an initially-confined population of cells which spread in two dimensions. The details of the equations governing the two-dimensional moment dynamics model are given in the supplementary material document. We apply our model to a square stencil assay, where cells are grown initially inside a square stencil. The assay is initiated by removing the stencil and allowing the cells to spread into the area surrounding the initially-confined population of cells. We model this process using an initial condition given by

$$C^G_{(i,j)}(0) = C^B_{(i,j)}(0) = \begin{cases} C_0 & i_1 \leq i < i_2,\ j_1 \leq j < j_2, \\ \epsilon & \text{elsewhere.} \end{cases} \tag{6.12}$$

Again, we make the assumption that both cell subpopulations are initially present at the same density, such as the traditional mean-field initial condition shown in Figure 6.4(a) with both subpopulations present with $C_0 = 0.1$ inside the square stencil. The discrete analogue of this initial condition for a single realisation of the discrete model is presented in Figure 6.4(b). We allow the discrete model to evolve until $t = 100$, and present a snapshot of the results in Figure 6.4(c). In the two-dimensional setting we observe the formation of clustering, particularly in the less motile G subpopulation. This kind of clustering is frequently observed in many different experimental situations, such as in Figure 6.1(c).

We perform many identically-prepared realisations of the discrete model and present the average density distributions, for both subpopulation B and subpopulation G, in Figures 6.4(d) and 6.4(e), respectively. As we might expect, the more motile subpopulation B spreads further away from the location of the initial condition than subpopulation G. Interestingly, although both cell subpopulations have the same rates of proliferation and death, subpopulation B has a higher maximum density. The difference between the density of the two subpopulations is reported in Figure 6.4(f) and we observe that, aside from minor fluctuations, we have $C^B_{(i,j)} > C^G_{(i,j)}$ across the domain. It is instructive to examine whether this qualitative behaviour is captured by the traditional mean-field and moment dynamics models. The traditional mean-field solutions for subpopulations B and G, presented in Figures 6.4(g) and 6.4(h), respectively, exhibit higher cell density than the averaged discrete data. In particular, according to the traditional mean-field model, subpopulation G has a maximum density of approximately 0.25 whereas the maximum density according to the averaged data from the discrete model is approximately 0.15. The difference between the density of the two subpopulations according to the traditional mean-field model, given in Figure 6.4(i), predicts that $C^G_{(i,j)} > C^B_{(i,j)}$ in large parts of the domain, which is precisely the opposite of what we observe in the averaged discrete data.

To investigate whether including spatial correlation addresses the limitations of the traditional mean-field model, we compare the predictions of our moment dynamics model with the averaged discrete model. We note that, in the two-dimensional case, lattice sites separated in both the x and y directions can be correlated and that the maximum separation in both the x and y direction is denoted by $r_{\max}$. The relevant solution of the moment dynamics model is presented in Figures 6.4(j) and 6.4(k) for subpopulations B and G, respectively. Visually, we observe that the moment dynamics model matches the averaged discrete data far better than the solution of the traditional mean-field model. Indeed, measuring the difference between the solution of the moment dynamics model and the averaged discrete data leads to estimates of E_{MD} that are approximately one order of magnitude lower than estimates of E_{MF}.

6.3.2 Case study 2: Invasion of one subpopulation into another subpopulation

Cell invasion occurs when one cell subpopulation moves through a distinct background cell subpopulation, such as tumour cells spreading through the stroma [3]. To model this kind of process we assume that the background cell subpopulation is initially spatially uniform and can be modelled as a one-dimensional process. Therefore, we assume that one cell subpopulation, C_i^G, is uniformly distributed at some initial density, C_0^G, while the other cell subpopulation is initially confined, so that we can mimic the kind of geometry we see in Figure 1(d). To achieve this we set

$$
\begin{aligned}
C_i^G(0) &= C_0^G, \quad 1 \leq i \leq X, \\
C_i^B(0) &= \begin{cases} \epsilon & 1 \leq i < i_1, \\ C_0^B & i_1 \leq i < i_2, \\ \epsilon & i_2 \leq i \leq X, \end{cases}
\end{aligned}
\tag{6.13}
$$

as the initial condition, where $\epsilon \ll 1$.

Twenty identically-prepared realisations of the one-dimensional discrete model, at $t = 0$, $t = 100$ and $t = 200$, are presented in Figure 6.5(a)-(c) and Figure 6.6(a)-(c), respectively. The difference between Figure 6.5 and Figure 6.6 is in the choice of parameters. In summary, subpopulation B is more motile than subpopulation G in Figure 6.5, whereas subpopulation G is more motile than subpopulation B in Figure 6.6. We compare the relative performance of the traditional mean-field and moment dynamics models for the relevant parameter choices in Figures 6.5(d)-(f) and Figures 6.6(d)-(f) at $t = 100$, and in Figures 6.5(g)-(i) and Figures 6.6(g)-(i) at $t = 200$. In general, we observe that the solution of the moment dynamics model provides an improved match to the averaged discrete data relative to the solution of the traditional mean-field model for both parameter choices. In particular, the solution of the moment dynamics model provides an improved approximation of the averaged density from the discrete model at the low density leading edge of the invading subpopulation in Figure 6.5 where $P_m^B > P_m^G$. This improvement offered by the moment dynamics model at the leading edge of the spreading population is of particular interest when considering surgical removal of tumours where it is essential to have a good understanding of the location of the leading edge of the spreading subpopulation [2, 40].

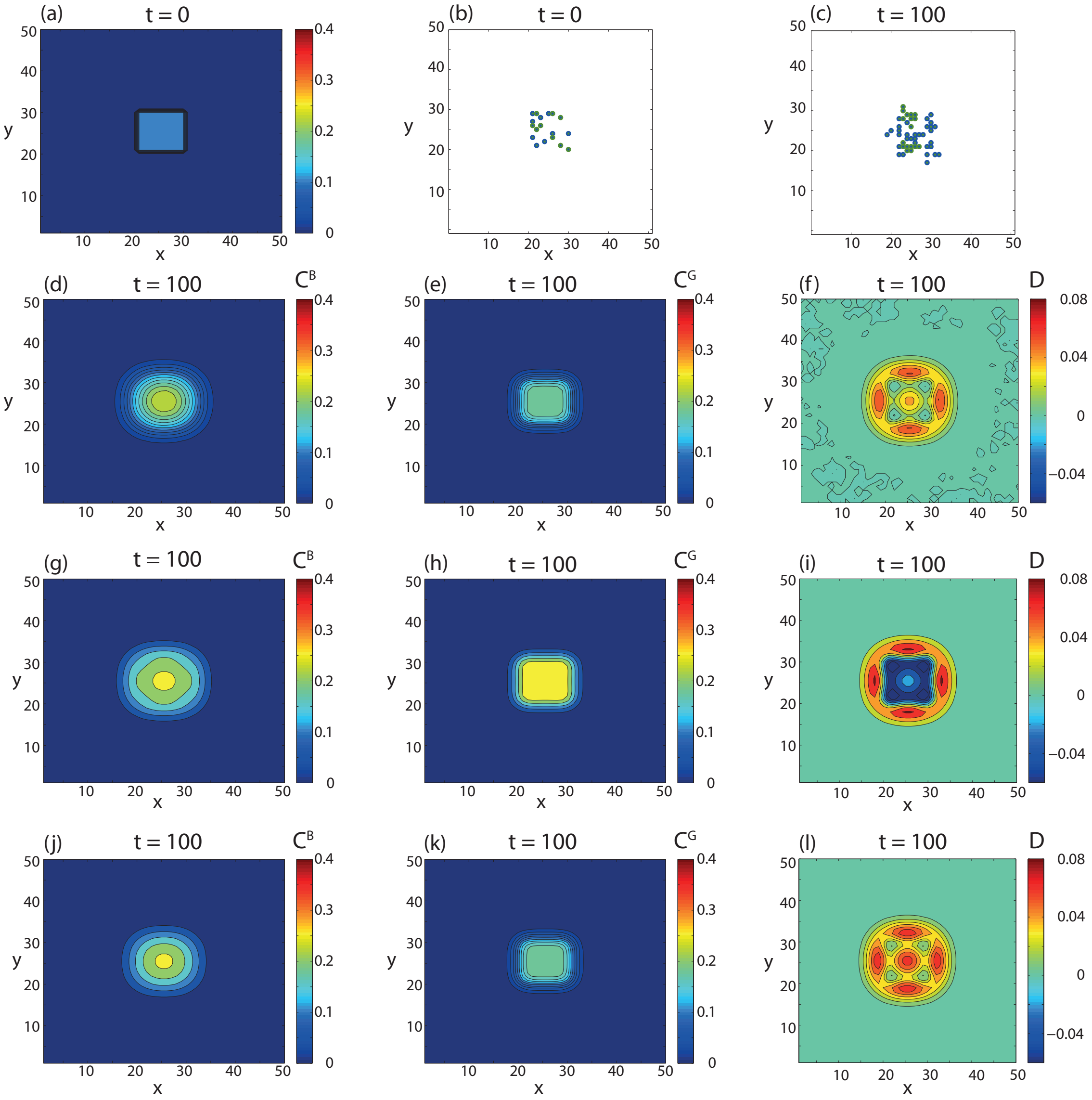

Figure 6.4: Two-dimensional model of a co-culture stencil assay. (a) Initial condition for the traditional mean-field and moment dynamics model. (b)-(c) Single realisation of the discrete model at (b) $t = 0$ and (c) $t = 100$. (d)-(e) Averaged density data from the discrete model for (d) subpopulation B and (e) subpopulation G at $t = 100$. (f) Averaged data from the discrete model describing the difference between the two subpopulations, $D = C^B - C^G$. (g)-(h) Traditional mean-field solution for (g) subpopulation B and (h) subpopulation G at $t = 100$. (i) Difference between the two subpopulations, $D = C^B - C^G$, for the traditional mean-field model. (j)-(k) Corrected mean-field solution for (j) subpopulation B and (k) subpopulation G at $t = 100$. (i) Difference between the two subpopulations, $C^B - C^G$, for the moment dynamics model. Parameters are $P_m^G = 0.01$, $P_m^B = 0.1$, $P_p^G = P_p^B = 0.05$, $P_d^G = P_d^B = 0.02$, $r_{\max} = 5$, $C_0 = 0.1$, $\epsilon = 10^{-8}$, $i_1 = j_1 = 81$, $i_2 = j_2 = 121$, $X = Y = 50$. Difference, given by Equation (6.10), between the mean-field solution and the averaged discrete model: 1.57×10^{-2} (subpopulation B) and 2.49×10^{-2} (subpopulation G). Averaged data from the discrete model corresponds to $M = 10^6$ identically prepared realisations. The discrepancy between the averaged discrete density data and the solution of the traditional mean-field and the moment dynamics models, given by Equation (6.10) are: (d) 3.90×10^{-3} (subpopulation B) and 9.48×10^{-4} (subpopulation G).

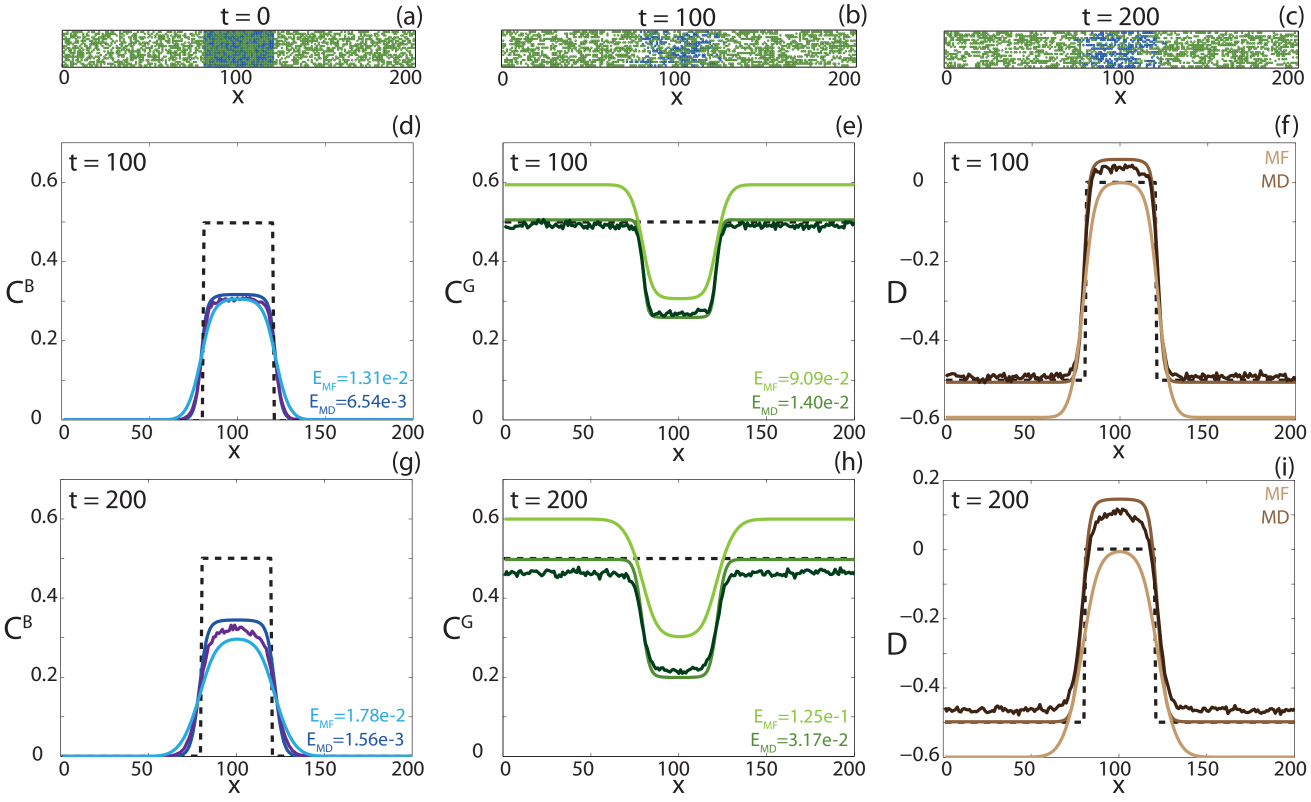

Figure 6.5: One-dimensional model of cell invasion. (a)-(c) Snapshots of 20 identically prepared realisations of the discrete model at (a) $t = 0$, (b) $t = 100$ and (c) $t = 200$. (d),(g) Comparison of the averaged discrete model (purple), corresponding mean-field solution (light blue) and moment dynamics solution (blue) for cell subpopulation B at (d) $t = 100$ and (g) $t = 200$. (e),(h) Comparison of the averaged discrete model (dark green), corresponding mean-field solution (light green) and moment dynamics solution (green) for cell subpopulation G at (e) $t = 100$ and (h) $t = 200$. (f),(i) Comparison of the averaged discrete model (dark brown), corresponding mean-field solution (light brown) and moment dynamics solution (brown) for the difference in cell subpopulations $D = C^B - C^G$ at (f) $t = 100$ and (i) $t = 200$. Parameters are $P_m^G = 0.1$, $P_m^B = 1$, $P_p^G = P_p^B = 0.05$, $P_d^G = P_d^B = 0.02$, $r_{\max} = 100$, $C_0^G = C_0^B = 0.5$, $\epsilon = 10^{-8}$, $i_1 = 81$, $i_2 = 121$, $X = 200$, $\Delta = 1$. Averaged data from the discrete model corresponds to $M = 10^4$ identically prepared realisations. In (d)–(i) the dashed lines corresponds to initial condition, and the discrepancy between the averaged discrete density data and the solution of the traditional mean-field and the moment dynamics models, E_{MF} and E_{MD}, respectively, are given.

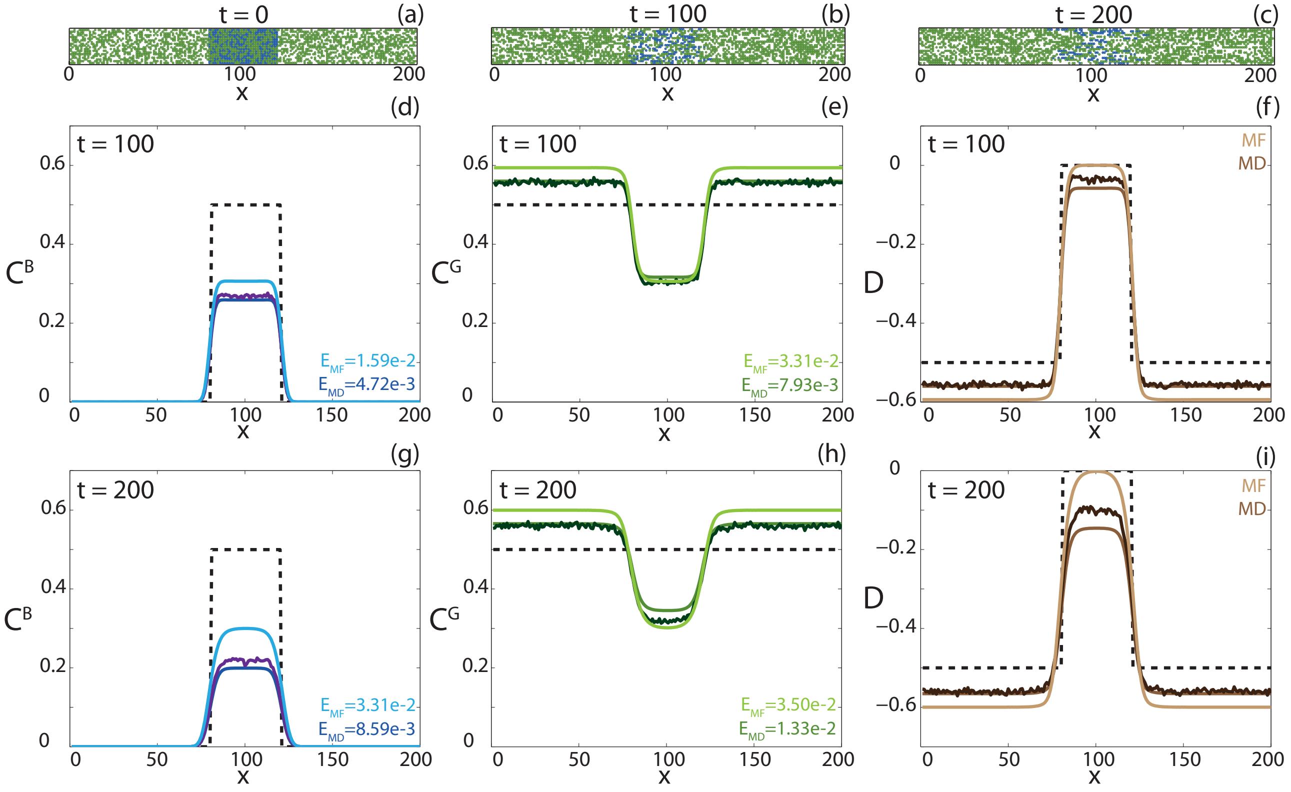

Figure 6.6: One-dimensional model of cell invasion. (a)-(c) Snapshots of 20 identically prepared realisations of the discrete model at (a) $t = 0$, (b) $t = 100$ and (c) $t = 200$. (d),(g) Comparison of the averaged discrete model (purple), corresponding mean-field solution (light blue) and moment dynamics solution (blue) for cell subpopulation B at (d) $t = 100$ and (g) $t = 200$. (e),(h) Comparison of the averaged discrete model (dark green), corresponding mean-field solution (light green) and moment dynamics solution (green) for cell subpopulation G at (e) $t = 100$ and (h) $t = 200$. (f),(i) Comparison of the averaged discrete model (dark brown), corresponding mean-field solution (light brown) and moment dynamics solution (brown) for the difference in cell subpopulations $D = C^B - C^G$, at (f) $t = 100$ and (i) $t = 200$. Parameters are $P_m^G = 1$, $P_m^B = 0.1$, $P_p^G = P_p^B = 0.05$, $P_d^G = P_d^B = 0.02$, $r_{\max} = 100$, $C_0^G = C_0^B = 0.5$, $\epsilon = 10^{-8}$, $i_1 = 81$, $i_2 = 121$, $X = 200$, $\Delta = 1$. Averaged data from the discrete model corresponds to $M = 10^4$ identically prepared realisations. In (d)–(i) the dashed lines corresponds to initial condition, and the discrepancy between the averaged discrete density data and the solution of the traditional mean-field and the moment dynamics models, E_{MF} and E_{MD}, respectively, are given.

P_m^B/P_m^G	P_p^B/P_p^G	P_d^B/P_d^G	P_p^B/P_m^B	P_p^G/P_m^G	P_d^B/P_p^B	P_d^G/P_p^G	Mean-field	Corrected mean-field
Large	Intermediate	Intermediate	Intermediate	Intermediate	Intermediate	Intermediate	**X**	✓
Intermediate	Large	Intermediate	Intermediate	Intermediate	Intermediate	Large	**X**	✓
Intermediate	Intermediate	Intermediate	Intermediate	Intermediate	Intermediate	Large	**X***	✓
Large	Large	Intermediate	Small	Small	Intermediate	Large	**X***	✓
Intermediate	Intermediate	Intermediate	Small	Small	Intermediate	Intermediate	✓	✓
Intermediate	Intermediate	Intermediate	Intermediate	Intermediate	Small	Intermediate	**X**	✓
Intermediate	Intermediate	Intermediate	Small	Small	Large	Large	**X***	✓
Intermediate	Intermediate	Intermediate	Large	Large	Intermediate	Intermedate	**X***	✓
Large	Intermediate	Intermediate	Small	Large	Intermediate	Intermediate	**X**	✓
Small	Large	Intermediate	Large	Small	Intermediate	Intermediate	**X***	✓

Table 6.2: Parameter ratios and the validity of the mean-field and moment dynamics models for describing the averaged discrete model for the cell invasion initial condition. Large indicates 10^1 or higher, intermediate indicates $5 \times 10^{-2} - 5 \times 10^0$, small indicates less than 10^{-2}. **X** denotes a model that is inappropriate for the corresponding parameter ratio while **X*** denotes a model that provides an accurate prediction for one subpopulation, but not both. The tick symbol denotes a model that matches the averaged discrete model well.

To examine the role of initial cell density we present an additional set of results in Figure 6.7 where we have reduced the initial cell density. The additional results in Figure 6.7 involve the same initial conditions, given by Equation (6.13), except that we set $C_0^G = C_0^B = 0.1$. Since the background density has been decreased, we observe that the invading subpopulation spreads further in Figure 6.7 than in the corresponding situations presented in Figure 6.5 and Figure 6.6. It is interesting that both the solutions of the mean-field and moment dynamics models are less accurate in describing the density of the background subpopulation for the lower density initial condition.

To provide more comprehensive insight into the relative performance of the moment dynamics model we also examine the match between the average density data and the solution of the traditional mean-field and the moment dynamics models over a range of parameter combinations. The results of this comparison are summarised in Table 6.2, where we see that the solution of the moment dynamics model matches the averaged density data from the discrete model better than the corresponding solution of the traditional mean-field model in each parameter regime considered. We also observe that the traditional mean-field model is appropriate for situations where the proliferation rate is small relative to the motility rate.

6.4 Discussion and conclusions

In this work we have considered developing mathematical models which describe the motion of populations containing distinct subpopulations. These kinds of processes are relevant to a range of biological and ecological applications including malignant spreading [27], wound healing [28] and the spread of invasive species [13]. Previous models of these processes typically focus on population-level PDE descriptions that neglect to explicitly account for individual-level behaviour [9, 23, 27, 39]. To partly address this limitation, other researchers use discrete mathematical models in conjunction with the associated population-level PDE description which is derived from the underlying stochastic process by invoking a mean-field approximation [15, 33]. While averaged density data

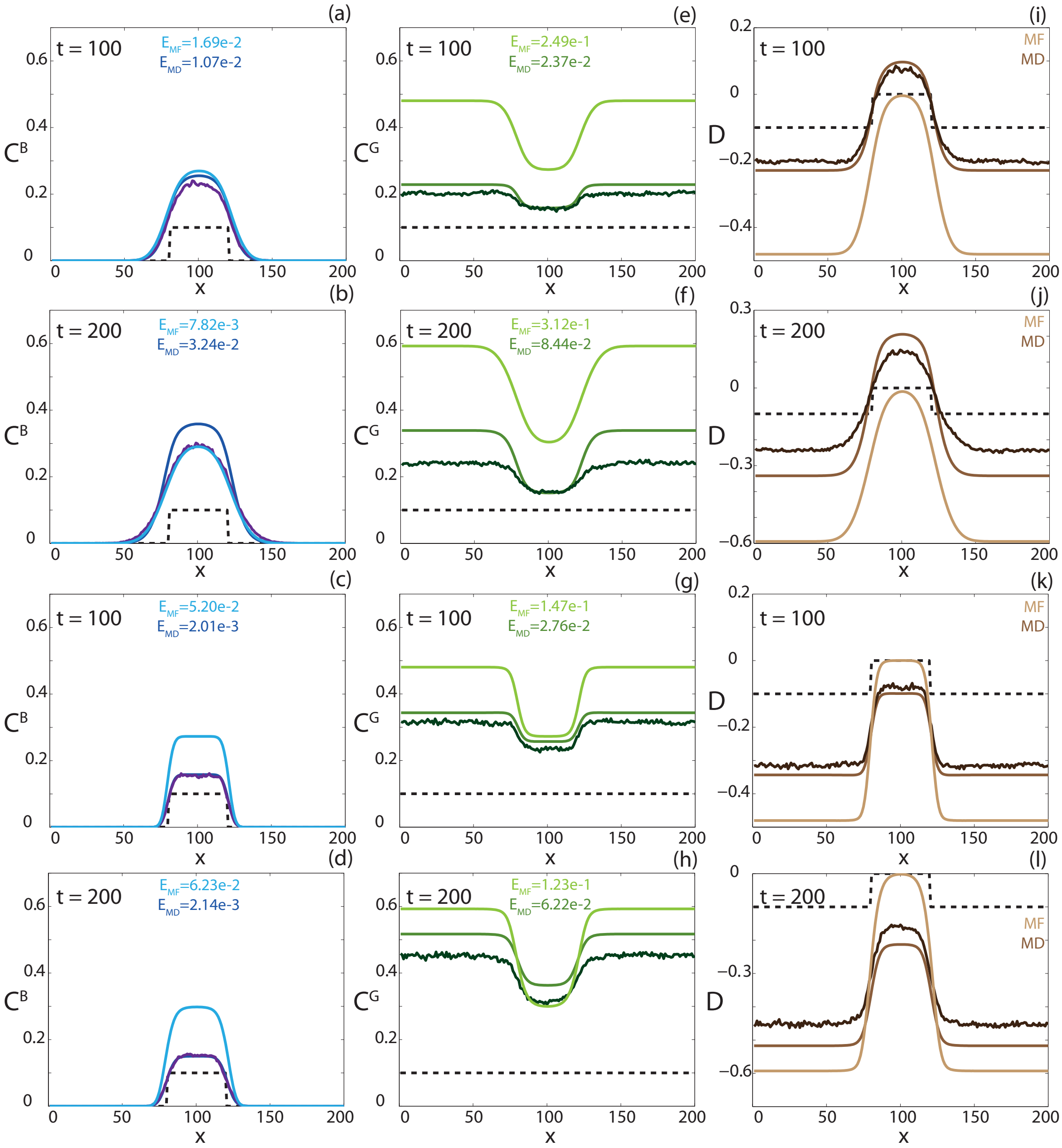

Figure 6.7: One-dimensional model of cell invasion. (a)-(d) Comparison of the averaged discrete model (purple), corresponding mean-field solution (light blue) and moment dynamics solution (blue) for cell subpopulation B at (a),(c) $t = 100$ and (b),(d) $t = 200$ for (a)-(b) parameter regime one and (c)-(d) parameter regime two. (e)-(h) Comparison of the averaged discrete model (purple), corresponding mean-field solution (light blue) and moment dynamics solution (blue) for cell subpopulation G at (e),(g) $t = 100$ and (f),(h) $t = 200$ for (e)-(f) parameter regime one and (g)-(h) parameter regime two. (i)-(l) Comparison of the averaged discrete model (purple), corresponding mean-field solution (light blue) and moment dynamics solution (blue) for cell subpopulation $D = C^B - C^G$, at (i),(k) $t = 100$ and (j),(l) $t = 200$ for (i)-(j) parameter regime one and (k)-(l) parameter regime two. Parameters used were $r_{\max} = 100$, $C_0^G = C_0^B = 0.1$, $\epsilon = 10^{-8}$, $i_1 = 81$, $i_2 = 121$, $X = 200$, $\Delta = 1$. Paramter regime one used $P_m^G = 0.1$, $P_m^B = 1$, $P_p^G = P_p^B = 0.05$, $P_d^G = P_d^B = 0.02$. Parameter regime two used $P_m^G = 1$, $P_m^B = 0.1$, $P_p^G = P_p^B = 0.05$, $P_d^G = P_d^B = 0.02$. Averaged data from the discrete model corresponds to $M = 10^4$ identically prepared realisations. In (a)–(l) the dashed lines corresponds to initial condition, and the discrepancy between the averaged discrete density data and the solution of the traditional mean-field and the moment dynamics models, E_{MF} and E_{MD}, respectively, are given.

from these kinds of stochastic models is known to match the solution of the associated mean-field PDE approximation in certain parameter regimes, it is well-known that mean-field PDE descriptions fail to match average density information from the stochastic process for parameter combinations where the discrete model leads to significant correlation and clustering effects [1].

Our study, in which we derive new moment dynamics models governing the motion of cell populations composed of interacting subpopulations, offers two improvements on previous approaches. First, our moment dynamics model approximately incorporates clustering and correlation, which are implicitly neglected in previous PDE-based descriptions. This is important since clustering and correlation effects are often observed in cell biology experiments [42]. We note that the clustering incorporated is not due to explicit cell-to-cell adhesion but from the nature of the proliferation mechanism. Second, our moment dynamics model is more computationally efficient to implement than using a large number of repeated stochastic realisations of the discrete model. By presenting a thorough comparison of the performance of a traditional mean-field model and the new moment dynamics model for two case studies we are able to summarise some of the general features of our model. While we always find that the moment dynamics model produces a more accurate description of the entire cell density profile than the traditional mean-field model, we also find that certain features of the processes are reliably predicted by the traditional mean-field framework.

There are several ways that our work could be extended. In all cases we always assume that the influence of cell-to-cell adhesion and cell-to-substrate adhesion is negligible. While these assumptions are relevant for certain types of cells, it is well known that other types of cells, such as glioma and melanoma, exhibit significant adhesion [15, 41, 42]. Therefore, it is of interest to examine how to incorporate the effects of cell-to-cell adhesion and cell-to-substrate adhesion in our moment dynamics framework. In this work we have only considered cell biology as an application of this framework. Future applications of this framework could include chemical kinetics [37] and predator-prey interactions [21]. A further extension would be to consider an off-lattice discrete process, such as the model presented by Middleton *et al.* [20]. Given that off-lattice discrete models are far more computationally expensive than lattice-based discrete models, the need for efficient and accurate mean-field descriptions of these processes is even more significant for off-lattice models than for lattice–based models. We leave these extensions for future analysis.

Acknowledgements: This work was financially supported by the Cooperative Research Centre for Wound Management Innovation and the Australian Research Council (FT130100148). We also appreciate the assistance of Emeritus Professor Sean McElwain.

Bibliography

[1] Baker, R. E. and Simpson, M. J. [2010], 'Correcting mean-field approximations for birth-death-movement processes', *Physical Review E* **82**(4), 041905.

[2] Beets-Tan, R., Beets, G., Vliegen, R., Kessels, A., Van Boven, H., De Bruine, A., Von Meyenfeldt, M., Baeten, C. and Van Engelshoven, J. [2001], 'Accuracy of magnetic resonance imaging in prediction of tumour-free resection margin in rectal cancer surgery', *The Lancet* **357**(9255), 497–504.

[3] Bhowmick, N. A. and Moses, H. L. [2005], 'Tumor–stroma interactions', *Current Opinion in Genetics & Development* **15**(1), 97–101.

[4] Binder, B. J. and Landman, K. A. [2009], 'Exclusion processes on a growing domain', *Journal of Theoretical Biology* **259**(3), 541–551.

[5] Chowdhury, D., Schadschneider, A. and Nishinari, K. [2005], 'Physics of transport and traffic phenomena in biology: from molecular motors and cells to organisms', *Physics of Life Reviews* **2**(4), 318–352.

[6] Codling, E. A., Plank, M. J. and Benhamou, S. [2008], 'Random walk models in biology', *Journal of the Royal Society Interface* **5**(25), 813–834.

[7] De Wever, O. and Mareel, M. [2003], 'Role of tissue stroma in cancer cell invasion', *The Journal of Pathology* **200**(4), 429–447.

[8] Fernando, A. E., Landman, K. A. and Simpson, M. J. [2010], 'Nonlinear diffusion and exclusion processes with contact interactions', *Physical Review E* **81**(1), 011903.

[9] Gatenby, R. A. and Gawlinski, E. T. [1996], 'A reaction-diffusion model of cancer invasion', *Cancer Research* **56**(24), 5745–5753.

[10] Gatenby, R. A., Gawlinski, E. T., Gmitro, A. F., Kaylor, B. and Gillies, R. J. [2006], 'Acid-mediated tumor invasion: a multidisciplinary study', *Cancer Research* **66**(10), 5216–5223.

[11] Gillespie, D. T. [1977], 'Exact stochastic simulation of coupled chemical reactions', *The Journal of Physical Chemistry* **81**(25), 2340–2361.

[12] Grima, R. [2008], 'Multiscale modeling of biological pattern formation', *Current Topics in Developmental Biology* **81**, 435–460.

[13] Hastings, A., Cuddington, K., Davies, K. F., Dugaw, C. J., Elmendorf, S., Freestone, A., Harrison, S., Holland, M., Lambrinos, J., Malvadkar, U. et al. [2005], 'The spatial spread of invasions: new developments in theory and evidence', *Ecology Letters* **8**(1), 91–101.

[14] Johnston, S. T., Simpson, M. J. and Baker, R. E. [2012], 'Mean-field descriptions of collective migration with strong adhesion', *Physical Review E* **85**(5), 051922.

[15] Khain, E., Katakowski, M., Charteris, N., Jiang, F. and Chopp, M. [2012], 'Migration of adhesive glioma cells: front propagation and fingering', *Physical Review E* **86**(1), 011904.

[16] Kroening, S. and Goppelt-Struebe, M. [2010], 'Analysis of matrix-dependent cell migration with a barrier migration assay', *Science Signaling* **3**(126), pl1.

[17] Law, R. and Dieckmann, U. [2000], 'A dynamical system for neighborhoods in plant communities', *Ecology* **81**(8), 2137–2148.

[18] Li, G., Satyamoorthy, K., Meier, F., Berking, C., Bogenrieder, T. and Herlyn, M. [2003], 'Function and regulation of melanoma–stromal fibroblast interactions: when seeds meet soil', *Oncogene* **22**(20), 3162–3171.

[19] Markham, D. C., Simpson, M. J., Maini, P. K., Gaffney, E. A. and Baker, R. E. [2013], 'Incorporating spatial correlations into multispecies mean-field models', *Physical Review E* **88**(5), 052713.

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

[21] Murrell, D. J. [2005], 'Local spatial structure and predator-prey dynamics: counterintuitive effects of prey enrichment', *The American Naturalist* **166**(3), 354–367.

[22] Oberringer, M., Meins, C., Bubel, M. and Pohlemann, T. [2007], 'A new in vitro wound model based on the co-culture of human dermal microvascular endothelial cells and human dermal fibroblasts', *Biology of the Cell* **99**(4), 197–207.

[23] Painter, K. J. and Sherratt, J. A. [2003], 'Modelling the movement of interacting cell populations', *Journal of Theoretical Biology* **225**(3), 327–339.

[24] Phillips, B. L., Brown, G. P., Greenlees, M., Webb, J. K. and Shine, R. [2007], 'Rapid expansion of the cane toad (bufo marinus) invasion front in tropical australia', *Austral Ecology* **32**(2), 169–176.

[25] Riahi, R., Yang, Y., Zhang, D. D. and Wong, P. K. [2012], 'Advances in wound-healing assays for probing collective cell migration', *Journal of Laboratory Automation* **17**(1), 59–65.

[26] Shampine, L. F. and Reichelt, M. W. [1997], 'The Matlab ODE suite', *SIAM Journal on Scientific Computing* **18**(1), 1–22.

[27] Sherratt, J. A. [2000], 'Wavefront propagation in a competition equation with a new motility term modelling contact inhibition between cell populations', *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* **456**(2002), 2365–2386.

[28] Sherratt, J. A. and Murray, J. [1990], 'Models of epidermal wound healing', *Proceedings of the Royal Society of London. Series B: Biological Sciences* **241**(1300), 29–36.

[29] Simpson, M. J. and Baker, R. E. [2011], 'Corrected mean-field models for spatially dependent advection-diffusion-reaction phenomena', *Physical Review E* **83**(5), 051922.

[30] Simpson, M. J., Haridas, P. and McElwain, D. L. S. [2014], 'Do pioneer cells exist?', *PLOS ONE* **9**(1), e85488.

[31] Simpson, M. J., Landman, K. A. and Bhaganagarapu, K. [2007], 'Coalescence of interacting cell populations', *Journal of Theoretical Biology* **247**(3), 525–543.

[32] Simpson, M. J., Landman, K. A. and Hughes, B. D. [2009], 'Multi-species simple exclusion processes', *Physica A: Statistical Mechanics and its Applications* **388**(4), 399–406.

[33] Simpson, M. J., Landman, K. A. and Hughes, B. D. [2010], 'Cell invasion with proliferation mechanisms motivated by time-lapse data', *Physica A: Statistical Mechanics and its Applications* **389**(18), 3779–3790.

[34] Simpson, M. J., Treloar, K. K., Binder, B. J., Haridas, P., Manton, K. J., Leavesley, D. I., McElwain, D. L. S. and Baker, R. E. [2013], 'Quantifying the roles of cell motility and cell proliferation in a circular barrier assay', *Journal of The Royal Society Interface* **10**(82), 20130007.

[35] Simpson, M. J., Zhang, D. C., Mariani, M., Landman, K. A. and Newgreen, D. F. [2007], 'Cell proliferation drives neural crest cell invasion of the intestine', *Developmental Biology* **302**(2), 553–568.

[36] Singer, A. [2004], 'Maximum entropy formulation of the Kirkwood superposition approximation', *The Journal of Chemical Physics* **121**(8), 3657–3666.

[37] Singh, A. and Hespanha, J. P. [2011], 'Approximate moment dynamics for chemically reacting systems', *Automatic Control, IEEE Transactions on* **56**(2), 414–418.

[38] Skellam, J. [1951], 'Random dispersal in theoretical populations', *Biometrika* pp. 196–218.

[39] Smallbone, K., Gavaghan, D. J., Gatenby, R. A. and Maini, P. K. [2005], 'The role of acidity in solid tumour growth and invasion', *Journal of Theoretical Biology* **235**(4), 476–484.

[40] Swan, G. W. [1975], 'A mathematical model for the density of malignant cells in the spread of cancer in the uterus', *Mathematical Biosciences* **25**(3), 319–329.

[41] Treloar, K. K. and Simpson, M. J. [2013], 'Sensitivity of edge detection methods for quantifying cell migration assays', *PLOS ONE* **8**(6), e67389.

[42] Treloar, K. K., Simpson, M. J., Binder, B. J., McElwain, D. L. S. and Baker, R. E. [2014], 'Assessing the role of spatial correlations during collective cell spreading', *Scientific Reports* **4**(5713).

[43] Walter, M., Wright, K. T., Fuller, H., MacNeil, S. and Johnson, W. E. B. [2010], 'Mesenchymal stem cell-conditioned medium accelerates skin wound healing: an in vitro study of fibroblast and keratinocyte scratch assays', *Experimental Cell Research* **316**(7), 1271–1281.

6.5 Supplementary material

6.5.1 Two-dimensional mean-field model

We consider the change in occupancy of the lattice site (i, j) by accounting for all possible motility, proliferation and death events, and obtain

$$\begin{aligned}\frac{\mathrm{d}C^k_{(i,j)}}{\mathrm{d}t} =& \frac{P^k_m}{4}\Big[C^k_{(i-1,j)}\Phi_{(i,j)} + C^k_{(i+1,j)}\Phi_{(i,j)} + C^k_{(i,j-1)}\Phi_{(i,j)} + C^k_{(i,j+1)}\Phi_{(i,j)} \\ & - C^k_{(i,j)}\Phi_{(i-1,j)} - C^k_{(i,j)}\Phi_{(i+1,j)} - C^k_{(i,j)}\Phi_{(i,j-1)} - C^k_{(i,j)}\Phi_{(i,j+1)}\Big] \\ & + \frac{P^k_p}{4}\Big[C^k_{(i-1,j)}\Phi_{(i,j)} + C^k_{(i+1,j)}\Phi_{(i,j)} + C^k_{(i,j-1)}\Phi_{(i,j)} + C^k_{(i,j+1)}\Phi_{(i,j)}\Big] - P^k_d C^k_{(i,j)},\end{aligned}$$

where $C^k_{(i,j)}$ is the average lattice occupancy of site (i, j) for species $k \in K$ and $\Phi_{(i,j)} = 1 - \sum_K C^k_{(i,j)}$ represents the probability that site (i, j) is unoccupied.

6.5.2 One-dimensional moment dynamics model

We consider the change in occupancy of the lattice site i by accounting for all possible motility, proliferation and death events, in terms of the one- and two-point distribution functions. We then rewrite the distribution functions in terms of the correlation function [1] and obtain

$$\begin{aligned}\frac{\mathrm{d}C_i^G}{\mathrm{d}t} =&\frac{P_m^G}{2}\Bigg[C_{i-1}^G + C_{i+1}^G - 2C_i^G + C_i^B\left\{2C_i^G - C_{i-1}^G F_{i-1}^{G,B}(\Delta) - C_{i+1}^G F_i^{B,G}(\Delta)\right\}\\ &- C_i^G\left\{2C_i^B - C_{i-1}^B F_{i-1}^{B,G}(\Delta) - C_{i+1}^B F_i^{G,B}(\Delta)\right\}\Bigg]\\ &+ \frac{P_p^G}{2}\Bigg[C_{i-1}^G\left\{1 - C_i^G F_{i-1}^{G,G}(\Delta) - C_i^B F_{i-1}^{G,B}(\Delta)\right\}\\ &+ C_{i+1}^G\left\{1 - C_i^G F_i^{G,G}(\Delta) - C_i^B F_i^{B,G}(\Delta)\right\}\Bigg] - P_d^G C_i^G,\end{aligned}$$

where $F_i^{a,b}(r\Delta)$ is the correlation between sites i and m, such that $r\Delta = (m-i)\Delta$, with lattice site i occupied by subpopulation a and lattice site m occupied by subpopulation b.

Similarly, we consider the change in occupancy for the lattice pair $i,\ i+1$ and rewrite the distribution functions in terms of the correlation function, using the Kirkwood superposition approximation to truncate the system of equations. For the case where the lattice sites are nearest-neighbours occupied by the same subpopulation we obtain

$$
\begin{aligned}
\frac{\mathrm{d}F_i^{G,G}(\Delta)}{\mathrm{d}t} =& - F_i^{G,G}(\Delta)\left[\frac{1}{C_i^G}\frac{\mathrm{d}C_i^G}{\mathrm{d}t} + \frac{1}{C_{i+1}^G}\frac{\mathrm{d}C_{i+1}^G}{\mathrm{d}t}\right] \\
&+ \frac{P_m^G}{2}\Bigg[\frac{C_{i-1}^G}{C_i^G}F_{i-1}^{G,G}(2\Delta) + \frac{C_{i+2}^G}{C_{i+1}^G}F_i^{G,G}(2\Delta) - 2F_L^{G,G}(\Delta) \\
&+ F_i^{G,G}(\Delta)\Big\{C_{i-1}^B F_{i-1}^{B,G}(\Delta)F_{i-1}^{B,G}(2\Delta) + C_{i+2}^B F_i^{G,B}(2\Delta)F_{i+1}^{G,B}(\Delta)\Big\} \\
&- \frac{C_{i+1}^B C_{i+2}^G}{C_{i+1}^G}F_i^{G,B}(\Delta)F_i^{G,G}(2\Delta)F_{i+1}^{B,G}(\Delta) \\
&- \frac{C_{i-1}^G C_i^B}{C_i^G}F_{i-1}^{G,B}(\Delta)F_{i-1}^{G,G}(2\Delta)F_i^{B,G}(\Delta)\Bigg] \\
&+ \frac{P_p^G}{2}\Bigg[\frac{1}{C_i^G} + \frac{1}{C_{i+1}^G} - 2F_i^{G,G}(\Delta) - \frac{C_i^B}{C_i^G}F_i^{B,G}(\Delta) - \frac{C_{i+1}^B}{C_{i+1}^G}F_i^{G,B}(\Delta) \\
&+ \frac{C_{i-1}^G F_{i-1}^{G,G}(2\Delta)}{C_i^G\left(1 - C_i^G - C_i^B\right)} \times \\
&\Big\{1 - C_i^G F_{i-1}^{G,G}(\Delta) - C_i^B F_{i-1}^{G,B}(\Delta)\Big\}\Big\{1 - C_i^G F_i^{G,G}(\Delta) - C_i^B F_i^{B,G}(\Delta)\Big\} \\
&+ \frac{C_{i+2}^G F_i^{G,G}(2\Delta)}{C_{i+1}^G\left(1 - C_{i+1}^G - C_{i+1}^B\right)} \times \\
&\Big\{1 - C_{i+1}^G F_i^{G,G}(\Delta) - C_{i+1}^B F_i^{G,B}(\Delta)\Big\}\Big\{1 - C_{i+1}^G F_{i+1}^{G,G}(\Delta) - C_{i+1}^B F_{i+1}^{B,G}(\Delta)\Big\}\Bigg] \\
&- 2P_d^G F_i^{G,G}(\Delta).
\end{aligned}
$$

We consider the same approach for the case where nearest-neighbour lattices sites are occupied by distinct subpopulations and obtain

$$
\begin{aligned}
\frac{\mathrm{d}F_i^{G,B}(\Delta)}{\mathrm{d}t} = & -F_i^{G,B}(\Delta)\left[\frac{1}{C_i^G}\frac{\mathrm{d}C_i^G}{\mathrm{d}t}+\frac{1}{C_{i+1}^B}\frac{\mathrm{d}C_{i+1}^B}{\mathrm{d}t}\right]+\frac{P_m^G}{2}\Bigg[\frac{C_{i-1}^G}{C_i^G}F_{i-1}^{G,B}(2\Delta)-F_i^{G,B}(\Delta) \\
& -\frac{C_{i-1}^G C_i^B}{C_i^G}F_{i-1}^{G,B}(\Delta)F_{i-1}^{G,B}(2\Delta)F_i^{B,B}(\Delta)+C_{i-1}^B F_{i-1}^{B,G}(\Delta)F_{i-1}^{B,B}(2\Delta)F_i^{G,B}(\Delta)\Bigg] \\
& +\frac{P_m^B}{2}\Bigg[\frac{C_{i+2}^B}{C_{i+1}^B}F_i^{G,B}(2\Delta)-F_i^{G,B}(\Delta)-\frac{C_{i+1}^G C_{i+2}^B}{C_{i+1}^B}F_i^{G,G}(\Delta)F_i^{G,B}(2\Delta)F_{i+1}^{G,B}(\Delta) \\
& +C_{i+2}^G F_i^{G,B}(\Delta)F_i^{G,G}(2\Delta)F_{i+1}^{B,G}(\Delta)\Bigg]+\frac{P_p^G}{2}\Bigg[\frac{C_{i-1}^G F_{i-1}^{G,B}(2\Delta)}{C_i^G\left(1-C_i^G-C_i^B\right)}\times \\
& \Big\{1-C_i^G F_{i-1}^{G,G}(\Delta)-C_i^B F_{i-1}^{G,B}(\Delta)\Big\}\Big\{1-C_i^G F_i^{G,B}(\Delta)-C_i^B F_i^{B,B}(\Delta)\Big\}\Bigg] \\
& \frac{P_p^B}{2}\Bigg[\frac{C_{i+2}^B F_i^{G,B}(2\Delta)}{C_{i+1}^B\left(1-C_{i+1}^G-C_{i+1}^B\right)}\times \\
& \Big\{1-C_{i+1}^G F_i^{G,G}(\Delta)-C_{i+1}^B F_i^{G,B}(\Delta)\Big\}\Big\{1-C_{i+1}^G F_{i+1}^{G,B}(\Delta)-C_{i+1}^B F_{i+1}^{B,B}(\Delta)\Big\}\Bigg] \\
& -P_d^G F_i^{G,B}(\Delta)-P_d^B F_i^{G,B}(\Delta).
\end{aligned}
$$

We then consider the case where lattice sites are separated by an arbitrary distance $r\Delta = (m-i)\Delta$ and are occupied by the same subpopulation. We obtain

$$
\begin{aligned}
\frac{\mathrm{d}F_i^{G,G}(r\Delta)}{\mathrm{d}t} &= -F_i^{G,G}(r\Delta)\Bigg[\frac{1}{C_i^G}\frac{\mathrm{d}C_i^G}{\mathrm{d}t} + \frac{1}{C_m^G}\frac{\mathrm{d}C_m^G}{\mathrm{d}t}\Bigg] + \frac{P_m^G}{2}\Bigg[\frac{C_{i-1}^G}{C_i^G}F_{i-1}^{G,G}(r\Delta+\Delta) \\
&+ \frac{C_{i+1}^G}{C_i^G}F_{i+1}^{G,G}(r\Delta-\Delta) + \frac{C_{m-1}^G}{C_m^G}F_i^{G,G}(r\Delta-\Delta) + \frac{C_{m+1}^G}{C_m^G}F_i^{G,G}(r\Delta+\Delta) \\
&- 4F_i^{G,G}(r\Delta) + F_i^{G,G}(r\Delta)\Big\{C_{i-1}^B F_{i-1}^{B,G}(\Delta)F_{i-1}^{B,G}(r\Delta+\Delta) \\
&+ C_{i+1}^B F_i^{G,B}(\Delta)F_{i+1}^{B,G}(r\Delta-\Delta) + C_{m-1}^B F_i^{G,B}(r\Delta-\Delta)F_{m-1}^{B,G}(\Delta) \\
&+ C_{m+1}^B F_i^{G,B}(r\Delta+\Delta)F_m^{G,B}(\Delta)\Big\} \\
&- \frac{C_i^B}{C_i^G}F_i^{B,G}(r\Delta)\Big\{C_{i-1}^G F_{i-1}^{G,B}(\Delta)F_{i-1}^{G,G}(r\Delta+\Delta) \\
&+ C_{i+1}^G F_i^{B,G}(\Delta)F_{i+1}^{G,G}(r\Delta-\Delta)\Big\} - \frac{C_m^B}{C_m^G}F_i^{G,B}(r\Delta)\times \\
&\Big\{C_{m-1}^G F_i^{G,G}(r\Delta-\Delta)F_{m-1}^{G,B}(\Delta) + C_{m+1}^G F_i^{G,G}(r\Delta+\Delta)F_m^{B,G}(\Delta)\Big\}\Bigg] \\
&+ \frac{P_p^G}{2}\Bigg[\frac{1 - C_i^G F_i^{G,G}(r\Delta) - C_i^B F_i^{B,G}(r\Delta)}{C_i^G\left(1 - C_i^G - C_i^B\right)}\Bigg\langle C_{i-1}^G F_{i-1}^{G,G}(r\Delta+\Delta)\times \\
&\Big\{1 - C_i^G F_{i-1}^{G,G}(\Delta) - C_i^B F_{i-1}^{G,B}(\Delta)\Big\} \\
&+ C_{i+1}^G F_{i+1}^{G,G}(r\Delta-\Delta)\Big\{1 - C_i^G F_i^{G,G}(\Delta) - C_i^B F_i^{B,G}(\Delta)\Big\}\Bigg\rangle \\
&+ \frac{1 - C_m^G F_i^{G,G}(r\Delta) - C_m^B F_i^{G,B}(r\Delta)}{C_m^G\left(1 - C_m^G - C_m^B\right)}\times \\
&\Bigg\langle C_{m-1}^G F_i^{G,G}(r\Delta-\Delta)\Big\{1 - C_m^G F_{m-1}^{G,G}(\Delta) - C_m^B F_{m-1}^{G,B}(\Delta)\Big\} \\
&+ C_{m+1}^G F_i^{G,G}(r\Delta+\Delta)\Big\{1 - C_m^G F_m^{G,G}(\Delta) - C_m^B F_m^{B,G}(\Delta)\Big\}\Bigg\rangle\Bigg] \\
&- 2P_d^G F_i^{G,G}(r\Delta).
\end{aligned}
$$

Finally, we consider the case for lattice sites separated by an arbitrary distance $r\Delta$ that are occupied by distinct subpopulations, and obtain

$$\begin{aligned}
\frac{\mathrm{d}F_i^{G,B}(r\Delta)}{\mathrm{d}t} = &- F_i^{G,B}(r\Delta)\left[\frac{1}{C_i^G}\frac{\mathrm{d}C_i^G}{\mathrm{d}t} + \frac{1}{C_m^B}\frac{\mathrm{d}C_m^B}{\mathrm{d}t}\right] + \frac{P_m^G}{2}\Bigg[\frac{C_{i-1}^G}{C_i^G}F_{i-1}^{G,B}(r\Delta+\Delta) \\
&+ \frac{C_{i+1}^G}{C_i^G}F_{i+1}^{G,B}(r\Delta-\Delta) - 2F_i^{G,B}(r\Delta) - \frac{C_i^B}{C_i^G}F_i^{B,B}(r\Delta) \times \\
&\Big\{C_{i-1}^G F_{i-1}^{G,B}(\Delta)F_{i-1}^{G,B}(r\Delta+\Delta) + C_{i+1}^G F_i^{B,G}(\Delta)F_{i+1}^{G,B}(r\Delta-\Delta)\Big\} + \\
&F_i^{G,B}(r\Delta)\Big\{C_{i-1}^B F_{i-1}^{B,G}(\Delta)F_{i-1}^{B,B}(r\Delta+\Delta) + C_{i+1}^B F_i^{G,B}(\Delta)F_{i+1}^{B,B}(r\Delta-\Delta)\Big\}\Bigg] \\
&+ \frac{P_m^B}{2}\Bigg[\frac{C_{m-1}^B}{C_m^B}F_i^{G,B}(r\Delta-\Delta) + \frac{C_{m+1}^B}{C_m^B}F_i^{G,B}(r\Delta+\Delta) - 2F_i^{G,B}(r\Delta) \\
&- \frac{C_m^G}{C_m^B}F_i^{G,G}(r\Delta)\Big\{C_{m-1}^B F_i^{G,B}(r\Delta-\Delta)F_{m-1}^{B,G}(\Delta) \\
&+C_{m+1}^B F_i^{G,B}(r\Delta+\Delta)F_m^{G,B}(\Delta)\Big\} + F_i^{G,B}(r\Delta)\times \\
&\Big\{C_{G,m-1}F_i^{G,G}(r\Delta-\Delta)F_{m-1}^{G,B}(\Delta) + C_{m+1}^G F_i^{G,G}(r\Delta+\Delta)F_m^{B,G}(\Delta)\Big\}\Bigg] \\
&+ \frac{P_p^G}{2}\Bigg[\frac{1 - C_i^G F_i^{G,B}(r\Delta) - C_i^B F_i^{B,B}(r\Delta)}{C_i^G\left(1 - C_i^G - C_i^B\right)} \times \\
&\Big\langle C_{i-1}^G F_{i-1}^{G,B}(r\Delta+\Delta)\Big\{1 - C_i^G F_{i-1}^{G,G}(\Delta) - C_i^B F_{i-1}^{G,B}(\Delta)\Big\} \\
&+ C_{i+1}^G F_{i+1}^{G,B}(r\Delta-\Delta)\Big\{1 - C_i^G F_i^{G,G}(\Delta) - C_i^B F_i^{B,G}(\Delta)\Big\}\Big\rangle\Bigg] \\
&+ \frac{P_p^B}{2}\Bigg[\frac{1 - C_m^G F_i^{G,G}(r\Delta) - C_m^B F_i^{G,B}(r\Delta)}{C_m^B\left(1 - C_m^G - C_m^B\right)} \times \\
&\Big\langle C_{m-1}^B F_{G,B}^i(r\Delta-\Delta)\Big\{1 - C_m^G F_{m-1}^{B,G}(\Delta) - C_m^B F_{m-1}^{B,B}(\Delta)\Big\} \\
&+ C_{m+1}^B F_i^{G,B}(r\Delta+\Delta)\left\{1 - C_m^G F_m^{G,B}(\Delta) - C_m^B F_m^{B,B}(\Delta)\right\}\Big\rangle\Bigg] \\
&- P_d^G F_i^{G,B}(r\Delta) - P_d^B F_i^{G,B}(r\Delta).
\end{aligned}$$

6.5.3 Two-dimensional moment dynamics model

We now consider the two-dimensional case, where lattice sites are indexed by (i, j), and consider all possible motility, proliferation and death events, in terms of the one- and two-point distribution functions. Rewriting in terms of the correlation function, we obtain

$$
\begin{aligned}
\frac{\mathrm{d}C^G_{(i,j)}}{\mathrm{d}t} =& \frac{P^G_m}{4}\Bigg[C^G_{(i-1,j)} + C^G_{(i+1,j)} + C^G_{(i,j-1)} + C^G_{(i,j+1)} - 4C^G_{(i,j)} \\
&- C^B_{(i,j)}\Big\{C^G_{(i-1,j)}F^{G,B}_{(i-1,j)}(\Delta,0) + C^G_{(i+1,j)}F^{B,G}_{(i,j)}(\Delta,0) \\
&+ C^G_{(i,j-1)}F^{G,B}_{(i,j-1)}(0,\Delta) + C^G_{(i,j+1)}F^{B,G}_{(i,j)}(0,\Delta)\Big\} \\
&+ C^G_{(i,j)}\Big\{C^B_{(i-1,j)}F^{B,G}_{(i-1,j)}(\Delta,0) + C^B_{(i+1,j)}F^{G,B}_{(i,j)}(\Delta,0) \\
&+ C^B_{(i,j-1)}F^{B,G}_{(i,j-1)}(0,\Delta) + C^B_{(i,j+1)}F^{G,B}_{(i,j)}(0,\Delta)\Big\}\Bigg] + \\
&\frac{P^G_p}{4}\Bigg[C^G_{(i-1,j)}\Big\{1 - C^G_{(i,j)}F^{G,G}_{(i-1,j)}(\Delta,0) - C^B_{(i,j)}F^{G,B}_{(i-1,j)}(\Delta,0)\Big\} \\
&+ C^G_{(i+1,j)}\Big\{1 - C^G_{(i,j)}F^{G,G}_{(i,j)}(\Delta,0) - C^B_{(i,j)}F^{B,G}_{(i,j)}(\Delta,0)\Big\} \\
&+ C^G_{(i,j-1)}\Big\{1 - C^G_{(i,j)}F^{G,G}_{(i,j-1)}(0,\Delta) - C^B_{(i,j)}F^{G,B}_{(i,j-1)}(0,\Delta)\Big\} \\
&+ C^G_{(i,j+1)}\Big\{1 - C^G_{(i,j)}F^{G,G}_{(i,j)}(0,\Delta) - C^B_{(i,j)}F^{B,G}_{(i,j)}(0,\Delta)\Big\}\Bigg] - P^G_d C^G_{(i,j)},
\end{aligned}
$$

where $F^{a,b}_{(i,j)}(r\Delta, s\Delta)$ corresponds to the correlation between site (i, j) and site (m, n), such that $r\Delta = (m - i)\Delta$ in the x direction and $s\Delta = (n - j)\Delta$ in the y direction, and where site (i, j) is occupied by subpopulaton a and site (m, n) is occupied by subpopulation b. We now consider the change in the occupancy in nearest-neighbour lattice pairs, occupied by the same subpopulation and rewrite the three-point distribution functions using the Kirkwood superposition approximation and the correlation function, and obtain

$$
\begin{aligned}
&\frac{\mathrm{d}F^{G,G}_{(i,j)}(0,\Delta)}{\mathrm{d}t} = -F^{G,G}_{(i,j)}(0,\Delta)\Bigg[\frac{1}{C^G_{(i,j)}}\frac{\mathrm{d}C^G_{(i,j)}}{\mathrm{d}t} + \frac{1}{C^G_{(i,j+1)}}\frac{\mathrm{d}C^G_{(i,j+1)}}{\mathrm{d}t}\Bigg] + \frac{P^G_m}{4}\Bigg[\frac{C^G_{(i-1,j)}}{C^G_{(i,j)}}F^{G,G}_{(i-1,j)}(\Delta,\Delta) \\
&+ \frac{C^G_{(i+1,j)}}{C^G_{(i,j)}}F^{G,G}_{(i+1,j)}(-\Delta,\Delta) + \frac{C^G_{(i,j-1)}}{C^G_{(i,j)}}F^{G,G}_{(i,j-1)}(0,2\Delta) + \frac{C^G_{i-1,j+1}}{C^G_{(i,j+1)}}F^{G,G}_{i-1,j+1}(\Delta,-\Delta) + \frac{C^G_{i+1,j+1}}{C^G_{(i,j+1)}}F^{G,G}_{(i,j)}(\Delta,\Delta) \\
&+ \frac{C^G_{i,j+2}}{C^G_{(i,j+1)}}F^{G,G}_{(i,j)}(0,2\Delta) - 6F^{G,G}_{(i,j)}(0,\Delta) + \frac{C^B_{(i,j)}}{C^G_{(i,j)}}F^{B,G}_{(i,j)}(0,\Delta)\Big\{C^G_{(i-1,j)}F^{G,B}_{(i-1,j)}(\Delta,0)F^{G,G}_{(i-1,j)}(\Delta,\Delta) \\
&+ C^G_{(i+1,j)}F^{B,G}_{(i,j)}(\Delta,0)F^{G,G}_{(i+1,j)}(-\Delta,\Delta) + C^G_{(i,j-1)}F^{G,B}_{(i,j-1)}(0,\Delta)F^{G,G}_{(i,j-1)}(0,2\Delta)\Big\} \\
&+ \frac{C^B_{(i,j+1)}}{C^G_{(i,j+1)}}F^{G,B}_{(i,j)}(0,\Delta)\Big\{C^G_{i-1,j+1}F^{G,G}_{i-1,j+1}(\Delta,-\Delta)F^{G,B}_{i-1,j+1}(\Delta,0) + C^G_{i+1,j+1}F^{G,G}_{(i,j)}(\Delta,\Delta)F^{B,G}_{(i,j+1)}(\Delta,0) \\
&+ C^G_{i,j+2}F^{G,G}_{(i,j)}(0,2)F^{B,G}_{(i,j+1)}(0,1)\Big\} - F^{G,G}_{(i,j)}(0,\Delta)\Big\{C^B_{(i-1,j)}F^{B,G}_{(i-1,j)}(\Delta,0)F^{B,G}_{(i-1,j)}(\Delta,\Delta) \\
&+ C^B_{(i+1,j)}F^{G,B}_{(i,j)}(\Delta,0)F^{B,G}_{(i+1,j)}(-\Delta,\Delta) + C^B_{(i,j-1)}F^{B,G}_{(i,j-1)}(0,\Delta)F^{B,G}_{(i,j-1)}(0,2\Delta) + C^B_{i-1,j+1}F^{B,G}_{i-1,j+1}(\Delta,-\Delta)F^{B,G}_{i-1,j+1}(\Delta,0) \\
&+ C^B_{i+1,j+1}F^{G,B}_{(i,j)}(\Delta,\Delta)F^{G,B}_{(i,j+1)}(\Delta,0) + C^B_{i,j+2}F^{B,G}_{(i,j)}(0,2\Delta)F^{G,B}_{(i,j+1)}(0,\Delta)\Big\}\Bigg] + \frac{P^G_p}{4}\Bigg[\frac{1}{C^G_{(i,j)}} + \frac{1}{C^G_{(i,j+1)}} - 2F^{G,G}_{(i,j)}(\Delta,0) \\
&- \frac{C^B_{(i,j)}}{C^G_{(i,j)}}F^{B,G}_{(i,j)}(0,\Delta) - \frac{C^B_{(i,j+1)}}{C^G_{(i,j+1)}}F^{G,B}_{(i,j)}(0,\Delta) + \frac{1 - C^G_{(i,j)}F^{G,G}_{(i,j)}(0,\Delta) - C^B_{(i,j)}F^{B,G}_{(i,j)}(0,\Delta)}{C^G_{(i,j)}\Big(1 - C^G_{(i,j)} - C^B_{(i,j)}\Big)}\Big\langle C^G_{(i-1,j)}F^{G,G}_{(i-1,j)}(\Delta,\Delta)\times \\
&\Big\{1 - C^G_{(i,j)}F^{G,G}_{(i-1,j)}(\Delta,0)C^B_{(i,j)}F^{G,B}_{(i-1,j)}(\Delta,0)\Big\} + C^G_{(i+1,j)}F^{G,G}_{(i+1,j)}(-\Delta,\Delta)\Big\{1 - C^G_{(i,j)}F^{G,G}_{(i,j)}(\Delta,0) - C^B_{(i,j)}F^{B,G}_{(i,j)}(\Delta,0)\Big] \\
&+ C^G_{(i,j-1)}F^{G,G}_{(i,j-1)}(0,2\Delta)\Big\{1 - C^G_{(i,j)}F^{G,G}_{(i,j-1)}(0,\Delta) - C^B_{(i,j)}F^{G,B}_{(i,j-1)}(0,\Delta)\Big\}\Big\rangle + \frac{1 - C^G_{(i,j+1)}F^{G,G}_{(i,j)}(0,\Delta) - C^B_{(i,j+1)}F^{G,B}_{(i,j)}(0,\Delta)}{C^G_{(i,j+1)}\Big(1 - C^G_{(i,j+1)} - C^B_{(i,j+1)}\Big)}\times \\
&\Big\langle C^G_{i-1,j+1}F^{G,G}_{i-1,j+1}(\Delta,-\Delta)\Big\{1 - C^G_{(i,j+1)}F^{G,G}_{i-1,j+1}(\Delta,0) - C^B_{(i,j+1)}F^{G,B}_{i-1,j+1}(\Delta,0)\Big\} + C^G_{i+1,j+1}F^{G,G}_{(i,j)}(\Delta,\Delta)\times \\
&\Big\{1 - C^G_{(i,j+1)}F^{G,G}_{(i,j+1)}(\Delta,0) - C^B_{(i,j+1)}F^{B,G}_{(i,j+1)}(\Delta,0)\Big\} \\
&+ C^G_{i,j+2}F^{G,G}_{(i,j)}(0,2\Delta)\Big\{1 - C^G_{(i,j+1)}F^{G,G}_{(i,j+1)}(0,\Delta) - C^B_{(i,j+1)}F^{B,G}_{(i,j+1)}(0,\Delta)\Big\}\Big\rangle\Bigg] - 2P^G_d F^{G,G}_{(i,j)}(0,\Delta),
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{\mathrm{d}F^{G,G}_{(i,j)}(\Delta,0)}{\mathrm{d}t} = -F^{G,G}_{(i,j)}(\Delta,0)\Bigg[\frac{1}{C^G_{(i,j)}}\frac{\mathrm{d}C^G_{(i,j)}}{\mathrm{d}t} + \frac{1}{C^G_{(i+1,j)}}\frac{\mathrm{d}C^G_{(i+1,j)}}{\mathrm{d}t}\Bigg] + \frac{P^G_m}{4}\Bigg[\frac{C^G_{(i,j-1)}}{C^G_{(i,j)}}F^{G,G}_{(i,j-1)}(\Delta,\Delta) \\
&+ \frac{C^G_{(i,j+1)}}{C^G_{(i,j)}}F^{G,G}_{(i,j+1)}(\Delta,-\Delta) + \frac{C^G_{(i-1,j)}}{C^G_{(i,j)}}F^{G,G}_{(i-1,j)}(2\Delta,0) + \frac{C^G_{i+1,j-1}}{C^G_{(i+1,j)}}F^{G,G}_{i+1,j-1}(-\Delta,\Delta) + \frac{C^G_{i+1,j+1}}{C^G_{(i+1,j)}}F^{G,G}_{(i,j)}(\Delta,\Delta) \\
&+ \frac{C^G_{i+2,j}}{C^G_{(i+1,j)}}F^{G,G}_{(i,j)}(2\Delta,0) - 6F^{G,G}_{(i,j)}(\Delta,0) + \frac{C^B_{(i,j)}}{C^G_{(i,j)}}F^{B,G}_{(i,j)}(\Delta,0)\Big[C^G_{(i,j-1)}F^{G,B}_{(i,j-1)}(0,\Delta)F^{G,G}_{(i,j-1)}(\Delta,\Delta) \\
&+ C^G_{(i,j+1)}F^{B,G}_{(i,j)}(0,\Delta)F^{G,G}_{(i,j+1)}(\Delta,-\Delta) + C^G_{(i-1,j)}F^{G,B}_{(i-1,j)}(\Delta,0)F^{G,G}_{(i-1,j)}(2\Delta,0)\Big] \\
&+ \frac{C^B_{(i+1,j)}}{C^G_{(i+1,j)}}F^{G,B}_{(i,j)}(\Delta,0)\Big[C^G_{i+1,j-1}F^{G,G}_{i+1,j-1}(-\Delta,\Delta)F^{G,B}_{i+1,j-1}(0,\Delta) + C^G_{i+1,j+1}F^{G,G}_{(i,j)}(\Delta,\Delta)F^{B,G}_{(i+1,j)}(0,\Delta) \\
&+ C^G_{i+2,j}F^{G,G}_{(i,j)}(2,0)F^{B,G}_{(i+1,j)}(1,0)\Big] - F^{G,G}_{(i,j)}(\Delta,0)\Big[C^B_{(i,j-1)}F^{B,G}_{(i,j-1)}(0,\Delta)F^{B,G}_{(i,j-1)}(\Delta,\Delta) \\
&+C^B_{(i,j+1)}F^{G,B}_{(i,j)}(0,\Delta)F^{B,G}_{(i,j+1)}(\Delta,-\Delta) + C^B_{(i-1,j)}F^{B,G}_{(i-1,j)}(\Delta,0)F^{B,G}_{(i-1,j)}(2\Delta,0) + C^B_{i+1,j-1}F^{B,G}_{i+1,j-1}(-\Delta,\Delta)F^{B,G}_{i+1,j-1}(0,\Delta)+ \\
&C^B_{i+1,j+1}F^{G,B}_{(i,j)}(\Delta,\Delta)F^{G,B}_{(i+1,j)}(0,\Delta) + C^B_{i+2,j}F^{B,G}_{(i,j)}(2\Delta,0)F^{G,B}_{(i+1,j)}(\Delta,0)\Big]\Bigg] + \frac{P^G_p}{4}\Bigg[\frac{1}{C^G_{(i,j)}} + \frac{1}{C^G_{(i+1,j)}} - 2F^{G,G}_{(i,j)}(0,\Delta) \\
&- \frac{C^B_{(i,j)}}{C^G_{(i,j)}}F^{B,G}_{(i,j)}(\Delta,0) - \frac{C^B_{(i+1,j)}}{C^G_{(i+1,j)}}F^{G,B}_{(i,j)}(\Delta,0) + \frac{1 - C^G_{(i,j)}F^{G,G}_{(i,j)}(\Delta,0) - C^B_{(i,j)}F^{B,G}_{(i,j)}(\Delta,0)}{C^G_{(i,j)}\Big(1 - C^G_{(i,j)} - C^B_{(i,j)}\Big)}\Big\rangle C^G_{(i,j-1)}F^{G,G}_{(i,j-1)}(\Delta,\Delta)\times \\
&\Big\{1 - C^G_{(i,j)}F^{G,G}_{(i,j-1)}(0,\Delta)C^B_{(i,j)}F^{G,B}_{(i,j-1)}(0,\Delta)\Big\} + C^G_{(i,j+1)}F^{G,G}_{(i,j+1)}(\Delta,-\Delta)\Big\{1 - C^G_{(i,j)}F^{G,G}_{(i,j)}(0,\Delta) - C^B_{(i,j)}F^{B,G}_{(i,j)}(0,\Delta)\Big\} \\
&+ C^G_{(i-1,j)}F^{G,G}_{(i-1,j)}(2\Delta,0)\Big\{1 - C^G_{(i,j)}F^{G,G}_{(i-1,j)}(\Delta,0) - C^B_{(i,j)}F^{G,B}_{(i-1,j)}(\Delta,0)\Big\}\Big\rangle + \frac{1 - C^G_{(i+1,j)}F^{G,G}_{(i,j)}(\Delta,0) - C^B_{(i+1,j)}F^{G,B}_{(i,j)}(\Delta,0)}{C^G_{(i+1,j)}\Big(1 - C^G_{(i+1,j)} - C^B_{(i+1,j)}\Big)}\times \\
&\Big\langle C^G_{i+1,j-1}F^{G,G}_{i+1,j-1}(-\Delta,\Delta)\Big\{1 - C^G_{(i+1,j)}F^{G,G}_{i+1,j-1}(0,\Delta) - C^B_{(i+1,j)}F^{G,B}_{i+1,j-1}(0,\Delta)\Big\} + C^G_{i+1,j+1}F^{G,G}_{(i,j)}(\Delta,\Delta)\times \\
&\Big\{1 - C^G_{(i+1,j)}F^{G,G}_{(i+1,j)}(0,\Delta) - C^B_{(i+1,j)}F^{B,G}_{(i+1,j)}(0,\Delta)\Big\} \\
&+ C^G_{i+2,j}F^{G,G}_{(i,j)}(2\Delta,0)\Big\{1 - C^G_{(i+1,j)}F^{G,G}_{(i+1,j)}(\Delta,0) - C^B_{(i+1,j)}F^{B,G}_{(i+1,j)}(\Delta,0)\Big\}\Big\rangle\Bigg] - 2P^G_d F^{G,G}_{(i,j)}(\Delta,0).
\end{aligned}
$$

Using the same approach for nearest-neighbour lattice sites that are occupied by distinct subpopulations, we obtain

$$
\begin{aligned}
&\frac{\mathrm{d}F^{G,B}_{(i,j)}(0,\Delta)}{\mathrm{d}t} = -F^{G,B}_{(i,j)}(0,\Delta)\left[\frac{1}{C^G_{(i,j)}}\frac{\mathrm{d}C^G_{(i,j)}}{\mathrm{d}t} + \frac{1}{C^B_{(i,j+1)}}\frac{\mathrm{d}C^B_{(i,j+1)}}{\mathrm{d}t}\right] + \frac{P^G_m}{4}\Bigg[\frac{C^G_{(i-1,j)}}{C^G_{(i,j)}}F^{G,B}_{(i-1,j)}(\Delta,\Delta) \\
&+ \frac{C^G_{(i+1,j)}}{C^G_{(i,j)}}F^{G,B}_{(i+1,j)}(-\Delta,\Delta) - 3F^{G,B}_{(i,j)}(0,\Delta) + \frac{C^B_{(i,j-1)}}{C^G_{(i,j)}}F^{G,G}_{(i,j-1)}(0,2\Delta) \\
&- \frac{C^B_{(i,j)}}{C^G_{(i,j)}}F^{B,B}_{(i,j)}(0,\Delta)\Big\{C^G_{(i-1,j)}F^{G,B}_{(i-1,j)}(\Delta,0)F^{G,B}_{(i-1,j)}(\Delta,\Delta) + C^G_{(i+1,j)}F^{B,G}_{(i,j)}(\Delta,0)F^{G,B}_{(i+1,j)}(-\Delta,\Delta) \\
&+ C^G_{(i,j-1)}F^{G,B}_{(i,j-1)}(0,\Delta)F^{G,B}_{(i,j-1)}(0,2\Delta)\Big\} + F^{G,B}_{(i,j)}(0,\Delta)\Big\{C^B_{(i-1,j)}F^{B,G}_{(i-1,j)}(\Delta,0)F^{B,B}_{(i-1,j)}(\Delta,\Delta) \\
&+ C^B_{(i+1,j)}F^{G,B}_{(i,j)}(\Delta,0)F^{B,B}_{(i+1,j)}(-\Delta,\Delta) + C^B_{(i,j-1)}F^{B,G}_{(i,j-1)}(0,\Delta)F^{B,B}_{(i,j-1)}(0,2\Delta)\Big\}\Bigg] + \\
&\frac{P^B_m}{4}\Bigg[\frac{C^B_{i-1,j+1}}{C^B_{(i,j+1)}}F^{B,G}_{i-1,j+1}(\Delta,-\Delta) + \frac{C^B_{i+1,j+1}}{C^B_{(i,j+1)}}F^{G,B}_{(i,j)}(\Delta,\Delta) + \frac{C^B_{i,j+2}}{C^B_{(i,j+1)}}F^{G,B}_{(i,j)}(0,2\Delta) - 3F^{G,B}_{(i,j)}(0,\Delta) \\
&+ \frac{C^G_{(i,j+1)}}{C^B_{(i,j+1)}}F^{G,G}_{(i,j)}(0,\Delta)\Big\{C^B_{i-1,j+1}F^{B,G}_{i-1,j+1}(\Delta,-\Delta)F^{B,G}_{i-1,j+1}(\Delta,0) + C^B_{i+1,j+1}F^{G,B}_{(i,j)}(\Delta,\Delta)F^{G,B}_{(i,j+1)}(\Delta,0) \\
&+ C^B_{i,j+2}F^{G,B}_{(i,j)}(0,2\Delta)F^{G,B}_{(i,j+1)}(0,\Delta)\Big\} + F^{G,B}_{(i,j)}(0,\Delta)\Big\{C^G_{i-1,j+1}F^{G,G}_{i-1,j+1}(\Delta,-\Delta)F^{G,B}_{i-1,j+1}(\Delta,0) \\
&+ C^G_{i+1,j+1}F^{G,G}_{(i,j)}(\Delta,\Delta)F^{B,G}_{(i,j+1)}(\Delta,0) + C^G_{i,j+2}F^{G,G}_{(i,j)}(0,2\Delta)F^{B,G}_{(i,j+1)}(0,\Delta)\Big\}\Bigg] + \\
&\frac{P^G_p}{4}\Bigg[\frac{1 - C^G_{(i,j)}F^{G,B}_{(i,j)}(0,\Delta) - C^B_{(i,j)}F^{B,B}_{(i,j)}(0,\Delta)}{C^G_{(i,j)}\left(1 - C^G_{(i,j)} - C^B_{(i,j)}\right)}\Bigg\langle C^G_{(i-1,j)}F^{G,B}_{(i-1,j)}(\Delta,\Delta)\Big\{1 - C^G_{(i,j)}F^{G,G}_{(i-1,j)}(\Delta,0) - C^B_{(i,j)}F^{G,B}_{(i-1,j)}(\Delta,0)\Big\} \\
&+ C^G_{(i+1,j)}F^{G,B}_{(i+1,j)}(-\Delta,\Delta)\Big\{1 - C^G_{(i,j)}F^{G,B}_{(i,j)}(0,\Delta) - C^B_{(i,j)}F^{B,B}_{(i,j)}(0,\Delta)\Big\} \\
&+ C^G_{(i,j-1)}F^{G,B}_{(i,j-1)}(0,2\Delta)\Big\{1 - C^G_{(i,j)}F^{G,G}_{(i,j-1)}(0,\Delta) - C^B_{(i,j)}F^{G,B}_{(i,j-1)}(0,\Delta)\Big\}\Bigg\rangle\Bigg] + \\
&\frac{P^B_p}{4}\Bigg[\frac{1 - C^G_{(i,j+1)}F^{G,G}_{(i,j)}(0,\Delta) - C^B_{(i,j+1)}F^{G,B}_{(i,j)}(0,\Delta)}{C^B_{(i,j+1)}\left(1 - C^G_{(i,j+1)} - C^B_{(i,j+1)}\right)}\Bigg\langle C^B_{i-1,j+1}F^{B,G}_{i-1,j+1}(\Delta,-\Delta)\times \\
&\Big\{1 - C^G_{(i,j+1)}F^{B,G}_{i-1,j+1}(\Delta,0) - C^B_{(i,j+1)}F^{B,B}_{i-1,j+1}(\Delta,0)\Big\} + C^B_{i+1,j+1}F^{G,B}_{(i,j)}(\Delta,\Delta)\times \\
&\Big\{1 - C^G_{(i,j+1)}F^{G,B}_{(i,j+1)}(\Delta,0) - C^B_{(i,j+1)}F^{B,B}_{(i,j+1)}(\Delta,0)\Big\} + C^B_{i,j+2}F^{G,B}_{(i,j)}(0,2\Delta)\Big\{1 - C^G_{(i,j+1)}F^{G,B}_{(i,j+1)}(0,\Delta) - \\
&C^B_{(i,j+1)}F^{B,B}_{(i,j+1)}(0,\Delta)\Big\}\Bigg\rangle\Bigg] - P^G_d F^{G,B}_{(i,j)}(0,\Delta) - P^B_d F^{G,B}_{(i,j)}(0,\Delta),
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{\mathrm{d}F^{G,B}_{(i,j)}(\Delta,0)}{\mathrm{d}t} = -F^{G,B}_{(i,j)}(\Delta,0)\left[\frac{1}{C^G_{(i,j)}}\frac{\mathrm{d}C^G_{(i,j)}}{\mathrm{d}t} + \frac{1}{C^B_{(i+1,j)}}\frac{\mathrm{d}C^B_{(i+1,j)}}{\mathrm{d}t}\right] + \frac{P^G_m}{4}\Bigg[\frac{C^G_{(i,j-1)}}{C^G_{(i,j)}}F^{G,B}_{(i,j-1)}(\Delta,\Delta) \\
&+ \frac{C^G_{(i,j+1)}}{C^G_{(i,j)}}F^{G,B}_{(i,j+1)}(\Delta,-\Delta) - 3F^{G,B}_{(i,j)}(\Delta,0) + \frac{C^B_{(i-1,j)}}{C^G_{(i,j)}}F^{G,G}_{(i-1,j)}(2\Delta,0) \\
&- \frac{C^B_{(i,j)}}{C^G_{(i,j)}}F^{B,B}_{(i,j)}(\Delta,0)\Big\{C^G_{(i,j-1)}F^{G,B}_{(i,j-1)}(0,\Delta)F^{G,B}_{(i,j-1)}(\Delta,\Delta) + C^G_{(i,j+1)}F^{B,G}_{(i,j)}(0,\Delta)F^{G,B}_{(i,j+1)}(\Delta,-\Delta) \\
&+ C^G_{(i-1,j)}F^{G,B}_{(i-1,j)}(\Delta,0)F^{G,B}_{(i-1,j)}(2\Delta,0)\Big\} + F^{G,B}_{(i,j)}(\Delta,0)\Big\{C^B_{(i,j-1)}F^{B,G}_{(i,j-1)}(0,\Delta)F^{B,B}_{(i,j-1)}(\Delta,\Delta) \\
&+ C^B_{(i,j+1)}F^{G,B}_{(i,j)}(0,\Delta)F^{B,B}_{(i,j+1)}(\Delta,-\Delta) + C^B_{(i-1,j)}F^{B,G}_{(i-1,j)}(\Delta,0)F^{B,B}_{(i-1,j)}(2\Delta,0)\Big\}\Bigg] + \\
&\frac{P^B_m}{4}\Bigg[\frac{C^B_{i+1,j-1}}{C^B_{(i+1,j)}}F^{B,G}_{i+1,j-1}(-\Delta,\Delta) + \frac{C^B_{i+1,j+1}}{C^B_{(i+1,j)}}F^{G,B}_{(i,j)}(\Delta,\Delta) + \frac{C^B_{i+2,j}}{C^B_{(i+1,j)}}F^{G,B}_{(i,j)}(2\Delta,0) - 3F^{G,B}_{(i,j)}(\Delta,0) \\
&+ \frac{C^G_{(i+1,j)}}{C^B_{(i+1,j)}}F^{G,G}_{(i,j)}(\Delta,0)\Big\{C^B_{i+1,j-1}F^{B,G}_{i+1,j-1}(-\Delta,\Delta)F^{B,G}_{i+1,j-1}(0,\Delta) + C^B_{i+1,j+1}F^{G,B}_{(i,j)}(\Delta,\Delta)F^{G,B}_{(i+1,j)}(0,\Delta) \\
&+C^B_{i+2,j}F^{G,B}_{(i,j)}(2\Delta,0)F^{G,B}_{(i+1,j)}(\Delta,0)\Big\} + F^{G,B}_{(i,j)}(\Delta,0)\Big\{C^G_{i+1,j-1}F^{G,G}_{i+1,j-1}(-\Delta,\Delta)F^{G,B}_{i+1,j-1}(0,\Delta) \\
&+ C^G_{i+1,j+1}F^{G,G}_{(i,j)}(\Delta,\Delta)F^{B,G}_{(i+1,j)}(0,\Delta) + C^G_{i+2,j}F^{G,G}_{(i,j)}(2\Delta,0)F^{B,G}_{(i+1,j)}(\Delta,0)\Big\}\Bigg] + \\
&\frac{P^G_p}{4}\Bigg[\frac{1 - C^G_{(i,j)}F^{G,B}_{(i,j)}(\Delta,0) - C^B_{(i,j)}F^{B,B}_{(i,j)}(\Delta,0)}{C^G_{(i,j)}\left(1 - C^G_{(i,j)} - C^B_{(i,j)}\right)}\Big\langle C^G_{(i,j-1)}F^{G,B}_{(i,j-1)}(\Delta,\Delta)\Big\{1 - C^G_{(i,j)}F^{G,G}_{(i,j-1)}(0,\Delta) - C^B_{(i,j)}F^{G,B}_{(i,j-1)}(0,\Delta)\Big\} \\
&+ C^G_{(i,j+1)}F^{G,B}_{(i,j+1)}(\Delta,-\Delta)\Big\{1 - C^G_{(i,j)}F^{G,B}_{(i,j)}(\Delta,0) - C^B_{(i,j)}F^{B,B}_{(i,j)}(\Delta,0)\Big\} \\
&+ C^G_{(i-1,j)}F^{G,B}_{(i-1,j)}(2\Delta,0)\Big\{1 - C^G_{(i,j)}F^{G,G}_{(i-1,j)}(\Delta,0) - C^B_{(i,j)}F^{G,B}_{(i-1,j)}(\Delta,0)\Big\}\Big\rangle\Bigg] + \\
&\frac{P^B_p}{4}\Bigg[\frac{1 - C^G_{(i+1,j)}F^{G,G}_{(i,j)}(\Delta,0) - C^B_{(i+1,j)}F^{G,B}_{(i,j)}(\Delta,0)}{C^B_{(i+1,j)}\left(1 - C^G_{(i+1,j)} - C^B_{(i+1,j)}\right)}\Big\langle C^B_{i+1,j-1}F^{B,G}_{i+1,j-1}(-\Delta,\Delta)\times \\
&\Big\{1 - C^G_{(i+1,j)}F^{B,G}_{i+1,j-1}(0,\Delta) - C^B_{(i+1,j)}F^{B,B}_{i+1,j-1}(0,\Delta)\Big\} + C^B_{i+1,j+1}F^{G,B}_{(i,j)}(\Delta,\Delta)\times \\
&\Big\{1 - C^G_{(i+1,j)}F^{G,B}_{(i+1,j)}(0,\Delta) - C^B_{(i+1,j)}F^{B,B}_{(i+1,j)}(0,\Delta)\Big\} + C^B_{i+2,j}F^{G,B}_{(i,j)}(2\Delta,0)\Big\{1 - C^G_{(i+1,j)}F^{G,B}_{(i+1,j)}(\Delta,0) - \\
&C^B_{(i+1,j)}F^{B,B}_{(i+1,j)}(\Delta,0)\Big\}\Big\rangle\Bigg] - P^G_d F^{G,B}_{(i,j)}(\Delta,0) - P^B_d F^{G,B}_{(i,j)}(\Delta,0).
\end{aligned}
$$

We consider lattice pairs separated by a distance $r\Delta = (m-i)\Delta$ in the x-direction and a distance $s\Delta = (n-j)\Delta$ in the y-direction, occupied by the same subpopulation. Using the same approach, we obtain

$$
\begin{aligned}
&\frac{\mathrm{d}F^{G,G}_{(i,j)}(r\Delta,s\Delta)}{\mathrm{d}t} = -F^{G,G}_{(i,j)}(r\Delta,s\Delta)\left[\frac{1}{C^G_{(i,j)}}\frac{\mathrm{d}C^G_{(i,j)}}{\mathrm{d}t} + \frac{1}{C^G_{(m,n)}}\frac{\mathrm{d}C^G_{(m,n)}}{\mathrm{d}t}\right] + \frac{P^G_m}{4}\Bigg[\frac{C^G_{(i-1,j)}}{C^G_{(i,j)}}F^{G,G}_{(i-1,j)}(r\Delta+\Delta,s\Delta)\\
&+\frac{C^G_{(i+1,j)}}{C^G_{(i,j)}}F^{G,G}_{(i+1,j)}(r\Delta-\Delta,s\Delta) + \frac{C^G_{(i,j-1)}}{C^G_{(i,j)}}F^{G,G}_{(i,j-1)}(r\Delta,s\Delta+\Delta) + \frac{C^G_{(i,j+1)}}{C^G_{(i,j)}}F^{G,G}_{(i,j+1)}(r\Delta,s\Delta-\Delta)\\
&+\frac{C^G_{(m-1,n)}}{C^G_{(m,n)}}F^{G,G}_{(i,j)}(r\Delta-\Delta,s\Delta) + \frac{C^G_{(m+1,n)}}{C^G_{(m,n)}}F^{G,G}_{(i,j)}(r\Delta+\Delta,s\Delta) + \frac{C^G_{(m,n-1)}}{C^G_{(m,n)}}F^{G,G}_{(i,j)}(r\Delta,s\Delta-\Delta)\\
&+\frac{C^G_{(m,n+1)}}{C^G_{(m,n)}}F^{G,G}_{(i,j)}(r\Delta,s\Delta+\Delta) - 8F^{G,G}_{(i,j)}(r\Delta,s\Delta) - \frac{C^B_{(i,j)}}{C^G_{(i,j)}}F^{B,G}_{(i,j)}(r\Delta,s\Delta)\times\\
&\Big\{C^G_{(i-1,j)}F^{G,B}_{(i-1,j)}(\Delta,0)F^{G,G}_{(i-1,j)}(r\Delta+\Delta,s\Delta) + C^G_{(i+1,j)}F^{B,G}_{(i,j)}(\Delta,0)F^{G,G}_{(i+1,j)}(r\Delta-\Delta,s\Delta)\\
&+C^G_{(i,j-1)}F^{G,B}_{(i,j-1)}(0,\Delta)F^{G,G}_{(i,j-1)}(r\Delta,s\Delta+\Delta) + C^G_{(i,j+1)}F^{B,G}_{(i,j)}(0,\Delta)F^{G,G}_{(i,j+1)}(r\Delta,s\Delta-\Delta)\Big\}\\
&-\frac{C^B_{(m,n)}}{C^G_{(m,n)}}F^{G,B}_{(i,j)}(r\Delta,s\Delta)\Big\{C^G_{(m-1,n)}F^{G,G}_{(i,j)}(r\Delta-\Delta,s\Delta)F^{G,B}_{(m-1,n)}(\Delta,0) + C^G_{(m+1,n)}F^{G,G}_{(i,j)}(r\Delta+\Delta,s\Delta)F^{B,G}_{(m,n)}(\Delta,0)\\
&+C^G_{(m,n-1)}F^{G,G}_{(i,j)}(r\Delta,s\Delta-\Delta)F^{G,B}_{(m,n-1)}(0,\Delta) + C^G_{(m,n+1)}F^{G,G}_{(i,j)}(r\Delta,s\Delta+\Delta)F^{B,G}_{(m,n)}(0,\Delta)\Big\}\\
&+F^{G,G}_{(i,j)}(r\Delta,s\Delta)\Big\{C^B_{(i-1,j)}F^{B,G}_{(i-1,j)}(\Delta,0)F^{B,G}_{(i-1,j)}(r\Delta+\Delta,s\Delta) + C^B_{(i+1,j)}F^{G,B}_{(i,j)}(\Delta,0)F^{B,G}_{(i+1,j)}(r\Delta-\Delta,s\Delta)\\
&+C^B_{(i,j-1)}F^{B,G}_{(i,j-1)}(0,\Delta)F^{B,G}_{(i,j-1)}(r\Delta,s\Delta+\Delta) + C^B_{(i,j+1)}F^{G,B}_{(i,j)}(0,\Delta)F^{B,G}_{(i,j+1)}(r\Delta,s\Delta-\Delta)\\
&+C^B_{(m-1,n)}F^{G,B}_{(i,j)}(r\Delta-\Delta,s\Delta)F^{B,G}_{(m-1,n)}(\Delta,0) + C^B_{(m+1,n)}F^{G,B}_{(i,j)}(r\Delta+\Delta,s\Delta)F^{G,B}_{(m,n)}(\Delta,0)\\
&+C^B_{(m,n-1)}F^{G,B}_{(i,j)}(r\Delta,s\Delta-\Delta)F^{B,G}_{(m,n-1)}(0,\Delta) + C^B_{(m,n+1)}F^{G,B}_{(i,j)}(r\Delta,s\Delta+\Delta)F^{G,B}_{(m,n)}(0,\Delta)\Big\}\Bigg]+\\
&\frac{P^G_p}{4}\Bigg[\frac{1-C^G_{(i,j)}F^{G,G}_{(i,j)}(r\Delta,s\Delta)-C^B_{(i,j)}F^{B,G}_{(i,j)}(r\Delta,s\Delta)}{C^G_{(i,j)}\left(1-C^G_{(i,j)}-C^B_{(i,j)}\right)}\Bigg\langle C^G_{(i-1,j)}F^{G,G}_{(i-1,j)}(r\Delta+\Delta,s\Delta)\times\\
&\Big\{1-C^G_{(i,j)}F^{G,G}_{(i-1,j)}(\Delta,0)-C^B_{(i,j)}F^{G,B}_{(i-1,j)}(\Delta,0)\Big\} + C^G_{(i+1,j)}F^{G,G}_{(i+1,j)}(r\Delta-\Delta,s\Delta)\times\\
&\Big\{1-C^G_{(i,j)}F^{G,G}_{(i,j)}(\Delta,0)-C^B_{(i,j)}F^{B,G}_{(i,j)}(\Delta,0)\Big\} + C^G_{(i,j-1)}F^{G,G}_{(i,j-1)}(r\Delta,s\Delta+\Delta)\times\\
&\Big\{1-C^G_{(i,j)}F^{G,G}_{(i,j-1)}(0,\Delta)-C^B_{(i,j)}F^{G,B}_{(i,j-1)}(0,\Delta)\Big\} + C^G_{(i,j+1)}F^{G,G}_{(i,j+1)}(r\Delta,s\Delta-\Delta)\times\\
&\Big\{1-C^G_{(i,j)}F^{G,G}_{(i,j)}(0,\Delta)-C^B_{(i,j)}F^{B,G}_{(i,j)}(0,\Delta)\Big\}\Bigg\rangle\\
&+\frac{1-C^G_{(m,n)}F^{G,G}_{(i,j)}(r\Delta,s\Delta)-C^B_{(m,n)}F^{B,G}_{(i,j)}(r\Delta,s\Delta)}{C^G_{(m,n)}\left(1-C^G_{(m,n)}-C^B_{(m,n)}\right)}\Bigg\langle C^G_{(m-1,n)}F^{G,G}_{(i,j)}(r\Delta-\Delta,s\Delta)\times\\
&\Big\{1-C^G_{(m,n)}F^{G,G}_{(m-1,n)}(\Delta,0)-C^B_{(m,n)}F^{G,B}_{(m-1,n)}(\Delta,0)\Big\} + C^G_{(m+1,n)}F^{G,G}_{(i,j)}(r\Delta+\Delta,s\Delta)\times\\
&\Big\{1-C^G_{(m,n)}F^{G,G}_{(m,n)}(\Delta,0)-C^B_{(m,n)}F^{B,G}_{(m,n)}(\Delta,0)\Big\} + C^G_{(m,n-1)}F^{G,G}_{(i,j)}(r\Delta,s\Delta-\Delta)\times\\
&\Big\{1-C^G_{(m,n)}F^{G,G}_{(m,n-1)}(0,\Delta)-C^B_{(m,n)}F^{G,B}_{(m,n-1)}(0,\Delta)\Big\} + C^G_{(m,n+1)}F^{G,G}_{(i,j)}(r\Delta,s\Delta+\Delta)\times\\
&\Big\{1-C^G_{(m,n)}F^{G,G}_{(m,n)}(0,\Delta)-C^B_{(m,n)}F^{B,G}_{(m,n)}(0,\Delta)\Big\}\Bigg\rangle\Bigg] - 2P^G_dF^{G,G}_{(i,j)}.
\end{aligned}
$$

Finally, for lattice pairs separated by a distance $r\Delta$ in the x-direction and a distance $s\Delta$ in the y-direction, occupied by distinct subpopulations, we obtain

$$
\begin{aligned}
&\frac{\mathrm{d}F^{G,B}_{(i,j)}(r\Delta, s\Delta)}{\mathrm{d}t} = -F^{G,B}_{(i,j)}(r\Delta, s\Delta)\left[\frac{1}{C^G_{(i,j)}}\frac{\mathrm{d}C^G_{(i,j)}}{\mathrm{d}t} + \frac{1}{C^B_{(m,n)}}\frac{\mathrm{d}C^B_{(m,n)}}{\mathrm{d}t}\right] + \frac{P^G_m}{4}\Bigg[\frac{C^G_{(i-1,j)}}{C^G_{(i,j)}}F^{G,B}_{(i-1,j)}(r\Delta+\Delta, s\Delta) \\
&+ \frac{C^G_{(i+1,j)}}{C^G_{(i,j)}}F^{G,B}_{(i+1,j)}(r\Delta-\Delta, s\Delta) + \frac{C^G_{(i,j-1)}}{C^G_{(i,j)}}F^{G,B}_{(i,j-1)}(r\Delta, s\Delta+\Delta) + \frac{C^G_{(i,j+1)}}{C^G_{(i,j)}}F^{G,B}_{(i,j+1)}(r\Delta, s\Delta-\Delta) \\
&- 4F^{G,B}_{(i,j)}(r\Delta, s\Delta) - \frac{C^B_{(i,j)}}{C^G_{(i,j)}}F^{B,B}_{(i,j)}(r\Delta, s\Delta)\Big\{C^G_{(i-1,j)}F^{G,B}_{(i-1,j)}(\Delta,0)F^{G,B}_{(i-1,j)}(r\Delta+\Delta, s\Delta) \\
&+ C^G_{(i+1,j)}F^{B,G}_{(i,j)}(\Delta,0)F^{G,B}_{(i+1,j)}(r\Delta-\Delta, s\Delta) + C^G_{(i,j-1)}F^{G,B}_{(i,j-1)}(0,\Delta)F^{G,B}_{(i,j-1)}(r\Delta, s\Delta+\Delta) \\
&+ C^G_{(i,j+1)}F^{B,G}_{(i,j)}(0,\Delta)F^{G,B}_{(i,j+1)}(r\Delta, s\Delta-\Delta)\Big\} + F^{G,B}_{(i,j)}(r\Delta, s\Delta)\Big\{C^B_{(i-1,j)}F^{B,G}_{(i-1,j)}(\Delta,0)F^{B,B}_{(i-1,j)}(r\Delta+\Delta, s\Delta) \\
&+ C^B_{(i+1,j)}F^{G,B}_{(i,j)}(\Delta,0)F^{B,B}_{(i+1,j)}(r\Delta-\Delta, s\Delta) + C^B_{(i,j-1)}F^{B,G}_{(i,j-1)}(0,\Delta)F^{B,B}_{(i,j-1)}(r\Delta, s\Delta+\Delta) \\
&+ C^B_{(i,j+1)}F^{G,B}_{(i,j)}(0,\Delta)F^{B,B}_{(i,j+1)}(r\Delta, s\Delta-\Delta)\Big\}\Bigg] + \\
&\frac{P^B_m}{4}\Bigg[\frac{C^B_{(m-1,n)}}{C^B_{(m,n)}}F^{G,B}_{(i,j)}(r\Delta-\Delta, s\Delta) + \frac{C^B_{(m+1,n)}}{C^B_{(m,n)}}F^{G,B}_{(i,j)}(r\Delta+\Delta, s\Delta) + \frac{C^B_{(m,n-1)}}{C^B_{(m,n)}}F^{G,B}_{(i,j)}(r\Delta, s\Delta-\Delta) \\
&+ \frac{C^B_{(m,n+1)}}{C^B_{(m,n)}}F^{G,B}_{(i,j)}(r\Delta, s\Delta+\Delta) - 4F^{G,B}_{(i,j)}(r\Delta, s\Delta) - \frac{C^G_{(m,n)}}{C^B_{(m,n)}}F^{G,G}_{(i,j)}(r\Delta, s\Delta)\times \\
&\Big\{C^B_{(m-1,n)}F^{G,B}_{(i,j)}(r\Delta-\Delta, s\Delta)F^{B,G}_{(m-1,n)}(\Delta,0) + C^B_{(m+1,n)}F^{G,B}_{(i,j)}(r\Delta+\Delta, s\Delta)F^{G,B}_{(m,n)}(\Delta,0) \\
&+ C^B_{(m,n-1)}F^{G,B}_{(i,j)}(r\Delta, s\Delta-\Delta)F^{B,G}_{(m,n-1)}(0,\Delta) + C^B_{(m,n+1)}F^{G,B}_{(i,j)}(r\Delta, s\Delta+\Delta)F^{G,B}_{(m,n)}(0,\Delta)\Big\} \\
&+ F^{G,B}_{(i,j)}(r\Delta, s\Delta)\Big\{C^G_{(m-1,n)}F^{G,G}_{(i,j)}(r\Delta-\Delta, s\Delta)F^{G,B}_{(m-1,n)}(\Delta,0) + C^G_{(m+1,n)}F^{G,G}_{(i,j)}(r\Delta+\Delta, s\Delta)F^{B,G}_{(m,n)}(\Delta,0) \\
&+ C^G_{(m,n-1)}F^{G,G}_{(i,j)}(r\Delta, s\Delta-\Delta)F^{G,B}_{(m,n-1)}(0,\Delta)C^G_{(m,n+1)}F^{G,G}_{(i,j)}(r\Delta, s\Delta+\Delta)F^{B,G}_{(m,n+1)}(0,\Delta)\Big\}\Bigg] + \\
&\frac{P^G_p}{4}\Bigg[\frac{1 - C^G_{(i,j)}F^{G,B}_{(i,j)}(r\Delta, s\Delta) - C^B_{(i,j)}F^{B,B}_{(i,j)}(r\Delta, s\Delta)}{C^G_{(i,j)}\left(1 - C^G_{(i,j)} - C^B_{(i,j)}\right)}\times \\
&\Bigg\langle C^G_{(i-1,j)}F^{G,B}_{(i-1,j)}(r\Delta+\Delta, s\Delta)\Big\{1 - C^G_{(i,j)}F^{G,G}_{(i-1,j)}(\Delta,0) - C^B_{(i,j)}F^{G,B}_{(i-1,j)}(\Delta,0)\Big\} \\
&+ C^G_{(i+1,j)}F^{G,B}_{(i+1,j)}(r\Delta-\Delta, s\Delta)\Big\{1 - C^G_{(i,j)}F^{G,G}_{(i,j)}(\Delta,0) - C^B_{(i,j)}F^{B,G}_{(i,j)}(\Delta,0)\Big\} \\
&+ C^G_{(i,j-1)}F^{G,B}_{(i,j-1)}(r\Delta, s\Delta+\Delta)\Big\{1 - C^G_{(i,j)}F^{G,G}_{(i,j-1)}(0,\Delta) - C^B_{(i,j)}F^{G,B}_{(i,j-1)}(0,\Delta)\Big\} \\
&+ C^G_{(i,j+1)}F^{G,B}_{(i,j+1)}(r\Delta, s\Delta-\Delta)\Big\{1 - C^G_{(i,j)}F^{G,G}_{(i,j)}(0,\Delta) - C^B_{(i,j)}F^{B,G}_{(i,j)}(0,\Delta)\Big\}\Bigg\rangle\Bigg] + \\
&\frac{P^B_p}{4}\Bigg[\frac{1 - C^G_{(m,n)}F^{G,G}_{(i,j)}(r\Delta, s\Delta) - C^B_{(m,n)}F^{G,B}_{(i,j)}(r\Delta, s\Delta)}{C^B_{(m,n)}\left(1 - C^G_{(m,n)} - C^B_{(m,n)}\right)}\times \\
&\Bigg\langle C^B_{(m-1,n)}F^{G,B}_{(i,j)}(r\Delta-\Delta, s\Delta)\Big\{1 - C^G_{(m,n)}F^{B,G}_{(m-1,n)}(\Delta,0) - C^B_{(m,n)}F^{B,B}_{(m-1,n)}(\Delta,0)\Big\} \\
&+ C^B_{(m+1,n)}F^{G,B}_{(i,j)}(r\Delta+\Delta, s\Delta)\Big\{1 - C^G_{(m,n)}F^{G,B}_{(m,n)}(\Delta,0) - C^B_{(m,n)}F^{B,B}_{(m,n)}(\Delta,0)\Big\} \\
&+ C^B_{(m,n-1)}F^{G,B}_{(i,j)}(r\Delta, s\Delta-\Delta)\Big\{1 - C^G_{(m,n)}F^{B,G}_{(m,n-1)}(0,\Delta) - C^B_{(m,n)}F^{B,B}_{(m,n-1)}(0,\Delta)\Big\} \\
&+ C^B_{(m,n+1)}F^{G,B}_{(i,j)}(r\Delta, s\Delta+\Delta)\Big\{1 - C^G_{(m,n)}F^{G,B}_{(m,n)}(0,\Delta) - C^B_{(m,n)}F^{B,B}_{(m,n)}(0,\Delta)\Big\}\Bigg\rangle\Bigg] - P^G_d F^{G,B}_{(i,j)} - P^B_d F^{G,B}_{(i,j)}.
\end{aligned}
$$

Numerical method

In all cases the equations describing the time rate of change of $F_i^{a,b}$ is undefined if either $C_i^a = 0$ or $C_i^a + C_i^b = 1$. Physically, this corresponds to the case where either some part of the domain is completely vacant or is completely occupied to confluence. There are two ways of dealing with this complication:

1. Following the approach of Simpson and Baker [28] we can use a hybrid technique where we implement the mean-field model for all lattice sites where the density is below some user-defined small threshold value, ϵ, and greater than $1 - \epsilon$. For all others lattice sites we implement the moment dynamics model. This ensures that the equations governing the correlation functions are well-defined and we apply the traditional mean-field approximation for all remaining lattice sites on the domain.
2. It is useful to note that for many practical problems the density of agents is never precisely zero or precisely equal to the carrying capacity density (e.g. see Figure 1(d) in the main document). In these situations the correlation functions are always defined.

While Simpson and Baker's [27] hybrid method ensures that we never encounter any issue with the correlation functions being undefined, this method can be computationally expensive to implement since it requires that we determine the precise regions where the moment dynamics and traditional mean-field models are to be evaluated. For several of the problems we consider in this work, namely the results presented in Figures 6.3, 6.4 and 6.7 in the main document, we choose initial conditions and parameter combinations which ensure that we did not encounter this complication. However, we are also interested in an idealised problem where the initial density is occupied to confluence, which means that we need to implement the hybrid method. The density is only equal to the carrying capacity for the first timestep for problems with $P_d > 0$, as crowding effects imply that proliferation events cannot take place when the density is equal to the carrying capacity. Therefore, the mean-field approximation is only required initially and we do not need to track the region where the moment dynamics model is valid, which reduces to computational cost of the method. The results for this idealised problem are presented in Figures 6.2, 6.5 and 6.6 in the main document.

Correlation function solutions

We present the correlation functions for the one-dimensional model of a co-culture scratch assay in Figures 6.8 and 6.9 for $C_0 = 0.5$ and $C_0 = 0.1$, respectively. We observe that in both cases $F^{G,G}(r)$ and $F^{B,B}(r)$ are significantly higher than the standard assumption that $F^{a,b}(r) \equiv 1$, particularly for short distance. The magnitude of the discrepancy between the standard assumption and the calculated correlation values provides an explanation for the significant difference between the mean-field and moment dynamics models presented in the main document. The self-correlation functions ($F^{G,G}(r)$ and $F^{B,B}(r)$) are greater than one, indicating that lattice pairs are more likely to be occupied by the same subpopulation. This result is intuitive given that crowding effects do not allow agents to “jump” over other agents and that proliferation creates nearest-neighbour pairs of the same subpopulation. Subsequently, the correlation functions ($F^{G,B}(r)$ and $F^{G,B}(r)$) are negatively correlated, that is, less than one. We observe that the magnitude of F is a decreasing function of r, which suggests that our assumption that there exists $r = r_{\max}$ such that $F^{a,b}(r_{\max}) \equiv 1$.

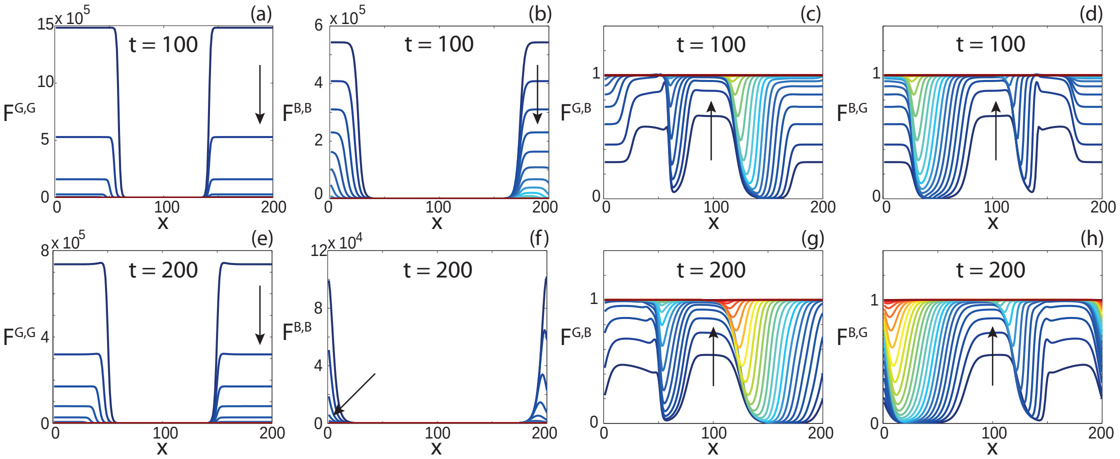

Figure 6.8: Correlation functions for the one-dimensional model of a co-culture scratch assay. Correlation between lattice sites separated by distance r, $r \in [1, 100]$ for: (a),(e) both lattice sites occupied by subpopulation G, (b),(f) both lattice sites occupied by subpopulation B, (c),(g) lattice site i occupied by subpopulation G and lattice site m occupied by subpopulation B, (d)-(h) lattice site i occupied by subpopulation B and lattice site m occupied by subpopulation G at $t = 100$ and $t = 200$, respectively. Parameters used were $P_m^G = 0.1$, $P_m^B = 1$, $P_p^G = P_p^B = 0.05$, $P_d^G = P_d^B = 0.02$, $r_{\max} = 100$, $C_0 = 0.5$, $\epsilon = 10^{-8}$, $i_1 = 81$, $i_2 = 121$, $X = 200$, $\Delta = 1$. For ease of interpretation we present every fourth correlation solution curve. Arrow indicates direction of increasing r.

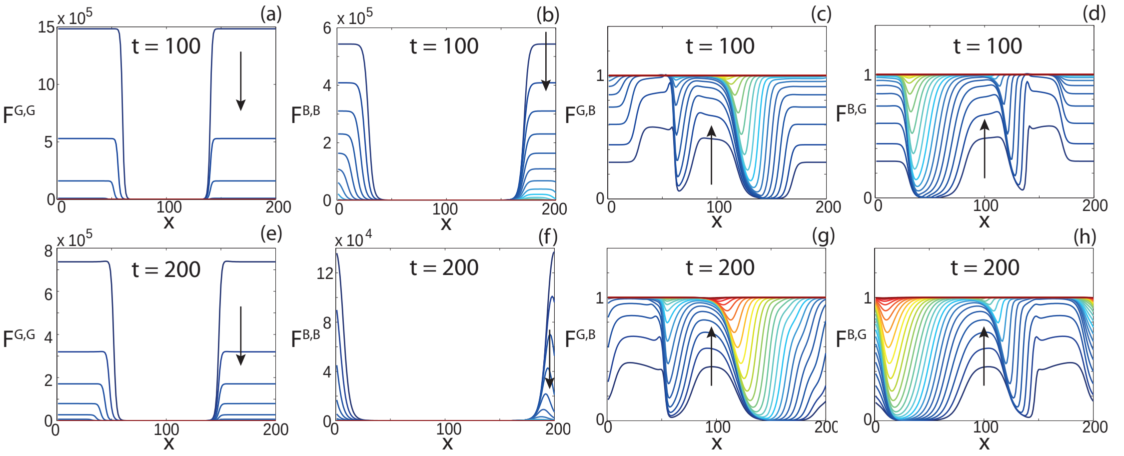

Figure 6.9: Correlation functions for the one-dimensional model of a co-culture scratch assay. Correlation between lattice sites separated by distance r, $r \in [1, 100]$ for: (a),(e) both lattice sites occupied by subpopulation G, (b),(f) both lattice sites occupied by subpopulation B, (c),(g) lattice site i occupied by subpopulation G and lattice site m occupied by subpopulation B, (d)-(h) lattice site i occupied by subpopulation B and lattice site m occupied by subpopulation G at $t = 100$ and $t = 200$, respectively. Parameters used were $P_m^G = 0.1$, $P_m^B = 1$, $P_p^G = P_p^B = 0.05$, $P_d^G = P_d^B = 0.02$, $r_{\max} = 100$, $C_0 = 0.1$, $\epsilon = 10^{-8}$, $i_1 = 81$, $i_2 = 121$, $X = 200$, $\Delta = 1$. For ease of interpretation we present every fourth correlation solution curve. Arrow indicates direction of increasing r.

Chapter 7

Filling the gaps: a robust description of adhesive birth-death-movement processes

A paper published in *Physical Review E.*

Johnston, S. T., Baker, R. E. and Simpson, M. J. [2016], ‘Filling the gaps: A robust description of adhesive birth-death-movement processes’, *Physical Review E* **93**(4), 042413.

abstract

Existing continuum descriptions of discrete adhesive birth-death-movement processes provide accurate predictions of the average discrete behavior for limited parameter regimes. Here we present an alternative continuum description in terms of the dynamics of groups of contiguous occupied and vacant lattice sites. Our method provides more accurate predictions, is valid in parameter regimes that could not be described by previous continuum descriptions, and provides information about the spatial clustering of occupied sites. Furthermore, we present a simple analytic approximation of the spatial clustering of occupied sites at late time, when the system reaches its steady-state configuration.

7.1 Introduction

Birth, death and movement of individuals are key components of collective behavior, relevant to tissue repair [11, 15, 21, 30, 31] and polymer aggregation [18]. Lattice-based random walks are often used to describe these processes [5–7, 12, 15]. For example, Deroulers *et al.* [6] use a random walk to model the migration of glioma cells, while Hackett-Jones *et al.* [12] use a random walk to describe cellular aggregation patterns arising from nonlocal interactions. Mackie *et al.* [18] are interested in molecular aggregation and use random walks to examine the formation of micelles. In the context of active transport, Illien *et al.* [13] use lattice-based random walks to model diffusion in a crowded single-file environment. While these stochastic random walk models provide insight into collective behavior, performing a sufficient number of stochastic realizations to obtain representative average behavior can be computationally expensive. Furthermore, relying solely upon stochastic simulations provides little opportunity for analyzing how parameters in the simulation are related to the collective behavior.

There is significant interest in continuum descriptions that produce robust approximations of the average behavior. Additionally, there is considerable interest in analytic approaches that describe the average steady-state behavior of the random walk model, particularly for the one-dimensional asymmetric exclusion process [7]. To address both the analytic intractability and the computational expense associated with obtaining representative average behavior of random walk models, mean-field (MF) descriptions of lattice-based adhesive birth-death-movement (ABDM) processes have been developed [1, 8, 21]. Unfortunately, MF descriptions only apply in limited parameter regimes due to the neglect of spatial correlations [2, 11, 14, 17, 20]. Specifically, MF descriptions only apply when adhesion (or repulsion) is sufficiently weak, and the ratio of both the birth and death rates to the movement rate is sufficiently small [2, 14]. To address this, corrected mean-field (CMF) descriptions that approximately incorporate spatial correlations have been proposed [2, 14, 23, 25, 29]. While CMF descriptions provide an accurate approximation of the average behavior for a wider range of parameter regimes, CMF descriptions are invalid when the rates of birth and death are not sufficiently small compared to the rate of movement [2]. Here we interpret ABDM processes in terms of groups of contiguous occupied and vacant lattice sites, and present the corresponding continuum description. We note that our description is limited to spatially-independent initial conditions. This implies that our description is translationally invariant and cannot describe processes that are initially spatially dependent. Our method provides accurate predictions of the average behavior in parameter regimes where MF and CMF descriptions are invalid. Furthermore, our description provides information about spatial clustering and we give a simple analytic approximation of the spatial clustering of occupied sites when the system has reached steady-state.

7.2 Discrete model

We consider a periodic one-dimensional lattice-based random walk, where each site may be occupied by, at most, one agent [4]. Isolated agents undergo birth, death and movement events at constant rates P_p, P_d and P_m, respectively [2]. During a birth event, an agent attempts to place a daughter at a randomly-selected nearest-neighbor site. This event is successful provided that the selected site is vacant. A death event removes an agent and the associated site becomes vacant. During a movement event, an agent attempts to move to a randomly-selected nearest-neighbor site. This selection is unbiased if both nearest-neighbor sites are vacant. If one nearest-neighbor site is occupied, the vacant nearest-neighbor site is selected with probability $(1 - \alpha)/2$, where $\alpha \in [-1, 1]$ represents the strength of agent-agent adhesion ($\alpha > 0$) or repulsion ($\alpha < 0$) [1, 14]. We use the Gillespie algorithm [10] to simulate the random walk. To generate representative behavior, we perform M identically-prepared realizations of the random walk to calculate the average of the summary statistic of interest.

7.3 Chain-and-Gap description

Here we interpret the system as a combination of groups of contiguous occupied and vacant sites of length $i \in [1, N]$, where N is the number of sites. We refer to groups of contiguous occupied sites as *chains* and groups of contiguous vacant sites as *gaps*. For example, Fig. 7.1(a) shows a lattice configuration that contains four chains; one each of length one, two, three and four, and four gaps; two each of length one and two. Instead of considering how a birth, death or movement event affects the occupancy of an individual site, we consider how these events change the lengths of the chains and gaps. The success of birth and movement events depends on the "type" of agent, that is, whether an agent has zero, one or two nearest-neighbor agents. We refer to these as *single*, *edge* or *middle* agents, respectively. Single agents can move to, or place a daughter agent at, both nearest-neighbor sites, whereas edge agents can move to, or place a daughter agent at, only one nearest-neighbor site. Due to crowding, it is impossible for middle agents to undergo successful movement or birth events. Potential death events are not affected by crowding. The net birth rate is $P_p(2N_S(t)+N_E(t))/2-P_d(N_S(t)+N_E(t)+N_M(t))$, where $N_S(t)$, $N_E(t)$ and $N_M(t)$ are the number of single, edge and middle agents, respectively. We can calculate $N_S(t)$, $N_E(t)$ and $N_M(t)$ provided that we know the number of chains of length $n \in [1, N]$, which we denote $C_n(t)$. For all $n \geq 2$, there are two edge agents and $n-2$ middle agents per chain. The number of single agents is $C_1(t)$. We denote the number of gaps of length $m \in [1, N]$ by $G_m(t)$. For notational convenience, we henceforth refer to N_S, N_E, N_M, C_n and G_m without explicitly noting temporal dependence.

There are eleven distinct birth, death or movement events, denoted by $B_j,\ j = 1, 2, 3, 4$, $D_j,\ j = 1, 2, 3$ and $M_j,\ j = 1, 2, 3, 4$. In Figs. 7.1(b)-(k), we demonstrate how each of these events affects the configuration of agents in Fig. 7.1(a). For an arbitrary configuration, the influence of each event can be described as follows. We note that an increase/decrease in C_n or G_m refers to an increase/decrease of one.

Event B_1: A single agent undergoes birth and places a daughter agent into a gap of length one, which is next to a chain of length n (Fig. 7.1(b)). C_1, G_1 and C_n decrease, C_{n+2} increases.

Event B_2: A single agent undergoes birth and places a daughter agent into a gap of length $m \geq 2$ (Fig. 7.1(c)). C_1 and G_m decrease, C_2 and G_{m-1} increase.

Event B_3: An edge agent, part of a chain of length n_1, undergoes birth and places a daughter agent into a gap of length one, which is next to a chain of length n_2 (Fig. 7.1(b)). C_{n_1}, C_{n_2} and G_1 decrease, $C_{n_1+n_2+1}$ increases.

Event B_4: An edge agent, part of a chain of length n, undergoes birth and places a daughter agent into a gap of length $m \geq 2$ (Fig. 7.1(d)). C_n and G_m decrease, C_{n+1} and G_{m-1} increase.

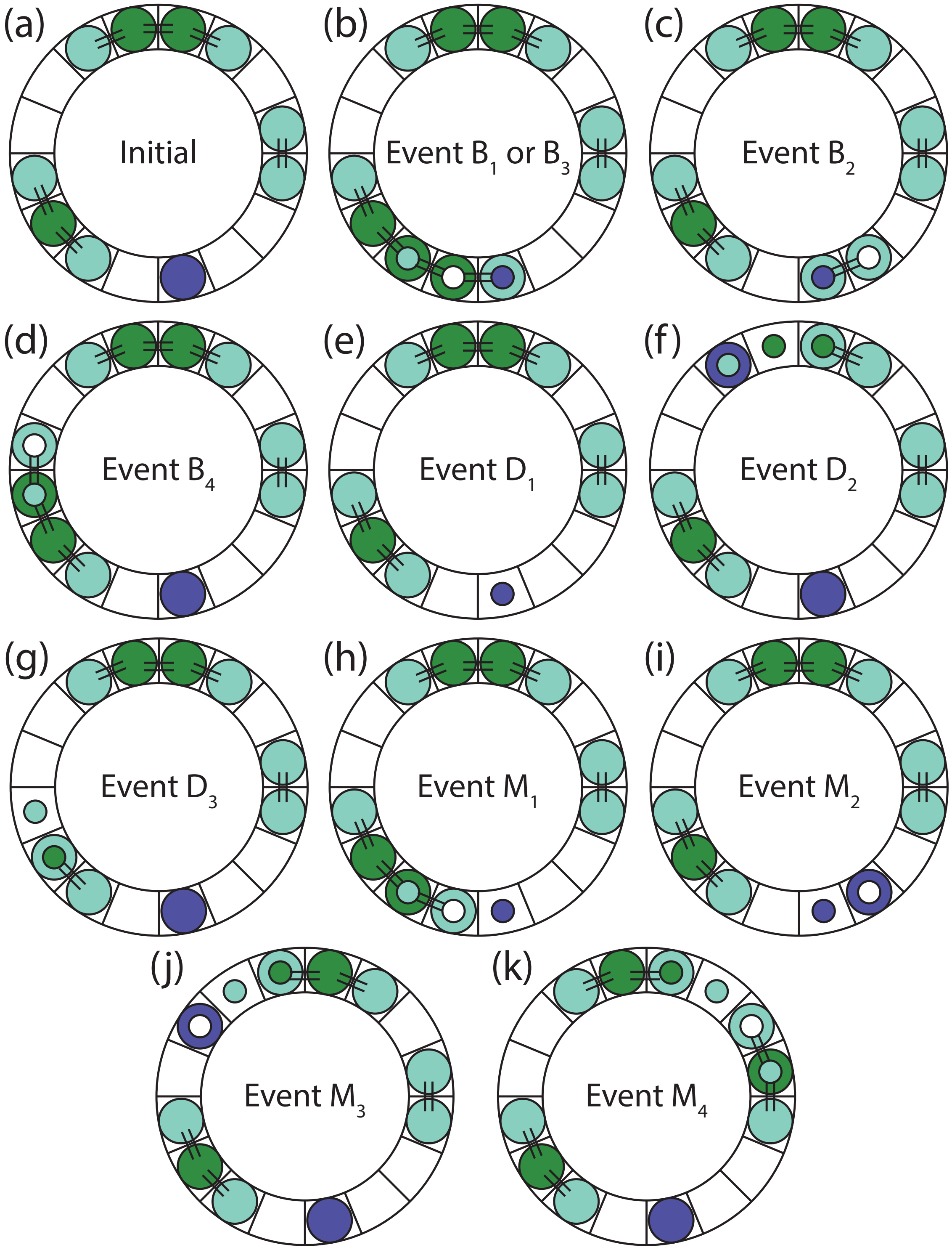

Figure 7.1: (a) Initial lattice configuration containing single (purple), edge (blue) and middle (green) agents. Lattice configuration after potential (b)-(d) birth, (e)-(g) death and (h)-(k) movement events. Inset circles represent the initial lattice configuration. Dashed lines represent agents influenced by agent-agent adhesion or repulsion.

Event D_1: A single agent, with neighbor gaps of length m_1 and m_2, undergoes death (Fig. 7.1(e)). C_1, G_{m_1} and G_{m_2} decrease, $G_{m_1+m_2+1}$ increases.

Event D_2: A middle agent, part of a chain of length n, undergoes death (Fig. 7.1(f)). C_n decreases, G_1, C_{n_1} and C_{n_2} increase, where $n_1 + n_2 + 1 = n$. We note that the expected increase of C_{n_1} and C_{n_2} is uniform for all $n_1, n_2 \in [1, n-2],\ n_1 + n_2 + 1 = n$.

Event D_3: An edge agent, part of a chain of length n, which is next to a gap of length m, undergoes death (Fig. 7.1(g)). C_n and G_m decrease, C_{n-1} and G_{m+1} increase.

Event M_1: A single agent undergoes movement into a gap of length one, which is next to a chain of length n (Fig. 7.1(h)). The single agent is also next to a gap of length m. C_1, C_n, G_1 and G_m decrease, C_{n+1} and G_{m+1} increase.

Event M_2: A single agent undergoes movement into a gap of length $m_1 \geq 2$, away from a gap of length m_2 (Fig. 7.1(i)). G_{m_1} and G_{m_2} decrease, G_{m_1-1} and G_{m_2+1} increase.

Event M_3: An edge agent, part of a chain of length n, undergoes movement into a gap of length $m \geq 2$ (Fig. 7.1(j)). C_n and G_m decrease, C_1, C_{n-1}, G_1 and G_{m-1} increase.

Event M_4: An edge agent, part of a chain of length n_1, undergoes movement into a gap of length one, which is next to a chain of length n_2 (Fig. 7.1(k)). C_{n_1} and C_{n_2} decrease, C_{n_1-1} and C_{n_2+1} increase.

We obtain transition rates between possible states of the system by considering the result of each potential event and the rate at which it occurs and, subsequently, obtain a system of ordinary differential equations (ODEs) describing $\mathrm{d}C_n/\mathrm{d}t$ and $\mathrm{d}G_m/\mathrm{d}t$, $n, m \in [1, N]$. For single agents, birth and movement events are never aborted due to crowding, and, therefore, the rate at which each of these events occurs is P_pC_1 and P_mC_1, respectively. We note that α does not influence single agents. For edge agents, birth events are aborted, on average, half the time. However, there are two edge agents for each C_n, $n \geq 2$, which implies that the rate at which these events occur for a chain of length n is P_pC_n. We follow a similar process for the rate of movement events. However, the probability of a successful movement event for edge agents is influenced by α and hence the rate of movement for a chain of length n is $(1-\alpha)P_mC_n$ for $n \geq 2$. Death events are never aborted due to crowding, which implies that the rate at which death events occur for a chain of length n is nP_dC_n. These rates describe ABDM events in terms of the agent type but do not describe the rate and result of possible sub-events. For example, Events B_1 and B_2 are both birth events for single agents, but occur depending on whether the gap that the daughter agent is placed into is length one (Event B_1) or greater (Event B_2). The proportion of single agent birth events that are Event B_1 is equivalent to the probability that, given that we have selected a single agent, there is a neighboring gap of length one. We denote this probability by $P(G_1|C_1)$. Due to the periodic geometry, a single agent can be next to gaps of length $j \in A(C_1)$, where $A(C_1) = \{i \in \mathbb{Z} | 1 \leq i \leq N-3\} \cup \{N-1\}$. We make the assumption that $P(G_1|C_1) = G_1/\sum_{j \in A(C_1)} G_j$, that is, the ratio of gaps of length one to all gaps that can exist next to a chain of length one. The expected change of both C_1 and G_1, due to Event B_1, is $P_pC_1P(G_1|C_1)/2$. Event B_1 also affects C_n and C_{n+2} and, as such, we require the expected change of C_n and C_{n+2}, due to Event B_1. To obtain an expression for the expected change for C_n and C_{n+2}, due to Event B_1, we consider the probability that, given that we have a single agent with a neighboring gap of length one, the other neighboring chain, with respect to the gap, is of length n (Appendix A). We can calculate the average outcome of Event B_1 by considering potential n values and the associated transition rate for each n. Combining all possible events with their respective rates gives the governing system of ODEs (Appendix A).

Our current framework can be extended to describe additional details. However, in this first presentation of the Chain-and-Gap (C&G) framework we focus on the most straightforward periodic one-dimensional case, with translationally-invariant initial conditions. Other boundary conditions could be considered; however, this would require that the number of chains and gaps of each length is known at each lattice site, and hence the number of equations in the governing system would be $\mathcal{O}(N^2)$ rather than $\mathcal{O}(N)$. For example, we could consider non-periodic geometry, with specific conditions imposed on the two boundary lattice sites. We acknowledge that our approach is currently limited to one-dimensional processes. While this restricts the type of physical applications that our approach can describe, there is considerable interest in one-dimensional exclusion processes throughout the physical and life sciences [1, 3, 13]. Furthermore, the influence of the spatial correlations is most pronounced in one dimension [2] and, therefore, it is logical to focus on the development of new approximations for one-dimensional processes.

7.4 Traditional continuum descriptions

Typically, continuum descriptions of ABDM processes are obtained by considering the time rate of change of the average occupancy of sites [2, 8, 14, 25]. These descriptions, which describe the number of occupied sites, $S(t)$, take the form

$$\frac{\mathrm{d}S(t)}{\mathrm{d}t} = P_p S(t)\left(1 - \frac{F(t)S(t)}{N}\right) - P_d S(t), \tag{7.1}$$

where $F(t)$ represents the nearest-neighbor correlation [2, 26]. Traditional MF descriptions implicitly assume that $F(t) \equiv 1$ [26]. For CMF descriptions, which approximate spatial correlation between sites by considering the evolution of pairs [24, 32], $F(t)$ is described by a system of ODEs [2]. This system is truncated through an appropriate moment closure method, such as the Kirkwood Superposition Approximation [28, 33].

7.5 Results

To examine whether our C&G description provides an accurate approximation of the lattice-based ABDM process, we first calculate the evolution of the number of occupied sites. The C&G system of ODEs is solved using an adaptive Runge-Kutta method with a strict truncation error control of 10^{-8} [22]. The number of occupied sites in the C&G description is $S_{\mathrm{C\&G}} = \sum_{i=1}^{N} iC_i$. In Figs. 7.2(a)-(h) we present occupancy evolution curves obtained from the averaged discrete model, and the MF, CMF and C&G descriptions for

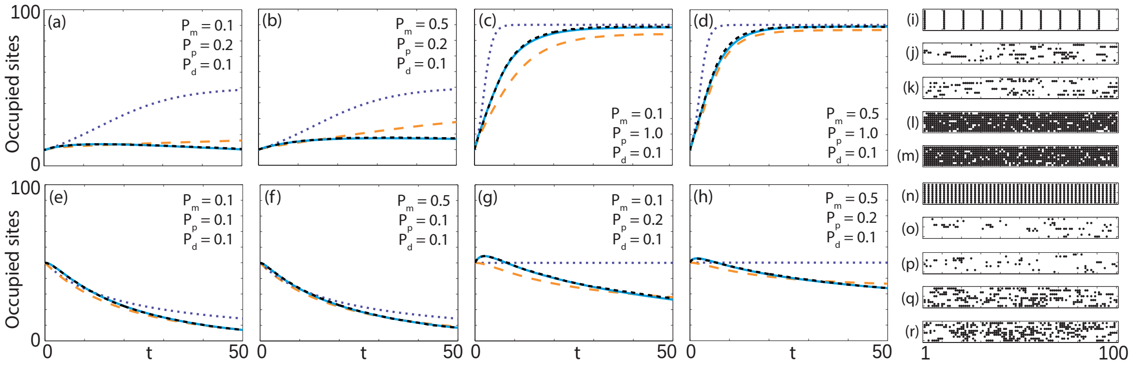

Figure 7.2: Comparison between averaged discrete (black, dashed), MF (purple, dotted), CMF (orange, long dashed) and C&G (cyan, solid) results for a suite of parameter regimes and (a)-(d)10 and (e)-(f) 50 uniformly distributed initially occupied sites. For all results $M = 1000$, $N = 100$ and $\alpha = 0$. (i)-(r) Ten discrete snapshots for (i),(n) the two initial conditions and (j)-(m),(o)-(r) the parameter regimes in (a)-(d),(e)-(h), respectively, at $t = 50$.

	Fig. 2(a)	Fig. 2(b)	Fig. 2(c)	Fig. 2(d)
Discrete model (single realization)	0.08 s	0.19 s	1.40 s	1.91 s
Discrete model (1000 realizations)	76.7 s	192.9 s	1401.7 s	1907.3 s
C&G description	21.1 s	28.2 s	36.8 s	39.1 s

	Fig. 2(e)	Fig. 2(f)	Fig. 2(g)	Fig. 2(h)
Discrete model (single realization)	0.10 s	0.24 s	0.23 s	0.48 s
Discrete model (1000 realizations)	99.4 s	241.9 s	233.5 s	483.1 s
C&G description	26.8 s	27.9 s	30.3 s	27.5 s

Table 7.1: Comparison of the time taken to perform: (i) a single realization of the discrete model; (ii) 1000 realizations of the discrete model, and (iii) a numerical solution of the C&G system of equations for the parameter values in Figs. 2(a)-(h). All solutions are obtained using a single 3.0 GHz Intel i7-3540M desktop processor.

a suite of parameter regimes and two different initial conditions. All results in Fig. 7.2 are without adhesion or repulsion ($\alpha = 0$) and the discrete initial conditions are presented in Figs. 7.2(i),(n). For all parameter regimes and initial conditions the C&G description provides extremely accurate predictions, even when both the MF and CMF descriptions are invalid due to the emergence of significant spatial correlations [2]. We illustrate correlations between sites by presenting snapshots of the clustering in the discrete model in Figs. 7.2(j)-(m),(o)-(r) at $t = 50$. The C&G description implicitly accounts for spatial correlations by considering contiguous occupied sites as a single object, which necessarily has neighboring vacant sites. Both the transient and steady-state behavior of the system is correctly predicted by the C&G description. For example, in Fig. 7.2(b), the MF description inaccurately predicts both the transient and steady-state behavior. The CMF model predicts the transient behavior accurately for $t < 20$, but provides inaccurate predictions for $t > 20$ and, consequently, the steady-state behavior. The C&G description correctly predicts that at steady-state there are, on average, approximately 18 occupied sites, as well as accurately describing the transient behavior of the system.

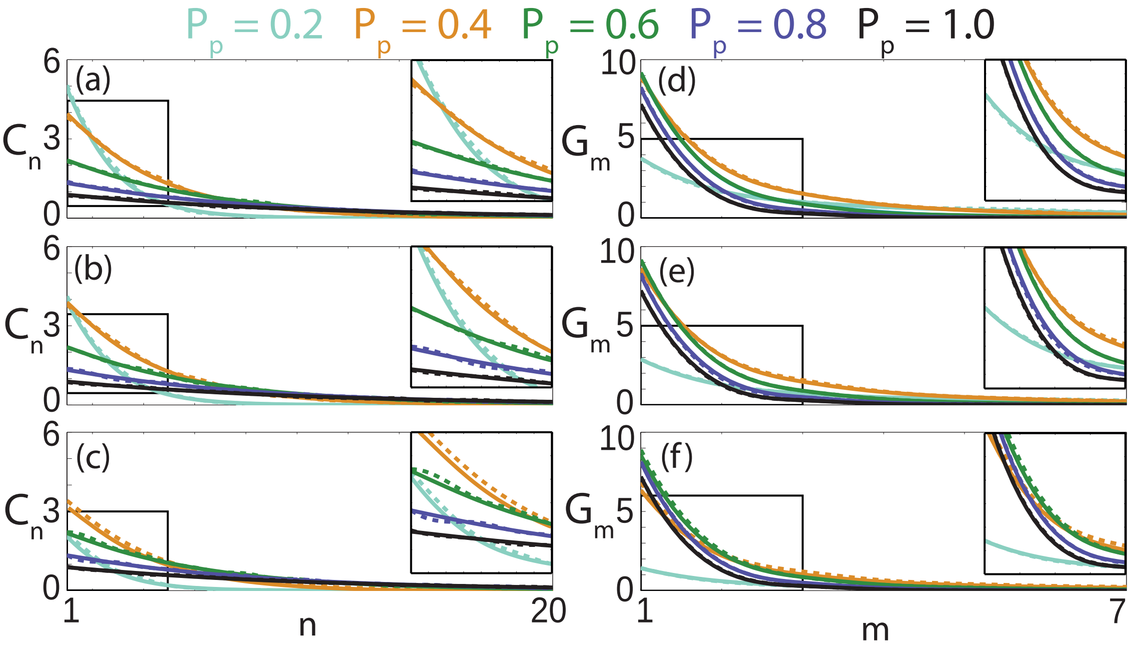

Figure 7.3: Comparison between averaged discrete (dashed) and C&G (solid) results for (a)-(c) C_n and (d)-(f) G_m for a suite of P_p values at $t = 50$. Initially, we have (a),(d) 50, (b),(e) 25 and (c),(f) 10 uniformly distributed occupied sites. For all results, $P_m = P_d = 0.1$, $\alpha = 0$, $M = 1000$ and $N = 100$. Inset boxes highlight regions of particular interest.

We compare the time required to perform a single realization of the discrete model, 1000 realizations of the discrete model and to solve the C&G system of equations in Table 7.1. We note that in all cases it is significantly faster to perform a single realization of the discrete model. However, it is significantly slower to perform sufficiently many realizations to obtain meaningful average behavior, compared to obtaining the numerical solution to the C&G system of equations.

7.5.1 Clustering

Additionally, the C&G description provides information about the clustering of both occupied and vacant sites. In Fig. 7.3 we compare predictions from the C&G description with the averaged discrete model for C_n and G_m for a range of parameter regimes and three different initial conditions. As neither the MF or CMF description contains this information, we are unable to provide comparisons with the predictions from these descriptions. We observe that the C&G description provides accurate predictions for all cases, describing the monotonically decreasing relationship between both C_n and n, and G_m and m. An increase in P_p reduces C_1 and increases the number of longer chains. As expected, we observe the opposite behavior for the gaps. We note that the C&G description provides slightly less accurate predictions for the lowest number of initially occupied sites (Figs. 7.3(c),(f)), a result that is consistent with other continuum descriptions [2].

In Figs. 7.4(a)-(c), we present occupancy evolution profiles, C_n, and G_m, for different α. The number of occupied sites decreases with α, due to the clustering that arises in the

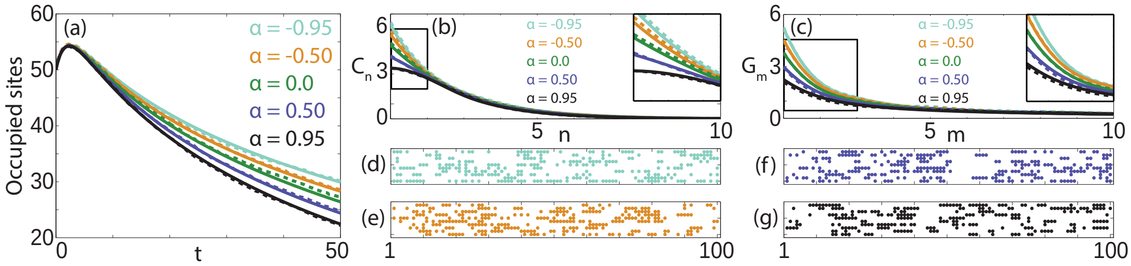

Figure 7.4: Comparison between averaged discrete (dashed) and C&G (solid) results for (a) occupied sites, (b) C_n and (c) G_m for a range of adhesion and repulsion values. (d)-(g) Ten snapshots of the discrete model for (d) $\alpha = -0.95$, (e) $\alpha = -0.5$, (f) $\alpha = 0.5$ and (g) $\alpha = 0.95$, at $t = 50$. Initially, we have 50 uniformly distributed occupied sites. For all results, $P_m = P_d = 0.1$, $P_p = 0.2$, $M = 1000$ and $N = 100$. Inset boxes highlight regions of particular interest.

presence of adhesion, which we observe in Figs. 7.4(b)-(c), where there are significantly fewer chains and gaps of short length compared to when $\alpha \leq 0$. Snapshots of the discrete model in Figs. 7.4(d)-(g) confirm this, as longer chains are present when $\alpha > 0$. For all adhesion and repulsion values the C&G predictions match the averaged discrete model extremely well.

7.5.2 Spatial correlations

Traditional continuum descriptions of ABDM processes describe the evolution of the number of occupied sites, $S(t)$. If there is no adhesion/repulsion present in the discrete model, traditional continuum descriptions are described by Equation (7.1) [2, 26]. Traditional MF descriptions implicitly assume that the spatial correlation between the occupancy of nearest-neighbor lattice sites, $F(t) \equiv 1$, while CMF descriptions approximate $F(t)$ by explicitly considering the dynamics of pairs of lattice sites [2]. In the CMF description

$$F(t) = \frac{\rho^{(2)}(\sigma_i, \sigma_{i+1})}{\rho^{(1)}(\sigma_i)\rho^{(1)}(\sigma_{i+1})},$$

where $\rho^{(2)}(\sigma_i, \sigma_{i+1})$ is the probability that both sites i and $i+1$ are occupied and $\rho^{(1)}(\sigma_i)$ is the probability that site i is occupied [2]. It is possible to reconstruct $F(t)$ from the C&G description as we can express the probabilities that certain sites are occupied. Since we consider problems with translationally-invariant initial conditions, the probability that a site is occupied is independent of position, and hence

$$\rho^{(1)}(\sigma) = \frac{1}{N}\sum_{i=1}^{N} iC_i.$$

The probability that two nearest-neighbor sites are occupied follows a similar argument. Every chain of length $2 \leq n \leq N-1$ consists of $n-1$ pairs of occupied nearest-neighbor sites, while a chain of length N consists of N pairs of occupied nearest-neighbor sites.

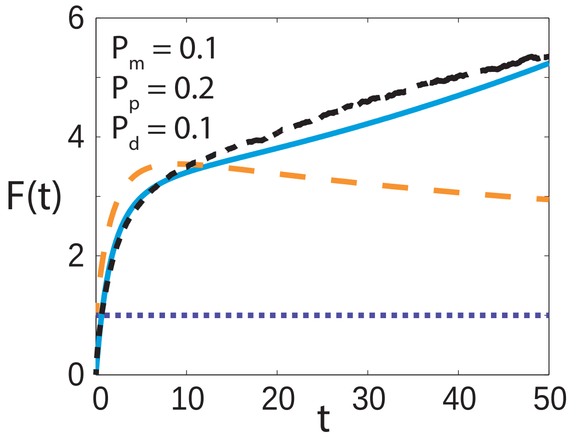

Figure 7.5: Comparison of $F(t)$ for the averaged discrete model (black, dashed), and the MF (purple, dotted), CMF (orange, long dashed) and C&G (cyan, solid) descriptions. Here $P_m = 0.1$, $P_p = 0.2$, $P_d = 0.1$, $\alpha = 0$, $N = 100$, $M = 10^4$. Initially we have 10 uniformly distributed occupied lattice sites.

As our domain is periodic, there are N possible pairs of nearest-neighbor sites. The probability that two nearest-neighbor sites are occupied is therefore

$$\rho^{(2)}(\sigma_i, \sigma_{i+1}) = \frac{1}{N}\left[\sum_{j=2}^{N-1}(j-1)C_j + NC_N\right].$$

In Figure 7.5 we demonstrate that the CMF description fails to describe the nearest-neighbor correlation present in the discrete model, whereas the C&G description correctly predicts the dynamics of the nearest-neighbor correlation. Furthermore, we demonstrate that both the CMF and C&G estimates of the nearest-neighbor correlation are significantly different to the traditional implicit assumption that $F(t) \equiv 1$. The CMF description relies on an approximation to obtain a closed system of equations describing the correlations, which may result in inaccurate predictions, whereas the C&G description accounts for all possible chain sizes and reconstructs the correlations from this information.

7.5.3 Steady-state approximation

In Fig. 7.6(a) we present C_n obtained from the C&G description at late-time (Appendix C) on a $\log_{10}$ scale. These plots are approximately linear, implying $C_n \approx \beta^{n-1}C_1$, $\beta > 0$. At steady-state, the net birth rate is zero and $P_p(2N_S + N_E)/2 = P_d(N_S + N_E + N_M)$. If we express N_S, N_E and N_M in terms of C_n and make a power series approximation in terms of β, we obtain $\beta = 1 - P_d/P_p$. For the power series to converge, we require $|\beta| < 1$. As expected, this implies that non-trivial steady states exist only for $P_p > P_d$. Full details of this argument are presented in Appendix C. In Fig. 7.6(b) we compare the averaged discrete model at late-time and the steady-state approximation and observe that the approximation is extremely accurate.

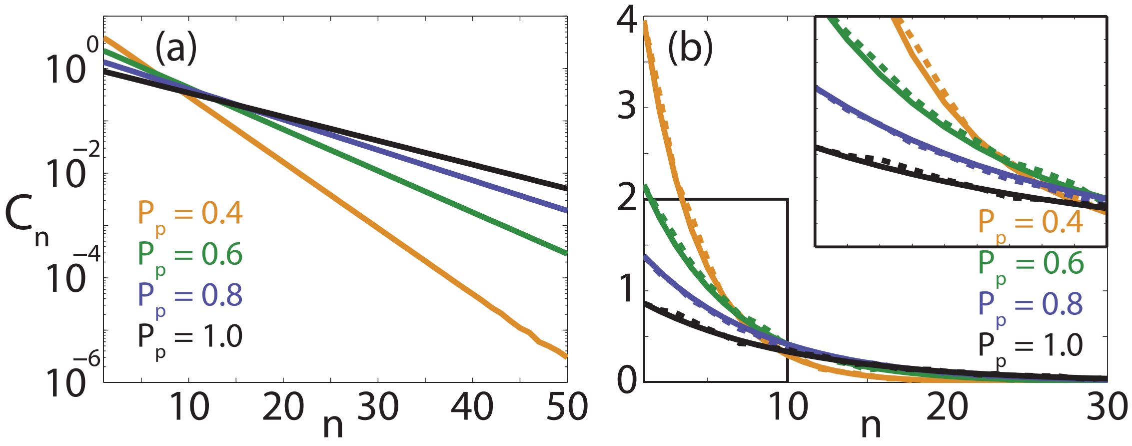

Figure 7.6: (a) C_n, calculated from the C&G description, on a $\log_{10}$ scale at steady-state for different P_p values. (b) Steady-state approximation for C_n (solid) compared with the averaged discrete results (dashed). For all results, $P_m = P_d = 0.1$, $\alpha = 0$, $M = 1000$ and $N = 100$. Inset box highlights a region of particular interest.

7.6 Conclusions

We present a powerful approach to describe lattice-based ABDM processes in terms of groups of contiguous occupied and vacant sites. Our C&G description provides more accurate predictions than either traditional or corrected mean-field descriptions and, unlike these descriptions, does not require that $P_p/P_m \ll 1$ and $P_d/P_m \ll 1$ to give accurate predictions [2]. Additionally, our description provides predictions about the clustering present in the system, unlike previous continuum descriptions, and these predictions match the averaged results from the discrete model extremely well, even with significant adhesion or repulsion. Furthermore, we derive a simple analytic approximation of the spatial clustering of occupied sites when the system has reached steady-state, and demonstrate that this approximation is accurate.

We could extend the description presented here in several ways. One extension could be to introduce spatial dependence [13], which would allow us to simulate processes such as scratch assays. Scratch assays are a common experimental procedure used to study collective cell behavior, where the initial distribution of cells depends on spatial location [15]. This extension would require $\mathcal{O}(N^2)$ ODEs, rather than the $\mathcal{O}(N)$ ODEs in the description presented here, to describe the location and length of the chains and gaps, and we leave the description of these processes for future work. Alternatively, we could introduce a chemical species and couple the ABDM parameters to the chemical concentration to model cellular nutrient uptake [9]. As the C&G description describes the dynamics of every possible chain and gap it would be relatively straightforward to consider non-constant parameters (P_m, P_p, P_d, α) that change with, for example, time, chain or gap length. Another extension would be to derive and analyze the partial differential equation approximation of the C&G description, similar to the work carried out by Markham *et* *al.* [19]. This extension would require a Taylor series expansion of C_n and G_m in an appropriate limit, such that n and m can be treated as continuous. We leave this extension for future analysis.

The calibration of traditional continuum descriptions to experimental ABDM data to obtain parameter estimates results in incorrect estimates when spatial correlations become significant [27]. Calibrating stochastic models to experimental data is significantly more computationally expensive [16], compared to dealing with continuum approximations. Non-Bayesian data calibration techniques have also been proposed [15]. However, these approaches can also be computationally expensive as they require the calculation of average discrete behavior across a potentially large parameter space. It would therefore be of interest to investigate the accuracy of parameter estimates obtained from calibrating the C&G description to experimental data in the future.

Acknowledgements. This work is supported by the Australian Research Council (FT130100148). We appreciate the assistance of Professor Sean McElwain, and the support provided by the High Performance Computing Centre at QUT. We also thank the three referees for their helpful comments.

Bibliography

[1] Anguige, K. and Schmeiser, C. [2009], 'A one-dimensional model of cell diffusion and aggregation, incorporating volume filling and cell-to-cell adhesion', *Journal of Mathematical Biology* **58**(3), 395–427.

[2] Baker, R. E. and Simpson, M. J. [2010], 'Correcting mean-field approximations for birth-death-movement processes', *Physical Review E* **82**(4), 041905.

[3] Borodin, A., Corwin, I. and Sasamoto, T. [2014], 'From duality to determinants for q-TASEP and ASEP', *Annals of Probability* **42**(6), 2314–2382.

[4] Chowdhury, D., Schadschneider, A. and Nishinari, K. [2005], 'Physics of transport and traffic phenomena in biology: from molecular motors and cells to organisms', *Physics of Life Reviews* **2**(4), 318–352.

[5] Codling, E. A., Plank, M. J. and Benhamou, S. [2008], 'Random walk models in biology', *Journal of the Royal Society Interface* **5**(25), 813–834.

[6] Deroulers, C., Aubert, M., Badoual, M. and Grammaticos, B. [2009], 'Modeling tumor cell migration: from microscopic to macroscopic models', *Physical Review E* **79**(3), 031917.

[7] Derrida, B., Evans, M., Hakim, V. and Pasquier, V. [1993], 'Exact solution of a 1D asymmetric exclusion model using a matrix formulation', *Journal of Physics A* **26**(7), 1493.

[8] Fernando, A. E., Landman, K. A. and Simpson, M. J. [2010], 'Nonlinear diffusion and exclusion processes with contact interactions', *Physical Review E* **81**(1), 011903.

[9] Gerlee, P. and Anderson, A. R. A. [2008], 'A hybrid cellular automaton model of clonal evolution in cancer: the emergence of the glycolytic phenotype', *Journal of Theoretical Biology* **250**(4), 705–722.

[10] Gillespie, D. T. [1977], 'Exact stochastic simulation of coupled chemical reactions', *Journal of Physical Chemistry* **81**(25), 2340–2361.

[11] Grima, R. [2008], 'Multiscale modeling of biological pattern formation', *Current Topics in Developmental Biology* **81**, 435–460.

[12] Hackett-Jones, E. J., Landman, K. A. and Fellner, K. [2012], 'Aggregation patterns from nonlocal interactions: Discrete stochastic and continuum modeling', *Physical Review E* **85**(4), 041912.

[13] Illien, P., Bénichou, O., Mejía-Monasterio, C., Oshanin, G. and Voituriez, R. [2013], 'Active transport in dense diffusive single-file systems', *Physical Review Letters* **111**(3), 038102.

[14] Johnston, S. T., Simpson, M. J. and Baker, R. E. [2012], 'Mean-field descriptions of collective migration with strong adhesion', *Physical Review E* **85**(5), 051922.

[15] Johnston, S. T., Simpson, M. J. and McElwain, D. L. S. [2014], 'How much information can be obtained from tracking the position of the leading edge in a scratch assay?', *Journal of the Royal Society Interface* **11**(97), 20140325.

[16] Johnston, S. T., Simpson, M. J., McElwain, D. L. S., Binder, B. J. and Ross, J. V. [2014], 'Interpreting scratch assays using pair density dynamics and approximate Bayesian computation', *Open Biology* **4**(9), 140097.

[17] Law, R. and Dieckmann, U. [2000], 'A dynamical system for neighborhoods in plant communities', *Ecology* **81**(8), 2137–2148.

[18] Mackie, A. D., Panagiotopoulos, A. Z. and Szleifer, I. [1997], 'Aggregation behavior of a lattice model for amphiphiles', *Langmuir* **13**(19), 5022–5031.

[19] Markham, D. C., Simpson, M. J. and Baker, R. E. [2013], 'Simplified method for including spatial correlations in mean-field approximations', *Physical Review E* **87**(6), 062702.

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

[21] Nardini, J. T., Chapnick, D. A., Liu, X., Bortz, D. et al. [n.d.], 'The effects of MAPK activity on cell-cell adhesion during wound healing', *arXiv preprint* .

[22] Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. [2007], *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, Cambridge University Press.

[23] Sharkey, K. J., Fernandez, C., Morgan, K. L., Peeler, E., Thrush, M., Turnbull, J. F. and Bowers, R. G. [2006], 'Pair-level approximations to the spatio-temporal dynamics of epidemics on asymmetric contact networks', *Journal of Mathematical Biology* **53**(1), 61–85.

[24] Sharkey, K. J., Kiss, I. Z., Wilkinson, R. R. and Simon, P. L. [2015], 'Exact equations for SIR epidemics on tree graphs', *Bulletin of Mathematical Biology* **77**(4), 614–645.

[25] Simpson, M. J. and Baker, R. E. [2011], 'Corrected mean-field models for spatially dependent advection-diffusion-reaction phenomena', *Physical Review E* **83**(5), 051922.

[26] Simpson, M. J., Landman, K. A. and Hughes, B. D. [2010], 'Cell invasion with proliferation mechanisms motivated by time-lapse data', *Physica A: Statistical Mechanics and its Applications* **389**(18), 3779–3790.

[27] Simpson, M. J., Sharp, J. A. and Baker, R. E. [2014], 'Distinguishing between mean-field, moment dynamics and stochastic descriptions of birth–death–movement processes', *Physica A: Statistical Mechanics and its Applications* **395**, 236–246.

[28] Singer, A. [2004], 'Maximum entropy formulation of the Kirkwood superposition approximation', *Journal of Chemical Physics* **121**(8), 3657–3666.

[29] Taylor, M., Simon, P. L., Green, D. M., House, T. and Kiss, I. Z. [2012], 'From markovian to pairwise epidemic models and the performance of moment closure approximations', *Journal of Mathematical Biology* **64**(6), 1021–1042.

[30] Turner, S. and Sherratt, J. A. [2002], 'Intercellular adhesion and cancer invasion: a discrete simulation using the extended potts model', *Journal of Theoretical Biology* **216**(1), 85–100.

[31] Turner, S., Sherratt, J. A., Painter, K. J. and Savill, N. J. [2004], 'From a discrete to a continuous model of biological cell movement', *Physical Review E* **69**(2), 021910.

[32] Wilkinson, R. R. and Sharkey, K. J. [2014], 'Message passing and moment closure for susceptible-infected-recovered epidemics on finite networks', *Physical Review E* **89**(2), 022808.

[33] Zahn, K., Maret, G., Ruß, C. and von Grünberg, H. H. [2003], 'Three-particle correlations in simple liquids', *Physical Review Letters* **91**(11), 115502.

7.7 Supplementary material

7.7.1 Chain-and-gap description

If we consider all possible events that change the number of chains and gaps of length $n \in [1, N]$, where N is the number of lattice sites, and the rate at which each event occurs we obtain the following system of ordinary differential equations.

Chains

$$\begin{aligned}\frac{\mathrm{d}C_1}{\mathrm{d}t} =& P_m\Bigg[(1-\alpha)C_2 + \sum_{i=2}^{N-2}\Big\{(1-\alpha)C_i\Big\} - \frac{1}{2}\Big\{Y(C_1,G_1,C_1) + \sum_{j=1}^{N-3}\big(Y(C_1,G_1,C_j)\big)\Big\} \\ & - \frac{1-\alpha}{2}\sum_{i=2}^{N-3}\Big\{Y(C_i,G_1,C_1) + \sum_{j=1}^{N-i-2}\big(Y(C_i,G_1,C_j)\big)\Big\}\Bigg] + \\ & P_p\Bigg[-C_1 - \frac{1}{2}\sum_{i=1}^{N-3}\Big\{Y(C_i,G_1,C_1)\Big\}\Bigg] + P_d\Bigg[-C_1 + 2\sum_{i=2}^{N-1}\Big\{C_i\Big\}\Bigg],\end{aligned}$$

$$\begin{aligned}\frac{\mathrm{d}C_k}{\mathrm{d}t} =& P_m\Bigg[(1-\alpha)(C_{k+1}-C_k) + \frac{1}{2}\Big\{Y(C_1,G_1,C_{k-1}) - Y(C_1,G_1,C_k)\Big\} \\ & + \frac{1-\alpha}{2}\sum_{i=2}^{N-k-2}\Big\{Y(C_i,G_1,C_{k-1}) - Y(C_i,G_1,C_k)\Big\} + \frac{1-\alpha}{2}Y(C_{N-k-1},G_1,C_{k-1})\Bigg] \\ & + P_p\Bigg[C_{k-1} - C_k - \frac{1}{2}\sum_{i=1}^{N-k-1}\Big\{Y(C_{k-1},G_1,C_i)\Big\} - \frac{1}{2}\sum_{i=1}^{N-k-2}\Big\{Y(C_i,G_1,C_k)\Big\} + \\ & \frac{1}{2}\sum_{i=1}^{k-2}\Big\{Y(C_i,G_1,C_{k-i-1})\Big\}\Bigg] + P_d\Bigg[-kC_k + 2\sum_{i=k+1}^{N-1}\Big\{C_i\Big\}\Bigg], \\ & \text{for } k = 2,\ \ldots,\ N-3,\end{aligned}$$

$$\begin{aligned}\frac{\mathrm{d}C_{N-2}}{\mathrm{d}t} =& P_m\Bigg[\frac{1}{2}Y(C_1,G_1,C_{N-3}) - (1-\alpha)C_{N-2}\Bigg] \\ & + P_p\Bigg[C_{N-3} - C_{N-2} - \frac{1}{2}Y(C_{N-3},G_1,C_1) + \frac{1}{2}\sum_{i=1}^{N-4}\Big\{Y(C_i,G_1,C_{N-i-3})\Big\}\Bigg] \\ & + P_d\Bigg[-(N-2)C_{N-2} + 2C_{N-1}\Bigg],\end{aligned}$$

$$\frac{\mathrm{d}C_{N-1}}{\mathrm{d}t} = P_p\Bigg[C_{N-2} - C_{N-1} + \frac{1}{2}\sum_{i=1}^{N-3}Y(C_i,G_1,C_{N-i-2})\Bigg] + P_d\Bigg[-(N-1)C_{N-1} + NC_N\Bigg],$$

$$\frac{\mathrm{d}C_N}{\mathrm{d}t} = P_p\Bigg[C_{N-1}\Bigg] + P_d\Bigg[-NC_N\Bigg],$$

where

$$Y(C_i, G_1) = C_i\left[P(G_1|C_i) + \sum_{j=1}^{N-i-2} P(G_j|C_i)P(G_i|C_i, G_j)\right],$$

represents the number of chains of length i that are next to a gap of length one and

$$\begin{aligned} Y(C_i, G_1, C_j) =& C_i\Bigg[P(G_1|C_i)P(C_j|C_i, G_1) \\ &+ \sum_{k=1}^{N-i-j-1} P(G_k|C_i)P(G_1|C_i, G_k)P(C_j|C_i, G_k, G_1)\Bigg], \end{aligned}$$

represents the number of chains of length i that are separated from a chain of length j by a chain of length one.

$$P(G_j|C_i) = \frac{G_j}{\sum_m G_m},\ m \in A_G(C_i),$$

represents the probability that, provided we have selected a chain of length i, there is a neighboring gap of length j, where

$$A_G(C_i) = \{m \in \mathbb{Z}|1 \leq m \leq N - j - 2\} \cup \{N - j\}.$$

are the possible gap lengths, given that we have selected a chain of length i.

$$P(G_k|C_i, G_j) = \frac{G_k}{\sum_m G_m},\ m \in A_G(C_i, G_j),$$

represents the probability that, provided we have selected a chain of length i with a neighboring gap of length j, there is a neighboring gap of length k, where

$$A_G(C_i, G_j) = \{m \in \mathbb{Z}|1 \leq m \leq N - i - j - 1\},$$

are the possible gaps, given that we have selected a chain of length i with a neighboring gap of length j.

$$P(C_k|C_i, G_j) = \frac{C_k}{\sum_m C_m},\ m \in A_C(C_i, G_j),$$

represents the probability that, provided we have selected a chain of length i with a neighboring gap of length j, there is a neighboring chain of length k, where

$$A_C(C_i, G_j) = \{m \in \mathbb{Z}|1 \leq m \leq N - i - j - 1\},$$

are the possible chains, given that we have selected a chain of length i with a neighboring gap of length j.

$$P(C_l|C_i, G_j, G_k) = \frac{C_l}{\sum_m C_m}, \; m \in A_C(C_i, G_j, G_k),$$

represents the probability that, provided we have selected a chain of length i with neighboring gaps of length j and k, there is a neighboring chain of length l, where

$$A_C(C_i, G_j, G_k) = \{m \in \mathbb{Z} | 1 \leq m \leq N - i - j - k - 2\} \cup \{N - i - j - k\},$$

are the possible chains, given that we have selected a chain of length i with neighboring gaps of length j and k. We note that the denominator in the probability expressions can be zero. However, this occurs if and only if the numerator is zero and, in terms of probability, is intuitive to interpret as zero.

Gaps

$$\begin{aligned}
\frac{\mathrm{d}G_1}{\mathrm{d}t} =& P_m \Bigg[-\frac{1}{2} \sum_{i=1}^{N-3} Y(C_1, G_1, C_i) + (1-\alpha) \sum_{i=2}^{N-2} \Big\{ C_i \Big\} - \frac{1}{2} \sum_{i=1}^{N-3} \Big\{ Y(C_1, G_1, C_i) \Big\} \\
& - \frac{1-\alpha}{2} \sum_{i=2}^{N-3} \sum_{j=1}^{N-i-2} \Big\{ Y(C_i, G_1, C_j) \Big\} + \frac{1}{2} \sum_{i=1}^{N-4} \Big\{ Y(C_1, G_2, C_i) \Big\} \\
& + \frac{1-\alpha}{2} \sum_{i=2}^{N-4} \sum_{j=1}^{N-i-3} \Big\{ Y(C_i, G_2, C_j) \Big\} + (1-\alpha) Y(C_{N-2}, G_2) \Bigg] + P_p \Bigg[G_2 - G_1 \Bigg] \\
& + P_d \Bigg[N C_N + \sum_{i=3}^{N-1} \Big\{ (i-2) C_i \Big\} - \sum_{i=2}^{N-3} \sum_{j=1}^{N-i-2} \Big\{ Y(C_i, G_1, C_j) \Big\} \\
& - 2Y(C_{N-1}, G_1) - \sum_{i=1}^{N-3} \Big\{ Y(C_1, G_1, C_i) \Big\} \Bigg],
\end{aligned}$$

$$\begin{aligned}
\frac{\mathrm{d}G_k}{\mathrm{d}t} =& P_m\Bigg[\frac{1}{2}\sum_{i=1}^{N-k-1}\Big\{Y(C_1,G_{k-1},C_i)\Big\} - \frac{1}{2}\sum_{i=1}^{N-k-2}\Big\{Y(C_1,G_k,C_i)\Big\} \\
& -\frac{1}{2}\sum_{i=1}^{N-k-2}\Big\{Y(C_1,G_k,C_i)\Big\} - \frac{1-\alpha}{2}\sum_{i=2}^{N-k-2}\sum_{j=1}^{N-k-i-1}\Big\{Y(C_i,G_k,C_j)\Big\} \\
& +\sum_{i=1}^{N-k-3}\Big\{Y(C_1,G_{k+1},C_i)\Big\} + \frac{1-\alpha}{2}\sum_{i=2}^{N-k-3}\sum_{j=1}^{N-k-i-2}\Big\{Y(C_i,G_{k+1},C_j)\Big\} \\
& +(1-\alpha)\big(Y(C_{N-k-1},G_{k+1}) - Y(C_{N-k},G_k)\big)\Bigg] + \\
& P_p\Bigg[G_{k+1}-G_k\Bigg] + P_d\Bigg[\sum_{i=2}^{N-k-1}\sum_{j=1}^{N_k-1}\Big\{Y(C_i,G_{k-1},C_j)\Big\} + 2Y(C_{N-k+1},G_{k-1}) \\
& -\sum_{i=2}^{N-k-2}\sum_{j=1}^{N-k-i-1}\Big\{Y(C_i,G_k,C_j)\Big\} - 2Y(C_{N-k},G_k) \\
& -\sum_{i=1}^{N-k-1}Y(C_1,G_k,C_i) + \frac{1}{2}\sum_{i=1}^{k-2}\Big\{Y(C_1,G_i,G_{k-i-1})\Big\}\Bigg], \qquad \text{for } k=2,\ \ldots,\ N-3,
\end{aligned}$$

$$\begin{aligned}
\frac{\mathrm{d}G_{N-2}}{\mathrm{d}t} =& P_m\Bigg[\frac{1}{2}Y(C_1,G_{N-3},C_1) - (1-\alpha)Y(C_2,G_{N-2})\Bigg] + P_p\Bigg[G_{N-1}-G_{N-2}\Bigg] + \\
& P_d\Bigg[2Y(C_3,G_{N-3}) - 2Y(C_2,G_{N-2}) + \frac{1}{2}\sum_{i=1}^{N-4}Y(C_1,G_i,G_{N-i-3})\Bigg],
\end{aligned}$$

$$\frac{\mathrm{d}G_{N-1}}{\mathrm{d}t} = P_p\Bigg[-G_{N-1}\Bigg] + P_d\Bigg[2Y(C_2,G_{N-2}) - G_{N-1} + \frac{1}{2}\sum_{i=1}^{N-3}Y(C_1,G_i,G_{N-i-2})\Bigg],$$

$$\frac{\mathrm{d}G_N}{\mathrm{d}t} = P_d\Bigg[G_{N-1}\Bigg],$$

where

$$Y(C_1,G_i,G_j) = C_1\big[P(G_i|C_1)P(G_j|C_1,G_i) + P(G_j|C_1)P(G_i|C_1,G_j)\big]$$

represents the number of chains of length one next to both a gap of length i and a gap of length j.

7.7.2 Small birth and death rates

Baker and Simpson [14] demonstrate that both mean-field and corrected mean-field descriptions provides predictions that match the average discrete behavior well, provided that $P_p/P_m \ll 1$ and $P_d/P_m \ll 1$. In Figure 7.7 we show that the C&G description provides similarly accurate predictions in these parameter regimes.

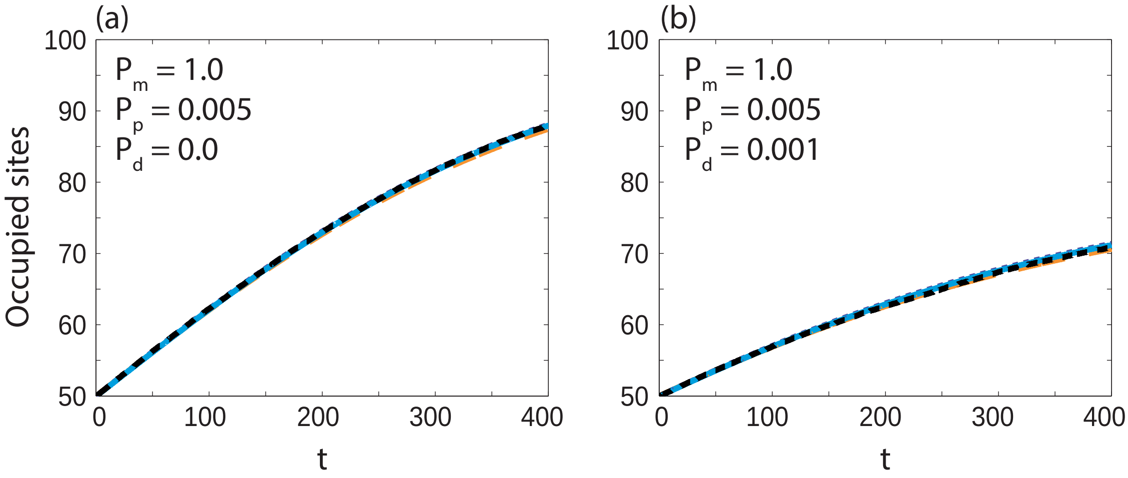

Figure 7.7: Comparison between averaged discrete (black, dashed), mean-field (purple), corrected mean-field (orange) and C&G (cyan) results for parameter regimes where $P_p/P_m \ll 1$ and $P_d/P_m \ll 1$. For all results $M = 1000$, $N = 100$ and $\alpha = 0$. Initially we have 50 occupied lattice sites.

7.7.3 Steady-state approximation

We solve the C&G model until late-time, that is, where $\mathrm{d}C_n/\mathrm{d}t \approx 0$ and $\mathrm{d}G_n/\mathrm{d}t \approx 0$. We note that the time required for $\mathrm{d}C_n/\mathrm{d}t \approx 0$ and $\mathrm{d}G_n/\mathrm{d}t \approx 0$ is dependent on the parameters and the initial condition. In practice, we solve the C&G model until the solution is approximately constant with respect to time, rather than algebraically solving for the steady-state. For the results presented in Fig. 7.5 in the main document, the C&G model was approximately at steady-state.

At steady-state, the net birth rate must be zero and hence

$$\frac{P_p}{2}(2N_S + N_E) = P_d(N_S + N_E + N_M).$$

We know that $N_S = C_1$ and that there are two edge agents and $n-2$ middle agents for chains of $n \geq 2$, except the special case where $n = N$ and all agents are middle agents. Therefore,

$$N_E = 2\sum_{i=2}^{N-1} C_i, \;\; N_M = NC_N + \sum_{i=2}^{N-1}(i-2)C_i.$$

If we rewrite the net birth rate in terms of C_n we obtain

$$\frac{P_p}{2}\left[2C_1 + 2\sum_{i=2}^{N-1} C_i\right] = P_d\left[C_1 + 2\sum_{i=2}^{N-1} C_i + NC_N + \sum_{i=2}^{N-1}(i-2)C_i\right],$$

and, simplifying,

$$P_p\left[\sum_{i=1}^{N-1} C_i\right] = P_d\left[\sum_{i=1}^{N} iC_i\right].$$

Given that C_n is linear on a $\log_{10}$ scale at late-time, we make the assumption that $C_n = \beta^{n-1}C_1$, $\beta > 0$, and, subsequently,

$$P_p C_1 \left[\sum_{i=1}^{N-1} \beta^{i-1} \right] = P_d C_1 \left[\sum_{i=1}^{N} i\beta^{i-1} \right].$$

In the limit $N \to \infty$, both the left- and right-hand side of the net birth rate have a power series representation. As N is finite in our model, we therefore have an approximation, namely,

$$\sum_{i=1}^{N-1} \beta^{i-1} \approx \frac{1}{1-\beta},$$

and

$$\sum_{i=1}^{N} i\beta^{i-1} \approx \frac{1}{(1-\beta)^2},$$

which require $|\beta| < 1$ for convergence. The net birth rate can therefore be expressed as

$$\frac{P_p C_1}{1-\beta} = \frac{P_d C_1}{(1-\beta)^2},$$

which can be rearranged to show that

$$\beta = 1 - \frac{P_d}{P_p}.$$

We note that we compared this approximation with β calculated from the C&G model when $\mathrm{d}C_n/\mathrm{d}t \approx 0$ and $\mathrm{d}G_m/\mathrm{d}t \approx 0$, and that the approximation matches the solution of the C&G model very well.

7.7.4 Numerical methods

The system of governing equations is implemented and solved in C with an adaptive Runge-Kutta method [29]. To increase computational efficiency, the probability terms are evaluated and stored at the beginning of each time step so that the probability terms are not calculated repeatedly. Subsequently, the code produces solutions to the system of governing equations within seconds on a desktop computer.

Chapter 8

Co-operation, competition and crowding: a discrete framework linking Allee kinetics, nonlinear diffusion, shocks and sharp-fronted travelling waves

A paper published in *Scientific Reports*.

Johnston, S. T., Baker, R. E., McElwain, D. L. S. and Simpson, M. J. [2017], 'Co-operation, competition and crowding: A discrete framework linking Allee kinetics, nonlinear diffusion, shocks and sharp-fronted travelling waves', *Scientific Reports* **7**, 421134.

8.1 Abstract

Invasion processes are ubiquitous throughout cell biology and ecology. During invasion, individuals can become isolated from the bulk population and behave differently. We present a discrete, exclusion-based process that models the birth, death and movement of individuals. The model distinguishes between individuals that are part of, or are isolated from, the bulk population by imposing different rates of birth, death and movement. This enables the simulation of various co-operative or competitive mechanisms, where there is a positive or negative benefit associated with being part of the bulk population, respectively. The mean-field approximation of the discrete process gives rise to 22 different classes of partial differential equation, which include Allee kinetics and nonlinear diffusion. Here we examine the ability of each class of partial differential equation to support travelling wave solutions and interpret the long time behaviour in terms of the individual-level parameters. For the first time we show that the strong Allee effect and nonlinear diffusion can result in shock-fronted travelling waves. We also demonstrate how differences in group and individual motility rates can influence the persistence of a population and provide conditions for the successful invasion of a population.

8.2 Introduction

Processes where individuals invade, and subsequently colonise, a region of space are prevalent in cell biology and ecology [2, 11, 19–21, 33, 48, 52, 61, 66, 68]. In cell biology, wound healing involves the invasion of fibroblasts into the wound space for tissue regeneration [48]. The invasion of glioma cells throughout the brain can lead to the formation of malignant brain tumours (glioblastoma) [2, 19, 66]. In ecology, the introduction and subsequent invasion of an alien species is a significant factor contributing to the extinction of native species [11, 68].

During invasion, individuals that become separated from the bulk population have been observed to have different behaviours to individuals within the bulk population [5, 13, 22, 52, 67]. This is intuitive in ecological processes, as a decrease in the number of individuals within the bulk population can reduce the number of potential mates [13, 57, 67, 70] or lessen the efficacy of predator avoidance [22, 57, 67]. In cell biology, individual micrometastases have been observed to have reduced growth rates below a threshold size, which suggests that the presence of additional cells enhances the birth rate [5].

Continuum mathematical models of invasion processes have been studied extensively since the Fisher-Kolmogorov model was first proposed in 1937 [1, 4, 7, 9, 15–17, 26, 34, 36, 41, 49, 54, 67]. The Fisher-Kolmogorov model is a partial differential equation (PDE) description of the evolution of population density, where the temporal change in population density is attributed to a combination of linear diffusion and logistic growth [17, 36]. The Fisher-Kolmogorov model has been applied to various problems in cell biology and ecology [30, 44, 45, 63]. The logistic growth term implies that the population density will always tend toward the carrying capacity [49]. This prediction does not reflect the observation that isolated individuals can experience a reduction in their birth rate [22]. This effect, known as the Allee effect, has two known forms. First, the strong Allee effect, where the growth rate is negative for sufficiently low densities [28, 67]. Second, the weak Allee effect, where the growth rate is reduced, but remains positive, at low densities [67]. Reaction-diffusion PDEs incorporating linear diffusion and Allee growth kinetics have been proposed and analysed [4, 15, 16, 26, 34, 41, 54, 67]. A key feature of interest for models of invasion is whether the PDE supports a travelling wave solution, where a wave front of constant shape moves through space with a constant speed. The sign of the wave speed indicates whether successful invasion occurs, and the magnitude of the wave speed provides an estimate of how quickly a population invades or recedes. More complicated descriptions of invasion processes with either Fisher or Allee kinetics and density-dependent nonlinear diffusion have been proposed, with the motivation of describing spatial aggregation or segregation [14, 38, 42, 43, 55, 56].

A key feature of the Fisher-Kolmogorov model, and many extensions thereof, is that these PDE models are typically derived using phenomenologically-based arguments without incorporating information from an underlying stochastic description of individual-level behaviours. In this work we consider a relatively straightforward lattice-based discrete birth-death-movement model. An important characteristic of the model is that it explicitly accounts for crowding effects by only allowing for one agent per lattice site. Additionally, the rates at which birth, death and movement events occur depend on whether an agent is part of a group of agents or is isolated. We demonstrate that the standard continuum approximation of this discrete model can lead to either logistic or Allee kinetics, in an appropriate parameter regime. Furthermore, we demonstrate that imposing a different motility rate for agents that are isolated, compared to other agents, leads to a variety of density-dependent nonlinear diffusion functions. Previous studies have examined many different types of phenomenologically-based PDEs that are motivated in an *ad hoc* fashion. In contrast, our PDE description arises from a single, relatively simple, physically-motivated model. In Table 8.1 we highlight this generality, as the single discrete model gives rise to 22 different classes of PDE that describe the population-level behaviour.

While several of these PDEs have been studied previously, for completeness we examine the ability of each class of PDE to support travelling wave solutions. For certain classes of PDE, we present details of the travelling wave solutions for the first time. Interestingly, we obtain travelling wave solutions for PDEs that have nonlinear diffusivity functions with regions of negative diffusivity. Furthermore, we show that the strong Allee effect combined with these diffusivity functions can lead to novel shock-fronted travelling wave solutions. As these diffusivity functions are obtained directly from a discrete model, we can determine which competitive/co-operative individual-level mechanisms result in shock-fronted travelling wave solutions. Similarly, we are able to interpret the influence of motility on the persistence of a population, and highlight how this influence varies nonlinearly with the carrying capacity density. More generally, we provide new insight into the long time behaviour of an invasive population in terms of its individual-level properties.

	Nonlinear diffusivity	Degenerate diffusivity	Number of non-degenerate zeroes	Negative diffusivity	Source Term	Grouped agent death	Relevant previous analysis
Case 1	x	x	0	x	Fisher	x	[1, 4, 7, 9, 17, 26, 36, 49]
Case 2: Sub-case 1	✓	x	0	x	Fisher	x	[23–25]
Case 2: Sub-case 2	✓	✓	0	x	Fisher	x	[46, 55, 60, 62]
Case 2: Sub-case 3	✓	x	2	✓	Fisher	x	[14]
Case 2: Sub-case 4	✓	✓	1	✓	Fisher	x	[42]
Case 3	x	x	0	x	Fisher	✓	[1, 4, 7, 9, 17, 26, 36]
Case 4: Sub-case 1	✓	x	0	x	Fisher	✓	[23–25]
Case 4: Sub-case 2	✓	✓	0	x	Fisher	✓	[46, 55, 60, 62]
Case 4: Sub-case 3	✓	x	2	✓	Fisher	✓	[14]
Case 4: Sub-case 4	✓	✓	1	✓	Fisher	✓	[42]
Case 4: Sub-case 5	✓	x	1	✓	Fisher	✓	[42]
Case 5	x	x	0	x	Allee	x	[4, 15, 16, 26, 34, 41, 54, 67]
Case 6: Sub-case 1	✓	x	0	x	Allee	x	[46, 50]
Case 6: Sub-case 2	✓	✓	0	x	Allee	x	[47, 56]
Case 6: Sub-case 3	✓	x	2	✓	Allee	x	[38]
Case 6: Sub-case 4	✓	✓	1	✓	Allee	x	[43]
Case 7	x	x	0	x	Allee	✓	[4, 15, 16, 26, 34, 41, 54, 67]
Case 8: Sub-case 1	✓	x	0	x	Allee	✓	[46, 50]
Case 8: Sub-case 2	✓	✓	0	x	Allee	✓	[47, 56]
Case 8: Sub-case 3	✓	x	2	✓	Allee	✓	[38]
Case 8: Sub-case 4	✓	✓	1	✓	Allee	✓	[43]
Case 8: Sub-case 5	✓	x	1	✓	Allee	✓	[43]

Table 8.1: Different classes of PDE associated with the discrete model in appropriate parameter regimes. An Allee source term can correspond to either the weak, strong or reverse Allee effect. Degenerate diffusivity refers to the case where $F(C^*) = R(C^*) = 0$ for some value C^*. A detailed analysis of all cases is presented in the Supplementary Material.

8.3 Results

We consider a discrete lattice-based exclusion process where agents undergo birth, death and movement events. We distinguish between isolated agents and grouped agents by imposing different rates of birth, death and movement depending on whether an agent has zero or at least one nearest-neighbour agent, respectively. A more detailed description of the discrete model is presented in the Methods. To derive a continuum limit PDE description of the discrete model [31, 65] we consider the change in occupancy of a lattice site j during a single time step of duration τ, and obtain

$$
\begin{aligned}
\delta C_j =& \frac{P_m^i}{2}\Big[C_{j-1}(1-C_j)(1-C_{j-2}) + C_{j+1}(1-C_j)(1-C_{j+2}) - 2C_j(1-C_{j-1})(1-C_{j+1})\Big] \\
&+ \frac{P_m^g}{2}\Big[C_{j-1}(1-C_j) + C_{j+1}(1-C_j) - C_j(1-C_{j-1}) - C_j(1-C_{j+1})\Big] \\
&- \frac{P_m^g}{2}\Big[C_{j-1}(1-C_j)(1-C_{j-2}) + C_{j+1}(1-C_j)(1-C_{j+2}) \\
&- 2C_j(1-C_{j-1})(1-C_{j+1})\Big] \\
&+ \frac{P_p^i}{2}\Big[C_{j-1}(1-C_j)(1-C_{j-2}) + C_{j+1}(1-C_j)(1-C_{j+2})\Big] \\
&+ \frac{P_p^g}{2}\Big[C_{j-1}(1-C_j) + C_{j+1}(1-C_j)\Big] \\
&- \frac{P_p^g}{2}\Big[C_{j-1}(1-C_j)(1-C_{j-2}) + C_{j+1}(1-C_j)(1-C_{j+2})\Big] \\
&- P_d^i\Big[C_j(1-C_{j-1})(1-C_{j+1})\Big] - P_d^g\Big[C_j\Big] + P_d^g\Big[C_j(1-C_{j-1})(1-C_{j+1})\Big]. \qquad (8.1)
\end{aligned}
$$

Here, C_j represents the probability that the site j is occupied and, therefore, $1-C_j$ represents the probability that the site is vacant [65]. Furthermore, as products of probabilities are interpreted as net transition probabilities, the usual assumption that the occupancy of lattice sites are independent is made [6, 29, 31, 32, 64].

Note that C_j is the total occupancy of site j, that is, the sum of the occupancy of isolated agents and the occupancy of grouped agents at that site. We now interpret the terms on the right-hand side of Equation (8.1) in terms of the physical change in lattice occupancy. The positive terms proportional to P_m^i correspond to isolated agents moving into site j, while the negative terms correspond to isolated agents moving out of site j. Each term consists of three factors. For the negative terms, these factors are the probability that site j is occupied, and the probabilities that sites $j-1$ and $j+1$ are vacant. For the positive terms, the three factors are the probability that site $j \pm 1$ is occupied, and the probabilities that sites j and $j \pm 2$ are vacant. The third factor is required to ensure that the term describes isolated agents. The positive/negative terms proportional to the first P_m^g term on the right-hand side of Equation (8.1) correspond to grouped agents moving in/out of site j. These terms consist of two factors; the probability that the selected site is occupied and the probability that the target site is vacant. The second P_m^g term ensures that the isolated agents are not counted twice. The remaining terms can be interpreted similarly; products of probabilities that specific sites are occupied or vacant that describe the change of occupancy of a site in response to a birth or death event.

To obtain a PDE description we divide Equation (8.1) by τ and consider C_j as a continuous function, $C(x,t)$. We expand $C(x,t)$ in a Taylor series around $x = j\Delta$, truncating terms of $\mathcal{O}(\Delta^3)$, where Δ is the lattice spacing [31, 65]. Taking the limit $\Delta \to 0$ and $\tau \to 0$ such that Δ^2/τ is held constant [12, 27, 65] gives

$$\begin{aligned}\frac{\partial C}{\partial t} = D_g\frac{\partial^2 C}{\partial x^2} + \Big(D_i - D_g\Big)\frac{\partial}{\partial x}\left(\Big(1-4C+3C^2\Big)\frac{\partial C}{\partial x}\right) + \lambda_g C\Big(1-C\Big) \\ + \Big(\lambda_i - \lambda_g\Big)C\Big(1-C\Big)^2 - K_g C - \Big(K_i - K_g\Big)C\Big(1-C\Big)^2, \qquad (8.2)\end{aligned}$$

where

$$\begin{aligned}D_g = \lim_{\Delta,\tau\to 0}\frac{P_m^g\Delta^2}{2\tau}, \quad D_i = \lim_{\Delta,\tau\to 0}\frac{P_m^i\Delta^2}{2\tau}, \quad \lambda_g = \lim_{\tau\to 0}\frac{P_p^g}{\tau}, \quad \lambda_i = \lim_{\tau\to 0}\frac{P_p^i}{\tau}, \\ K_g = \lim_{\tau\to 0}\frac{P_d^g}{\tau}, \quad K_i = \lim_{\tau\to 0}\frac{P_d^i}{\tau}, \qquad (8.3)\end{aligned}$$

with the further assumption that P_p^i, P_p^g, P_d^i, P_d^g are $\mathcal{O}(\tau)$ [65]. The individual-level parameters are treated as being interchangeable with the continuum-level parameters as defined in (8.3). All implementations of the discrete model in this work have $\Delta = \tau = 1$.

It is convenient to write Equation (8.2) in conservation form

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x}\left(F(C)\frac{\partial C}{\partial x}\right) + R(C), \qquad (8.4)$$

where

$$F(C) = D_i(1-4C+3C^2) + D_g(4C-3C^2), \qquad (8.5)$$

is the nonlinear diffusivity function, and

$$R(C) = \lambda_g C(1-C) + (\lambda_i - \lambda_g - K_i + K_g)C(1-C)^2 - K_g C, \qquad (8.6)$$

is the source/sink term.

The aims of this work are to first illustrate that the very different types of behaviour encoded in the discrete model are also reflected in the solution of Equation (8.2). Once we have demonstrated this connection, we focus on examining travelling wave solutions for the 22 different classes embedded within Equation (8.2), as summarised in Table 8.1. In the main document we highlight novel and key results for specific classes of PDEs resulting from the discrete model, and provide relevant discussion about the implications of the long time population behaviour. A more thorough investigation of the travelling wave solutions arising from all 22 classes of PDEs is presented in the Supplementary Material.

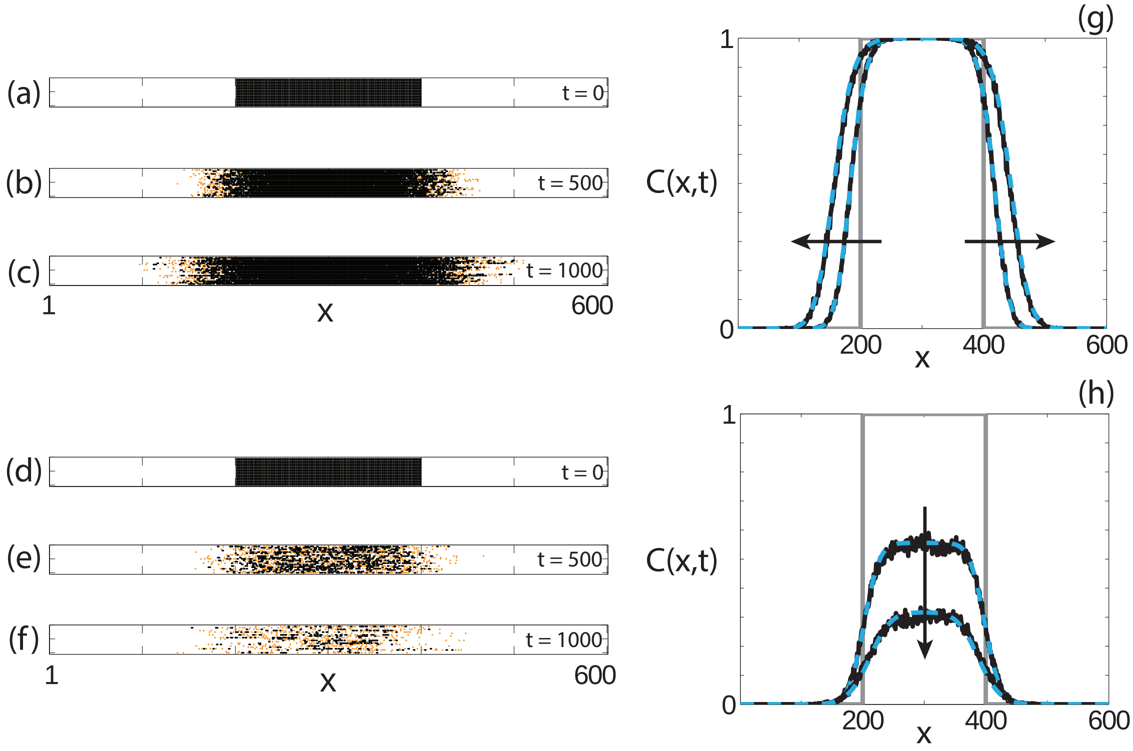

Figure 8.1: Comparison of the discrete model and the continuum approximation. (a)-(f) 20 identically-prepared realisations of the discrete model at (a), (d) $t = 0$; (b), (e) $t = 500$; (c), (f) $t = 1000$. The discrete model simulations correspond to (a)-(c) $P_m^i = P_m^g = 1$, $P_p^i = P_p^g = 0.005$, $P_d^i = 0.002$, $P_d^g = 0$; (d)-(f) $P_m^i = P_m^g = 1$, $P_p^i = P_p^g = 0.005$, $P_d^i = 0.008$, $P_d^g = 0.002$. For all simulations $\tau = 1$, $\Delta = 1$. (g)-(h) Comparisons between the averaged discrete model (black, solid) and the numerical solution of Equation (8.2) (cyan, dashed) at $t = 0$, $t = 500$ and $t = 1000$ for the parameters in (a)-(c) and (d)-(f), respectively. The grey lines indicate the initial condition and the arrow indicates the direction of increasing time. For all discrete solutions, $M = 1000$, $X = 600$, $\Delta = \tau = 1$. For all continuum solutions, $\delta x = 1$, $\delta t = 0.1$, $\epsilon = 10^{-6}$.

Twenty identically-prepared realisations of the discrete model are presented in Figures 8.1(a)-(f) for two different parameter regimes. In the first parameter regime, where $P_d^g = 0$, the initially-occupied region of the lattice remains fully occupied, as shown in Figures 8.1(a)-(c). When we introduce $P_d^g > 0$, as shown in Figures 8.1(d)-(f), the initially-occupied region of the lattice becomes partially vacant as time increases. We also compare the average discrete behaviour and the corresponding numerical solution of Equation (8.2) in Figures 8.1(g)-(h). This comparison shows that the solution of the continuum PDE matches the average discrete behaviour well, and predicts both the spread of the agent population in Figure 8.1(g) and the decrease in agent density in Figure 8.1(h).

The governing PDE, Equation (8.2), can be simplified in specific parameter regimes. While several of these simplified PDEs have been studied extensively, we summarise all non-trivial cases for completeness. It is instructive to consider each case and discuss the implications of the long term behaviour in terms of the discrete model parameters, as previous derivations of these PDEs have arisen from a variety of *ad hoc* arguments rather than working with a single unifying model. In Table 8.1 we summarise the salient features of 22 different special cases of Equation (8.2). The nonlinear diffusivity function, $F(C)$, has four key properties:

- $F(C)$ can either be a constant, or a function of the density of individuals;
- $F(C)$ can be degenerate, which implies that at one or more densities, C^*, we have $F(C^*) = R(C^*) = 0$;
- $F(C)$ can be zero at values of C^* that are non-degenerate, that is, $F(C^*) = 0,\ R(C^*) \neq 0$. In our model, this can occur at either zero, one or two different values of C; and
- $F(C)$ can be negative for an interval of C values.

The source term, $R(C)$, has two key properties:

- $R(C)$ can represent either Fisher kinetics (logistic growth) or Allee kinetics (bistable); and
- the grouped agent death rate, P_d^g, can be zero or non-zero. If the rate is non-zero, the carrying capacity density is reduced.

There are three different types of Allee kinetics considered in this work; weak, strong and reverse. We consider these three kinetics together for brevity, as it is relatively simple to change the parameter regime to alter the type of Allee effect without changing the competitive/co-operative mechanism described. The reverse Allee effect, which we describe here for the first time, refers to a growth rate that is reduced at high density, compared to logistic kinetics, but remains positive.

8.3.1 Fisher kinetics

The choice of whether the birth and death mechanisms imposed in the discrete model are neutral or are competitive/co-operative determines the form of the source term. If both $P_p^i = P_p^g$ and $P_d^i = P_d^g$, and hence both the birth and death mechanisms are neutral, the source term represents Fisher kinetics and Equation (8.2) simplifies to

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x}\left(F(C)\frac{\partial C}{\partial x}\right) + \lambda C(1-C) - KC, \tag{8.7}$$

where $\lambda = \lambda_i = \lambda_g$ and $K = K_i = K_g$. Transforming Equation (8.7) into travelling wave co-ordinates $z = x - vt$, where v is a constant wave speed and $-\infty < z < \infty$, results in

$$v\frac{\mathrm{d}C}{\mathrm{d}z} + F(C)\frac{\mathrm{d}^2C}{\mathrm{d}z^2} + (D_i - D_g)(6C-4)\left(\frac{\mathrm{d}C}{\mathrm{d}z}\right)^2 + \lambda C(1-C) - KC = 0,\ -\infty < z < \infty. \tag{8.8}$$

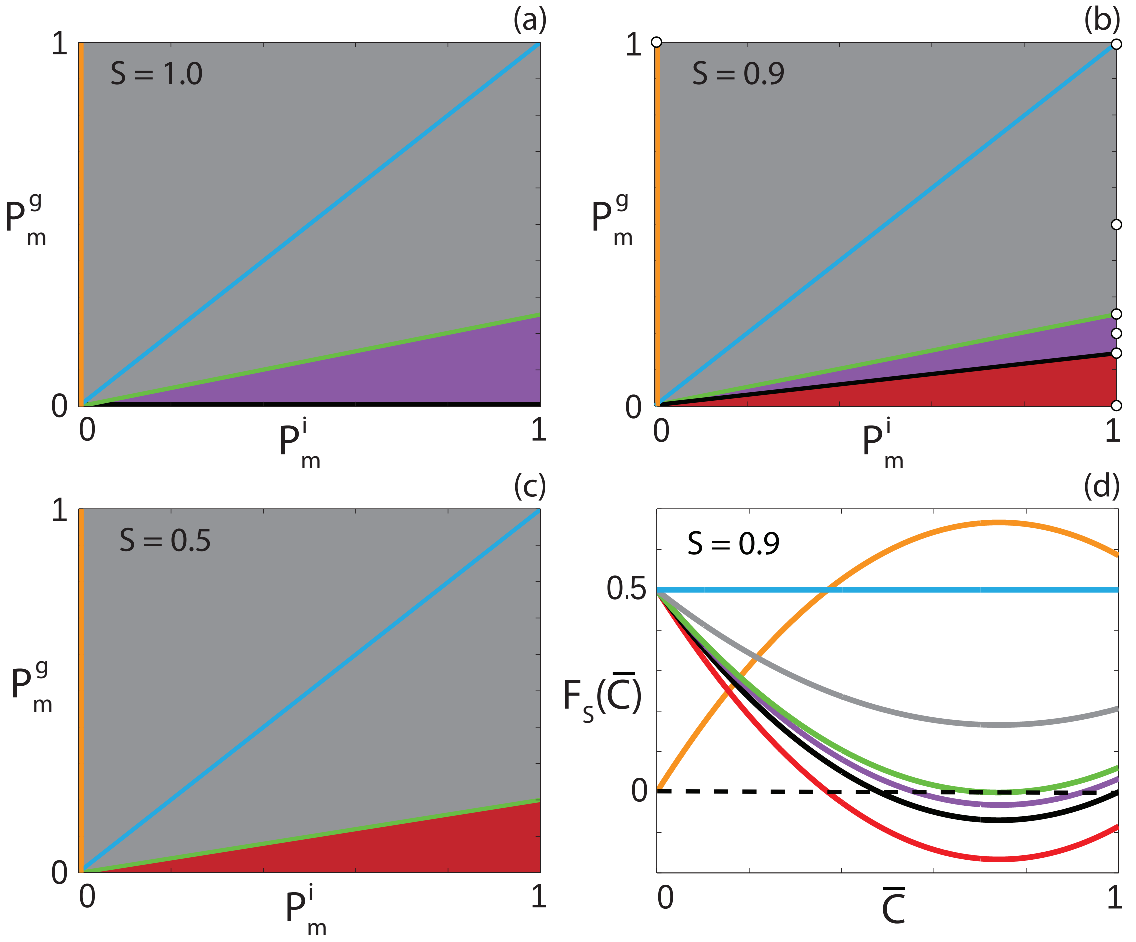

Figure 8.2: Classification of $F_s(\overline{C})$ for different carrying capacity densities. (a)-(c) Type of $F_s(\overline{C})$ function for $0 \leq \overline{C} \leq 1$ for the parameter space $P_m^i \in [0,1]$ and $P_m^g \in [0,1]$ with (a) $S = 1.0$; (b) $S = 0.9$; (c) $S = 0.5$. Grey regions correspond to parameter pairs that result in strictly positive $F_s(\overline{C})$, purple regions correspond to parameter pairs that result in positive-negative-positive $F_s(\overline{C})$ and red regions correspond to parameter pairs that result in positive-negative $F_s(\overline{C})$. Cyan, orange and black lines correspond to constant, extinction-degenerate non-negative and capacity-degenerate positive-negative $F_s(\overline{C})$ curves, respectively. (d) Example $F_s(\overline{C})$ for each region in (b). The white circles in (b) denote the parameter pairs used to generate the curves in (d).

Substituting $U = \mathrm{d}C/\mathrm{d}z$ allows Equation (8.8) to be expressed as a system of ordinary differential equations (ODEs)

$$\frac{\mathrm{d}C}{\mathrm{d}z} = U, \tag{8.9}$$

$$\frac{\mathrm{d}U}{\mathrm{d}z} = \frac{-vU - (D_i - D_g)(6C-4)U^2 - \lambda C(1-C) + KC}{F(C)}. \tag{8.10}$$

The equilibrium points of Equations (8.9)-(8.10) occur at $(C,U) = (0,0)$ and $(C,U) = (S,0)$, where $S = (\lambda - K)/\lambda$. The range of physically relevant C values correspond to $0 \leq C \leq S$. Hence the carrying capacity density, S, determines the numbers of times that $F(C) = 0$ for physically relevant C values. As such, we introduce a new variable $\overline{C} = C/S$ such that the agent density is scaled by the carrying capacity and the zeroes of $R(\overline{C})$ occur at $\overline{C} = 0$ and $\overline{C} = 1$.

Transforming Equation (8.7) in terms of $\overline{C}$, we obtain

$$\frac{\partial \overline{C}}{\partial t} = \frac{\partial}{\partial x}\left(F_s(\overline{C})\frac{\partial \overline{C}}{\partial x}\right) + (\lambda - K)\overline{C}(1-\overline{C}), \tag{8.11}$$

where $F_s(\overline{C}) = F(S\overline{C}) = D_i(1 - 4S\overline{C} + 3(S\overline{C})^2) + D_g(4S\overline{C} - 3(S\overline{C})^2)$. Equation (8.11) is the Fisher-Kolmogorov equation in terms of $\overline{C}$ with a nonlinear diffusivity function, $F_s(\overline{C})$. This new nonlinear diffusivity function has different properties depending on S, D_i and D_g. To highlight this, Figure 8.2 shows the (P_m^i, P_m^g) parameter space for three different choices of S and the qualitative behaviour of the corresponding $F_s(\overline{C})$ function. For all S values, parameter pairs that result in a constant $F_s(\overline{C})$ are highlighted in cyan. All parameter pairs that result in $F_s(\overline{C}) > 0$ for $0 \leq \overline{C} \leq 1$ are denoted by the grey regions. This type of diffusivity function is referred to as *strictly positive*. Similarly, for all S values, there are parameter pairs that result in $F_s(\overline{0}) = 0$, and $F_s(\overline{C}) > 0$ otherwise, which are highlighted in orange. We refer to this type of diffusivity function as *extinction-degenerate non-negative*.

For $S = 1$, presented in Figure 8.2(a), $P_m^i > 4P_m^g$, denoted in purple, results in an interval $\alpha < \overline{C} < \beta$, $\alpha < \beta < 1$, where $F_s(\overline{C}) < 0$. We refer to this type of nonlinear diffusivity function as *positive-negative-positive*. Decreasing S to 0.9, presented in Figure 8.2(b), we observe that the purple region again occurs for $P_m^i > 4P_m^g$. However, if $P_m^g < 0.145P_m^i$, highlighted in red, $F_s(\overline{C}) < 0$ for $\omega < \overline{C} \leq 1$, and hence $F_s(\overline{C})$ has only one zero in $0 \leq \overline{C} \leq 1$. This type of nonlinear diffusivity function is not observed with $S = 1$ and we refer to it as *positive-negative*. Specifically, this behaviour occurs when $(16 - (6S-4)^2)P_m^g < (4 - (6S-4)^2)P_m^i$ and $P_m^i > 4P_m^g$. Furthermore, this implies that for $S < 2/3$ there are no (P_m^i, P_m^g) values that correspond to positive-negative-positive $F_s(\overline{C})$. A choice of (P_m^i, P_m^g) that demonstrates this is shown in Figure 8.2(c). Unlike in Figures 8.2(a)-(b), we see that there is no purple region. Finally, if $(16 - (6S-4)^2)P_m^g = (4 - (6S-4)^2)P_m^i$, highlighted in black, $F_s(1) = 0$ and $F_s(\overline{C}) < 0$ for $\omega < \overline{C} < 1$, which we refer to as *capacity-degenerate positive-negative*. Note that for $S < 1/3$, $F_s(\overline{C}) \geq 0$ for $0 \leq \overline{C} \leq 1$. An example $F_s(\overline{C})$ curve for each type of diffusivity function is presented in Figure 8.2(d). PDE models that contain diffusivity functions with a region of negative diffusivity have been considered previously. However, these models either do not contain a source term or consider source terms that do not support travelling wave solutions [8, 40, 69]. Hence the model and analysis considered in this work is significantly different to the previous studies.

For all combinations of neutral, competitive and co-operative mechanisms that give rise to a reaction-diffusion equation with Fisher kinetics we examine the ability of the equation to give rise to long time travelling wave solutions. While details of the travelling wave solutions for certain types of diffusivity functions have been presented previously, we summarise the key features of the travelling wave solutions in tabular form for all cases for completeness in Table 8.2. For the cases where solution profiles have not been presented

Diffusivity function classification	**Travelling wave**	**Front type**	**Direction**
Constant	Yes	Smooth	Positive
Strictly positive	Yes	Smooth	Positive
Extinction-degenerate non-negative	Yes	Sharp	Positive
Positive-negative-positive	Yes	Smooth	Positive
Capacity-degenerate positive-negative	Yes	Smooth	Positive
Positive-negative	Yes	Smooth	Positive

Table 8.2: Classification of travelling wave solutions arising from different classes of PDEs with Fisher kinetics. Highlighted entries refer to cases analysed in detail in the manuscript; the other cases are analysed in the Supplementary Material.

previously, we provide more detailed discussion. A detailed analysis for each case is presented in the Supplementary Material.

Positive-negative-positive nonlinear diffusivity function

The first diffusivity function we examine in detail is the positive-negative-positive nonlinear diffusivity function, where $F_s(\overline{C}) < 0$ for an interval $\alpha < \overline{C} < \beta$. The simplest positive-negative-positive $F_s(\overline{C})$ occurs where $P_d^i = P_d^g = 0$ and hence $S = 1$. For these parameters, $F_s(\overline{C}) = F(C)$. Note that introducing non-zero P_d^i and P_d^g merely scales the governing equation and hence extending this analysis to cases with non-zero agent death is straightforward, provided that $F_s(\overline{C})$ has two zeroes on the interval $0 < \overline{C} < 1$. Parameters that result in a positive-negative-positive $F_s(\overline{C})$ are highlighted in purple in Figure 8.2 and, for this case, with $P_d^i = P_d^g = 0$, occur when $P_m^i > 4P_m^g$. For positive-negative-positive $F(C)$, Equation (8.10) is singular at $C = \alpha$ and $C = \beta$, where the interval of $F(C) < 0$ is given by

$$\alpha = \frac{2}{3} - \frac{\sqrt{(P_m^i)^2 - 5P_m^i P_m^g + 4(P_m^g)^2}}{3(P_m^i - P_m^g)} < C < \beta = \frac{2}{3} + \frac{\sqrt{(P_m^i)^2 - 5P_m^i P_m^g + 4(P_m^g)^2}}{3(P_m^i - P_m^g)}. \tag{8.12}$$

The singularities at $C = \alpha$ and $C = \beta$ cannot be removed using a stretching transformation (Supplementary Material) since $R(\alpha) \neq 0$ and $R(\beta) \neq 0$. However, it is possible for $\mathrm{d}U/\mathrm{d}z$ to be finite at $C = \alpha$ and $C = \beta$ if U_α and U_β exist such that

$$\lim_{C \to \alpha} \left[\frac{-vU_\alpha - (D_i - D_g)(6C - 4)U_\alpha^2 - \lambda C(1-C)}{D_i(1 - 4C + 3C^2) + D_g(4C - 3C^2)} \right], \tag{8.13}$$

$$\lim_{C \to \beta} \left[\frac{-vU_\beta - (D_i - D_g)(6C - 4)U_\beta^2 - \lambda C(1-C)}{D_i(1 - 4C + 3C^2) + D_g(4C - 3C^2)} \right], \tag{8.14}$$

are both finite. This requires the numerator in the expressions (8.13)-(8.14) vanish at $C = \alpha$ and $C = \beta$, respectively. As such, U_α and U_β are obtained by solving the system

$$0 = -vU_\alpha - (D_i - D_g)(6\alpha - 4)U_\alpha^2 - \lambda\alpha(1-\alpha), \tag{8.15}$$

$$0 = -vU_\beta - (D_i - D_g)(6\beta - 4)U_\beta^2 - \lambda\beta(1-\beta), \tag{8.16}$$

resulting in $U_\alpha = -(v \pm \sqrt{v^2 - 4F'(\alpha)R(\alpha)})/2F'(\alpha)$ and $U_\beta = -(v \pm \sqrt{v^2 - 4F'(\beta)R(\beta)})/2F'(\beta)$. We note that as $R(C) \geq 0$ for $0 \leq C \leq 1$, and that $F'(\alpha) \leq 0$ for all possible α values, U_α will be real-valued. Subsequently, we have a wave speed condition that $v \geq 2\sqrt{F'(\beta)R(\beta)}$, as $F'(\beta) \geq 0$ for all possible β values. Ferracuti *et al.* [14] prove that the minimum wave speed, v^*, is greater than a threshold value, which in turn is greater than $\max\{R'(0)F(0), F'(\beta)R(\beta)\}$. Therefore, U_β will also always be real-valued.

Applying L'Hopital's Rule to Equation (8.10), we obtain

$$\lim_{C\to\alpha} \left.\frac{\mathrm{d}U}{\mathrm{d}z}\right|_{U=U_\alpha} = \lim_{C\to\alpha}\left[\frac{6(D_i - D_g)U_\alpha^2 + \lambda(1-2C)}{(D_g - D_i)(6C-4)}\right], \tag{8.17}$$

$$\lim_{C\to\beta} \left.\frac{\mathrm{d}U}{\mathrm{d}z}\right|_{U=U_\beta} = \lim_{C\to\beta}\left[\frac{6(D_i - D_g)U_\beta^2 + \lambda(1-2C)}{(D_g - D_i)(6C-4)}\right], \tag{8.18}$$

which are finite provided that $\alpha \neq 2/3$ and $\beta \neq 2/3$. For the system of Equations (8.9)-(8.10), we have two straight lines in the phase plane where $\mathrm{d}U/\mathrm{d}z$ is infinite, at $C = \alpha$ and $C = \beta$. These kind of lines have previously been called *walls of singularities* for hyperbolic models related to chemotactic and haptotactic invasion [51]. For a smooth solution trajectory joining the two equilibrium points on opposite sides of the wall of singularities, we require that the trajectory passes through the wall of singularities. This implies that the solution trajectory must pass through the wall of singularities at the special points, (α, U_α) and (β, U_β), known as *holes in the wall* [51, 71]. Otherwise, a smooth heteroclinic orbit between $(1, 0)$ and $(0, 0)$ cannot exist, as $\lim_{C\to\alpha}|U| \to \infty$ and $\lim_{C\to\beta}|U| \to \infty$. As U_α and U_β are real valued and the limits in Equations (11)-(12) are finite, the holes in the wall always exist for Fisher kinetics.

We superimpose the numerical solution of Equation (8.7) in (C, U) co-ordinates on the phase plane for the system (8.9)-(8.10) in Figures 8.3(a) and 8.3(d). Details of the numerical techniques used to solve Equation (8.7) and to generate the phase planes are given in the Methods. The numerical solution appears to form a heteroclinic orbit between $(1, 0)$ and $(0, 0)$ in both cases, and passes through the holes in the wall of singularities, denoted using purple circles. Continuum models with negative diffusivity and no source terms have been relatively well studied, and exhibit shock behaviour across the region of negative diffusion [39, 72]. Interestingly, our solution does not include a shock and is instead smooth through the region of negative diffusion.

The numerical solution of Equation (8.7) at $t = 100$ and $t = 200$ is shown in Figures 8.3(b) and 8.3(e), confirming that the waveform does not change with time. To quantify the wave speed we calculate the time evolution of the leading edge, $L(t) = x_f$ such that $C(x_f, t) \approx 1 \times 10^{-4}$. If the solution of Equation (8.7) forms a travelling wave, $L(t)$ will tend to a straight line with slope v, as $t \to \infty$. In Figures 8.3(c) and 8.3(f), we observe that $L(t)$ is approximately linear with slope v, and hence the solution of Equation (8.7) moves with approximately constant speed at long times. Overall, these features suggest

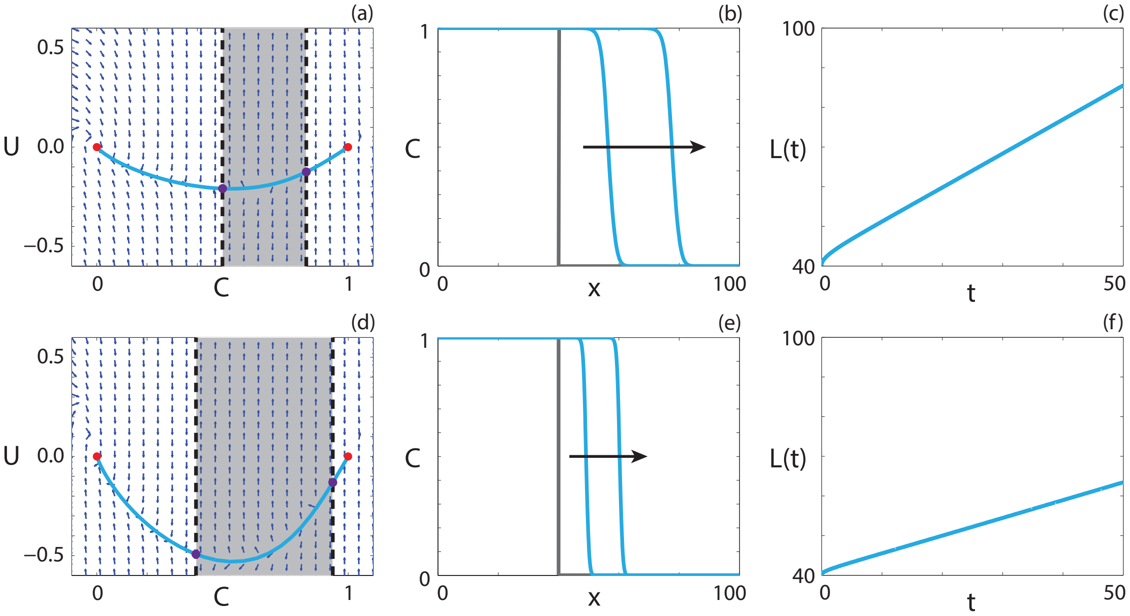

Figure 8.3: Travelling wave behaviour for Equation (8.7) **with positive-negative-positive** $F(C)$ **(Case 2.3)**. (a), (d) Phase plane for the system (8.9)-(8.10) with the numerical solution of Equation (8.7), in (C, U) co-ordinates, superimposed. The grey region corresponds to values of C where $F(C) < 0$. The dashed black lines denote a wall of singularities. Red circles correspond to equilibrium points and purple circles correspond to holes in the wall. (b), (e) Numerical solution of Equation (8.7) at $t = 100$ and $t = 200$. The grey lines indicate the initial condition and the arrows indicate the direction of increasing time. (c), (f) The time evolution of the position of the leading edge of the travelling wave solution, $L(t)$. All results are obtained using $P_d^i = P_d^g = 0$, $\delta x = 0.01$, $\delta t = 0.01$, $\epsilon = 10^{-6}$ and (a)-(c) $P_m^i = 0.5$, $P_m^g = 0.1$, $P_p^i = P_p^g = 0.75$, $v = 0.864$; (d)-(f) $P_m^i = 0.1$, $P_m^g = 0.01$, $P_p^i = P_p^g = 1.0$, $v = 0.448$.

that the solution of Equation (8.7) with positive-negative-positive $F(C)$ approaches a travelling wave.

Capacity-degenerate positive-negative nonlinear diffusivity function

For the special case where $P_m^g = 0$, $F_s(1) = 0$. Again, we consider the case with zero agent death for simplicity, and note that it is straightforward to extend the analysis for cases with non-zero agent death. As $F(C)$ is degenerate at $C = 1$, it is intuitive to expect there could be sharp-fronted travelling wave solutions, with the sharp front near $C = 1$, similar to the results in [62] and in the Supplementary Material. However, unlike these cases, we have an interval $1/3 < C < 1$ where $F(C) < 0$. To determine whether this negative diffusivity influences the presence of sharp fronts, we follow the approach of Maini *et al.* [42], who show that the existence of travelling waves for reaction-diffusion equations with capacity-degenerate positive-negative $F(C)$ can be determined by considering the existence of travelling waves for

$$\frac{\partial C}{\partial \hat{t}} = \frac{\partial^2 C}{\partial x^2} + F(C)R(C), \ \hat{t} \geq 0. \tag{8.19}$$

The restriction on $\hat{t}$ implies $F(C) > 0$. As $F(C) < 0$ for $1/3 < C < 1$, Equation (8.19) is only equivalent to Equation (8.7) for $0 \leq C \leq 1/3$. For $1/3 \leq C \leq 1$, Equation (8.7) is

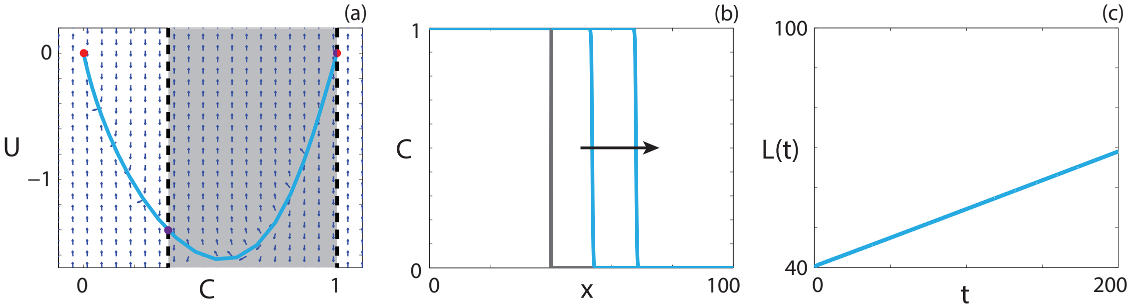

Figure 8.4: Travelling wave behaviour for Equation (8.7) **with capacity-degenerate positive-negative** $F(C)$ **(Case 2.4)**. (a) Phase plane for the system (8.9)-(8.10) with the numerical solution of Equation (8.7), in (C, U) co-ordinates, superimposed. The grey region corresponds to values of C where $F(C) < 0$. The dashed black lines denote two walls of singularities. Red circles correspond to equilibrium points and purple circles correspond to holes in the wall. (b) Numerical solution of Equation (8.7) at $t = 100$ and $t = 200$. The grey lines indicate the initial condition and the arrow indicates the direction of increasing time. (c) The time evolution of the position of the leading edge of the travelling wave solution, $L(t)$. All results are obtained using $P_m^i = 0.01$, $P_m^g = 0$, $P_p^i = P_p^g = 1.0$, $P_d^i = P_d^g = 0$, $\delta x = 0.01$, $\delta t = 0.01$, $\epsilon = 10^{-6}$, $v = 0.1433$.

equivalent to

$$\frac{\partial C}{\partial \hat{t}} = \frac{\partial^2 C}{\partial x^2} + \hat{F}(C)\hat{R}(C), \ \hat{t} \geq 0, \tag{8.20}$$

where $\hat{F}(C) = -F(1-C)$ and $\hat{R}(C) = R(1-C)$ [42]. Equations (8.19)-(8.20) have minimum travelling wave speeds v_0^* and v_1^*, respectively. Maini *et al.* [42] prove that sharp fronts in the travelling wave near $C = 1$ only exist if $F(1) = 0$ and $v_1^* < v_0^*$. The first condition is obviously satisfied, while the second can be determined by considering the behaviour of the equivalent ordinary differential systems in travelling wave coordinates in the neighbourhood of the equilibrium points. Both equations have minimum wave speed conditions, $v_0^* = v_1^* = 2\sqrt{\lambda D_i}$, to obtain physically-relevant heteroclinic orbits, and hence travelling wave solutions with a sharp region near $C = 1$ do not exist.

Travelling wave behaviour for a parameter regime with $F(1) = 0$ is shown in Figure 8.4. The equilibrium point at $(1, 0)$ is also a hole in the wall. The solution trajectory forms a heteroclinic orbit between $(1, 0)$ and $(0, 0)$, and passes through the region of C where $F(C) < 0$. Although $F(1) = 0$, we do not observe a solution trajectory corresponding to a sharp front, as for capacity-degenerate non-negative $F(C)$ (Supplementary Material). This result is consistent with the analysis of Maini *et al.* [42]. The numerical solution of Equation (8.4), presented in Figure 8.4(b), has a relatively steep front but is not sharp near $C = 1$. As $L(t)$, presented in Figure 8.4(c), becomes linear as t increases and the waveform in Figure 8.4(b) are consistent, the numerical solution of Equation (8.7) with $F(1) = 0$ appears to form a classic travelling wave.

Positive-negative nonlinear diffusivity function

The positive-negative case, where $F_s(\overline{C}) > 0$ for $0 \leq \overline{C} < \omega$ and $F_s(\overline{C}) < 0$ for $\omega < \overline{C} \leq 1$, cannot occur with $K = 0$. It is instructive to examine whether sta-

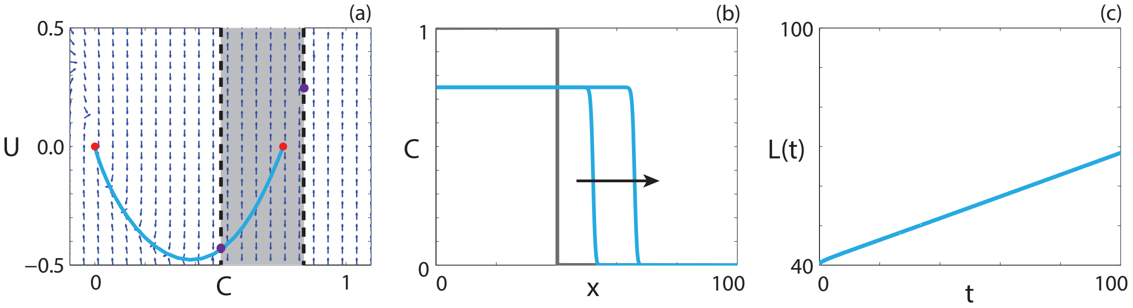

Figure 8.5: Travelling wave behaviour for Equation (8.7) **with positive-negative** $F_s(\overline{C})$ **(Case 4.5)**. (a) Phase plane for the system (8.9)-(8.10) with the numerical solution of Equation (8.7), in (C, U) co-ordinates, superimposed. The grey region corresponds to values of $\overline{C}$ where $F_s(\overline{C}) < 0$. The dashed black lines denote a wall of singularities. Red circles correspond to equilibrium points and purple circles correspond to holes in the wall. (b) Numerical solution of Equation (8.7) at $t = 50$ and $t = 100$. The grey lines indicate the initial condition and the arrow indicates the direction of increasing time. (c) The time evolution of the position of the leading edge of the travelling wave solution. All results are obtained using $P_m^i = 0.05$, $P_m^g = 0.01$, $P_p^i = P_p^g = 1.0$, $P_d^i = P_d^g = 0.25$, $\delta x = 0.1$, $\delta t = 0.01$, $\epsilon = 10^{-6}$, $v = 0.2760$.

ble travelling wave solutions of Equation (8.7) exist in such a case, as the non-zero equilibrium point now occurs in the region where $F_s(\overline{C}) < 0$. If we perform standard linear analysis on Equations (8.9)-(8.10), the Jacobian at $(S, 0)$ has eigenvalues $\xi = (-v \pm \sqrt{v^2 + 4F(S)(\lambda(2S-1)+K)})/2F(S)$, which implies that the equilibrium point is an unstable node provided $v > 2\sqrt{-F(S)(\lambda(2S-1)+K)}$. The negative sign is present as $F(S) < 0$ for positive-negative $F_s(\overline{C})$. The Jacobian at $(0, 0)$ has eigenvalues $\xi = (-v \pm \sqrt{v^2 - 4D_i(\lambda - K)})/2D_i$, which is a stable node provided that $v > 2\sqrt{(\lambda - K)D_i}$. While there are infinitely many solution trajectories out of the unstable node, we require that the solution trajectory passes through the hole in the wall, and hence there is a single solution trajectory that forms a heteroclinic orbit.

Travelling wave behaviour for Equation (8.7) with positive-negative $F_s(\overline{C})$ is shown in Figure 8.5. The numerical solution of Equation (8.7), in (C, U) co-ordinates, passes through the wall of singularities where Equation (8.10) is finite and forms a heteroclinic orbit between $(S, 0)$ and $(0, 0)$. The travelling wave front is of classic type, a result predicted by the analysis performed by Maini *et al.* [42] as $F_s(0) \neq 0$ and $F_s(1) \neq 0$.

8.3.2 Allee kinetics

If the birth and death mechanisms are either competitive or co-operative, that is, $P_p^i \neq P_p^g$ and $P_d^i \neq P_d^g$, then the source term represents an Allee effect [67] and hence Equation (8.2) can be expressed as

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x}\left(F(C)\frac{\partial C}{\partial x}\right) + (K_i - K_g - \lambda_i + \lambda_g)A_1 C\left(1 - \frac{C}{A_1}\right)\left(C - A_2\right), \qquad (8.21)$$

where

$$
\begin{aligned}
A_1 &= \frac{2\lambda_i - \lambda - 2K_i + 2K_g - \sqrt{\lambda_g^2 + 4K_g(\lambda_i - \lambda_g - K_i + K_g)}}{2(\lambda_i - \lambda_g - K_i + K_g)}, \\
A_2 &= \frac{2\lambda_i - \lambda - 2K_i + 2K_g + \sqrt{\lambda_g^2 + 4K_g(\lambda_i - \lambda_g - K_i + K_g)}}{2(\lambda_i - \lambda_g - K_i + K_g)}.
\end{aligned}
\tag{8.22}
$$

Note that either $\lambda_g \geq 2(K_g + \sqrt{Kg(K_i - \lambda_i)})$ or $\lambda_i > K_i$ must be satisfied or $R(C) \leq 0$ for $0 \leq C \leq 1$ and the population will tend to extinction. In travelling wave co-ordinates, Equation (8.21) is

$$
v\frac{\mathrm{d}C}{\mathrm{d}z} + F(C)\frac{\mathrm{d}^2C}{\mathrm{d}z^2} + F'(C)\left(\frac{\mathrm{d}C}{\mathrm{d}z}\right)^2 + (K_i - K_g - \lambda_i + \lambda_g)A_1C\left(1 - \frac{C}{A_1}\right)\left(C - A_2\right) = 0, \quad -\infty < z < \infty, \tag{8.23}
$$

and, making the substitution $U = \mathrm{d}C/\mathrm{d}z$, it corresponds to

$$
\frac{\mathrm{d}C}{\mathrm{d}z} = U, \tag{8.24}
$$

$$
\frac{\mathrm{d}U}{\mathrm{d}z} = -\frac{vU}{F(C)} - \frac{(D_i - D_g)(6C - 4)U^2}{F(C)} - \frac{(K_i - K_g - \lambda_i + \lambda_g)A_1C}{F(C)}\left(1 - \frac{C}{A_1}\right)\left(C - A_2\right). \tag{8.25}
$$

If $P_d^g = 0$, then $A_1 = 1$, and the source term in (8.21) simplifies to $R(C) = rC(1-C)(C-A)$, where $r = K_i - \lambda_i + \lambda_g$ is the intrinsic growth rate and $A = (K_i - \lambda_i)/(K_i - \lambda_i + \lambda_g)$ is the Allee parameter [67].

A new variable $\overline{C} = C/A_1$ is introduced such that the range of physically relevant $\overline{C}$ values corresponds to $0 \leq \overline{C} \leq 1$. Substituting $\overline{C}$ into Equation (8.21) results in

$$
\frac{\partial \overline{C}}{\partial t} = \frac{\partial}{\partial x}\left(F_A(\overline{C})\frac{\partial \overline{C}}{\partial x}\right) + (K_i - K_g - \lambda_i + \lambda_g)A_1^2\overline{C}(1 - \overline{C})(\overline{C} - \overline{A}), \tag{8.26}
$$

where $F_A(\overline{C}) = F(A_1\overline{C}) = D_i(1 - 4A_1\overline{C}^2 + 3A_1^2\overline{C}^2) + D_g(4A_1\overline{C} - 3A_1^2\overline{C}^2)$ and

$$
\overline{A} = \frac{A_2}{A_1} = \frac{2\lambda_i - \lambda - 2K_i + 2K_g + \sqrt{\lambda_g^2 + 4K_g(\lambda_i - \lambda_g - K_i + K_g)}}{2\lambda_i - \lambda - 2K_i + 2K_g - \sqrt{\lambda_g^2 + 4K_g(\lambda_i - \lambda_g - K_i + K_g)}}. \tag{8.27}
$$

The transformed nonlinear diffusivity, $F_A(\overline{C})$, has the same characteristics as $F_s(\overline{C})$, presented in Figure 8.2, albeit in terms of the scaled Allee carrying capacity, A_1. For $P_m^i = P_m^g$, $F_A(\overline{C})$ represents linear diffusion. Reaction-diffusion equations with linear diffusion and either weak or strong Allee kinetics have been well-studied [4, 15, 16, 26, 34, 41, 54, 67]. For additional details we refer the reader to [67]. Weak Allee kinetics correspond to $(K_i - K_g - \lambda_i + \lambda_g) > 0$ and $\overline{A} < 0$, and represent a growth rate that is inhibited at

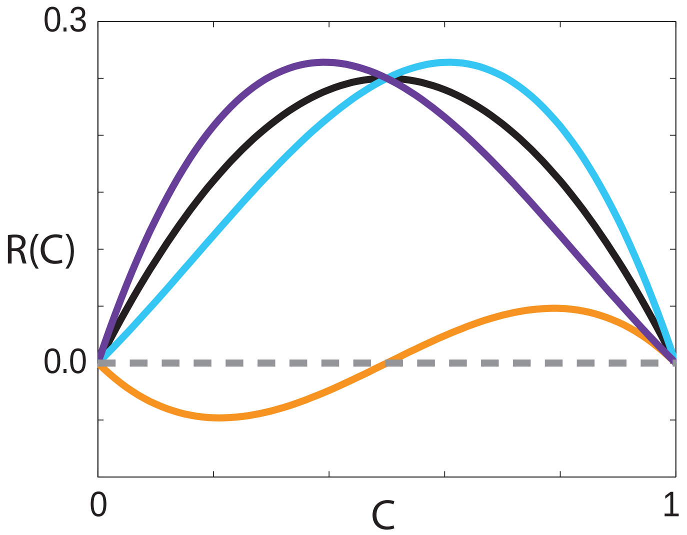

Figure 8.6: Comparison of source terms. $R(C)$ corresponding to the weak Allee effect with $r = 1$, $A = -0.5$ (cyan), strong Allee effect with $r = 1$, $A = 0.5$ (orange), reverse Allee effect with $r = -1$, $A = 1.5$ (purple) and logistic growth with $r = 1$ (black).

low densities, compared to logistic growth, but remains positive. Strong Allee kinetics correspond to $(K_i - K_g - \lambda_i + \lambda_g) > 0$ and $0 < \overline{A} < 1$ [67], and represent a growth rate that is negative beneath a threshold value, and positive otherwise. Interestingly, a third type of Allee kinetics can arise from the parameter values chosen in the discrete model, that has not been considered previously. If $(K_i - K_g - \lambda_i + \lambda_g) < 0$ and $\overline{A} > 1$, the growth rate is non-negative for all relevant $\overline{C}$ values but is inhibited at high densities, compared to logistic growth, rather than low densities like the weak Allee effect. We term this type of growth term the reverse Allee effect. Representative source terms showing the three types of Allee effect are compared with a logistic source term in Figure 8.6.

For all combinations of neutral, competitive and co-operative mechanisms that give rise to a reaction-diffusion equation with Allee kinetics we examine the ability of the equation to give rise to long time travelling wave solutions. Furthermore, the three types of Allee effect arising from the discrete model are considered. Several of these cases have been presented and examined previously, but we present details about the travelling wave solutions for all combinations of diffusivity functions and Allee effects in Table 8.3. A detailed analysis for each case is presented in the Supplementary Material.

Persistence and extinction

A key question of interest for a particular class of PDE is whether the population described persists or becomes extinct in the long time limit. In all cases with Fisher kinetics with $\lambda > K$, the source term is positive for $0 \leq C \leq 1$, and subsequently the population persists and spreads. As the kinetics representing an Allee effect can contain a source term that is negative for an interval of C, it is less obvious whether the minimum wave speed is positive or negative, corresponding to persistence or extinction, respectively.

Diffusivity function classification	**Allee effect**	**Travelling wave**	**Front type**	**Direction**
Constant	Weak/Reverse	Yes	Smooth	Positive
Constant	Strong	Yes	Smooth	Positive
Constant	Strong	Yes	Smooth	Negative
Strictly positive	Weak/Reverse	Yes	Smooth	Positive
Strictly positive	Strong	Yes	Smooth	Positive
Strictly positive	Strong	Yes	Smooth	Negative
Extinction-degenerate non-negative	Weak/Reverse	Yes	Sharp	Positive
Extinction-degenerate non-negative	Strong	Yes	Sharp	Positive
Extinction-degenerate non-negative	Strong	Yes	Smooth	Negative
Positive-negative-positive	Weak/Reverse	Yes	Smooth	Positive
Positive-negative-positive	Strong	Yes	Shock	Positive
Positive-negative-positive	Strong	Yes	Shock	Negative
Capacity-degenerate positive-negative	Weak/Reverse	Yes	Smooth	Positive
Capacity-degenerate positive-negative	Strong	No	N/A	N/A
Positive-negative	Weak/Reverse	Yes	Smooth	Positive
Positive-negative	Strong	No	N/A	N/A

Table 8.3: Classification of travelling wave solutions arising from different classes of PDEs with Allee kinetics. Highlighted entries refer to cases analysed in detail in the manuscript; the other cases are analysed in the Supplementary Material.

For the case with constant $F(C)$ and $P_d^g = 0$, the minimum wave speed for Equation (8.21) with $A < -1/2$ is $v^* = 2\sqrt{(\lambda_i - K_i)D}$ and hence the population persists, provided $\lambda_i > K_i$. Introducing $P_d^g \neq 0$ results in the same minimum wave speed, provided that $\overline{A} < -1/2$. This implies that introducing grouped agent death at a rate that does not result in a population tending to extinction has no influence on the invasion speed of the population. Specifically, the condition for $A < -1/2$ with $K_g = 0$ corresponds to $3(\lambda_i - K_i) > \lambda_g$. It can be shown that, with $3(\lambda_i - K_i) > \lambda_g$, we require $3K_g < \lambda_g$ for $\overline{A} < -1/2$. This implies that there is a range of K_g values that result in a travelling wave with a minimum wave speed that is independent of both K_g and λ_g. Interestingly, this suggests that if a control is implemented that increases the death rate of grouped agents, there is a threshold value of P_d^g for the control to influence the invasion speed and the subsequent persistence of the population. Introducing a non-zero K_g value for a parameter regime that results in the strong Allee effect with $K_g = 0$ never changes the type of Allee effect. It is possible to go from a weak Allee effect to a reverse Allee effect by introducing a non-zero K_g value. Non-zero K_g values correspond to a decreased benefit for grouped agents, which explains why the source term, previously a weak Allee effect, becomes the reverse Allee effect, corresponding to inhibited growth at high density.

The reaction-diffusion equation with constant $F_A(\overline{C})$ and the strong Allee effect, corresponding to $0 < A_2 < A_1 \leq 1$, has a unique wave speed $v = 2\sqrt{(K_i - K_g - \lambda_i + \lambda_g)D}\,(A_1/2 - A_2)$ [41]. This implies that for $A_2 > A_1/2$, $v < 0$ and $v > 0$ otherwise. Furthermore, the same wave speed applies for $-A_1/2 < A_2 < 0$ [41]. For both intervals, the minimum wave speed depends on the K_g value, and hence implementing any kind of partial eradication of the grouped agents will either reduce the speed of invasion or cause the extinction of the population.

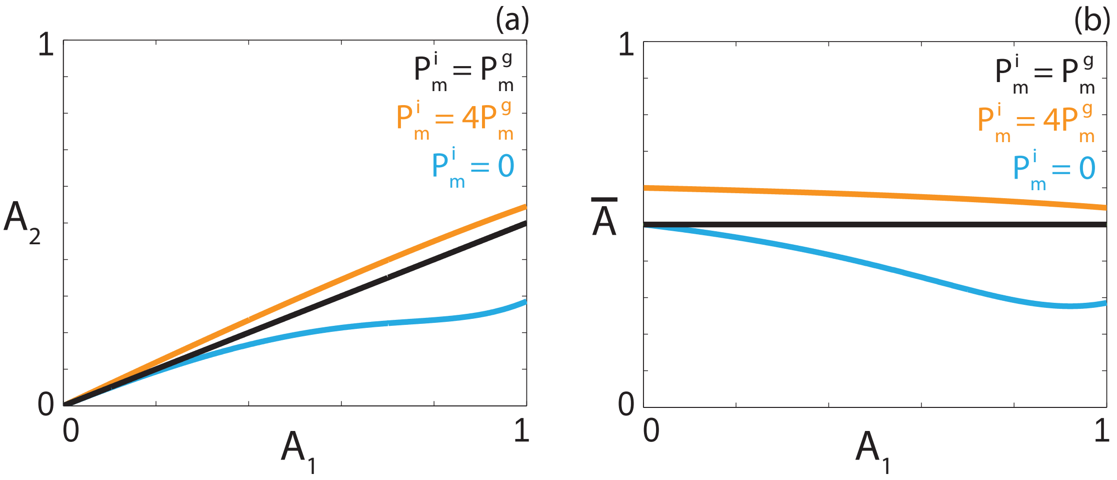

Figure 8.7: Persistence threshold. Persistence threshold as a function of the carrying capacity A_1, expressed as (a) an explicit value; (b) a proportion of the carrying capacity for three different diffusivities, corresponding to $P^i_m = P^g_m$ (black), $P^i_m = 4P^g_m$ (orange) and $P^i_m = 0$ (cyan).

For cases where $F_A(\overline{C}) \geq 0$ for $0 \leq C \leq 1$ and $F_A(\overline{C})$ is not constant, we follow the approach of Hadeler to establish whether the minimum wave speed is positive, and hence the population persists [23–25]. The integral condition for the wave speed to be positive,

$$\int_0^1 (K_i - K_g - \lambda_i + \lambda_g) A_1^2 (D_i(1 - 4A_1\overline{C}^2 + 3A_1^2\overline{C}^2) + D_g(4A_1\overline{C} - 3A_1^2\overline{C}^2))\overline{C}(1-\overline{C})(\overline{C}-\overline{A})\,\mathrm{d}\overline{C} > 0, \tag{8.28}$$

corresponds to

$$D_i(5 - 10\overline{A} + 6A_1^2 - 9A_1A_2 - 12A_1 + 20A_2) - D_g(6A_1^2 - 9A_1A_2 - 12A_1 + 20A_2) > 0. \tag{8.29}$$

If $D_i = D_g$, then $\overline{A} > 1/2$ leads to $v < 0$. For the strong Allee effect, $A_1 > A_2 = \overline{A}A_1$, we can determine the threshold value for the persistence of the population, namely,

$$\overline{A} < \frac{5D_i + (D_i - D_g)(6A_1^2 - 12A_1)}{10D_i + (D_i - D_g)(9A_1^2 - 20A_1)}. \tag{8.30}$$

Considering the two limiting cases for strictly positive $F_A(\overline{C})$, where $D_i = 0$ and $D_i = 4D_g$, $\overline{A}$ takes on a value of $(6A_1^2 - 12A_1)/(9A_1^2 - 20A_1)$ and $(18A_1^2 - 36A_1 + 20)/(27A_1^2 - 60A_1 + 30)$, respectively. These values reduce to 6/11 and 2/7 in the case that $A_1 = 1$, corresponding to $K_g = 0$. Therefore, populations with isolated agents that are more motile than grouped agents are less susceptible to extinction. To illustrate how the threshold value changes with A_1, P^i_m and P^g_m, Figure 8.7 shows the maximum A_2 and $\overline{A}$ values for three different P^i_m and P^g_m combinations. The A_2 value corresponds to the persistence threshold for a given A_1 value. The $\overline{A}$ value can be interpreted as the highest proportion of a given A_1 value that will result in the persistence of the population. For example, in Figure 8.7(a), we see that with $P^i_m = 0$ and $A_1 = 0.5$ we require $A_2 < 0.194$ for persistence. This corresponds to $\overline{A} < 0.388$.

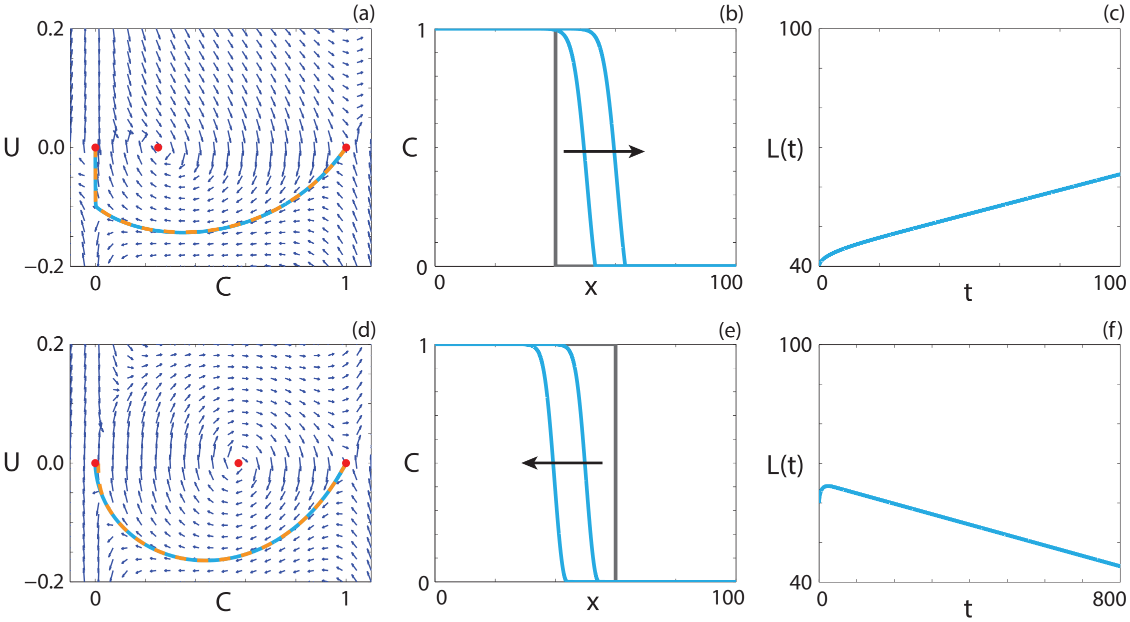

Figure 8.8: Travelling wave behaviour for Equation (8.21) **with the strong Allee effect and extinction-degenerate non-negative** $F(C)$ **(Case 6.2)**. (a), (d) Phase plane for the system (8.24)-(8.25) with the numerical solution of Equations (8.21) (cyan, solid) and (8.23) (orange, dashed), in (C, U) co-ordinates, superimposed. Red circles correspond to equilibrium points. (b), (e) Numerical solution of Equation (8.21) calculated at (b) $t = 50$ and $t = 100$; (e) $t = 400$ and $t = 800$. The grey lines indicate the initial condition and the arrows indicate the direction of increasing time. (c), (f) The time evolution of $L(t)$. All results are obtained with $\delta x = 0.01$, $\delta t = 0.005$, $\epsilon = 10^{-6}$, $P_m^i = 0$, $P_m^g = 1.0$,$P_d^g = 0$, (a)-(c) $P_p^i = 0.4$, $P_p^g = 0.3$, $P_d^i = 0.5$, $v = 0.199$; (d)-(f) $P_p^i = 0.4$, $P_p^g = 0.3$, $P_d^i = 0.8$, $v = -0.026$.

Extinction-degenerate non-negative nonlinear diffusivity function

Travelling wave behaviour for the strong Allee effect with extinction-degenerate non-negative $F(C)$ is shown in Figure 8.8. The numerical solution of Equation (8.21) with $A = 1/4$, in Figures 8.8(a)-(c), leads to a sharp-fronted travelling wave solution near $C = 0$ with $v > 0$. With $A = 1/4$, we expect to obtain $v > 0$. For a parameter regime that results in $A = 4/7$, we obtain a travelling wave solution of Equation (8.21) with $v < 0$ (Figures 8.8(d)-(f)). Interestingly, the sharp front near $C = 0$ is not present for the strong Allee effect with $v < 0$.

Positive-negative-positive nonlinear diffusivity function

A positive-negative-positive $F(C)$, where there is an interval $\alpha < C < \beta$ where $F(C) < 0$, corresponds to parameter pairs highlighted in purple in Figure 8.2(a). Kuzmin and Ruggerini [38] examine reaction-diffusion equations with similar properties for the strong Allee effect, in the context of diffusion-aggregation models, and provide conditions for smooth travelling wave solutions to exist. For a solution with $v > 0$, we require $A < \alpha$ [38] and

$$\int_0^{\alpha} F(C)R(C) \; \mathrm{d}C > 0. \tag{8.31}$$

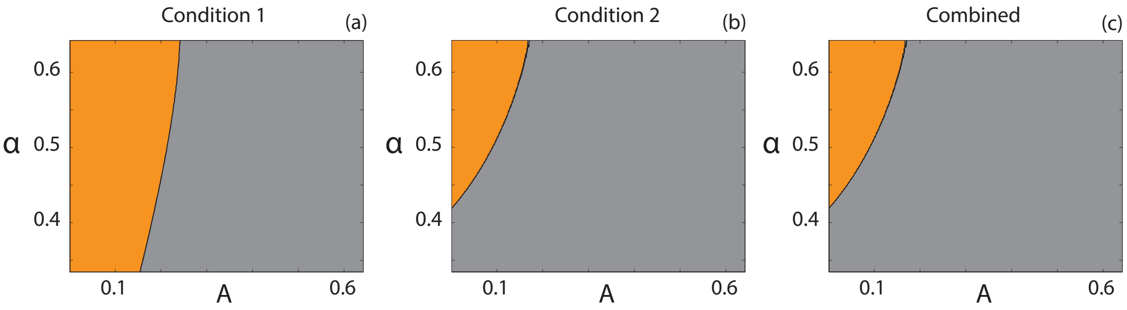

Figure 8.9: Parameter pairs that satisfy Kuzmin and Ruggerini's Conditions. (a) Condition (8.31); (b) Condition (8.32); (c) Conditions (8.31)-(8.32) combined. Orange regions correspond to parameter pairs that satisfy the respective condition(s), whereas grey regions correspond to parameter pairs that do not.

Furthermore, we require [38]

$$3\int_0^{\alpha} F(C)R(C)\ \mathrm{d}C \geq \max\{\Phi(\sigma), \Phi(\rho)\}, \tag{8.32}$$

where

$$\begin{aligned}
&\Phi(y) = 8\alpha^2 y + 4\sqrt{4\alpha^2y^2 - 2m\alpha^3 y},\\
&\sigma = \sup_{C\in[\alpha,\beta)}\left[\frac{F(C)R(C)}{C-\beta}\right],\ \rho = \sup_{C\in(\beta,1]}\left[\frac{F(C)R(C)}{C-\beta}\right],\ \text{and}\\
&m = \min_{C\in[0,A]}\left[F(C)R(C)\right].
\end{aligned}$$

A suite of P_m^g values with $P_m^i = 1$ that correspond to $1/3 < \alpha < 2/3$ are considered for parameter regimes that result in $A < \alpha$. Figures 8.9(a)-(c) show the parameter pairs, (A, α), that satisfy Condition (8.31), Condition (8.32) and Conditions (8.31)-(8.32) simultaneously, respectively. Orange regions represent parameter pairs where the condition is satisfied and grey regions represent parameter pairs where the condition is not satisfied. These results suggest that smooth travelling wave solutions should exist for certain choices of parameters. Interestingly, all parameter pairs that satisfy Condition (8.31) also satisfy Condition (8.32).

For Equation (8.7) with positive-negative-positive $F(C)$, smooth travelling wave solutions that pass through holes in the wall of singularities for positive-negative-positive $F(C)$ are obtained. The minimum wave speed bound presented by Ferracuti *et al.* [14] implies that the locations of the holes in the wall occur are real-valued for the wave speed arising from the Heaviside initial condition. As such, to obtain smooth travelling wave solutions of Equation (8.21) with positive-negative-positive $F(C)$, we might expect that the wave speed satisfies $v > 2\sqrt{F'(\beta)R(\beta)}$, such that the holes in the wall at $C = \beta$ are real-valued.

Following the approach used for Equation (8.7) with positive-negative-positive $F(C)$, it is simple to demonstrate that both the weak and reverse Allee effect have real-valued holes

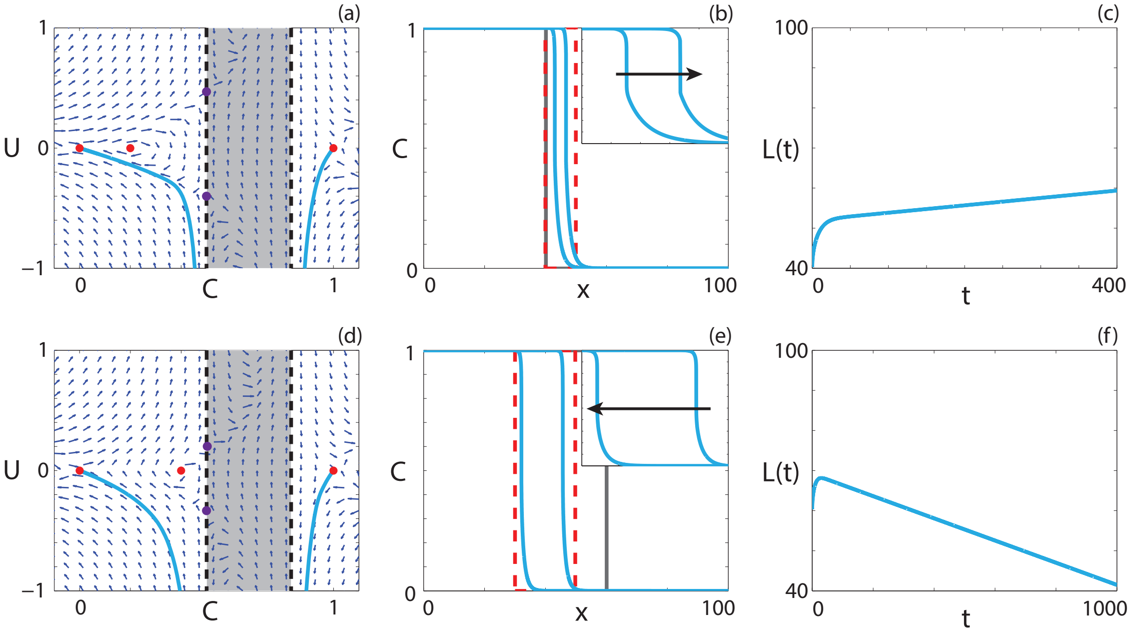

Figure 8.10: Travelling wave behaviour for Equation (8.21) **with the strong Allee effect and positive-negative-positive** $F(C)$ **(Case 6.3)**. (a), (d) Phase plane for the system (8.24)-(8.25) with the numerical solution of Equation (8.21) (cyan, solid), in (C, U) co-ordinates, superimposed. The dashed black lines denote a wall of singularities. Red circles correspond to equilibrium points and purple circles correspond to holes in the wall. (b), (e) Numerical solution of Equation (8.21) calculated at (b) $t = 200$ and $t = 400$; (e) $t = 500$ and $t = 1000$. The grey lines indicate the initial condition and the arrows indicate the direction of increasing time. The insets correspond to the areas within the red dashed lines, and highlight the shocks. (c), (f) The time evolution of $L(t)$. All results are obtained with $\delta x = 0.05$, $\delta t = 0.001$, $\epsilon = 10^{-6}$, $P_d^g = 0$, (a)-(c) $P_m^i = 0.5$, $P_m^g = 0.1$, $P_p^i = 0.5$, $P_p^g = 0.4$, $P_d^i = 0.6$, $v = 0.009$; (d)-(f) $P_m^i = 0.5$, $P_m^g = 0.1$, $P_p^i = 0.4$, $P_p^g = 0.2$, $P_d^i = 0.5$, $v = -0.028$.

in the wall (Supplementary Material). We now examine numerical solutions of Equation (8.21) with the strong Allee effect. For parameter regimes that give rise to wave speeds that satisfy $v > 2\sqrt{F'(\beta)R(\beta)}$, numerical travelling wave solutions could not be found. While the condition for real-valued holes in the wall is satisfied, the zeroes of Equation (8.25) are imaginary for a certain interval of $C > \beta$. This corresponds to a nullcline that is not real-valued for certain C values.

We now consider parameter regimes corresponding to the strong Allee effect with the additional restriction that $v < 2\sqrt{F'(C)R(C)}$ for $2/3 < C \leq 1$. For all P_m^i and P_m^g that give rise to a positive-negative-positive $F(C)$, holes in the wall at $C = \beta$ do not exist and, as such, we do not expect to obtain smooth solutions. Interestingly, we observe travelling wave solutions with shocks such that the solution never enters the region $\alpha < C < \beta$. An example of a shock-fronted travelling wave solution for the strong Allee effect with both $v > 0$ and $v < 0$ is shown in Figures 8.10(a)-(c) and Figures 8.10(d)-(f), respectively. Solutions of diffusion equations, without any source terms, that contain shocks have been reported previously [39, 72]. Similarly, shock-fronted travelling wave solutions arise in other kinds of models, including multispecies models of combustion [18] and haptotactic cell migration [71]. However, the models presented here are very different as our model contains a source term and no advection term, and it is therefore of interest to determine the properties of the reaction-diffusion equation that lead to shock-fronted travelling wave solutions.

Capacity-degenerate positive-negative nonlinear diffusivity function

Capacity-degenerate positive-negative $F(C)$, where $F(1) = 0$, arises if $P_m^g = 0$ and includes an interval $1/3 < C < 1$ where $F(C) < 0$. For the corresponding case with Fisher kinetics, despite the degenerate nature of the nonlinear diffusivity function at $C = 1$, we did not obtain solutions with a sharp front near $C = 1$. Instead, the solution passes through the region of negative diffusivity and a hole in the wall at $C = 1/3$, leading to smooth travelling wave solutions. As such, we expect similar solutions for both the weak and reverse Allee effect due to the qualitatively similar behaviour of the $R(C)$ function. It is of interest to examine whether smooth or shock-fronted travelling wave solutions arise from Equation (8.21) for the strong Allee effect and capacity-degenerate positive-negative $F(C)$, as for the positive-negative-positive $F(C)$ no smooth travelling wave solutions could be found.

As expected, smooth travelling wave solutions for both the weak and reverse Allee effects with capacity-degenerate positive-negative $F(C)$ are obtained. The solution behaviour for both the weak and reverse Allee effects are presented in the Supplementary Material. For the strong Allee effect, we examined a considerable number of parameter regimes and initial conditions and were unable to find travelling wave solutions.

Positive-negative nonlinear diffusivity function

For the case where $F_A(\overline{C})$ has exactly one zero on the interval $0 \leq \overline{C} \leq 1$ at $C = \omega$, Maini *et al.* [43] examine the existence of travelling wave solutions, and provide the necessary conditions for existence,

$$A_2 < \omega, \qquad v > 0, \qquad \int_0^\omega F(C)R(C)\,\mathrm{d}C > 0, \tag{8.33}$$

where $F(\omega) = 0$ and $0 < \omega < 1$. For the strong Allee effect in this parameter regime, the third part of Condition (8.33) corresponds to

$$\begin{aligned}&D_i(20(A_1 + A_2)\omega - 30A_1A_2\omega - 15\omega^2)+\\&(D_i - D_g)((84A_1 + 36A_2)\omega^3 - (45A_1A_2 + 60A_1 + 60A_2)\omega^2 - 30\omega^4 + 80A_1A_2\omega) > 0.\end{aligned} \tag{8.34}$$

Equation (8.21) is equivalent to

$$\frac{\partial C}{\partial \hat{t}} = \frac{\partial^2 C}{\partial x^2} + (K_i - K_g - \lambda_i + \lambda_g)A_1F(C)C\left(1 - \frac{C}{A_1}\right)\left(C - A_2\right),\ \hat{t} \geq 0, \tag{8.35}$$

on the interval $0 \leq C < \omega$, and equivalent to

$$\frac{\partial C}{\partial \hat{t}} = \frac{\partial^2 C}{\partial x^2} + (K_i - K_g - \lambda_i + \lambda_g) A_1 \hat{F}(C)(1-C)\left(1 - \frac{1-C}{A_1}\right)\left(1 - A_2 - C\right), \; \hat{t} \geq 0, \quad (8.36)$$

where $\hat{F}(C) = -F(1-C)$, on the interval $\omega < C \leq A_1$. The final necessary and sufficient condition from Maini *et al.* [43] for the existence of travelling wave solutions is that the minimum wave speed for Equation (8.35), v_1^*, is greater than, or equal to, the minimum wave speed for Equation (8.36), v_2^*. On the interval $0 \leq C < \omega$, Equation (8.21) has a strictly positive $F(C)$, where $F(C) \leq D_i$, and strong Allee kinetics. Hence, the minimum wave speed for Equation (8.35) has an upper bound, $v_1^* \leq \sqrt{2(\lambda_i - K_i)D_i}(1/2 - A_2)$. On the interval $\omega < C < A_1$ Equation (8.36) has a source term qualitatively similar to the Fisher-Kolmogorov equation and hence a lower bound for the minimum wave speed exists [43], $v_2^* \geq 2\sqrt{-F(A_1)(\lambda^2 + 4K_g(\lambda_i - \lambda_g - K_i + K_g))^{1/2}}$. For all parameter regimes considered that correspond to the strong Allee effect with positive-negative $F_A(\overline{C})$ we never observe a case where the upper bound for v_1^* is higher than the lower bound for v_2^* and hence the conditions required for travelling wave solutions are not met. As expected, numerical solutions of Equation (8.21) in these parameter regimes did not lead to travelling wave behaviour.

8.4 Discussion

In this work we present a discrete lattice-based model of birth, death and movement. The model is an exclusion process, and hence it explicitly incorporates crowding by allowing no more than one agent per site. A key feature of the model is that birth, death and movement rates depend on whether an agent is isolated or whether it is part of a group of agents. The discrete model can, therefore, be used to describe co-operative or competitive mechanisms [13, 22, 52, 67]. These kinds of mechanisms are thought to be relevant to many applications in cell biology [5, 37, 52, 58] and ecology [13, 22, 67]. By considering different combinations of parameters, the continuum limit PDE approximation of the discrete model leads to 22 different cases. These cases are reaction-diffusion equations with either Fisher kinetics or Allee kinetics, and a variety of density-dependent nonlinear diffusivity functions (Table 8.1). This approach also leads to a new kind of Allee effect, which we call the reverse Allee effect, where the growth rate is inhibited at high density. Although some of the PDEs that we consider have been investigated previously [1, 4, 7, 9, 14–17, 26, 34, 36, 38, 41–43, 49, 54–56, 67], they have never been linked together before using a single modelling framework. The presence of Allee kinetics allows for the more realistic description of biological and ecological phenomena, as the standard reaction-diffusion model with Fisher kinetics predicts either the population tending to extinction everywhere or the spread of the population in the form of a travelling wave. In comparison, Allee kinetics can describe population retreat, as well as shocks in the invading front of a population. Well-defined edges are thought to be present in invasive tumours [58], which can be described with travelling waves containing shocks.

In this work, we summarise properties of the long time travelling wave solutions for all classes of PDEs arising from our discrete model. For certain PDEs, where only existence of travelling wave solutions has been considered previously, we present numerical solutions here for the first time. We find that PDE models with density-dependent nonlinear diffusivity functions that have regions of negative diffusivity require a sufficiently non-negative source term to support smooth travelling wave solutions. Furthermore, there appears to be a threshold proliferation value, depending on the rate of motility, that must be exceeded for travelling wave solutions to be observed numerically. However, we do not comment on the putative relationship between the parameters in the discrete model and the existence of travelling wave solutions in the continuum limit PDE. Interestingly, for the strong Allee effect, shock-fronted travelling wave solutions are obtained. Following arguments presented in [43], we show that smooth travelling wave solutions cannot be obtained for certain types of nonlinear diffusivity functions and the strong Allee effect. We describe how nonlinear diffusion can either hinder or promote the persistence of a population, depending on the relative motility rates of the isolated and grouped agents. Interestingly, the motility rates affect the persistence differently for different carrying capacities. This relationship could provide insight into the requirements for a cell population, for example, to persist in the presence of a chemical treatment.

The six birth, death and motility rate parameters in the discrete model allow for the interpretation of the results in terms of whether individuals are part of, or isolated from, the bulk population. For example, a parameter regime corresponding to the strong Allee effect with constant diffusivity and no grouped agent death leads to the same travelling wave speed in the PDE description as a parameter regime corresponding to the strong Allee effect with constant diffusivity and a non-zero rate of grouped agent death, up to a threshold. This implies that a sufficiently strong intervention strategy aimed at grouped agents must be implemented if the goal of the intervention is to slow or halt the invasion of a population.

The work presented here suggests several avenues for future research. This work could be generalised by considering a two- or three–dimensional discrete process and deriving the continuum limit PDE descriptions in higher dimensions. This kind of higher-dimensional model might provide a more accurate description of real world observations where one-dimensional travelling wave solutions might not apply. In this work, numerical travelling wave solutions for each class of PDE are examined, but the formal stability of these travelling wave solutions is not considered. Another approach for analysing the discrete model would be to consider a coupled multispecies PDE model by accounting for the density of isolated agents and the density of grouped agents separately. This approach would lead to a system of two coupled PDEs instead of a single PDE for the total agent density. However, instead of working with coupled multispecies PDEs, we have taken the simplest and most fundamental approach of considering a single PDE description of the total population. In addition, a significant number of mechanisms could be implemented into the discrete model, such as cell-to-cell adhesion/repulsion [3, 31] or directed migration of isolated agents, such as chemotaxis [35]. We leave these extensions for future analysis.

8.5 Methods

8.5.1 Discrete model

We consider a one-dimensional lattice-based random walk with X sites and lattice spacing Δ [12]. Each site may be occupied by, at most, one agent [6, 10, 65]. The number of agents at time t is $N(t)$. Agents attempt to undergo birth, death and movement events. During a birth event, an agent attempts to place a daughter agent at a randomly selected nearest-neighbour site. This event is successful provided that the selected site is vacant. During a death event, an agent is removed from the lattice. During a movement event, an agent attempts to move to a randomly selected nearest-neighbour site. This event is successful provided that the selected site is vacant. We distinguish between types of agents based on the number of occupied nearest-neighbour sites for each agent [29]. We refer to agents with zero occupied nearest-neighbour sites as *isolated agents*, and agents with one or two occupied nearest-neighbour sites as *grouped agents*. This approach allows us to specify different birth, death and movement rates for isolated and grouped agents.

Different parameter choices can be used to impose either co-operative or competitive mechanisms, where an increase in local agent density provides a positive or negative benefit, respectively. Specifically, in situations where the group motility or group proliferation rates are higher than the isolated motility or isolated proliferation rates, respectively, we interpret this choice of parameters as a model of co-operation. Similarly, in situations where the group motility or group proliferation rates are lower than the isolated motility or isolated proliferation rates, respectively, we interpret this as a model of competition.

During each time step of duration τ, $N(t)$ agents are selected at random, one at a time, with replacement, and are given the opportunity to undergo a movement event. The constant probability that a selected agent attempts to undergo a movement event is P_m^i for an isolated agent and P_m^g for a grouped agent. We repeat this process for both birth and death events, with respective constant probabilities P_p^i and P_d^i for isolated agents and P_p^g and P_d^g for an agent within a group. At the end of each time step we update $N(t+\tau)$. To obtain the average agent density at each lattice site we perform M identically-prepared realisations of the discrete model and average the binary lattice occupancy at each lattice site at each time step. In any single realisation of the discrete model we have $C_j = 1$ when site j is occupied and $C_j = 0$ when site j is vacant. To evaluate the average occupancy of any lattice site we consider an ensemble of M identically-prepared realisations and calculate $\langle C_j \rangle = \sum_{m=1}^{M} C_j^m / M$.

8.5.2 Numerical techniques

Here we describe the techniques used to obtain numerical solutions of Equation (8.2), the corresponding ODE in travelling wave co-ordinates, and to generate the phase planes in (C, U) co-ordinates.

Partial differential equations

To obtain numerical solutions of Equation (8.2), we first spatially discretise Equation (8.2) onto a grid with uniform grid spacing δx by approximating the spatial derivatives with a central finite difference approximation. A backward Euler approximation with constant time steps of duration δt is used to approximate the temporal derivative. The resulting system of nonlinear algebraic equations is solved using Picard iteration with absolute convergence tolerance ϵ. The resulting system of tridiagonal algebraic equations is solved using the Thomas algorithm [53]. All results presented correspond to sufficiently small choices of δx, δt and ϵ so that the numerical solutions are grid independent. In all cases consider zero-flux boundary conditions are considered, and the finite domain is sufficiently large such that the numerical solution of Equation (8.2) does not interact with the boundaries on the time scale of the numerical simulations. All numerical solutions correspond to a Heaviside initial condition with $C = 1$ for $x \leq X_0$, and $C = 0$ otherwise.

Ordinary differential equations

The second order ODEs in the travelling wave co-ordinates are solved using Matlab's `ode45` routine [59]. This routine implements an adaptive Runge-Kutta method with relative error tolerance of 10^{-3} and an absolute error tolerance of 10^{-6} [59]. Travelling wave ODEs that contain a singularity are not solved numerically. Therefore, for these singular problems we obtain only the numerical solution of the PDE and present this solution in the transformed (C, U) travelling wave co-ordinate system.

Phase planes

To generate phase planes we substitute $U = \mathrm{d}C/\mathrm{d}z$ into the second order travelling wave ODE to obtain a system of two first-order ODEs. The phase plane is constructed by considering 22 equally-spaced values of C and 22 equally spaced values of U to calculate both $\mathrm{d}C/\mathrm{d}z$ and $\mathrm{d}U/\mathrm{d}z$ at all $22 \times 22 = 484$ pairs of (C, U) values. In each phase plane the same 22 equally spaced values of C on the interval $0 \leq C \leq 1$ are considered. However, depending on the steepness of the waveform, we choose a different interval of U to construct the phase plane, and this choice is made to accommodate the heteroclinic orbit. The phase planes are constructed using Matlab's `quiver` function. The location of the equilibrium points, where $\mathrm{d}C/\mathrm{d}z = \mathrm{d}U/\mathrm{d}z = 0$ are superimposed. Furthermore, in many cases the expression for $\mathrm{d}U/\mathrm{d}z$ has a rational form, $\mathrm{d}U/\mathrm{d}z = G(C, U)/H(C, U)$. In these cases both the wall of singularities $(H(C, U) = 0)$ and the locations of the holes in the wall $(H(C, U) = G(U, C) = 0)$ are also superimposed.

Acknowledgements

This work is supported by the Australian Research Council (DP140100249, FT130100148). We thank the referees for their helpful comments.

Author contributions

STJ, REB, DLSM and MJS conceived the experiments, STJ performed the experiments, STJ, REB, DLSM and MJS analysed the results, STJ and MJS wrote the manuscript. All authors read and approved the final version of the manuscript.

Additional information

Competing financial interests

The authors declare no competing financial interests.

Bibliography

[1] Ablowitz, M. J. and Zeppetella, A. [1979], 'Explicit solutions of Fisher's equation for a special wave speed', *Bulletin of Mathematical Biology* **41**(6), 835–840.

[2] Alvord Jr, E. C. and Shaw, C. M. [1991], *Neoplasms affecting the nervous system of the elderly*, Oxford University Press.

[3] Anguige, K. and Schmeiser, C. [2009], 'A one-dimensional model of cell diffusion and aggregation, incorporating volume filling and cell-to-cell adhesion', *Journal of Mathematical Biology* **58**(3), 395–427.

[4] Aronson, D. G. and Weinberger, H. F. [1978], 'Multidimensional nonlinear diffusion arising in population genetics', *Advances in Mathematics* **30**(1), 33–76.

[5] Axelrod, R., Axelrod, D. E. and Pienta, K. J. [2006], 'Evolution of cooperation among tumor cells', *Proceedings of the National Academy of Sciences* **103**(36), 13474–13479.

[6] Baker, R. E. and Simpson, M. J. [2010], 'Correcting mean-field approximations for birth-death-movement processes', *Physical Review E* **82**(4), 041905.

[7] Bramson, M. [1983], 'Convergence of solutions of the Kolmogorov equation to travelling waves', **44**, 285.

[8] Briscoe, B. K., Lewis, M. A. and Parrish, S. E. [2002], 'Home range formation in wolves due to scent marking', *Bulletin of Mathematical Biology* **64**(2), 261–284.

[9] Canosa, J. [1973], 'On a nonlinear diffusion equation describing population growth', *IBM Journal of Research and Development* **17**(4), 307–313.

[10] Chowdhury, D., Schadschneider, A. and Nishinari, K. [2005], 'Physics of transport and traffic phenomena in biology: from molecular motors and cells to organisms', *Physics of Life Reviews* **2**(4), 318–352.

[11] Clavero, M. and García-Berthou, E. [2005], 'Invasive species are a leading cause of animal extinctions', *Trends in Ecology and Evolution* **20**(3), 110–110.

[12] Codling, E. A., Plank, M. J. and Benhamou, S. [2008], 'Random walk models in biology', *Journal of the Royal Society Interface* **5**(25), 813–834.

[13] Courchamp, F., Clutton-Brock, T. and Grenfell, B. [1999], 'Inverse density dependence and the Allee effect', *Trends in Ecology and Evolution* **14**(10), 405–410.

[14] Ferracuti, L., Marcelli, C. and Papalini, F. [2009], 'Travelling waves in some reaction-diffusion-aggregation models', *Advances in Dynamical Systems and Applications* **4**(1), 19–33.

[15] Fife, P. C. [1979], 'Long time behavior of solutions of bistable nonlinear diffusion equations', *Archive for Rational Mechanics and Analysis* **70**(1), 31–36.

[16] Fife, P. C. and McLeod, J. B. [1977], 'The approach of solutions of nonlinear diffusion equations to travelling front solutions', *Archive for Rational Mechanics and Analysis* **65**(4), 335–361.

[17] Fisher, R. A. [1937], 'The wave of advance of advantageous genes', *Annals of Eugenics* **7**(4), 355–369.

[18] Forbes, L. K. and Derrick, W. [2001], 'A combustion wave of permanent form in a compressible gas', *The ANZIAM Journal* **43**(01), 35–58.

[19] Gabriely, G., Wurdinger, T., Kesari, S., Esau, C. C., Burchard, J., Linsley, P. S. and Krichevsky, A. M. [2008], 'MicroRNA 21 promotes glioma invasion by targeting matrix metalloproteinase regulators', *Molecular and Cellular Biology* **28**(17), 5369–5380.

[20] Gerlee, P. and Nelander, S. [2012], 'The impact of phenotypic switching on glioblastoma growth and invasion', *PLOS Computational Biology* **8**(6), e1002556.

[21] Gerlee, P. and Nelander, S. [2016], 'Travelling wave analysis of a mathematical model of glioblastoma growth', *Mathematical Biosciences* **276**, 75–81.

[22] Godin, J. G. J. [1986], 'Antipredator function of shoaling in teleost fishes: a selective review', *Le Naturaliste Canadien* **113**, 241–250.

[23] Hadeler, K. P. [1982], Travelling fronts and free boundary value problems, *in* 'Numerical Treatment of Free Boundary Value Problems', Springer, pp. 90–107.

[24] Hadeler, K. P. [1983], 'Free boundary problems in biological models', *Free boundary problems: Theory and applications* **2**, 664–671.

[25] Hadeler, K. P. [1987], Traveling fronts in parabolic and hyperbolic equations, *in* 'Dynamical Systems', Springer, pp. 154–164.

[26] Hadeler, K. P. and Rothe, F. [1975], 'Travelling fronts in nonlinear diffusion equations', *Journal of Mathematical Biology* **2**(3), 251–263.

[27] Hughes, B. D. [1995], *Random Walks and Random Environments*, Vol. 1, Clarendon Press; Oxford University Press.

[28] Ito, H., Kakishima, S., Uehara, T., Morita, S., Koyama, T., Sota, T., Cooley, J. R. and Yoshimura, J. [2015], 'Evolution of periodicity in periodical cicadas', *Scientific Reports* **5**, 14094.

[29] Johnston, S. T., Baker, R. E. and Simpson, M. J. [2016], 'Filling the gaps: A robust description of adhesive birth-death-movement processes', *Physical Review E* **93**(4), 042413.

[30] Johnston, S. T., Shah, E. T., Chopin, L. K., McElwain, D. L. S. and Simpson, M. J. [2015], 'Estimating cell diffusivity and cell proliferation rate by interpreting IncuCyte ZOOM assay data using the Fisher-Kolmogorov model', *BMC Systems Biology* **9**(38), 38.

[31] Johnston, S. T., Simpson, M. J. and Baker, R. E. [2012], 'Mean-field descriptions of collective migration with strong adhesion', *Physical Review E* **85**(5), 051922.

[32] Johnston, S. T., Simpson, M. J. and Baker, R. E. [2015], 'Modelling the movement of interacting cell populations: a moment dynamics approach', *Journal of Theoretical Biology* **370**, 81–92.

[33] Johnston, S. T., Simpson, M. J. and McElwain, D. L. S. [2014], 'How much information can be obtained from tracking the position of the leading edge in a scratch assay?', *Journal of The Royal Society Interface* **11**(97), 20140325.

[34] Keitt, T. H., Lewis, M. A. and Holt, R. D. [2001], 'Allee effects, invasion pinning, and species borders', *The American Naturalist* **157**(2), 203–216.

[35] Keller, E. F. and Segel, L. A. [1971], 'Model for chemotaxis', *Journal of Theoretical Biology* **30**(2), 225–234.

[36] Kolmogorov, A. N., Petrovsky, I. G. and Piskunov, N. S. [1937], 'Étude de léquation de la diffusion avec croissance de la quantité de matiere et son application un probleme biologique', *Moscow University Mathematics Bulletin* **1**, 1–25.

[37] Korolev, K. S., Xavier, J. B. and Gore, J. [2014], 'Turning ecology and evolution against cancer', *Nature Reviews Cancer* **14**(5), 371–380.

[38] Kuzmin, M. and Ruggerini, S. [2011], 'Front propagation in diffusion-aggregation models with bi-stable reaction', *Discrete and Continuous Dynamical Systems: Series B* **16**(3), 819–833.

[39] Landman, K. A. and White, L. R. [2011], 'Terraced spreading of nanofilms under a nonmonotonic disjoining pressure', *Physics of Fluids* **23**(1), 012004.

[40] Lewis, M. [1994], 'Spatial coupling of plant and herbivore dynamics: the contribution of herbivore dispersal to transient and persistent "waves" of damage', *Theoretical Population Biology* **45**(3), 277–312.

[41] Lewis, M. A. and Kareiva, P. [1993], 'Allee dynamics and the spread of invading organisms', *Theoretical Population Biology* **43**(2), 141–158.

[42] Maini, P. K., Malaguti, L., Marcelli, C. and Matucci, S. [2006], 'Diffusion-aggregation processes with mono-stable reaction terms', *Discrete and Continuous Dynamical Systems: Series B* **6**(5), 1175–1189.

[43] Maini, P. K., Malaguti, L., Marcelli, C. and Matucci, S. [2007], 'Aggregative movement and front propagation for bi-stable population models', *Mathematical Models and Methods in Applied Sciences* **17**(09), 1351–1368.

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

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

[46] Malaguti, L. and Marcelli, C. [2003], 'Sharp profiles in degenerate and doubly degenerate Fisher-KPP equations', *Journal of Differential Equations* **195**(2), 471–496.

[47] Malaguti, L., Marcelli, C. and Matucci, S. [2004], 'Front propagation in bistable reaction-diffusion-advection equations', *Advances in Differential Equations* **9**(9-10), 1143–1166.

[48] Martin, P. [1997], 'Wound healing – aiming for perfect skin regeneration', *Science* **276**(5309), 75–81.

[49] Murray, J. D. [2002], *Mathematical Biology I: An Introduction*, Springer, New York.

[50] Padrón, V. [2004], 'Effect of aggregation on population recovery modeled by a forward-backward pseudoparabolic equation', *Transactions of the American Mathematical Society* **356**(7), 2739–2756.

[51] Pettet, G. J., McElwain, D. L. S. and Norbury, J. [2000], 'Lotka-volterra equations with chemotaxis: Walls, barriers and travelling waves', *Mathematical Medicine and Biology* **17**(4), 395–413.

[52] Poujade, M., Grasland-Mongrain, E., Hertzog, A., Jouanneau, J., Chavrier, P., Ladoux, B., Buguin, A. and Silberzan, P. [2007], 'Collective migration of an epithelial monolayer in response to a model wound', *Proceedings of the National Academy of Sciences* **104**(41), 15988–15993.

[53] Press, W. H. [2007], *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, Cambridge University Press.

[54] Rothe, F. [1981], 'Convergence to pushed fronts', *Journal of Mathematics* **11**(4), 617–634.

[55] Sánchez-Garduño, F. and Maini, P. K. [1994], 'Existence and uniqueness of a sharp travelling wave in degenerate non-linear diffusion Fisher-KPP equations', *Journal of Mathematical Biology* **33**(2), 163–192.

[56] Sánchez-Garduño, F. and Maini, P. K. [1997], 'Travelling wave phenomena in non-linear diffusion degenerate Nagumo equations', *Journal of Mathematical Biology* **35**(6), 713–728.

[57] Semmens, B. X., Semmens, D. J., Thogmartin, W. E., Wiederholt, R., López-Hoffman, L., Diffendorfer, J. E., Pleasants, J. M., Oberhauser, K. S. and Taylor, O. R. [2016], 'Quasi-extinction risk and population targets for the Eastern, migratory population of monarch butterflies (danaus plexippus)', *Scientific Reports* **6**, 23265.

[58] Sewalt, L., Harley, K., van Heijster, P. and Balasuriya, S. [2016], 'Influences of Allee effects in the spreading of malignant tumours', *Journal of Theoretical Biology* **394**, 77–92.

[59] Shampine, L. F. and Reichelt, M. W. [1997], 'The Matlab ODE suite', *SIAM Journal on Scientific Computing* **18**(1), 1–22.

[60] Sherratt, J. A. [2010], 'On the form of smooth-front travelling waves in a reaction-diffusion equation with degenerate nonlinear diffusion', *Mathematical Modelling of Natural Phenomena* **5**(5), 64–79.

[61] Sherratt, J. A., Lewis, M. A. and Fowler, A. C. [1995], 'Ecological chaos in the wake of invasion.', *Proceedings of the National Academy of Sciences* **92**(7), 2524–2528.

[62] Sherratt, J. A. and Marchant, B. P. [1996], 'Nonsharp travelling wave fronts in the Fisher equation with degenerate nonlinear diffusion', *Applied Mathematics Letters* **9**(5), 33–38.

[63] Sherratt, J. A. and Murray, J. D. [1990], 'Models of epidermal wound healing', *Proceedings of the Royal Society of London B: Biological Sciences* **241**(1300), 29–36.

[64] Simpson, M. J. and Baker, R. E. [2011], 'Corrected mean-field models for spatially dependent advection-diffusion-reaction phenomena', *Physical Review E* **83**(5), 051922.

[65] Simpson, M. J., Landman, K. A. and Hughes, B. D. [2010], 'Cell invasion with proliferation mechanisms motivated by time-lapse data', *Physica A: Statistical Mechanics and its Applications* **389**(18), 3779–3790.

[66] Swanson, K. R., Bridge, C., Murray, J. D. and Alvord, E. C. [2003], 'Virtual and real brain tumors: using mathematical modeling to quantify glioma growth and invasion', *Journal of the Neurological Sciences* **216**(1), 1–10.

[67] Taylor, C. M. and Hastings, A. [2005], 'Allee effects in biological invasions', *Ecology Letters* **8**(8), 895–908.

[68] Townsend, C. R. [1996], 'Invasion biology and ecological impacts of brown trout *salmo trutta* in New Zealand', *Biological Conservation* **78**(1), 13–22.

[69] Turchin, P. [1989], 'Population consequences of aggregative movement', *The Journal of Animal Ecology* pp. 75–100.

[70] Wang, J., Yang, Y., Yang, F., Li, Y., Li, L., Lin, D., He, T., Liang, B., Zhang, T., Lin, Y. et al. [2016], 'A framework for the assessment of the spatial and temporal patterns of threatened coastal delphinids', *Scientific Reports* **6**, 19883.

[71] Wechselberger, M. and Pettet, G. J. [2010], 'Folds, canards and shocks in advection–reaction–diffusion models', *Nonlinearity* **23**(8), 1949.

[72] Witelski, T. P. [1995], 'Shocks in nonlinear diffusion', *Applied Mathematics Letters* **8**(5), 27–32.

8.6 Supplementary material

This supplementary material for the manuscript 'Co-operation, competition and crowding: a discrete framework linking Allee kinetics, nonlinear diffusion, shocks and sharp-fronted travelling waves' contains a more detailed analysis of all of the different cases of partial differential equations (PDEs) that arise from the continuum description of the discrete model of co-operation and competition. Eight cases of different combinations of co-operative, competitive or neutral mechanisms are now considered systematically. Some of these cases involve distinct sub-cases so that, in total, we consider 22 different classes of PDE models of invasion. The properties of each of these distinct cases are presented in Table 8.1 in the main document. Note that certain cases are examined in the main document and hence there is some unavoidable overlap between the two documents, as here we examine all cases in a systematic manner.

8.6.1 Equal motility rates, equal proliferation rates, no agent death.

For $P_m^i = P_m^g$, $P_p^i = P_p^g$ and $P_d^i = P_d^g = 0$, there is no co-operative or competitive mechanism. This gives $F(C) = D = D_i = D_g$, and $R(C) = \lambda C(1-C)$, where $\lambda = \lambda_i = \lambda_g$. Therefore, Equation (8.2) reduces to the Fisher-Kolmogorov equation [17, 36, 49]

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} + \lambda C(1-C). \tag{8.37}$$

As the source term is non-negative for all physical values of $C \geq 0$, the agent population will always eventually reach the carrying capacity.

The Fisher-Kolmogorov equation has been studied extensively [1, 4, 7, 9, 17, 26, 36, 49]. Here we present the key results in the context of examining the long time travelling wave solution. We seek right moving travelling waves in the co-ordinate $z = x - vt$, $-\infty < z < \infty$, where v is a constant wave speed [49]. Transforming Equation (8.37) into the travelling wave co-ordinate gives

$$D\frac{\mathrm{d}^2 C}{\mathrm{d}z^2} + v\frac{\mathrm{d}C}{\mathrm{d}z} + \lambda C(1-C) = 0, \quad -\infty < z < \infty. \tag{8.38}$$

With $U = \mathrm{d}C/\mathrm{d}z$, Equation (8.38) can be expressed as a system of ordinary differential equations (ODEs),

$$\frac{\mathrm{d}C}{\mathrm{d}z} = U, \tag{8.39}$$

$$\frac{\mathrm{d}U}{\mathrm{d}z} = -\frac{v}{D}U - \frac{\lambda}{D}C(1-C). \tag{8.40}$$

This system has two equilibrium points: $(C,U) = (0,0)$, and $(C,U) = (1,0)$. The linear stability of these equilibrium points can be analysed by examining the eigenvalues of the Jacobian at each equilibrium point. At $(0,0)$ the characteristic equation has solutions $\xi = (-v \pm \sqrt{v^2 - 4\lambda D})/2D$, implying that $(0,0)$ is a stable node provided that $v > 2\sqrt{\lambda D}$,

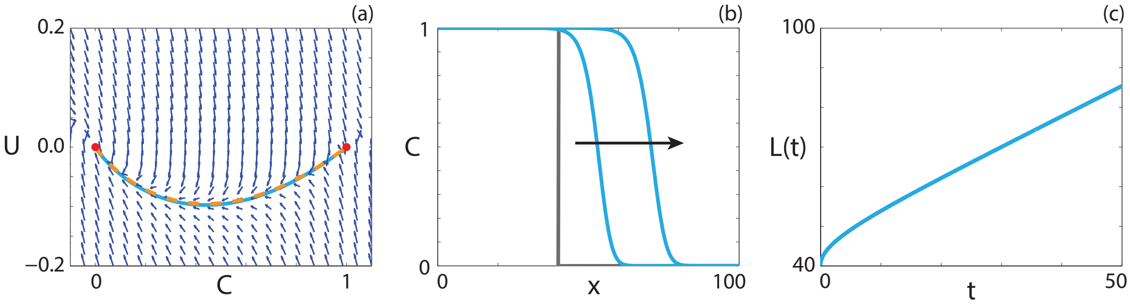

Figure 8.11: Travelling wave behaviour for the Fisher-Kolmogorov model (Case 1). (a) Phase plane for the system (8.39)-(8.40). Red circles denote equilibrium points. The numerical solutions of Equation (8.37) (cyan, solid) and Equation (8.38) (orange, dashed), in (C, U) co-ordinates, are superimposed. (b) Numerical solution of Equation (8.37) at $t = 25$ and $t = 50$ (blue). The grey line indicates the initial condition and the arrow indicates the direction of increasing time. (c) The time evolution of $L(t)$. All results are obtained using $P_m^i = P_m^g = 1$, $P_p^i = P_p^g = 0.3$, $P_d^g = P_d^i = 0$, $\delta x = 0.1$, $\delta t = 0.01$, $\epsilon = 10^{-6}$, $v = 0.768$.

and a stable spiral (focus) otherwise, as λ and D are both positive. We therefore have a minimum wave speed condition, $v^* = 2\sqrt{\lambda D}$, that must be satisfied otherwise the solution trajectory will enter non-physical regions of the phase plane [49]. The Jacobian of the linearised system at $(1, 0)$ has eigenvalues $\xi = (-v \pm \sqrt{v^2 + 4\lambda D})/2D$, implying that $(1, 0)$ is a saddle point.

The phase plane and associated heteroclinic orbit for Equations (8.39)-(8.40) are shown in Figure 8.11(a). Details of the numerical techniques used to solve Equation (8.38) and to generate the phase planes are given in the Methods (Main Document). Provided $v \geq 2\sqrt{\lambda D}$ we observe a heteroclinic orbit between $(1, 0)$ and $(0, 0)$. The numerical solution of Equation (8.38) and the numerical solution of Equation (8.37), transformed into (C, U) co-ordinates, are superimposed, showing a good match. This result is unsurprising, as Equation (8.38) is solved using the minimum wave speed, $v = v^* = 2\sqrt{\lambda D}$, and the numerical solution of Equation (8.37) evolves from a Heaviside initial condition, which is known to approach a travelling wave moving at the minimum wave speed [49]. The numerical solution of Equation (8.37) at $t = 25$ and $t = 50$ is shown in Figure 8.11(b), confirming that the waveform does not change with time. To quantify the wave speed we calculate the time evolution of the leading edge, $L(t) = x_f$ such that $C(x_f, t) \approx 1 \times 10^{-4}$. If the solution of Equation (8.37) forms a travelling wave, $L(t)$ will tend to a straight line with slope v, as $t \to \infty$. In Figure 8.11(c), we observe that $L(t)$ is approximately linear with slope v, and hence the solution of Equation (8.37) moves with approximately constant speed at late time. Overall, these features suggest that the solution of Equation (8.37) is a travelling wave.

8.6.2 Different motility rates, equal proliferation rates, no agent death.

If $P_m^i \neq P_m^g$ the governing PDE contains a nonlinear diffusivity term. Since the agent birth rate is independent of agent type and agents do not die, we consider the same source

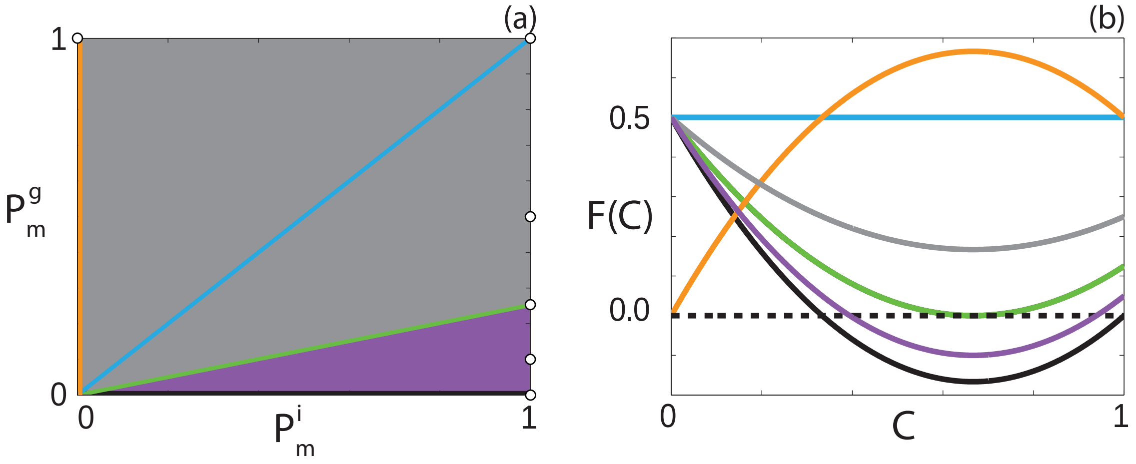

Figure 8.12: Classification of $F(C)$. (a) Type of $F(C)$ function for $0 \leq C \leq 1$ for the parameter space $P_m^i \in [0,1]$ and $P_m^g \in [0,1]$. The grey region represents parameter pairs that result in only positive $F(C)$, and the purple region represents parameter pairs that result in negative $F(C)$ for an interval of C. The orange line represents parameter pairs that result in $F(0) = 0$, the green line represents parameter pairs that result in $F(2/3) = 0$, the cyan line represents parameter pairs that result in constant $F(C)$, and the black line represents parameter pairs that result in negative $F(C)$ for an interval of C with $F(1) = 0$. (b) Example $F(C)$ for each region in (a). Positive $F(C)$ (grey), corresponding to $P_m^i = 1$ and $P_m^g = 0.5$. Negative $F(C)$ for an interval of C (purple), corresponding to $P_m^i = 1$ and $P_m^g = 0.1$. Negative $F(C)$ for an interval of C with $F(1) = 0$ (black), corresponding to $P_m^i = 1$ and $P_m^g = 0.1$. $F(0) = 0$ (orange), corresponding to $P_m^i = 0$ and $P_m^g = 1$. $F(2/3) = 0$ (green), corresponding to $P_m^i = 1$ and $P_m^g = 0.25$. Constant $F(C)$ (cyan), corresponding to $P_m^i = 1$ and $P_m^g = 1$. The white circles in (a) denote the parameter pairs used to generate the curves in (b).

term as for Case 1. Again, there are no competitive or co-operative mechanisms associated with birth/death but it could be either advantageous ($P_m^i > P_m^g$) or disadvantageous ($P_m^i < P_m^g$) for an individual to be isolated from the bulk population, if the goal for the population is to invade unoccupied space. In this parameter regime, Equation (8.2) simplifies to

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x}\left(F(C)\frac{\partial C}{\partial x}\right) + \lambda C(1-C), \tag{8.41}$$

where $\lambda = \lambda_i = \lambda_g$ and $F(C) = D_i(1 - 4C + 3C^2) + D_g(4C - 3C^2)$.

$F(C)$ has different properties depending on the choice of P_m^i and P_m^g. To illustrate this, we present the (P_m^i, P_m^g) parameter space in Figure 8.12(a), and highlight regions of different behaviour of $F(C)$. If $P_m^i > 4P_m^g$, there will be an interval, $1/3 \leq \alpha < C < \beta \leq 1$, centred around $C = 2/3$, where $F(C) < 0$. Specifically, this interval is given by

$$\alpha = \frac{2}{3} - \frac{\sqrt{(P_m^i)^2 - 5P_m^i P_m^g + 4(P_m^g)^2}}{3(P_m^i - P_m^g)} < C < \beta = \frac{2}{3} + \frac{\sqrt{(P_m^i)^2 - 5P_m^i P_m^g + 4(P_m^g)^2}}{3(P_m^i - P_m^g)}. \tag{8.42}$$

All parameter pairs that result in $F(C) < 0$, which we refer to as *positive-negative-positive*, correspond to the purple region in Figure 8.12(a), and an example $F(C)$ curve is given in Figure 8.12(b). Parameter pairs that result in $F(C) < 0$ with $F(1) = 0$ correspond to the black line in Figure 8.12(a), and an example $F(C)$ curve is given in Figure 8.12(b). We refer to this type of nonlinear diffusivity function as *capacity-degenerate positive-negative*.

It is relevant for us to remark that nonlinear diffusivity functions with negative regions can lead to shocks in the solution of nonlinear diffusion equations without any source term [39, 72]. Therefore, it is instructive to consider whether shock-fronted travelling waves exist with Fisher-Kolmogorov kinetics.

For specific parameter regimes, $F(C)$ is degenerate at $C = 0$, that is, $F(0) = R(0) = 0$. This type of nonlinear diffusivity function, which we refer to as *extinction-degenerate non-negative*, leads to sharp-fronted travelling waves, provided that $F(C) \geq 0$ for $0 \leq C \leq 1$ [55, 60, 62]. For Equation (8.41), this corresponds to $P_m^i = 0$. The parameter pairs that satisfy this condition correspond to the orange line in Figure 8.12(a), and a typical $F(C)$ curve is given in Figure 8.12(b). The special case $P_m^i = P_m^g$ leads to a constant diffusivity, and parameter pairs that satisfy lie along the cyan line in Figure 8.12(a). A typical $F(C)$ curve for this case is presented in Figure 8.12(b). For all other parameter pairs $F(C) > 0$, which we refer to as *strictly positive*, and these parameter pairs correspond to the grey region in Figure 8.12(a), for which an example $F(C)$ curve is shown in Figure 8.12(b).

We look for a right moving travelling wave solution of Equation (8.41) in terms of the co-ordinate $z = x - vt$. Transforming Equation (8.41) into travelling wave co-ordinates, we obtain

$$v\frac{\mathrm{d}C}{\mathrm{d}z} + F(C)\frac{\mathrm{d}^2C}{\mathrm{d}z^2} + (D_i - D_g)(6C-4)\left(\frac{\mathrm{d}C}{\mathrm{d}z}\right)^2 + \lambda C(1-C) = 0, \quad -\infty < z < \infty. \tag{8.43}$$

Making the substitution $U = \mathrm{d}C/\mathrm{d}z$ gives

$$\frac{\mathrm{d}C}{\mathrm{d}z} = U, \tag{8.44}$$

$$\frac{\mathrm{d}U}{\mathrm{d}z} = \frac{-vU - (D_i - D_g)(6C-4)U^2 - \lambda C(1-C)}{F(C)}. \tag{8.45}$$

We now consider the properties of the travelling wave solutions for several sub-cases within Case 2. Unlike the Fisher-Kolmogorov equation, the minimum wave speed is unknown and hence all phase planes presented in this section are generated with v obtained from the numerical solution of Equation (8.41) at sufficiently late time.

Strictly positive nonlinear diffusivity function

If $F(C) > 0$ for $0 \leq C \leq 1$, Equations (8.44)-(8.45) are not singular for $0 \leq C \leq 1$. Hence the linear analysis performed for Case 1 is valid in terms of the position and stability of the equilibrium points. The exception is the minimum wave speed condition for the equilibrium point at $(0, 0)$ to be a stable node, which becomes $v > 2\sqrt{\lambda D_i}$, which is always positive [23–25].

Solutions of Equation (8.41), illustrating travelling wave behaviour for two different $F(C)$ functions are given in Figures 8.13(a)-(c) and Figures 8.13(d)-(f), respectively. In both cases the solution trajectory in the phase plane, Figure 8.13(a) and Figure 8.13(d), forms a

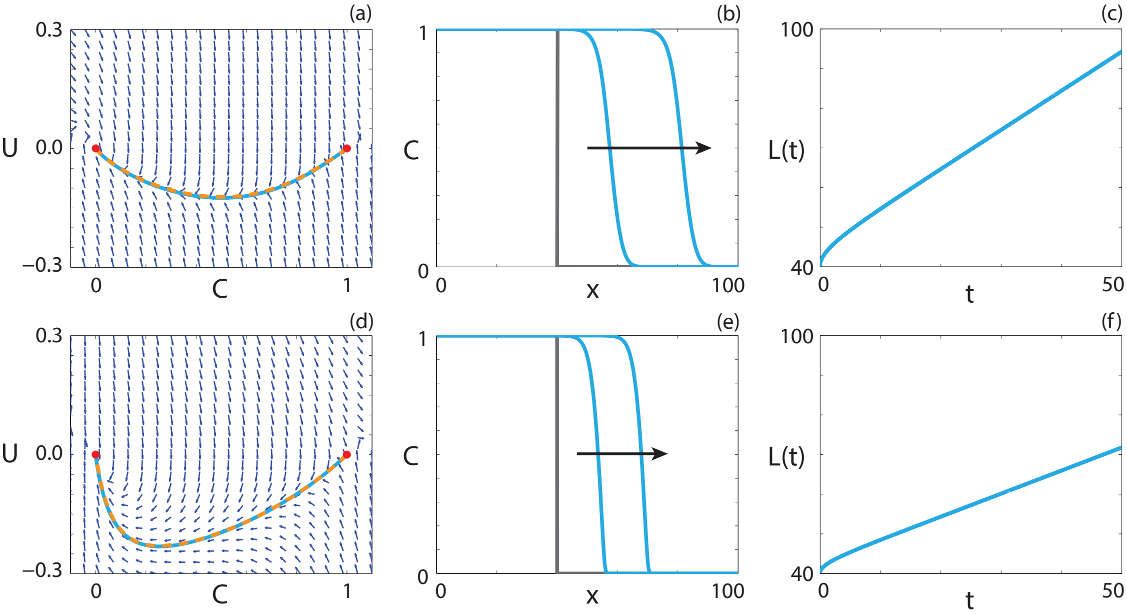

Figure 8.13: Travelling wave behaviour for Equation (8.41) **with strictly positive** $F(C)$ **(Case 2.1)**. (a), (d) Phase plane for the system (8.44)-(8.45). Red circles denote equilibrium points. The numerical solutions of Equation (8.41) (cyan, solid) and Equation (8.43) (orange, dashed), in (C, U) co-ordinates, are superimposed. (b), (e) Numerical solution of Equation (8.41) at $t = 25$ and $t = 50$ (blue). The grey lines indicate the initial condition and the arrows indicate the direction of increasing time. (c), (f) The time evolution of the position of the leading edge of the travelling wave solution. All results are obtained using $P_p^i = P_p^g = 0.5$, $P_d^i = P_d^g = 0$, $\delta x = 0.1$, $\delta t = 0.01$, $\epsilon = 10^{-6}$ and (a)-(c) $P_m^i = 1.0$, $P_m^g = 0.5$, $v = 0.992$; (d)-(f) $P_m^i = 0.2$, $P_m^g = 0.8$, $v = 0.584$.

heteroclinic orbit between $(1, 0)$ and $(0, 0)$. Interestingly, the waveform in Figure 8.13(e), with $P_m^g > P_m^i$, is relatively sharp near $C = 0$. If $P_m^g > P_m^i$, $F(C)$ is concave up with a minimum value of $P_m^i/2$ at $C = 0$, for $0 \leq C \leq 1$, whereas $F(C)$ has a minimum value at $C = 2/3$ for $P_m^i > P_m^g$. This suggests that $F(0)$ has considerable influence on the waveform.

The observed wave speed in Figure 8.13(a), $v = 0.992$, is close to the predicted minimum wave speed $v^* = 2\sqrt{\lambda D_i} = 1$, whereas the observed wave speed in Figure 8.13(b), $v = 0.584$, is greater than the predicted minimum wave speed $v^* = 0.447$. To determine whether v^* provides an accurate prediction of the observed wave speed, we calculate the long time numerical solution of Equation (8.41) and measure v for a suite of P_m^i and P_m^g values. Predicted minimum wave speeds and observed wave speeds are compared in Figure 8.14. In Figure 8.14(a), the predicted wave speed is accurate for all P_m^g values with $P_m^i = 0.8$. Interestingly, with $P_m^i = 0.2$, the predicted wave speed is only accurate for $P_m^g \leq 0.4$. Setting $P_m^g = 1$ and varying P_m^i we observe, in Figure 8.14(b), that the prediction is accurate for $P_m^i \geq 0.5$. Hence, it appears that for $P_m^i \geq 2P_m^g$ the minimum wave speed condition is accurate. For $P_m^i < 2P_m^g$ the grouped agents may have more successful movement events than the individual agents. Therefore, the dominant contribution to the invasion of the population may be attributed to the grouped agents,

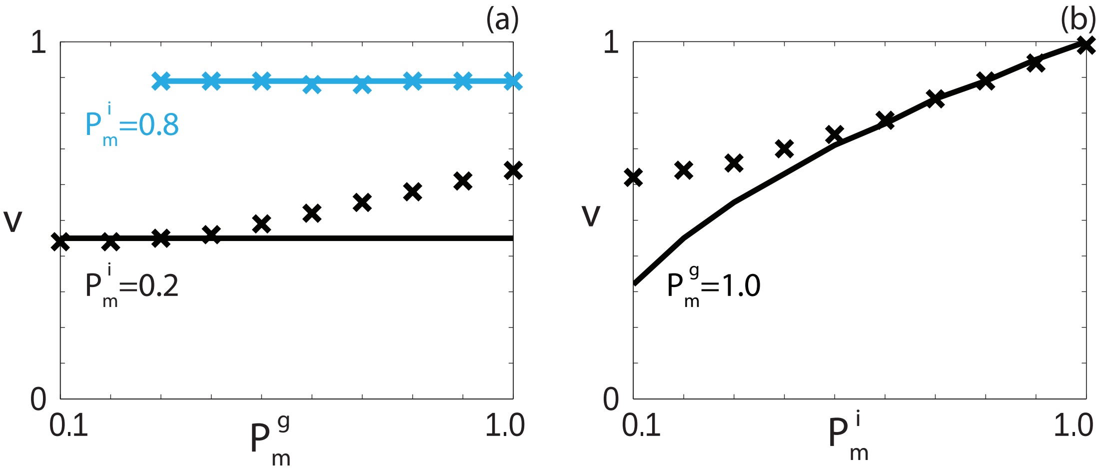

Figure 8.14: Wave speed comparison for Case 2.1. Comparison of the minimum wave speed condition (solid) and the observed wave speed at sufficiently late time (crosses) for (a) constant P_m^g and a suite of P_m^i values; (b) constant P_m^i and a suite of P_m^g values. All results are obtained using $P_p^i = P_p^g = 0.5$, $P_d^i = P_d^g = 0$, $\delta x = 0.1$, $\delta t = 0.01$, $\epsilon = 10^{-6}$ and the Heaviside initial condition.

which could explain why the minimum wave speed, which depends on P_m^i, does not provide a good estimate of the observed wave speed in these parameter regimes.

Extinction-degenerate non-negative nonlinear diffusivity function

The case where $F(0) = R(0) = 0$, and $F(C) > 0$ for $0 < C \leq 1$, occurs when $P_m^i = 0$. Under these conditions Equations (8.44)-(8.45) simplify to

$$\frac{\mathrm{d}C}{\mathrm{d}z} = U, \tag{8.46}$$

$$\frac{\mathrm{d}U}{\mathrm{d}z} = \frac{1}{D_g(4C - 3C^2)}\bigg(- vU + D_g(6C-4)U^2 - \lambda C(1-C)\bigg), \quad -\infty < z < \infty. \tag{8.47}$$

Note that Equation (8.47) is singular at $C = 0$ and, furthermore, that $R(0) = 0$. Hence we apply a stretching transformation

$$\zeta = \int_0^z \frac{1}{D_g(4C(s) - 3C(s)^2)}\,\mathrm{d}s, \tag{8.48}$$

to remove the singularity, which gives

$$\frac{\mathrm{d}C}{\mathrm{d}\zeta} = D_g U(4C - 3C^2), \tag{8.49}$$

$$\frac{\mathrm{d}U}{\mathrm{d}\zeta} = -vU + D_g(6C-4)U^2 - \lambda C(1-C), \;\; \zeta \geq 0. \tag{8.50}$$

Equations (8.49)-(8.50) have equilibrium points at $(C, U) = (1, 0)$, $(C, U) = (0, 0)$ and $(C, U) = (0, -v/4D_g)$. The additional equilibrium point in the transformed system corresponds to a solution trajectory approaching $C = 0$ with a non-zero slope. Performing linear analysis to determine the eigenvalues of the Jacobian at the steady states, we find

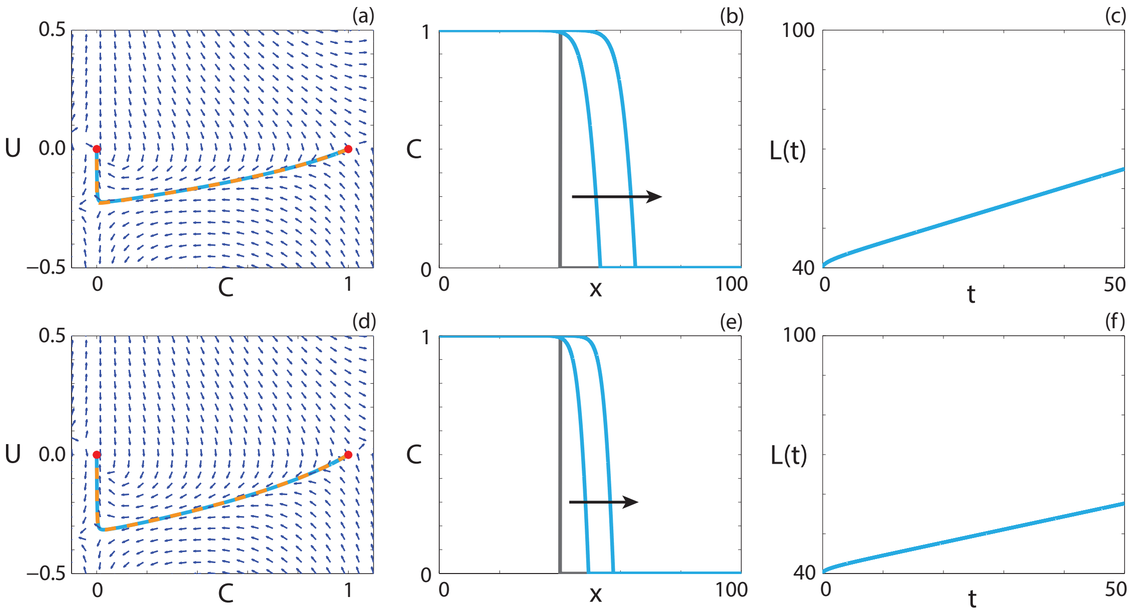

Figure 8.15: Travelling wave behaviour for Equation (8.41) **with extinction-degenerate non-negative** $F(C)$ **(Case 2.2)**. (a), (d) Phase plane for the system (8.46)-(8.47). Red circles denote equilibrium points. The numerical solutions of Equations (8.41) (cyan, solid) and (8.43) (orange, dashed), in (C, U) co-ordinates, are superimposed. (b), (e) Numerical solution of Equation (8.41) at $t = 25$ and $t = 50$. The grey lines indicate the initial condition and the arrows indicate the direction of increasing time. (c), (f) The time evolution of the position of the leading edge of the travelling wave solution. All results are obtained using $P_m^i = 0$, $P_p^i = P_p^g = 0.3$, $P_d^g = P_d^i = 0$, $\delta x = 0.01$, $\delta t = 0.01$, $\epsilon = 10^{-6}$ and (a)-(c) $P_m^g = 1.0$, $v = 0.463$; (d)-(f) $P_m^g = 0.5$, $v = 0.328$.

that the characteristic equation at $(0, 0)$ has solutions $\xi = 0$ and $\xi = v$, implying that $(0, 0)$ is an improper node. Sánchez-Garduño and Maini [55] investigate the stability of this equilibrium point and find that the equilibrium point is a saddle node. The characteristic equation at $(1, 0)$ has solutions $\xi = (-v \pm \sqrt{v^2 + 4\lambda D_g})/2$, implying that $(1, 0)$ is a saddle point. Finally the characteristic equation of the equilibrium point at $(0, -v/4D_g)$ has eigenvalues $\xi = \pm v$, implying that $(0, -v/4D_g)$ is a saddle point. A critical value v^* exists such that $v < v^*$ results in no travelling wave solution, $v = v^*$ results in a sharp-fronted travelling wave and $v > v^*$ results in a classic (smooth) travelling wave [55].

Numerical solutions illustrating travelling wave behaviour for Equation (8.41) with $P_m^i = 0$ are given in Figure 8.15. In the phase plane for both cases, Figure 8.15(a) and Figure 8.15(d), the solution trajectory tends to the origin with $\mathrm{d}U/\mathrm{d}C$ large and negative. The corresponding numerical solutions of Equation (8.41), presented in Figure 8.15(b) and Figure 8.15(e), approach a travelling wave solution with a sharp front near $C = 0$. This result is expected as the Heaviside initial condition results in the minimum wave speed that, for a degenerate diffusivity function, results in a sharp-fronted wave [62].

Positive-negative-positive nonlinear diffusivity function

In order for $F(C)$ to change sign twice, that is, $F(C) < 0$ for $1/3 \leq \alpha < C < \beta \leq 1$ and $F(C) \geq 0$ otherwise for $0 \leq C \leq 1$, the parameters must lie within the purple region in Figure 8.12(a). In this situation, Equations (8.44)-(8.45) are undefined at $C = \alpha$ and $C = \beta$, and these singularities cannot be removed using a stretching transformation since $R(\alpha) \neq 0$ and $R(\beta) \neq 0$. However, it is possible for $\mathrm{d}U/\mathrm{d}z$ to be finite at $C = \alpha$ and $C = \beta$ if U_α and U_β exist such that

$$\lim_{C \to \alpha} \left[\frac{-vU_\alpha - (D_i - D_g)(6C-4)U_\alpha^2 - \lambda C(1-C)}{D_i(1-4C+3C^2) + D_g(4C-3C^2)} \right], \tag{8.51}$$

$$\lim_{C \to \beta} \left[\frac{-vU_\beta - (D_i - D_g)(6C-4)U_\beta^2 - \lambda C(1-C)}{D_i(1-4C+3C^2) + D_g(4C-3C^2)} \right], \tag{8.52}$$

are both finite. This requires the numerator in the expressions (8.51)-(8.52) vanish at $C = \alpha$ and $C = \beta$, respectively. As such, U_α and U_β are obtained by solving the system

$$0 = -vU_\alpha - (D_i - D_g)(6\alpha - 4)U_\alpha^2 - \lambda\alpha(1-\alpha), \tag{8.53}$$

$$0 = -vU_\beta - (D_i - D_g)(6\beta - 4)U_\beta^2 - \lambda\beta(1-\beta), \tag{8.54}$$

resulting in $U_\alpha = -(v \pm \sqrt{v^2 - 4F'(\alpha)R(\alpha)})/2F'(\alpha)$ and $U_\beta = -(v \pm \sqrt{v^2 - 4F'(\beta)R(\beta)})/2F'(\beta)$. We note that, as $R(C) \geq 0$ for $0 \leq C \leq 1$, and $F'(\alpha) \leq 0$ for all possible α values, U_α will be real-valued. Subsequently, we have a wave speed condition that $v \geq 2\sqrt{F'(\beta)R(\beta)}$, as $F'(\beta) \geq 0$ for all possible β values. Ferracuti *et al.* [14] prove that the minimum wave speed, v^*, is greater than a threshold value, which in turn is greater than $\max\{R'(0)F(0), F'(\beta)R(\beta)\}$. Therefore, U_β will also always be real-valued.

Applying L'Hopital's Rule to Equation (8.45), we obtain

$$\lim_{C \to \alpha} \left. \frac{\mathrm{d}U}{\mathrm{d}z} \right|_{U = U_\alpha} = \lim_{C \to \alpha} \left[\frac{6(D_i - D_g)U_\alpha^2 + \lambda(1-2C)}{(D_g - D_i)(6C-4)} \right], \tag{8.55}$$

$$\lim_{C \to \beta} \left. \frac{\mathrm{d}U}{\mathrm{d}z} \right|_{U = U_\beta} = \lim_{C \to \beta} \left[\frac{6(D_i - D_g)U_\beta^2 + \lambda(1-2C)}{(D_g - D_i)(6C-4)} \right], \tag{8.56}$$

which are finite provided that $\alpha \neq 2/3$ and $\beta \neq 2/3$. For the system of Equations (8.44)-(8.45), we have two straight lines in the phase plane where $\mathrm{d}U/\mathrm{d}z$ is infinite, at $C = \alpha$ and $C = \beta$. These kind of lines have previously been called *walls of singularities* for hyperbolic models related to chemotactic and haptotactic invasion [51]. For a smooth solution trajectory to exist between two equilibrium points on opposite sides of the wall of singularities, we require that the trajectory passes through the wall of singularities. This implies that the solution trajectory must pass through the wall of singularities at the special points, (α, U_α) and (β, U_β), known as holes in the wall [51, 71]. Otherwise, a smooth heteroclinic orbit between $(1, 0)$ and $(0, 0)$ cannot exist, as $\lim_{C \to \alpha} |U| \to \infty$ and $\lim_{C \to \beta} |U| \to \infty$. As U_α and U_β are real valued and the limits in Equations (20)-(21) are finite, the holes in the wall always exist for Fisher kinetics.

Ferracuti *et al.* [14] prove that travelling wave solutions exist for reaction-diffusion equations with positive-negative-positive $F(C)$ and Fisher kinetics, however travelling wave

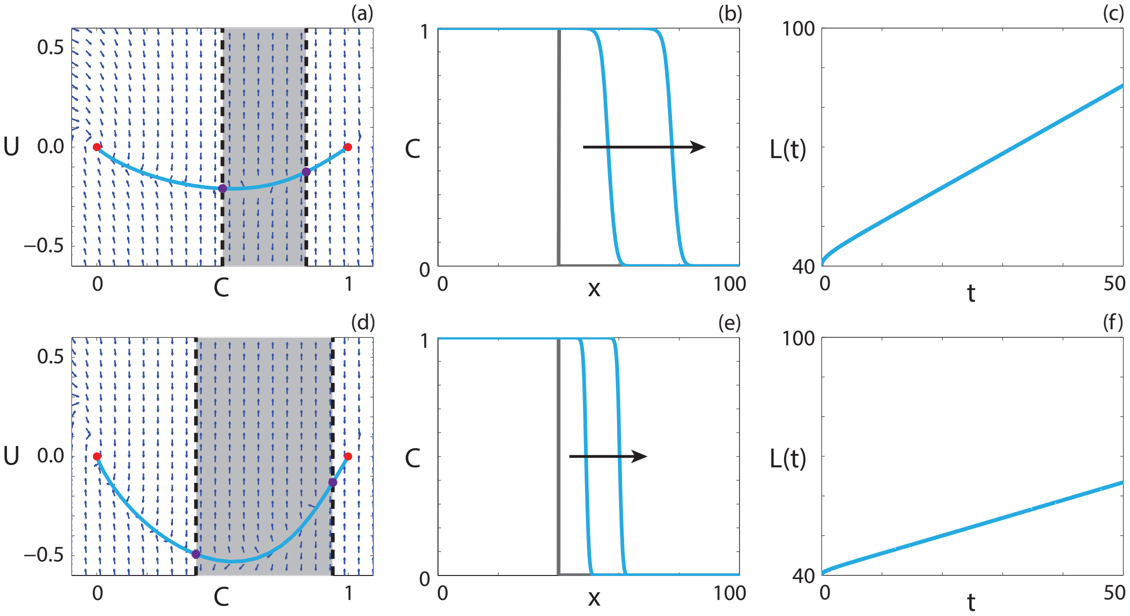

Figure 8.16: Travelling wave behaviour for Equation (8.41) **with positive-negative-positive** $F(C)$ **(Case 2.3)**. (a), (d) Phase plane for the system (8.44)-(8.45) with the numerical solution of Equation (8.41), in (C, U) co-ordinates, superimposed. The grey region corresponds to values of C where $F(C) < 0$. The dashed black lines denote a wall of singularities. Red circles correspond to equilibrium points and purple circles correspond to holes in the wall. (b), (e) Numerical solution of Equation (8.41) at $t = 100$ and $t = 200$. The grey lines indicate the initial condition and the arrows indicate the direction of increasing time. (c), (f) The time evolution of the position of the leading edge of the travelling wave solution, $L(t)$. All results are obtained using $P_d^i = P_d^g = 0$, $\delta x = 0.01$, $\delta t = 0.01$, $\epsilon = 10^{-6}$ and (a)-(c) $P_m^i = 0.5$, $P_m^g = 0.1$, $P_p^i = P_p^g = 0.75$, $v = 0.864$; (d)-(f) $P_m^i = 0.1$, $P_m^g = 0.01$, $P_p^i = P_p^g = 1.0$, $v = 0.448$.

profiles arising from the PDE are not presented. An upper bound on the minimum wave speed is stated as [14]

$$v^* = \max\{v_1, v_2, v_3\}, \tag{8.57}$$

where

$$F'(0)R(0) + F(0)R'(0) \leq \frac{v_1^2}{4} \leq \sup_{C\in(0,\alpha]} \left[\frac{F(C)R(C)}{C}\right], \tag{8.58}$$

$$F'(\beta)R(\beta) + F(\beta)R'(\beta) \leq \frac{v_2^2}{4} \leq \sup_{C\in[\alpha,\beta)} \left[\frac{F(C)R(C)}{C-\beta}\right], \tag{8.59}$$

$$F'(\beta)R(\beta) + F(\beta)R'(\beta) \leq \frac{v_3^2}{4} \leq \sup_{C\in(\beta,1]} \left[\frac{F(C)R(C)}{C-\beta}\right], \tag{8.60}$$

and the prime denotes ordinary differentiation with respect to C. Numerical solutions of Equation (8.41) with $P_m^i > 4P_m^g$ are presented in Figure 8.16. We superimpose the numerical solution of Equation (8.41) in (C, U) co-ordinates on the phase plane for the system (8.44)-(8.45) in Figures 8.16(a) and 8.16(d). The numerical solution forms a heteroclinic orbit between $(1, 0)$ and $(0, 0)$ in both cases, and passes through the holes in the wall of singularities, denoted using purple circles. Continuum models with negative diffusivity and no source terms have been relatively well studied, and exhibit shock behaviour across

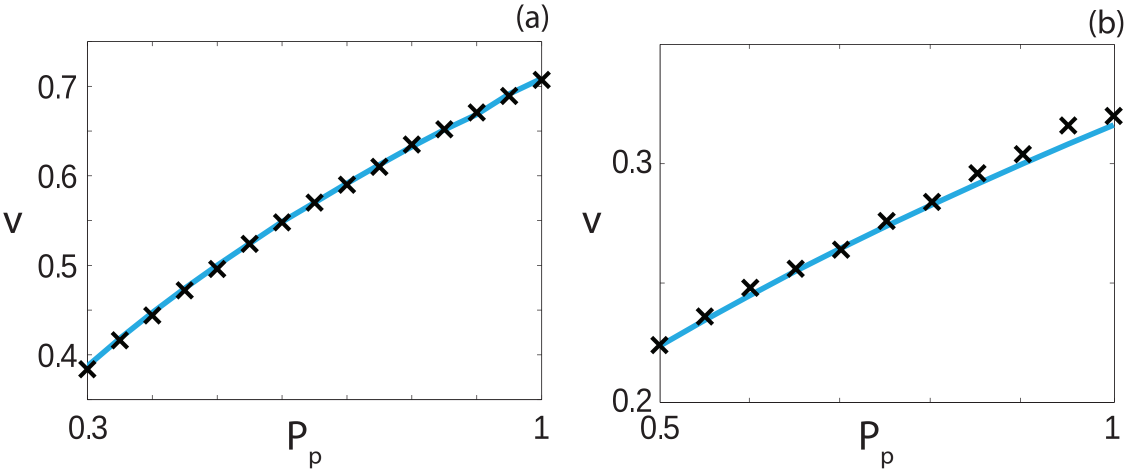

Figure 8.17: Wave speed comparison for Case 2.3. Comparison of the observed wave speed and the upper bound of the minimum wave speed obtained from the relationship in Equation (8.57) for a suite of P_p values. All results are obtained using $P_d^i = P_d^g = 0$, $\delta x = 0.01$, $\delta t = 0.01$, $\epsilon = 10^{-6}$, and (a) $P_m^i = 0.25$, $P_m^g = 0.05$; (b) $P_m^i = 0.05$, $P_m^g = 0.005$, and the Heaviside initial condition. In all cases Equation (8.58) provided the estimate of the minimum wave speed. Crosses correspond to the observed wave speed and the solid line corresponds to the upper bound of the wave speed.

the region of negative diffusion [39, 72]. Interestingly, our solution does not include a shock and is instead smooth through the region of negative diffusion.

Numerical solutions of Equation (8.41) are presented in Figures 8.16(b) and 8.16(e), which appear to take the form of travelling waves. The observed wave speeds, $v = 0.864$ and $v = 0.456$, in Figure 8.16(c) and Figure 8.16(f), respectively, are well approximated by the upper bound on the minimum wave speed presented by Ferracuti *et al.* [14]. The bound provides values for the minimum wave speed of $v^* = 0.866$ and $v^* = 0.447$, respectively. We might expect that the observed wave speeds correspond to the minimum wave speeds since the initial conditions for the numerical solutions are given by a Heaviside initial condition.

The observed wave speed, obtained from the long time numerical solutions of Equation (8.41), and the upper bound on the minimum wave speed, given by Equation (8.57), are shown in Figure 8.17 for a suite of P_p values and two positive-negative-positive $F(C)$ functions. In all cases the bound provides an accurate prediction of the observed wave speed.

Capacity-degenerate positive-negative nonlinear diffusivity function

For the special case where $P_m^g = 0$, $F(1) = 0$. As $F(C)$ is degenerate at $C = 1$, it is intuitive to expect there could be sharp-fronted travelling wave solutions, with the sharp front near $C = 1$, similar to the results in Figure 8.15. However, unlike for the parameter regimes in Figure 8.15, we have an interval $1/3 < C < 1$ where $F(C) < 0$. To determine whether this negative diffusivity influences the presence of sharp fronts, we follow the approach of Maini *et al.* [42], who show that the existence of travelling waves

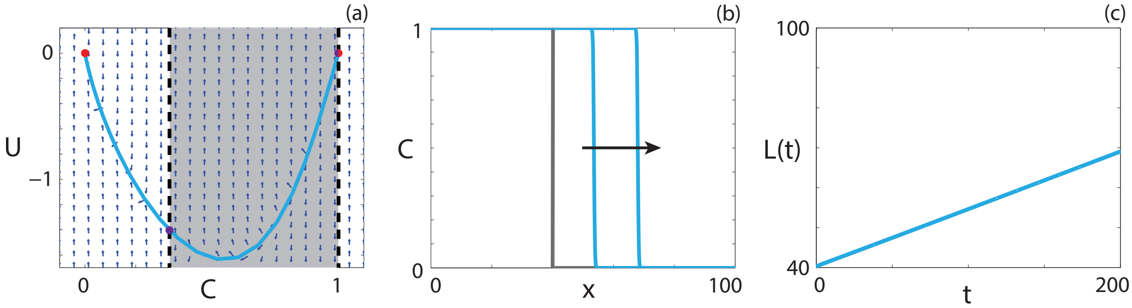

Figure 8.18: Travelling wave behaviour for Equation (8.41) **with capacity-degenerate positive-negative** $F(C)$ **(Case 2.4).** (a) Phase plane for the system (8.44)-(8.45) with the numerical solution of Equation (8.41), in (C, U) co-ordinates, superimposed. The grey region corresponds to values of C where $F(C) < 0$. The dashed black lines denote a wall of singularities. Red circles correspond to equilibrium points and purple circles correspond to holes in the wall. (b) Numerical solution of Equation (8.41) at $t = 100$ and $t = 200$. The grey lines indicate the initial condition and the arrow indicates the direction of increasing time. (c) The time evolution of the position of the leading edge of the travelling wave solution, $L(t)$. All results are obtained using $P_m^i = 0.01$, $P_m^g = 0$, $P_p^i = P_p^g = 1.0$, $P_d^i = P_d^g = 0$, $\delta x = 0.01$, $\delta t = 0.01$, $\epsilon = 10^{-6}$, $v = 0.1433$.

for reaction-diffusion equations with capacity-degenerate positive-negative $F(C)$ can be determined by considering the existence of travelling waves for

$$\frac{\partial C}{\partial \hat{t}} = \frac{\partial^2 C}{\partial x^2} + F(C)R(C), \ \hat{t} \geq 0. \tag{8.61}$$

The restriction on $\hat{t}$ implies that $F(C) > 0$. As $F(C) < 0$ for $1/3 < C < 1$, Equation (8.61) is only equivalent to Equation (8.41) for $0 \leq C \leq 1/3$. For $1/3 \leq C \leq 1$, Equation (8.41) is equivalent to

$$\frac{\partial C}{\partial \hat{t}} = \frac{\partial^2 C}{\partial x^2} + \hat{F}(C)\hat{R}(C), \ \hat{t} \geq 0, \tag{8.62}$$

where $\hat{F}(C) = -F(1-C)$ and $\hat{R}(C) = R(1-C)$ [42]. Equations (8.61)-(8.62) have minimum travelling wave speeds v_0^* and v_1^*, respectively. Maini *et al.* [42] prove that sharp fronts in the travelling wave near $C = 1$ only exist if $F(1) = 0$ and $v_1^* < v_0^*$. The first condition is obviously satisfied, while the second can be determined through linear analysis of Equations (8.61)-(8.62) in travelling wave co-ordinates. Both equations have minimum wave speed conditions, $v_0^* = v_1^* = 2\sqrt{\lambda D_i}$, to obtain physically-relevant heteroclinic orbits, and hence travelling wave solutions with a sharp region near $C = 1$ do not exist.

Travelling wave behaviour for a parameter regime with $F(1) = 0$ is shown in Figure 8.18. The equilibrium point at $(1, 0)$ is also a hole in the wall. The solution trajectory forms a heteroclinic orbit between $(1, 0)$ and $(0, 0)$, and moves through the region of C where $F(C) < 0$. Although $F(1) = 0$, we do not observe a solution trajectory corresponding to a sharp front, as we observed in Figure 8.15, where $F(0) = 0$. This result is consistent with the analysis of Maini *et al.* [42]. The numerical solution of Equation (8.18), presented in Figure 8.18(b), has a relatively steep front but is not sharp near $C = 1$. As $L(t)$, presented in Figure 8.18(c), becomes linear as t increases and the waveform in Figure

8.18(b) are consistent, the numerical solution of Equation (8.41) with $F(1) = 0$ appears to form a classic travelling wave.

8.6.3 Equal motility rates, equal proliferation rates, equal death rates.

For $P_d^i = P_d^g > 0$, with $P_m^i = P_m^g$ and $P_p^i = P_p^g$, there are no competitive or co-operative mechanisms. In this case, Equation (8.2) becomes

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} + \lambda C(1-C) - KC, \tag{8.63}$$

where $D = D_i = D_g$, $\lambda = \lambda_i = \lambda_g$ and $K = K_i = K_g$. The corresponding ODE in travelling wave co-ordinates is

$$D\frac{\mathrm{d}^2 C}{\mathrm{d}z^2} + v\frac{\mathrm{d}C}{\mathrm{d}z} + \lambda C(1-C) - KC = 0, \ -\infty < z < \infty, \tag{8.64}$$

and, with $U = \mathrm{d}C/\mathrm{d}z$, we obtain

$$\frac{\mathrm{d}C}{\mathrm{d}z} = U, \tag{8.65}$$

$$\frac{\mathrm{d}U}{\mathrm{d}z} = -\frac{v}{D}U - \frac{\lambda}{D}C(1-C) + \frac{K}{D}C. \tag{8.66}$$

The source term in Equation (8.63) is non-positive for all relevant C values if $K > \lambda$, and negative for $C > (\lambda - K)/\lambda$ otherwise. Hence the population will never reach the original carrying capacity of unity. The new carrying capacity can be determined by considering the zeros of the source term, which occur at $C = 0$ and $C = (\lambda - K)/\lambda$. Introducing a new variable, $\overline{C} = \lambda C/(\lambda - K)$, and rewriting Equation (8.63) in terms of $\overline{C}$ we obtain

$$\frac{\partial \overline{C}}{\partial t} = D\frac{\partial^2 \overline{C}}{\partial x^2} + (\lambda - K)\overline{C}(1-\overline{C}). \tag{8.67}$$

Equation (8.67) is the Fisher-Kolmogorov equation in terms of the new variable, $\overline{C}$, with an intrinsic growth rate $(\lambda - K)$. As such, the analysis performed for Case 1 is applicable here and we obtain information about the stability of the equilibrium points, as well as the minimum wave speed required for physically meaningful travelling wave solutions. If $\lambda > K$, the minimum wave speed is $v^* = 2\sqrt{(\lambda - K)D}$. If $K > \lambda$, there is only one physically relevant equilibrium point, $C = 0$, and hence the population will tend to extinction and travelling wave solutions do not exist.

Travelling wave behaviour for two parameter regimes with $\lambda > K$ are illustrated in Figure 8.19. The phase plane for $K = 0.1$, presented in Figure 8.19(a), displays qualitatively similar behaviour to the phase plane for $K = 0.2$, presented in Figure 8.19(d). Unsurprisingly, the unstable equilibrium point moves closer to zero as K approaches λ. The numerical solutions of Equation (8.63), presented in Figures 8.19(b) and 8.19(e), have significantly different densities behind the wave fronts. However, both travelling wave fronts represent heteroclinic orbits between $(C, U) = ((\lambda - K)/\lambda, 0)$ and $(C, U) = (0, 0)$.

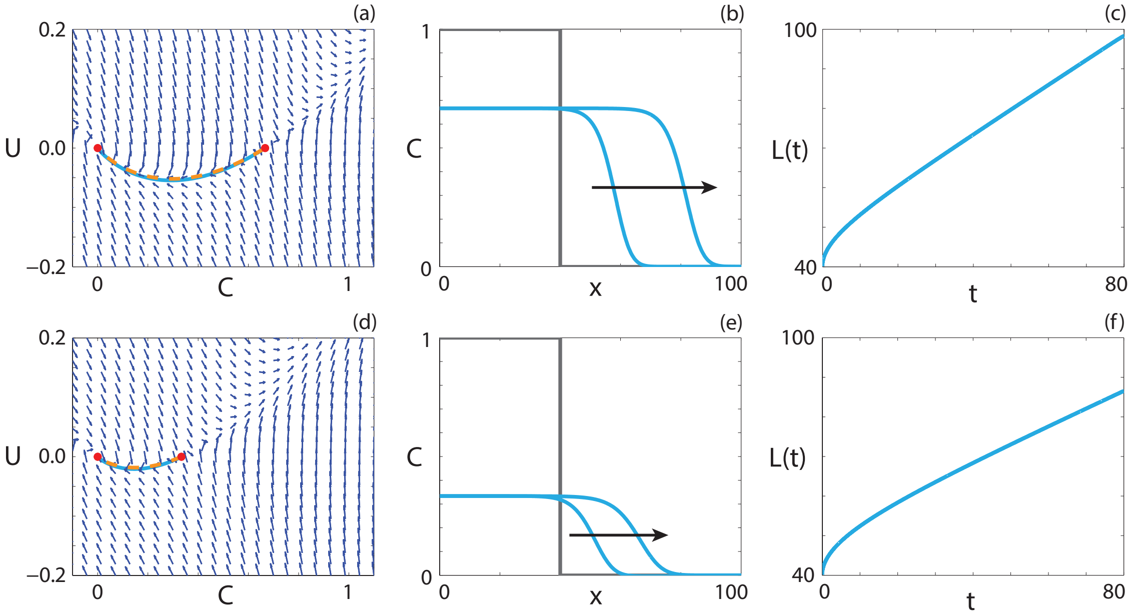

Figure 8.19: Travelling wave behaviour for the scaled Fisher-Kolmogorov model (Case 3). (a), (d) Phase plane for the system (8.65)-(8.66). Red circles denote equilibrium points. The numerical solutions of Equations (8.63) (cyan, solid) and (8.64) (orange, dashed), in (C, U) co-ordinates, are superimposed. (b), (e) Numerical solution of Equation (8.63) at $t = 40$ and $t = 80$. The grey lines indicate the initial condition and the arrows indicate the direction of increasing time. (c), (f) The time evolution of the position of the leading edge of the travelling wave solution. All results are obtained using $P_m^i = P_m^g = 1$, $P_p^i = P_p^g = 0.3$, $\delta x = 0.1$, $\delta t = 0.01$, $\epsilon = 10^{-6}$ and (a)-(c) $P_d^g = P_d^i = 0.1$, $v = 0.615$; (d)-(f) $P_d^g = P_d^i = 0.2$, $v = 0.445$.

Interestingly, the two travelling wave fronts have approximately the same support, even though the waveform is significantly shallower for the case with $K = 0.2$. Results in Figures 8.19(c) and (f) show that both solutions approach travelling waves as t increases, and that increasing the death rate reduces the wave speed.

8.6.4 Different motility rates, equal proliferation rates, equal death rates.

For $P_p^i = P_p^g$, $P_d^i = P_d^g$ and $P_m^i \neq P_m^g$, the co-operative or competitive mechanism arises due to the difference in the rate of motility. In this case, the governing PDE is

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x}\left(F(C)\frac{\partial C}{\partial x}\right) + \lambda C(1 - C) - KC, \tag{8.68}$$

where $F(C) = D_i(1 - 4C + 3C^2) + D_g(4C - 3C^2)$, $\lambda = \lambda_i = \lambda_g$ and $K = K_i = K_g$. Equation (8.68) corresponds to

$$v\frac{\mathrm{d}C}{\mathrm{d}z} + F(C)\frac{\mathrm{d}^2C}{\mathrm{d}z^2} + (D_i - D_g)(6C - 4)\left(\frac{\mathrm{d}C}{\mathrm{d}z}\right)^2 + \lambda C(1 - C) - KC = 0, \; -\infty < z < \infty, \tag{8.69}$$

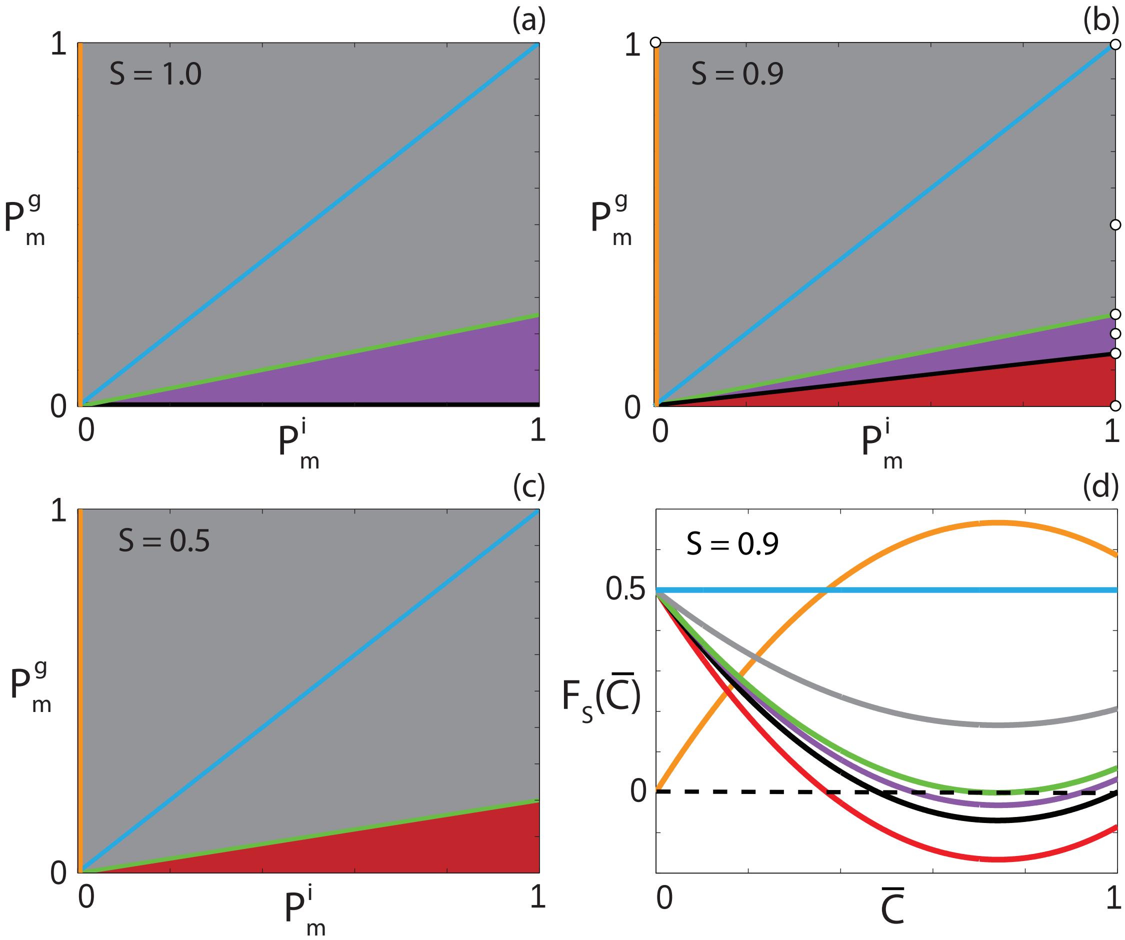

Figure 8.20: Classification of $F_s(\overline{C})$ for different carrying capacity densities. (a)-(c) Type of $F_s(\overline{C})$ function for $0 \leq \overline{C} \leq 1$ for the parameter space $P_m^i \in [0,1]$ and $P_m^g \in [0,1]$ with (a) $S = 1.0$; (b) $S = 0.9$; (c) $S = 0.5$. Grey regions correspond to parameter pairs that result in strictly positive $F_s(\overline{C})$, purple regions correspond to parameter pairs that result in positive-negative-positive $F_s(\overline{C})$ and red regions correspond to parameter pairs that result in positive-negative $F_s(\overline{C})$. Cyan, orange and black lines correspond to constant, extinction-degenerate non-negative and capacity-degenerate positive-negative $F_s(\overline{C})$ curves, respectively. (d) Example $F_s(\overline{C})$ for each region in (b). The white circles in (b) denote the parameter pairs used to generate the curves in (d).

in travelling wave co-ordinates, and with the substitution $U = \mathrm{d}C/\mathrm{d}z$, we obtain

$$\frac{\mathrm{d}C}{\mathrm{d}z} = U, \tag{8.70}$$

$$\frac{\mathrm{d}U}{\mathrm{d}z} = \frac{-vU - (D_i - D_g)(6C-4)U^2 - \lambda C(1-C) + KC}{F(C)}. \tag{8.71}$$

The system of Equations (8.70)-(8.71) has equilibrium points $(C,U) = (0,0)$ and $(C,U) = (S,0)$, where $S = (\lambda - K)/\lambda$. Increasing K causes a decrease in the carrying capacity, S. If $K > \lambda$, the non-zero equilibrium point occurs at a negative C value and hence only one physically relevant equilibrium point exists, implying that the population will become extinct. Hence we only investigate the behaviour of parameter regimes where $\lambda > K$.

We introduce the variable $\overline{C} = C/S$ such that the agent density is scaled by the carrying capacity and, subsequently, the zeros of the source term occur at $\overline{C} = 0$ and $\overline{C} = 1$. This approach allows us to repeat the analysis for Case 2 with a different $F(C)$. We transform Equation (8.68) in terms of $\overline{C}$ to obtain

$$\frac{\partial \overline{C}}{\partial t} = \frac{\partial}{\partial x}\left(F(S\overline{C})\frac{\partial \overline{C}}{\partial x}\right) + (\lambda - K)\overline{C}(1-\overline{C}). \tag{8.72}$$

If we define $\overline{U} = \mathrm{d}\overline{C}/\mathrm{d}z$, Equation (8.72) corresponds to the system,

$$\frac{\mathrm{d}\overline{C}}{\mathrm{d}z} = \overline{U}, \tag{8.73}$$

$$\frac{\mathrm{d}\overline{U}}{\mathrm{d}z} = \frac{-v\overline{U} - (D_i - D_g)S(6S\overline{C}-4)\overline{U}^2 - (\lambda - K)\overline{C}(1-\overline{C})}{F(S\overline{C})}, \quad -\infty < z < \infty. \tag{8.74}$$

The transformed nonlinear diffusivity function

$$F_s(\overline{C}) = F(S\overline{C}) = D_i(1 - 4S\overline{C} + 3(S\overline{C})^2) + D_g(4S\overline{C} - 3(S\overline{C})^2), \tag{8.75}$$

has different properties depending on S, D_i and D_g. To highlight this, Figure 8.20 shows the (P_m^i, P_m^g) parameter space for three different S values and the qualitative behaviour of the corresponding $F_s(\overline{C})$ function. For $S = 1$, presented in Figure 8.20(a), we recover the nonlinear diffusivity function examined for Case 2, where for $P_m^i > 4P_m^g$, denoted in purple, there is an interval $\alpha < \overline{C} < \beta$, $\alpha < \beta < 1$, where $F_s(\overline{C}) < 0$. Decreasing S to 0.9, presented in Figure 8.20(b), we observe that the purple region again occurs for $P_m^i > 4P_m^g$. However, if $P_m^g < 0.145P_m^i$, highlighted in red, $F_s(\overline{C}) < 0$ for $\omega < \overline{C} \leq 1$, and hence $F_s(\overline{C})$ has only one zero in $0 \leq \overline{C} \leq 1$. This type of nonlinear diffusivity function is not observed for Case 2 and we refer to it as *positive-negative*. Specifically, this behaviour occurs when $(16 - (6S-4)^2)P_m^g < (4 - (6S-4)^2)P_m^i$ and $P_m^i > 4P_m^g$. Furthermore, this implies that for $S < 2/3$ there are no (P_m^i, P_m^g) values that correspond to positive-negative-positive $F_s(\overline{C})$. An example of this (P_m^i, P_m^g) parameter space is shown in Figure 8.20(c). For $S < 1/3$, $F_s(\overline{C}) \geq 0$ for $0 \leq \overline{C} \leq 1$. Parameter pairs that correspond to extinction-degenerate non-negative $F_s(\overline{C})$ (orange) and constant $F_s(\overline{C})$ (cyan) exist for all S values.

Strictly positive nonlinear diffusivity function

If $F_s(\overline{C}) > 0$ for $0 \leq \overline{C} \leq 1$, linear analysis leads to a minimum wave speed $v^* = 2\sqrt{(\lambda - K)D_i}$ for the equilibrium point at $(0, 0)$ to be stable. Intuitively, the wave speed is positive provided that the rate of birth is greater than the rate of death.

Travelling wave behaviour for Equation (8.68) with strictly positive $F_s(\overline{C})$ and $S = 0.5$ is shown in Figures 8.21(a)-(c). Similar to the strictly positive $F(C)$ considered in Case 2.1, there is a heteroclinic orbit between the two equilibrium points, implying that the solution of Equation (8.68) forms a travelling wave. Intuitively, introducing agent death

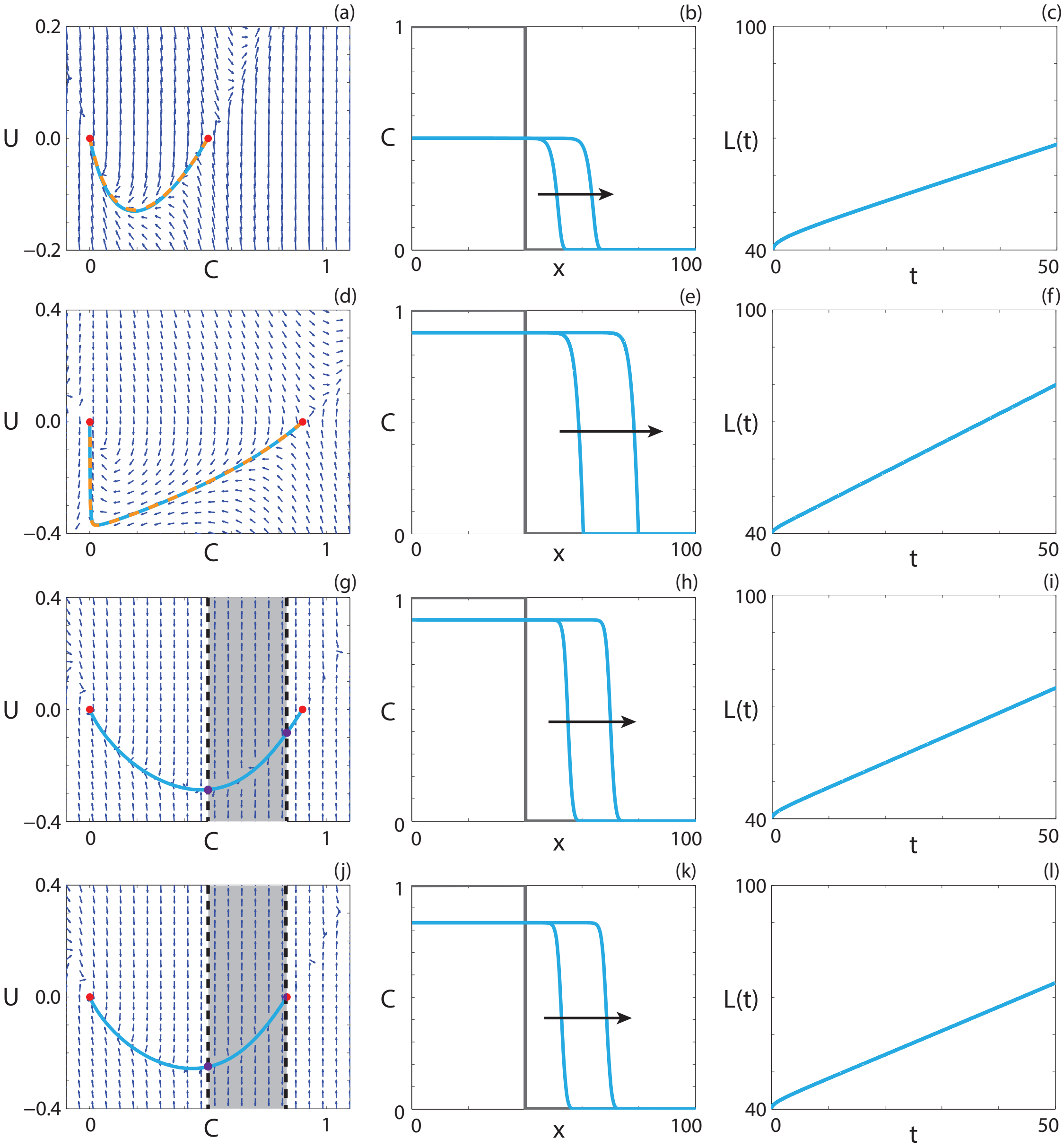

Figure 8.21: Travelling wave behaviour for Equation (8.68) **for various** $F_s(\overline{C})$. We consider (a)-(c) strictly positive $F_s(\overline{C})$ (Case 4.1), (d)-(f) extinction-degenerate non-negative $F_s(\overline{C})$ (Case 4.2), (g)-(i) positive-negative-positive $F(S\overline{C})$ (Case 4.3), and (j)-(l) capacity-degenerate positive-negative $F_s(\overline{C})$ (Case 4.4). (a), (d), (g), (j) Phase plane for the system (8.70)-(8.71) with the numerical solution of Equations (8.68) (cyan, solid) and (8.69) (orange, dashed), in (C, U) co-ordinates, superimposed. The grey regions correspond to values of $\overline{C}$ where $F_s(\overline{C}) < 0$. The dashed black lines denote a wall of singularities. Red circles correspond to equilibrium points and purple circles correspond to holes in the wall. (b), (e), (h), (k) Numerical solution of Equation (8.68) at $t = 25$ and $t = 50$. The grey lines indicate the initial condition and the arrows indicate the direction of increasing time. (c), (f), (i), (l) The time evolution of the position of the leading edge of the travelling wave solution. Parameters used are (a)-(c) $P_m^i = 0.25$, $P_m^g = 0.5$, $P_p^i = P_p^g = 1$, $P_d^i = P_d^g = 0.5$, $v = 0.504$; (d)-(f) $P_m^i = 0$, $P_m^g = 1$, $P_p^i = P_p^g = 1$, $P_d^i = P_d^g = 0.1$, $v = 0.777$; (g)-(i) $P_m^i = 0.25$, $P_m^g = 0.05$, $P_p^i = P_p^g = 1$, $P_d^i = P_d^g = 0.1$, $v = 0.672$; (j)-(l) $P_m^i = 0.25$, $P_m^g = 0.05$, $P_p^i = P_p^g = 1$, $P_d^i = P_d^g = 1/6$, $v = 0.648$. All results are obtained using $\delta x = 0.1$, $\delta t = 0.01$, $\epsilon = 10^{-6}$.

has the effect of reducing the density sufficiently far behind the wave front, and we observe that the non-zero equilibrium point now occurs at $C = S \leq 1$.

Extinction-degenerate non-negative nonlinear diffusivity function

If $F_s(0) = 0$, $F_s(\overline{C}) > 0$ for $0 < \overline{C} \leq 1$, and $R(0) = 0$, a stretching transformation is applied to remove the singularity in Equations (8.73)-(8.74), giving

$$\frac{\mathrm{d}\overline{C}}{\mathrm{d}\zeta} = D_g S(4\overline{C} - 3S\overline{C}^2)\overline{U}, \tag{8.76}$$

$$\frac{\mathrm{d}\overline{U}}{\mathrm{d}\zeta} = -v\overline{U} + D_g S(6S\overline{C} - 4)\overline{U}^2 - (\lambda - K)\overline{C}(1 - \overline{C}), \ \zeta \geq 0, \tag{8.77}$$

as $D_i = 0$. There are now three equilibrium points: $(\overline{C}, \overline{U}) = (0,0)$; $(\overline{C}, \overline{U}) = (1,0)$; and $(\overline{C}, \overline{U}) = (0, -v/4SD_g)$. As for Case 2.2, the saddle-saddle connection between $(1,0)$ and $(0, -v/4SD_g)$ only occurs for a unique wave speed, v^*, which implies that sharp-fronted solutions exist only for the minimum wave speed.

Sharp-fronted travelling wave solutions of Equation (8.68) with extinction-degenerate non-negative $F_s(\overline{C})$ and $S = 0.9$ are shown in Figures 8.21(d)-(f). Introducing agent death does not change the qualitative behaviour compared to the corresponding case with $K = 0$ (Case 2.2). Specifically, $\mathrm{d}C/\mathrm{d}z$ approaches $C = 0$ with a non-zero value and hence the wave front is sharp near $C = 0$. Again, the density behind the wave front decreases such that $C = S$, corresponding to the non-zero equilibrium point.

Positive-negative-positive nonlinear diffusivity function

For positive-negative-positive $F_s(\overline{C})$, the analysis in Case 2.3 holds provided that $\lambda > K$. Specifically, the minimum wave speed condition proved by Ferracuti *et al.* [14] implies that real-valued holes in the wall will be present for the scaled Fisher-Kolmogorov equation with $\lambda > K$. In turn, this suggests that the smooth travelling wave solutions passing through the region of negative diffusivity observed for Case 2.3 will be present with non-zero K. Travelling wave behaviour for Equation (8.68) with positive-negative-positive $F_s(\overline{C})$ is demonstrated in Figures 8.21(g)-(i). The travelling wave solution behaviour is similar to the behaviour in the corresponding case with $K = 0$ (Case 2.3), with the exception of the reduced carrying capacity.

Capacity-degenerate positive-negative nonlinear diffusivity function

The capacity-degenerate positive-negative diffusivity case, where $F_s(1) = R(S) = 0$, $F_s(\overline{C}) < 0$ for $\omega < \overline{C} < 1$ and $F_s(\overline{C}) \geq 0$ otherwise, might be thought to lead to travelling wave solutions with a sharp front near the carrying capacity density [42]. Similar to the approach for Case 2.4, we consider the conditions proposed by Maini *et al.* [42]. Again, we satisfy the condition that $F_s(1) = 0$. However, the minimum wave speed for the transformation of Equation (8.41) in $0 \leq \overline{C} < \omega$ is the same as the minimum wave

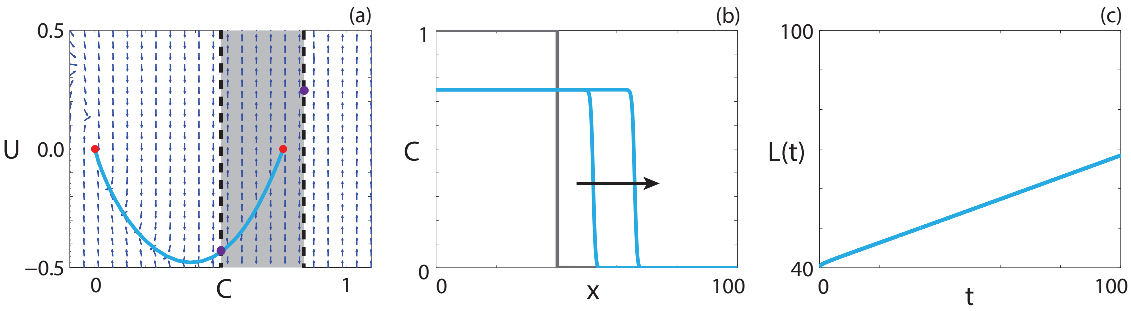

Figure 8.22: Travelling wave behaviour for Equation (8.68) **with positive-negative** $F_s(\overline{C})$ **(Case 4.5)**. (a) Phase plane for the system (8.70)-(8.71) with the numerical solution of Equation (8.68), in (C, U) co-ordinates, superimposed. The grey region corresponds to values of $\overline{C}$ where $F_s(\overline{C}) < 0$. The dashed black lines denote a wall of singularities. Red circles correspond to equilibrium points and purple circles correspond to holes in the wall. (b) Numerical solution of Equation (8.68) at $t = 50$ and $t = 100$. The grey lines indicate the initial condition and the arrow indicates the direction of increasing time. (c) The time evolution of the position of the leading edge of the travelling wave solution. All results are obtained using $P_m^i = 0.05$, $P_m^g = 0.01$, $P_p^i = P_p^g = 1.0$, $P_d^i = P_d^g = 0.25$, $\delta x = 0.1$, $\delta t = 0.01$, $\epsilon = 10^{-6}$, $v = 0.2760$.

speed for the transformation of Equation (8.41) in $\omega < \overline{C} \leq 1$. As such, we do not expect that Equation (8.68) will approach a travelling wave solution with a sharp front near $\overline{C} = 1$. We present travelling wave behaviour for Equation (8.68) with capacity-degenerate positive-negative $F_s(\overline{C})$ in Figures 8.21(j)-(l) and, as anticipated, observe that the travelling wave solution is a classic front.

Positive-negative nonlinear diffusivity function

The positive-negative case, where $F_s(\overline{C}) > 0$ for $0 \leq \overline{C} < \omega$ and $F_s(\overline{C}) < 0$ for $\omega < \overline{C} \leq 1$, cannot occur with $K = 0$. It is instructive to examine whether stable travelling wave solutions of Equation (8.68) exist, as the non-zero equilibrium point now occurs in the region where $F_s(\overline{C}) < 0$. If we perform standard linear analysis on Equations (8.70)-(8.71), the Jacobian at $(S, 0)$ has eigenvalues $\xi = (-v \pm \sqrt{v^2 + 4F(S)(\lambda(2S-1)+K)})/2F(S)$, which implies that the equilibrium point is an unstable node provided $v > 2\sqrt{-F(S)(\lambda(2S-1)+K)}$. The negative sign is present as $F(S) < 0$ for positive-negative $F_s(\overline{C})$. The Jacobian at $(0, 0)$ has eigenvalues $\xi = (-v \pm \sqrt{v^2 - 4D_i(\lambda - K)})/2D_i$, which is a stable node provided that $v > 2\sqrt{(\lambda - K)D_i}$. While there are infinitely many solution trajectories out of the unstable node, we require that the solution trajectory passes through the hole in the wall, and hence there is a single solution trajectory that forms a heteroclinic orbit.

Travelling wave behaviour for Equation (8.68) with positive-negative $F_s(\overline{C})$ is shown in Figure 8.22. The numerical solution of Equation (8.68), in (C, U) co-ordinates, passes through the wall of singularities where Equation (8.71) is finite and forms a heteroclinic orbit between $(S, 0)$ and $(0, 0)$. The travelling wave front is of classic type, a result predicted by the analysis performed by Maini *et al.* [42] as $F_s(0) \neq 0$ and $F_s(1) \neq 0$.

8.6.5 Equal motility rates, no grouped agent death.

Results in Figure 8.1 (Main Document) indicate that restricting death events to isolated agents significantly change the behaviour of the agent population. This represents a co-operative mechanism, as there is a benefit to being in close proximity to another agent. In the case where $P_d^i \neq 0$ and $P_d^g = 0$, the source term can be expressed as an Allee effect [67]

$$R(C) = rC(1-C)(C-A), \tag{8.78}$$

where

$$r = K_i - \lambda_i + \lambda_g, \tag{8.79}$$

is the intrinsic growth rate, and

$$A = \frac{K_i - \lambda_i}{K_i - \lambda_i + \lambda_g}, \tag{8.80}$$

is the Allee parameter. It follows that Equation (8.2) becomes

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} + (K_i - \lambda_i + \lambda_g)C(1-C)\left(C - \frac{K_i - \lambda_i}{K_i - \lambda_i + \lambda_g}\right). \tag{8.81}$$

If $K_i > \lambda_i$, $R(C)$ represents the strong Allee effect, $A > 0$ [67]. The strong Allee effect has bistable growth kinetics, namely, $R(C) < 0$ for $0 < C < A$ and $R(C) > 0$ for $A < C < 1$. For low densities there are significantly more isolated agents than grouped agents, which corresponds to negative growth if $K_i > \lambda_i$. If $\lambda_i > K_i$, and $\lambda_g > 0$, $R(C) > 0$ for $0 < C < 1$. There are two possibilities for this case: $r > 0$ and $r < 0$. If $r > 0$ then $A < 0$ and hence the growth rate is inhibited at low density, but remains positive, which corresponds to the weak Allee effect [67]. For the case where $r < 0$ and $R(C) > 0$, we obtain $A > 1$ for all parameter combinations. Interestingly, this implies that the growth rate is inhibited at high density, but remains positive. This situation does not correspond to either the weak or strong Allee effect, and we term this behaviour the reverse Allee effect. It is not possible to have a combination of parameters that results in $r < 0$ and $0 < A < 1$ as all of our parameters are non-negative. Representative source terms showing the three types of Allee effect are compared with a logistic source term in Figure 8.23.

For $P_m^i = P_m^g$, we have linear diffusion in Equation (8.81). Reaction-diffusion equations with linear diffusion and either weak or strong Allee kinetics have been well-studied [4, 15, 16, 26, 34, 41, 54, 67]. We briefly present results here and interpret these in the context of examining the long time travelling wave solution. For additional details we refer the reader to [67]. We look for solutions in the travelling wave co-ordinate $z = x - vt$. The existence of such solutions has been examined previously and requirements for the initial conditions to converge to a travelling wave solution have been found for both the case where $A < 0$ and where $0 < A < 1$ [15]. Transforming Equation (8.81) into travelling

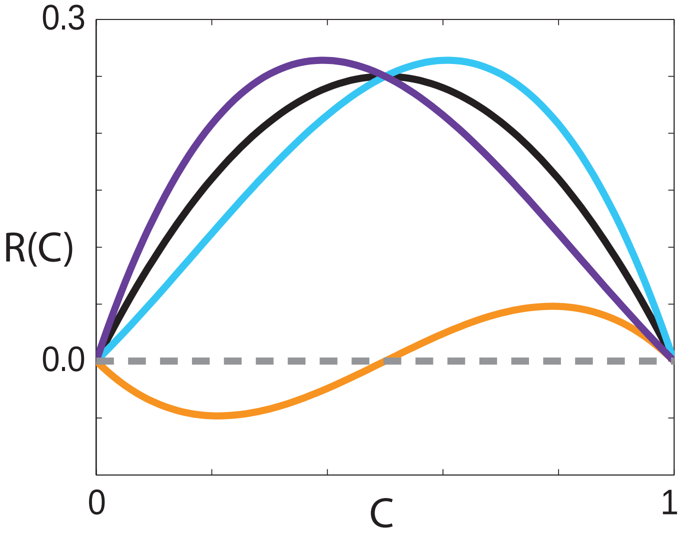

Figure 8.23: Comparison of source terms. $R(C)$ corresponding to the weak Allee effect with $r = 1$, $A = -0.5$ (cyan), strong Allee effect with $r = 1$, $A = 0.5$ (orange), reverse Allee effect with $r = -1$, $A = 1.5$ (purple) and logistic growth with $r = 1$ (black).

wave co-ordinates we obtain

$$D\frac{\mathrm{d}^2C}{\mathrm{d}z^2} + v\frac{\mathrm{d}C}{\mathrm{d}z} + (K_i - \lambda_i + \lambda_g)C(1-C)\left(C - \frac{K_i - \lambda_i}{K_i - \lambda_i + \lambda_g}\right) = 0, \quad -\infty < z < \infty, \tag{8.82}$$

where $D = D_g = D_i$. If $U = \mathrm{d}C/\mathrm{d}z$, Equation (8.82) can be expressed as

$$\frac{\mathrm{d}C}{\mathrm{d}z} = U, \tag{8.83}$$

$$\frac{\mathrm{d}U}{\mathrm{d}z} = \frac{-vU}{D} - \frac{(K_i - \lambda_i + \lambda_g)C(1-C)(C-A)}{D}. \tag{8.84}$$

This system of equations has three equilibrium points: $(C, U) = (0, 0)$; $(C, U) = (1, 0)$; and $(C, U) = (A, 0)$. We are only concerned with physically relevant equilibrium points, where $0 \leq C \leq 1$. If $A < 0$ or $A > 1$, there are only two physically relevant equilibrium points as the equilibrium point at $(C, U) = (A, 0)$ has no physical meaning. Performing standard linear stability analysis by examining the eigenvalues of Jacobian of the system, the characteristic equation at $(0, 0)$ has solutions $\xi = (-v \pm \sqrt{v^2 + 4(K_i - \lambda_i)D})/2D$, which implies that the equilibrium point is a stable node when $\lambda_i > K_i$, provided the wave speed satisfies $v > 2\sqrt{(\lambda_i - K_i)D}$. If the wave speed does not satisfy this condition then the equilibrium point is a stable spiral, which implies that the heteroclinic orbit enters non-physical values of C. If $K_i > \lambda_i$, $(0, 0)$ is a saddle point. The characteristic equation for the equilibrium point at $(1, 0)$ has solutions $\xi = (-v \pm \sqrt{v^2 + 4\lambda_g D})/2D$, which implies that the equilibrium point is a saddle point, as λ_g is non-negative. A heteroclinic orbit between $(1, 0)$ and $(0, 0)$ exists for a unique wave speed [26]. The equilibrium point at $(A, 0)$ has a characteristic equation with solutions $\xi = (-v \pm \sqrt{v^2 - 4\lambda_g AD})/2D$. As we are only concerned with physically realistic equilibrium points, that is, where $0 < A < 1$, the equilibrium point $(A, 0)$ will be a stable node provided that the minimum wave speed

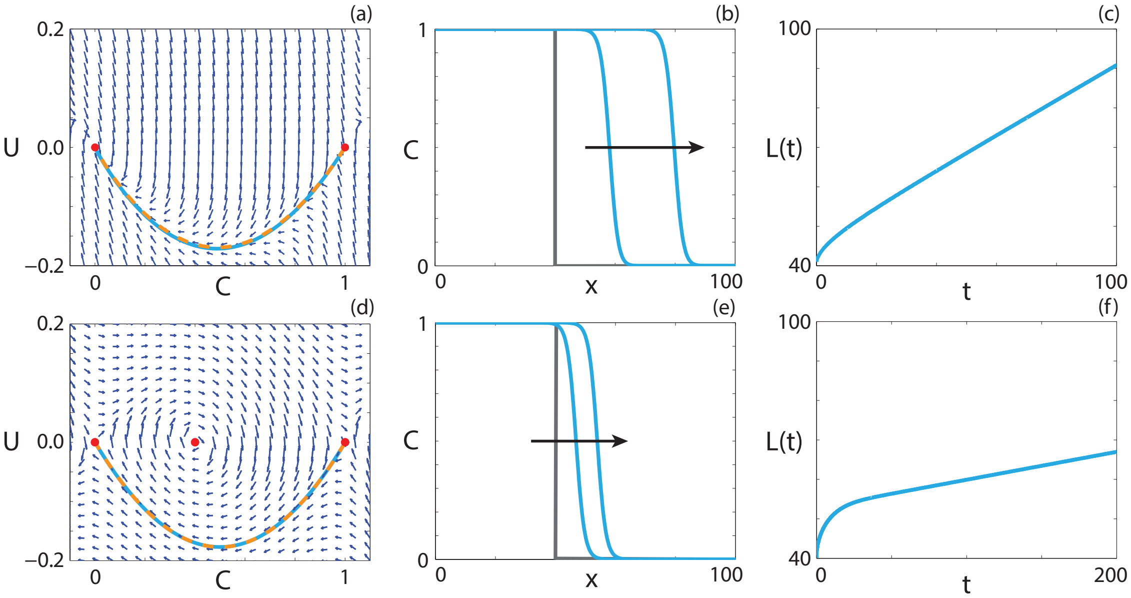

Figure 8.24: Travelling wave behaviour for the (a)-(c) weak Allee effect and the (d)-(f) strong Allee effect with a constant $F(C)$ (Case 5). (a), (d) Phase plane for the system (8.83)-(8.84) with the numerical solution of Equations (8.81) (cyan, solid) and (8.82) (orange, dashed), in (C, U) co-ordinates, superimposed. Red circles correspond to equilibrium points. (b), (e) Numerical solution of Equation (8.81) calculated at (b) $t = 50$ and $t = 100$; (e) $t = 100$ and $t = 200$. The grey lines indicate the initial condition and the arrows indicate the direction of increasing time. (c), (f) The time evolution of $L(t)$. All results are obtained with $\delta x = 0.1$, $\delta t = 0.01$, $\epsilon = 10^{-6}$, $P_d^g = 0$, (a)-(c) $P_m^i = 0.5$, $P_m^g = 0.5$, $P_p^i = 0.7$, $P_p^g = 0.4$, $P_d^i = 0.5$, $v = 0.44$; (d)-(f) $P_m^i = 1.0$, $P_m^g = 1.0$, $P_p^i = 0.3$, $P_p^g = 0.3$, $P_d^i = 0.5$, $v = 0.072$.

$v > 2\sqrt{\lambda_g AD}$ is satisfied, and a stable spiral otherwise. The spiral behaviour does not cause the solution trajectory to become non-physical and hence this wave speed condition is not required to obtain physically meaningful solutions.

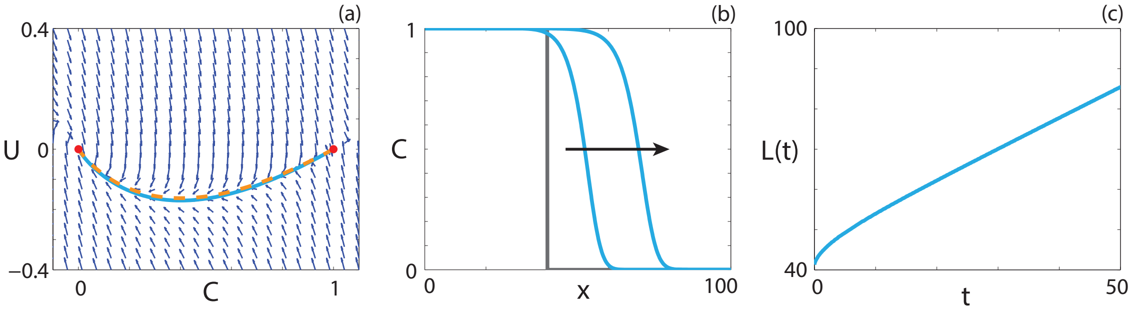

Figure 8.25: Travelling wave behaviour for Equation (8.81) **with the reverse Allee effect and constant** $F(C)$ **(Case 5).** (a) Phase plane for the system (8.83)-(8.84) with the numerical solution to Equations (8.81) (cyan, solid) and (8.82) (orange, dashed), in (C, U) co-ordinates, superimposed. Red circles correspond to equilibrium points. (b) Numerical solution to Equation (8.81) calculated at $t = 25$ and $t = 50$. The grey lines indicate the initial condition and the arrow indicates the direction of increasing time. (c) The time evolution of the position of the leading edge of the wave front. All results are obtained with $\delta x = 0.1$, $\delta t = 0.01$, $\epsilon = 10^{-6}$, $P_m^i = 1.0$, $P_m^g = 1.0$, $P_p^i = 0.6$, $P_p^g = 0.2$, $P_d^i = 0.3$, $P_d^g = 0$, $v = 0.756$.

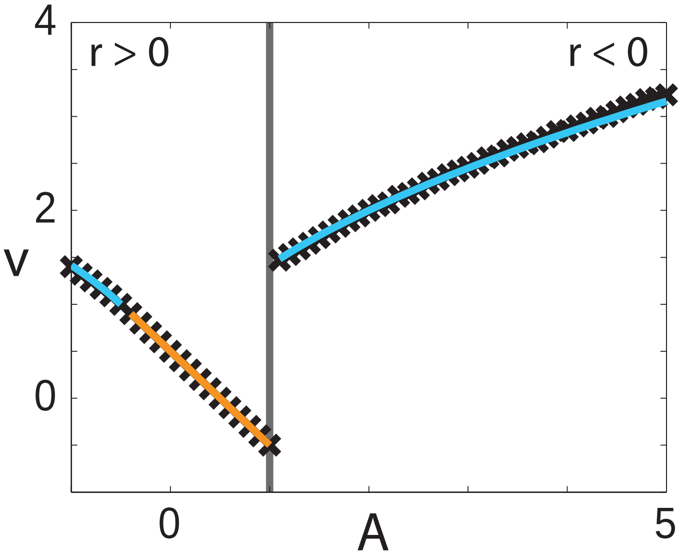

Figure 8.26: Wave speed comparison for Case 5. Comparison of observed wave speeds obtained from the numerical solution of Equation (8.81). For $A \leq 1$ results are obtained using $r = 1$, whereas for $A > 1$ results are obtained using $r = -1$. The cyan line corresponds to $v = 2\sqrt{-ArD}$ and the orange line corresponds to $v = 2\sqrt{rD}(1/2 - A)$.

Solutions that display travelling wave behaviour for the weak Allee effect are presented in Figures 8.24(a)-(c). There is a heteroclinic solution trajectory for Equations (8.83)-(8.84) between the two equilibrium points, and the numerical solution of Equation (8.81) matches the solution trajectory in (C, U) co-ordinates. The results in Figures 8.24(b)-(c) suggest that the numerical solution of Equation (8.81) with the weak Allee effect approaches a travelling wave solution. Since the source term for the reverse Allee effect is qualitatively similar to the weak Allee effect, we do not present solutions here. Solution behaviour for the reverse Allee effect can be found in Figure 8.25.

A travelling wave solution of Equation (8.81) in a parameter regime that results in a strong Allee effect is now considered. The phase plane for the system (8.83)-(8.84), presented in Figure 8.24(d), has three physically meaningful equilibrium points, and the equilibrium point at $(0, 0)$ is unstable, unlike in Figure 8.24(d). However, there is still a heteroclinic orbit between $(1, 0)$ and $(0, 0)$. Unlike for the weak Allee effect, the wave speed that admits this solution trajectory is unique [26]. The numerical solution of Equation (8.81) shows the solution approaches a travelling wave, although the approach is slower than for the weak Allee effect. This is intuitive, as the growth rate for the weak Allee effect is non-negative, while the strong Allee effect has regions of negative growth.

It is instructive to consider how v depends on A. We calculate the numerical solution of Equation (8.81) for a range of A values with $r = 1$ for $A \leq 1$, and $r = -1$ for $A > 1$, and use the numerical solution to calculate v at sufficiently late time. We consider $r < 0$ for $A > 1$ because, due to our parameters being non-negative, $A < 1$ for $r > 0$. The minimum wave speed for the travelling wave solution is known for $A \leq 0$, namely, $v^* = 2\sqrt{-ArD}$

for $A \leq -1/2$ and $v^* = 2\sqrt{rD}(1/2 - A)$ for $-1/2 \leq A \leq 0$ [41]. For $0 < A < 1$ there is a unique wave speed, $v = \sqrt{2rD}(1/2 - A)$ [41]. Consequently, for $A > 1/2$, the population will tend to extinction because the travelling wave speed is negative [41]. For the case where $A > 1$ it is unclear whether there is a minimum wave speed condition. A comparison between the observed wave speed for each A value and the predicted minimum wave speed is given in Figure 8.26. The predicted wave speeds match the observed wave speeds well for $A \leq 1$. For $A > 1$ we superimpose the wave speed prediction $v = 2\sqrt{-ArD}$, and observe that the predictions match the numerical wave speeds well. For the case $A > 1$ we require that $\lambda_i > K_i$, and hence the minimum wave speed condition is the same as for the weak Allee effect.

8.6.6 Different motility rates, no grouped agent death.

For $P_m^i \neq P_m^g$ and $P_d^g = 0$, there is a co-operative mechanism in terms of increased survival for agents in close proximity to other agents. In this parameter regime Equation (8.2) becomes

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x}\left(F(C)\frac{\partial C}{\partial x}\right) + (K_i - \lambda_i + \lambda_g)C(1-C)\left(C - \frac{K_i - \lambda_i}{K_i - \lambda_i + \lambda_g}\right), \tag{8.85}$$

where $F(C) = D_i(1 - 4C + 3C^2) + D_g(4C - 3C^2)$. Note that $F(C)$ is the same as in Case 2 and, as such, encodes the same four types of qualitative behaviour. To examine the long term travelling wave behaviour of Equation (8.85), we transform Equation (8.85) into travelling wave co-ordinates, $z = x - vt$, giving

$$v\frac{\mathrm{d}C}{\mathrm{d}z} + F(C)\frac{\mathrm{d}^2C}{\mathrm{d}z^2} + (D_i - D_g)(6C - 4)\left(\frac{\mathrm{d}C}{\mathrm{d}z}\right)^2 + R(C) = 0, \ -\infty < z < \infty, \tag{8.86}$$

where $R(C) = (K_i - \lambda_i + \lambda_g)C(1-C)(C - (K_i - \lambda_i)/(K_i - \lambda_i + \lambda_g))$. Making the substitution $U = \mathrm{d}C/\mathrm{d}z$ results in

$$\frac{\mathrm{d}C}{\mathrm{d}z} = U, \tag{8.87}$$

$$\frac{\mathrm{d}U}{\mathrm{d}z} = \frac{-vU - (D_i - D_g)(6C-4)U^2(K_i - \lambda_i + \lambda_g) - R(C))}{F(C)}. \tag{8.88}$$

Equation (8.88) is singular if $F(C) = 0$ for $0 \leq C \leq 1$. It is therefore of interest to determine whether travelling wave solutions can be found for each class of $F(C)$.

Strictly positive nonlinear diffusivity function

Strictly positive $F(C)$ corresponds to parameters in the grey region of Figure 8.12(a). As the wave speed for an arbitrary A is not determined by linear analysis for the Allee equation with constant diffusivity, we follow the approach of Hadeler [23–25] to determine a condition for $v > 0$. Making the transformation

$$\hat{z} = \int_0^z \frac{1}{D_i(1-4C(s)+3C(s)^2)+D_g(4C(s)-3C(s)^2)}\,\mathrm{d}s, \tag{8.89}$$

results in

$$\frac{\mathrm{d}C}{\mathrm{d}\hat{z}} = \hat{U}, \tag{8.90}$$

$$\frac{\mathrm{d}\hat{U}}{\mathrm{d}\hat{z}} = -v\hat{U} - \lambda\Big(D_i(1-4C+3C^2)+D_g(4C-3C^2)\Big) \times \\ (K_i-\lambda_i+\lambda_g)C(1-C)\Big(C-\frac{K_i-\lambda_i}{K_i-\lambda_i+\lambda_g}\Big). \tag{8.91}$$

For $v > 0$ we require that the transformed source term is, on average, positive, which corresponds to [23–25, 46]

$$\int_0^1 (K_i-\lambda_i+\lambda_g)\Big(D_i\big(1-4C+3C^2\big)+D_g\big(4C-3C^2\big)\Big)C(1-C)\Big(C-\frac{K_i-\lambda_i}{K_i-\lambda_i+\lambda_g}\Big)\mathrm{d}C > 0. \tag{8.92}$$

Condition (8.92) is equivalent to

$$(6\lambda_g - 5K_i + 5\lambda_i)D_g - \lambda_g D_i > 0. \tag{8.93}$$

For the case with $r > 0$, Condition (8.93) is equivalent to $(A-1)D_i + (6-11A)D_g > 0$. Since $D_i < 4D_g$ for $F(C) > 0$ on $0 \leq C \leq 1$, it is trivial to see that for $A < 0$, $v > 0$. Interestingly, for $A > 0$, the threshold A value for the population to persist increases if $P_m^g > P_m^i$, and decreases otherwise. For example, if $P_m^i = 0$ then $A < 6/11$ leads to persistence, higher than the threshold A value in the case with constant $F(C)$. Alternatively, as $P_m^i \to 4P_m^g$, $A \to 2/7$. This implies that populations where isolated agents are significantly more motile than grouped agents are more susceptible to extinction. This result is intuitive, as the parameter regime considered here describes a co-operative benefit, namely, a reduced death rate for agents in close proximity to other agents. Finally, for the reverse Allee case, where $r < 0$ and $A > 1$, Condition (8.93) is always satisfied and the population persists.

Travelling wave solutions for the strong Allee effect with a strictly positive $F(C)$ are shown in Figure 8.27. For the strong Allee effect, with parameters that correspond to $A = 1/4$, presented in Figures 8.27(a)-(c), we observe a heteroclinic orbit between $(1, 0)$ and $(0, 0)$. The numerical solution of Equation (8.85) in this parameter regime approaches a travelling wave solution with $v > 0$. However, if we consider a parameter regime that corresponds to the strong Allee effect with $A = 4/9$, presented in Figures 8.27(d)-(f), we observe that, while a heteroclinic orbit between $(1, 0)$ and $(0, 0)$ exists, it corresponds to a negative wave speed. As a consequence, the population tends to extinction in a birth/death parameter

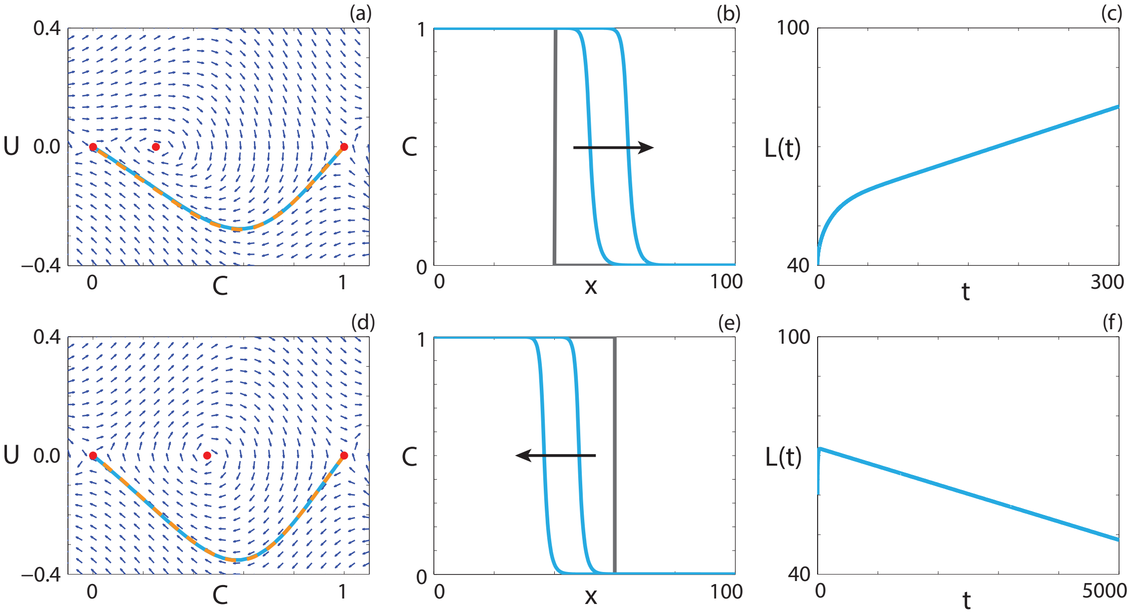

Figure 8.27: Travelling wave behaviour for the Equation (8.85) **with the strong Allee effect and strictly positive** $F(C)$ **(Case 6.1)**. (a), (d) Phase plane for the system (8.87)-(8.88) with the numerical solution of Equations (8.85) (cyan, solid) and (8.86) (orange, dashed), in (C, U) co-ordinates, superimposed. Red circles correspond to equilibrium points. (b), (e) Numerical solution of Equation (8.85) calculated at (b) $= 150$ and $t = 300$; (e) $t = 2500$ and $t = 5000$. The grey lines indicate the initial condition and the arrows indicate the direction of increasing time. (c), (f) The time evolution of $L(t)$. All results are obtained with $\delta x = 0.1$, $\delta t = 0.01$, $\epsilon = 10^{-6}$, $P_d^g = 0$, (a)-(c) $P_m^i = 1.0$, $P_m^g = 0.5$, $P_p^i = 0.4$, $P_p^g = 0.3$, $P_d^i = 0.5$, $v = 0.084$; (d)-(f) $P_m^i = 1.0$, $P_m^g = 0.5$, $P_p^i = 0.4$, $P_p^g = 0.3$, $P_d^i = 0.65$, $v = -0.004$.

regime that would otherwise result in the persistence of the population if the diffusivity is constant. As both the weak and reverse Allee effect are qualitatively similar to Fisher kinetics, numerical solutions are not presented here. Numerical solutions can be found in Figure 8.28.

Extinction-degenerate non-negative nonlinear diffusivity function

The case where $F(0) = 0$ corresponds to parameters along the orange line in Figure 8.12(a). Sánchez-Garduño and Maini [56] demonstrate that Condition (8.92) must be satisfied for travelling wave solutions to have $v > 0$. Furthermore, there is a critical wave speed that results in a sharp-fronted travelling wave [56]. From the results obtained for Case 6.1, Condition (8.92) is always satisfied for $A < 0$ or $A > 1$. For parameter regimes where $0 < A < 1$ the choice of P_m^i and P_m^g influences whether Condition (8.92) is satisfied. To obtain an extinction-degenerate diffusivity we require that $P_m^i = 0$. Hence (8.93) implies that for $A < 6/11$ the wave speed will be positive. To obtain a positive wave speed with constant $F(C)$, we require $A < 1/2$, which implies that the population is more likely to persist in an parameter regime that leads to extinction-degenerate non-negative $F(C)$.

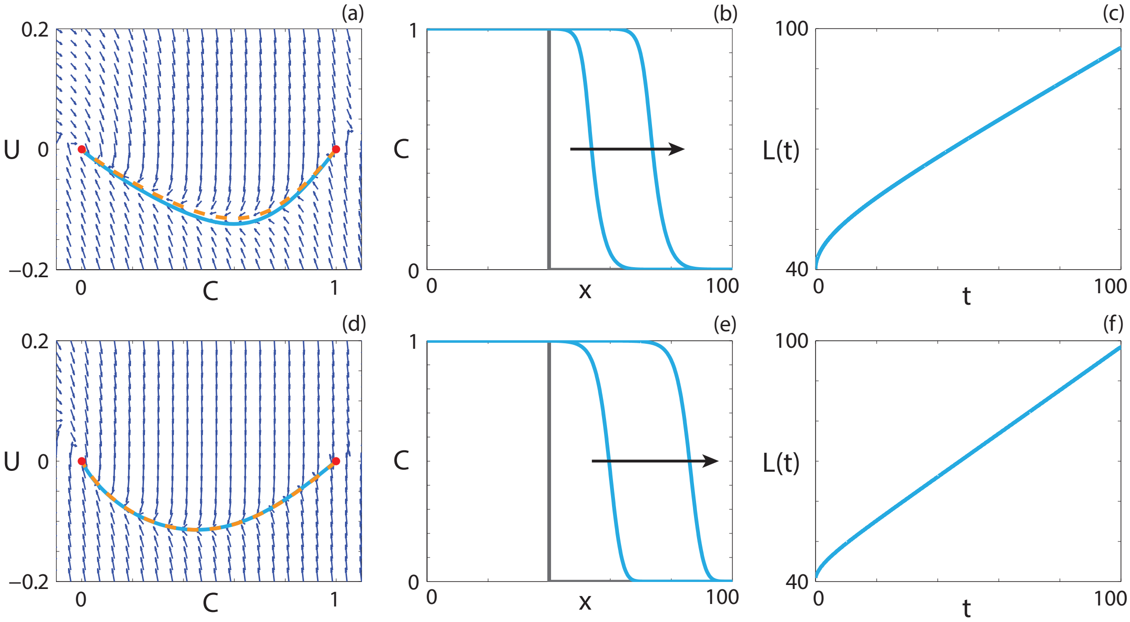

Figure 8.28: Travelling wave behaviour for Equation (8.85) **with the (a)-(c) weak Allee effect and the (d)-(f) reverse Allee effect and strictly positive** $F(C)$ **(Case 6.1)**. (a),(d) Phase plane for the system (8.87)-(8.88) with the numerical solution to Equations (8.85) (cyan, solid) and (8.86) (orange, dashed), in (C, U) co-ordinates, superimposed. Red circles correspond to equilibrium points. (b),(e) Numerical solution to Equation (8.85) calculated at $t = 50$ and $t = 100$. The grey lines indicate the initial condition and the arrows indicate the direction of increasing time. (c),(f) The time evolution of the position of the leading edge of the wave front. All results are obtained with $\delta x = 0.1$, $\delta t = 0.01$, $\epsilon = 10^{-6}$, $P_d^g = 0$, (a)-(c) $P_m^i = 1.0$, $P_m^g = 0.5$, $P_p^i = 0.4$, $P_p^g = 0.3$, $P_d^i = 0.3$, $v = 0.448$; (d)-(f) $P_m^i = 0.5$, $P_m^g = 0.25$, $P_p^i = 0.6$, $P_p^g = 0.2$, $P_d^i = 0.3$, $v = 0.536$.

Travelling wave behaviour for the strong Allee effect with extinction-degenerate non-negative $F(C)$ is shown in Figure 8.29. The numerical solution of Equation (8.85) with $A = 1/4$, in Figures 8.29(a)-(c), leads to a sharp-fronted travelling wave solution near $C = 0$ with $v > 0$. With $A = 1/4$, we expect to obtain $v > 0$. For a parameter regime that results in $A = 4/7$, we obtain a travelling wave solution of Equation (8.85) with $v < 0$ (Figures 8.29(d)-(f)). Interestingly, the sharp front near $C = 0$ is not present for the strong Allee effect with $v < 0$, unlike with $v > 0$, where the wave front is smooth. We present travelling wave behaviour for both the weak Allee effect and the reverse Allee effect in Figure 8.30.

Positive-negative-positive nonlinear diffusivity function

A positive-negative-positive $F(C)$, where there is an interval $\alpha < C < \beta$ where $F(C) < 0$, corresponds to parameter pairs highlighted in purple in Figure 8.12(a). Kuzmin and Ruggerini [38] examine reaction-diffusion equations with similar properties for the strong Allee effect, in the context of diffusion-aggregation models, and provide conditions for smooth travelling wave solutions to exist. For a solution with $v > 0$, we require $A < \alpha$ [38] and

$$\int_0^{\alpha} F(C)R(C) \, \mathrm{d}C > 0. \tag{8.94}$$

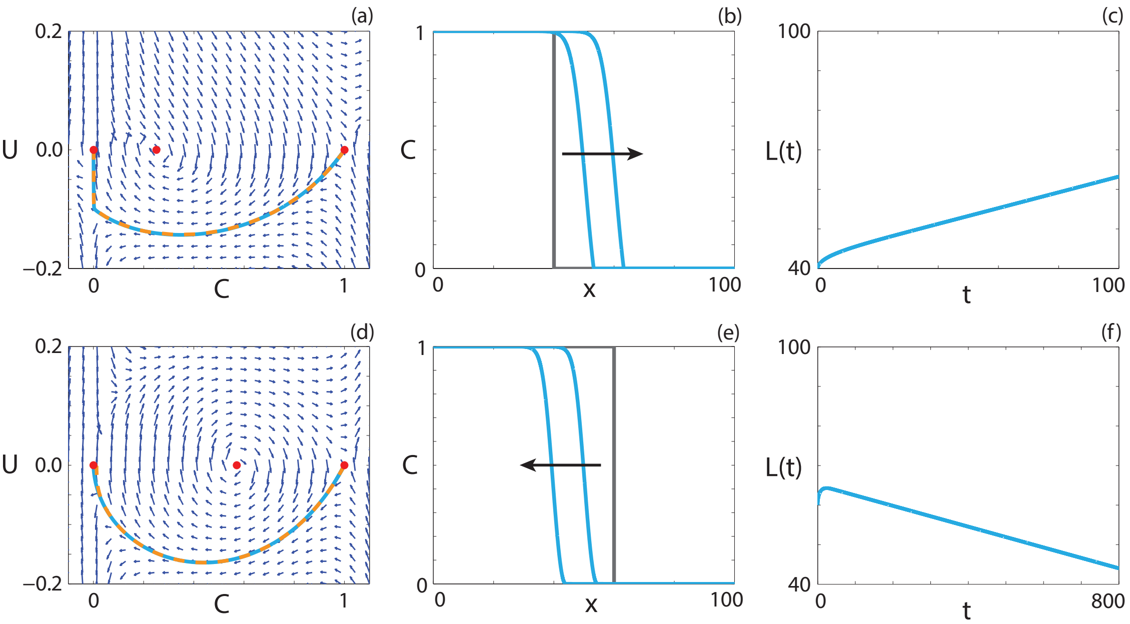

Figure 8.29: Travelling wave behaviour for Equation (8.85) **with the strong Allee effect and extinction-degenerate non-negative** $F(C)$ **(Case 6.2)**. (a), (d) Phase plane for the system (8.87)-(8.88) with the numerical solution of Equations (8.85) (cyan, solid) and (8.86) (orange, dashed), in (C, U) co-ordinates, superimposed. Red circles correspond to equilibrium points. (b), (e) Numerical solution of Equation (8.85) calculated at (b) $t = 50$ and $t = 100$; (e) $t = 400$ and $t = 800$. The grey lines indicate the initial condition and the arrows indicate the direction of increasing time. (c), (f) The time evolution of $L(t)$. All results are obtained with $\delta x = 0.01$, $\delta t = 0.005$, $\epsilon = 10^{-6}$, $P_m^i = 0$, $P_m^g = 1.0$,$P_d^g = 0$, (a)-(c) $P_p^i = 0.4$, $P_p^g = 0.3$, $P_d^i = 0.5$, $v = 0.199$; (d)-(f) $P_p^i = 0.4$, $P_p^g = 0.3$, $P_d^i = 0.8$, $v = -0.026$.

Furthermore, we require [38]

$$3\int_0^{\alpha} F(C)R(C)\ \mathrm{d}C \geq \max\{\Phi(\sigma), \Phi(\rho)\}, \tag{8.95}$$

where

$$\begin{aligned}
&\Phi(y) = 8\alpha^2 y + 4\sqrt{4\alpha^2 y^2 - 2m\alpha^3 y},\\
&\sigma = \sup_{C\in[\alpha,\beta)}\left[\frac{F(C)R(C)}{C-\beta}\right],\ \ \rho = \sup_{C\in(\beta,1]}\left[\frac{F(C)R(C)}{C-\beta}\right],\ \text{and}\\
&m = \min_{C\in[0,A]}\left[F(C)R(C)\right].
\end{aligned}$$

A suite of P_m^g values with $P_m^i = 1$, which correspond to $1/3 < \alpha < 2/3$, are considered for parameter regimes that result in $A < \alpha$. Figures 8.31(a)-(c) show the parameter spaces, (A, α), that satisfy Condition (8.94), Condition (8.95) and Conditions (8.94)-(8.95) simultaneously, respectively. Orange regions represent parameter pairs where the condition is satisfied and grey regions represent parameter pairs where the condition is not satisfied. These results suggest that smooth travelling wave solutions should exist for certain choices of parameters. Interestingly, all parameter pairs that satisfy Condition (8.94) also satisfy Condition (8.95).

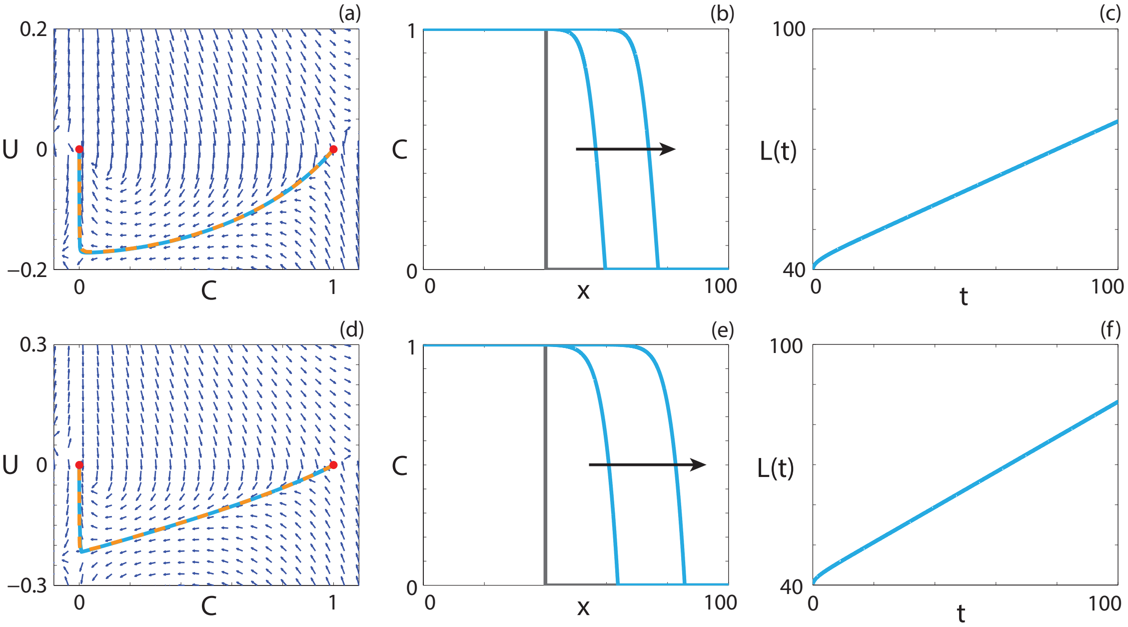

Figure 8.30: Travelling wave behaviour for Equation (8.85) **with the (a)-(c) weak Allee effect and the (d)-(f) reverse Allee effect and extinction-degenerate non-negative** $F(C)$ **(Case 6.2)**. (a),(d) Phase plane for the system (8.87)-(8.88) with the numerical solution to Equations (8.85) (cyan, solid) and (8.86) (orange, dashed), in (C, U) co-ordinates, superimposed. Red circles correspond to equilibrium points. (b),(e) Numerical solution to Equation (8.85) calculated at $t = 50$ and $t = 100$. The grey lines indicate the initial condition and the arrows indicate the direction of increasing time. (c),(f) The time evolution of the position of the leading edge of the wave front. All results are obtained with $\delta x = 0.1$, $\delta t = 0.01$, $\epsilon = 10^{-6}$, $P_m^i = 0$, $P_m^g = 1.0$, $P_d^g = 0$, (a)-(c) $P_p^i = 0.4$, $P_p^g = 0.3$, $P_d^i = 0.3$, $v = 0.347$; (d)-(f) $P_p^i = 0.6$, $P_p^g = 0.2$, $P_d^i = 0.3$, $v = 0.438$.

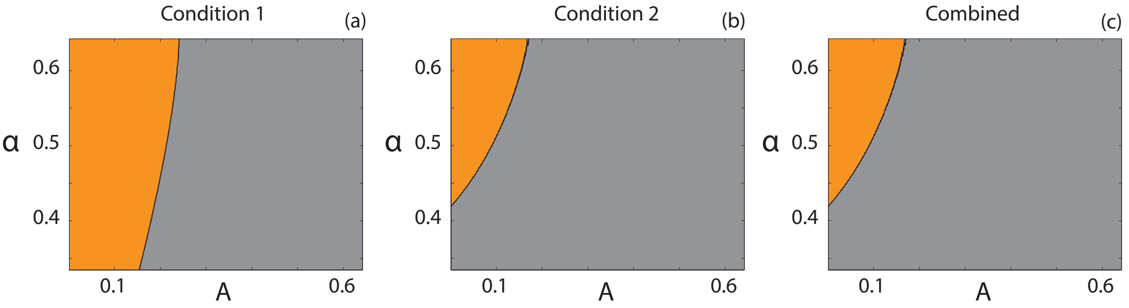

Figure 8.31: Parameter pairs that satisfy Kuzmin and Ruggerini's Conditions [38]. (a) Condition (8.94); (b) Condition (8.95); (c) Conditions (8.94)-(8.95) combined. Orange regions correspond to parameter pairs that satisfy the respective condition(s), whereas grey regions correspond to parameter pairs that do not.

For Case 2.3 and Case 4.3, smooth travelling wave solutions that pass through holes in the wall of singularities for positive-negative-positive $F(C)$ are obtained. The minimum wave speed bound presented by Ferracuti *et al.* [14] implies that the location of the holes in the wall occur are real-valued for the wave speed arising from the Heaviside initial condition. As such, to obtain smooth travelling wave solutions of Equation (8.85) with positive-negative-positive $F(C)$, we might expect that the wave speed satisfies $v > 2\sqrt{F'(\beta)R(\beta)}$, such that the holes in the wall at $C = \beta$ are real-valued.

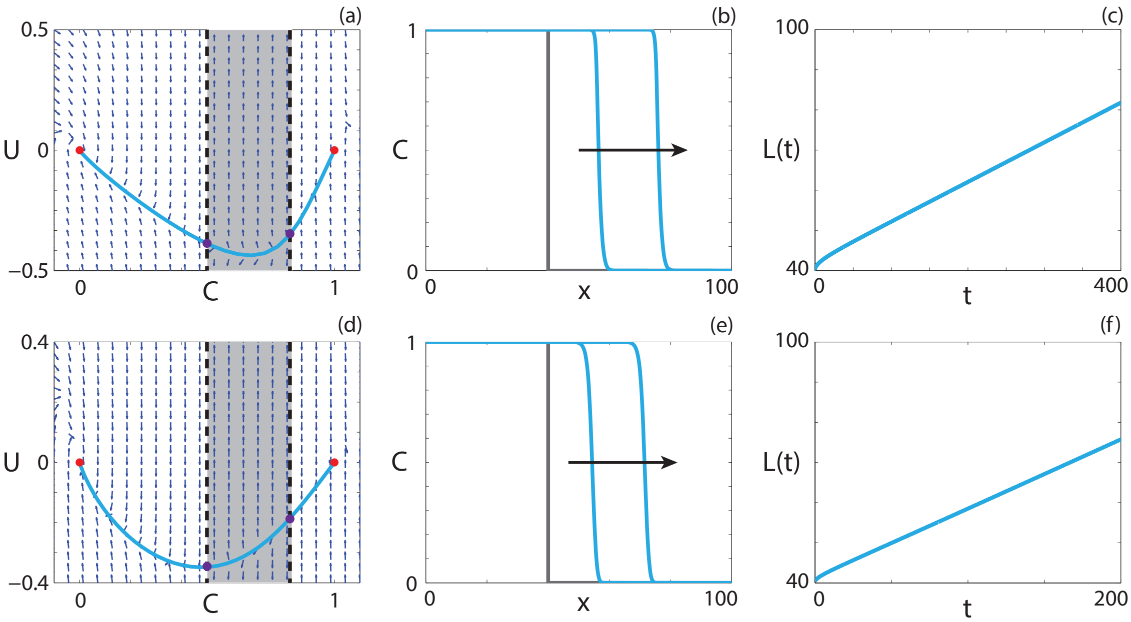

Figure 8.32: Travelling wave behaviour for Equation (8.85) **with the (a)-(c) weak Allee effect and the (d)-(f) reverse Allee effect and positive-negative-positive** $F(C)$ **(Case 6.3)**. (a),(d) Phase plane for the system (8.87)-(8.88) with the numerical solution to Equation (8.85) (cyan, solid), in (C, U) co-ordinates, superimposed. The dashed black lines denote a wall of singularities. Red circles correspond to equilibrium points and purple circles correspond to holes in the wall. (b),(e) Numerical solution to Equation (8.85) calculated at (b) $t = 200$ and $t = 400$; (e) $t = 100$ and $t = 200$. The grey lines indicate the initial condition and the arrows indicate the direction of increasing time. (c),(f) The time evolution of the position of the leading edge of the wave front. All results are obtained with $\delta x = 0.1$, $\delta t = 0.01$, $\epsilon = 10^{-6}$, $P_d^g = 0$, (a)-(c) $P_m^i = 0.05$, $P_m^g = 0.01$, $P_p^i = 0.4$, $P_p^g = 0.3$, $P_d^i = 0.3$, $v = 0.098$; (d)-(f) $P_m^i = 0.5$, $P_m^g = 0.1$, $P_p^i = 0.6$, $P_p^g = 0.2$, $P_d^i = 0.3$, $v = 0.172$.

Following the approach used for Case 2.3, it is simple to demonstrate that both the weak and reverse Allee effect have real-valued holes in the wall. As such, we observe heteroclinic orbits between $(1, 0)$ and $(0, 0)$ that pass through the holes in the wall, and present the corresponding travelling wave solutions in Figure 8.32. We now examine numerical solutions of Equation (8.85) with the strong Allee effect. For parameter regimes that give rise to wave speeds that satisfy $v > 2\sqrt{F'(\beta)R(\beta)}$, numerical travelling wave solutions could not be found. While the condition for real-valued holes in the wall is satisfied, the zeros of Equation (8.88) are imaginary for a certain interval of $C > \beta$. This corresponds to a nullcline that is not real-valued for certain C values.

We now consider parameter regimes corresponding to the strong Allee effect with the additional restriction that $v < 2\sqrt{F'(C)R(C)}$ for $2/3 < C \leq 1$. For all P_m^i and P_m^g that give rise to a positive-negative-positive $F(C)$, holes in the wall at $C = \beta$ do not exist and, as such, we do not expect to obtain smooth solutions. Interestingly, we observe travelling wave solutions with shocks such that the solution never enters the region $\alpha < C < \beta$. An example of a shock-fronted travelling wave solution for the strong Allee effect with both $v > 0$ and $v < 0$ is shown in Figures 8.33(a)-(c) and Figures 8.33(d)-(f), respectively.

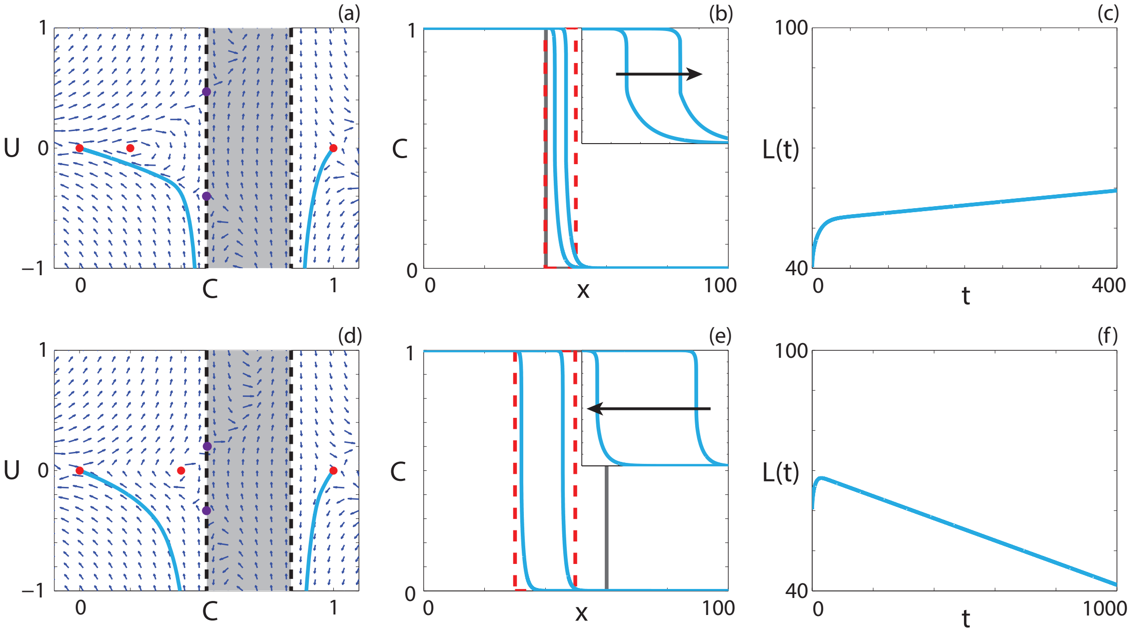

Figure 8.33: Travelling wave behaviour for Equation (8.85) **with the strong Allee effect and positive-negative-positive** $F(C)$ **(Case 6.3).** (a), (d) Phase plane for the system (8.87)-(8.88) with the numerical solution of Equation (8.85) (cyan, solid), in (C, U) co-ordinates, superimposed. The dashed black lines denote a wall of singularities. Red circles correspond to equilibrium points and purple circles correspond to holes in the wall. (b), (e) Numerical solution of Equation (8.85) calculated at (b) $t = 200$ and $t = 400$; (e) $t = 500$ and $t = 1000$. The grey lines indicate the initial condition and the arrows indicate the direction of increasing time. The insets correspond to the areas within the red dashed lines, and highlight the shocks. (c), (f) The time evolution of $L(t)$. All results are obtained with $\delta x = 0.05$, $\delta t = 0.001$, $\epsilon = 10^{-6}$, $P_d^g = 0$, (a)-(c) $P_m^i = 0.5$, $P_m^g = 0.1$, $P_p^i = 0.5$, $P_p^g = 0.4$, $P_d^i = 0.6$, $v = 0.009$; (d)-(f) $P_m^i = 0.5$, $P_m^g = 0.1$, $P_p^i = 0.4$, $P_p^g = 0.2$, $P_d^i = 0.5$, $v = -0.028$.

Solutions of diffusion equations, without any source terms, that contain shocks have been reported previously [39, 72]. Similarly, shock-fronted travelling wave solutions arise in other kinds of models, including multispecies models of combustion [18] and haptotactic cell migration [71]. However, the models presented here are very different, and it is therefore of interest to determine the properties of the reaction-diffusion equation that lead to shock-fronted travelling wave solutions.

Capacity-degenerate positive-negative nonlinear diffusivity function

Capacity-degenerate positive-negative $F(C)$, where $F(1) = 0$, arises if $P_m^g = 0$ and includes an interval $1/3 < C < 1$ where $F(C) < 0$. For Case 2.4, despite the degenerate nature of the nonlinear diffusivity function at $C = 1$, we did not obtain solutions with a sharp front near $C = 1$. Instead, the solution passes through the region of negative diffusivity and a hole in the wall at $C = 1/3$, leading to smooth travelling wave solutions. As such, we expect similar solutions for both the weak and reverse Allee effect due to the qualitatively similar behaviour of the $R(C)$ function. It is of interest to examine whether smooth or shock-fronted travelling wave solutions arise from Equation (8.85) for the strong Allee effect, as for the positive-negative-positive diffusivity examined for Case 6.3 no smooth travelling wave solutions could be found.

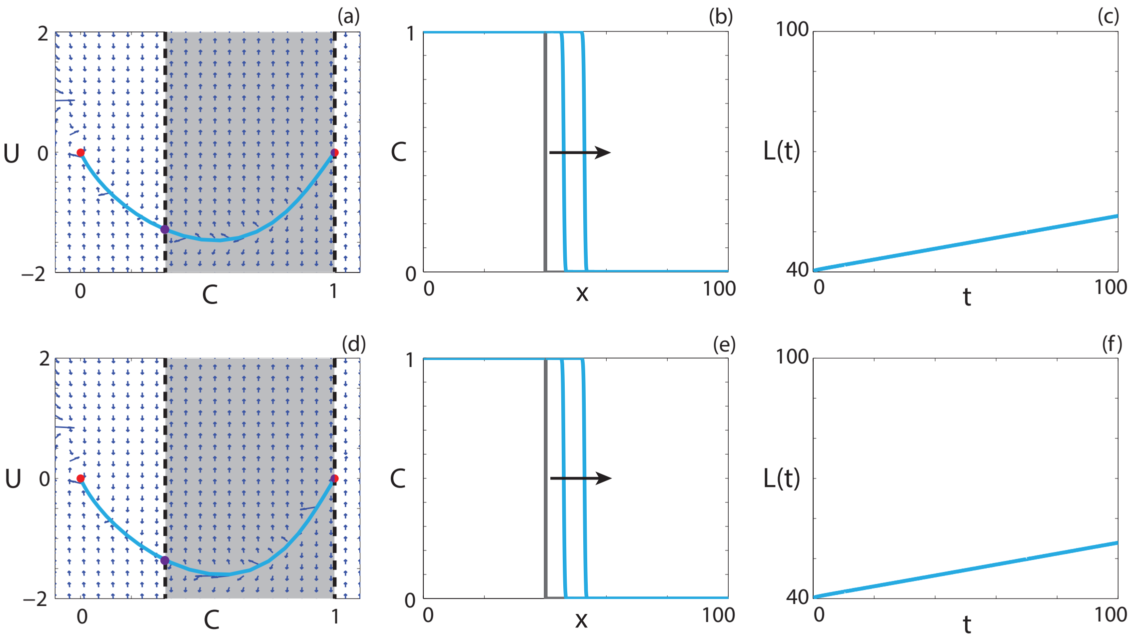

Figure 8.34: Travelling wave behaviour for the (a)-(c) weak Allee effect and the (d)-(f) reverse Allee effect with capacity-degenerate $F(C)$ (Case 6.4). (a), (d) Phase plane for the system (8.87)-(8.88) with the numerical solution of Equation (8.85) (cyan, solid), in (C, U) co-ordinates, superimposed. The dashed black lines denote a wall of singularities. Red circles correspond to equilibrium points and purple circles correspond to holes in the wall. (b), (e) Numerical solution of Equation (8.85) calculated at $t = 50$ and $t = 100$. The grey lines indicate the initial condition and the arrows indicate the direction of increasing time. (c), (f) The time evolution of $L(t)$. All results are obtained with $\delta x = 0.1$, $\delta t = 0.01$, $\epsilon = 10^{-6}$, $P_m^i = 0.01$, $P_m^g = 0$, $P_p^i = 1.0$, $P_d^i = 0.1$, $P_d^g = 0$, (a)-(c) $P_p^g = 0.8$, $v = 0.098$; (d)-(f) $P_p^g = 0.95$, $v = 0.136$.

As expected, smooth travelling wave solutions for both the weak and reverse Allee effects with capacity-degenerate positive-negative $F(C)$ are obtained. The solution behaviour for both the weak and reverse Allee effects are presented in Figure 8.34. For the strong Allee effect, we examine a considerable number of parameter regimes and initial conditions and are unable to find travelling wave solutions.

8.6.7 Equal motility rates, different death rates.

Without the restriction that only isolated agents are able to undergo death events ($P_d^g \neq 0$), death events can be considered as either a co-operative mechanism ($P_d^i > P_d^g$), such as group defence against predation, or a competitive mechanism ($P_d^i < P_d^g$), where a population is more easily discovered and eradicated, compared to an isolated individual. In these parameter regimes, Equation (8.2) can be expressed as

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} + (K_i - K_g - \lambda_i + \lambda_g)A_1 C\left(1 - \frac{C}{A_1}\right)\left(C - A_2\right), \tag{8.96}$$

where

$$A_1 = \frac{2\lambda_i - \lambda - 2K_i + 2K_g - \sqrt{\lambda_g^2 + 4K_g(\lambda_i - \lambda_g - K_i + K_g)}}{2(\lambda_i - \lambda_g - K_i + K_g)},$$
$$A_2 = \frac{2\lambda_i - \lambda - 2K_i + 2K_g + \sqrt{\lambda_g^2 + 4K_g(\lambda_i - \lambda_g - K_i + K_g)}}{2(\lambda_i - \lambda_g - K_i + K_g)}, \tag{8.97}$$

provided that $\lambda_g \geq 2(K_g + \sqrt{Kg(K_i - \lambda_i)})$ or $\lambda_i > K_i$. If this is not satisfied, $R(C) \leq 0$ for $0 \leq C \leq 1$ and the population will tend to extinction. The corresponding ODE in travelling wave co-ordinates is

$$v\frac{\mathrm{d}C}{\mathrm{d}z} + D\frac{\mathrm{d}^2C}{\mathrm{d}z^2} + (K_i - K_g - \lambda_i + \lambda_g)A_1C\left(1 - \frac{C}{A_1}\right)\left(C - A_2\right) = 0, \quad -\infty < z < \infty, \tag{8.98}$$

and, making the substitution $U = \mathrm{d}C/\mathrm{d}z$, results in

$$\frac{\mathrm{d}C}{\mathrm{d}z} = U, \tag{8.99}$$
$$\frac{\mathrm{d}U}{\mathrm{d}z} = -\frac{vU}{D} - \frac{(K_i - K_g - \lambda_i + \lambda_g)A_1C}{D}\left(1 - \frac{C}{A_1}\right)\left(C - A_2\right). \tag{8.100}$$

Introducing a new variable $\overline{C} = C/A_1$ which, upon substitution into Equation (8.96), results in

$$\frac{\partial \overline{C}}{\partial t} = D\frac{\partial^2 \overline{C}}{\partial x^2} + (K_i - K_g - \lambda_i + \lambda_g)A_1^2\overline{C}(1 - \overline{C})\left(\overline{C} - \overline{A}\right), \tag{8.101}$$

where

$$\overline{A} = \frac{A_2}{A_1} = \frac{2\lambda_i - \lambda - 2K_i + 2K_g + \sqrt{\lambda_g^2 + 4K_g(\lambda_i - \lambda_g - K_i + K_g)}}{2\lambda_i - \lambda - 2K_i + 2K_g - \sqrt{\lambda_g^2 + 4K_g(\lambda_i - \lambda_g - K_i + K_g)}}. \tag{8.102}$$

Equation (8.101) is a reaction-diffusion equation with Allee kinetics in terms of the scaled variable $\overline{C}$. Both the carrying capacity and Allee parameter are scaled by A_1, which influences the maximum population density as well as the threshold density required for positive growth. Following the analysis for Case 5, the minimum wave speed for Equation (8.101) with $\overline{A} < -1/2$ is $v^* = 2\sqrt{(\lambda_i - K_i)D}$. Interestingly, this implies that introducing grouped agent death at a rate that does not result in a population tending to extinction has no influence on the invasion speed of the population. Specifically, the condition for $A < -1/2$ in Case 5 corresponds to $3(\lambda_i - K_i) > \lambda_g$. It can be shown that, with $3(\lambda_i - K_i) > \lambda_g$, we require $3K_g < \lambda_g$ for $\overline{A} < -1/2$. This implies that there is a range of K_g values that result in a travelling wave with a minimum wave speed that is independent of both K_g and λ_g. Interestingly, this suggests that if a control is implemented that increases the death rate of grouped agents, there is a threshold value for the control to influence the invasion speed and the subsequent persistence of the population. Introducing a non-zero K_g value for a parameter regime that results in the strong Allee effect with $K_g = 0$ never changes the type of Allee effect. Hence it is possible

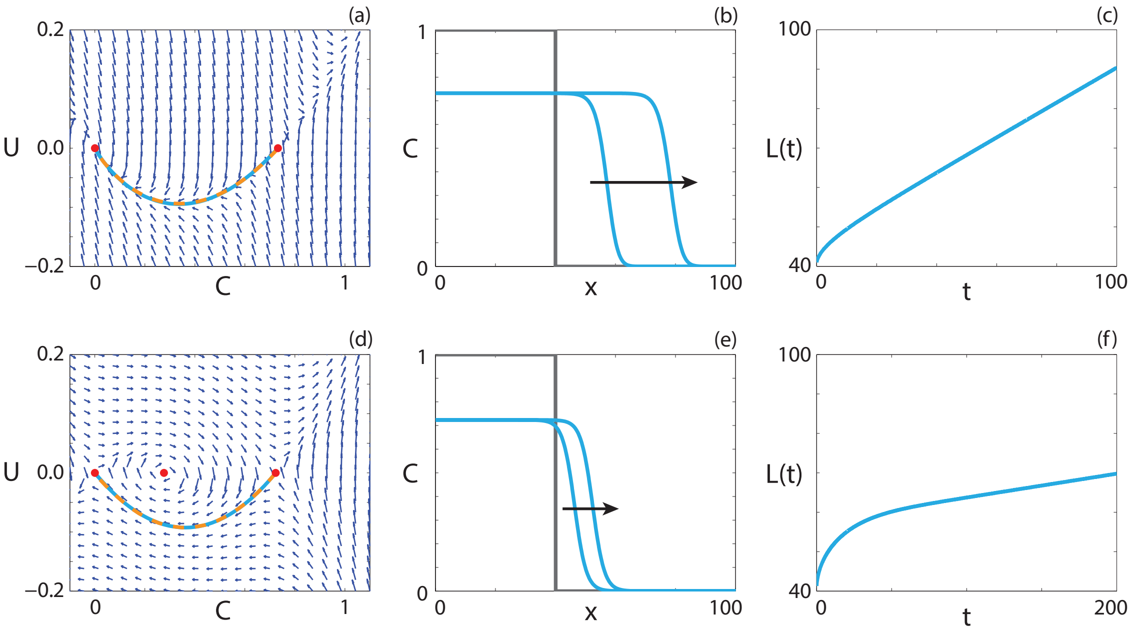

Figure 8.35: Travelling wave behaviour for the (a)-(c) weak Allee effect and the (d)-(f) strong Allee effect with constant $F(C)$ (Case 7). (a), (d) Phase plane for the system (8.99)-(8.100) with the numerical solution of Equations (8.96) (cyan, solid) and (8.98) (orange, dashed), in (C, U) co-ordinates, superimposed. Red circles correspond to equilibrium points. (b), (e) Numerical solution of Equation (8.96) calculated at (b) $t = 50$ and $t = 100$; (e) $t = 100$ and $t = 200$. The grey lines indicate the initial condition and the arrows indicate the direction of increasing time. (c), (f) The time evolution of $L(t)$. All results are obtained with $\delta x = 0.1$, $\delta t = 0.01$, $\epsilon = 10^{-6}$, $P_d^g = 0.1$, (a)-(c) $P_m^i = 0.5$, $P_m^g = 0.5$, $P_p^i = 0.7$, $P_p^g = 0.4$, $P_d^i = 0.5$, $v = 0.44$; (d)-(f) $P_m^i = 1.0$, $P_m^g = 1.0$, $P_p^i = 0.7$, $P_p^g = 0.5$, $P_d^i = 0.8$, $v = 0.06$.

to go from a weak Allee effect to a reverse Allee effect by introducing a non-zero K_g value. Non-zero K_g values correspond to a decreased benefit for grouped agents, which explains why the source term, previously a weak Allee effect, becomes the reverse Allee effect, corresponding to inhibited growth at high density.

For the strong Allee effect, corresponding to $0 < A_2 < A_1 \leq 1$, the unique wave speed is $v = 2\sqrt{(K_i - K_g - \lambda_i + \lambda_g)D}(A_1/2 - A_2)$ [41]. This implies that for $A_2 > A_1/2$, $v < 0$ and $v > 0$ otherwise. Furthermore, the same wave speed applies for $-A_1/2 < A_2 < 0$ [41]. For both intervals, the minimum wave speed does depend on the K_g value, and hence implementing any kind of partial eradication of the grouped agents will either reduce the speed of invasion or cause the extinction of the population.

Travelling wave behaviour for the weak and strong Allee effect and constant $F(C)$ is shown in Figure 8.35. For both numerical solutions, calculated with $K_g = 0.1$, the carrying capacity is reduced by approximately 27%. With the exception of K_g, the parameters used to obtain the numerical solutions in Figures 8.35(a)-(c) are the same as in Figures 8.24(a)-(c) and we observe that, as expected, the wave speed is the same. This demonstrates that, while the carrying capacity is reduced, the population is able to

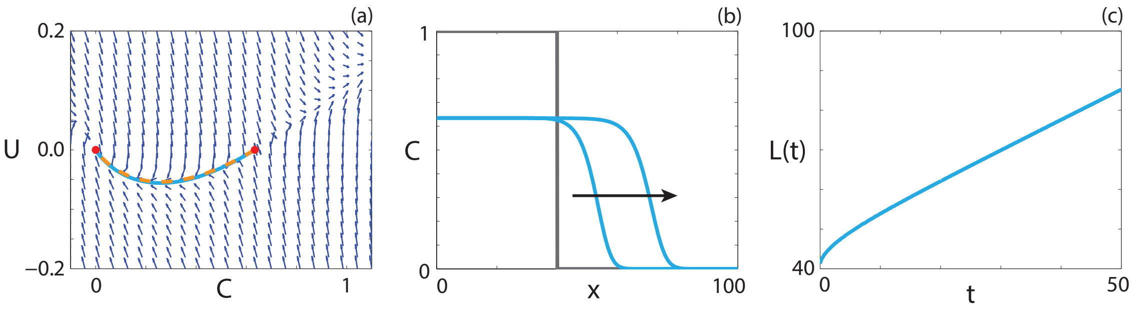

Figure 8.36: Travelling wave behaviour for Equation (8.96) **with the reverse Allee effect and constant** $F_s(\overline{C})$ **(Case 7)**. (a) Phase plane for the system (8.99)-(8.100) with the numerical solution to Equations (8.96) (cyan, solid) and (8.98) (orange, dashed), in (C, U) co-ordinates, superimposed. Red circles correspond to equilibrium points. (b) Numerical solution to Equation (8.96) calculated at $t = 25$ and $t = 50$. The grey lines indicate the initial condition and the arrow indicates the direction of increasing time. (c) The time evolution of the position of the leading edge of the wave front. All results are obtained with $\delta x = 0.1$, $\delta t = 0.01$, $\epsilon = 10^{-6}$, $P_m^i = 1.0$, $P_m^g = 1.0$, $P_p^i = 0.6$, $P_p^g = 0.2$, $P_d^i = 0.3$, $P_d^g = 0.1$, $v = 0.760$.

invade vacant space at the same speed, even though a control measure for the grouped agents has been implemented. Results for the reverse Allee effect are presented in Figure 8.34.

8.6.8 Different motility rates, different death rates.

Setting $P_m^i \neq P_m^g$ and $P_d^i \neq P_d^g \neq 0$ allows for significant flexibility in describing a combination of competitive and/or co-operative mechanisms, depending on the relevant motivation. In this case, Equation (8.2) can be expressed as

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x}\left(F(C)\frac{\partial C}{\partial x}\right) + (K_i - K_g - \lambda_i + \lambda_g)A_1 C\left(1 - \frac{C}{A_1}\right)\left(C - A_2\right), \qquad (8.103)$$

where $F(C) = D_i(1 - 4C + 3C^2) + D_g(4C - 3C^2)$. Note that, again, this simplification requires that $\lambda_g \geq 2(K_g + \sqrt{Kg(K_i - \lambda_i)})$ or $\lambda_i > K_i$, otherwise the population will tend to extinction. In travelling wave co-ordinates, Equation (8.103) is

$$v\frac{\mathrm{d}C}{\mathrm{d}z} + F(C)\frac{\mathrm{d}^2C}{\mathrm{d}z^2} + F'(C)\left(\frac{\mathrm{d}C}{\mathrm{d}z}\right)^2 + (K_i - K_g - \lambda_i + \lambda_g)A_1C\left(1 - \frac{C}{A_1}\right)\left(C - A_2\right) = 0,$$
$$-\infty < z < \infty, \qquad (8.104)$$

and, making the substitution $U = \mathrm{d}C/\mathrm{d}z$, it corresponds to

$$\frac{\mathrm{d}C}{\mathrm{d}z} = U, \qquad (8.105)$$
$$\frac{\mathrm{d}U}{\mathrm{d}z} = -\frac{vU}{F(C)} - \frac{(D_i - D_g)(6C - 4)U^2}{F(C)} - \frac{(K_i - K_g - \lambda_i + \lambda_g)A_1C}{F(C)}\left(1 - \frac{C}{A_1}\right)\left(C - A_2\right). \qquad (8.106)$$

Introducing the variable $\overline{C} = C/A_1$, Equation (8.103) can be written as

$$\frac{\partial \overline{C}}{\partial t} = \frac{\partial}{\partial x}\left(F_A(\overline{C})\frac{\partial \overline{C}}{\partial x}\right) + (K_i - K_g - \lambda_i + \lambda_g)A_1^2\overline{C}(1-\overline{C})(\overline{C}-\overline{A}), \qquad (8.107)$$

where $F_A(\overline{C}) = F(A_1\overline{C}) = D_i(1-4A_1\overline{C}^2+3A_1^2\overline{C}^2)+D_g(4A_1\overline{C}-3A_1^2\overline{C}^2)$. The transformed nonlinear diffusivity, $F_A(\overline{C})$, has the same characteristics as $F_s(\overline{C})$, presented in Figure 8.20, albeit in terms of the scaled Allee carrying capacity, A_1. Here we examine the five types of $F_A(\overline{C})$ for $A_1 \neq 1$.

Strictly positive nonlinear diffusivity function

For $F_A(\overline{C}) > 0$ on the interval $0 < \overline{C} \leq 1$, we follow the approach of Hadeler [23–25]. The integral condition for the wave speed to be positive,

$$\int_0^1 (K_i-K_g-\lambda_i+\lambda_g)A_1^2(D_i(1-4A_1\overline{C}^2+3A_1^2\overline{C}^2)+D_g(4A_1\overline{C}-3A_1^2\overline{C}^2))\overline{C}(1-\overline{C})(\overline{C}-\overline{A})\,\mathrm{d}\overline{C} > 0, \qquad (8.108)$$

corresponds to

$$D_i(5-10\overline{A}+6A_1^2-9A_1A_2-12A_1+20A_2)-D_g(6A_1^2-9A_1A_2-12A_1+20A_2) > 0. \qquad (8.109)$$

If $D_i = D_g$, then $\overline{A} > 1/2$ leads to $v < 0$. For the strong Allee effect, $A_1 > A_2 = \overline{A}A_1$, we can determine the threshold value for the persistence of the population, namely,

$$\overline{A} < \frac{5D_i + (D_i - D_g)(6A_1^2 - 12A_1)}{10D_i + (D_i - D_g)(9A_1^2 - 20A_1)}. \qquad (8.110)$$

Considering the two limiting cases, where $D_i = 0$ and $D_i = 4D_g$, $\overline{A}$ takes on a value of $(6A_1^2-12A_1)/(9A_1^2-20A_1)$ and $(18A_1^2-36A_1+20)/(27A_1^2-60A_1+30)$, respectively. These values reduce to 6/11 and 2/7 in the case that $A_1 = 1$, as in Case 6.1. To illustrate how the threshold value changes with A_1, P_m^i and P_m^g, Figure 8.37 shows the maximum A_2 and $\overline{A}$ values for three different P_m^i and P_m^g combinations. The A_2 value corresponds to the persistence threshold for a given A_1 value. The $\overline{A}$ value can be interpreted as the highest proportion of a given A_1 value that will result in the persistence of the population. For example, in Figure 8.37(a), we see that with $P_m^i = 0$ and $A_1 = 0.5$ we require $A_2 < 0.194$ for persistence. This corresponds to $\overline{A} < 0.388$.

Travelling wave behaviour for Equation (8.103) in a parameter regime corresponding to strictly positive $F_A(\overline{C})$ and the strong Allee effect is shown in Figures 8.38(a)-(c). This parameter regime leads to $A_1 = 0.723$ and $A_2 = 0.2764$, which is below the persistence threshold value of $A_2 = 0.315$ for this P_m^i and P_m^g combination, and hence the population persists.

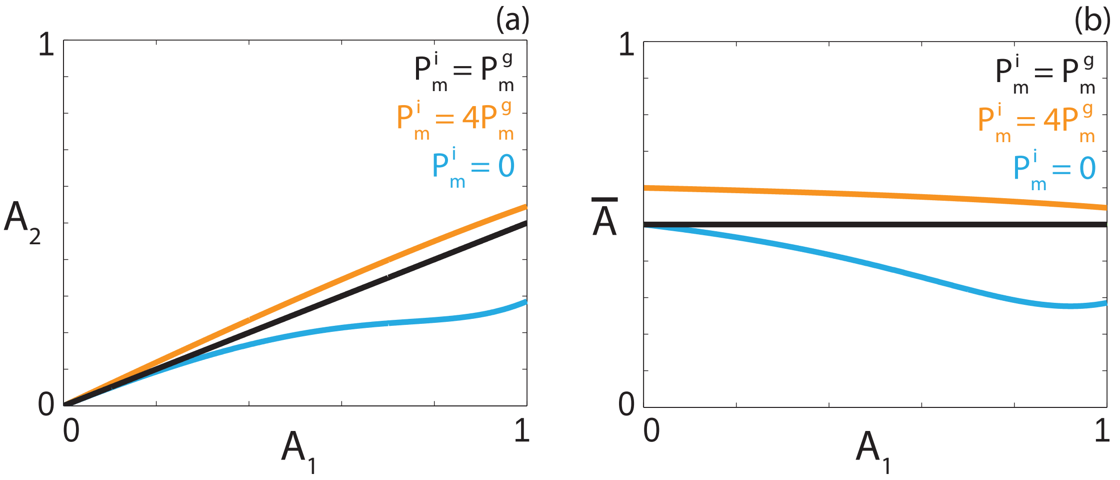

Figure 8.37: Persistence threshold. Persistence threshold as a function of the carrying capacity A_1, expressed as (a) an explicit value, and (b) a proportion of the carrying capacity for three different diffusivities, corresponding to $P_m^i = P_m^g$ (black), $P_m^i = 4P_m^g$ (orange) and $P_m^i = 0$ (cyan).

Extinction-degenerate non-negative nonlinear diffusivity function

For extinction-degenerate $F_A(\overline{C})$, $P_m^i = 0$. As such, the persistence threshold corresponds to $(6A_1^2 - 12A_1)/(9A_1^2 - 20A_1)$. For Case 6.2 we observe that sharp fronts for the strong Allee effect with a extinction-degenerate non-negative $F(C)$ only occur if $v > 0$. Hence for $\overline{A} < (6A_1^2 - 12A_1)/(9A_1^2 - 20A_1)$ Equation (8.103) should approach a sharp-fronted travelling wave solution with $v > 0$, and a smooth travelling solution with $v < 0$ otherwise. Results in Figures 8.38(d)-(f) show numerical solutions of Equation (8.103) with $A_1 = 0.723$ and $A_2 = 0.2764$, which satisfies the threshold for $v > 0$ and hence sharp-fronted travelling wave solutions exist. As expected, results in Figure 8.38(e) indicate that the solution of Equation (8.103) approaches a travelling wave with $v > 0$ and a sharp front near $C = 0$.

Positive-negative-positive nonlinear diffusivity function

For a positive-negative-positive $F_A(\overline{C})$, there are exactly two zeros at $C = \alpha$ and $C = \beta$. In Case 6.3 the strong Allee effect does not give rise to smooth travelling wave solutions, even with real-valued holes in the wall at $C = \alpha$ and $C = \beta$. However, interestingly, shock-fronted travelling wave solutions arise from the Heaviside initial condition. Again, we are unable to find numerical travelling wave solutions of Equation (8.103) in parameter regimes with real-valued holes in the wall. Shock-fronted travelling wave solutions of Equation (8.103) are given in Figures 8.38(g)-(i) where the observed wave speed is $v = 0.014 < 2\sqrt{\min\{F'(C)R(C)\}}$ on the interval $2A_1/3 < C < A_1$. Smooth travelling wave solutions obtained from the weak and reverse Allee effect are shown in Figure 8.39.

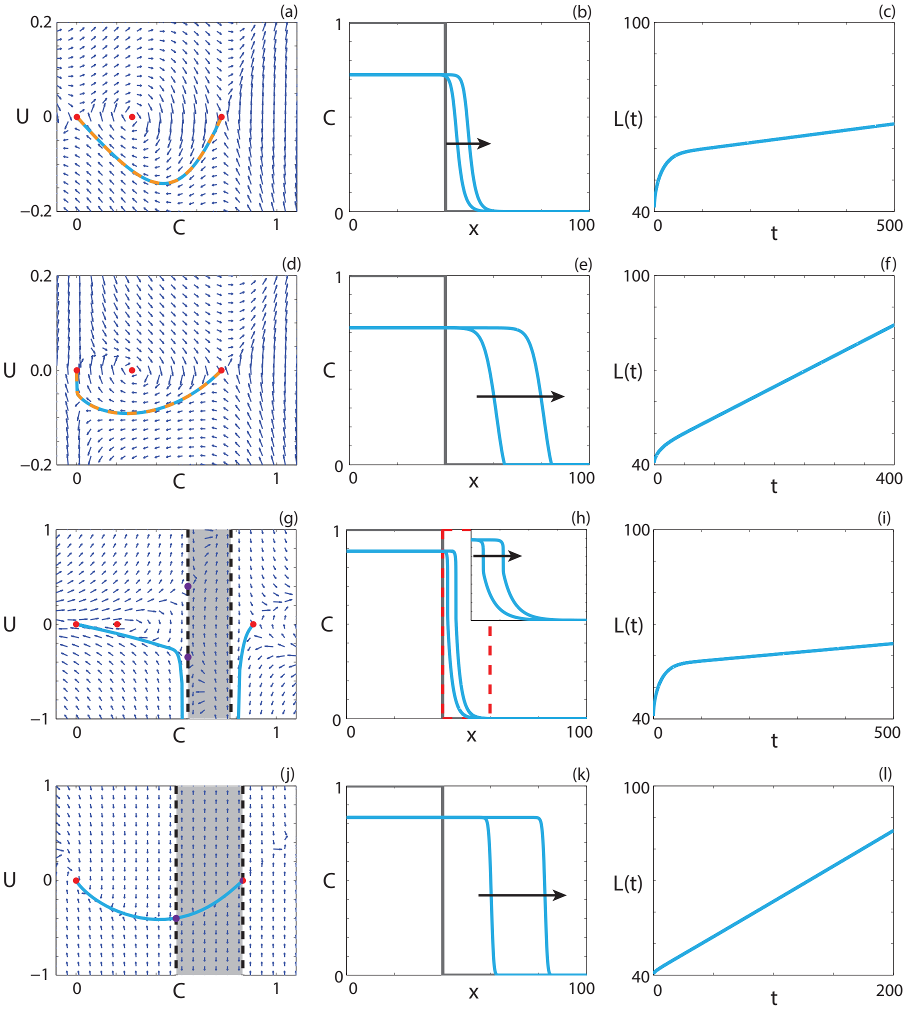

Figure 8.38: Travelling wave behaviour for Equation (8.103) **with various Allee effects and** $F_A(\overline{C})$. We consider the (a)-(c) scaled strong Allee effect with strictly positive $F_A(\overline{C})$ (Case 8.1), (d)-(f) scaled strong Allee effect with extinction-degenerate non-negative $F_A(\overline{C})$ (Case 8.2), (g)-(i) scaled strong Allee effect with positive-negative-positive $F_A(\overline{C})$ (Case 8.3) and, (j)-(l) scaled reverse Allee effect with capacity-degenerate $F_A(\overline{C})$ (Case 8.4). (a), (d), (g), (j) Phase plane for the system (8.105)-(8.106) with the numerical solution of Equations (8.103) (cyan, solid) and (8.104) (orange, dashed), in (C, U) co-ordinates, superimposed. Red circles correspond to equilibrium points. (b), (e), (h), (k) Numerical solution of Equation (8.103) calculated at (b) $t = 250$ and $t = 500$; (e) $t = 200$ and $t = 400$; (h) $t = 250$ and $t = 500$; (k) $t = 100$ and $t = 200$. The grey lines indicate the initial condition and the arrows indicate the direction of increasing time. The inset corresponds to the area within the red dashed lines, and highlights the shock. (c), (f), (i), (l) The time evolution of $L(t)$. All results are obtained with $\delta x = 0.05$, $\delta t = 0.001$, $\epsilon = 10^{-6}$, $P_d^g = 0.1$, (a)-(c) $P_m^i = 1.0$, $P_m^g = 0.5$, $P_p^i = 0.7$, $P_p^g = 0.8$, $P_d^i = 0.5$, $v = 0.02$; (d)-(f) $P_m^i = 0$, $P_m^g = 1.0$, $P_p^i = 0.7$, $P_p^g = 0.5$, $P_d^i = 0.8$, $v = 0.098$; (g)-(i) $P_m^i = 1.0$, $P_m^g = 0.23$, $P_p^i = 0.7$, $P_p^g = 0.5$, $P_d^i = 0.8$, $P_d^g = 0.05$, $v = 0.014$; (j)-(l) $P_m^i = 0.05$, $P_m^g = 0.01$, $P_p^i = 0.7$, $P_p^g = 0.5$, $P_d^i = 0.2$, $P_d^g = 6/70$, $v = 0.22$.

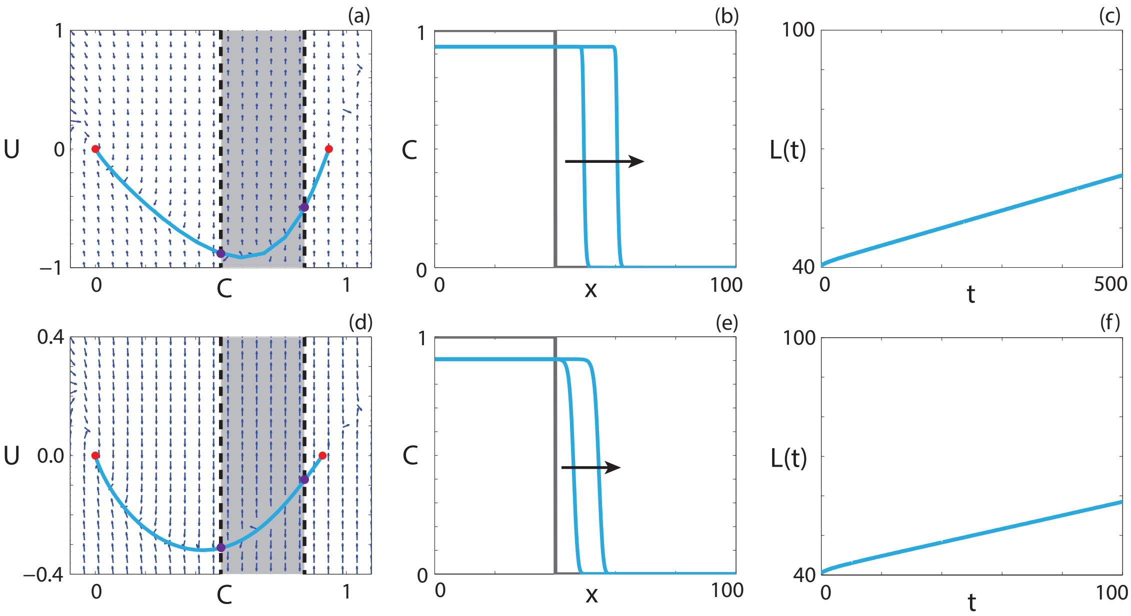

Figure 8.39: Travelling wave behaviour for Equation (8.103) **with the (a)-(c) weak Allee effect and the (d)-(f) reverse Allee effect with positive-negative-positive** $F_s(\overline{C})$ **(Case 8.3)**. (a), (d) Phase plane for the system (8.105)-(8.106) with the numerical solution to Equations (8.103) (cyan, solid) and (8.104) (orange, dashed), in (C, U) co-ordinates, superimposed. Red circles correspond to equilibrium points. (b), (e) Numerical solution to Equation (8.103) calculated at (b) $t = 250$ and $t = 500$; (e) $t = 50$ and $t = 100$. The grey lines indicate the initial condition and the arrow indicates the direction of increasing time. (c), (f) The time evolution of the position of the leading edge of the wave front. All results are obtained with $\delta x = 0.1$, $\delta t = 0.01$, $\epsilon = 10^{-6}$, (a)-(c) $P_m^i = 0.01$, $P_m^g = 0.002$, $P_p^i = 0.3$, $P_p^g = 0.4$, $P_d^i = 0.3$, $P_d^g = 0.02$, $v = 0.045$; (d)-(f) $P_m^i = 0.05$, $P_m^g = 0.01$, $P_p^i = 0.6$, $P_p^g = 0.2$, $P_d^i = 0.3$, $P_d^g = 0.02$, $v = 0.172$.

Capacity-degenerate positive-negative nonlinear diffusivity function

Capacity-degenerate positive-negative $F_A(\overline{C})$ requires $P_m^i = 0$ and, subsequently, $F_A(1) = 0$. Furthermore $F_A(\overline{C}) < 0$ for $\omega < C < S$. For Case 6.4 we found smooth travelling wave solutions for both the weak and reverse Allee effect with capacity-degenerate positive-negative $F(C)$ but could not obtain stable solutions for the strong Allee effect. As $F_A(\overline{C})$ is qualitatively similar to the $F(C)$ considered for Case 6.4 similar results are expected here.

Again, smooth travelling wave solutions of Equation (8.103) for both the weak (Figures 8.38(j)-(l)) and reverse (Figure 8.40) Allee effects are obtained. As for Case 6.4, we consider a variety of parameter regimes corresponding to the strong Allee effect with capacity-degenerate positive-negative $F_A(\overline{C})$, as well as a number of initial conditions, but are unable to find long time travelling wave-type solutions.

Positive-negative nonlinear diffusivity function

For the case where $F_A(\overline{C})$ has exactly one zero on the interval $0 \leq \overline{C} \leq 1$ at $C = \omega$, Maini *et al.* [43] examine the existence of travelling wave solutions, and provide the necessary

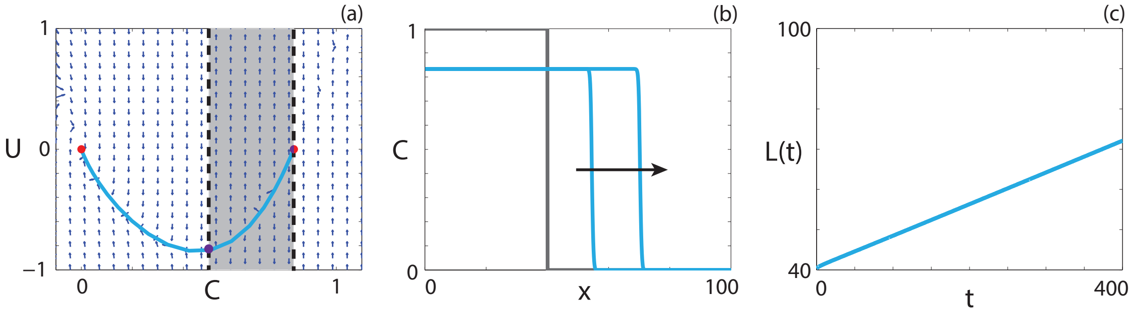

Figure 8.40: Travelling wave behaviour for Equation (8.103) **with the weak Allee effect and positive-negative capacity-degenerate** $F_s(\overline{C})$ **(Case 8.5)**. (a) Phase plane for the system (8.105)-(8.106) with the numerical solution to Equations (8.103) (cyan, solid) and (8.104) (orange, dashed), in (C, U) co-ordinates, superimposed. Red circles correspond to equilibrium points. (b) Numerical solution to Equation (8.103) calculated at $t = 200$ and $t = 400$. The grey lines indicate the initial condition and the arrow indicates the direction of increasing time. (c) The time evolution of the position of the leading edge of the wave front. All results are obtained with $\delta x = 0.1$, $\delta t = 0.01$, $\epsilon = 10^{-6}$, $P_m^i = 0.01$, $P_m^g = 0.002$, $P_p^i = 0.6$, $P_p^g = 0.5$, $P_d^i = 0.3$, $P_d^g = 0.08$, $v = 0.079$.

conditions for existence,

$$A_2 < \omega, \qquad v > 0, \qquad \int_0^{\omega} F(C)R(C)\ \mathrm{d}C > 0, \tag{8.111}$$

where $F(\omega) = 0$ and $0 < \omega < 1$. For the strong Allee effect in this parameter regime, the third part of Condition (8.111) corresponds to

$$\begin{aligned}&D_i(20(A_1 + A_2)\omega - 30A_1A_2\omega - 15\omega^2)+\\&(D_i - D_g)((84A_1 + 36A_2)\omega^3 - (45A_1A_2 + 60A_1 + 60A_2)\omega^2 - 30\omega^4 + 80A_1A_2\omega) > 0.\end{aligned} \tag{8.112}$$

As in Case 4.4, Equation (8.103) is equivalent to

$$\frac{\partial C}{\partial \hat{t}} = \frac{\partial^2 C}{\partial x^2} + (K_i - K_g - \lambda_i + \lambda_g)A_1F(C)C\left(1 - \frac{C}{A_1}\right)\left(C - A_2\right),\ \hat{t} \geq 0, \tag{8.113}$$

on the interval $0 \leq C < \omega$, and equivalent to

$$\frac{\partial C}{\partial \hat{t}} = \frac{\partial^2 C}{\partial x^2} + (K_i - K_g - \lambda_i + \lambda_g)A_1\hat{F}(C)(1-C)\left(1 - \frac{1-C}{A_1}\right)\left(1 - A_2 - C\right),\ \hat{t} \geq 0, \tag{8.114}$$

where $\hat{F}(C) = -F(1 - C)$, on the interval $\omega < C \leq A_1$. The final necessary and sufficient condition from Maini *et al.* [43] for the existence of travelling wave solutions is that the minimum wave speed for Equation (8.113), v_1^*, is greater than, or equal to, the minimum wave speed for Equation (8.114), v_2^*. On the interval $0 \leq C < \omega$, Equation (8.103) has a strictly positive $F_A(\overline{C})$, where $F_A(\overline{C}) \leq D_i$, and strong Allee kinetics. Hence, the minimum wave speed for Equation (8.113) has an upper bound, $v_1^* \leq \sqrt{2(\lambda_i - K_i)D_i}(1/2 - A_2)$. On the interval $\omega < C < A_1$ Equation (8.114) has a source term qualitatively similar to the Fisher-Kolmogorov equation and hence a lower bound for the minimum wave speed exists [43], $v_2^* \geq 2\sqrt{-F(A_1)(\lambda^2 + 4K_g(\lambda_i - \lambda_g - K_i + K_g))^{1/2}}$.

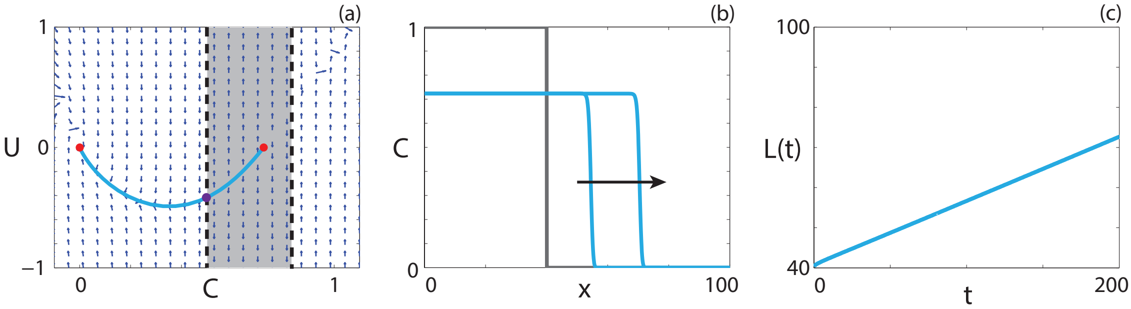

Figure 8.41: Travelling wave behaviour for Equation (8.103) **with the reverse Allee effect and positive-negative** $F_A(\overline{C})$ **(Case 8.5).** (a) Phase plane for the system (8.105)-(8.106) with the numerical solution of Equation (8.103) (cyan, solid), in (C, U) co-ordinates, superimposed. Red circles correspond to equilibrium points. (b) Numerical solution of Equation (8.103) calculated at $t = 100$ and $t = 200$. The grey lines indicate the initial condition and the arrows indicate the direction of increasing time. (c), (f) The time evolution of $L(t)$. All results are obtained with $\delta x = 0.1$, $\delta t = 0.01$, $\epsilon = 10^{-6}$, $P_d^g = 0.1$, $P_m^i = 0.025$, $P_m^g = 0.005$, $P_p^i = 0.7$, $P_p^g = 0.5$, $P_d^i = 0.2$, $P_d^g = 0.15$, $v = 0.16$.

For all parameter regimes considered that correspond to the strong Allee effect with positive-negative $F_A(\overline{C})$ we never observe a case where the upper bound for v_1^* is higher than the lower bound for v_2^* and hence the conditions required for travelling wave solutions are not met. As expected, numerical solutions of Equation (8.103) in these parameter regimes did not lead to travelling wave behaviour. For both the weak and the reverse Allee effect, we expect that solutions do exist as the source terms on both intervals are qualitatively equivalent to a Fisher source term.

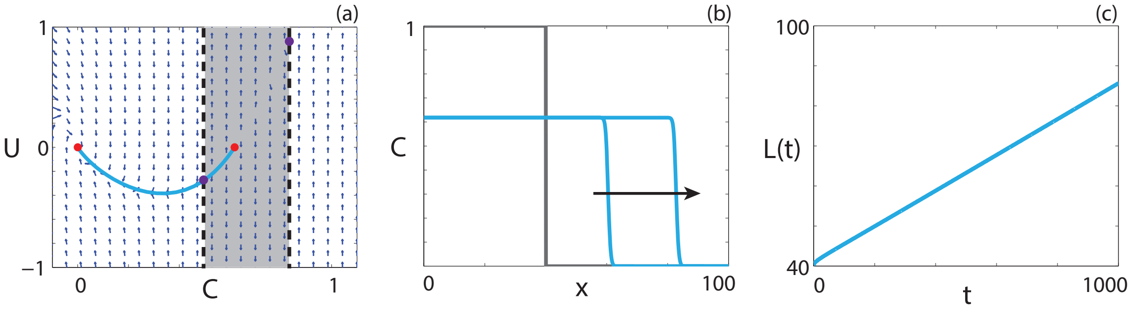

Figure 8.42: Travelling wave behaviour for Equation (8.103) **with the weak Allee effect and positive-negative** $F_s(\overline{C})$ **(Case 8.4).** (a) Phase plane for the system (8.105)-(8.106) with the numerical solution to Equations (8.103) (cyan, solid) and (8.104) (orange, dashed), in (C, U) co-ordinates, superimposed. Red circles correspond to equilibrium points. (b) Numerical solution to Equation (8.103) calculated at $t = 50$ and $t = 100$. The grey lines indicate the initial condition and the arrow indicates the direction of increasing time. (c) The time evolution of the position of the leading edge of the wave front. All results are obtained with $\delta x = 0.1$, $\delta t = 0.01$, $\epsilon = 10^{-6}$, $P_m^i = 0.01$, $P_m^g = 0.002$, $P_p^i = 0.4$, $P_p^g = 0.3$, $P_d^i = 0.3$, $P_d^g = 0.1$, $v = 0.045$.

Numerical solutions demonstrating the travelling wave behaviour of Equation (8.103) with positive-negative $F_A(\overline{C})$ and both the reverse Allee effect and the weak Allee effect are given in Figures 8.41 and 8.42, respectively.

Chapter 9

A new and accurate continuum description of moving fronts

A paper published in the *New Journal of Physics.*

Johnston, S. T., Baker, R. E. and Simpson, M. J. [2017], ‘A new and accurate continuum description of moving fronts’, *New Journal of Physics* **19**(3), 033010.

Abstract

Processes that involve moving fronts of populations are prevalent in ecology and cell biology. A common approach to describe these processes is a lattice-based random walk model, which can include mechanisms such as crowding, birth, death, movement and agent-agent adhesion. However, these models are generally analytically intractable and it is computationally expensive to perform sufficiently many realisations of the model to obtain an estimate of average behaviour that is not dominated by random fluctuations. To avoid these issues, both mean-field and corrected mean-field continuum descriptions of random walk models have been proposed. However, both continuum descriptions are inaccurate outside of limited parameter regimes, and corrected mean-field descriptions cannot be employed to describe moving fronts. Here we present an alternative description in terms of the dynamics of groups of contiguous occupied lattice sites and contiguous vacant lattice sites. Our description provides an accurate prediction of the average random walk behaviour in all parameter regimes. Critically, our description accurately predicts the persistence or extinction of the population in situations where previous continuum descriptions predict the opposite outcome. Furthermore, unlike traditional mean-field models, our approach provides information about the spatial clustering within the population and, subsequently, the moving front.

9.1 Introduction

Moving fronts feature ubiquitously throughout biological and ecological processes [1, 6, 12–14, 26, 32, 36–38, 41, 46]. The introduction of non-native species can result in a catastrophic invasion wave if the introduced species out-competes native fauna [6, 47]. For example, the cane toad *bufo marinus* was introduced to north-eastern Australia in 1935, and has subsequently invaded much of northern Australia due to a lack of natural predation [34, 37]. Similarly, malignant tumours spread through the invasion of previously-healthy tissue, such as glioma cells moving throughout the brain to form glioblastoma [1, 11, 12, 46].

Lattice-based random walk models that include crowding, birth, death, movement and agent-agent adhesion are commonly used to describe processes that involve moving fronts [7, 10, 19, 26, 27, 44, 45]. For example, these random walk models have been used to interpret *in vitro* cell biology experiments by considering the position of the leading edge of the cell front or the cell density profile [26, 27, 45]. Illien *et al.* [19] consider random walks in the context of single-file diffusion models of active transport, that is, transport that requires energy due to an opposing force. However, the stochastic nature of random walks makes it problematic to efficiently examine the collective behaviour of a population, as a large number of realisations of the random walk must be performed to reduce the influence of stochastic fluctuations. Furthermore, it is difficult to determine meaningful population behaviour through analysis of the discrete process. There is, therefore, considerable interest in approaches that are both analytically tractable and avoid the computational expense of repeated simulations.

A common technique to analyse random walk processes is to consider a deterministic, continuum approximation of the discrete process [2, 3, 8, 21, 24, 25, 27, 33, 42, 44, 48]. The standard approximation, known as a mean-field (MF) approximation, results in a partial differential equation (PDE) [8, 27, 44]. The resulting PDE is amenable to analysis but only provides an accurate approximation of the discrete process in an extremely limited set of parameter regimes where spatial correlations are weak [3, 16, 42]. Hence, using these approximations to model moving fronts may provide an inaccurate estimate of the velocity of the moving front if the spatial correlations are important. This inaccuracy could have significant consequences if, for example, the front velocity is used to make decisions about implementing ecological control measures. To address the influence of spatial correlations, alternative approximation techniques have been proposed [3, 21, 40, 42, 49]. The corrected mean-field (CMF) approximation, which explicitly describes pairwise correlations, results in a system of ordinary differential equations (ODEs) that accurately approximate the discrete process for a wider range of parameter regimes, compared to the MF approximation [3, 42]. However, the CMF description is still invalid in parameter regimes where spatial correlations are sufficiently strong [21]. Furthermore, the CMF cannot be used to study moving fronts, as the governing equations cannot be evaluated anywhere that has zero agent density [42], such as areas where the population has yet to invade. The chain-and-gap approach (C&G), proposed by Johnston *et al.* [21], considers the dynamics of groups of contiguous occupied and vacant sites, and results in a system of ODEs that provide an accurate approximation of the discrete process in all parameter regimes. These groups are termed *chains* and *gaps* for the contiguous occupied and vacant sites, respectively. Additionally, the C&G description provides information about the spatial clustering and patchiness present in the system. There is considerable interest in determining the influence of local spatial structure on the persistence of a species [29, 35]. However, the C&G description has previously only been applied to discrete processes that are, on average, spatially uniform [21]. As such, the C&G description is not currently suitable for describing processes that contain moving fronts.

Here the C&G description presented by Johnston *et al.* [21] is extended to incorporate spatial variation so that the description can be applied to moving fronts. We interpret the discrete process in terms of chains and gaps, noting that the left-most site in each chain or gap can occur at any lattice site. The corresponding system of ODEs is derived and presented, and we demonstrate that the numerical solution of the ODE system provides an accurate approximation of the average discrete behaviour in all cases, even in parameter regimes where both the MF and CMF descriptions are inaccurate. This allows for the robust prediction of whether a population persists or becomes extinct, as well as reliable estimates of the velocity of the moving front. In addition, for the first time, the C&G description has been extended to include rates of birth, death and movement that are dependent on the length of the chain an agent belongs to. Furthermore, the C&G description includes explicit information about the spatial clustering present within a moving front.

9.2 Random walk model

We consider a one-dimensional lattice-based random walk model where each lattice site may be occupied by, at most, one agent [5]. The lattice is interpreted as a combination of groups of contiguous occupied and vacant sites [21]. Agents on the lattice undergo birth, death and movement events. These events occur at rates P_p^n, P_d^n and P_m^n per unit time, respectively, where $n \in [1, N]$ is the length of the chain an agent belongs to and N is the total number of sites. During a potential birth event, an agent randomly selects a nearest-neighbour site and attempts to place a daughter agent at that site. The birth event is successful provided that the target site is vacant [3]. During a death event, an agent is removed from the lattice [3]. During a potential movement event, an agent selects a nearest-neighbour site and attempts to move to that site, and is successful provided that the target site is vacant [3]. The target site selection is unbiased if both nearest-neighbour sites are vacant. If one nearest-neighbour site is occupied, the vacant nearest-neighbour site is selected with probability $(1-\alpha)/2$, where $\alpha \in [-1, 1]$ [24]. The constant parameter α represents the strength of agent-agent adhesion/repulsion. Setting $\alpha = 0$ means that there is no agent-agent adhesion/repulsion, whereas setting $\alpha \neq 0$ simulates agent-agent adhesion ($\alpha > 0$) or repulsion ($\alpha < 0$) [24]. Note that α does not depend on the chain length.

Due to crowding, the success of birth and movement events depends on whether an agent has zero, one or two nearest-neighbour agents. These agents are referred to as *single*, *edge* or *middle* agents, respectively [21]. An example lattice configuration highlighting

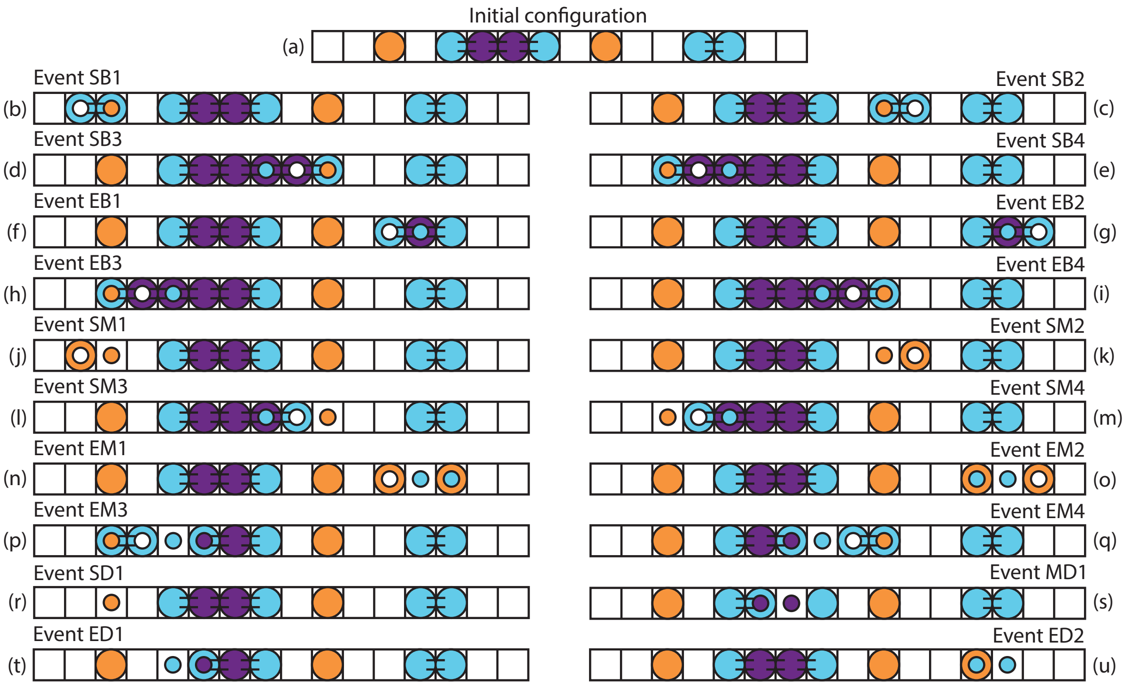

Figure 9.1: (a) Example lattice configuration with single (orange), edge (cyan) and middle (purple) agents. (b)-(u) Potential birth, death and movement events for the configuration of agents in (a), with subsequent change in agent type. Inset circles denote the agent type and location before the birth, death or movement event occurred. Lines connecting agents represent agent-agent adhesion/repulsion.

the different types of agents is presented in Figure 9.1. The necessary information to obtain the average numbers of single, edge and middle agents at each site $i \in [1, N]$ is encoded within the average number of chains of length $n \in [1, N]$ that contain the site i. The average number of single, edge and middle agents at site i are denoted $N_i^S(t)$, $N_i^E(t)$ and $N_i^M(t)$, respectively. The average number of chains at time t where the left-most agent in the chain is at site $i \in [1, N]$ and the length is $n \in [1, N-i+1]$ is denoted $C_i^n(t)$. Similarly, the average number of gaps at time t where the left-most vacant site in the gap is at site $i \in [1, N]$ and the length is $m \in [1, N-i+1]$ is denoted $G_i^m(t)$. The spatially-dependent restriction on the length is due to the choice of no-flux boundary conditions. Note that $N_i^S(t)$, $N_i^E(t)$, $N_i^M(t)$, $C_i^n(t)$ and $G_i^m(t)$ are all temporally-dependent but, for convenience, we do not explicitly include this dependence. The average numbers of edge, middle and single agents at site i are given by:

$$\begin{aligned}
N_i^E &= \sum_{j=2}^{N-i+1} C_i^j + \sum_{j=2}^{i-1} C_{i-j+1}^j, \\
N_i^M &= \sum_{j=1}^{i-1} \sum_{k=3}^{N-j+1} C_j^k, \quad \text{and} \\
N_i^S &= C_i^1.
\end{aligned}$$

Note that middle agents cannot exist at sites $i = 1$ and $i = N$ due to the choice of boundary conditions.

Similar to the approach of Johnston *et al.* [21], we consider how birth, death and movement events change the location and lengths of the chains and gaps, rather than the occupancy of an individual lattice site. This approach avoids making an assumption about the probability that a particular site is occupied or vacant, as the sites either side of a chain or gap are necessarily vacant or occupied, respectively. There are twenty different types of events that change either the location of the left-most agent in a chain, the length of a chain, or both. An example of each event is presented in Figure 9.1 for a particular configuration of agents. Events can have more than one potential outcome. For example, the potential outcomes of a single agent at site i undergoing a birth event can be classified into four groups. The daughter agent can be placed at site $i-1$ or $i+1$, and the gap that the daughter agent is placed in can be length one or greater. Here we detail each possible event and the subsequent change in configuration with respect to the number of chains and gaps. Events are referred to by a nomenclature describing the type of agent undergoing the event and the mechanism of the event itself, followed by a number highlighting the potential for multiple outcomes to arise from a specific event. For example, birth events for single agents are referred as SB events. Since there are four different types of SB events, we refer to these as SB1, SB2, SB3 and SB4 events. The details of each event are as follows:

Event SB1: A single agent at site i places a daughter agent at site $i-1$, where the gap that includes site $i-1$ is greater than length one. C_i^1 and G_j^{i-j} decrease, C_{i-1}^2 and G_j^{i-j-1} increase (Figure 9.1(b)).

Event SB2: A single agent at site i places a daughter agent at site $i+1$, where the gap that includes site $i+1$ is greater than length one. C_i^1 and G_{i+1}^j decrease, C_i^2 and G_{i+2}^{j-1} increase (Figure 9.1(c)).

Event SB3: A single agent at site i places a daughter agent at site $i-1$, where the gap that includes site $i-1$ is of length one. C_i^1, G_{i-1}^1 and C_j^{i-j-1} decrease, C_j^{i-j+1} increases (Figure 9.1(d)).

Event SB4: A single agent at site i places a daughter agent at site $i+1$, where the gap that includes site $i+1$ is of length one. C_i^1, G_{i+1}^1 and C_{i+2}^j decrease, C_i^{j+2} increases (Figure 9.1(e)).

Event EB1: An edge agent at site i places a daughter agent at site $i-1$, where the gap that includes site $i-1$ is greater than length one. C_i^j and G_k^{i-k} decrease, C_{i-1}^{j+1} and G_k^{i-k-1} increase, where $j \geq 2$ (Figure 9.1(f)).

Event EB2: An edge agent at site i places a daughter agent at site $i+1$, where the gap that includes site $i+1$ is greater than length one. C_j^{i-j+1} and G_{i+1}^k decrease, C_j^{i-j+2} and G_{i+2}^{k-1} increase, where $j \geq 2$ (Figure 9.1(g)).

Event EB3: An edge agent at site i places a daughter agent at site $i-1$, where the gap that includes site $i-1$ is of length one. C_i^j, G_{i-1}^1 and C_k^{i-k-1} decrease, C_k^{i+j-k} increases, where $j \geq 2$ (Figure 9.1(h)).

Event EB4: An edge agent at site i places a daughter agent at site $i+1$, where the gap that includes site $i+1$ is of length one. C_j^{i-j+1}, G_{i+1}^1 and C_{i+2}^k decrease, $C_j^{i-j+k+2}$ increases, where $j \geq 2$ (Figure 9.1(i)).

Event SM1: A single agent at site i moves to site $i-1$, where the gap that includes site $i-1$ is greater than length one. C_i^1, G_j^{i-j} and G_{i+1}^k decrease, C_{i-1}^1, G_j^{i-j-1} and G_i^{k+1} increase (Figure 9.1(j)).

Event SM2: A single agent at site i moves to site $i+1$, where the gap that includes site $i+1$ is greater than length one. C_i^1, G_j^{i-j} and G_{i+1}^k decrease, C_{i+1}^1, G_j^{i-j+1} and G_{i+2}^{k-1} increase (Figure 9.1(k)).

Event SM3: A single agent at site i moves to site $i-1$, where the gap that includes site $i-1$ is of length one. C_i^1, G_{i-1}^1, G_{i+1}^j and C_k^{i-k-1} decrease, C_k^{i-k} and G_i^{j+1} increase (Figure 9.1(l)).

Event SM4: A single agent at site i moves to site $i+1$, where the gap that includes site $i+1$ is of length one. C_i^1, G_{i+1}^1, G_j^{i-j} and C_{i+2}^k decrease, C_{i+1}^{k+1} and G_j^{i-j+1} increase (Figure 9.1(m)).

Event EM1: An edge agent at site i moves to site $i-1$, where the gap that includes site $i-1$ is greater than length one. C_i^j and G_k^{i-k} decrease, C_{i-1}^1, G_i^1, C_{i+1}^{j-1} and G_k^{i-k-1} increase, where $j \geq 2$ (Figure 9.1(n)).

Event EM2: An edge agent at site i moves to site $i+1$, where the gap that includes site $i+1$ is greater than length one. C_j^{i-j+1} and G_{i+1}^k decrease, C_{i+1}^1, G_i^1, C_j^{i-j} and G_{i+2}^{k-1} increase, where $j \geq 2$ (Figure 9.1(o)).

Event EM3: An edge agent at site i moves to site $i-1$, where the gap that includes site $i-1$ is of length one. C_i^j, G_{i-1}^1 and C_k^{i-k-1} decrease, C_{i+1}^{j-1}, G_i^1 and C_k^{i-k} increase, where $j \geq 2$ (Figure 9.1(p)).

Event EM4: An edge agent at site i moves to site $i+1$, where the gap that includes site $i+1$ is of length one. C_j^{i-j+1}, G_{i+1}^1 and C_{i+2}^k decrease, C_j^{i-j}, G_i^1 and C_{i+1}^{k+1} increase, where $j \geq 2$ (Figure 9.1(q)).

Event SD1: A single agent at site i dies. C_i^1, G_j^{i-j} and G_{i+1}^k decrease, $G_j^{i+k-j+1}$ increases (Figure 9.1(r)).

Event MD1: A middle agent at site i dies. C_j^k decreases, C_j^{i-j}, $C_{i+1}^{j+k-i-1}$ and G_i^1 increase, where $j \geq 3$ (Figure 9.1(s)).

Event ED1: An edge agent at site i dies, where site $i+1$ is occupied and site $i-1$ is vacant. C_i^j and G_k^{i-k} decrease, C_{i+1}^{j-1} and G_k^{i-k+1} increase, where $j \geq 2$ (Figure 9.1(t)).

Event ED2: An edge agent at site i dies, where site $i-1$ is occupied and site $i+1$ is vacant. C_j^{i-j+1} and G_{i+1}^k decrease, C_j^{i-j} and G_i^{k+1} increase, where $j \geq 2$ (Figure 9.1(u)).

To obtain expressions for the time rate of change of C_i^n and G_i^m, we consider the rate at which each event occurs at site i, and all possible results of each event. Birth events are always successful for single agents and, as such, occur at rate $P_p^1 C_i^1$. Similarly, movement events for single agents are always successful and are not influenced by agent-agent adhesion/repulsion. Therefore single agent movement events occur at rate $P_m^1 C_i^1$. Birth events for edge agents are, on average, unsuccessful half the time due to crowding. Hence, birth events occur at rates $P_p^n C_i^n/2$ for $n \in [2, N]$. Movement events for edge agents are influenced by both crowding and agent-agent adhesion/repulsion, and occur at rates $P_m^n C_i^n (1-\alpha)/2$ for $n \in [2, N]$. Neither birth or movement events can occur for middle agents. For all agent types, death events are not influenced by crowding and occur at rates

$P_d^n C_i^n$ for $n \in [1, N]$. Note that all rates are chain-length dependent. Practical examples that could be described using chain-length dependent rates arise in a variety of situations [17, 21]. For example, Hedayati *et al.* [17] demonstrate that nanoparticle-mediated heating causes cytotoxicity in prostate cancer cells, provided the volume of cells is above a threshold value. Furthermore, the cytotoxicity increases with the cell number. Our framework would be suitable for modelling this process, as we are able to impose death rates that are zero below a threshold length, and an increasing function with respect to length otherwise.

While the rates at which events occur for each mechanism and agent type combination are known, multiple events can occur for a mechanism and agent type combination. For example, there are four types of birth events for single agents and, as such, we require the proportion of single birth events that are SB1, SB2, SB3 and SB4 events. For a single birth event to be an SB1 event, the agent at site i must place a daughter agent at site $i-1$ and the gap that contains site $i-1$ must be of length two or greater. Note that the gap cannot include site i, as the selected agent occupies that site. The proportion of single birth events where the agent at site i selects a target site at $i-1$ is $1/2$. The proportion of single birth events where the agent selects a target site that is part of a gap of length two or greater is $\sum_{j=1}^{i-2} G_j^{i-j} / \sum_{j=1}^{i-1} G_j^{i-j}$, that is, the number of gaps including site $i-1$ that are of length two or greater divided by the total number of gaps including site $i-1$. Hence the rate at which SB1 events occur at site i, and subsequently decrease C_i^1 and increase C_{i-1}^2, is $(1/2)P_p^1 C_i^1 \sum_{j=1}^{i-2} G_j^{i-j} / \sum_{j=1}^{i-1} G_j^{i-j}$. SB1 events also decrease G_j^{i-j} and increase G_j^{i-j-1}, where $j \in [1, i-2]$. The proportion of SB1 events that change G_j^{i-j} and G_j^{i-j-1} for a specific $j \in [1, i-2]$ is $(1/2)P_p^1 C_i^1 G_j^{i-j} / \sum_{k=1}^{i-1} G_k^{i-k}$. Therefore, we can determine the expected rate of change for all chains and gaps affected by an SB1 event at site i. Following a similar process for all events we obtain transition rates for C_i^n and G_i^m, where $i \in [1, N]$, $n \in [1, N-i+1]$ and $m \in [1, N-i+1]$. The resulting system of ODEs is presented in Section 9.6.1 for C_i^n and Section 9.6.2 for G_i^m (Supplementary material).

9.2.1 Traditional mean-field descriptions

Traditional MF descriptions of lattice-based random walk models containing crowding, birth, death, movement and agent-agent adhesion do not have the flexibility to describe processes where the rate of birth, death and/or movement is arbitrarily chain-length dependent. However, with certain simplifying assumptions, MF descriptions of the discrete process can be derived [20, 44]. Here we examine two special cases of the discrete process, where continuum MF descriptions have been presented previously.

Special case 1: One simplifying assumption is that the birth, death and movement rates are independent of chain length. Hence we define $P_p = P_p^1 = P_p^2 = ... = P_p^N$, $P_d = P_d^1 = P_d^2 = ... = P_d^N$, and $P_m = P_m^1 = P_m^2 = ... = P_m^N$. There is, therefore, no positive or negative benefit associated with an agent being near other agents. Furthermore, there is no agent-agent adhesion/repulsion, and hence $\alpha = 0$. The MF description for this case takes the form of a reaction-diffusion equation, known as the Fisher-Kolmogorov model [9, 20, 28, 44],

$$\frac{\partial \rho(x,t)}{\partial t} = D\frac{\partial^2 \rho(x,t)}{\partial x^2} + (\lambda - K)\rho(x,t)(1-\rho(x,t)), \tag{9.1}$$

where

$$D = \lim_{\Delta \to 0} \frac{P_m \Delta^2}{2}, \ \lambda = P_p, \ K = P_d,$$

and $\rho(x,t)$ is the agent density [20, 44].

Special case 2: An alternative simplifying assumption is that birth, death and movement rates depend on whether an agent is isolated or not. An isolated agent has zero nearest-neighbours [20], corresponding to a chain of length one. All grouped agents, that is, agents with at least one nearest-neighbour, undergo birth, death and movement events at the same rate. Grouped agents correspond to agents that are part of a chain of length two or greater. Hence we define $P_p^G = P_p^2 = ... = P_p^N$, $P_d^G = P_d^2 = ... = P_d^N$, and $P_m^G = P_m^2 = ... = P_m^N$. This assumption introduces a potential positive or negative benefit associated with an agent being adjacent to other agents. For example, if $P_d^1 > P_d^G$ then isolated agents are more likely to die, compared to other agents, and hence there is a positive benefit associated with being adjacent to other agents. This allows for significant flexibility in modelling a variety of competitive or co-operative processes [20]. Again, there is no agent-agent adhesion/repulsion, and $\alpha = 0$. The MF description for this case is a reaction-diffusion equation with a nonlinear diffusivity function and Allee kinetics [20],

$$\frac{\partial \rho(x,t)}{\partial t} = \frac{\partial}{\partial x}\left(F(\rho(x,t))\frac{\partial \rho(x,t)}{\partial x}\right) + R(\rho(x,t)), \tag{9.2}$$

where

$$\begin{aligned} F(\rho(x,t)) &= D_I(1-4\rho(x,t)+3\rho(x,t)^2) + D_G(4\rho(x,t)-3\rho(x,t)^2), \\ R(\rho(x,t)) &= (\lambda_I - \lambda_G - K_I + K_G)\rho(x,t)(1-\rho(x,t))^2 \\ &\quad +\lambda_G \rho(x,t)(1-\rho(x,t)) - K_G \rho(x,t), \end{aligned}$$

and

$$\begin{aligned} D_I &= \lim\nolimits_{\Delta \to 0} \tfrac{P_m^1 \Delta^2}{2}, \ \lambda_I = P_p^1, \ K_I = P_d^1, \\ D_G &= \lim\nolimits_{\Delta \to 0} \tfrac{P_m^G \Delta^2}{2}, \ \lambda_G = P_p^G, \ K_G = P_d^G. \end{aligned}$$

While CMF descriptions have been presented for certain lattice-based random walk models, these descriptions are unsuitable for studying problems containing moving fronts as the system of governing ODEs is singular at zero agent density [3, 24, 42].

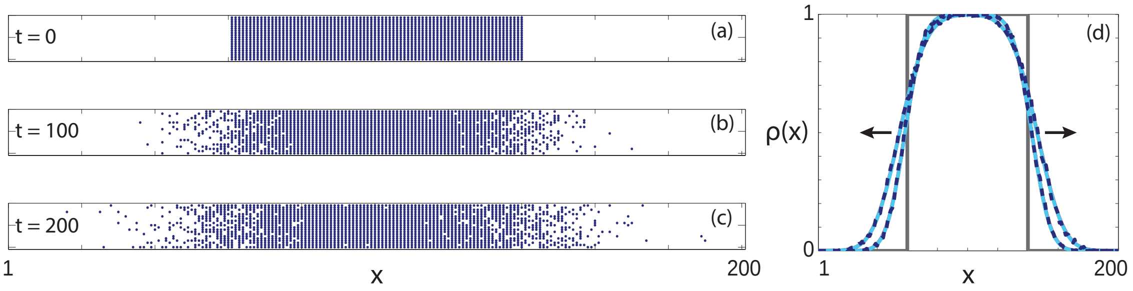

Figure 9.2: (a)-(c) Twenty identically-prepared realisations of the one-dimensional discrete model at (a) $t = 0$, (b) $t = 100$, and (c) $t = 200$, as indicated. (d) Comparisons of density profiles obtained from the averaged discrete model (blue, dashed), MF description (red, solid) and C&G description (cyan, solid). Note that the cyan and red curves are visually indistinguishable at this scale. All results are obtained using $P_m^i = 1$, $P_p^i = 0.005$, $P_d^i = 0$ for $i \in [1, 200]$, $N = 200$, $M = 10^3$ and $\Delta = 1$. These parameter choices correspond to Special case 1. Profiles are presented at $t = 100$ and $t = 200$, and the arrow indicates the direction of increasing time. Grey lines represent the initial condition.

9.3 Results

The solution of the MF model leads to a prediction of the average agent density profile as a function of position and time, whereas the C&G description provides the number of chains and gaps of all possible lengths as a function of location and time. Hence to compare the MF descriptions with the C&G description, it is necessary to reconstruct the agent density at each location, ρ_i, from C_i^n and G_i^m. The agent density at site i is the sum of all possible chains that include site i, namely,

$$\rho_i(t) = \sum_{j=1}^{i} \sum_{k=i-j+1}^{N-j+1} C_j^k. \tag{9.3}$$

An example of the output of the one-dimensional discrete model, illustrating twenty identically-prepared realisations, is presented in Figures 9.2(a)-(c) at $t = 0$, $t = 100$ and $t = 200$, respectively. Identically-prepared realisations refer to simulations of the discrete model performed with the same initial condition, parameter regime and boundary conditions. Initially, the domain is fully-occupied for $61 \leq x \leq 140$, and vacant otherwise. As time increases, the population spreads into the initially-vacant region. Note that Figures 9.2(a)-(c) each show twenty one-dimensional simulations, rather than one two-dimensional simulation. To obtain the average behaviour of the agent population, M identically-prepared realisations of the discrete model are performed and the binary lattice occupancy at each site i, $\hat{\rho}_i$, is calculated for each realisation. Note that the discrete model is simulated with the Gillespie algorithm [15]. The binary lattice occupancy is then averaged, giving $\rho_i = (1/M) \sum_M \hat{\rho}_i$. The averaged density profile from the discrete model is presented in Figure 9.2(d), with the numerical solutions to both Equation (9.1) and the C&G governing equations superimposed, at $t = 100$ and $t = 200$. Both continuum descriptions match the averaged discrete model predictions extremely well. Details of the numerical techniques used to solve the PDEs and systems of ODEs are presented in the Supplementary material.

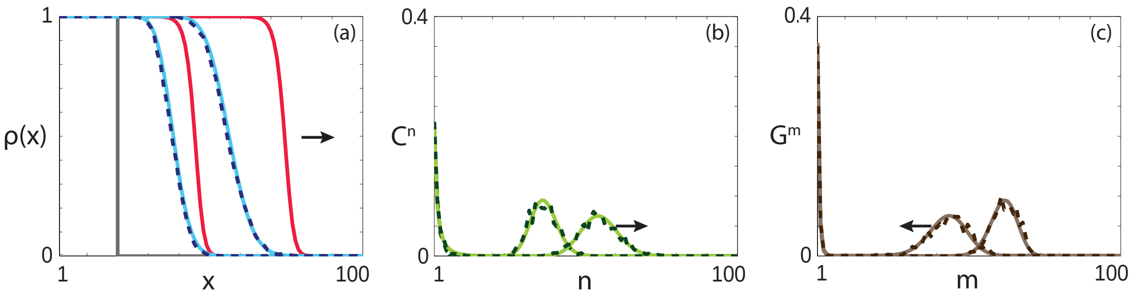

Figure 9.3: (a) Comparisons of density profiles obtained from the averaged discrete model (blue, dashed), MF description (red, solid) and C&G description (cyan, solid). (b) Comparison of the chain distribution obtained from the averaged discrete model (dark green, dashed) and C&G description (light green, solid). (c) Comparison of the gap distribution obtained from the averaged discrete model (dark brown, dashed) and C&G description (light brown, solid). All results are obtained using $P_m^i = 0.5$, $P_p^i = 0.5$, $P_d^i = 0$ for $i \in [1, 100]$, $N = 100$, $M = 10^3$ and $\Delta = 1$. These parameter choices correspond to Special case 1. Profiles are presented at $t = 100$ and $t = 200$, and the arrow indicates the direction of increasing time. Grey lines represent the initial condition.

The parameter regime considered in Figure 9.2 has $P_m/P_p \gg 1$ and $P_d = 0$, and, as such, we expect the MF description to approximate the average discrete behaviour well since this combination of parameters avoids the formation of significant clusters [42, 44]. We now consider a parameter regime where P_m/P_p is $\mathcal{O}(1)$. In such parameter regimes, the spatial correlations are significant and, subsequently, the MF description does not provide a valid approximation of the average discrete behaviour [21, 42]. To highlight this, a comparison between the average discrete behaviour and the numerical solution of Equation (9.1) is presented in Figure 9.3(a). For these results, the domain is initially fully-occupied for $x \leq 20$, and is vacant otherwise. The MF description predicts a front that is significantly sharper, and has a higher front speed, compared to the discrete model. In contrast, the numerical solution of the C&G description predicts the average discrete behaviour well, matching both the shape and position of the averaged discrete data. Furthermore, the distribution of chain and gap lengths, C^n and G^m, matches the observed average distribution of chains and gaps in the discrete model. As such, the C&G description provides an accurate estimate of the front shape and speed, as well as a valid prediction of the clustering of occupied and vacant sites in the system.

If the birth, death and movement rates depend on whether an agent has zero or at least one nearest-neighbour agent then the MF description of the discrete model is Equation (9.2) [20]. A comparison of the average discrete behaviour, the numerical solution of Equation (9.2) and the numerical solution of the C&G description in an appropriate parameter regime is presented in Figure 9.4(a). Interestingly, the MF description predicts that the agent population moves in the negative x direction, and would subsequently become extinct. In contrast, both the discrete model and the C&G description suggest that the agent population persists, and spreads in the positive x direction. Again, the C&G description matches the average discrete behaviour extremely well. The results in Figure 9.4(a) highlight the need for an accurate approximation. If the naive approach of implementing a MF approximation to describe the spread of an invasive species is taken, it

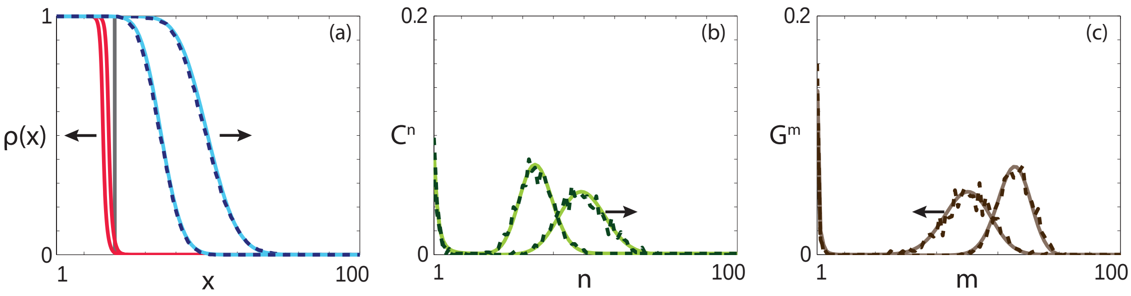

Figure 9.4: (a) Comparisons of density profiles obtained from the averaged discrete model (blue, dashed), MF description (red, solid) and C&G description (cyan, solid). (b) Comparison of the chain distribution obtained from the averaged discrete model (dark green, dashed) and C&G description (light green, solid). (c) Comparison of the gap distribution obtained from the averaged discrete model (dark brown, dashed) and C&G description (light brown, solid). All results are obtained using $P_m^1 = 0.5$, $P_p^1 = 0.4$, $P_d^1 = 0.7$, $P_m^i = 0.25$, $P_p^i = 0.3$, $P_d^i = 0$ for $i \in [2, 100]$, $N = 100$, $M = 10^3$ and $\Delta = 1$. These parameter choices correspond to Special case 2. Profiles are presented at $t = 50$ and $t = 100$, and the arrow indicates the direction of increasing time. Grey lines represent the initial condition.

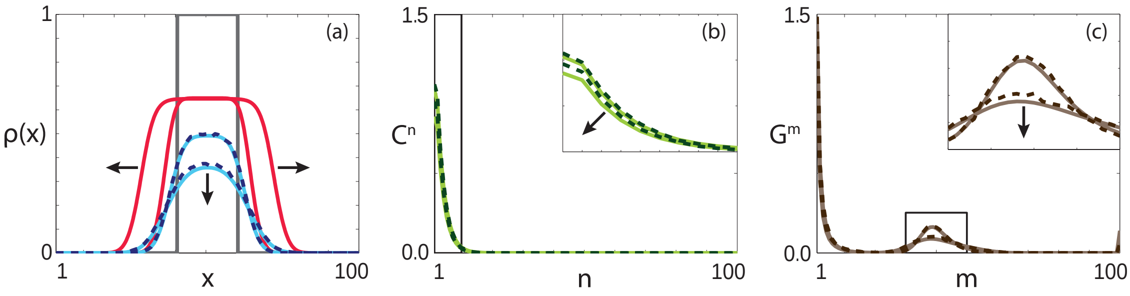

Figure 9.5: (a) Comparisons of density profiles obtained from the averaged discrete model (blue, dashed), MF description (red, solid) and C&G description (cyan, solid). (b) Comparison of the chain distribution obtained from the averaged discrete model (dark green, dashed) and C&G description (light green, solid). (c) Comparison of the gap distribution obtained from the averaged discrete model (dark brown, dashed) and C&G description (light brown, solid). All results are obtained using $P_m^1 = 0.5$, $P_p^1 = 0.45$, $P_d^1 = 0.3$, $P_m^i = 0.25$, $P_p^i = 0.3$, $P_d^i = 0.1$ for $i \in [2, 100]$, $N = 100$, $M = 10^4$, $\Delta = 1$. These parameter choices correspond to Special case 2. Profiles are presented at $t = 25$ and $t = 50$, and the arrow indicates the direction of increasing time. Insets highlight regions of particular interest. Grey lines represent the initial condition.

might be recommended that no culling measures are required to curtail the spread of the species. Obviously, such a recommendation is incorrect if the aim is to halt the invasion of the alien species. The clustering present in the system is highlighted in Figures 9.4(b)-(c), for chains and gaps, respectively. Intuitively, as time increases and the population spreads, the average chain length increases and the average gap length decreases. The C&G description predicts both the average chain and gap distributions in the discrete model well.

Introducing a non-zero death rate for chains of length two or greater reduces the carrying capacity in the MF description of the discrete model [20]. To determine whether this reduction is an accurate reflection of the average discrete behaviour, we present a comparison of the average discrete behaviour, and the numerical solutions of both Equation (9.2) and the C&G governing equations in Figure 9.5(a). Both the discrete model and

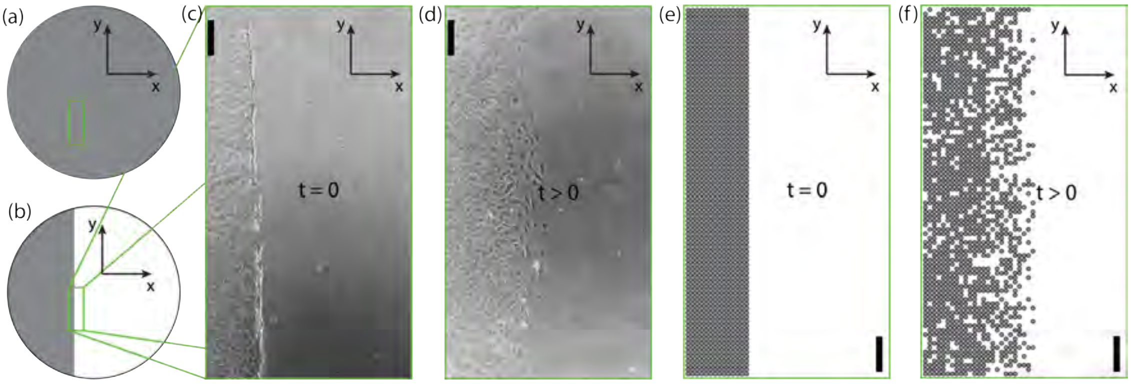

Figure 9.6: Schematic and experimental images of a scratch assay. (a) Schematic of a confluent cell population. (b) Schematic of a cell population after the scratch has been performed. Grey regions represent the confluent cell population and white regions represent the vacant area. (c)-(d) Experimental images of a 3T3 fibroblast scratch assay at (c) $t = 0$, and (d) $t > 0$. (e)-(f) Schematic of a mathematical model of a scratch assay at (e) $t = 0$, and (f) $t > 0$. Scale bar corresponds to 200 μm.

the C&G description predict that the peak agent density near $x = 50$ decreases between $t = 25$ and $t = 50$, whereas the MF description predicts that the peak agent density at this location is approximately constant, at $\rho = 0.646$ [20], after $t = 25$. Interestingly, both the discrete model and the C&G description predict that the population eventually goes extinct. In contrast, the MF description predicts that the population persists and spreads throughout the domain. Again, these results highlight the importance of implementing an accurate approximation to obtain meaningful conclusions, as well as the robust nature of the C&G description.

Note that all results presented here have been performed with $\alpha = 0$, and hence no agent-agent adhesion/repulsion. Simulations performed with $\alpha \neq 0$ (not presented) confirm that the C&G description accurately predicts the average discrete behaviour in all cases, even with strong agent-agent adhesion/repulsion.

9.4 Experimental case study

To highlight the insight provided about spatial clustering by the C&G description, we consider a case study motivated by a scratch assay. Scratch assays are widely used to observe the collective behaviour of a cell population in response to a model wound [30]. We present a schematic representation and experimental images of a scratch assay in Figure 9.6. In a scratch assay, a cell population is placed on a dish and allowed to grow to confluence. A portion of the cell population is then removed, and the remaining cells spread, through a combination of migration and proliferation, into the newly-vacant space [30]. A schematic representation of the confluent population before and after the scratch is performed is presented in Figures 9.6(a)-(b), with a typical experimental field of view highlighted in green. To mimic the geometry of this experiment, we consider $1 \leq x \leq 100$ and initially set $\rho(x) = 1$ for $x \leq 30$ and $\rho(x) = 0$, otherwise. Note that scratch assays are two-dimensional processes, as observed in the two experimental images of a scratch assay

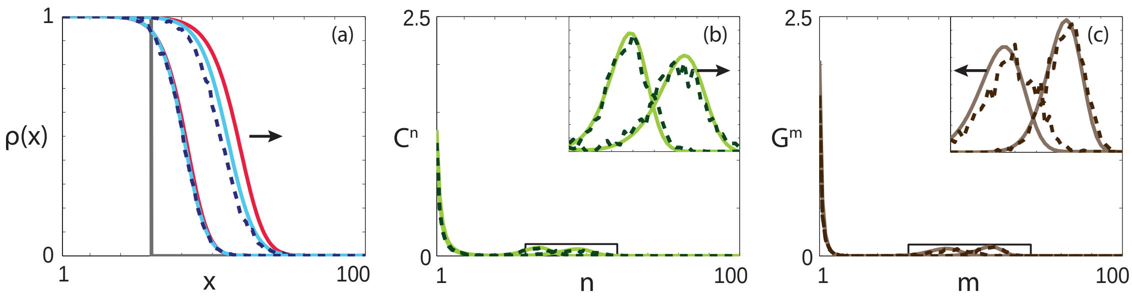

Figure 9.7: (a) Comparisons of density profiles obtained from the averaged discrete model (blue, dashed), MF description (red, solid) and C&G description (cyan, solid). (b) Comparison of the chain distribution obtained from the averaged discrete model (dark green, dashed) and C&G description (light green, solid). (c) Comparison of the gap distribution obtained from the averaged discrete model (dark brown, dashed) and C&G description (light brown, solid). All results are obtained using $P_m^i = 0.66$, $P_p^i = 0.056$, $P_d^i = 0$ for $i \in [1, 100]$, $N = 100$, $M = 10^3$, $\Delta = 1$. These parameter choices correspond to 3T3 fibroblast cells. Profiles are presented at $t = 75$ and $t = 150$, and the arrow indicates the direction of increasing time. Insets highlight regions of particular interest. Grey lines represent the initial condition.

for a 3T3 fibroblast population in Figures 9.6(c)-(d), and the corresponding schematics for this experiment in Figures 9.6(e)-(f). Full experimental details are given in [26].

As the experiment is approximately spatially-uniform in one direction, we can approximate the scratch assay with a one-dimensional model [23]. We consider two cell populations where parameter estimates for the cell motility and the cell proliferation rate have been presented previously: 3T3 fibroblast cells and MDA MB 321 breast cancer cells [43]. The investigation performed by Simpson *et al.* [43] resulted in parameter estimates of $P_p^i = 0.056$ h^{-1}, $P_m^i = 0.66$ h^{-1} and $P_d^i = 0$ h^{-1} for all i for 3T3 cells, and $P_p^i = 0.069$ h^{-1}, $P_m^i = 0.04$ h^{-1}, $P_d^i = 0$ h^{-1} for all i for 231 cells. Note that the ratio P_p^i/P_m^i is approximately one order of magnitude higher for the breast cancer cells compared to the fibroblasts, which implies that the spatial correlations between breast cancer cells will be more significant [42].

For the numerical solution corresponding to the 3T3 cell population, presented in Figure 9.7, both the MF description and the C&G description approximate the average discrete behaviour reasonably well. However, the C&G description provides additional information regarding the clustering present within the migrating cell population. The chain distribution, presented in Figure 9.7(b), suggests that the 3T3 population does not form significant clusters, as the majority of the chains are short length.

In contrast, for the numerical solution corresponding to the 231 cell population, presented in Figure 9.8, the C&G description accurately approximates the average discrete behaviour, while the MF description does not. This is result is intuitive as we observe that there is significantly more clustering present in the system, compared to the numerical solution corresponding to the 3T3 cell population. That is, the chain distribution in Figure 9.8(b) contains significantly fewer chains of short length, compared to the chain distribution in Figure 9.7(b). For example, at the times shown, C^1 is approximately 25 times higher in the 3T3 cell population, compared to the 231 cell population. Critically,

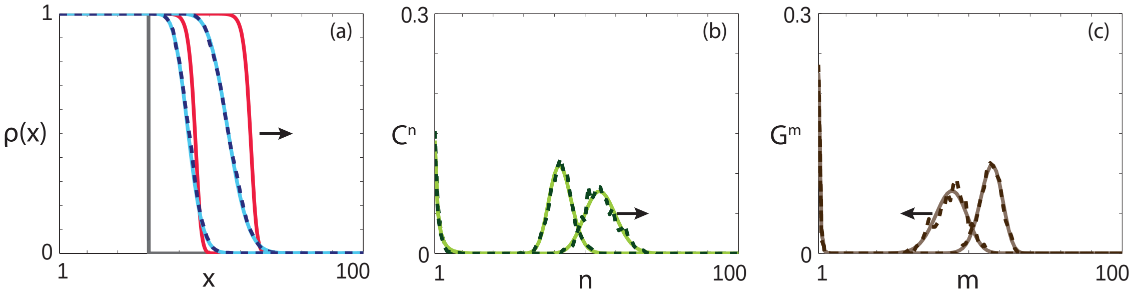

Figure 9.8: (a) Comparisons of density profiles obtained from the averaged discrete model (blue, dashed), MF description (red, solid) and C&G description (cyan, solid). (b) Comparison of the chain distribution obtained from the averaged discrete model (dark green, dashed) and C&G description (light green, solid). (c) Comparison of the gap distribution obtained from the averaged discrete model (dark brown, dashed) and C&G description (light brown, solid). All results are obtained using $P_m^i = 0.04$, $P_p^i = 0.069$, $P_d^i = 0$ for $i \in [1, 100]$, $N = 100$, $M = 10^3$, $\Delta = 1$. These parameter choices correspond MDA MB 321 breast cancer cells. Profiles are presented at $t = 250$ and $t = 500$, and the arrow indicates the direction of increasing time. Grey lines represent the initial condition.

	Fig. 2	Fig. 3	Fig. 4	Fig. 5	Fig. 6	Fig. 7
Discrete model (single realisation)	3.42	1.13	0.22	0.11	0.37	0.27
Discrete model (1000 realisations)	2838.1	954.4	134.1	39.1	313.6	177.9
C&G description	601.3	52.3	25.2	34.7	27.1	27.6

Table 9.1: Time in seconds taken to perform: (i) a single realisation of the discrete model; (ii) 1000 realisations of the discrete model, and (iii) a numerical solution of the C&G system of equations for the parameter values in Figures 9.2-9.7. All solutions are obtained using a single 3.0 GHz Intel i7-3540M desktop processor. Note that the computation time for 1000 realisations is lower than performing 1000 repeats of a single realisation due to the time associated with initial set-up.

the C&G description accurately predicts the experimental observation that 231 cells form clusters while 3T3 cells do not [43]. Specifically, during monolayer formation, 3T3 cells form an approximately spatially uniform monolayer while 231 cells develop into clusters [43].

9.5 Discussion and conclusions

Processes that involve moving fronts are common in cell biology and ecology [1, 6, 13, 14, 26, 32, 37, 38, 41, 46], and lattice-based random walks are widely employed to describe these processes [7, 18, 19, 26, 27, 44, 45]. Due to the stochastic nature of random walks, it can be computationally intractable to perform sufficiently many realisations of a random walk model to obtain average behaviour that is not dominated by fluctuations. Furthermore, it is difficult to extract meaningful information about population behaviour through analysis of the random walk. The standard approach to overcome these issues is to derive a MF description of the random walk [8, 27, 44]. However, this approach relies on the assumption that any spatial correlations within the random walk are weak [3, 42]. CMF descriptions that account for the spatial correlations have been proposed [3, 31, 42].

Unfortunately, these descriptions are not applicable to problems involving moving fronts as the ODEs governing the CMF description are singular in regions where the density of agents is zero [3, 42].

Here we develop and present an accurate continuum description for moving fronts associated with lattice-based random walks that contain crowding, birth, death, movement and agent-agent adhesion. We consider processes that are spatially variable, and include birth, death and movement rates that are chain-length dependent. Our C&G description provides predictions that match the average behaviour of the discrete model well in all parameter regimes. In contrast, the MF description is less flexible in terms of the birth, death and movement rates and only provides a valid approximation of the average discrete behaviour in extremely limited parameter regimes. Furthermore, for all cases considered in this work, the C&G description requires less computation time than performing 1000 realisations of the discrete model. A comparison between the time taken to perform a single realisation of the discrete model, 1000 realisations of the discrete model, and to obtain the numerical solution to the C&G system of equations is presented in Table 9.1.

For the special case where the rates of birth, death and movement are independent of the chain length, the MF description correctly predicts the persistence of the population but inaccurately predicts the front velocity. For the special case where the rates of birth, death and movement are different depending on whether the agents are part of a chain of length one, or are part of a chain of length two or greater, the MF description predicts persistence when the population becomes extinct, and predicts extinction when the population persists. In both these cases the C&G description accurately predicts the front velocity, and the persistence or extinction of the population. Furthermore, the C&G description provides information about the spatial clustering of both occupied and unoccupied sites, and the clustering predictions approximate the clustering observed in the discrete model accurately.

The work presented here could be extended in several ways. The influence of local spatial structure on the persistence of species is a key question in ecology [4, 18, 29, 35]. The C&G description provides an explicit estimate of the spatial clustering of both agents and unoccupied space. Therefore, it would be instructive to apply the C&G description to ecological processes to obtain insight into the clustering present in the system for parameter regimes where the agent population becomes extinct. Another approach would be to investigate a truncated system of governing equations, where there is a maximum chain or gap length. If there is prior knowledge about the long-time density of the system, an assumption could be made that chains or gaps above a threshold length could be neglected, and hence the system of equations could be truncated. This truncation would reduce the computational cost associated with obtaining a numerical solution to the governing equations. It would be insightful to examine the trade-off between the reduction in computational cost and the decreased accuracy caused by the truncation. Alternatively, the C&G description presented here could be calibrated to experimental data from the cell biology literature. For example, lattice-based random walks have been calibrated to *in vitro* cell biology experiments to obtain estimates of cell diffusivity and cell proliferation rates [26, 45]. However, the calibration of stochastic models to experimental data is computationally expensive [22, 26]. As the C&G description accurately approximates the average random walk behaviour in all parameter regimes, it would be instructive to determine whether similar cell diffusivity and cell proliferation rates could be obtained from calibration of the deterministic C&G description to experimental data, and to quantify the reduction in computation time to obtain the parameter estimates. However, we leave these extensions for future consideration.

Acknowledgements

We appreciate the assistance provided by Emeritus Professor Sean McElwain. This work is supported by the Australian Research Council (FT130100148, DP140100249). We thank two referees and the editor for their helpful comments.

Bibliography

[1] Alvord Jr, E. C. and Shaw, C. M. [1991], *Neoplasms affecting the nervous system of the elderly*, Oxford University Press.

[2] Anguige, K. and Schmeiser, C. [2009], ‘A one-dimensional model of cell diffusion and aggregation, incorporating volume filling and cell-to-cell adhesion’, *Journal of Mathematical Biology* **58**(3), 395–427.

[3] Baker, R. E. and Simpson, M. J. [2010], ‘Correcting mean-field approximations for birth-death-movement processes’, *Physical Review E* **82**(4), 041905.

[4] Brown, R. G., James, A. F., Pitchford, J. W. and Plank, M. J. [2012], ‘Habitat fragmentation: simple models for local persistence and the spread of invasive species’, *Journal of Theoretical Biology* **310**, 231–238.

[5] Chowdhury, D., Schadschneider, A. and Nishinari, K. [2005], ‘Physics of transport and traffic phenomena in biology: from molecular motors and cells to organisms’, *Physics of Life Reviews* **2**(4), 318–352.

[6] Clavero, M. and García-Berthou, E. [2005], ‘Invasive species are a leading cause of animal extinctions’, *Trends in Ecology and Evolution* **20**(3), 110–110.

[7] Deroulers, C., Aubert, M., Badoual, M. and Grammaticos, B. [2009], ‘Modeling tumor cell migration: from microscopic to macroscopic models’, *Physical Review E* **79**(3), 031917.

[8] Fernando, A. E., Landman, K. A. and Simpson, M. J. [2010], ‘Nonlinear diffusion and exclusion processes with contact interactions’, *Physical Review E* **81**(1), 011903.

[9] Fisher, R. A. [1937], ‘The wave of advance of advantageous genes’, *Annals of Eugenics* **7**(4), 355–369.

[10] Fort, J. and Pujol, T. [2007], ‘Time-delayed fronts from biased random walks’, *New Journal of Physics* **9**(7), 234.

[11] Fort, J. and Sole, R. V. [2013], 'Accelerated tumor invasion under non-isotropic cell dispersal in glioblastomas', *New Journal of Physics* **15**(5), 055001.

[12] Gabriely, G., Wurdinger, T., Kesari, S., Esau, C. C., Burchard, J., Linsley, P. S. and Krichevsky, A. M. [2008], 'MicroRNA 21 promotes glioma invasion by targeting matrix metalloproteinase regulators', *Molecular and Cellular Biology* **28**(17), 5369–5380.

[13] Gerlee, P. and Nelander, S. [2012], 'The impact of phenotypic switching on glioblastoma growth and invasion', *PLOS Computational Biology* **8**(6), e1002556.

[14] Gerlee, P. and Nelander, S. [2016], 'Travelling wave analysis of a mathematical model of glioblastoma growth', *Mathematical Biosciences* **276**, 75–81.

[15] Gillespie, D. T. [1977], 'Exact stochastic simulation of coupled chemical reactions', *Journal of Physical Chemistry* **81**(25), 2340–2361.

[16] Grima, R. [2008], 'Multiscale modeling of biological pattern formation', *Current Topics in Developmental Biology* **81**, 435–460.

[17] Hedayati, M., Thomas, O., Abubaker-Sharif, B., Zhou, H., Cornejo, C., Zhang, Y., Wabler, M., Mihalic, J., Gruettner, C., Westphal, F. et al. [2013], 'The effect of cell cluster size on intracellular nanoparticle-mediated hyperthermia: is it possible to treat microscopic tumors?', *Nanomedicine* **8**(1), 29–41.

[18] Hiebeler, D. E., Houle, J., Drummond, F., Bilodeau, P. and Merckens, J. [2016], 'Locally dispersing populations in heterogeneous dynamic landscapes with spatiotemporal correlations. i. block disturbance', *Journal of Theoretical Biology* **407**, 212–224.

[19] Illien, P., Bénichou, O., Mejía-Monasterio, C., Oshanin, G. and Voituriez, R. [2013], 'Active transport in dense diffusive single-file systems', *Physical Review Letters* **111**(3), 038102.

[20] Johnston, S. T., Baker, R. E., McElwain, D. L. S. and Simpson, M. J. [2016], 'Co-operation, competition and crowding: a discrete framework linking Allee kinetics, nonlinear diffusion, shocks and sharp-fronted travelling waves', *bioRxiv* p. 077743.

[21] Johnston, S. T., Baker, R. E. and Simpson, M. J. [2016], 'Filling the gaps: A robust description of adhesive birth-death-movement processes', *Physical Review E* **93**(4), 042413.

[22] Johnston, S. T., Ross, J. V., Binder, B. J., McElwain, D. L. S., Haridas, P. and Simpson, M. J. [2016], 'Quantifying the effect of experimental design choices for in vitro scratch assays', *Journal of Theoretical Biology* **400**, 19–31.

[23] Johnston, S. T., Shah, E. T., Chopin, L. K., McElwain, D. L. S. and Simpson, M. J. [2015], 'Estimating cell diffusivity and cell proliferation rate by interpreting IncuCyte ZOOM assay data using the Fisher-Kolmogorov model', *BMC Systems Biology* **9**(38), 38.

[24] Johnston, S. T., Simpson, M. J. and Baker, R. E. [2012], 'Mean-field descriptions of collective migration with strong adhesion', *Physical Review E* **85**(5), 051922.

[25] Johnston, S. T., Simpson, M. J. and Baker, R. E. [2015], 'Modelling the movement of interacting cell populations: a moment dynamics approach', *Journal of Theoretical Biology* **370**, 81–92.

[26] Johnston, S. T., Simpson, M. J. and McElwain, D. L. S. [2014], 'How much information can be obtained from tracking the position of the leading edge in a scratch assay?', *Journal of the Royal Society Interface* **11**(97), 20140325.

[27] Khain, E., Katakowski, M., Charteris, N., Jiang, F. and Chopp, M. [2012], 'Migration of adhesive glioma cells: front propagation and fingering', *Physical Review E* **86**(1), 011904.

[28] Kolmogorov, A. N., Petrovsky, I. G. and Piskunov, N. S. [1937], 'Étude de léquation de la diffusion avec croissance de la quantité de matiere et son application un probleme biologique', *Moscow University Mathematics Bulletin* **1**, 1–25.

[29] Law, R., Purves, D. W., Murrell, D. J. and Dieckmann, U. [2001], 'Causes and effects of small-scale spatial structure in plant populations', *British Ecological Society* **14**, 21–44.

[30] Liang, C.-C., Park, A. Y. and Guan, J.-L. [2007], 'In vitro scratch assay: a convenient and inexpensive method for analysis of cell migration in vitro', *Nature Protocols* **2**(2), 329–333.

[31] Markham, D. C., Simpson, M. J. and Baker, R. E. [2013], 'Simplified method for including spatial correlations in mean-field approximations', *Physical Review E* **87**(6), 062702.

[32] Martin, P. [1997], 'Wound healing–aiming for perfect skin regeneration', *Science* **276**(5309), 75–81.

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

[34] Mungomery, R. [1935], 'The giant american toad (bufo marinus)', *Cane Growers Quarterly Bulletin* **3**, 21–27.

[35] Murrell, D. J. [2010], 'When does local spatial structure hinder competitive coexistence and reverse competitive hierarchies?', *Ecology* **91**(6), 1605–1616.

[36] Nnetu, K. D., Knorr, M., Käs, J. and Zink, M. [2012], 'The impact of jamming on boundaries of collectively moving weak-interacting cells', *New Journal of Physics* **14**(11), 115012.

[37] Phillips, B. L., Brown, G. P., Greenlees, M., Webb, J. K. and Shine, R. [2007], 'Rapid expansion of the cane toad (bufo marinus) invasion front in tropical australia', *Austral Ecology* **32**(2), 169–176.

[38] Poujade, M., Grasland-Mongrain, E., Hertzog, A., Jouanneau, J., Chavrier, P., Ladoux, B., Buguin, A. and Silberzan, P. [2007], 'Collective migration of an epithelial monolayer in response to a model wound', *Proceedings of the National Academy of Sciences* **104**(41), 15988–15993.

[39] Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. [2007], *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, Cambridge University Press.

[40] Sharkey, K. J., Fernandez, C., Morgan, K. L., Peeler, E., Thrush, M., Turnbull, J. F. and Bowers, R. G. [2006], 'Pair-level approximations to the spatio-temporal

dynamics of epidemics on asymmetric contact networks', *Journal of Mathematical Biology* **53**(1), 61–85.

[41] Sherratt, J. A., Lewis, M. A. and Fowler, A. C. [1995], 'Ecological chaos in the wake of invasion.', *Proceedings of the National Academy of Sciences* **92**(7), 2524–2528.

[42] Simpson, M. J. and Baker, R. E. [2011], 'Corrected mean-field models for spatially dependent advection-diffusion-reaction phenomena', *Physical Review E* **83**(5), 051922.

[43] Simpson, M. J., Binder, B. J., Haridas, P., Wood, B. K., Treloar, K. K., McElwain, D. L. S. and Baker, R. E. [2013], 'Experimental and modelling investigation of monolayer development with clustering', *Bulletin of Mathematical Biology* **75**(5), 871–889.

[44] Simpson, M. J., Landman, K. A. and Hughes, B. D. [2010], 'Cell invasion with proliferation mechanisms motivated by time-lapse data', *Physica A* **389**(18), 3779–3790.

[45] Simpson, M. J., Treloar, K. K., Binder, B. J., Haridas, P., Manton, K. J., Leavesley, D. I., McElwain, D. L. S. and Baker, R. E. [2013], 'Quantifying the roles of cell motility and cell proliferation in a circular barrier assay', *Journal of the Royal Society Interface* **10**(82), 20130007.

[46] Swanson, K. R., Bridge, C., Murray, J. and Alvord Jr, E. C. [2003], 'Virtual and real brain tumors: using mathematical modeling to quantify glioma growth and invasion', *Journal of the Neurological Sciences* **216**(1), 1–10.

[47] Townsend, C. R. [1996], 'Invasion biology and ecological impacts of brown trout *salmo trutta* in New Zealand', *Biological Conservation* **78**(1), 13–22.

[48] Turner, S., Sherratt, J. A., Painter, K. J. and Savill, N. J. [2004], 'From a discrete to a continuous model of biological cell movement', *Physical Review E* **69**(2), 021910.

[49] Wilkinson, R. R. and Sharkey, K. J. [2014], 'Message passing and moment closure for susceptible-infected-recovered epidemics on finite networks', *Physical Review E* **89**(2), 022808.

9.6 Supplementary material

9.6.1 Governing equations

Chains

Here we present the time rate of change for each chain length and location, obtained by considering the potential outcomes of each type of birth, death and movement event.

$$
\begin{aligned}
\frac{\mathrm{d}C_1^1}{\mathrm{d}t} =& \frac{1}{2}\left[-P_p^1 C_1^1 - \sum_{i=1}^{N-2} P_p^1 Y(C_3^i, G_2^1, C_1^1)\right] + \left[\sum_{i=2}^{N}\left\{P_d^i C_1^i\right\} - P_d^1 C_1^1\right] \\
&+ \frac{P_m^1}{2}\left[-C_1^1 + C_2^1 - Y(C_3^1, G_2^1, C_1^1)\right] \\
&+ \frac{1-\alpha}{2}\left[P_m^2 C_1^2 + \sum_{i=2}^{N-1}\left\{P_m^i C_2^i\right\} - \sum_{i=2}^{N-2} P_m^i Y(C_3^i, G_2^1, C_1^1)\right],
\end{aligned}
$$

$$
\begin{aligned}
\frac{\mathrm{d}C_2^1}{\mathrm{d}t} =& \frac{1}{2}\left[-2P_p^1 C_2^1 - \sum_{i=1}^{N-3} P_p^i Y(C_4^i, G_3^1, C_2^1)\right] + \left[\sum_{i=2}^{N-1}\left\{P_d^i C_2^i\right\} + P_d^2 C_1^2 - P_d^1 C_2^1\right] \\
&+ \frac{P_m^1}{2}\left[C_1^1 - Y(C_1^1, G_2^1) - 2C_2^1 + C_3^1 - Y(C_3^1, G_2^1) - Y(C_4^1, G_3^1, C_2^1)\right] \\
&+ \frac{1-\alpha}{2}\left[P_m^2 C_2^2 + \sum_{i=2}^{N-2}\left\{P_m^i C_3^i - P_m^i Y(C_3^i, G_2^1)\right\} - \sum_{i=2}^{N-3} P_m^i Y(C_4^i, G_3^1, C_2^1)\right],
\end{aligned}
$$

$$
\begin{aligned}
\frac{\mathrm{d}C_k^1}{\mathrm{d}t} =& \frac{1}{2}\left[-2P_p^1 C_k^1 - \sum_{i=1}^{N-k-1} P_p^i Y(C_{k+2}^i, G_{k+1}^1, C_k^1) - \sum_{i=1}^{k-2} P_p^i Y(C_{k-i-1}^i, G_{k-1}^1, C_k^1)\right] \\
&+ \left[\sum_{i=2}^{k+1}\left\{P_d^i C_k^i\right\} + \sum_{i=2}^{k}\left\{P_d^i C_{k-i+1}^i\right\} - P_d^1 C_k^1\right] \\
&+ \frac{P_m^1}{2}\Bigg[-2C_k^1 + C_{k-1}^1 - Y(C_{k-1}^1, G_1^1) + C_{k+1}^1 - Y(C_{k+1}^1, G_k^1) \\
&- Y(C_{k-2}^1, G_{k-1}^1, C_k^1) - Y(C_{k+2}^1, G_{k+1}^1, C_k^1)\Bigg] \\
&+ \frac{1-\alpha}{2}\Bigg[P_m^2 C_k^2 + P_m^2 C_{k-1}^2 + \sum_{i=2}^{k-1}\left\{P_m^i C_{k-i}^i - P_m^i Y(C_{k-i}^i, G_k^1)\right\} \\
&+ \sum_{i=2}^{N-k}\left\{P_m^i C_{k+1}^i - P_m^i Y(C_{k+1}^i, G_k^1)\right\} - \sum_{i=2}^{k-2} P_m^i Y(C_{k-i-1}^i, G_{k-1}^1, C_k^1) \\
&- \sum_{i=2}^{N-k-1} P_m^i Y(C_{k+2}^i, G_{k+1}^1, C_k^1)\Bigg], \quad k = 3, \ldots, N-2,
\end{aligned}
$$

$$\begin{aligned}
\frac{\mathrm{d}C^1_{N-1}}{\mathrm{d}t} =& \frac{1}{2}\left[-2P^1_pC^1_{N-1} - \sum_{i=1}^{N-3} P^i_pY(C^i_{N-i-2}, G^1_{N-2}, C^1_{N-1})\right] \\
&+ \left[P^2_dC^2_{N-1} + \sum_{i=2}^{N-1}\left\{P^i_dC^i_{N-i}\right\} - P^1_dC^1_{N-1}\right] \\
&+ \frac{P^1_m}{2}\Bigg[C^1_N - Y(C^1_N, G^1_{N-1}) - 2C^1_{N-1} + C^1_{N-2} - Y(C^1_{N-2}, G^1_{N-1}) \\
&- Y(C^1_{N-3}, G^1_{N-2}, C^1_{N-1})\Bigg] \\
&+ \frac{1-\alpha}{2}\Bigg[P^2_mC^2_{N-2} + \sum_{i=2}^{N-2} P^i_m\Big\{C^i_{N-i-1} - Y(C^i_{N-i-1}, G^1_{N-1})\Big\} \\
&- \sum_{i=2}^{N-3} P^i_mY(C^i_{N-i-2}, G^1_{N-2}, C^1_{N-1})\Bigg],
\end{aligned}$$

$$\begin{aligned}
\frac{\mathrm{d}C^1_{N}}{\mathrm{d}t} =& \frac{1}{2}\left[-P^1_pC^1_N - \sum_{i=1}^{N-2} P^i_pY(C^i_{N-i-1}, G^1_{N-1}, C^1_N)\right] + \left[\sum_{i=2}^{N} P^i_d\Big\{C^i_{N-i+1}\Big\} - P^1_dC^1_N\right] \\
&+ \frac{P^1_m}{2}\left[-C^1_N + C^1_{N-1} - Y(C^1_{N-2}, G^1_{N-1}, C^1_N)\right] \\
&+ \frac{1-\alpha}{2}\left[P^2_mC^2_{N-1} + \sum_{i=2}^{N-1} P^i_m\Big\{C^i_{N-i}\Big\} - \sum_{i=2}^{N-2} P^i_mY(C^i_{N-i-1}, G^1_{N-1}, C^1_N)\right],
\end{aligned}$$

$$\begin{aligned}
\frac{\mathrm{d}C^L_1}{\mathrm{d}t} =& \frac{1}{2}\Bigg[P^{L-1}_pC^{L-1}_1 + P^{L-1}_pC^{L-1}_2 - P^{L-1}_pY(C^{L-1}_1, G^1_L) - P^L_pC^L_1 \\
&- \sum_{i=1}^{N-L-1} P^i_pY(C^i_{L+2}, G^1_{L+1}, C^L_1) + \sum_{i=1}^{L-2} P^i_pY(C^i_1, G^1_{i+1}, C^{L-i-1}_{i+2}) \\
&+ \sum_{i=1}^{L-2} P^i_pY(C^i_{L-i+1}, G^1_{L-i}, C^{L-i-1}_1)\Bigg] \\
&+ \left[\sum_{i=L+1}^{N} P^i_d\Big\{C^i_1\Big\} - P^L_dLC^L_1\right] + \frac{P^1_m}{2}\left[Y(C^1_{L+1}, G^1_L, C^{L-1}_1) - Y(C^1_{L+2}, G^1_{L+1}, C^L_1)\right] \\
&+ \frac{1-\alpha}{2}\Bigg[P^{L+1}_mC^{L+1}_1 - C^L_1 + \sum_{i=2}^{N-L} P^i_mY(C^i_{L+1}, G^1_L, C^{L-1}_1) \\
&- \sum_{i=2}^{N-L-1} P^i_mY(C^i_{L+2}, G^1_{L+1}, C^L_1)\Bigg], \quad L = 2, \dots, N-2,
\end{aligned}$$

$$\begin{aligned}
\frac{\mathrm{d}C_2^L}{\mathrm{d}t} ={}& \frac{1}{2}\Bigg[P_p^{L-1}C_2^{L-1} + P_p^{L-1}C_3^{L-1} - 2P_p^L C_2^L - P_p^{L-1}Y(C_2^{L-1}, G_{L+1}^1) \\
& - P_p^{L-1}Y(C_3^{L-1}, G_2^1) + \sum_{i=1}^{L-2} P_p^i Y(C_2^i, G_{i+2}^1, C_{i+3}^{L-i-1}) \\
& + \sum_{i=1}^{L-2} P_p^i Y(C_{L-i+2}^i, G_{L-i+1}^1, C_2^{L-i-1}) - \sum_{i=1}^{N-4} P_p^i Y(C_{L+3}^i, G_{L+2}^1, C_2^L)\Bigg] \\
& + \Bigg[\sum_{i=L+1}^{N-1} P_d^i\Big\{C_2^i\Big\} + P_d^{L+1}C_1^{L+1} - P_d^L L C_2^L\Bigg] \\
& + \frac{P_m^1}{2}\Bigg[Y(C_1^1, G_2^1, C_3^{L-1}) + Y(C_{L+2}^1, G_{L+1}^1, C_2^{L-1}) - Y(C_{L+3}^1, G_{L+2}^1, C_2^L)\Bigg] \\
& + \frac{1-\alpha}{2}\Bigg[P_m^{L+1}C_2^{L+1} - 2P_m^L C_2^L + \sum_{i=2}^{N-L-1} P_m^i Y(C_{L+2}^i, G_{L+1}^1, C_2^{L-1}) \\
& - \sum_{i=2}^{N-L-2} P_m^i Y(C_{L+3}^i, G_{L+2}^1, C_2^L)\Bigg], \quad L = 2, \ldots, N-3,
\end{aligned}$$

$$\begin{aligned}
\frac{\mathrm{d}C_k^L}{\mathrm{d}t} ={}& \frac{1}{2}\Bigg[P_p^{L-1}C_k^{L-1} + P_p^{L-1}C_{k+1}^{L-1} - 2P_p^L C_k^L - P_p^{L-1}Y(C_k^{L-1}, G_{L+k-1}^1) \\
& - P_p^{L-1}Y(C_{k+1}^{L-1}, G_k^1) + \sum_{i=1}^{L-2} P_p^i Y(C_k^i, G_{k+i}^1, C_{k+i+1}^{L-i-1}) \\
& + \sum_{i=1}^{L-2} P_p^i Y(C_{L+k-i}^i, G_{L+k-i-1}^1, C_k^{L-i-1}) - \sum_{i=1}^{k-2} P_p^i Y(C_{k-i-1}^i, G_{k-1}^1, C_k^L) \\
& - \sum_{i=1}^{N-L-k} P_p^i Y(C_{L+k+1}^i, G_{L+k}^1, C_k^L)\Bigg] \\
& + \Bigg[\sum_{i=L+1}^{N-k+1} P_d^i\Big\{C_k^i\Big\} + \sum_{i=L+1}^{L+k-1} P_d^i\Big\{C_{L+k-i}^i\Big\} - P_d^L L C_k^L\Bigg] + \frac{P_m^1}{2}\Bigg[Y(C_{k-1}^1, G_k^1, C_{k+1}^{L-1}) \\
& + Y(C_{L+k}^1, G_{L+k-1}^1, C_k^{L-1}) - Y(C_{k-2}^1, G_{k-1}^1, C_k^L) - Y(C_{L+k+1}^1, G_{L+k}^1, C_k^L)\Bigg] \\
& + \frac{1-\alpha}{2}\Bigg[P_m^{L+1}C_k^{L+1} + P_m^{L+1}C_{k-1}^{L+1} - 2P_m^L C_k^L + \sum_{i=2}^{k-1} P_m^i Y(C_{k-i}^i, G_k^1, C_{k+1}^{L-1}) \\
& + \sum_{i=2}^{N-L-k+1} P_m^i Y(C_{L+k}^i, G_{L+k-1}^1, C_k^{L-1}) - \sum_{i=2}^{k-2} P_m^i Y(C_{k-i-1}^i, G_{k-1}^1, C_k^L) \\
& - \sum_{i=2}^{N-L-k} P_m^i Y(C_{L+k+1}^i, G_{L+k}^1, C_k^L)\Bigg], \\
& L = 2, \ldots, N-4, \quad k = 3, \ldots, N-L-1,
\end{aligned}$$

$$\begin{aligned}
\frac{\mathrm{d}C^L_{N-L}}{\mathrm{d}t} =& \frac{1}{2}\Bigg[P_p^{L-1}C^{L-1}_{N-L} + P_p^{L-1}C^{L-1}_{N-L+1} - 2P_p^L C^L_{N-L} - P_p^{L-1}Y(C^{L-1}_{N-L}, G^1_{N-1}) \\
& - P_p^{L-1}Y(C^{L-1}_{N-L+1}, G^1_{N-L}) + \sum_{i=1}^{L-2} P_p^i Y(C^i_{N-L}, G^1_{N-L+i}, C^{L-i-1}_{N-L+i+1}) \\
& + \sum_{i=1}^{L-2} P_p^i Y(C^i_{N-i}, G^1_{N-i-1}, C^{L-i-1}_{N-L}) - \sum_{i=1}^{N-L-2} P_p^i Y(C^i_{N-L-i-1}, G^1_{N-L-1}, C^L_{N-L})\Bigg] \\
& + \Bigg[P_d^{L+1}C^{L+1}_{N-L} + \sum_{i=L+1}^{N-1} P_d^i \Big\{C^i_{N-i}\Big\} - P_d^L L C^L_{N-L}\Bigg] \\
& + \frac{P_m^1}{2}\Bigg[Y(C^1_N, G^1_{N-1}, C^{L-1}_{N-L}) + Y(C^1_{N-L-1}, G^1_{N-L}, C^{L-1}_{N-L+1}) \\
& - Y(C^1_{N-L-2}, G^1_{N-L-1}, C^L_{N-L})\Bigg] \\
& + \frac{1-\alpha}{2}\Bigg[P_m^{L+1}C^{L+1}_{N-L-1} - 2P_m^L C^L_{N-L} + \sum_{i=2}^{N-L-1} P_m^i Y(C^i_{N-L-i}, G^1_{N-L}, C^{L-1}_{N-L+1}) \\
& - \sum_{i=2}^{N-L-2} P_m^i Y(C^i_{N-L-i-1}, G^1_{N-L-1}, C^L_{N-L})\Bigg], \quad L = 2, \ldots, N-3,
\end{aligned}$$

$$\begin{aligned}
\frac{\mathrm{d}C^L_{N-L+1}}{\mathrm{d}t} =& \frac{1}{2}\Bigg[P_p^{L-1}C^{L-1}_{N-L+1} + P_p^{L-1}C^{L-1}_{N-L+2} - P_p^L C^L_{N-L+1} \\
& - P_p^{L-1}Y(C^{L-1}_{N-L+2}, G^1_{N-L+1}) - \sum_{i=1}^{N-L-1} P_p^i Y(C^i_{N-L-i}, G^1_{N-L}, C^L_{N-L+1}) \\
& + \sum_{i=1}^{L-2} P_p^i Y(C^i_{N-L+1}, G^1_{N-L+i+1}, C^{L-i-1}_{N-L+i+2}) \\
& + \sum_{i=1}^{L-2} P_p^i Y(C^i_{N-i+1}, G^1_{N-i}, C^{L-i-1}_{N-L+1})\Bigg] \\
& + \Bigg[\sum_{i=L+1}^{N} P_d^i \Big\{C^i_{N-i+1}\Big\} - P_d^L L C^L_{N-L+1}\Bigg] \\
& + \frac{P_m^1}{2}\Bigg[Y(C^1_{N-L}, G^1_{N-L+1}, C^{L-1}_{N-L+2}) - Y(C^1_{N-L-1}, G^1_{N-L}, C^L_{N-L})\Bigg] \\
& + \frac{1-\alpha}{2}\Big[P_m^{L+1}C^{L+1}_{N-L} - P_m^L C^L_{N-L+1} \\
& + \sum_{i=2}^{N-L} P_m^i Y(C^i_{N-L-i+1}, G^1_{N-L+1}, C^{L-1}_{N-L+2}) \\
& - \sum_{i=2}^{N-L-1} P_m^i Y(C^i_{N-L-i}, G^1_{N-L}, C^L_{N-L})\Bigg], \quad L = 2, \ldots, N-2,
\end{aligned}$$

$$
\begin{aligned}
\frac{\mathrm{d}C_2^{N-2}}{\mathrm{d}t} =& \frac{1}{2}\Bigg[P_p^{N-3}C_2^{N-3} + P_p^{N-3}C_3^{N-3} - 2P_p^{N-2}C_2^{N-2} \\
& - P_p^{N-3}Y(C_2^{N-3}, G_{N-1}^1) - P_p^{N-3}Y(C_3^{N-3}, G_2^1) \\
& + \sum_{i=1}^{N-4} P_p^i Y(C_2^i, G_{i+2}^1, C_{i+3}^{N-i-3}) + \sum_{i=1}^{N-4} P_p^i Y(C_{N-i}^i, G_{N-i-1}^1, C_2^{N-i-3})\Bigg] \\
& + \Bigg[P_d^{N-1}C_2^{N-1} + P_d^{N-1}C_1^{N-1} - P_d^{N-2}(N-2)C_2^{N-2}\Bigg] \\
& + \frac{P_m^1}{2}\Bigg[Y(C_1^1, G_2^1, C_3^{N-3}) + Y(C_N^1, G_{N-1}^1, C_2^{N-3})\Bigg] + \frac{1-\alpha}{2}\Bigg[-2P_m^{N-2}C_2^{N-2}\Bigg],
\end{aligned}
$$

$$
\begin{aligned}
\frac{\mathrm{d}C_1^{N-1}}{\mathrm{d}t} =& \frac{1}{2}\Bigg[P_p^{N-2}C_1^{N-2} + P_p^{N-2}C_2^{N-2} - P_p^{N-1}C_1^{N-1} + \sum_{i=1}^{N-3} P_p^i Y(C_1^i, G_{i+1}^1, C_{i+2}^{N-i-2}) \\
& + \sum_{i=1}^{N-3} P_p^i Y(C_{N-i}^i, G_{N-i-1}^1, C_1^{N-i-2})\Bigg] + \Big[P_d^N C_1^N - P_d^{N-1}(N-1)C_1^{N-1}\Big] \\
& + \frac{P_m^1}{2}\Bigg[Y(C_N^1, G_{N-1}^1, C_1^{N-2})\Bigg] + \frac{1-\alpha}{2}\Bigg[-P_m^{N-1}C_1^{N-1}\Bigg],
\end{aligned}
$$

$$
\begin{aligned}
\frac{\mathrm{d}C_2^{N-1}}{\mathrm{d}t} =& \frac{1}{2}\Bigg[P_p^{N-2}C_2^{N-2} + P_p^{N-2}C_3^{N-2} - P_p^{N-1}C_2^{N-1} - P_p^{N-2}Y(C_3^{N-2}, G_2^1) \\
& + \sum_{i=1}^{N-3} P_p^i Y(C_2^i, G_{i+2}^1, C_{i+3}^{N-i-2}) + \sum_{i=1}^{N-3} P_p^i Y(C_{N-i+1}^i, G_{N-i}^1, C_2^{N-i-2})\Bigg] \\
& + \Bigg[P_d^N C_1^N - P_d^{N-1}(N-1)C_2^{N-1}\Bigg] + \frac{P_m^1}{2}\Bigg[Y(C_1^1, G_2^1, C_3^{N-2})\Bigg] \\
& + \frac{1-\alpha}{2}\Bigg[-P_m^{N-1}C_2^{N-1}\Bigg],
\end{aligned}
$$

$$
\begin{aligned}
\frac{\mathrm{d}C_1^{N}}{\mathrm{d}t} =& \frac{1}{2}\Bigg[P_p^{N-1}C_1^{N-1} + P_p^{N-1}C_2^{N-1} + \sum_{i=1}^{N-2} P_p^i Y(C_1^i, G_{i+1}^1, C_{i+2}^{N-i-1}) \\
& + \sum_{i=1}^{N-2} P_p^i Y(C_{N-i+1}^i, G_{N-i}^1, C_1^{N-i-1})\Bigg] + P_d^N\Bigg[-NC_1^N\Bigg].
\end{aligned}
$$

Gaps

Here we present the time rate of change for each gap length and location, obtained by considering the potential outcomes of each type of birth, death and movement event.

$$\begin{aligned}\frac{\mathrm{d}G_1^1}{\mathrm{d}t} =& \frac{1}{2}\left[\sum_{i=1}^{N-2} P_p^i Y(C_3^i, G_1^2) - \sum_{i=1}^{N-1} P_p^i Y(C_2^i, G_1^1)\right] + \left[\sum_{i=2}^{N} P_d^i \Big\{C_1^i\Big\} - \sum_{i=1}^{N-1} Y(C_2^i, G_1^1)\right] \\ &+ \frac{P_m^1}{2}\left[C_1^1 - 2C_2^1 + Y(C_3^1, G_1^2)\right] + \frac{1-\alpha}{2}\left[-\sum_{i=2}^{N-1} P_m^i \Big\{C_2^i\Big\} + \sum_{i=2}^{N-2} P_m^i Y(C_3^i, G_1^2)\right],\end{aligned}$$

$$\begin{aligned}\frac{\mathrm{d}G_2^1}{\mathrm{d}t} =& \frac{1}{2}\left[\sum_{i=1}^{N-3} P_p^i Y(C_{4,}^i, G_2^2) - \sum_{i=1}^{N-2} P_p^i \Big\{Y(C_3^i, G_2^1)\Big\} - P_p^1 Y(C_1^1, G_2^1)\right] \\ &+ \left[\sum_{i=3}^{N} P_d^i \Big\{C_1^i\Big\} - P_d^1 Y(C_1^1, G_2^1) - \sum_{i=1}^{N-1} P_d^i Y(C_3^i, G_2^1)\right] \\ &+ \frac{P_m^1}{2}\left[- Y(C_1^1, G_2^1) - 2Y(C_3^1, G_2^1) + Y(C_4^1, G_2^2)\right] \\ &+ \frac{(1-\alpha)}{2}\left[P_m^2 C_1^2 + \sum_{i=2}^{N-1} P_m^i \Big\{C_2^i\Big\} - \sum_{i=2}^{N-2} P_m^i \Big\{Y(C_3^i, G_2^1)\Big\} + \sum_{i=2}^{N-3} P_m^i Y(C_4^i, G_2^2)\right],\end{aligned}$$

$$\begin{aligned}\frac{\mathrm{d}G_k^1}{\mathrm{d}t} =& \frac{1}{2}\left[\sum_{i=1}^{N-k-1} P_p^i Y(C_{k+2}^i, G_k^2)\right. \\ &\left.+ \sum_{i=1}^{k-2} P_p^i Y(C_{k-i-1}^i, G_{k-1}^2) - \sum_{i=1}^{N-k} P_p^i Y(C_{k+1}^i, G_k^1) - \sum_{i=1}^{k-1} P_p^i Y(C_{k-i}^i, G_k^1)\right] \\ &+ \left[\sum_{i=1}^{k-1} \sum_{j=k-i+2}^{N-i+1} P_d^j \Big\{C_i^j\Big\} - \sum_{i=1}^{k-1} P_d^i Y(C_{k-i}^i, G_k^1) - \sum_{i=1}^{N-k} P_d^i Y(C_{k+1}^i, G_k^1)\right] \\ &+ \frac{P_m^1}{2}\left[- 2Y(C_{k-1}^1, G_k^1) - 2Y(C_{k+1}^1, G_k^1) + Y(C_{k-2}^1, G_{k-1}^2) + Y(C_{k+2}^1, G_k^2)\right] \\ &+ \frac{1-\alpha}{2}\left[\sum_{i=2}^{k} P_m^i \Big\{C_{k-i+1}^i\Big\} + \sum_{i=2}^{N-k+1} P_m^i \Big\{C_k^i\Big\} - \sum_{i=2}^{k-1} P_m^i \Big\{Y(C_{k-i}^i, G_k^1)\Big\}\right. \\ &\left.- \sum_{i=2}^{N-k} P_m^i \Big\{Y(C_{k+1}^i, G_k^1)\Big\} + \sum_{i=2}^{k-2} P_m^i Y(C_{k-i-1}^i, G_{k-1}^2) + \sum_{i=2}^{N-k-1} P_m^i Y(C_{k+2}^i, G_k^2)\right], \\ & k = 3, \ldots, N-2,\end{aligned}$$

$$\begin{aligned}\frac{\mathrm{d}G^1_{N-1}}{\mathrm{d}t} =& \frac{1}{2}\left[\sum_{i=1}^{N-3} P^i_p\Big\{Y(C^i_{N-i-2}, G^2_{N-2})\Big\} - P^1_pY(C^1_N, G^{N-1}_1) - \sum_{i=1}^{k-1} P^i_pY(C^i_{N-i-1}, G^1_{N-1})\right]\\ &+ \left[\sum_{i=1}^{N-2} P^i_d\Big\{C_{N-i+1,i}\Big\} - \sum_{i=1}^{N-2} P^i_d\Big\{Y(C^i_{N-i-1}, G^1_{N-1})\Big\} - P^1_dY(C^1_N, G^1_k)\right]\\ &+ \frac{P^1_m}{2}\left[- 2Y(C^1_{N-2}, G^1_{N-1}) - Y(C^1_N, G^1_{N-1}) + Y(C^1_{N-3}, G^2_{N-2})\right]\\ &+ \frac{1-\alpha}{2}\left[\sum_{i=2}^{N-1} P^i_m\Big\{C^i_{N-i}\Big\} + P^2_mC^2_{N-1} - \sum_{i=2}^{N-2} P^i_m\Big\{Y(C^i_{N-1-i}, G^1_{N-1})\Big\}\right.\\ &\left.+ \sum_{i=2}^{N-3} P^i_mY(C^i_{N-i-2}, G^2_{N-2})\right],\end{aligned}$$

$$\begin{aligned}\frac{\mathrm{d}G^1_N}{\mathrm{d}t} =& \frac{1}{2}\left[\sum_{i=1}^{N-2} P^i_pY(C^i_{N-i-1}, G^2_{N-1}) - \sum_{i=1}^{N-1} P^i_pY(C^i_{N-i}, G^1_N)\right]\\ &+ \left[\sum_{i=2}^{N} P^i_dC^i_{N-i+1} - \sum_{i=1}^{N-1} P^i_dY(C^i_{N-i}, G^1_N)\right]\\ &+ \frac{P^1_m}{2}\left[C^1_N - 2C^1_{N-1} + Y(C^1_{N-2}, G^2_{N-1})\right]\\ &+ \frac{1-\alpha}{2}\left[- \sum_{i=2}^{N-1} P^i_m\Big\{C^i_{N-i}\Big\} + \sum_{i=2}^{N-2} P^i_mY(C^i_{N-i-1}, G^2_{N-1})\right],\end{aligned}$$

$$\begin{aligned}\frac{\mathrm{d}G^L_1}{\mathrm{d}t} =& \frac{1}{2}\left[\sum_{i=1}^{N-L-1} P^i_pY(C^i_{L+2}, G^{L+1}_1) - \sum_{i=1}^{N-L} P^i_pY(C^i_{L+1}, G^L_1)\right] + \left[P^1_dY(C^1_1, G^{L-1}_2)\right.\\ &\left.+ \sum_{i=2}^{N-L+1} P^i_dY(C^i_L, G^{L-1}_1) + \sum_{i=1}^{L-2} P^1_dY(C^1_{i+1}, G^i_1, G^{L-i-1}_{i+2}) - \sum_{i=1}^{N-L} P^i_dY(C^i_{L+1}, G^L_1)\right]\\ &+ \frac{P^1_m}{2}\left[Y(C^1_L, G^{L-1}_1) - 2Y(C^1_{L+1}, G^L_1) + Y(C^1_{L+2}, G^{L+1}_1)\right]\\ &+ \frac{1-\alpha}{2}\left[- \sum_{i=2}^{N-L} P^i_m\Big\{Y(C^i_{L+1}, G^L_1)\Big\} + \sum_{i=2}^{N-L-1} P^i_mY(C^i_{L+2}, G^{L+1}_1)\right],\\ &L = 2, \ldots, N-2,\end{aligned}$$

$$
\begin{aligned}
\frac{\mathrm{d}G_2^L}{\mathrm{d}t} =& \frac{1}{2}\Bigg[\sum_{i=1}^{N-L-2} P_p^i Y(C_{L+3}^i, G_2^{L+1}) - \sum_{i=1}^{N-L-1} P_p^i \Bigg\{ Y(C_{L+2}^i, G_2^L) \Bigg\} - P_p^1 Y(C_1^1, G_2^L)\Bigg] \\
&+ \Bigg[P_d^2 Y(C_1^2, G_3^{L-1}) + \sum_{i=2}^{N-L} P_d^i Y(C_{L+1}^i, G_2^{L-1}) + \sum_{i=1}^{L-2} P_d^1 \Big\{ Y(C_{i+2}^1, G_2^i, G_{i+3}^{L-i-1}) \Big\} \\
&- P_d^1 Y(C_1^1, G_2^L) - \sum_{i=1}^{N-L-1} P_d^i Y(C_{L+2}^i, G_2^L)\Bigg] + \frac{P_m^1}{2}\Bigg[Y(C_2^1, G_3^{L-1}) + Y(C_{L+1}^1, G_2^{L-1}) \\
&- Y(C_1^1, G_2^L) - 2Y(C_{L+2}^1, G_2^L) + Y(C_{L+3}^1, G_2^{L+1})\Bigg] \\
&+ \frac{1-\alpha}{2}\Bigg[-\sum_{i=2}^{N-L-1} P_m^i Y(C_{L+2}^i, G_2^L) + \sum_{i=2}^{N-L-2} P_m^i Y(C_{L+3}^i, G_2^{L+1})\Bigg], \\
& L = 2, \dots, N-3,
\end{aligned}
$$

$$
\begin{aligned}
\frac{\mathrm{d}G_k^L}{\mathrm{d}t} =& \frac{1}{2}\Bigg[\sum_{i=1}^{N-L-k} P_p^i Y(C_{L+k+1}^i, G_k^{L+1}) + \sum_{i=1}^{k-2} P_p^i Y(C_{k-i-1}^i, G_{k-1}^{L+1}) \\
&- \sum_{i=1}^{N-L-k+1} P_p^i Y(C_{L+k}^i, G_k^L) - \sum_{i=1}^{k-1} P_p^i Y(C_{k-i}^i, G_k^L)\Bigg] \\
&- \Bigg[\sum_{i=2}^{k} P_d^i Y(C_{k-i+1}^i, G_{k+1}^{L-1}) + \sum_{i=2}^{N-L-k+2} P_d^i Y(C_{L+k-1}^i, G_k^{L-1}) \\
&+ \sum_{i=1}^{L-2} P_d^1 Y(C_{k+i}^1, G_k^i, G_{k+i+1}^{L-i-1}) - \sum_{i=1}^{k-1} P_d^i Y(C_{k-i}^i, G_k^L) - \sum_{i=1}^{N-L-k+1} P_d^i Y(C_{L+k}^i, G_k^L)\Bigg] \\
&+ \frac{P_m^1}{2}\Bigg[Y(C_k^1, G_{k+1}^{L-1}) + Y(C_{L+k-1}^1, G_k^{L-1}) - 2Y(C_{k-1}^1, G_k^L) - 2Y(C_{L+k}^1, G_k^L) \\
&+ Y(C_{k-2}^1, G_{k-1}^{L+1}) + Y(C_{L+k+1}^1, G_k^{L+1})\Bigg] \\
&+ \frac{1-\alpha}{2}\Bigg[-\sum_{i=2}^{k-1} P_m^i Y(C_{k-i}^i, G_k^L) - \sum_{i=2}^{N-L-k+1} P_m^i Y(C_{L+k}^i, G_k^L) \\
&+ \sum_{i=2}^{k-2} P_m^i Y(C_{k-i-1}^i, G_{k-1}^{L+1}) + \sum_{i=2}^{N-L-k} P_m^i Y(C_{L+k+1}^i, G_k^{L+1})\Bigg], \\
& L = 2, \dots, N-4, \quad k = 3, \dots, N-L-1,
\end{aligned}
$$

$$
\begin{aligned}
\frac{\mathrm{d}G^L_{N-L}}{\mathrm{d}t} =& \frac{1}{2}\Bigg[\sum_{i=1}^{N-L-2} P_p^i\Big\{Y(C^i_{N-L-i-1}, G^{L+1}_{N-L-1})\Big\} - P_p^1 Y(C^1_N, G^L_{N-L}) \\
&- \sum_{i=1}^{N-L-1} P_p^i Y(C^i_{N-L-i}, G^L_{N-L})\Bigg] + \Bigg[\sum_{i=2}^{N-L} P_d^i\Big\{Y(C^i_{N-L-i+1}, G^{L-1}_{N-L+1})\Big\} \\
&+ P_d^2 Y(C^2_{N-1}, G^{L-1}_{N-L}) + \sum_{i=1}^{L-2} P_d^1 Y(C^1_{N-L+i}, G^i_{N-L}, G^{L-i-1}_{N-L+i+1}) \\
&- \sum_{i=1}^{N-L-1} P_d^i\Big\{Y(C^i_{N-L-i}, G^L_k)\Big\} - P_d^1 Y(C^1_N, G^L_{N-L})\Bigg] \\
&+ \frac{P_m^1}{2}\Bigg[Y(C^1_{N-L}, G^{L-1}_{N-L+1}) + Y(C^1_{N-1}, G^{L-1}_{N-L}) - 2Y(C^1_{N-L-1}, G^L_{N-L}) \\
&- Y(C^1_N, G^L_{N-L}) + Y(C^1_{N-L-2}, G^{L+1}_{N-L-1})\Bigg] \\
&+ \frac{1-\alpha}{2}\Bigg[-\sum_{i=2}^{N-L-1} P_m^i Y(C^i_{N-L-i}, G^L_{N-L} \\
&+ \sum_{i=2}^{N-L-2} P_m^i Y(C^i_{N-L-i-1}, G^{L+1}_{N-L-1})\Bigg], \quad L = 2, \ldots, N-3,
\end{aligned}
$$

$$
\begin{aligned}
\frac{\mathrm{d}G^L_{N-L+1}}{\mathrm{d}t} =& \frac{1}{2}\Bigg[\sum_{i=1}^{N-L-1} P_p^i Y(C^i_{N-L-i}, G^{L+1}_{N-L}) - \sum_{i=1}^{N-L} P_p^i Y(C^i_{N-L-i+1}, G^L_{N-L+1})\Bigg] \\
&+ \Bigg[P_d^1 Y(C^1_N, G^{L-1}_{N-L+1}) + \sum_{i=1}^{N-L+1} P_d^i Y(C^i_{N-L-i+2}, G^{L-1}_{N-L+2}) \\
&- \sum_{i=1}^{N-L} P_d^i Y(C^i_{N-L-i+1}, G^L_{N-L+1}) \\
&+ \sum_{i=1}^{L-2} P_d^1 Y(C^1_{N-L+i+1}, G^i_{N-L+1}, G^{L-i-1}_{N-L+i+2})\Bigg] \\
&+ \frac{P_m^1}{2}\Bigg[Y(C^1_{N-L+1}, G^{L-1}_{N-L+2}) - 2Y(C^1_{N-L}, G^L_{N-L+1}) + Y(C^1_{N-L-1}, G^{L+1}_{N-L})\Bigg] \\
&+ \frac{1-\alpha}{2}\Bigg[-\sum_{i=2}^{N-L} P_m^i Y(C^i_{N-L-i+1}, G^L_{N-L+1}) \\
&+ \sum_{i=2}^{N-L-1} P_m^i Y(C^i_{N-L-i}, G^{L+1}_{N-L})\Bigg], \quad L = 2, \ldots, N-2,
\end{aligned}
$$

$$\begin{aligned}
\frac{\mathrm{d}G_2^{N-2}}{\mathrm{d}t} =& \frac{1}{2}\Bigg[- P_p^1 Y(C_N^1, G_2^{N-2}) - P_p^1 Y(C_1^1, G_2^{N-2}) \Bigg] \\
&+ \Bigg[P_d^2 Y(C_1^2, G_3^{N-3}) + P_d^2 Y(C_{N-1}^2, G_2^{N-3}) \\
&+ \sum_{i=1}^{N-4} P_d^1 \Big\{ Y(C_{i+2}^1, G_2^i, G_{i+3}^{N-i-3}) \Big\} - P_d^1 Y(C_1^1, G_2^{N-2}) - P_d^1 Y(C_N^1, G_2^{N-2}) \Bigg] \\
&+ \frac{P_m^1}{2} \Bigg[Y(C_2^1, G_3^{N-3}) + Y(C_{N-1}^1, G_2^{N-3}) - Y(C_1^1, G_2^{N-2}) - Y(C_N^1, G_2^{N-2}) \Bigg],
\end{aligned}$$

$$\begin{aligned}
\frac{\mathrm{d}G_1^{N-1}}{\mathrm{d}t} =& \frac{1}{2}\Bigg[- P_p^1 Y(C_N^1, G_1^{N-1}) \Bigg] + \Bigg[P_d^1 Y(C_1^1, G_2^{N-2}) + P_d^2 Y(C_{N-1}^2, G_1^{N-2}) \\
&+ \sum_{i=1}^{N-3} P_d^1 \Big\{ Y(C_{i+1}^1, G_1^i, G_{i+2}^{N-i-2}) \Big\} - P_d^1 Y(C_N^1, G_1^{N-1}) \Bigg] \\
&+ \frac{P_m^1}{2} \Bigg[Y(C_{N-1}^1, G_1^{N-2}) - Y(C_N^1, G_1^{N-1}) \Bigg],
\end{aligned}$$

$$\begin{aligned}
\frac{\mathrm{d}G_2^{N-1}}{\mathrm{d}t} =& \frac{1}{2}\Bigg[- P_p^1 Y(C_1^1, G_2^{N-1}) \Bigg] + \Bigg[P_d^1 Y(C_N^1, G_2^{N-2}) + P_d^2 Y(C_1^2, G_3^{N-2}) \\
&- P_d^1 Y(C_1^1, G_2^{N-1}) + \sum_{i=1}^{N-3} P_d^1 Y(C_{i+2}^1, G_2^i, G_{i+3}^{N-i-2}) \Bigg] \\
&+ \frac{P_m^1}{2} \Bigg[Y(C_2^1, G_3^{N-2}) - Y(C_1^1, G_2^{N-1}) \Bigg],
\end{aligned}$$

$$\frac{\mathrm{d}G_1^N}{\mathrm{d}t} = P_d^1 \Bigg[Y(C_1^1, G_2^{N-1}) + Y(C_N^1, G_1^{N-1}) + \sum_{i=1}^{N-2} Y(C_{i+1}^1, G_1^i, G_{i+2}^{L-i-1}) \Bigg].$$

The Y function represents the number of configurations that contain the specified chains and gaps within the parentheses, and is defined as

$$Y(C_j^i, G_l^k) = C_j^i P(G_l^k | C_j^i),$$

where $P(G_l^k|C_j^i)$ represents the probability that a gap of length k at site l exists, given that there is a chain of length i at site j, and can be calculated as

$$P(G_l^k|C_j^i) = G_l^k / A_G(C_j^i),$$

where $A_G(C_j^i)$ are all the possible gaps that can exist on either side of the chain of length i at site j. If we are interested in gaps on the positive x side of the chain, we require

$l = i + j$ and $k \leq N - i - j + 1$. If we are interested in gaps on the negative x side of the chain, we require $l \leq j - 1$, $k = j - k$. Similarly,

$$Y(C_j^i, G_l^k, G_n^m) = C_j^i P(G_l^k | C_j^i) P(G_n^m | C_j^i, G_l^k),$$

where $P(G_n^m | C_j^i, G_l^k)$ is the probability a gap of length m at site n exists, given that there is a chain of length i at site j and a gap of length k at site l. Note that the gaps at site l and site n must necessarily be on opposing sides of the chain at site j and hence $P(G_n^m | C_j^i, G_l^k)$ is calculated in the same manner as $P(G_l^k | C_j^i)$.

9.6.2 Numerical techniques

Ordinary differential equations

The system of ODEs describing the dynamics of the chains and gaps, presented in Appendices A.1 and A.2, are solved using an adaptive Runge-Kutta method with a strict truncation error control of 10^{-6} [39]. All results presented are found to be insensitive to a reduction in the strict truncation error control.

Partial differential equations

The two mean-field descriptions of the discrete process, Equation (9.1) and Equation (9.2), are discretised onto a spatially uniform finite difference grid with grid spacing δx. The spatial derivative terms are approximated using a central finite difference approximation. We approximate the temporal derivative using the backward Euler method with constant time step δt, and the resulting system of nonlinear algebraic equations is solved using Picard iteration with absolute convergence tolerance ϵ. Finally, the system of tridiagonal algebraic equations is solved using the Thomas algorithm [39]. In all cases, $\delta x = 0.1$, $\delta t = 0.01$ and $\epsilon = 10^{-6}$.

Chapter 10

Conclusions and future work

Here we detail the outcomes and impact on the relevant pre-existing literature arising from each component of this thesis, as well as discuss potential directions for future research.

10.1 Summary

This thesis is separated into two distinct parts; the application of mathematical models to experimental data, and developing novel continuum models of collective behaviour. Both of these parts are motivated by the mathematical modelling of collective cell behaviour, and the potential for these models to inform experimental direction in an inexpensive and insightful manner.

Both of these parts involve the realisation of four sub-aims to achieve an overall aim. The first part, where the overall aim is to use a combination of mathematical modelling, experimental techniques and parameter recovery approaches to obtain quantitative insight into the cellular mechanisms governing collective cell behaviour, consists of

- Determining the requisite experimental measurements and procedure for applying the Fisher-Kolmogorov model to relevant experimental data such that reliable estimates of the cell carrying capacity, cell diffusivity and cell proliferation rate can be obtained from an IncuCyte ZOOM$^{\text{TM}}$ assay.
- Analysing the information present in the position of the leading edge of the cell population in a scratch assay to determine whether robust estimates of the cell diffusivity and cell proliferation rate can be obtained.
- Investigating whether approximate Bayesian computation is a valid technique for estimating cell parameters from experimental data obtained from a scratch assay and, if so, determining which experimental measurements contain sufficient information for robust parameter estimates.

- Quantifying the change in parameter estimates, and uncertainty associated with the parameter estimates, due to experimental design choices in a scratch assay using approximate Bayesian computation.

The second part, where the overall aim is to use mathematical modelling techniques to develop novel continuum descriptions of collective behaviour from individual-based models of motility, birth and death, consists of

- Developing a corrected mean-field description for multiple interacting subpopulations to obtain insight into the parameter regimes where corrected mean-field descriptions are required to accurately approximate lattice-based random walk models of multiple interacting subpopulations.
- Determining whether modelling the dynamics of groups of contiguous occupied lattice sites and contiguous unoccupied lattice sites provides a valid approximation of the average behaviour of birth-death-movement processes, in particular for parameter regimes where mean-field and corrected mean-field descriptions are inaccurate.
- Investigating how introducing competitive and co-operative mechanisms into lattice-based birth-death-movement processes changes the standard mean-field description of the process, and the implications on long-time behaviour for the population.
- Developing a model for the dynamics of groups of contiguous occupied lattice sites and contiguous unoccupied lattice sites for an invasive population and compare the predictions for the long-time behaviour of the population with predictions obtained from previously-proposed continuum descriptions.

Each of the eight sub-aims listed resulted in the publication of a peer-reviewed manuscript, corresponding to the work presented in Chapters 2-9. Here we summarise the outcomes of each of the publications in the context of the relevant in the literature, as well as the impact of each publication with respect to potential future work.

In Chapter 2, we investigate techniques for obtaining quantitative estimates of the cellular parameters governing the collective behaviour in an IncuCyte ZOOM$^{\text{TM}}$ assay. *In vitro* cell biology assays are commonly used to investigate the influence of chemical treatments on collective cell behaviour [31, 42, 49, 58, 65]. However, standard techniques for interpreting IncuCyte ZOOM$^{\text{TM}}$ assays use the area of the uncolonised region in the assay as a measurement, which does not provide information about the individual roles of cell motility and cell proliferation [14, 45, 46]. It was unclear whether the Fisher-Kolmogorov model could be used to provide quantitative estimates of the parameters governing collective cell migration in an IncuCyte ZOOM$^{\text{TM}}$ assay. We demonstrate that the Fisher-Kolmogorov model can be calibrated to IncuCyte ZOOM$^{\text{TM}}$ assays to provide individual estimates of the cell diffusivity, cell proliferation rate and the cell carrying capacity density, provided that the calibration is done in a methodical manner. Specifically, we first measure the cell carrying capacity density by counting the number of cells in several regions sufficiently far away from the initially-unoccupied region at the final experimental time. Second, we track the temporal evolution of the cell density in several regions sufficient far away from the initially-unoccupied region and calibrate a logistic growth model to obtain an estimate of the cell proliferation rate. Finally, we obtain an estimate of the cell diffusivity by calibrating the Fisher-Kolmogorov model to the position of the leading edge of the cell population, using our previously-obtained estimates of the cell carrying capacity density and cell proliferation rate.

The technique developed provides parameter estimates for a prostate cancer cell line (PC-3) [25] that are consistent with previously-reported parameter values and allows us to quantify the influence of different concentrations of epidermal growth factor (EGF) on the cell diffusivity, cell proliferation rate and the cell carrying capacity density [18]. Previously, it was unclear whether PC-3 cells are influenced by EGF, and if the cell population is influenced, how the cell parameters change with EGF concentration. We find that there is a monotonically increasing relationship between EGF concentration and cell diffusivity, and that an increase in EGF concentration results in an increase in the cell proliferation rate, up to a threshold concentration. The cell carrying capacity density appears to be unaffected by the application of EGF. As both the cell diffusivity and cell proliferation rate increase with the EGF concentration, the overall rate of spread for the PC-3 population is enhanced by the EGF concentration. Our technique provides a robust quantitative method for investigating the influence of chemical treatments on cell populations. This technique has been extended by Jin *et al.* [20] to investigate the influence of the initial cell density on the per capita growth rate of the cell population, and presents an interesting opportunity to examine the standard assumption that the cell diffusivity and cell proliferation rates are constant.

In Chapter 3, we examine the information about the cell motility and cell proliferation rate present in the position of the leading edge of a cell population in an *in vitro* scratch assay. The most common method for quantitatively interpreting scratch assays is to track the position of the leading edge [1, 5, 49]. Such techniques are inexpensive and do not require time-consuming approaches such as cell tracking or counting individual cells [47, 61]. Hence this approach requires less time to interpret the experimental images than the approach considered in Chapter 2, and would be more suitable for high throughput experiments. However, it was unclear whether unique estimates of the cell motility and cell proliferation rate could be obtained by measuring the position of the leading edge only. Isolating parameter estimates for the mechanisms governing the collective migration of a cell population is critical for quantifying the influence of potential intervention strategies, such as chemotherapeutic drugs or steroid treatment [12, 13, 29, 62]. We demonstrate that a naïve parameter recovery approach, where we compare the leading edge position obtained from an averaged discrete model of cell motility and cell proliferation with the leading edge position obtained from a scratch assay, does not result in unique parameter estimates.

Typically, cell proliferation occurs on a time scale of days and cell motility occurs on a time scale of minutes [28, 32], an phenomenon also observed in the cell population examined in Chapter 2. This implies that the short-time evolution of the position of the leading edge is dominated by motility. Using this information, we calibrate the discrete mathematical model to the experimental data for short-time, assuming that only motility events occur, to obtain a parameter estimate for the cell motility. With this estimate, we calibrate the discrete mathematical model to the experimental data at late time to obtain an estimate of the cell proliferation rate. We find that this approach results in unique estimates of the cell motility and the cell proliferation. We anticipate that this approach could be applied to other cell populations, or could be used to investigate the influence of chemical treatments on a particular cell population. The position of the leading edge of a cell population has subsequently been used in studies investigating inference techniques for parameter recovery [44] and models of keratinocyte wound healing dynamics [39].

In Chapter 4, we investigate approximate Bayesian computation (ABC) as a parameter estimation technique for *in vitro* scratch assays. ABC allows for the quantification of the uncertainty associated with the parameter estimates, rather than providing point estimates like traditional deterministic parameter recovery techniques [33]. While the approaches considered in Chapter 2 and Chapter 3 produce unique estimates of the cell diffusivity and cell proliferation rate, any uncertainty with these estimates is associated with experimental replicates or multiple experimental measurements. It was unclear whether ABC is a suitable technique for spatiotemporal processes such as scratch assays. Furthermore, it was unclear which summary statistic or experimental measurement contains sufficient information about the cell parameters. ABC requires a summary statistic to compare between the experimental data and the mathematical model [33]. Typically, this summary statistic cannot be high dimensional as the computational cost associated with the ABC algorithm increases with the dimensionality of the summary statistic. This implies that the position of the cells in the scratch assays cannot be considered as a summary statistic.

We examine a number of summary statistics including the pair correlation function [3], the number of cells, the counts of pair distances, and a combination of the number of cells and the pair correlation function. We find that the combination of summary statistics provides the most information about the cell motility and the cell proliferation rate. Furthermore, this information provides unimodal distributions for both cell parameters and hence estimates of the parameter values and their associated uncertainty can be quantified. Previous parameter recovery techniques, including the techniques presented in Chapter 2 and Chapter 3, either report no uncertainty in the measurement, or only include uncertainty associated with the experimental replicates [55, 56]. Our technique would allow for appropriate summary statistics to be determined for different experiments, and would provide quantifiable information about the underlying cellular mechanisms governing the collective behaviour in the experiment. The potential of ABC techniques to recover parameter estimates from cell biology experiments is highlighted by several follow-up studies that incorporate ABC in the context of cell biology experiments where cell-to-cell adhesion is important [44] or studies that estimate the cell diffusivity and cell proliferation rate separately [63].

In Chapter 5, we examine how the experimental design of an *in vitro* scratch assay influences the information obtained from the scratch assay about the cell parameters. The experimental design of scratch assays varies significantly throughout the literature and there is no apparent motivation about the choice of experimental design [1, 4, 37, 42, 50]. For example, the scratch can be performed such that only one cell front is observed [4, 32] or such that there are two opposingly-directed cell fronts [31, 43]. The number and timing of experimental observations reported varies similarly [1, 4, 37, 42, 50]. If we compare the scratch assays presented in Chapters 2-4, we see that the design of the scratch assays is not consistent, and we do not consider how the design choices may have influenced the parameter estimates. To investigate this, we implement our previously developed ABC parameter recovery approach (Chapter 4) to a scratch assay with one cell front and a scratch assay with two cell fronts. Furthermore, we consider capturing experimental observations at different combinations of time points.

We find that design choices influence both the parameters recovered and the uncertainty associated with each parameter. If the combinations of observation times used in the parameter recovery process include the final experimental observation there is less uncertainty about the parameter estimates, compared to combinations of observation times that include only short-time observations. Furthermore, there are diminishing returns associated with the information obtained about the parameters and the number of observation times. Using three observation times, not including the initial observation, provides a marginal increase in information obtained about the parameters, compared to using two observation times. As additional observation times increase both the computational effort required and the time taken to interpret the experimental image, this result provides guidance about reducing experimental cost while preserving sufficient information to obtain reliable parameter estimates. Additionally, our investigation finds that the experimental design containing one cell front contains less information about the cell parameters than the experimental design containing two cell fronts. However, the design with two fronts contains less information about the cell motility, with the additional information associated with the proliferation rate. As the amount of information obtained about the proliferation rate is significantly higher than the amount of information obtained about the motility rate, this additional information is less valuable than an equivalent increase in the information obtained about the motility rate. Hence we find that the experimental design with a single cell front provides more robust estimates than the experimental design with two cell fronts. Our approach provides a robust and quantifiable method for evaluating the impact of experimental design choices on both parameter estimates and their associated uncertainty, which previously had not been investigated. We anticipate that implementing similar ABC parameter recovery algorithms would provide guidance about experimental design choices for a variety of cell biology experiments.

The work presented in Chapters 2-5 provides guidance relating to the framework choices that can be made when performing a scratch assay to obtain quantitative estimates of the cell parameters governing the behaviour of the population. Decisions must be made regarding the experimental measurement recorded from the scratch assay, the experimental design of the scratch assay, the type of mathematical model used to interpret the experimental data and the type of parameter recovery technique used to interface the model and experimental data. We present a variety of frameworks that highlight the interplay between the experimental and/or mathematical complexity required in the implementation of the parameter recovery framework, and the information gained about the cell parameters. We show that relatively simple approaches produce unique point-estimates of the cell diffusivity and the cell proliferation rate, while additional information regarding the uncertainty can be gleaned through more sophisticated approaches. Preliminary investigations may not require this additional information, and hence the suite of frameworks presented in Chapters 2-5 allows for an informed choice of framework, depending on the level of detail about the cell parameters desired from the investigation.

In Chapter 6, we investigate the use of a corrected mean-field description for describing the movement, proliferation and death of two interacting and spreading subpopulations. Standard mean-field descriptions of these subpopulations are accurate only in extremely limited parameter regimes [2, 21, 51]. For simpler processes, such as movement, proliferation and death for a single population or the movement, proliferation and death for two interacting but spatially uniform subpopulations, corrected mean-field descriptions have been shown to robustly describe these processes in a wider range of parameter regimes [2, 21, 35, 51]. It was unclear in which parameter regimes a similar approach would provide an accurate description of the underlying discrete model for two spatially heterogeneous interacting subpopulations. Similarly, it was unclear which parameter regimes give rise to strong spatial correlations within the one subpopulation or between the two subpopulations.

We derive a corrected mean-field description for a discrete exclusion-based process of movement, proliferation and death for two interacting subpopulations and investigate how accurately the numerical solution of the description matches the average behaviour of the discrete model, compared to the standard mean-field description. In this investigation we compare the two descriptions for several geometries motivated by relevant experimental designs for a suite of parameter regimes. While the corrected mean-field description always outperforms the standard mean-field description, our investigation distinguishes between cases where the standard mean-field description provides an adequate approximation and cases where a corrected mean-field description is required. These results provide guidance for future models, with respect to the type of mean-field description required for a particular parameter regime.

In Chapter 7, we devise a novel interpretation of stochastic lattice-based exclusion processes containing birth, death and adhesive movement. Instead of interpreting the discrete model as a number of lattice sites that are either occupied or vacant, such as in Chapter 6, we interpret the model as a group of contiguous occupied lattice sites and contiguous vacant lattice sites, termed chain and gaps, respectively. Furthermore, we subsequently develop a continuum approximation of this interpretation of the discrete model such that the dynamics of the chains and gaps can be described in a deterministic manner. It was unclear whether this approximation and interpretation would adequately describe the average behaviour from the discrete model. Additionally, it was uncertain if the chains and gaps provided an accurate measure of the clustering present in the discrete model.

We find that the the chain-and-gap (C&G) description provides a more accurate approximation of the average discrete behaviour than previous mean-field and corrected mean-field descriptions in all parameter regimes [2]. We also distinguish between cases where mean-field and corrected mean-field descriptions provide an adequate approximation of the average discrete behaviour, and when the chain-and-gap description is required. This analysis informs modelling choices in a similar manner to the analysis performed in Chapter 6, albeit for a different underlying discrete model. The C&G description is also less computationally intensive than performing sufficiently many realisations of the discrete model to obtain average behaviour. Furthermore, the additional information contained within the C&G description provides an accurate measure of the spatial clustering present in the discrete model, unlike other continuum descriptions. Hence the C&G description provides, for the first time, a robust method for describing both the number of occupied sites and the spatial clustering. We anticipate that the C&G description will be a valuable tool for interpreting biological and ecological processes where spatial clustering and spatial correlations are important.

In Chapter 8, we introduce competitive and co-operative mechanisms into lattice-based models of birth, death and movement, and examine the corresponding mean-field PDE description of the discrete model. Standard birth, death and movement processes, such as those examined in Chapter 6 and Chapter 7, involve an assumption that agents are influenced by nearest-neighbour agents through exclusion only. By imposing different rates of birth, death and movement depending on whether an agent on a lattice has zero nearest-neighbour agents (isolated) or at least one nearest-neighbour agent (grouped) we are able to describe both competitive and co-operative mechanisms. It was uncertain how introducing such mechanisms would influence the mean-field PDE description, which is a linear reaction-diffusion equation with Fisher kinetics for the model without competitive or competitive mechanisms [52]. Furthermore, it was unclear what implications any change in the governing PDE would have with respect to the long-time behaviour of the population.

By introducing a competitive or co-operative movement mechanism we find that the diffusivity function in the PDE description is a nonlinear function that depends on the overall agent density, as well as the rates of motility for both isolated and grouped agents. Interestingly, in certain parameter regimes, this nonlinear function has regions of negative diffusivity. We find that introducing competitive/co-operative birth or death mechanisms results in Allee kinetics, where population growth is negative below a threshold density [57]. Unlike the PDE description for the model without competitive or co-operative mechanisms, this negative growth results in long-time behaviour where the population recedes and eventually undergoes extinction. Additionally, we find that combining the nonlinear diffusivity function with Allee kinetics leads to a rich variety of travelling wave behaviour, including travelling waves containing shocks and sharp fronts. Travelling waves containing shocks had not been previously observed for single species reaction-diffusion equations [30, 64]. As biological phenomena such as tumour invasion have been observed with shock fronts [48], our work provides insight into the type of PDEs that give rise to such fronts.

In Chapter 9, we extend the C&G description to incorporate populations with spatial heterogeneities such that the description is suitable for describing moving fronts. Additionally, we derive a generalised description of competitive and co-operative mechanisms, which allows the birth, death and motility events to occur at different rates depending on the length of the chain that the agent belongs to. This allows for a description of the discrete model with competitive and/or co-operative mechanisms, presented in Chapter 8, that does not rely on the standard mean-field approximation. However, it was unclear whether the C&G description would accurately approximate the average behaviour in the corresponding discrete model, particularly with respect to the persistence or extinction of the population. Furthermore, it was unclear whether the C&G description would provide similar long-time predictions of the population as previous mean-field descriptions.

We find that the modified C&G description provides an accurate approximation to the average discrete behaviour in all parameter regimes. Critically, the C&G description accurately predicts the persistence or extinction of the population, as we observe that previous mean-field descriptions do not predict the correct qualitative long-time behaviour of the population in certain parameter regimes. The C&G description also provides information about the spatial clustering present within the moving front of the population. Hence the C&G description is an insightful technique for describing processes in cell biology and ecology that involve the spreading of a population, and determining potential control measures that lead to either a reduction in the invasion rate or the extinction of the population.

The work presented in Chapters 6-9 forms a thorough investigation into discrete exclusion-based birth, death and movement processes, and the corresponding continuum approximation required to accurately approximate the average discrete behaviour. We consider a number of different discrete processes, with respect to the number of subpopulations, the type of birth, death and movement mechanisms, and the geometry. For each process, we highlight where standard mean-field approximations are adequate, and where more sophisticated approximations are required. As such, Chapters 6-9 provides guidance about making an appropriate continuum approximation for any future models that involve birth, death and movement processes. In particular, we highlight the need for approximations that account for the spatial clustering present in the underlying discrete model. Previous continuum approximations are unable to explicitly describe spatial clustering and hence the C&G approximation presented in this thesis is an important tool for describing the average discrete behaviour in models that contain significant spatial clustering.

Overall, our work highlights the usefulness of mathematical models for obtaining quantitative insight into the mechanisms governing the collective behaviour of a cell population. Random walk models are relatively simple to implement and provide a flexible tool for modelling the mechanisms of individual cells. Continuum approximations of these random walk models allow for analysis of the behaviour of the cell population, provided the approximation made is appropriate. We present a thorough investigation of different continuum approximations and parameter regimes, and demonstrate when particular approximations are valid. Furthermore, we implement parameter recovery techniques that involve the calibration of random walk models to experimental data to recover parameter estimates for cellular mechanisms. As such, the work contained within this thesis demonstrates the power of using random walk models to describe collective cell behaviour, as these models are suitable for both parameter recovery and analysis of the population behaviour.

10.2 Future work

Each of the studies comprising this thesis has resulted in several potential avenues for future research. We discuss some of these ideas here and note that details regarding other ideas can be found in the discussion and conclusions sections of Chapters 2-9.

- **Extend the Fisher-Kolmogorov model for IncuCyte ZOOM$^{\text{TM}}$ assays to incorporate chemotactic effects**
  In our investigation we did not consider the influence of chemotaxis on the migration of the PC-3 cell population in the IncuCyte ZOOM$^{\text{TM}}$ assay. It is possible that a chemical gradient forms due to the non-uniform initial distribution of the cell population, which may lead to directed cellular motion along the chemical gradient. Mathematical models that allow for the influence of movement along a chemical gradient have been presented previously [26]. To calibrate such a model to IncuCyte ZOOM$^{\text{TM}}$ assay data, we would require a more sophisticated parameter estimation process to account for the four additional parameters in the model. For details of the potential model, see the Conclusions section of Chapter 2. Furthermore, additional experimental measurements would have to be recorded such that the chemical gradient throughout the experimental domain was known. Such an investigation would be instructive to determine whether chemotaxis plays an instrumental role in IncuCyte ZOOM$^{\text{TM}}$ assays.
- **Incorporate cell-to-cell adhesion to describe the collective migration of an adhesive cell population, and subsequently attempt to recover the cell parameters from leading edge data**
  The parameter recovery process implemented to obtain cell parameter estimates from leading edge data implicitly assumes that the cell population exhibits no cell-to-cell adhesion. We deliberately considered 3T3 fibroblasts for this investigation as these cells do not form clusters due to cell-to-cell adhesion [59]. It is known

that certain cell types, such as melanoma cells, exhibit strong cell-to-cell adhesion [28, 60]. Random walk models that mimic the behaviour of cell-to-cell adhesion through an additional parameter describing the strength of adhesion or repulsion have been investigated previously [21, 28]. It would be of interest to perform a scratch assay for a cell type that exhibits strong cell-to-cell adhesion and record the evolution of the position of the leading edge. Similarly, it would be of interest to combine this data with the extended random walk model to attempt to recover unique parameter estimates for the cell motility rate, cell proliferation rate and the cell-to-cell adhesion strength.

- **Relax the limitations of lattice-based movement and proliferation in the ABC parameter recovery process**
  Lattice-based random walk models of movement and proliferation impose a restriction on the location of agents, as the agents can only exist on a predefined lattice. Such a restriction is obviously unrealistic if the random walk model is applied to experimental data. For our ABC parameter recovery process we impose this limitation as lattice-free random walk models are more computationally intensive than lattice-based random walk models [24, 40, 41] and we require thousands of realisations of the random walk model to obtain reliable parameter estimates. While more computationally intensive, it would be instructive to interface a lattice-free random walk model with our ABC parameter recovery process, to elucidate whether the restriction on cell location influences the parameter estimates. If the parameter estimates are independent of the lattice restriction, the more efficient lattice-based random walk can be implemented in future parameter estimation processes. Conversely, if the parameter estimates depend on whether the agent locations are restricted to lattice sites, the lattice-free random walk may be required to obtain robust parameter estimates.
- **Introduce model parameters that are either temporally-dependent or local cell density-dependent and attempt to recover the parameter functions**
  During our investigation of the choice of combinations of observation times used in the ABC parameter recovery process we observed that the estimates of both the cell diffusivity and cell proliferation rate appear to depend on the choice of final observation time. Similar observations have been reported previously [9, 16]. It is not obvious what causes the parameters to be non-constant, nor is it obvious what form any putative parameter function would take. It would therefore be instructive to investigate whether a temporal or cell density dependence is present for either the cell diffusivity or cell proliferation rate. After a variable dependence is established, it would be instructive to attempt to ascertain an approximate relationship between the variable and the parameter through a similar ABC technique.
- **Extend the interactions of the corrected mean-field description to describe chemical kinetics or predator-prey interactions**
  The terms in the corrected mean-field description of two interacting subpopulations undergoing birth, death and movement that involve both subpopulations arise solely

from the exclusion-based nature of the discrete model. Processes such as chemical kinetics and predator-prey interactions involve the direct interaction between the two subpopulations [38, 54]. Hence the two-point distribution functions would involve the probability that two neighbouring sites are occupied by agents belonging to the different subpopulations, rather than the probability that an agent has a vacant neighbouring site. It would be instructive to derive the corrected mean-field description for this type of interaction and examine how the spatial correlations differ compared to the previous interactions. Furthermore, it would be interesting to apply the new corrected mean-field model to a relevant case study from ecology or chemistry [38, 54].

- **Derive and analyse the partial differential equation approximation of the C&G description**
  The governing equations of the C&G description are a system of ODEs, where the evolution of the number of chains and gaps of each length is described by an ODE. Both the mean-field description and corrected mean-field description have similar governing equations, where the average occupancy of each lattice site is governed by an ODE [51]. Unfortunately due to the number of governing equations, it is typically difficult to analyse the system. To address this, PDE approximations of the system of equations have been proposed for both mean-field and corrected mean-field descriptions [21, 34, 52]. Expanding the C&G governing equations with a Taylor series in an appropriate limit, such that both the chain length and the gap length could be treated as continuous variables, would result in a PDE approximation of the C&G governing equations. The resulting system would be two coupled PDEs describing the number of chains and the number of gaps, which is more to amenable to analysis than the large system of ODEs.
- **Implement isolated and grouped agents as separate subpopulations and analyse the resulting coupled multispecies partial differential equation model**
  The mean-field description of lattice-based co-operative/competitive birth, death and movement considered the most fundamental approach of modelling the sum of the density of isolated agents and the density of grouped agents. It would be possible to consider isolated agents and grouped agents as distinct subpopulations, which would result in a coupled multispecies PDE model instead of a single PDE model. The coupled system of PDEs contains additional information with respect to the relative densities of isolated and grouped agents. Furthermore, it is unclear how the two subpopulations combine to form a travelling wave in terms of overall agent density. Hence it would be instructive to perform a similar investigation into the types of travelling waves arising from the coupled system of PDEs.
- **Investigate the potential of the C&G description with regard to parameter recovery by calibrating the C&G description to experimental data**
  It has been observed previously that traditional continuum descriptions can result in incorrect parameter estimates when calibrated to relevant experimental data if spatial correlations are significant [53]. As the C&G description accurately reflects

the average behaviour in the discrete model for all parameter regimes, even in the presence of strong spatial correlations, it would be interesting to calibrate the C&G description to relevant experimental data and examine the recovered parameter estimates. If the parameter estimates are consistent with the estimates obtained from calibrating the underlying stochastic discrete model to the experimental data then there would be significant benefit in implementing a parameter recovery process with the C&G description. Stochastic parameter recovery techniques are computationally intensive as both Bayesian techniques [23] and techniques that involve the calculation of the average discrete behaviour across a potentially large parameter space require a large number of realisations of the discrete model [22]. Deterministic calibration techniques are more computationally efficient but require an accurate and robust deterministic continuum description.

- **Examine the viability of extending the C&G description into higher dimensions**
  The C&G description has been shown to provide an excellent approximation of the average behaviour of lattice-based birth, death and movement processes. However, this description is currently limited to one-dimensional processes due to the complexity of the governing equations. Further, the natural description of a chain or gap in terms of its location and length does not have not a clear analogue in higher dimensions. The C&G description arises from considering events based on the number of occupied nearest-neighbour sites, and how these events affect the length and location of chains and gaps. In higher dimensions, events can still depend on the number of occupied nearest-neighbour sites but it is not straightforward to describe the effect of these events on a two- or three-dimensional chain- or gap-equivalent. Hence any higher dimensional description is unlikely to explicitly describe all possible configurations of contiguous occupied sites and contiguous unoccupied sites. Instead, the description may be limited to describing the evolution of the number of occupied and unoccupied sites with $1, ..., 2d$ occupied nearest-neighbour sites, where d is the dimension. It would be instructive to investigate whether this reduced description provides a similarly accurate approximation of the average discrete behaviour in two and three dimensions, compared to the C&G description in one dimension.

10.3 Future Directions

Mathematical models provide a flexible and powerful tool for supplementing insight obtained from cell biology experiments. With the ever-increasing amount of computational power available, *in silico* experiments, based on sound mathematical and biological knowledge, now form an important part of the toolbox for attempting to understand biological processes. In particular, inference methods that were previously computationally intractable now prove instructive for quantifying the influence of particular mechanisms on collective cell behaviour. However, there is still significant work to be done in the field. Given the wide variety of models available, it is critical that an appropriate selection of model is made given the underlying biological mechanisms. For example, in the context of individual-based models, the types of available model include *lattice-based models*, where agents are restricted to a pre-defined lattice of the same size as the agent [8, 10, 11, 15, 17, 27, 52], *patch-based or voxel models*, where agents exist within a voxel that is larger than the agents, and agents can transition between voxels [6, 7, 36], and *lattice-free models*, where agents move through continuous space [24, 40]. Further, each of these models can incorporate exclusion effects, increasing the number of models that could be implemented to describe an experiment. While each of these types of models has proved valid in certain cases, it is critical that the limitations of each model are well described such that an informed choice of model can be made. For example, if the cell line in the system of interest displays a distribution of sizes then imposing a constant diameter in a lattice-based model is inappropriate [7]. Studies that provide comparisons between modelling approaches and highlight the assumptions required for each approach to be valid would prove beneficial for the more widespread adoption of modelling techniques in biological investigations.

Relatively simple mathematical models of collective cell behaviour have had success in explaining experimental data, but it is clear that these models do not reflect all of the cell behaviour. There is, therefore, a need to develop more complex mathematical models to incorporate this behaviour, such as the putative dependence of cellular parameters on time after wounding or passage number [19, 20]. Care must be taken, however, that these future models are developed in a manner that allows for the experimental validation of proposed mechanisms. There is great potential for mathematical models to uncover biological information currently hidden in experimental data, if the models are implemented in a rigorous and biologically-consistent manner. Further, it is imperative that these models are presented in a way that is accessible for the wide range of researchers working in biology, mathematical biology and computational biology, to ensure that the models are adopted and the potential insight is realised.

10.4 Final Remarks

Developing mathematical techniques for extracting quantitative information from cell biology experiments is critical for robustly analysing the influence of putative chemical treatments and experimental design choices. In this thesis we have presented several modelling frameworks for obtaining reliable estimates of the cell diffusivity and the cell proliferation rate from a variety of experimental data. While all of the frameworks provide parameter estimates that are consistent, increasing either the experimental or mathematical complexity of the framework can provide additional information about the parameters. Investigating the level of complexity required to accurately model relevant experimental procedures, and to subsequently recover reliable parameter estimates, will prove instructive for future combinations of mathematical and experimental investigations. Similarly, obtaining insight into the role of spatial correlations in lattice-based models, and their subsequent influence on the type of continuum description required to accurately approximate the average behaviour, will inform future modelling choices. Both types of mathematical techniques can be utilised to further understand the processes governing phenomena such as tumour invasion, wound healing and the invasion of alien species.

Bibliography

[1] Aichele, K., Bubel, M., Deubel, G., Pohlemann, T. and Oberringer, M. [2013], ‘Bromelain down-regulates myofibroblast differentiation in an in vitro wound healing assay’, *Naunyn-Schmiedeberg’s Archives of Pharmacology* **386**, 853–763.

[2] Baker, R. E. and Simpson, M. J. [2010], ‘Correcting mean-field approximations for birth-death-movement processes’, *Physical Review E* **82**(4), 041905.

[3] Binder, B. J. and Simpson, M. J. [2013], ‘Quantifying spatial structure in experimental observations and agent-based simulations using pair-correlation functions’, *Physical Review E* **88**(2), 022705.

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

[5] Chigurupati, S., Mughal, M. R., Okun, E., Das, S., Kumar, A., McCaffery, M., Seal, S. and Mattson, M. P. [2012], ‘Effects of cerium oxide nanoparticles on the growth of keratinocytes, fibroblasts and vascular endothelial cells in cutaneous wound healing’, *Biomaterials* **34**, 2194–2201.

[6] Cianci, C., Smith, S. and Grima, R. [2016], ‘Molecular finite-size effects in stochastic models of equilibrium chemical systems’, *The Journal of Chemical Physics* **144**(8), 084101.

[7] Cianci, C., Smith, S. and Grima, R. [2017], ‘Capturing brownian dynamics with an on-lattice model of hard-sphere diffusion’, *Physical Review E* **95**(5), 052118.

[8] Codling, E. A., Plank, M. J. and Benhamou, S. [2008], ‘Random walk models in biology’, *Journal of the Royal Society Interface* **5**, 813–834.

[9] Curtis, C. W. and Bortz, D. M. [2012], ‘Propagation of fronts in the Fisher-Kolmogorov equation with spatially varying diffusion’, *Physical Review E* **86**, 066108.

[10] Deroulers, C., Aubert, M., Badoual, M. and Grammaticos, B. [2009], ‘Modeling tumor cell migration: from microscopic to macroscopic models’, *Physical Review E* **79**(3), 031917.

[11] Derrida, B., Evans, M., Hakim, V. and Pasquier, V. [1993], ‘Exact solution of a 1D asymmetric exclusion model using a matrix formulation’, *Journal of Physics A* **26**(7), 1493.

[12] Fishman, D. A., Liu, Y., Ellerbroek, S. M. and Stack, M. S. [2001], ‘Lysophosphatidic acid promotes matrix metalloproteinase (MMP) activation and MMP-dependent invasion in ovarian cancer cells’, *Cancer Research* **61**(7), 3194–3199.

[13] Gingis-Velitski, S., Loven, D., Benayoun, L., Munster, M., Bril, R., Voloshin, T., Alishekevitz, D., Bertolini, F. and Shaked, Y. [2011], ‘Host response to short-term,

single-agent chemotherapy induces matrix metalloproteinase-9 expression and accelerates metastasis in mice', *Cancer Research* **71**(22), 6986–6996.

[14] Gujral, T. S., Chan, M., Peshkin, L., Sorger, P. K., Kirschner, M. W. and MacBeath, G. [2014], 'A noncanonical frizzled2 pathway regulates epithelial-mesenchymal transition and metastasis', *Cell* **159**, 844–856.

[15] Hackett-Jones, E. J., Landman, K. A. and Fellner, K. [2012], 'Aggregation patterns from nonlocal interactions: Discrete stochastic and continuum modeling', *Physical Review E* **85**(4), 041912.

[16] Hammond, J. F. and Bortz, D. M. [2011], 'Analytical solutions to Fishers equation with time-variable coefficients', *Applied Mathematics and Computation* **218**, 2497–2508.

[17] Illien, P., Bénichou, O., Mejía-Monasterio, C., Oshanin, G. and Voituriez, R. [2013], 'Active transport in dense diffusive single-file systems', *Physical Review Letters* **111**(3), 038102.

[18] Jarrard, D. F., Blitz, B. F., Smith, R. C., Patai, B. L. and Rukstalis, D. B. [1994], 'Effect of epidermal growth factor on prostate cancer cell line PC3 growth and invasion', *Prostate* **24**, 46–53.

[19] Jin, W., Penington, C. J., McCue, S. W. and Simpson, M. J. [2017], 'A computational modelling framework to quantify the effects of passaging cell lines', *bioRxiv* p. 161265.

[20] Jin, W., Shah, E. T., Penington, C. J., McCue, S. W., Maini, P. K. and Simpson, M. J. [2016], 'Logistic proliferation of cells in scratch assays is delayed', *bioRxiv* p. 077388.

[21] Johnston, S. T., Simpson, M. J. and Baker, R. E. [2012], 'Mean-field descriptions of collective migration with strong adhesion', *Physical Review E* **85**(5), 051922.

[22] Johnston, S. T., Simpson, M. J. and McElwain, D. L. S. [2014], 'How much information can be obtained from tracking the position of the leading edge in a scratch assay?', *Journal of the Royal Society Interface* **11**(97), 20140325.

[23] Johnston, S. T., Simpson, M. J., McElwain, D. L. S., Binder, B. J. and Ross, J. V. [2014], 'Interpreting scratch assays using pair density dynamics and approximate Bayesian computation', *Open Biology* **4**, 140097.

[24] Johnston, S. T., Simpson, M. J. and Plank, M. J. [2013], 'Lattice-free descriptions of collective motion with crowding and adhesion', *Physical Review E* **88**(6), 062720.

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

[26] Keller, E. F. and Segel, L. A. [1971], 'Model for chemotaxis', *Journal of Theoretical Biology* **30**(2), 225–234.

[27] Khain, E., Katakowski, M., Charteris, N., Jiang, F. and Chopp, M. [2012], 'Migration of adhesive glioma cells: Front propagation and fingering', *Physical Review E* **86**(1), 011904.

[28] Khain, E., Katakowski, M., Hopkins, S., Szalad, A., Zheng, X., Jiang, F. and Chopp, M. [2011], 'Collective behavior of brain tumor cells: the role of hypoxia', *Physical Review E* **83**(3), 031920.

[29] Knecht, D. A., LaFleur, R. A., Kahsai, A. W., Argueta, C. E., Beshir, A. B. and Fenteany, G. [2010], 'Cucurbitacin I inhibits cell motility by indirectly interfering with actin dynamics', *PLOS ONE* **5**(11), e14039.

[30] Landman, K. A. and White, L. R. [2011], 'Terraced spreading of nanofilms under a nonmonotonic disjoining pressure', *Physics of Fluids* **23**(1), 012004.

[31] Liang, C.-C., Park, A. Y. and Guan, J.-L. [2007], 'In vitro scratch assay: a convenient and inexpensive method for analysis of cell migration in vitro', *Nature Protocols* **2**(2), 329–333.

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

[33] Marjoram, P., Molitor, J., Plagnol, V. and Tavaré, S. [2003], 'Markov chain Monte Carlo without likelihoods', *Proceedings of the National Academy of Sciences* **100**(26), 15324–15328.

[34] Markham, D. C., Simpson, M. J. and Baker, R. E. [2013], 'Simplified method for including spatial correlations in mean-field approximations', *Physical Review E* **87**(6), 062702.

[35] Markham, D. C., Simpson, M. J., Maini, P. K., Gaffney, E. A. and Baker, R. E. [2013], 'Incorporating spatial correlations into multispecies mean-field models', *Physical Review E* **88**(5), 052713.

[36] McKane, A. J. and Newman, T. J. [2004], 'Stochastic models in population biology and their deterministic analogs', *Physical Review E* **70**(4), 041902.

[37] Meier, F. M.-P., Frommer, K. W., Peters, M. A., Brentano, F., Lefèvre, S., Schröder, D., Kyburz, D., Steinmeyer, J., Rehart, S., Gay, S. et al. [2012], 'Visfatin/pre-b-cell colony-enhancing factor (PBEF), a proinflammatory and cell motility-changing factor in rheumatoid arthritis', *Journal of Biological Chemistry* **287**(34), 28378–28385.

[38] Murrell, D. J. [2005], 'Local spatial structure and predator-prey dynamics: counter-intuitive effects of prey enrichment', *The American Naturalist* **166**(3), 354–367.

[39] Nardini, J. T., Chapnick, D. A., Liu, X. and Bortz, D. M. [2016], 'Modeling keratinocyte wound healing dynamics: cell–cell adhesion promotes sustained collective migration', *Journal of Theoretical Biology* **400**, 103–117.

[40] Plank, M. J. and Simpson, M. J. [2012], 'Models of collective cell behaviour with crowding effects: comparing lattice-based and lattice-free approaches', *Journal of the Royal Society Interface* **9**(76), 2983–2996.

[41] Plank, M. J. and Simpson, M. J. [2013], 'Lattice-free models of cell invasion: Discrete simulations and travelling waves', *Bulletin of Mathematical Biology* **75**(11), 2150–2166.

[42] Rausch, V., Liu, L., Apel, A., Rettig, T., Gladkich, J., Labsch, S., Kallifatidis, G., Kaczorowski, A., Groth, A., Gross, W. et al. [2012], 'Autophagy mediates survival of pancreatic tumour-initiating cells in a hypoxic microenvironment', *The Journal of Pathology* **227**(3), 325–335.

[43] Riahi, R., Yang, Y., Zhang, D. D. and Wong, P. K. [2012], ‘Advances in wound-healing assays for probing collective cell migration’, *Journal of Laboratory Automation* **17**(1), 59–65.

[44] Ross, R., Baker, R., Ford, M., Mort, R. and Yates, C. [2016], ‘Using approximate Bayesian computation to quantify cell-cell adhesion parameters in a cell migratory process’, *bioRxiv* p. 068791.

[45] Salomon, C., Kobayashi, M., Ashman, K., Sobrevia, L., Mitchell, M. D. and Rice, G. E. [2013], ‘Hypoxia-induced changes in the bioactivity of cytotrophoblast-derived exosomes’, *PLOS ONE* **8(11)**, e79636.

[46] Salomon, C., Torres, M. J., Kobayashi, M., Scholz-Romero, K., Sobrevia, L., Dobierzewska, A., Illanes, S. E., Mitchell, M. D. and Rice, G. E. [2014], ‘A gestational profile of placental exosomes in maternal plasma and their effects on endothelial cell migration’, *PLOS ONE* **9(6)**, e98667.

[47] Sengers, B. G., Please, C. P. and Oreffo, R. O.-C. [2007], ‘Experimental characterization and computational modelling of two-dimensional cell spreading for skeletal regeneration’, *Journal of the Royal Society Interface* **4**(17), 1107–1117.

[48] Sewalt, L., Harley, K., van Heijster, P. and Balasuriya, S. [2016], ‘Influences of Allee effects in the spreading of malignant tumours’, *Journal of Theoretical Biology* **394**, 77–92.

[49] Sharili, A.-S., Allen, S., Smith, K., Price, J. and McGonnell, I. M. [2013], ‘Snail2 promotes osteosarcoma cell motility through remodelling of the actin cytoskeleton and regulates tumor development’, *Cancer Letters* **333**, 170–179.

[50] Shibata, A., Tanabe, E., Inoue, S., Kitayoshi, M., Okimoto, S., Hirane, M., Araki, M., Fukushima, N. and Tsujiuchi, T. [2013], ‘Hydrogen peroxide stimulates cell motile activity through LPA receptor-3 in liver epithelial WB-F344 cells’, *Biochemical and Biophysical Research Communications* **433**, 317–321.

[51] Simpson, M. J. and Baker, R. E. [2011], ‘Corrected mean-field models for spatially dependent advection-diffusion-reaction phenomena’, *Physical Review E* **83**, 051922.

[52] Simpson, M. J., Landman, K. A. and Hughes, B. D. [2010], ‘Cell invasion with proliferation mechanisms motivated by time-lapse data’, *Physica A: Statistical Mechanics and its Applications* **389**(18), 3779–3790.

[53] Simpson, M. J., Sharp, J. A. and Baker, R. E. [2014], ‘Distinguishing between mean-field, moment dynamics and stochastic descriptions of birth–death–movement processes’, *Physica A: Statistical Mechanics and its Applications* **395**, 236–246.

[54] Singh, A. and Hespanha, J. P. [2011], ‘Approximate moment dynamics for chemically reacting systems’, *Automatic Control, IEEE Transactions on* **56**(2), 414–418.

[55] Swanson, K. R. [2008], ‘Quantifying glioma cell growth and invasion in vitro’, *Mathematical and Computer Modelling* **47**(5), 638–648.

[56] Swanson, K. R., Bridge, C., Murray, J.-D. and Alvord Jr, E. C. [2003], ‘Virtual and real brain tumors: using mathematical modeling to quantify glioma growth and invasion’, *Journal of the Neurological Sciences* **216**(1), 1–10.

[57] Taylor, C. M. and Hastings, A. [2005], ‘Allee effects in biological invasions’, *Ecology Letters* **8**(8), 895–908.

[58] Teppo, S., Sundquist, E., Vered, M., Holappa, H., Parkkisenniemi, J., Rinaldi, T., Lehenkari, P., Grenman, R., Dayan, D., Risteli, J. et al. [2012], 'The hypoxic tumor microenvironment regulates invasion of aggressive oral carcinoma cells', *Experimental Cell Research* **319**, 376–389.

[59] Todaro, G. J. and Green, H. [1963], 'Quantitative studies of the growth of mouse embryo cells in culture and their development into established lines', *The Journal of Cell Biology* **17**(2), 299–313.

[60] Treloar, K. K., Simpson, M. J., Haridas, P., Manton, K. J., Leavesley, D. I., McElwain, D.-L. S. and Baker, R. E. [2013], 'Multiple types of data are required to identify the mechanisms influencing the spatial expansion of melanoma cell colonies', *BMC Systems Biology* **7**(1), 137.

[61] Tremel, A., Cai, A., Tirtaatmadja, N., Hughes, B.-D., Stevens, G.-W., Landman, K.-A. and O'Connor, A.-J. [2009], 'Cell migration and proliferation during monolayer formation and wound healing', *Chemical Engineering Science* **64**(2), 247–253.

[62] Upton, Z., Wallace, H. J., Shooter, G. K., van Lonkhuyzen, D. R., Yeoh-Ellerton, S., Rayment, E. A., Fleming, J. M., Broszczak, D., Queen, D., Sibbald, R. G. et al. [2011], 'Human pilot studies reveal the potential of a vitronectin: growth factor complex as a treatment for chronic wounds', *International Wound Journal* **8**(5), 522–532.

[63] Vo, B. N., Drovandi, C. C., Pettitt, A. N. and Simpson, M. J. [2015], 'Quantifying uncertainty in parameter estimates for stochastic models of collective cell spreading using approximate Bayesian computation', *Mathematical Biosciences* **263**, 133–142.

[64] Witelski, T. P. [1995], 'Shocks in nonlinear diffusion', *Applied Mathematics Letters* **8**(5), 27–32.

[65] Zhao, X., Li, D.-C., Zhu, X.-G., Gan, W.-J., Li, Z., Xiong, F., Zhang, Z.-X., Zhang, G.-B., Zhang, X.-G. and Zhao, H. [2013], 'B7-H3 overexpression in pancreatic cancer promotes tumor progression', *International Journal of Molecular Medicine* **31**(2), 283–291.