

Modelling cell migration, proliferation and interactions on growing domains



Robert Ross
Pembroke College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Trinity Term 2016

Supervisors:

Professor Ruth E. Baker, Wolfson Centre for Mathematical Biology,
Mathematical Institute, University of Oxford

Dr Christian A. Yates, Centre for Mathematical Biology, Department
of Mathematical Sciences, University of Bath

Abstract

Motile and proliferative cell populations in growing tissues are a common feature of mammalian embryonic development. The aim of the work presented in this thesis is to develop models that better describe the behaviour of motile and proliferative cell populations on growing domains. We motivate our study by investigating the behaviour of melanoblasts, a migratory and proliferative cell population responsible for fur pigmentation in the embryonic mouse. We then develop mathematical models to describe cell populations with cell-cell interactions such as adhesion and repulsion, cell populations where spatial correlations established by cell proliferation play an important role, and scenarios where cell populations consist of different species. We also develop methods to improve the identification of motility parameters in migratory cell populations. The processes we model often occur at developmental phases that are difficult to observe experimentally, and so mathematical and computational modelling can play a central role in understanding the behaviour of motile and proliferative cell populations in growing tissues.

Acknowledgements

First and foremost I would like to acknowledge my supervisors Ruth and Kit. I am grateful to them for both their patience and high standards. I would also like to acknowledge Richard, who I collaborated with to produce much of the contents of Chapters 2 and 7 of this thesis. It is important to acknowledge all those at the Doctoral Training Centre who secured funding for the Systems Biology course. Finally, I would like to acknowledge my Mum, Dad, Sister and Aadhya, as without them none of this would matter anyway.

Contents

List of Figures	iv
List of Tables	vii
1 Introduction	1
1.1 The role of mathematical and computational modelling in developmen- tal biology research	1
1.2 Individual-based and continuum models	2
1.2.1 Individual-based models	3
1.2.2 Continuum models	5
1.3 Thesis outline	6
2 Modelling neural crest stem cell migration on a growing domain	9
Overview	9
2.1 The development of neural crest stem cells	10
2.1.1 Phenotypes associated with melanoblast colonisation of the dor- sal lateral epithelium	12
2.2 Individual-based model	16
2.2.1 Inferring parameters for the individual-based model from ex- perimental data	18
2.3 Results	22
2.4 Summary	32
3 Modelling cell-cell interactions and cell pathlines on a growing do- main	35
Overview	35
3.1 Introduction	36
3.2 Model formulation	38
3.2.1 Individual-based model	38
3.2.2 Transition probabilities that represent cell-cell interactions . .	40
3.2.3 Population-level density continuum equations	46
3.2.4 Cell trajectories	51
3.3 Comparing continuum approximations to individual-based model results	54
3.3.1 Static domain	56
3.3.2 Growing domain	62
3.3.3 Tagged agent trajectories	65

3.4	Summary	68
4	Modelling spatial correlations on a growing domain in one dimension	71
	Overview	71
4.1	Introduction	72
4.2	Model	74
4.2.1	One-dimensional individual-based model and the domain growth mechanism	74
4.2.2	System of ordinary differential equations	75
4.2.3	Agent motility and proliferation	80
4.3	Results	87
4.3.1	Motile agents on a growing domain	90
4.3.2	Motile, proliferative agents on a growing domain	93
4.4	Integrating domain growth into correlation functions	96
4.4.1	Correlation function	99
4.4.2	Results for correlation ordinary differential equation model . .	100
4.5	Summary	104
5	Modelling spatial correlations on a growing domain in two dimensions	106
	Overview	106
5.1	Introduction	107
5.2	Model	107
5.2.1	IBM and domain growth mechanisms	108
5.2.2	Individual density functions	109
5.2.3	Pairwise density functions	113
5.3	Results	129
5.3.1	Agents on a growing domain	130
5.3.2	Agent motility, proliferation and death	131
5.3.3	Agent motility and proliferation	133
5.4	Summary	142
6	Modelling spatial correlations on a growing domain in two dimensions with two species	144
	Overview	144
6.1	Introduction	145
6.2	Model	146
6.2.1	IBM containing two species	146
6.2.2	Individual density functions	147
6.2.3	Pairwise density functions	148
6.3	Results	152
6.3.1	Uniform domain growth	156
6.3.2	Non-uniform domain growth	164
6.4	Summary	168

7	Model parameterisation for a simple wound healing experiment	171
	Overview	171
7.1	Introduction	172
7.2	Methods	176
	7.2.1 Individual-based model	176
	7.2.2 Cell-cell adhesion models	177
	7.2.3 Summary statistics	179
	7.2.4 Approximate Bayesian computation	182
7.3	Results	184
	7.3.1 Ideal experiment	184
	7.3.2 Practically realisable experiment	188
	7.3.3 Improving the experimental design	197
7.4	Summary	198
8	Discussion	201
	8.1 Conclusions and future work	202
	Appendices	207
A	Appendices for Chapter 3	207
	A.1 Example transition probabilities	207
B	Appendices for Chapter 5	209
	B.1 Individual density functions	209
	B.2 Pairwise density functions	211
C	Appendices for Chapter 6	214
	C.1 Individual density functions	214
	C.2 Pairwise density functions	216
D	Appendices for Chapter 7	219
	D.1 Experimental methods	219
	D.2 Further variance plots for models A and B for the PCF summary statistic	220
	Bibliography	224

List of Figures

1.1	A lattice-based model.	4
2.1	The migration path of melanoblasts in the developing mouse embryo.	11
2.2	Chimeric mouse.	13
2.3	The original hypothesis for the mechanism of stripe formation in chimeric mice.	14
2.4	<i>LaacZ</i> mouse embryo.	15
2.5	Belly spot phenotype in a mouse.	16
2.6	The domain growth mechanism for the IBM.	18
2.7	An example of the initial configuration of agents in the IBM.	20
2.8	The diffusion coefficient and cell cycle time in the IBM.	20
2.9	A time-lapse sequence of melanoblasts migrating in ex vivo culture of E14.5 skin.	23
2.10	The PCF analysis on experimental data for melanoblast spatial organisation at E14.5.	26
2.11	A comparison of the experimental melanoblast densities with results from the IBM for E10.5 to E15.5.	27
2.12	Two examples of diffusive <i>LaacZ</i> patterns described in Wilkie et al. [6]	28
2.13	Identification of stripes in the IBM.	29
2.14	Stripe intensity vs. number of starting genotypes.	30
2.15	An example of a chimeric pattern from a single simulation of the IBM initialised with either agents of two genotypes in equal proportions or with twenty-one differently coloured genotypes.	31
2.16	The belly-spotting phenotype in the IBM is more sensitive to a decrease in melanoblast proliferation rate.	32
2.17	Model phenotypes for different values of T_c	33
3.1	The nearest neighbourhood occupancy sets.	43
3.2	Myopic neighbourhood example.	45
3.3	Concavity measure.	56
3.4	Agent population density and expected displacement for agents with blind interactions.	57
3.5	Agent population density and expected displacement for cells with repulsive and myopic interactions.	60
3.6	The evolution of the SEM for the position of the tagged agent for all agent interactions.	62

3.7	Cell density and expected displacement for cells with adhesive and repulsive cell interactions on a growing domain.	63
3.8	The expected displacement trajectory of a tagged agent in the positive x direction for all interactions with respect to time.	65
3.9	Agent trajectory concavity measure with respect to adhesion parameter α for a tagged agent initially situated at (114, 10).	66
4.1	A one-dimensional domain with periodic boundary conditions can be represented as a circle.	76
4.2	The evolution of the individual densities, $c^L(t)$, for two different sets of parameters and initial conditions.	91
4.3	The evolution of the pairwise densities, $\rho_2^L(r, t)$, for two different sets of parameters and initial conditions.	92
4.4	The evolution of the individual densities, $c^L(t)$, with agent proliferation for two different sets of parameters and initial conditions.	94
4.5	The evolution of the pairwise densities, $\rho_2^L(r, t)$, with agent proliferation for two different sets of parameters and initial conditions.	95
4.6	The relative error, calculated with Eq. (4.34), in the individual densities due to approximations Eq. (4.32) and (4.33) for different domain lengths.	98
4.7	The evolution of the macroscopic agent density as given by Eqs. (4.24), (4.41) and (4.42).	102
4.8	The macroscopic agent density with agent proliferation.	103
5.1	Before and after a growth event for both (a) GM1 and (b) GM2. . . .	110
5.2	The two types of configuration of lattice sites: (a) colinear and (b) diagonal.	113
5.3	Measuring the distance of a lattice site from the boundary.	119
5.4	(Colour online): The evolution of the colinear and diagonal components of the pairwise density functions for (a) GM1 and (b) GM2. $P_{gx} = P_{gy} = 0.1$	131
5.5	For GM1 including pairwise correlations in a correlation ODE model provides a more accurate approximation of the evolution of the macroscopic density in the IBM than the MFA.	134
5.6	Increasing the growth rates, P_{gx} and P_{gy} , causes the pairwise correlations to augment for GM1 in the correlation ODE model.	135
5.7	For GM2 including pairwise correlations in a correlation ODE model provides a more accurate approximation of the results of the evolution of the macroscopic density in the IBM.	136
5.8	For GM2 increasing the growth rates, P_{gx} and P_{gy} , causes the pairwise correlations to shrink in the correlation ODE model.	137
5.9	Altering the initial size of the domain does not cause the evolution of the agent density to change for GM1, but does for GM2.	139
5.10	Altering the initial size of the domain does not change the evolution of the pairwise correlations in the case of GM1.	140

5.11	Altering the initial size of the domain changes the evolution of the pairwise correlations in the case of GM2.	141
6.1	Including the effects of pairwise correlations renders the correlations ODE model (Eq. (6.15)) able to accurately approximate the averaged results from the IBM, whereas the standard MFA (Eq. (6.16)) cannot.	157
6.2	Including the effects of pairwise correlations renders the correlations ODE model (Eq. (6.15)) able to accurately approximate the averaged results from the IBM, whereas the standard MFA (Eq. (6.16)) cannot.	157
6.3	The evolution of the correlation functions for $F_{A,A}$, $F_{B,B}$ and $F_{A,B}$ with exponential growth via GM1.	160
6.4	The evolution of the correlation functions for $F_{A,A}$, $F_{B,B}$ and $F_{A,B}$ with exponential growth via GM2.	161
6.5	Altering the initial domain size causes the evolution of both species to change in GM2 but not in GM1.	162
6.6	Including the effects of pairwise correlations renders the correlations ODE model (Eqs. (6.19) and (6.21)) able to accurately approximate the averaged results from the IBM, whereas the standard MFA cannot.	163
6.7	The non-uniform domain growth examples.	165
6.8	Apical domain growth.	166
6.9	Enteric domain growth.	167
7.1	The initial condition in the IBM for the ideal experiment.	177
7.2	Posterior distributions for model A for an ideal experiment with different summary statistics.	187
7.3	Typical initial frame of the experimental data.	189
7.4	Posterior distributions for simulations of the experiment described in Section 2.5 using the PCF as a summary statistic for an IBM of size $L_x = 23$ and $L_y = 23$	194
7.5	The variance in the PCF synthetic data for model B with $P_m = 0.5$, $\alpha = 0.25$ and different IBM domain sizes.	194
7.6	Increasing the size of the simulation domain while keeping the initial column densities the same.	195
7.7	Posterior distributions for simulations of the experiment using the PCF as a summary statistic for an IBM simulated on a domain of size $L_x = 23$ by $L_y = 23$ with synthetic data generated from 500 repeats.	198
D.1	The variance in PCF synthetic data for model A with $P_m = 0.5$, $\alpha = 0.1$ for different IBM domain sizes.	221
D.2	The variance in the synthetic data for the adhesion model B with $P_m = 0.5$, $\alpha = 0.25$ for different IBM domain sizes.	222
D.3	The variance in PCF synthetic data for model B with $P_m = 0.5$, $\alpha = -0.1$ for different IBM domain sizes.	223

List of Tables

2.1	IBM parameters.	18
3.1	Population-level density equations for adhesive, repulsive and myopic agent interactions.	49
3.2	Expected displacement of a tagged agent initially at $[p_x(0), p_y(0)]$ for adhesive, repulsive and myopic agent interactions.	53
7.1	Average initial condition for the IBM.	190

List of Acronyms

ABC	Approximate Bayesian computation
CSR	Complete spatial randomness
DFFT	Discrete fast Fourier transform
DLE	Dorsal lateral epithelium
IBM	Individual-based model
KSA	Kirkwood superposition approximation
LHS	Left-hand side
LPE	Largest percentage error
MFA	Mean-field approximation
MSD	Mean square displacement
NCSCs	Neural crest stem cells
ODE	Ordinary differential equation
PCF	Pair-correlation function
PDE	Partial differential equation
RHS	Right-hand side
RSUM	Random sequential update method
SEM	Standard error of the mean
SSA	Stochastic simulation algorithm

Chapter 1

Introduction

Embryonic development requires the orchestrated movement and proliferation of cells in growing tissues. For instance, cranial neural crest stem cells, a subset of a migratory cell population that give rise to a diverse lineage, have been shown to migrate along the growing cranofacial region in embryonic chickens [1–3]. Similarly melanoblasts, neural crest precursors to melanocytes, have been shown to migrate through the growing dorsal lateral epithelium (DLE) in the embryonic mouse [4–6]. Therefore, the inclusion of domain growth in models of the aforementioned cell migration processes is essential. To this end, the aim of the work presented in this thesis is to develop models that better describe the behaviour of motile and proliferative cells on growing domains.

1.1 The role of mathematical and computational modelling in developmental biology research

Mathematical and computational modelling are important research tools in developmental biology. One reason for this is that certain aspects of biological systems remain difficult to study with experimental techniques, for instance domain growth [7]. Biological domain growth is difficult to study *in vitro*, as experimentally recreating domain growth is often not possible. Meanwhile, *in vivo* experiments studying domain growth are typically not feasible. Although only part of the solution, math-

ematical and computational modelling are able to alleviate some of the difficulties associated with implementing experiments in developmental biology. First and foremost, mathematical and computational modelling can help with experimental design and hypothesis testing. For instance, theoretical tools can be used to augment the information content of experiments and help improve parameter identifiability from experimental data [8, 9]. Mathematical and computational modelling can also test the feasibility of a range of hypotheses given what is known about a biological system. This can save both time and money for experimental biologists, as hypotheses can be rejected without the need for experimental testing. Secondly, mathematical and computational modelling can also generate new hypotheses about biological systems that can be tested experimentally (albeit with the caveat that these experiments may not be currently possible).

The contents of this thesis display both of the aforementioned roles of mathematical and computational research in developmental biology. In Chapters 2 and 7 we use computational methods in conjunction with experimental data to test hypotheses and improve experimental design [10, 11]. In Chapters 3, 4, 5 and 6 we develop new mathematical methods, and from these methods generate novel hypotheses about the effects of domain growth on populations of motile and proliferative cells [12–15].

1.2 Individual-based and continuum models

Throughout this thesis a combination of individual-based models (IBMs) and continuum models are presented. The approach where both IBMs and continuum models are used in tandem for addressing a research problem is widespread in mathematical and computational modelling in biology [16–41]. The reasons for using both IBMs and continuum models in tandem are manifold; the most important of which, in relation to the work presented in this thesis, are discussed below.

1.2.1 Individual-based models

All of the IBMs presented in this thesis take the form of agent-based (individual) discrete (lattice-based) random-walk (stochastic) models with volume exclusion. Figure 1.1 displays the typical set-up of the lattice-based models we implement throughout this thesis.

In all of the IBMs presented in this thesis a single agent is allowed per lattice site, from which it can move or proliferate into adjacent sites. Throughout this thesis the agents represent cells, although they do not necessarily have to. Agent movement and proliferation events occur with specified probabilities per unit time. If an agent attempts to move or proliferate into a site that is already occupied by another agent, the event is aborted. This aspect of the IBM is referred to as ‘volume exclusion’. The IBMs presented in this thesis are simulated stochastically. By this it is meant that different realisations of the same IBM will generally yield different results. Two different stochastic simulation algorithms (SSAs) are implemented in this thesis: i) the Gillespie algorithm [42]; and ii) the random sequential update method (RSUM) [43]. These two SSAs will be explained when necessary in the text.

The suitability of IBMs for representing cell populations is readily apparent. IBMs provide an intuitive representation of cells, and allow for complex behaviours representing biological processes, such as cell-cell interactions and volume exclusion, to be easily assigned to agents in the IBM [12, 33–35]. Importantly, IBMs can capture the effects of spatial correlations and heterogeneity in agent populations, and the ramifications spatial correlations can have on processes such as cell proliferation [18–23, 44]. As such, IBMs have played an important role in research focussing on cell migratory processes [2, 3, 10]. For example, our study using an IBM has shown, in the case of melanoblasts, that spatial correlations established by melanoblast proliferation es-

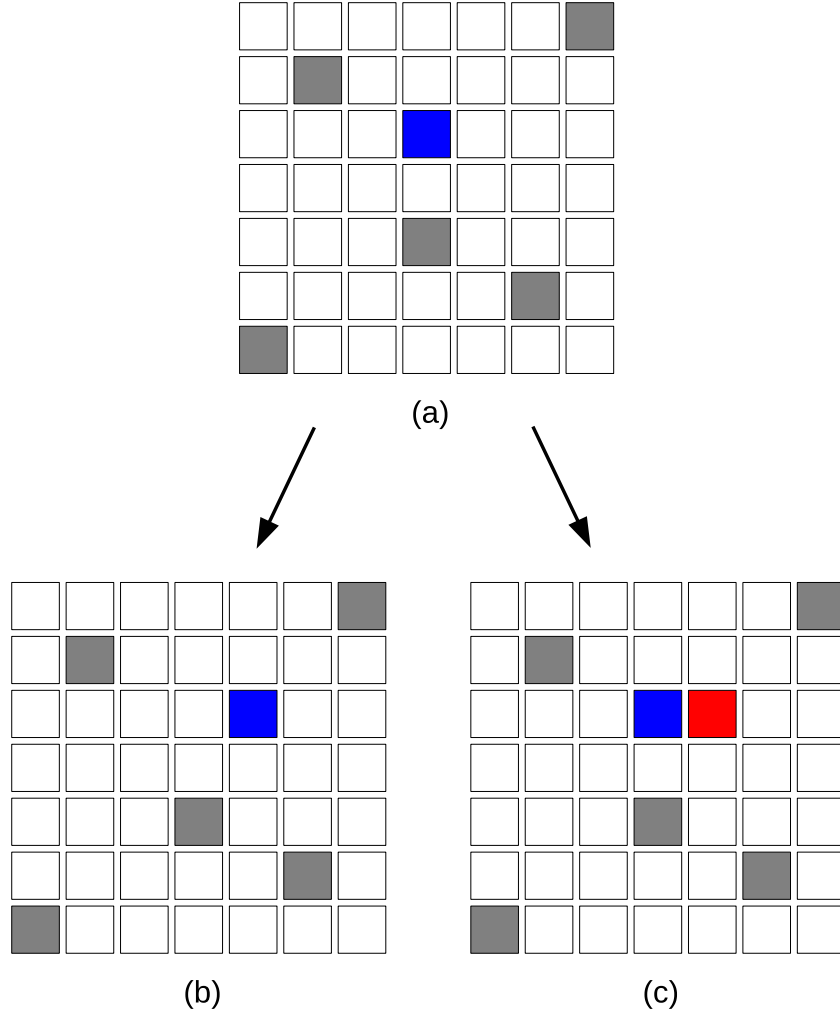


Figure 1.1: A lattice-based model. The lattice sites occupied by an agent are indicated by coloured sites (blue, grey and red). Only one agent is allowed per site, and this agent is able to move or proliferate into adjacent sites at a given rate. For instance, the agent in blue in (a) has been chosen to attempt a movement or proliferation event. This blue agent can move or proliferate into any one of the four empty adjacent sites next to it. If this blue agent moves in the positive horizontal direction (right) the result is the configuration displayed in (b). If this agent proliferates into the lattice site in the positive horizontal direction the result is the configuration displayed in (c), where the red agent is the blue agent's progeny.

establish fur patterning phenotypes [10]. Beyond studies in cell migration, IBMs have been widely applied to the study of other biological phenomena. For instance, IBMs have been used to study spatial structure in cell aggregates [44, 45], the growth of

cell colonies [18], enteric colonic growth [24], and infection spreading and ecological interactions [16, 46].

Other IBMs include off-lattice models [1–3, 47]. In off-lattice IBMs agents are not confined to a lattice, and can occupy any location in space in the IBM domain. Off-lattice IBMs are often more realistic than lattice-based IBMs, in that agents are able to move continuously in space, and do not have to ‘jump’ fixed distances. However, what gain off-lattice IBMs potentially make in biological realism is generally countered by a reduction in their mathematical tractability, compared to lattice-based IBMs [47–51]. Off-lattice IBMs are also typically more computationally expensive than lattice-based IBMs, which can be an important consideration if a large number of IBM repeats are required to generate results.

1.2.2 Continuum models

Throughout this thesis we derive continuum models that accurately approximate certain features of the IBMs we present. The reason we derive continuum models to approximate averaged IBM results is that continuum models are often useful for forming a more thorough understanding of a process, which is not always apparent (or feasible) from merely studying a large number of simulations of an IBM. Chapters 4, 5 and 6 of this thesis especially highlight how deriving a continuum model can conceptually simplify a complicated system, and provide an analytical understanding not readily apparent from examining the results of the corresponding IBM.

Importantly, all continuum models presented in this thesis are derived directly from considering the averaged behaviour of the IBM. This is generally done by formulating a probability master equation describing the evolution of the probability that a system is in a certain state at a given time. The derivation of continuum models from

IBMs typically utilise moment closure approximations, with the choice of moment closure dependent on the process being modelled. The assumptions made in deriving a particular continuum model will be made clear when necessary in the text.

1.3 Thesis outline

We now give a brief summary of the remaining chapters. In Chapter 2 we motivate the use of mathematical and computational modelling in developmental biology research with a study of how melanoblasts colonise the epidermis during murine development [10]. We show, through the use of an IBM parameterised by experimental data, that a simple mechanism consisting of undirected cell migration and proliferation, in tandem with domain growth, can account for the colonisation of the epidermis by melanoblasts in the embryonic mouse.

Cell-cell interactions, such as cell-cell adhesion or repulsion, are not included in our IBM of melanoblast colonisation of the DLE presented in Chapter 2. In the case of melanoblast colonisation of the DLE this is because complex cell-cell interactions are *not* thought to exist between melanoblasts. However, it is well understood that cell-cell interactions *do* play a significant role in some cell migratory processes, often during embryonic development [52–54]. Therefore, in Chapter 3 we conduct a theoretical study into how cell migration is affected by cell-cell interactions, and include the effects of both domain growth and a chemoattractant [12]. In doing so we compare results from IBMs and continuum models describing the spatial and temporal evolution of a cell population with adhesive and repellent cell-cell interactions, on both static (non-growing) and growing domains. From our studies we suggest that cell-cell interactions can be inferred from population density characteristics in the presence of a chemoattractant, and that these population density characteristics are conserved on both static and growing domains. We also study the expected displacement of

a tagged cell, and show that different types of cell-cell interactions give rise to cell trajectories with distinct characteristics.

In Chapter 3 we assume agent proliferation is negligible in our IBM and corresponding continuum models. However, cell proliferation is of obvious importance in developmental biology, indeed in Chapter 2 we present an example of how a *proliferative* cell population migrates along a growing domain. Agent proliferation in IBMs establishes spatial correlations between agents, and these spatial correlations can lead to some continuum approximations of IBMs being inaccurate. Therefore, the central topic of Chapters 4, 5 and 6 is deriving accurate continuum approximations of IBMs that include agent proliferation and domain growth.

In Chapter 4 we present methods to include the effects of domain growth on the evolution of spatial correlations between agents in a continuum approximation of a one-dimensional model of cell motility and proliferation [13]. We include the effect of spatial correlations by deriving a system of ordinary differential equations (ODEs) that describe the expected evolution of individual and pairwise density functions for agents on a growing domain. We then demonstrate how to simplify this system of ODEs by using an appropriate approximation. This simplification allows domain growth to be included in models describing the evolution of spatial correlations between agents in a tractable manner.

In order to describe cell migratory processes more realistically it is necessary to extend the methods presented in Chapter 4 to two dimensions. Therefore, in Chapters 5 and 6 we present methods to include the effects of domain growth on the evolution of spatial correlations between agents in a continuum approximation of a two-dimensional model of cell motility and proliferation [14]. In Chapter 5 we show that, depending

on the way in which domain growth is implemented, different steady state densities are predicted for otherwise identical agent populations. Furthermore, we demonstrate that the way in which domain growth is implemented can cause the initial size of the IBM domain to impact the evolution of the agent density. In Chapter 6 we extend the results presented in Chapter 5 to show that, depending on the way in which domain growth is implemented, different species dominate in two species simulations [15]. Continuum approximations of the IBM that ignore spatial correlations cannot capture any of the aforementioned behaviour, while those that account for spatial correlations can. The results in Chapters 5 and 6 show that the mechanism used to implement domain growth can determine the long term behaviour of agent populations, and in certain circumstances, establish patterning.

To use the modelling approaches developed in Chapters 3 to 6 to represent biological systems of interest, it is often necessary to accurately parameterise them. Therefore, in Chapter 7 we develop methods to identify parameters for cell motility and adhesion in a simple wound-healing experiment [11]. Building on the results of Chapter 3 we implement different cell-cell interactions in an IBM simulating a wound-healing experiment, and address issues of parameter identifiability and experimental design. We show that, given an *ideal* experimental design, cell motility and adhesion parameters can be successfully retrieved from synthetic data using approximate Bayesian computation (ABC). Using an IBM we then improve the design of a practically realisable experiment, and show that identification of model parameters is more accurate if the domain upon which the experiment is performed is expanded, as opposed to increasing the number of repeats of the experiment.

We conclude the thesis in Chapter 8 with a short discussion of the main results, and consider some possible extensions to the work presented here.

Chapter 2

Modelling neural crest stem cell migration on a growing domain

Overview

In this chapter we present an IBM calibrated by experimental measurements to investigate colonisation of the dorsal lateral epithelium (DLE) by melanoblasts in the embryonic mouse. Our observations indicate that melanoblast colonisation likely proceeds through a process of undirected migration, proliferation and tissue expansion. Our IBM also explains the formation of murine pigmentation patterns such as diffuse *LaacZ*¹ patterns, chimeric stripes, and belly spots via a common and, hitherto unrecognised, mechanism.

The majority of the work in this chapter appears in the paper: R. L. Mort^{*2}, R. J. H. Ross^{*}, K. J. Hainey, O. Harrison, M. A. Keighren, G. Landini, R. E. Baker, K. J. Painter, I. J. Jackson and C. A. Yates. “Reconciling diverse mammalian pigmentation patterns with a fundamental mathematical model”. *Nature Communications*, 7(10288), 2016. R. J. H. Ross performed all the computational modelling. R. L. Mort, O. J. Harrison, K. J. Hainey and M. A. Keighren performed the wet lab work,

¹Throughout this chapter we implement the standard convention whereby a gene is italicised, and its corresponding protein is not. For example, *LaacZ* is the gene, and LaacZ is the protein.

^{2*} These authors contributed equally to this work.

and are the people referred to as experimental collaborators in this chapter.

2.1 The development of neural crest stem cells

Neural crest stem cells (NCSCs) are highly motile and invasive multipotent cells that contribute to many tissues in the developing embryo. They originate in a region between the neural tube and the ectoderm after closure of the neural tube, and migrate along well-defined pathways in the developing embryo to populate a variety of sites [55]. The position along the neural tube at which NCSCs originate determines the cell lineage they become. For instance, depending on the position NCSCs originate they can become cranial, cardiac or trunk NCSCs, in addition to other types of NCSC [1, 55].

Given the fundamental role of NCSCs in human embryonic development, the failure of NCSCs to correctly develop is associated with many human birth defects. Two such defects are neurofibromatosis type 1 and piebaldism [56]. One in 3000 babies born in the UK is diagnosed with neurofibromatosis type 1, and it has been shown that these children have an elevated risk of developing certain cancers [57, 58]. Therefore, the study of NCSCs is not only important to understand correct human development, but also has clinical applications regarding understanding the factors that contribute to human disease.

In this chapter we study the development of *trunk* NCSCs. Trunk NCSCs are the progenitors of cells that become the dorsal root ganglion and adrenal medulla, and another population that become melanoblasts, which go on to become melanocytes [55]³. Melanoblasts originate in a region between the neural tube and the ectoderm

³In this chapter we will technically be modelling the migration of melanoblasts. Trunk NCSCs are referred to as melanoblasts once they begin colonisation of the DLE.

in the trunk region of the back, before colonising the DLE. In Fig. 2.1 the region of the embryonic mouse referred to as the DLE is displayed. During the colonisation of the DLE by melanoblasts the DLE is growing. Therefore, melanoblast colonisation of the DLE constitutes a migratory cell population on a growing domain.

Many genetic mutations that disrupt the colonisation of the DLE by melanoblasts have been discovered [59, 60]. This makes melanoblast colonisation of the DLE a powerful experimental system to work with, as many genetic manipulations can be performed. Therefore, we use the colonisation of the DLE by melanoblasts as a model system to try and further understand both the correct, and incorrect, development of NCSCs.

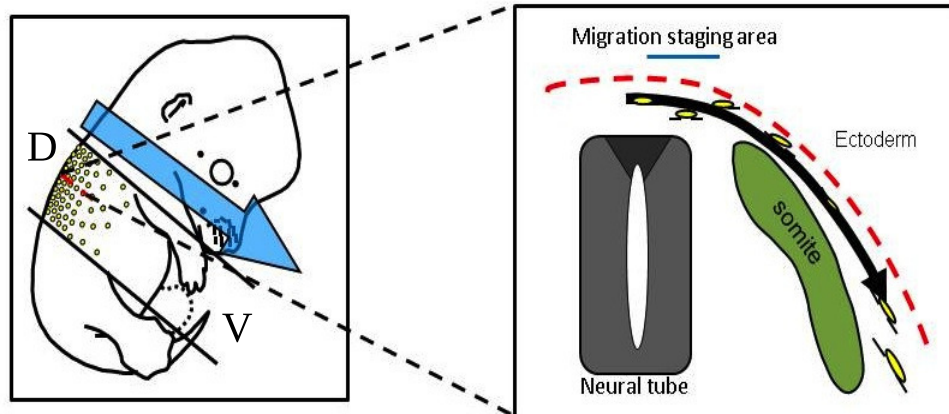


Figure 2.1: The migration path of melanoblasts in the developing mouse embryo. D indicates the dorsal side of the mouse embryo, V indicates the ventral of the mouse embryo, and the blue arrow indicates the direction of **dorsoventral** migration by the melanoblasts. The surface of the abdominal region between the two straight black lines in the left panel is the DLE. The right panel shows this same region from the perspective of a cross section through the axial axis, and in this right panel the DLE is the region between the somite and ectoderm.

We now formalise the research hypothesis we address in this chapter, and explain why computational modelling was required to address this research hypothesis.

The research hypothesis

We hypothesise that undirected movement and proliferation of melanoblasts, and displacement of melanoblasts by growth of the DLE is all that is required to colonise the DLE by E15.5 (15.5 days post-conception) in a mouse embryo. Implicit in our hypothesis is that colonisation of the DLE by melanoblasts does not require any extracellular co-ordination such as complex cell-cell interactions, or directional cues, such as that induced by a chemoattractant. This is different to other populations of migrating NCSCs, which have been shown to depend on cell-cell interactions and chemoattractants for successful migration [1–3, 61].

Why use computational modelling for this project?

Current experimental techniques only allow our collaborators to record live-imaging of melanoblasts in the DLE from E14.5 onwards. By this time melanoblasts have largely colonised the DLE. Therefore, the role of our IBM was to describe melanoblast colonisation of the DLE from E10.5 onwards (E10.5 corresponds to the beginning of melanoblast colonisation of the DLE), and verify the hypothesis that undirected movement and proliferation of melanoblasts, in tandem with displacement of melanoblasts by growth of the DLE, is all that is required to colonise the DLE by E15.5.

2.1.1 Phenotypes associated with melanoblast colonisation of the dorsal lateral epithelium

We now describe the phenotypic properties associated with melanoblast colonisation of the DLE that emerge from the IBM.

Pigmentation patterns

As melanoblasts play a role in the development of hair follicles in the DLE, many murine pigmentation phenotypes depend on melanoblast behaviour. Stripes of colour

extending dorsoventrally are a common feature of mouse chimeras [62]. A chimera is a single organism composed of cells from different zygotes. The genome from each zygote may determine a different coat colour, and so a common feature of chimeric organisms is that they have two coat colours. An example of a chimeric mouse with two coat colours can be seen in Fig. 2.2.

Chimeric stripes were originally taken as evidence of the *directed* dorsoventral migration of melanoblasts as they colonise the DLE, and led researchers to conclude that melanoblast colonisation of the DLE is *organised*, and that melanoblasts *do not* mix significantly when colonising the DLE. That is, each stripe consists of melanoblasts that are all progeny of the same progenitor melanoblast. This mechanism of melanoblast colonisation of the DLE is displayed in Fig. 2.3.



Figure 2.2: Chimeric mouse. In this mouse the striped pattern is caused by having melanoblasts from different genetic lineages localising to hair follicles. The dorsoventral stripes on this mouse were originally taken as evidence of directed dorsoventral migration of melanoblasts while colonising the DLE.

More recent research into melanoblast migration revealed a *diffuse* distribution of melanoblasts with high levels of axial mixing post-colonisation of the DLE. By diffuse it is meant that a melanoblast and its progeny are distributed sparsely across

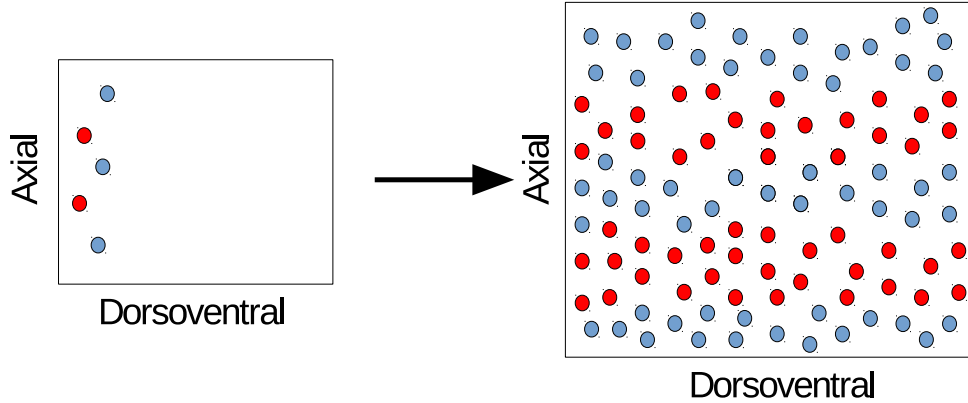


Figure 2.3: The original hypothesis for the mechanism of stripe formation in chimeric mice. The initial configuration of melanoblasts is preserved as the DLE grows and the melanoblasts proliferate and migrate dorsoventrally in an organised manner. Red circles indicate melanoblasts of one genotype and blue circles indicate melanoblasts of another genotype.

the DLE, and are not confined to a narrow region, as may be expected if melanoblast migration is directed. Wilkie et al. [6] presented tracing of single melanoblast clones using a marking method we refer to as *LaacZ*. These *LaacZ* marker gene patterns are caused by ‘switching-on’ the *LaacZ* marker gene in a melanoblast at a random time during embryonic development. All progeny of this melanoblast also express the *LaacZ* marker gene, leading to the collection of marked melanoblasts in the DLE that have all descended from the same melanoblast. An example of this diffuse distribution of melanoblasts, all of which share a common progenitor cell, can be seen in Fig. 2.4.

The existence of both chimeric stripes and *LaacZ* patterns has been a contentious issue in the field of melanoblast migration [6]. The stripes associated with chimeric patterns have led some researchers to argue that melanoblast migration across the DLE is organised [62], such that melanoblasts migrate and proliferate dorsoventrally while colonising the DLE, with little axial mixing (see Fig. 2.3). Meanwhile, the dif-

fuse nature of *LaacZ* patterns suggest that melanoblast colonisation of the DLE is less structured, with melanoblasts from the same progenitor cell being spread out across the DLE [6]. Here we provide evidence to suggest that the contention surrounding this issue is due to an erroneous interpretation of the original chimera experiments [62].



Figure 2.4: *LaacZ* mouse embryo at E15.5. In this mouse embryo the light blue dots indicate melanoblasts with *LaacZ* marker gene expression. All of these melanoblasts share a common progenitor cell.

Piebaldism

In mice piebaldism results in a white spots on the back and belly [59], an example of the belly-spotting phenotype is shown in Fig. 2.5. This belly-spotting phenotype is caused by the failure of melanoblasts to localise with hair follicles in the DLE. Kit, a cytokine receptor expressed on the surface of stem cells, is known to be important in melanoblast migration. It is thought Kit accelerates the localisation of melanoblasts to hair follicles in the DLE, and mice heterogeous for a mutation in *Kit* have white back and belly spots [55], a piebaldism phenotype. The mechanism by which mutation of *Kit* causes belly spots is not well characterised. Through a combination of experiments and modelling we will show in this chapter that it is likely that impaired

proliferation in melanoblasts is the cause of white belly spots in mice.



Figure 2.5: Belly spot phenotype in a mouse, also referred to as piebaldism. The white belly spot indicates where melanoblasts have failed to colonise the DLE during embryonic development.

2.2 Individual-based model

In this section we describe the IBM we use to model melanoblast colonisation of the DLE. Following this we explain how we parameterise this IBM using experimental data, and the phenotypic properties associated with melanoblast colonisation of the DLE that we expect to emerge in the IBM simulation results.

We use an IBM with volume exclusion on a two-dimensional square lattice of size $N_x(t)$ by $N_y(t)$ to model the DLE. Although the DLE is a curved region of a mouse embryo, for simplicity we choose to represent it by a rectangle, as if flattened out. $N_x(t)$ denotes the length of the domain (number of sites) dorsoventrally at time t , and $N_y(t)$ denotes the length of the domain (number of sites) axially at time t . The lattice spacing is Δ , with Δ^2 understood to represent the area of a melanoblast⁴, and time, t , evolves continuously according to the Gillespie algorithm. Each agent

⁴By the area of a melanoblast we mean the region of the two-dimensional domain the melanoblast occupies in the IBM.

(melanoblast) is assigned to a lattice site, from which it can move or place progeny into adjacent sites. Attempted agent movement or proliferation events occur with rate P_m or P_p per unit time, respectively. That is, $P_m\delta t$ is the probability of a given agent attempting to move in an infinitesimally small time interval δt . If an agent attempts to move or proliferate into a site that is already occupied, the event is aborted.

Throughout the simulation the IBM domain grows in both the dorsoventral and the axial directions. To model domain growth we employ a stochastic ‘pushing’ growth mechanism, as described by Binder and Landman [25]. This domain growth mechanism is illustrated in Fig. 2.6. The insertion of new lattice sites into the domain occurs with rates P_{ga} and P_{gd} per unit time, for growth in the axial and dorsoventral directions, respectively. When a ‘growth event’ occurs in the dorsoventral direction, for each row of the lattice one new site is added in a column, which is selected uniformly at random. To accommodate the new sites, in each row, the sites to the right of the added site are shifted a distance Δ rightwards carrying their contents with them (that is, agents move with their sites). Likewise, for axial growth one new site is added to each column in a row, which is selected uniformly at random, and the appropriate sites are shifted upwards. Growth is linear in both the dorsoventral and axial directions in the IBM, as evidenced by experimental data [10].

All domain boundaries in our IBM of melanoblast colonisation of the DLE have no-flux boundary conditions. We have very little information about melanoblast behaviour beyond the boundaries of our domain. Therefore, the simplest assumption to make is no-flux boundary conditions.

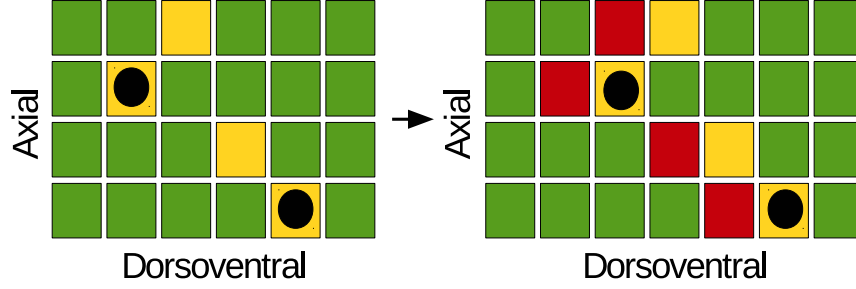


Figure 2.6: The domain growth mechanism for the IBM. An example of a domain growth event in which domain growth is in the dorsoventral direction (horizontal direction) for a two-dimensional lattice. In each row the yellow site has been chosen to undergo a growth event. Following this the yellow site is moved to the right with its contents, for instance an agent (represented by a black circle). The red sites are the new lattice sites and are always initially empty.

2.2.1 Inferring parameters for the individual-based model from experimental data

We now describe how we infer parameters from experimental data for our IBM. Following this we will describe the phenotypic properties associated with melanoblast colonisation of the DLE that emerge from the IBM. The parameters we implement in the IBM are listed in Table 2.1.

Table 2.1: IBM parameters.

Simulation parameter	Parameter value
Axial growth (P_{ga})	$0.0056 \mu\text{m min}^{-1}$ (one site added to each column per growth event)
Dorsoventral growth (P_{gd})	$0.0246 \mu\text{m min}^{-1}$ (one site added to each row per growth event)
Motility rate (P_m)	0.0412 min^{-1}
Proliferation rate (P_p)	0.00165 min^{-1}
Site length (Δ)	$38 \mu\text{m}$
Initial axial length	$\approx 1634 \mu\text{m}$ (43 sites in the IBM)
Initial dorsoventral length	$\approx 1178 \mu\text{m}$ (31 sites in the IBM)

The size of a melanoblast in the individual-based model

In the IBM one agent is defined to occupy a $38\text{ }\mu\text{m}$ by $38\text{ }\mu\text{m}$ square (i.e. $\Delta = 38\text{ }\mu\text{m}$), which is the area of a lattice site in the simulation. This is a biologically realistic estimate for the size of a melanoblast, as confirmed by experiments [10]. This means for complete colonisation of the domain, that is, every site in the domain is occupied by an agent, our simulation has 693 cells per mm^2 . This is the complete colonisation density because the approximate number of melanoblasts that would fit into a mm^2 is $(1000/38)^2 \approx 693$. This density value for complete colonisation of the IBM domain also agrees with experimental data.

The initial distribution of agents in the individual-based model

In agreement with experimental data our IBM is initialised in the following manner: at E10.5 ($t = 0$) we initialise 21 agents on the domain such that, on average, a third of the agents are between sites 12 and 32 of the axial axis, and the remaining two thirds are evenly distributed between sites 1 and 11, and 33 and 43, corresponding to a slight under-representation of agents in the middle of the axial axis. These agents are distributed so that 95% of the initial agents are between sites 8 and 17 of the dorsoventral axis. All agents are distributed between sites 8 and 19 of the dorsoventral axis. An example initial configuration of the IBM is shown in Fig. 2.7.

The melanoblast motility parameter in the individual-based model

Our collaborators examined the mean square displacement (MSD) of melanoblasts in the DLE at various densities. They found the MSD of melanoblasts to be linearly proportional to time, indicating that Fickian diffusion is an appropriate description of melanoblast migration. Our collaborators also examined the relationship between melanoblast density and the diffusion coefficient, D , and found a strong negative correlation, i.e. melanoblasts diffuse faster when they are less densely packed, as

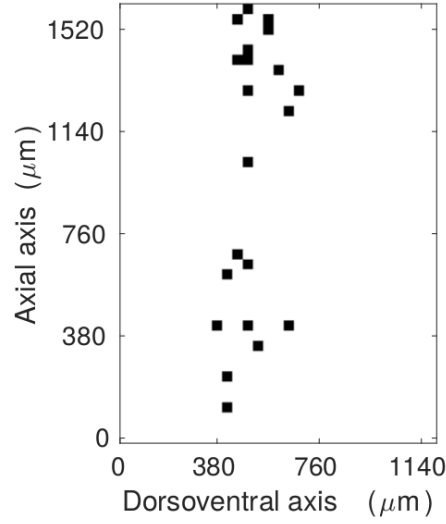


Figure 2.7: An example of the initial configuration of the agents in the IBM. 21 agents are initialised in accordance with an experimentally ascertained distribution as described in the text.

illustrated in Fig. 2.8.

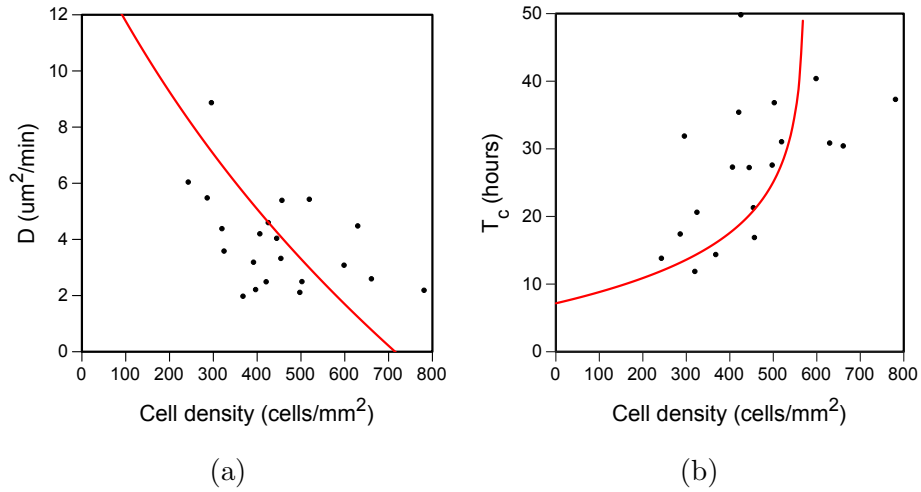


Figure 2.8: (a) The relationship between the diffusion coefficient, D , and the agent density is recapitulated as an emergent property of the IBM (red line) in agreement with experimental data (black dots). (b) Similarly, the relationship between T_c (average minimum melanoblast cell cycle in the IBM) and agent density described is also recapitulated as an emergent property of the IBM (red line) in agreement with experimental data (black dots).

To determine the same relationship between melanoblast density and the diffusion

coefficient in our IBM we proceeded as follows: we tracked agents moving on a 646 μm by 646 μm domain (corresponding to 17 by 17 sites) with periodic boundary conditions. This domain size corresponds, approximately, to the field of view of the microscope used to collect the experimental data on melanoblast movement. At $t = 0$ in the simulation, a number of agents (from 1 to 289, representing all possible non-zero agent densities) are initialised with positions chosen uniformly at random throughout the domain. These agents are allowed to move (but not to proliferate, so as to keep the agent density constant) for a simulation duration equivalent to 400 minutes of real time. This process was repeated 100 times for each agent density to guarantee enough data for an accurate estimation of the MSD of the agent population. In each simulation the resulting agent tracks were used to characterise the MSD⁵. To determine the agent movement rate, P_m , we compared the relationship between the agent density and the *effective* diffusion coefficient for the experimental data to those for the IBM for a range of different values of P_m . By effective diffusion coefficient we mean a diffusion coefficient that produces the correct relationship between the diffusion coefficient and agent density in our IBM, as in agreement with experimental data. We chose the value of P_m that gave the best fit to the experimental data (smallest least squares error, l^2 norm), as can be seen in Fig. 2.8. This value P_m is given in Table 2.1. To convert between P_m and the effective diffusion coefficient, D , we use the following formula

$$D = \frac{P_m \Delta^2}{4}. \quad (2.1)$$

Equation (2.1) demonstrates the relationship between D , P_m and Δ in the IBM [63].

⁵For agents that crossed a boundary the portion of their tracks before they crossed a boundary was included in the analysis, however, after this their tracks were ignored. This is because these agents would be lost in the experimental system. These agents are not lost to the simulation however, as the domain boundaries are periodic, and so a constant density was maintained.

The melanoblast proliferation parameter in the individual-based model

The parameter for agent proliferation, P_p , was chosen such that the average minimum melanoblast cell cycle in the IBM, T_c , is 7 hours. This produced a similar relationship between cell density and proliferation in the IBM that is exhibited in the experimental data, as can be seen in Fig. 2.8. To convert between P_p and T_c we use the following formula

$$T_c = \frac{\log 2}{(60)(P_p)}. \quad (2.2)$$

Equation (2.2) is derived in the following manner: if we assume the melanoblast population grows exponentially, the doubling time of a melanoblast population is given by

$$2 = e^{P_p T_c}. \quad (2.3)$$

If we solve Eq. (2.3) for T_c we obtain Eq. (2.2) (the dimension of P_p is min^{-1}).

2.3 Results

The colonisation of the domain by melanoblasts

To better quantify the migration of melanoblasts during colonisation of the DLE our collaborators performed *ex vivo* live imaging of mouse embryonic skin at E14.5. A time-lapse sequence of melanoblasts in the mouse embryonic skin at E14.5 can be seen in Fig. 2.9. Experiments showed that melanoblasts are polarized during migration, and move along their longest axis (often referred to as the Feret's diameter). The Feret's angle is therefore assumed to be a good proxy for the direction of a cell's migration. Our collaborators demonstrated that the distribution of the Feret's angles in the melanoblast population in the DLE is uniform, consistent with completely

random melanoblast orientation and undirected migration [10].

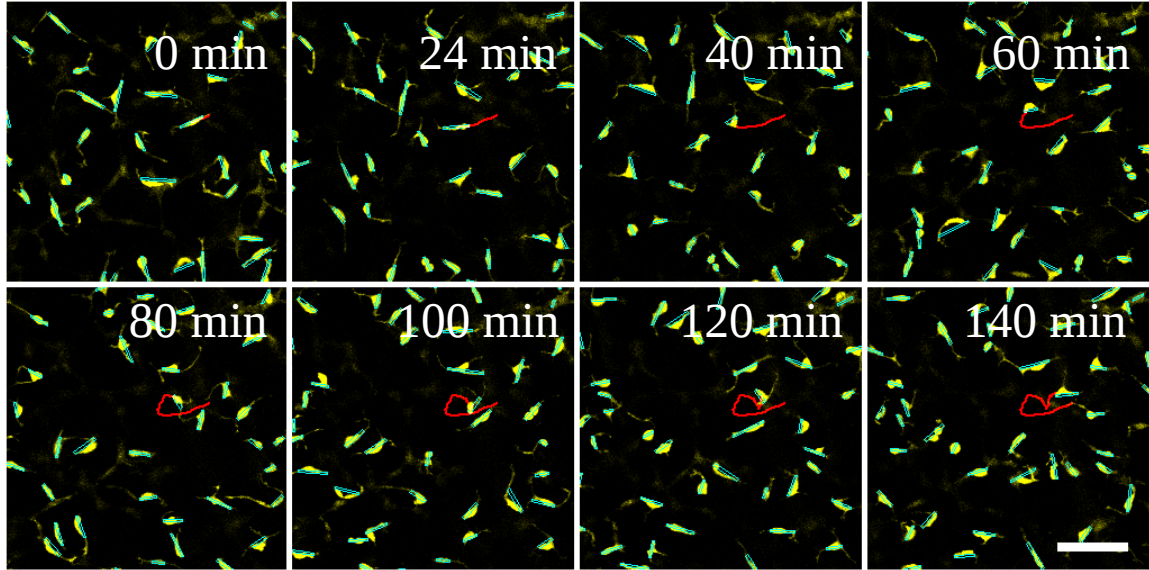


Figure 2.9: A time-lapse sequence of melanoblasts migrating in ex vivo culture of E14.5 skin. The Feret's diameter of each cell body is indicated in cyan. The path of a single migrating cell is indicated in red. Melanoblasts migrate along their Feret's diameter - the longest distance between any two points along a given boundary. Scale bar is 50 μm .

To supplement the Feret's angle analysis we examined the melanoblast spatial distribution at E14.5 in the DLE with the pair-correlation function (PCF) [44, 45]. The PCF is a summary statistic that provides a measure of clustering between cells in experimental images. To calculate the PCF from the experimental images we proceeded as follows: experimental images provided us with coordinates of each individual melanoblast's centre of mass on a two-dimensional domain of 512×512 pixels. Each pixel is 1.26 μm , and so the microscope field of view was $645.12 \mu\text{m} \times 645.12 \mu\text{m}$. We mapped each melanoblast centre of mass to a 512×512 lattice, whereby if an agent occupies a lattice site this indicates the centre of mass of a melanoblast in the experimental image. To map the position of cells in the experimental images, where melanoblast position is a continuous variable, (x, y) , to a discrete lattice site,

(i, j) , we use the following formulae

$$i = \lceil x \rceil, \quad j = \lceil y \rceil, \quad (2.4)$$

where $\lceil \cdot \rceil$ denotes the ceiling function. To avoid possible boundary effects at the edge of the microscope field we extracted a central portion of the lattice for the PCF analysis. This central portion of the lattice was all lattice sites indexed by both $129 \leq i \leq 384$ and $129 \leq j \leq 384$. This central portion of the lattice is therefore 256×256 lattice sites, and we denote its horizontal length by N_x^p and its vertical length by N_y^p .

To compute the PCF we require $p_x(m)$, which we define as the number of *nonperiodic* counts of occupied lattice pairs for each horizontal pair distance $m = 1, \dots, N_x^p - 1$. We use j^k to denote the *column* position of agent k , and j^l to denote the *column* position of agent l . This means $p_x(m)$ is given by

$$p_x(m) = \sum_{k=1}^A \sum_{l=k+1}^A \mathbb{1}\{|j^k - j^l| = m\}, \quad \forall m = 1, \dots, N_x^p - 1, \quad (2.5)$$

where A is the number of agents and $\mathbb{1}$ is the indicator function such that it is equal to one if $|j^k - j^l| = m$, and is zero otherwise. We also require $p_y(m)$, which we define as the number of *nonperiodic* counts of occupied lattice pairs for each vertical pair distance $m = 1, \dots, N_y^p - 1$. We use i^k to denote the *row* position of agent k , and i^l to denote the *row* position of agent l . This means $p_y(m)$ is given by

$$p_y(m) = \sum_{k=1}^A \sum_{l=k+1}^A \mathbb{1}\{|i^k - i^l| = m\}, \quad \forall m = 1, \dots, N_y^p - 1. \quad (2.6)$$

Binder and Simpson [45] demonstrated that is necessary to normalise Eqs. (2.5) and (2.6) to account for volume exclusion (and the nonperiodic pair count). The

normalisation term for Eqs. (2.5) and (2.6) are, respectively,

$$\hat{p}_x(m) = N_y^2(N_x - m)\rho\hat{\rho}, \quad \forall m = 1, \dots, N_x - 1, \quad (2.7)$$

and

$$\hat{p}_y(m) = N_x^2(N_y - m)\rho\hat{\rho}, \quad \forall m = 1, \dots, N_y - 1, \quad (2.8)$$

where $\rho = A/(N_x^p N_y^p)$, and $\hat{\rho} = (A - 1)/(N_x^p N_y^p - 1)$. Equations (2.7) and (2.8) describe the expected number of pairs of occupied lattice sites, for each nonperiodic pair distance, m , in an agent population distributed uniformly at random on the IBM domain. The horizontal component of the PCF, $q_x(m)$, is given by

$$q_x(m) = \frac{p_x(m)}{\hat{p}_x(m)}, \quad (2.9)$$

and the vertical component of the PCF, $q_y(m)$, is given by

$$q_y(m) = \frac{p_y(m)}{\hat{p}_y(m)}, \quad (2.10)$$

where $q_x(m)$, the horizontal component of the PCF, is a measure of how far $p_x(m)$ departs from describing the expected number of occupied lattice pairs for each horizontal distance of an agent population spatially distributed uniformly at random on the IBM domain, and $q_y(m)$ is an analogous measure for each vertical distance.

Finally, for a square domain we define the average PCF as

$$F(m) = \frac{1}{2S} \sum_s^S q_x^s(m) + q_y^s(m), \quad (2.11)$$

where $q_x^s(m)$ and $q_y^s(m)$ are the pair-correlation functions for the horizontal and ver-

tical directions, respectively, for the s^{th} sample of S total experimental images. The total number of experimental images $S = 440$.

The PCF analysis of the melanoblast spatial distribution at E14.5 showed a volume exclusion effect below 28 μm as can be seen in Fig. 2.10, consistent with volume exclusion at the scale of an individual melanoblast⁶. Above this distance the spacing of melanoblasts conformed to complete spatial randomness (CSR). By CSR it is meant the spacing of melanoblasts conforms to expected number of pairs of occupied lattice sites, for each pair distance m , in an agent population distributed uniformly at random on the IBM domain. It has been shown that cell-cell interactions, for instance, cell-cell adhesion or repulsion, can cause patterning between cells at distances greater than the cell volume [44]. The existence of CSR at distances greater than the cell volume suggests that no cell-cell interactions of this kind are present between melanoblasts. We therefore conclude that the melanoblast distribution at E14.5 is spatially random (above 28 μm), and cell-cell interactions between melanoblasts does not contribute to their colonisation of the DLE.

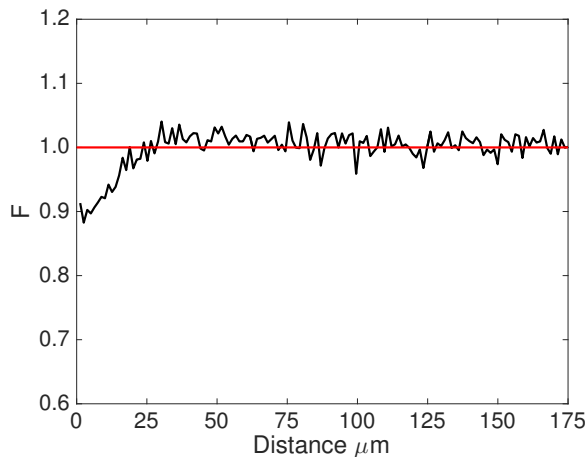


Figure 2.10: The PCF analysis on experimental data for melanoblast spatial organisation at E14.5. The red line indicates the value of F for CSR ($F = 1$).

⁶The reason for the difference between the volume exclusion demonstrated by the PCF analysis and the size of a melanoblast in the IBM is because melanoblasts are typically not symmetric.

To reiterate, our hypothesis is that undirected melanoblast movement and proliferation, in tandem with domain growth, is sufficient for melanoblast colonisation of the DLE. Using the experimentally established parameters listed in Table 2.1 our model predicts colonisation of the DLE. That is, the averaged agent density in the model at all stages of embryonic development closely fits the experimental data, as can be seen in Fig. 2.11.

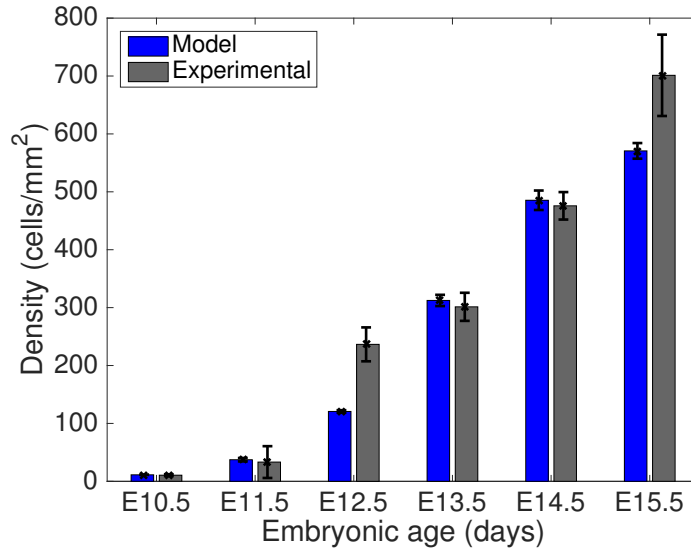


Figure 2.11: A comparison of the experimental melanoblast densities with results from the IBM for E10.5 to E15.5 (IBM results are an ensemble average from 100 repeats).

Chimeric and mosaic patterns formed by a common mechanism

To examine the emergence of pigmentation patterns in the IBM we simulated both mice with *LaacZ* marker genes and chimeric mice. To simulate the *LaacZ* marker gene we proceeded in the following manner: at a time point chosen uniformly at random during the simulation, we chose a single agent uniformly at random from amongst all the agents that populate the domain at that time to have the *LaacZ* marker gene. This single agent is ‘marked’, and all of this agent’s progeny inherit this mark. At the end of the simulation all marked agents are plotted to give a *LaacZ* pattern. To compare the *LaacZ* pattern generated by the IBM and those previously

described, the *LaacZ* patterns from both from the IBM and Wilkie et al. [6] are displayed in Fig. 2.12. It can be seen that the *LaacZ* IBM patterns appear qualitatively similar to the patterns previously described. A shortage of experimental data regarding *LaacZ* patterns prohibited a more quantitative comparison of the *LaacZ* patterns generated by the IBM and those observed experimentally. A quantitative description of *LaacZ* patterns is therefore deserving of further study, the possibility of which we discuss in this chapter's summary.

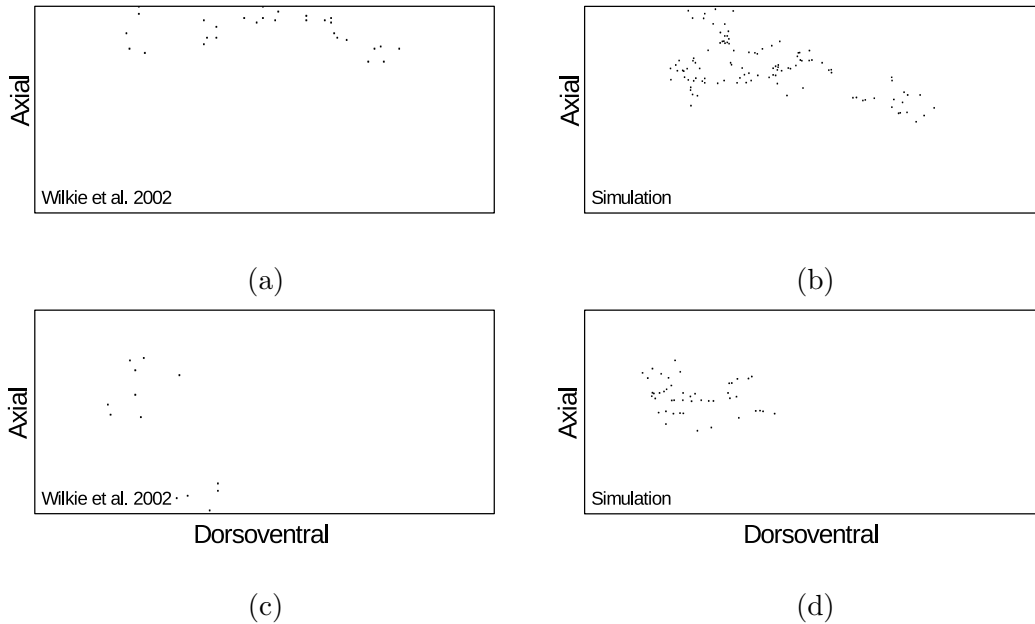


Figure 2.12: Two examples of diffusive *LaacZ* patterns described in Wilkie et al. [6] redrawn with permission from *Development* (replotted here in (a) and (c) - *LaacZ* expressing melanoblasts shown in black). In (b) and (d) patterns qualitatively similar to those in Wilkie et al. [6] generated by randomly labelling a single agent during a simulation and tracing all its progeny.

To simulate chimeric mice in our IBM we initialized our simulations with two differently labelled agent populations (representing two different genotypes) and tracked the positions of their progeny over time. When a simulation was complete, we assigned the value +1 (associated with light grey cells in Fig. 2.13 (a)) to one of the agent types and -1 (associated with black cells in Fig. 2.13 (a)) to the other (while empty lattice sites are assigned the value '0').

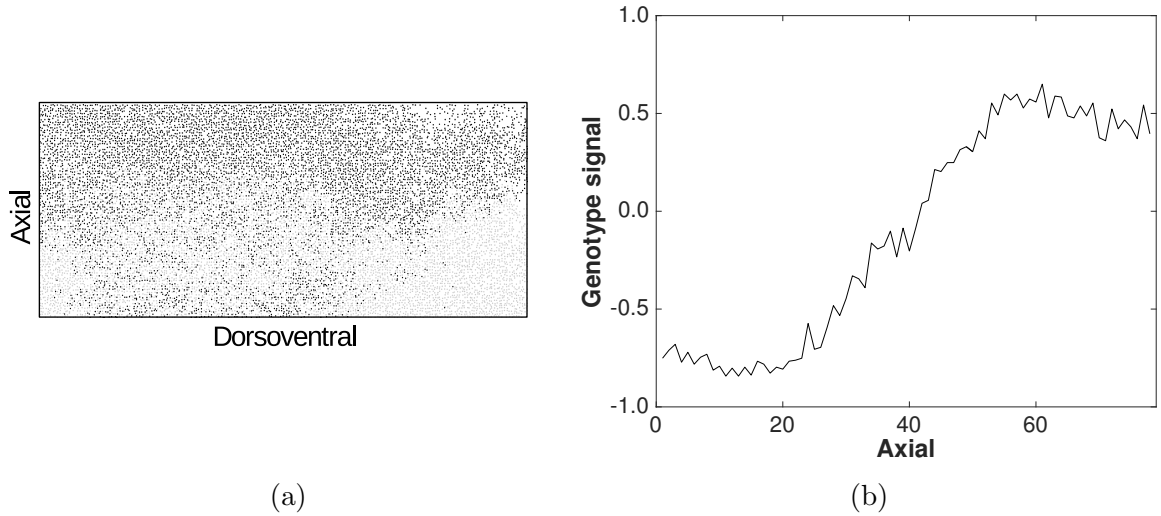


Figure 2.13: Identification of stripes in the IBM. (a) A stripe-like pattern from a single simulation of the IBM initialised with agents of two melanoblast genotypes (black and light grey) in equal proportions. (b) The axial profile of the genotype signal for the simulation depicted in (a).

To quantify the presence of stripe-like patterns in our IBM we averaged the values associated with the agent populations on each row for a given simulation. This provides a measure of the proportion of each agent genotype in each row, which we call the ‘genotype signal’. The genotype signal for Fig. 2.13 (a) can be seen in Fig. 2.13 (b). We repeat this process for each of the possible divisions of the twenty-one initial cells into two non-overlapping subsets, which we call a ‘genotype ratio’⁷. We then apply the discrete fast Fourier transform (DFFT) to our genotype signal (Fig. 2.13 (b)). We repeat this process 100 times and generate an average DFFT for each different initial genotype ratio. For simulations without stripe-like patterns no dominant frequency is identifiable. However, in the case where different agent genotype populations are not well mixed and have formed dorsoventral stripes, a single, low frequency is identifiable that relates directly to the periodicity of the stripes in the simulation. This means a dominant frequency (corresponding to the maximum value of the averaged DFFT) can be identified for each initial genotype ratio. We call this

⁷For instance, (1 light grey cell, 20 black cells), (2,19), (3,18), etc.

dominant frequency the ‘stripe intensity’. This method allows us to systematically identify the presence of stripes.

Following the above procedure we find that stripes are most apparent when we initialise simulations with only two genotypes. This corresponds to most experimental patterns in mice consisting of only two colours. As we increase the number of starting genotypes the stripe intensity decreases, as can be seen in Fig. 2.14.

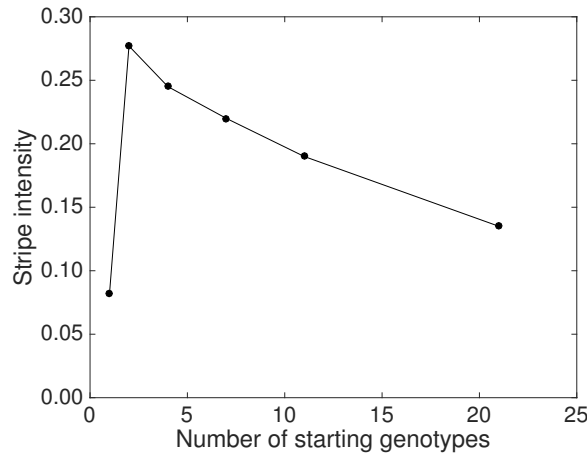


Figure 2.14: Stripe intensity *vs.* number of starting genotypes. When only two genotypes are present stripes are most apparent. As the number of genotypes is increased ‘stripe intensity’ is reduced (where stripe intensity is defined as the dominant frequency emerging from the mean DFFT of the intensity profile of 100 simulations of the IBM).

These results demonstrate that the IBM is able to produce both chimeric stripes and diffuse *LaacZ* patterns, and that this is through a common mechanism for melanoblast colonisation of the DLE. Given the described analysis of stripe formation in the IBM, we hypothesise that the stripes are formed by the coalescence of clusters of melanoblasts with different progenitor cells. This means that stripe formation does not proceed through the directed dorsoventral migration of descendent melanoblasts, but rather stripes form from a favourable arrangement of melanoblasts of the same genotype in the initial conditions that are elongated dorsoventrally by the growth in that direction. Consequently, increasing the number of differently labelled

melanoblast populations, akin to increasing the genotypes in the chimera, removes the stripe-like pattern in the chimeras. For example, compare the panels in Fig. 2.15, showing two-colour and multicolour depictions of a single simulation. In the two-colour chimera a stripe-like pattern is discernible, whereas in the twenty-one colour chimera it is not.

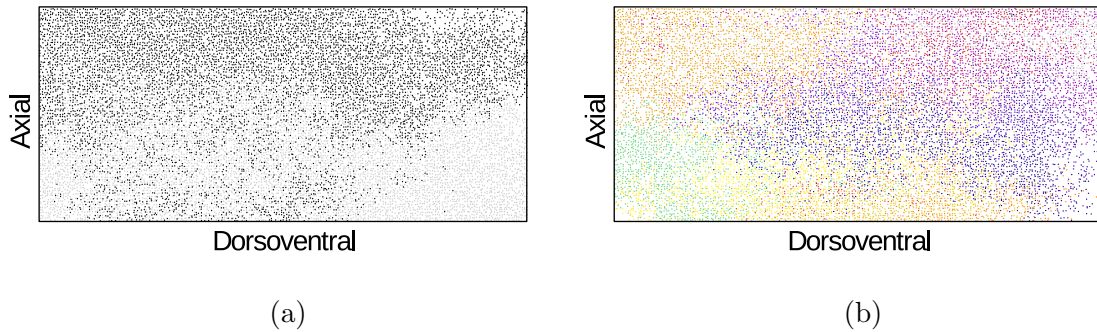


Figure 2.15: Example chimeric patterns from single simulations of the IBM initialised with either agents of two genotypes in equal proportions, or with twenty-one differently coloured genotypes. (a) and (b) are generated from a single simulation. In the case of chimeras with two genotypes (colours) melanoblasts of the same genotype, but from different initial progenitor cells, are more likely to create continuous bands of a single colour when compared with twenty-one genotype chimeras. This coalescing of melanoblasts of like genotype in the two-genotype chimera creates stripe-like patterns.

Belly spots arise from reduced proliferation

Belly spot formation has long been proposed as a product of either altered proliferation or migration in melanoblasts [55]. Experiments by our collaborators had confirmed that *Kit* mutant melanoblasts have similar motility to wild type melanoblasts. However, *Kit* mutants were shown experimentally to have a longer cell cycle, i.e. to proliferate more slowly. To compare the effect of agent motility and proliferation rate on domain colonisation we performed a parameter sweep comparing different values of D and T_c in our IBM. This revealed that domain colonisation in our IBM is markedly more sensitive to changes in T_c than in D , as can be seen in Fig. 2.16. We proceeded to test this conclusion for single realisations of the IBM. Accordingly,

small increases in cell cycle time in the IBM resulted in a failure of agents to fully populate the domain and produced a white belly spot that mimics the belly spot seen *in vivo*, as can be seen in Fig. 2.17. As the cell cycle time is increased in the IBM belly spots are increasingly observed, and at longer cell cycle times dorsal spots are also observed. This is consistent with existing mutant phenotypes known to affect melanoblast proliferation, such as *Rac1* [59] or *Magoh* [60].

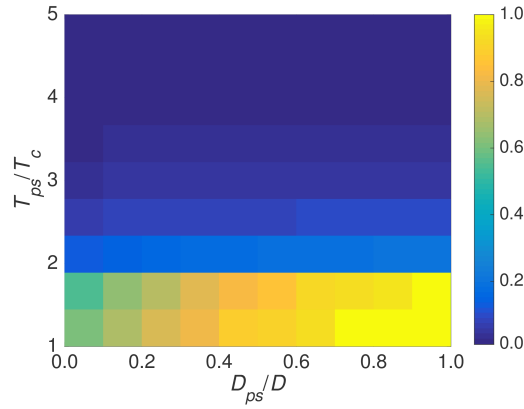


Figure 2.16: The bellyspotting phenotype in the IBM is more sensitive to a decrease in melanoblast proliferation rate than diffusion coefficient as indicated by the yellow region (blue: low probability of colonisation; yellow: high probability of colonisation). The parameter sweep compares colonisation in the IBM for different values of diffusion, D_{ps} , (normalised by the diffusion coefficient value used in our simulations, D) and cell cycle time, T_{ps} (normalised by the cell cycle time used in our simulations, T_c). 100 repeats of the IBM were performed for each combination.

2.4 Summary

Neural crest stem cell migration is a fundamental developmental process, the failure of which leads to human birth abnormalities. Therefore, understanding NCSC development has important implications for the treatment of a variety of human diseases. In this chapter we have used melanoblast colonisation of the DLE as a model with which to study the the development of NCSCs. We have shown, using a combination of experiments and analysis, that migrating melanoblasts do not have

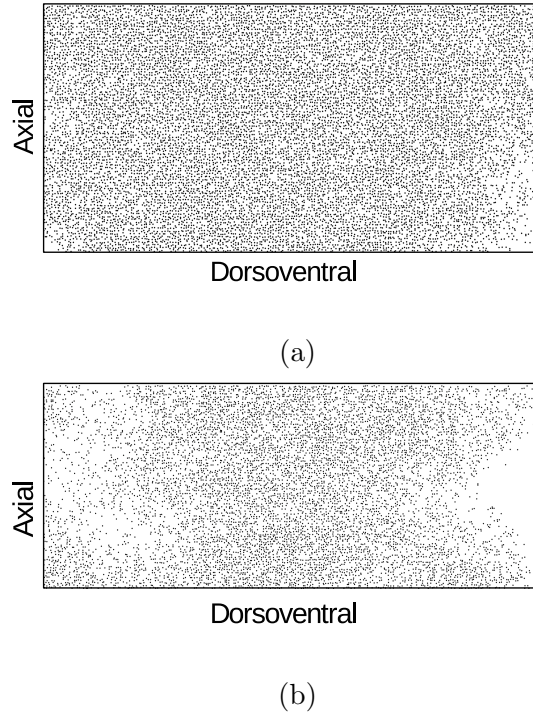


Figure 2.17: Model phenotypes for different values of T_c . (a) $T_c = 7$ hours, (b) $T_c = 10$ hours. An increase in T_c , from 7 to 10 hours, results in a ventral belly spot at E15.5 in our simulations similar to the pattern observed in *Kit* mutant mice. Each plot represents one side of the embryo extending from the dorsal most aspect on the left to the ventrum on the right. (a) Wild type ($T_c = 7$ hours), (b) *Kit* mutant ($T_c = 10$ hours).

preferred directionality, but rather diffuse and proliferate throughout the DLE in a density-dependent manner. Undirected melanoblast migration and proliferation, in combination with domain growth, is sufficient for the colonisation of the DLE by melanoblasts. Our IBM suggests that the formation of diffuse *LaacZ* patterns and chimeric stripes can arise through a *common* mechanism. Further experiments, using stochastic labelling methods such as brainbow and confetti [64], are required to definitively confirm our findings concerning the formation of pigmentation patterns in mice. Our IBM, in conjunction with experimental data, also suggests that impaired proliferation in melanoblasts is the cause of bellyspotting in mice rather than impaired migration. This suggests piebaldism is caused by a defect in melanoblast proliferation and not motility.

To conclude; melanoblast colonisation of the DLE, belly spot formation, chimeric stripes, and diffuse *LaacZ* patterns are all explained by a simple model incorporating random melanoblast migration with proliferation, in conjunction with domain growth, during the course of colonisation. Importantly, colonisation and the observed phenotypes are produced without the need for more complex cell-cell interactions or extracellular signals. This result has broader implications for cell behaviour in other NCSC lineages and their associated neurocristopathies, in that other NCSC populations may exhibit similar behaviours.

.....

In this chapter we have studied how a simple mechanism of undirected cell migration and proliferation, in tandem with domain growth, can account for melanoblast colonisation of the DLE. Cell-cell interactions, such as cell-cell adhesion or repulsion, are not present in our IBM of melanoblast colonisation of the DLE. In the case of melanoblast colonisation of the DLE this is because cell-cell interactions are not thought to exist between melanoblasts. However, it is well understood that cell-cell interactions *do* play a significant role in some cell migratory processes, often during embryonic development [52–54]. Therefore, in Chapter 3 we conduct a theoretical study as to how cell-cell interactions affect cell migratory processes on both static and growing domains. The aim of this study is to better understand the effect of cell-cell interactions on cell migratory processes.

Chapter 3

Modelling cell-cell interactions and cell pathlines on a growing domain

Overview

In this chapter we compare IBMs and continuum models describing the spatial and temporal evolution of a cell population with different forms of cell-cell interactions, such as cell-cell adhesion or repulsion. We study cell-cell interactions on static (non-growing) and growing domains, and also in the presence of a chemoattractant. We suggest that cell-cell interactions can be inferred from cell population density characteristics in the presence of a chemoattractant, and these population density characteristics for different cell-cell interactions are conserved on both static and growing domains¹. We also study the expected displacement of a tagged cell in the cell population and show, depending on the form of cell-cell interaction, the expected displacement of a tagged cell has different characteristics. Our results are relevant for researchers who study the role of cell-cell interactions in cell migratory processes, in so far as we suggest that different types of cell-cell interactions could be identified from cell population density and individual cell trajectory data.

The majority of the work in this chapter appears in the paper: R. J. H. Ross, C.

¹By population density characteristics we mean simple measurements that can be applied to the cell population to identify the cell-cell interactions present at the individual (cell) level.

A. Yates and R. E. Baker. “Inference of cell–cell interactions from population density characteristics and cell trajectories on static and growing domains”. *Mathematical Biosciences*, 264:108–118, 2015. R. J. H. Ross performed all the computational and mathematical modelling presented in this chapter.

3.1 Introduction

It is widely understood that cell-cell interactions play a significant role in cell migration [52–54, 65]. For example, multiple different cell-cell interaction mechanisms have been identified as promoting metastasis in breast cancer [65], and repulsive interactions mediated via ephrins on the surface of NCSCs have been shown to be important in organising NCSC migration [52]. Therefore, understanding interactions between motile cells is important for understanding cell migration in biological systems. However, although significant progress has been made in the identification and quantification of cell-cell interactions, work to understand the effect different cell-cell interactions have on cell motility at both individual and population levels is still ongoing [66–71].

Throughout this chapter we use mathematical and computational modelling to elucidate simple measurements that could be made during experiments in order to discern cell-cell interactions. To do so we use an IBM with volume exclusion to represent a motile cell population on a two-dimensional domain [43]. As outlined in the introduction we use an IBM as it is straightforward to incorporate a large number of cell-cell interactions, the focus of our study [33, 34, 38]. We also study these cell-cell interactions in the presence of domain growth and motility bias. By motility bias we mean the directed migration of a cell caused, for example, by a chemoattractant [72–74]. Chemoattractants have been shown to play a major role in cell migration. For instance, chemoattractants are known to be essential for the correct migration of

cranial, cardiac and enteric NCSCs [1, 55, 61, 73]. Meanwhile, growth is an essential component of biological systems, especially in the context of development, when large scale tissue formation is taking place [75]. This means that the inclusion of domain growth into models is necessary to understand cell migration in many contexts.

Importantly, we derive continuum approximations of our IBM for the spatial and temporal evolution of the cell population density, and the expected displacement of tagged cells on both static and growing domains [31, 33, 34, 38]. These continuum approximations in many cases are extensions of previously derived continuum models [33, 34, 38]. As outlined in Chapter 1, the derivation of continuum models is important because IBMs are often computationally expensive and analytically intractable.

By applying simple measurements to our simulation results we show that cell-cell interactions can be inferred from population density characteristics in the presence of motility bias, and these population density characteristics for different cell-cell interactions are conserved on both static and growing domains. Different cell-cell interactions can also be inferred from the characteristics of the trajectory of a tagged cell. This means the results presented here are relevant to experimentalists in the field of cell biology, as the study of population density and cell trajectory characteristics could be used to distinguish between different cell-cell interactions.

The outline of this chapter is as follows: in Section 3.2 we introduce our IBM of migratory cells on a growing domain. We also postulate mathematical representations of different types of cell-cell interactions, and derive corresponding continuum approximations of our IBM. In Section 3.3 we compare the accuracy of these continuum approximations with their IBM counterparts. We do this for a range of cell-cell interactions, for both static and growing domains, and show that different cell-cell

interactions can be inferred from population density and cell trajectory characteristics using simple measurements. In Section 3.4 we summarise the results presented in this chapter.

3.2 Model formulation

In this section we first introduce our IBM and the cell-cell interactions that we employ throughout this chapter. We then derive continuum approximations of our IBM, and compare the IBM simulation results with our continuum approximation solutions.

3.2.1 Individual-based model

We implement an IBM on a two-dimensional square lattice with lattice spacing Δ [46] and size $N_x(t)$ by $N_y(t)$. The lattice spacing Δ is understood to represent the length of a cell (the area the cell occupies is Δ^2). For IBM simulations in which the domain does not grow, the number of sites in the x and y directions is constant, so that $N_x(t) = N_x$ and $N_y(t) = N_y$. All simulations are performed with periodic boundary conditions at $y = 0$ and $y = N_y(t)$, and no-flux boundary conditions at $x = 0$ and $x = N_x(t)$. Furthermore, in all simulations displayed here, domain growth only occurs along the x axis.

The lattice is populated by agents, which represent biological cells. Each agent is assigned to a lattice site, from which it can move into an adjacent site. If an agent attempts to move into a site that is already occupied, the movement event is aborted. In this chapter we neglect cell proliferation, death and differentiation as they add significant complications to our methodology and provide little extra insight into the scenarios we study here. However, in Chapters 4, 5 and 6 the impact of cell proliferation is studied in detail.

At each time step, τ , all agents on the lattice have the opportunity to move to an adjacent site with probability P_m . On a square lattice, with an agent positioned at site ($row = i$, $column = j$) and not on a boundary, adjacent sites would consist of sites $(i, j - 1)$, $(i - 1, j)$, $(i, j + 1)$ and $(i + 1, j)$. To relate site coordinates to continuous Cartesian coordinates, we use $x = i\Delta$, $y = j\Delta$. This means the length of the x axis is $L_x(t) = N_x(t)\Delta$, where $N_x(t)$ is the number of sites in the x direction. An analogous relationship exists for $L_y(t)$. For all simulations presented here $P_m = 1$. The random sequential update method (RSUM) we employ to evolve the system forward in time means that if there are M agents in the domain, M sequential random agent movements to an adjacent site are attempted at each time step with replacement [43]. This means each agent is chosen on average once per time step, however, in some time steps an agent may be selected more than once or not at all.

Key to both the IBM we implement and continuum models we derive in this chapter is the growth mechanism we employ. This growth mechanism is the same as the one employed in Chapter 2, except that in this chapter the domain only grows in the horizontal direction. We briefly describe the growth mechanism again for clarity, and describe some small differences in its implementation in this chapter. Our growth mechanism is as follows: at each time step one site is selected uniformly at random from each row of the domain to undergo growth. This means that domain growth is linear. It also means that all realisations of the IBM grow to the same length, and so in this respect our growth mechanism is deterministic. We choose linear domain growth as it has been shown to occur in many biological systems [10, 24, 26, 76]. However, this growth mechanism is equally applicable to other types of isotropic domain growth, such as exponential growth [31]. As in Chapter 2 if site (i, j) is selected for growth, the new site is inserted at (i, j) , and the ‘selected’ site is moved to $(i, j + 1)$, taking its contents (i.e. an agent or no agent) with it. This means new sites are

unoccupied initially, and that all sites to the right of a site that undergoes growth are ‘pushed’ one lattice spacing, Δ , in the positive x direction, along with their contents².

3.2.2 Transition probabilities that represent cell-cell interactions

As outlined in the introduction of this chapter it is known that cell-cell interactions play an important role in cell migration [52–54, 65]. Therefore, we implement a number of cell-cell interactions in both our IBM and continuum models. These four types of cell-cell interaction, which we term blind, adhesive, repulsive and myopic, are designed to imitate cell-cell interactions in biological systems [33, 34, 53, 65]. We also study the effect of motility bias, for example induced by a chemoattractant, as it is known to play a significant role in some cell migratory processes [61, 72, 74]. It is important to note that the way in which we model these cell-cell interactions is not unique. However, given the focus of our study to discriminate between different types of cell-cell interactions we have chosen models of cell-cell interactions that have previously been published [33, 34].

To define our cell-cell interactions we first introduce some notation [34]. For any site v on the square lattice we define the nearest neighbourhood set $\mathcal{N}\{v\}$. This set contains all lattice sites that share a boundary with v , so that $N = |\mathcal{N}\{v\}| = 4$ (away from the boundary) for a two-dimensional square lattice. This neighbourhood is typically referred to as a von Neumann neighbourhood, and these are the sites from/to which agent movements into and out of site v can be made.

Throughout this chapter we use standard conservation arguments to derive our continuum approximations of our IBMs [35, 39]. The occupancy of site v is denoted by C_v , with $C_v = 1$ for an occupied site and $C_v = 0$ for an unoccupied site. By aver-

²This growth mechanism is illustrated in Fig. 2.6.

aging over many identically prepared realisations of our IBM we obtain the expected occupancy of site v , denoted by $\langle C_v \rangle$. This quantity represents the probability of occupancy of site v . Continuum approximations of our IBMs will describe the evolution of these expected occupancies in the IBM. Importantly, we assume that adjacent sites are independent in the derivation of the continuum approximations in this chapter³. This approach is well-established in the literature [34, 38, 77, 78].

Blind agent interactions

In ‘blind’ agent interactions the agent does not consider the occupancy status of adjacent sites before attempting a move. In this sense the agent is ‘blind’ to its environment. The transition probability for a blind agent, in which an agent attempts to move to one of the N nearest neighbour sites with probability $P_m(1 + \rho_{v \rightarrow v'})/N$ is written as

$$T(v'|v)_B = \frac{P_m(1 + \rho_{v \rightarrow v'})}{N}(1 - C_{v'}), \quad (3.1)$$

such that

$$0 \leq \sum_{v' \in \mathcal{N}\{v\}} \frac{P_m(1 + \rho_{v \rightarrow v'})}{N} \leq 1. \quad (3.2)$$

In Eq. (3.1) $T(v'|v)_B$ is the probability that an agent will successfully move from site v to site $v' \in \mathcal{N}\{v\}$ during $[t, t + \tau)$, given it is at v at time t (the subscript B denotes blind) [38], and $\rho_{v \rightarrow v'}$ is the bias present in moving from site v to v' . For our purposes $\rho_{v \rightarrow v'}$ will be one of $\pm\rho_x$ or $\pm\rho_y$, with $\pm\rho_x$ the bias for movements in the positive/negative x directions and $\pm\rho_y$ for movements in the y direction. For example, a movement from site $(i, j) \rightarrow (i, j + 1)$ would have a bias of ρ_x . Conversely, a movement from site $(i, j) \rightarrow (i, j - 1)$ would have a bias of $-\rho_x$. We define $\rho_{v \rightarrow v'}$

³This assumption will be disregarded in Chapters 4, 5 and 6.

$\in [-1, 1]$ and, given $0 \leq P_m \leq 1$, Eq. (3.2) is a condition that must be satisfied so that the probability of an agent attempting to move if selected does not exceed one.

We now proceed to define more complicated agent interactions. To do so we adopt a transition probability notation similar to that used by Landman and Fernando [34], with some modifications. From our nearest neighbourhood set $\mathcal{N}\{v\}$ we define two further sets. $\mathcal{O}\{v\}$ is defined as the set of all nearest neighbour occupancy configurations in the set $\mathcal{N}\{v\} \setminus \{v'\}$. $\mathcal{O}_n\{v\}$ is then defined as a subset of $\mathcal{O}\{v\}$ containing exactly n occupied nearest neighbour sites from the possible $N - 1$ sites in the set $\mathcal{N}\{v\} \setminus \{v'\}$. We then use $\mathcal{O}_n^k\{v\}$ to index the possible choices of sites in $\mathcal{O}_n\{v\}$, where the superscript k denotes one of the $\binom{N-1}{n}$ possible choices of n distinct sites in $\mathcal{O}_n\{v\}$. Put simply, these represent all the possible combinations of occupied sites in $\mathcal{N}\{v\} \setminus \{v'\}$, and are the elements of set $\mathcal{O}_n\{v\}$. Additionally, we define $\mathcal{U}_n^k\{v\}$ to be $(\mathcal{N}\{v\} \setminus \{v'\}) \setminus \mathcal{O}_n^k\{v\}$. Therefore, $\mathcal{U}\{v\}$ can be understood as the set of all possible combinations for unoccupied sites in the nearest neighbourhood set of site v . These combinations are displayed in Fig. 3.1. With this notation we can now define more complicated agent interactions for our IBM.

Adhesive/repulsive agent interactions

In adhesive and repulsive agent interactions, the number of sites that are occupied in the nearest neighbourhood of an agent influence the probability of that agent completing a move. In this sense the agent is not ‘blind’, as it takes its nearest neighbourhood into account before attempting a move. For adhesive interactions the success of a move decreases if the agent is in contact with other agents that it can adhere to. Conversely, for repulsive interactions the success of a move increases if an agent is in contact with other agents. In adhesive and repulsive agent interactions an agent attempts to move to one of its N nearest neighbours with probability

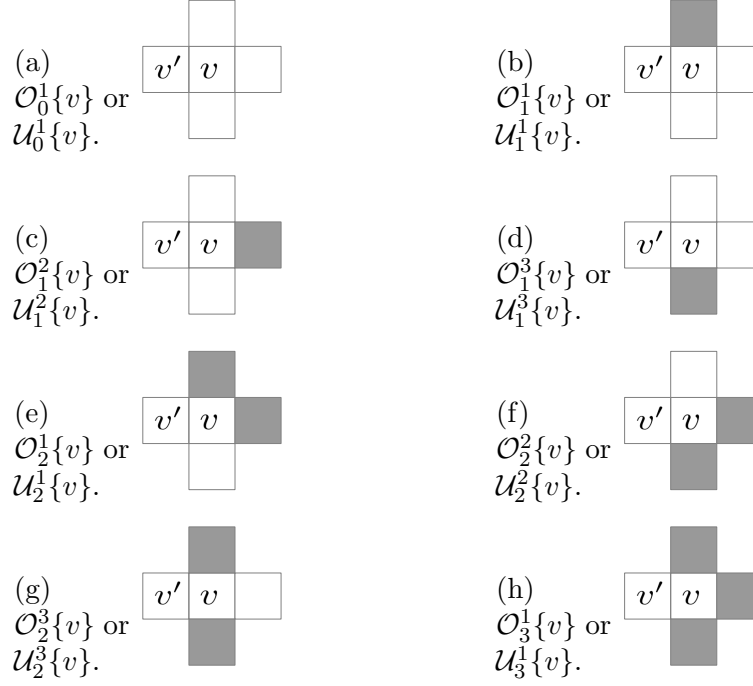


Figure 3.1: The nearest neighbourhood occupancy sets. The shaded sites indicate a site that is occupied by a cell.

$P_m(1 + \rho_{v \rightarrow v'})(1 + \alpha)^n/N$, where $n \leq N$ signifies the number of nearest neighbours of v that are occupied, $-1 \leq \alpha < 0$ for adhesive agent interactions, and $0 < \alpha \leq 1$ for repulsive agent interactions. This transition probability was first suggested by Khain et al. [33], and has been successfully applied to IBMs of wound-healing and cell aggregation [19, 33, 35].

We denote the occupancy of sites u and w by C_u and C_w , respectively, and let $u \in \mathcal{O}_n^k\{v\}$ and $w \in \mathcal{U}_n^k\{v\}$ represent all configurations of occupied and unoccupied members of the nearest neighbour set of site v , respectively. For a two-dimensional lattice the transition probability for adhesive/repulsive agent interactions with bias

can be written as

$$T(v'|v)_{A/R} = (1 - C_{v'}) \sum_{n=0}^{N-1} \frac{P_m(1 + \alpha)^n(1 + \rho_{v \rightarrow v'})}{N} \left[\sum_{k=1}^{\binom{N-1}{n}} \prod_{u \in \mathcal{O}_n^k\{v\}} C_u \prod_{w \in \mathcal{U}_n^k\{v\}} (1 - C_w) \right], \quad (3.3)$$

such that

$$0 \leq \sum_{v' \in \mathcal{N}\{v\}} \left[\frac{P_m(1 + \alpha)^n(1 + \rho_{v \rightarrow v'})}{N} \right] \leq 1. \quad (3.4)$$

Here, the subscript A/R denotes adhesive/repulsive agent interactions (adhesion or repulsion depends on the sign of α). Note that if $\alpha = 0$ Eq. (3.3) is equivalent to Eq. (3.1), the blind transition probability. This is because the sum of the probabilities of all possible neighbourhood configurations is equal to unity, i.e. the neighbourhood must be in some configuration. Once again, Eq. (3.4) is a condition that must be satisfied so that the probability of an agent attempting to move during $[t, t + \tau)$ does not exceed one.

Myopic agent interactions

The final interaction type we study is myopic agent interactions. A myopic agent never attempts to move into an occupied site [63]. Instead, an agent's probability of moving to an occupied site is redistributed amongst non-occupied sites. That is

$$\sum_{w \in \mathcal{U}_n^k\{v\}} T(w|v) = 1 \quad \forall k, n \in [0, \dots, N-1], \quad (3.5)$$

for an agent selected to move in the IBM. For a transition probability describing myopic agent interactions without bias this means an agent at site v attempts to move into an unoccupied site in $\mathcal{N}\{v\}$ with equal probability $P_m/(N - n)$, where n

signifies the number of nearest neighbours of v that are occupied. If bias is introduced, then not only is the number of occupied neighbours important, but also the position of the occupied neighbours. An example of myopic agent interactions is displayed in Fig. 3.2. For Fig. 3.2 (a), none of the nearest neighbour sites of v are occupied, so an agent at v would move left into site v' with probability

$$T(v'|v) = \frac{(1 - \rho_x)}{4}.$$

However, in Fig. 3.2 (b), a site in the nearest neighbourhood of v is occupied,

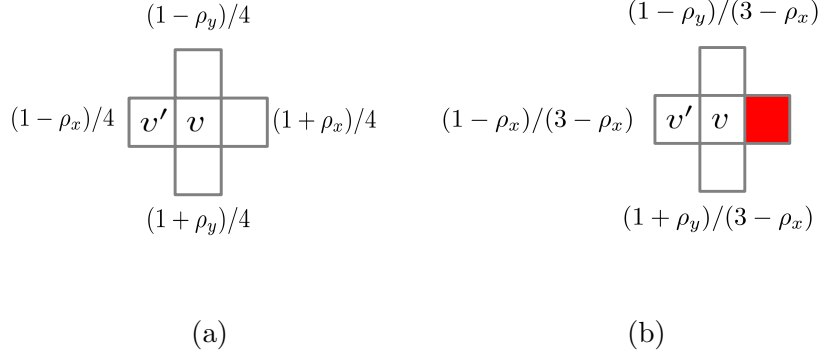


Figure 3.2: Myopic neighbourhood example. In panel (a) an empty nearest neighbourhood is displayed. With this configuration $T(v'|v) = (1 - \rho_x)/4$. In panel (b), a nearest neighbourhood in which one site is occupied is shown. The transition probability associated with this is $T(v'|v) = (1 - \rho_x)/(3 - \rho_x)$.

and so the probability of moving to this occupied site from v must be redistributed amongst the unoccupied sites. Therefore, an agent at site v would move to site v' with probability

$$T(v'|v) = \frac{\frac{(1 - \rho_x)}{4}}{\frac{(1 - \rho_x)}{4} + \frac{(1 + \rho_y)}{4} + \frac{(1 - \rho_y)}{4}} = \frac{(1 - \rho_x)}{(3 - \rho_x)}.$$

The transition probability for a myopic agent is

$$T(v'|v)_M = (1 - C_{v'}) \times \sum_{n=0}^{N-1} \left[\sum_{k=1}^{\binom{N-1}{n}} \left[\frac{(1 + \rho_{v \rightarrow v'})}{(1 + \rho_{v \rightarrow v'}) + \sum_{w \in \mathcal{U}_n^k\{v\}} (1 + \rho_{v \rightarrow w})} \prod_{u \in \mathcal{O}_n^k\{v\}} C_u \prod_{w \in \mathcal{U}_n^k\{v\}} (1 - C_w) \right] \right]. \quad (3.6)$$

The subscript M denotes that this is the transition rate for myopic agent interaction.

3.2.3 Population-level density continuum equations

We have described a range of transition probabilities that represent cell-cell interactions. In this section we derive corresponding continuum approximations for the spatial and temporal evolution of the population density and expected displacement of a tagged agent on both static and growing domains, and then compare our approximations with the results of our IBM simulations.

Static domains

We begin by deriving a continuum approximation for a static domain with agents that undergo blind interactions [39]. To derive our continuum approximation we proceed as follows. If the change in the average occupancy of site v during the time interval $[t, t + \tau)$ is denoted by $\delta \langle C_v(t) \rangle$, and the change in average occupancy due to a transition from $v \rightarrow v'$ is written as $T(v'|v)_B \langle C_v \rangle$ [39], then by summing over all members of the nearest neighbourhood set $\mathcal{N}\{v\}$ the discrete conservation equation for agents with blind agent interactions can be written as

$$\delta \langle C_v(t) \rangle = \sum_{v' \in \mathcal{N}\{v\}} [T(v|v')_B \langle C_{v'} \rangle - T(v'|v)_B \langle C_v \rangle]. \quad (3.7)$$

In Eq. (3.7) our transition rate, $T(v|v')_B$, is now treated as an ensemble average transition rate for blind agent interactions. That is, Eq. (3.7) gives the expected change in occupancy over many realisations, assuming independence of lattice site occupancy.

To derive a partial differential equation (PDE) describing the spatio-temporal evolution of the agent population density from Eq. (3.7) we make a change of variables, such that $v \rightarrow (x, y)$ and $\langle C_v(t) \rangle \rightarrow C(x, y, t)$, and then perform Taylor series expansions on Eq. (3.7). It has previously been shown [27, 38, 79] that Eq. (3.7) is related to a PDE in the limit as $\Delta \rightarrow 0$ and $\tau \rightarrow 0$, such that Δ^2/τ is held constant. Thus, the spatial and temporal evolution of the continuous density, $C(x, y, t)$, for a population with blind agent interactions on a two-dimensional square lattice is governed by

$$\frac{\partial C}{\partial t} = D \nabla^2 C - \nabla \cdot [\boldsymbol{\mu} C(1 - C)], \quad (3.8)$$

where $\boldsymbol{\mu} = (\mu_x, \mu_y)$ is the advective velocity with components

$$\mu_x = \lim_{\Delta, \tau \rightarrow 0} \left(\frac{P_m \rho_x \Delta}{2\tau} \right), \quad \mu_y = \lim_{\Delta, \tau \rightarrow 0} \left(\frac{P_m \rho_y \Delta}{2\tau} \right), \quad (3.9)$$

and diffusivity is given by

$$D = \lim_{\Delta, \tau \rightarrow 0} \left(\frac{P_m \Delta^2}{4\tau} \right), \quad (3.10)$$

which implies $\rho_x = O(\Delta)$ and $\rho_y = O(\Delta)$ [63].

If we follow the same procedure for the adhesive, repulsive and myopic agent interaction transition probabilities, that is Eqs. (3.3) and (3.6), we can derive equations for the spatial and temporal evolutions of the density, $C(x, y, t)$, for adhesive, repulsive

and myopic interactions. Further details are given in Appendix A. These equations are listed in Table 3.1.

It is of interest to note that the cell-cell interactions presented in this chapter can lead to negative diffusion for some values of C . This has been previously reported, and in these cases the PDE models may give rise to solutions containing shocks [35]. Given our choices for the values of the parameters in this chapter this is not a situation we encounter, however it is something that one should be aware of when implementing the techniques presented in this chapter.

Growing domains

We now add domain growth into the continuum approximations of our IBM. To do so we name the two horizontally adjacent sites of v , i.e. $(i, j - 1)$ and $(i, j + 1)$ as v_L and v_R , respectively. Similarly, the two vertically adjacent sites of v , i.e. $(i - 1, j)$ and $(i + 1, j)$ are defined as v_U and v_D , respectively. Ignoring agent movement, we can write a difference equation to account for the effect of growth in the x direction on the average occupancy status of site v , that is

$$\delta\langle C_v(t) \rangle = -\frac{P_g i}{N_x(t)}\langle C_v(t) \rangle + \frac{P_g(i-1)}{N_x(t)}\langle C_{v_L}(t) \rangle. \quad (3.13)$$

Here, P_g represents the probability of a growth event in the time duration $[t, t + \tau)$. The terms on the right-hand side (RHS) of Eq. (3.13) represent the ways in which a growth event can affect the occupancy status of site v . The first term on the RHS of Eq. (3.13) represents site v being occupied and a growth event occurring to the left of, and including, site v . The second term on the RHS of Eq. (3.13) represents a growth event occurring to the left of site v and causing site v to become/remain occupied. If we transform Eq. (3.13) from discrete to continuous variables and perform Taylor

Agent interaction	Population-level density equation
Adhesion ($\alpha > 0$) and repulsion ($\alpha < 0$)	$ \begin{aligned} \frac{\partial C}{\partial t} = & D \nabla \cdot [(1 - 8\alpha C + (5\alpha + 13\alpha^2)C^2 - (10\alpha^2 + 6\alpha)^3 C^3 + 5\alpha^3 C^4) \nabla C] \\ & - \nabla \cdot \left[\mu C \left(\frac{1}{5}(5\alpha^3 - 2\alpha^2)C^4 - \frac{1}{4}(4\alpha^3 + 10\alpha^2 - 4\alpha)C^3 \right. \right. \\ & \left. \left. + \frac{1}{3}(9\alpha^2 + 5\alpha - 2)C^2 - 3\alpha C + 1 \right) \right]. \end{aligned} \tag{3.11} $
Myopic	$ \begin{aligned} \frac{\partial C}{\partial t} = & D \nabla \cdot \left[\left(1 + \frac{8}{3}(C + C^2 + C^3) - 5C^4 \right) \nabla C \right] \\ & - \nabla \cdot \left[\mu C \left(\frac{1}{9}(13C^4 - 12C^3 - 6C^2 - 4C) + 1 \right) \right]. \end{aligned} \tag{3.12} $

Table 3.1: Population-level density equations for adhesive, repulsive and myopic agent interactions.

series expansions, noting that

$$N_x(t) = L_x(t)/\Delta, \quad (3.14)$$

and

$$P_g\Delta = L_x(t + \tau) - L_x(t), \quad (3.15)$$

we obtain

$$\frac{\partial C}{\partial t} = -L'_x(t) \frac{\partial}{\partial x} \left[\frac{x}{L_x(t)} C \right] + \Delta \frac{L'_x(t)}{2} \frac{\partial^2}{\partial x^2} \left[\frac{x}{L_x(t)} C \right]. \quad (3.16)$$

Here, $L'_x(t)$ represents the growth rate of our domain in the positive x direction. This is the same equation derived by Hywood et al. [31]. Importantly, in the derivation of Eq. (3.16) we do not explicitly take $\Delta, \tau \rightarrow 0$. This is because if we did the first term on the RHS of Eq. (3.16) would be divergent. This approach is not ideal, however it can be argued that this approach is reasonable as Δ is non-zero in the IBM simulations. This growth term can now be added to the previous equations describing the evolution of population density, Eqs. (3.8), (3.11) and (3.12), to represent the effects of domain growth. For instance, with the addition of growth Eq. (3.16) becomes

$$\frac{\partial C}{\partial t} + L'_x(t) \frac{\partial}{\partial x} \left[\frac{x}{L_x(t)} C \right] - \Delta \frac{L'_x(t)}{2} \frac{\partial^2}{\partial x^2} \left[\frac{x}{L_x(t)} C \right] = D\nabla^2 C - \nabla \cdot [\boldsymbol{\mu} C(1 - C)]. \quad (3.17)$$

We can combine motility and growth terms in this way as the motility and growth terms would be independent of each other in the derivation of Eq. (3.17).

3.2.4 Cell trajectories

Initially, we derive equations for the expected displacement of a tagged agent on a two-dimensional square lattice on a static domain.

Static domains

The expected displacement of a blind tagged agent at site v in the positive x direction, given its position along the x axis is p_x at time t , during the time-interval $[t, t + \tau)$ is [38]

$$\delta p_x = \Delta T(v_R|v)_B - \Delta T(v_L|v)_B. \quad (3.18)$$

Equation (3.18) states that the expected transition probability in the positive x direction multiplied by the jump length, minus the expected transition probability in the negative x direction multiplied by the jump length, gives the expected displacement of a blind tagged agent in the positive x direction. Equivalently, the expected displacement of a blind tagged agent in the positive y direction, given its position along the y axis is p_y at time t , is

$$\delta p_y = \Delta T(v_U|v)_B - \Delta T(v_D|v)_B. \quad (3.19)$$

If we rearrange Eq. (3.19) and divide by τ , with Δ^2/τ held constant as $\Delta, \tau \rightarrow 0$ simultaneously, it has been shown by Simpson et al. [38] that for an agent with blind agent interactions

$$\frac{dp_x}{dt} = -2D \frac{\partial C}{\partial x} + \mu_x(1 - C), \quad (3.20)$$

and

$$\frac{dp_y}{dt} = -2D \frac{\partial C}{\partial y} + \mu_y(1 - C), \quad (3.21)$$

where C , D , μ_x and μ_y are defined as previously.

If we follow this same procedure with the relevant transition probabilities, Eqs. (3.3) and (3.6), we can derive equations for the expected displacement of agents with adhesive, repulsive and myopic cell interactions. The resulting equations for the expected displacement of a tagged agent are listed in Table 3.2. Solving Eqs. (3.20)-(3.25) gives the expected displacement of a tagged agent initially at $[p_x(0), p_y(0)]$. As can be seen from Eqs. (3.20)-(3.25) the rate of change in the expected displacement of an agent depends explicitly on the average agent density. It should also be noted that if $\alpha = 0$ in Eqs. (3.22) and (3.23) we recapitulate the expected displacement equation for a blind agent.

Growing domains

To add growth to Eqs. (3.20)-(3.25) we proceed in the following manner. Ignoring agent movement, the expected displacement of an agent at position $p_x(t)$, at time t , in the positive horizontal direction during the time interval $[t, t + \tau)$ due to growth is

$$\delta p_x = P_g \Delta \frac{p_x}{L_x(t)}. \quad (3.26)$$

On dividing Eq. (3.26) by τ and taking $\Delta, \tau \rightarrow 0$, Eq. (3.26) becomes

$$\frac{dp_x}{dt} = L_x(t)' \frac{p_x}{L_x(t)}. \quad (3.27)$$

Agent interaction	Expected displacement equation in the x direction
Adhesion/repulsion	$\frac{dp_x}{dt} = -2D(1-\alpha)(\alpha C - 1)^2 \frac{\partial C}{\partial x} + \mu_x(C - 1)(\alpha C - 1)^3. \quad (3.22)$ $\frac{dp_y}{dt} = -2D(1-\alpha)(\alpha C - 1)^2 \frac{\partial C}{\partial y} + \mu_y(C - 1)(\alpha C - 1)^3. \quad (3.23)$
Myopic	$\frac{dp_x}{dt} = -2D\left(\frac{4}{3}\right)(1 + C + C^2) \frac{\partial C}{\partial x} + \frac{\mu_x}{9}(C - 1)^2(9 + 14C + 13C^2). \quad (3.24)$ $\frac{dp_y}{dt} = -2D\left(\frac{4}{3}\right)(1 + C + C^2) \frac{\partial C}{\partial y} + \frac{\mu_y}{9}(C - 1)^2(9 + 14C + 13C^2). \quad (3.25)$

Table 3.2: Expected displacement of a tagged agent initially at $[p_x(0), p_y(0)]$ for adhesive, repulsive and myopic agent interactions.

As with the population-level density equations we simply add this growth term to the expected displacement equations previously derived. For instance, with domain growth Eq. (3.20) becomes

$$\frac{dp_x}{dt} = -2D \frac{\partial C}{\partial x} + \mu_x(1 - C) + L'_x(t) \frac{p_x}{L_x(t)}. \quad (3.28)$$

Like Eq. (3.17), we can combine motility and growth terms in this way as they would be independent of each other in the derivation of Eq. (3.28).

3.3 Comparing continuum approximations to individual-based model results

We now compare our continuum approximations with the results of our IBM. The results presented here are not exhaustive, however they serve to present our key findings. For static domains simulations are performed on a square lattice with $N_x(t) = 200$ and $N_y(t) = 20$, with $\Delta = 1$. For growing domains one site is added to each row per time step for all simulations (i.e. $P_g = 1$)⁴. All sites between $85 \leq x \leq 115$ are initially occupied with agents, and an agent at site $(114, 10)$ is tagged. We chose this initial location for our tagged agent as it emphasises the differences between different types of agent interaction, as compared to other initial locations. All simulations have periodic boundary conditions at $y = 0$ and $y = N_y(t)$, and no-flux boundary conditions at $x = 0$ and $x = N_x(t)$. Together these conditions allow us to treat the system as translationally invariant in the y axis, hence we can write $\langle C(x, y, t) \rangle = \langle C(x, t) \rangle$.

In all simulations we consider three values of bias parameter in the x direction, $\rho_x = 0, 0.1, 0.5$, with $\rho_y = 0$ for all simulations. The density profiles from the IBM, $\langle C(x, t) \rangle$, are obtained by averaging the column occupancy for R identically

⁴In the case of $P_g < 1$, for instance $P_g = 0.5$, one site would be added to each row every two time steps. This could either be implemented deterministically or stochastically.

prepared realisations of the IBM. That is,

$$\langle C(x, t) \rangle = \frac{1}{R} \sum_{r=1}^R C_r(x, t), \quad (3.29)$$

where $C_r(x, t)$ is defined as the averaged column occupancy in a realisation r at time t . We also calculate the average displacement of the tagged agent over R identically prepared realisations of the IBM, that is

$$\langle x(t) \rangle = \frac{1}{R} \sum_{r=1}^R x_r(t), \quad (3.30)$$

where $x_r(t)$ represents the displacement of a tagged agent in realisation r at time t . Comparisons of simulation and continuum results are presented at $t = 100$ ($R = 100$) for static domains, and $t = 200$ ($R = 500$) for growing domains. We generate additional realisations on the growing domain as we simulate the IBM for a longer time, and the growth mechanism introduces additional stochasticity to the IBM results.

Finally, we introduce a measure for the ‘concavity’ of the expected displacement profile of our tagged agents. To do so we draw a straight line between the initial and final expected displacement of a tagged agent, and subtract the area beneath the straight line from the area beneath the expected displacement trajectories of the agent (using the trapezium rule). This difference is then normalised with respect to the displacement the agent has travelled. The concavity measure is illustrated graphically in Fig. 3.3. We will show that with this measure it is possible to identify different types of agent interaction.

On a static domain the population-level continuum Eqs. (3.8), (3.11) and (3.12) are solved with a centred-in-space finite difference method and implicit Euler method with Picard iteration with $\delta t = 0.05$ and $\delta x = 0.5$ for time integration, and abso-

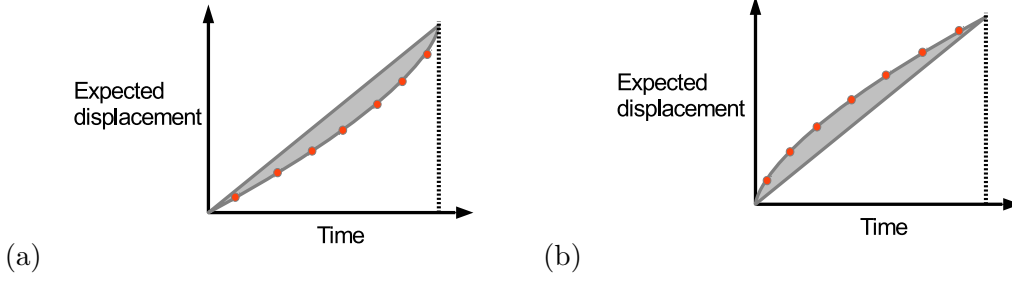


Figure 3.3: Concavity measure. The dotted line is the trajectory of a tagged agent. A straight line has been drawn connecting the start and final point of the tagged agent. The region between these two lines is shaded and used to calculate the concavity measure. This difference is then normalised with respect to the distance the agent has travelled. (a) Negative concavity, (b) positive concavity.

lute error tolerance $\epsilon = 10^{-8}$. From the solutions of these equations we approximate $\partial C/\partial x$, $\partial C/\partial y$ and C using linear interpolation, and then solve Eqs. (3.20), (3.22) and (3.24) with initial position $[p_x(0), p_y(0)]$, as in [38].

For simulations on a growing domain we implement the change of variables used by Crampin et al. [80], that is $\xi = x/L(t)$ and $t = \tau$. This allows Eqs. (3.8), (3.11) and (3.12) to be solved on a fixed domain [25, 80]. The growing domain equations are solved with a centred-in-space finite difference method and implicit Euler method with Picard iteration with $\delta\tau = 0.05$ and $\delta\xi = 0.1$, and absolute error tolerance $\epsilon = 10^{-8}$.

3.3.1 Static domain

We begin by testing the accuracy of our continuum approximations of our IBM on static domains.

Adhesive agent interactions

Figure 3.4 row (a)-(b) shows results from IBM simulations and the corresponding continuum approximation for blind agent interactions. The averaged density profile

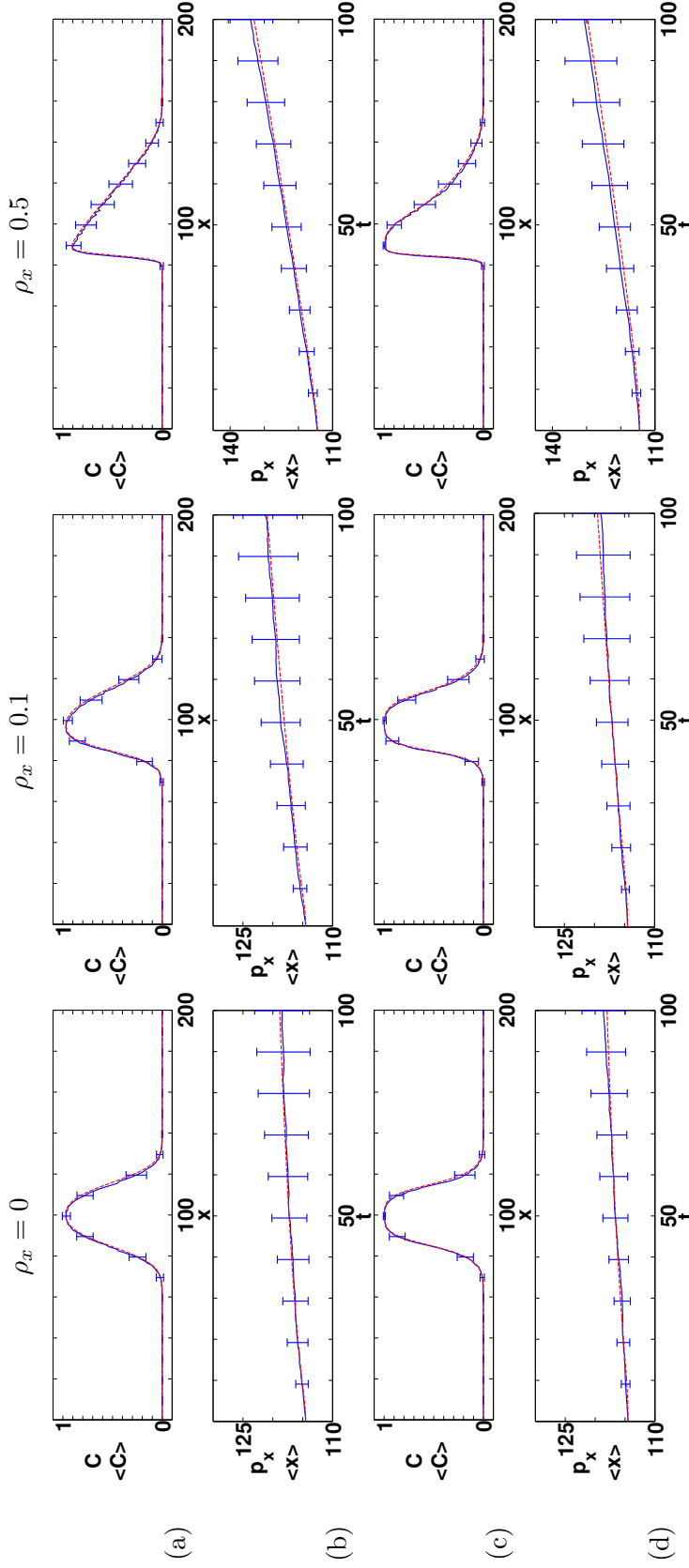


Figure 3.4: Agent population density and expected displacement for agents with blind and adhesive interactions at $t = 100$. Row (a) shows the averaged density profiles, $\langle C \rangle$ (blue), of the IBM simulation in comparison with the solution of Eq. (3.8), C (red), for blind agent interactions. Row (b) shows the expected displacement of the tagged agent for blind interactions. IBM simulation results (blue) are compared with continuum Eq. (3.20) (red). Row (c) shows the averaged density profiles, $\langle C \rangle$, of the IBM simulation in comparison with the solution of Eq. (3.11), C , for adhesive interactions ($\alpha = -0.2$). Row (d) shows the expected displacement for the tagged agent for adhesive interactions. Simulation results are compared with continuum Eq. (3.22). Error bars display the standard error of the mean of the IBM simulations.

of the IBM simulations, taken from 100 repeats, compares well with the solution of Eq. (3.8) for all values of ρ_x , as can be seen in Fig. 3.4 row (a) (largest percentage error (LPE) for $(\rho_x = 0, \rho_x = 0.1, \rho_x = 0.5)$ is $(0.13, 0.14, 0.06)$)⁵. The effect of the non-linear advection term introduced into Eq. (3.8) by a non-zero ρ_x is seen in the asymmetry of the density plots, whereby the agent density skews to the right and reduces in height more quickly compared to agents without motility bias. Figure 3.4 row (b) shows that the expected displacement of the tagged agent predicted by the solution of Eq. (3.20) compares well with the IBM simulation for all values of ρ_x (0.01, 0.03, 0.02).

Figure 3.4 row (c) shows a comparison of the IBM simulation and continuum approximation solution for agents with adhesive interactions with $\alpha = -0.2$. Like the blind case, the averaged density profiles of the IBM simulations and the solution of Eq. (3.11) compare excellently for all values of ρ_x for this value of α (0.08, 0.07, 0.10). However, compared to agents with blind interactions, agents with adhesive interactions spread less, due to fewer movements being attempted. Importantly, the density profile for the agent adhesion process is convex on its RHS in the presence of motility bias (the direction of positive bias). This is due to two reasons: (i) agents with adhesive interactions have dispersed less by $t = 100$; (ii) the adhesive interaction means once an agent has left a high density region it is more likely to move again, resulting in the density skewing in the direction of positive motility bias. Finally, Fig. 3.4 row (d) shows that the evolution of the expected displacement for the tagged agent predicted by Eq. (3.22) compares well with the simulation data for all values of ρ_x (0.02, 0.02, 0.02). In the case of agent adhesion the expected displacement of the agent is less than for an agent with blind interactions, as would be expected.

⁵Throughout this thesis we quantify the error between the IBM results and continuum models using the largest percentage error (LPE). This is the largest percentage (relative) difference between an IBM simulation and the continuum model results at any point during the simulation.

Repulsive agent interactions

We now study repulsive-type agent interactions, that is, agents with repulsive and myopic interactions. Figure 3.5 row (a) shows the averaged density profile for the IBM simulation and continuum approximation solution for agents with repulsive interactions for $\alpha = 0.2$. Compared to agents with blind interactions, agents with repulsive interactions spread more due to movement probability being ‘redistributed’ from occupied sites to unoccupied sites in the nearest neighbourhood set. As with blind and adhesive interactions the density profiles of the averaged simulation data and solution of Eq. (3.11) compare excellently for all values of ρ_x for $\alpha = 0.2$ (0.05, 0.06, 0.04). This excellent agreement between the continuum approximation and IBM is maintained for any $0 < \alpha < 0.2$. When this threshold is passed the condition stipulated by Eq. (3.4) is violated and the continuum approximation becomes inaccurate (dependent on the value of the motility bias, ρ_x).

In contrast to agents with adhesive interactions, the density profile for agents with repulsive interactions is concave on the RHS of the agent population in the presence of bias. This is because agents with repulsive interactions have dispersed more by $t = 100$. This is an important observation as visualising the profile of the cells at the edge of a cell population is experimentally tractable. We suggest that looking at the shape of the leading edge could aid researchers in identifying cell behaviours and interactions. For instance, a chemoattractant could be placed in the vicinity of a cell population *in vitro*, and the density profile at the edge of this population could then be measured to test whether there are adhesive or repulsive interactions between the cells. Figure 3.5 row (b) shows that the evolution of the expected displacement for the tagged agent predicted by Eq. (3.22) compares well with the simulation data for all values of ρ_x (0.02, 0.01, 0.03). As predicted, the expected displacement of an agent

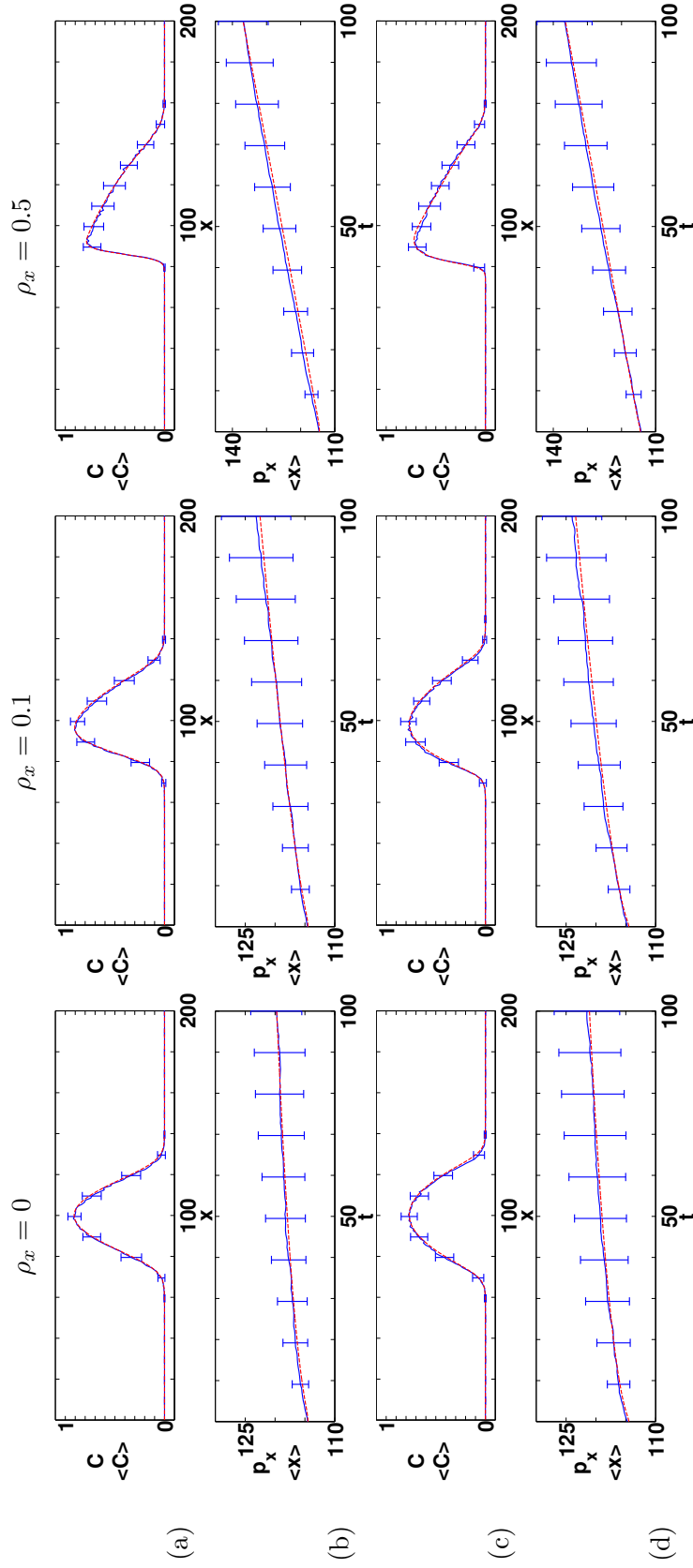


Figure 3.5: Agent population density and expected displacement for agents with repulsive and myopic interactions at $t = 100$. Parameters and figure descriptions are the same as those in Fig. 3.4. Rows (a)-(b) are repulsive agent interactions ($\alpha = 0.2$) and Eq. (3.11) and (3.22), respectively. Rows (c)-(d) are myopic agent interactions and Eq. (3.12) and (3.24), respectively.

with repulsive interactions is greater than would be expected for an agent with blind or adhesive interactions. Therefore, the magnitude of the displacement of individual agents is indicative of the type of interactions between agents.

Finally, Fig. 3.5 row (c) shows the IBM simulations and continuum approximation solutions for agents with myopic interactions. Like all other agent interactions the density profiles of the averaged simulation data and the solution of Eq. (3.11) compare excellently for all values of ρ_x (0.02, 0.06, 0.06). As with agents with repulsive interactions the RHS of the agent population exhibits a concave profile. The myopic agent interaction can be thought of in some sense to be a ‘maximised’ repulsive interaction. The expected displacement Eq. (3.24) also agrees well with the simulation data for all values of ρ_x , as can be seen from Fig. 3.5 row (d) (0.01, 0.02, 0.01). A tagged agent with myopic interactions exhibits the greatest expected displacement by $t = 100$ compared to the other interactions, which we would expect.

In Fig. 3.6 the evolution of the standard error of the mean (SEM) for the displacement of the tagged agent in the IBM for all agent interactions has been plotted against time with and without motility bias. It can be seen that without bias the repulsive-type interactions exhibit a larger SEM in the displacement of the tagged agent at all time points in the IBM, when compared to blind and adhesive interactions. Therefore, we could expect cells with repulsive-type cell-cell interactions to exhibit greater variance in their displacement in cell trajectory experiments. In the presence of motility bias, however, the SEM of different agent interactions becomes less distinguishable. This is the opposite to what is found when considering the population density, in which case it is in the presence of motility bias that different agent interactions become distinguishable. This difference between the collective and individual migration of agents is an interesting observation.

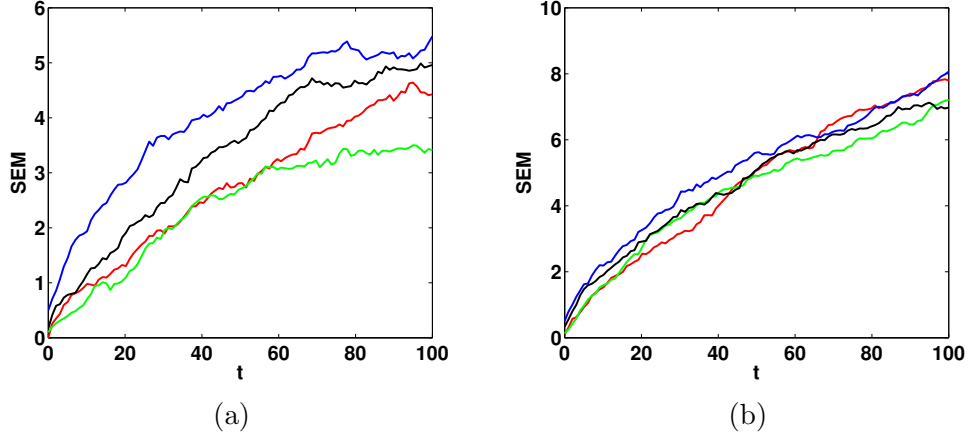


Figure 3.6: The evolution of the SEM for the position of the tagged agent for all agent interactions. The SEM was calculated from 100 repeats of the IBM for each agent interaction. (a) $\rho_x = 0$, (b) $\rho_x = 0.5$. Blind (red), adhesive (green), repulsive (black), myopic (blue).

3.3.2 Growing domain

We now compare the solutions of our continuum approximations to our IBM simulations on a growing domain. Figure 3.7 row (a) shows a comparison of the IBM and continuum model results for agents with adhesive interactions (0.02, 0.03, 0.10). The density profiles of the averaged IBM data and the solution of Eq. (3.11) are not as accurate as compared to a static domain for adhesive interactions (although the inaccuracy is only minor). This is noticeable on the left-hand side (LHS) of the agent population for $\rho_x = 0.5$. However, the agent population retains its convex character in the presence of motility bias on its RHS. Figure 3.7 row (b) shows that the expected displacement of the tagged agent predicted by the solution of Eq. (3.22) compares well with the simulation data for all values of ρ_x (0.01, 0.01, 0.02). Compared with a static domain the displacement of the tagged agent is much greater, as we would expect due to domain growth.

Figure 3.7 row (c) shows the spatial and temporal evolution of the population density for agents with repulsive interactions on a growing domain (0.02, 0.02, 0.00). The

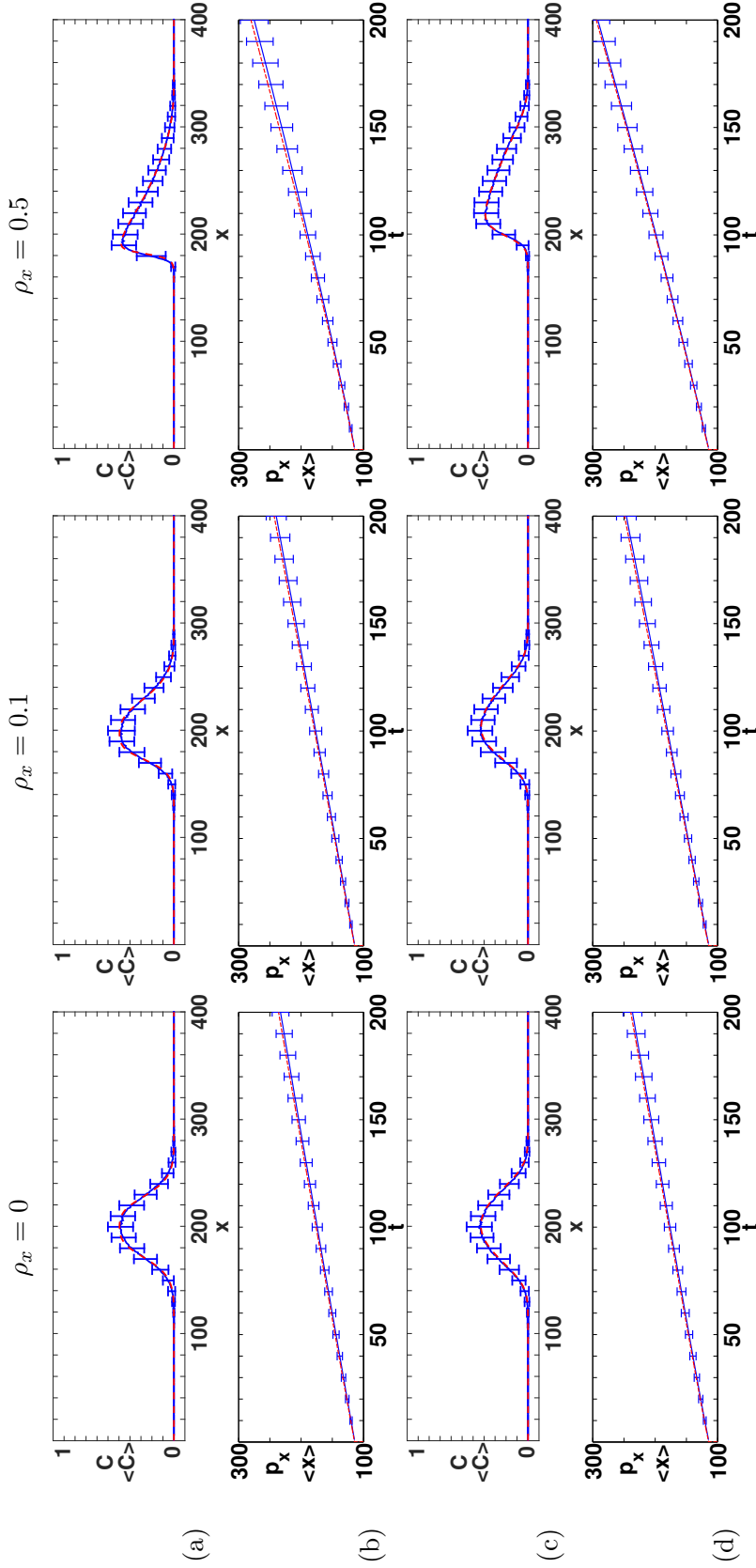


Figure 3.7: Agent density and expected displacement for agents with adhesive and repulsive interactions on a growing domain at $t = 200$. Row (a) shows the averaged density profiles, $\langle C \rangle$ (blue), of the IBM simulations in comparison with the solution of Eq. (3.11), C (red), for adhesive interactions ($\alpha = -0.2$). Row (b) shows the expected displacement for the tagged agent for adhesive interactions. IBM results (blue) are compared with continuum Eq. (3.22) (red). Row (c) shows the averaged density profiles, $\langle C \rangle$, of the IBM in comparison with the solution of Eq. (3.11), C , for repulsive interactions ($\alpha = 0.2$). Row (d) shows the expected displacement for the tagged agent for repulsive interactions. IBM results are compared with continuum Eq. (3.22). Error bars are the SEM of the IBM simulation.

averaged density profile of the IBM simulations compares excellently with the continuum solution of the growth analogue of Eq. (3.11) for all values of ρ_x . This is different to the adhesion process, where the continuum solution does not accurately predict the averaged density profile on the LHS of the population. Like the non-growing domain the density profile for agents with repulsive interactions is concave on the RHS of the agent population. Therefore, both agents with adhesive and repulsive interactions maintain their population density characteristics caused by motility bias in the presence of growth. Finally, as before, Fig. 3.7 row (d) shows that the evolution of the expected displacement of the tagged agent predicted by Eq. (3.22) compares excellently with the IBM simulations for all values of ρ_x (0.01, 0.01, 0.00).

Summary of population density results

These results, for all agent interactions, have shown that our continuum approximations of our IBM are accurate for static domains, and for the parameters values presented here could be used as our model instead of our IBM. This is important as IBM simulations can be time-consuming and difficult to complete parameter sweeps with. However, with the addition of domain growth our continuum approximations for the spatial and temporal evolution of the population density for agents with adhesive interactions were less accurate, especially on the LHS of the agent population. We found this to be the case for agents with myopic interactions too, but not for agents with blind interactions, which gave results that were similar in accuracy to agents with repulsive interactions. Finally, our continuum approximations for the expected displacement of a tagged agent were accurate for all four interactions for the parameter values presented here (data not shown for blind and myopic interactions).

3.3.3 Tagged agent trajectories

We now briefly discuss the results of our continuum approximations for the expected displacement of a tagged agent in more detail, and illustrate how they could be used to identify different agent interactions. Figure 3.8 shows the expected displacement of the tagged agent for all interactions when $\rho_x = 0$. All four trajectories have inflection

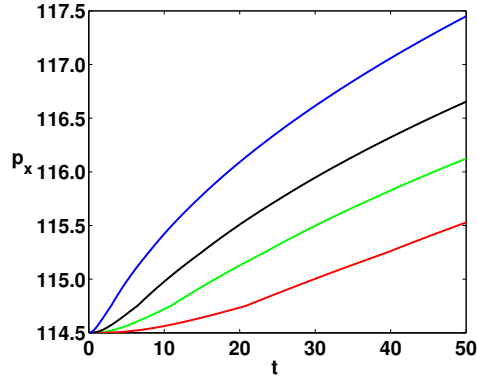


Figure 3.8: The expected displacement trajectory of a tagged agent in the positive x direction for all interactions with respect to time. Eqs. (3.20), (3.22) and (3.24) with $\rho_x = 0$ and on a static domain. Blind (green), adhesive (red), repulsive (black), myopic (blue).

points. In the case of blind agent interactions this signifies the point at which the tagged agent passes the inflection point in the agent population density profile. This can be seen most easily by looking at Eq. (3.20) with $\rho_x = 0$. If we take the derivative of Eq. (3.20) with respect to time we obtain

$$\frac{d^2 p_x}{dt^2} = -2D \frac{\partial}{\partial t} \left(\frac{\partial C}{\partial x} \right). \quad (3.31)$$

Additionally, we know that

$$\frac{\partial C}{\partial x} \leq 0 \quad \forall t, \quad \forall x > 100. \quad (3.32)$$

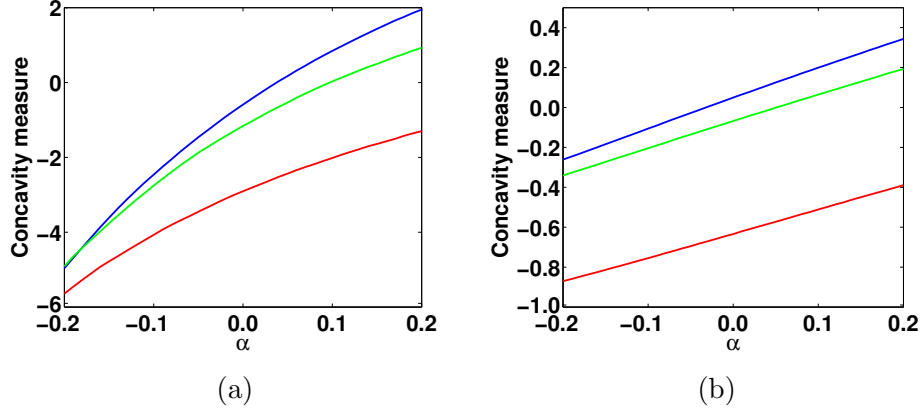


Figure 3.9: Agent trajectory concavity measure with respect to adhesion parameter α for a tagged agent initially situated at (114, 10) in the continuum model. (a) Static domain, $\rho_x = 0$ (blue), $\rho_x = 0.1$ (green), $\rho_x = 0.5$ (red). (b) Growing domain, same colour scheme as (a).

So for d^2p_x/dt^2 (Eq. (3.31)) to change sign we require $\partial/\partial t(\partial C/\partial x)$ to change sign at the position of the tagged agent. This occurs when the tagged agent passes the inflection point in the agent population density profile. This observation, although qualitative, could be of some use to researchers trying to identify cell properties through cell trajectories. Cells with repulsive cell-cell interactions may be expected to exhibit a convex trajectory in their displacement for a shorter amount of time (see black line, Fig. 3.8), after which they exhibit a concave trajectory. Conversely, the expected displacement trajectory of cells with adhesive cell-cell interactions exhibit a convex shape for a longer amount of time compared to cells with repulsive cell-cell interactions (see red line, Fig. 3.8).

Figure 3.9 displays how the concavity measure of a tagged agent depends on the adhesion parameter α . Specifically, Fig. 3.9 (a) shows the concavity measure for a static domain at $t = 50$. The blue line shows the non-biased case, $\rho_x = 0$. We can see that for most positive values of α (i.e. repulsive interactions) the trajectory has a positive concavity measure. However, as α approaches zero, that is, interactions become blind, the concavity measure becomes negative, as for adhesive interactions the

tagged agent trajectory is convex. Therefore, we suggest that the concavity measure could be used to identify repulsive and adhesive cell-cell interactions experimentally. In the presence of a strong motility bias, $\rho_x = 0.5$, the concavity measure is negative for all values of α . This is important as it suggests that repulsive interactions may not exhibit concave profiles in the presence of motility bias. Meanwhile, Fig. 3.9 (b) shows the same continuum approximation solutions, with the addition of growth, at $t = 50$. Here we see that while growth reduces the magnitude of the concavity measure of the trajectory of a tagged agent with repulsive interactions, it does not make the concavity measure negative for repulsive interactions. Therefore, the concavity measure associated with different agent interactions are conserved in the presence of the growth mechanism implemented in this chapter.

It is important to discuss how the concavity measurement will depend on the time period over which it is measured. This is because the rate of change of the concavity measure reverses in sign in value after an inflection point in the cell trajectory. If we observed a cell with adhesive cell-cell interactions with no directional bias on a static domain at the edge of a cell mass, such as in Fig. 3.4 row (d), we would eventually expect an inflection point in its expected displacement trajectory (see Fig. 3.8, red line) [38]. Therefore, it would be necessary to study cell trajectories over an appropriate timescale. The relationship between the variance in cell trajectory data and the timescale of measurement has been previously studied, and it was shown that collecting large numbers of repeats over short timescales is most effective in reducing the variance of cell trajectory data [81]. Therefore, given the nature of cell trajectory experiments we would only recommend use of the concavity measure over short time scales with multiple repeats, as with long time scales both theoretical and experimental difficulties with cell trajectory experiments arise. However, it stands to reason that over short timescales the different concavity properties of cell trajectories with

adhesive and repulsive cell-cell interactions presented here could be viable, as long as the experiment is effectively controlled.

3.4 Summary

In this chapter we have extended previously formulated IBMs and continuum models for a range of agent interactions, representing known cell-cell interactions in biological systems. These have been implemented on both static and growing domains. The density profiles resulting from different types of interactions are differently shaped. Agents with myopic and repulsive cell-cell interactions display concave profiles in the presence of a motility bias, whereas agents with adhesive interactions exhibit a convex profile in the presence of motility bias. These agent density profiles were conserved on both static and growing domains, suggesting these profile differences would be conserved in the presence of domain growth, an important consideration when considering whether experimental data collected *in vitro* is applicable to growing domain environments. Visualising the profile of the cells at the edge of a cell population in the presence of a chemoattractant is potentially experimentally tractable, and so this knowledge could help researchers identify different cell-cell interactions.

We have also presented continuum equations describing the expected displacement of a tagged agent with different interaction mechanisms. By studying the expected trajectory of a tagged agent with different agent interactions (Fig. 3.8) we showed that different agent interactions give rise to different trajectory profiles and magnitudes. Agents with repulsive interactions exhibited concave trajectories compared to agents with adhesive interactions. In the presence of strong motility bias the differences in the trajectories of tagged agents with different agent interactions was lost. Moreover, agents with repulsive-type interactions exhibited greater expected displacement compared to agents with adhesive interactions. This is an obvious but important con-

sideration when analysing cell trajectory data, and could be used to identify cell-cell interactions. We also found that repulsive-type agent interactions exhibited a greater SEM when compared to blind or adhesive interactions. Therefore, the SEM could also be used as a potential means by which to distinguish between repulsive-type cell-cell interactions. The difference in the magnitude of the SEM for tagged agents with different agent interactions was lost in the presence of a strong motility bias. This highlights how understanding the effect of different agent interactions in the presence of motility bias is not straightforward. Ironically, the presence of motility bias is what allows us to distinguish between different agent interactions when looking at the population density. However, the presence of motility bias obscures the differences between the trajectories of tagged agents with different agent interactions. This is an interesting difference between analysing population and individual agent level statistics from our individual-based and continuum models.

In this chapter the parameter values we use mean that our PDE models never exhibit negative diffusion. However, this is a phenomenon that has been previously examined [35] and is something to be aware of when developing models for cell-cell interactions if using the framework presented in this chapter.

.....

In this chapter we have ignored the effect of agent proliferation in our IBM and continuum models. Cell proliferation is of obvious importance in developmental biology, indeed in Chapter 2 we studied an example of a proliferative cell population on a growing domain. Chapter 2 also demonstrated how agent proliferation generates spatial correlations between agents in the IBM, and in the case of melanoblast colonisation of the DLE, these spatial correlations are hypothesised to underpin stripe formation in chimeric mice.

The spatial correlations generated by agent proliferation in IBMs can lead to some continuum approximations of IBMs being inaccurate. This means the derivation of accurate continuum models including agent proliferation often requires the use of different moment closure approximations to those presented thus far. The central topic of the next three chapters will be deriving accurate continuum approximations of IBMs that include agent proliferation, for IBMs with growing domains.

Chapter 4

Modelling spatial correlations on a growing domain in one dimension

Overview

In this chapter we present methods to include the effects of domain growth on the evolution of spatial correlations between agent locations in a continuum approximation of a one-dimensional lattice-based model of cell motility and proliferation. This is important as the inclusion of spatial correlations in continuum models of cell motility and proliferation without domain growth has previously been shown to be essential for their accuracy in certain scenarios [18, 21–23]. We include the effect of spatial correlations by deriving a system of ODEs that describe the expected evolution of individual and pairwise density functions for agents on a growing domain. We then demonstrate the simplification of this system of ODEs by using an appropriate approximation. This simplification allows domain growth to be included in models describing the evolution of spatial correlations between agents in a tractable manner.

The majority of the work in this chapter appears in the paper: R. J. H. Ross, C. A. Yates and R. E. Baker. “The effect of domain growth on spatial correlations”. *Physica A*, 466:334-345, 2016. R. J. H. Ross performed all the computational and mathematical modelling presented in this chapter.

4.1 Introduction

In Chapters 2 and 3 we have seen how IBMs can play a significant role in research into cell migratory processes. In Chapter 2, we demonstrated that an IBM parameterised by experimental data could be used to validate a scientific hypothesis regarding colonisation of the DLE by melanoblasts. In Chapter 3 we saw that IBMs allow complex behaviours, such as cell-cell interactions, to be easily assigned to agents in an IBM [12, 33–35]. The effects of cell-cell interactions on cell migratory processes can then be studied with the IBM, potentially leading to the generation of new hypotheses that can be tested experimentally. Importantly, IBMs also capture spatial correlations established by cell proliferation, and the ramifications spatial correlations can have on density-dependent processes such as cell migration and proliferation [18–23, 44, 50, 51, 82, 83]. An example of this is presented in Chapter 2, whereby the coalescence of clusters of agents, established by agent proliferation, formed chimeric stripes in the IBM. The main focus of the next three chapters will be the study of proliferative agent populations on growing domains and, in particular, modelling the evolution of the spatial correlations established by agent proliferation in these populations.

In Chapter 3 we demonstrated that IBMs are often amenable to approximation by population-level continuum models. Accurate continuum approximations of IBMs are important tools for understanding biological systems as, in contrast to IBMs, they generally allow for more mathematical analysis. As discussed in the thesis introduction, this analysis can be crucial in forming a mechanistic understanding of biological systems, which is not always apparent (or feasible) from simply studying the averaged results of a large number of repeats of an IBM. However, in certain scenarios standard mean-field continuum descriptions of IBMs, such as those describing the expected evolution of the population density, suffer from the limitation that they neglect

to incorporate the impact of spatial correlations and clustering. Therefore, in order to derive accurate continuum approximations of IBMs it is often necessary to include the effects of spatial correlations in continuum models [18–23, 50, 51, 78, 82, 84–87]. Furthermore, having the mathematical tools to directly compute spatial correlations allows them to be analysed, which can give important insights into the biological process being studied. For instance, spatial correlations indicative of different types of cell-cell interactions can be observed in cell populations [44, 45, 88], and spatial correlations between cells are thought to play an important role in tumour growth [89].

In this chapter we examine how domain growth affects the evolution of individual and pairwise density functions for agents in an IBM. A large body of literature already exists concerning the evolution of individual and pairwise density functions on static domains [18–23, 50, 51, 82, 83], the most striking examples of which show that standard mean-field PDE descriptions can be wholly insufficient approximations of the evolution of the agent density in IBMs in certain scenarios [18, 21]. We therefore also demonstrate the integration of the results presented in this chapter into pre-existing models. In doing so we simplify the implementation of the methods we present so that they can be more easily applied to the study of complex systems.

The outline of this chapter is as follows: to begin we introduce our one-dimensional IBM and domain growth mechanism in Section 4.2.1. We then define the individual and pairwise density functions, and derive a system of ODEs describing the evolution of the individual and pairwise density functions with respect to time on a growing domain in Section 4.2.2. To test the accuracy of this system of ODEs we compare its numerical solution with ensemble averages of the individual and pairwise agent density functions from the IBM for a range of initial conditions and parameter values in Section 4.3. In Section 4.4 we integrate domain growth into existing models for

calculating the evolution of pairwise spatial correlations. These models are typically used to correct mean-field approximations (MFAs) for the evolution of the agent density in an IBM by taking spatial correlations into account. In Section 4.5 we conclude with a summary of the results presented in this chapter.

4.2 Model

In this section we first introduce the IBM and the domain growth mechanism we employ throughout this chapter. We then introduce the individual and pairwise density functions and derive a system of ODEs describing the evolution of these functions.

4.2.1 One-dimensional individual-based model and the domain growth mechanism

We use an agent-based, discrete random-walk model on a one-dimensional regular lattice with lattice spacing Δ [46] and length $L(t)$, where $L(t)$ is an integer describing the number of lattice sites. Throughout this chapter the lattice site spacing, Δ , is always taken equal to one¹. All analysis and simulations are performed with periodic boundary conditions. Each agent is assigned to a lattice site, from which it can move or proliferate into an adjacent site. If an agent attempts to move into a site that is already occupied, the movement event is aborted. Similarly, if an agent attempts to proliferate into a site that is already occupied, the proliferation event is aborted. Time is evolved continuously, in accordance with the Gillespie algorithm [90], such that movement, proliferation and growth events are modelled as exponentially distributed ‘reaction events’ in a Markov chain. Attempted agent movement or proliferation events occur with rates P_m or P_p per unit time, respectively. That is, $P_m \delta t$ is the probability of an agent attempting to move in the next infinitesimally

¹Note, however, that Δ does not have to be equal to one for the results presented in this chapter to hold.

small time interval δt . Throughout this chapter the initial agent distribution for all simulations is achieved by populating lattice sites uniformly at random until the required initial density is achieved².

We employ a stochastic growth mechanism in one dimension in which the insertion of new lattice sites into the domain occurs with rate $P_g L(t)$ per unit time. That is, each individual lattice site undergoes growth events with rate $P_g > 0$. When a growth event occurs a site on the domain is selected uniformly at random and a new (empty) lattice site is inserted at the selected site, as can be seen in Fig. 4.1. For notational convenience from this point on we write $L(t)$ as L .

Figure 4.1 also displays how the distance ‘ r ’ between two lattice sites on the domain is defined. This distance r is measured between the centres of the lattice sites and in the clockwise direction. This means that on a domain of length L there are L counts of each distance $r \in [1, \dots, L - 1]$ between pairs of lattice sites.

4.2.2 System of ordinary differential equations

We begin by introducing the density functions. A density function, ρ_n^L , is defined as the probability that n sites have given occupancies when there are L sites in the domain [18–23]. For instance, a pairwise density function, $\rho_2^L(A_j, A_k; t)$, is the probability that sites j and k are both occupied by an agent ‘ A ’ at time t on a domain of length L (i.e. contains L sites)³. An individual density function, $\rho_1^L(A_j; t)$, is the probability that site j is occupied by an agent ‘ A ’ at time t on a domain of length L .

²An alternative method to generate the same average initial density in the simulations would be to populate each lattice site uniformly at random with the probability of the initial density required. This method was also implemented and found to make no difference to the results (not shown).

³In our notation we explicitly write the pairwise density function with ‘ A ’ to indicate the occupancy of a lattice site by an agent. This is because when we introduce agent proliferation into the model terms such as $\rho_2^L(0_j, A_k; t)$ will be present in the equations, where $\rho_2^L(0_j, A_k; t)$ is taken to mean the probability that site j is unoccupied and site k is occupied by an agent at time t . This notation is also easily extendable to multispecies models [21], which will be necessary in due course.

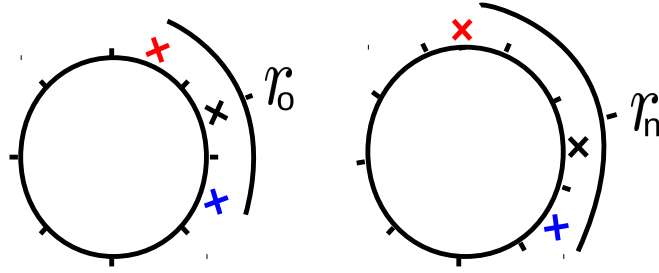


Figure 4.1: A one-dimensional domain with periodic boundary conditions can be represented as a circle. A site marked by a black cross has been chosen to undergo a growth event. Following this event the site marked by the black cross and its contents are moved one lattice spacing in the clockwise direction on the circle and a new lattice site is inserted. The new lattice site is empty. The distance between the lattice site marked by the red cross and that marked by the blue cross is measured in the clockwise direction. In this example the distance between the two sites initially, r_o (r old), is two. Following the growth event the distance between these lattice sites, r_n (r new), is three.

In the course of this derivation the pairwise density functions will be rewritten in terms of the distance between two sites, r . We can do this as we assume translational invariance of the density functions throughout this chapter. This is a valid assumption because the initial condition used in our IBM simulations is achieved by populating sites uniformly at random until we have achieved a required initial density. Two important consequences of assuming translational invariance are

$$\rho_1^L(A_j; t) = \rho_1^L(A_{j+\ell}; t), \quad \forall j, \ell \in [1, \dots, L], \quad (4.1)$$

and

$$\rho_2^L(A_j, A_k; t) = \rho_2^L(A_{j+\ell}, A_{k+\ell}; t), \quad \forall j, k, \ell \in [1, \dots, L]. \quad (4.2)$$

Equation (4.1) simply means that the probability of a site being occupied is independent of its location on the domain. Similarly, Eq. (4.2) means that the probability of two sites a distance r apart being occupied is independent of the location of the two

sites on the domain (where $j + r = k$)⁴. Importantly, Eqs. (4.1) and (4.2) allow us to simplify the following derivation greatly as $\rho_2^L(A_j, A_k; t)$ can be written as a function of the distance between two sites, so that

$$\rho_2^L(A_j, A_k; t) = \rho_2^L(A_j, A_{j+r}; t) = \rho_2^L(r; t). \quad (4.3)$$

This ‘abuse’ of notation, whereby ρ_2^L is rewritten as a function of the distance between two lattice sites, as opposed to the lattice sites themselves, will prove useful in the following derivation.

To begin with, we only model the effects of *domain growth* on the density functions (the effects of agent *motility* and *proliferation* on the density functions have been derived previously and will be included in due course). On a growing domain the individual probability density functions evolve according to

$$\sum_{j=1}^L \rho_1^L(A_j; t + \delta t) = (1 - \delta t P_g L) \sum_{j=1}^L \rho_1^L(A_j; t) + \delta t P_g \sum_{j=1}^{L-1} (L - 1) \rho_1^{L-1}(A_j; t) + \mathcal{O}(\delta t^2). \quad (4.4)$$

That is, the sum of individual probability density functions on a domain of length L at $[t + \delta t)$ corresponds to the following terms on the RHS of Eq. (4.4): i) the sum of individual probability density functions at time t multiplied by the probability that no growth event occurs in $[t, t + \delta t)$; and ii) the sum of individual probability density functions at time t on a domain of length $L - 1$ multiplied by the probability that a growth event occurs.

To simplify Eq. (4.4) we rewrite $\rho_1^L(A_j; t)$ as $c^L(t)$ (and $\rho_1^{L-1}(A_j; t)$ as $c^{L-1}(t)$), which

⁴Translational invariance in the initial conditions also means $\rho_2^L(A_j, A_k; t)$ is symmetric under permutation, i.e. $\rho_2^L(A_j, A_k; t) = \rho_2^L(A_k, A_j; t)$. Equations (4.1) and (4.2) also mean $j + \ell \bmod L$ is the location of site $j + \ell$.

we can do due to Eq. (4.1), and obtain

$$c^L(t + \delta t) = (1 - \delta t P_g L) c^L(t) + \delta t P_g \left(\frac{L-1}{L} \right) (L-1) c^{L-1}(t) + \mathcal{O}(\delta t^2). \quad (4.5)$$

Finally, if we rearrange Eq. (4.5) and take the limit as $\delta t \rightarrow 0$, we arrive at

$$\frac{dc^L(t)}{dt} = -P_g L c^L(t) + P_g \left(\frac{L-1}{L} \right) (L-1) c^{L-1}(t). \quad (4.6)$$

Equation (4.6) is a system of coupled ODEs for the evolution of the individual agent density for each domain length, L . That is, the probability of a site being occupied on a domain of length L at time t . It is of interest to note that Eq. (4.6) is exactly solvable. A generating function approach can be employed to rewrite Eq. (4.6) as a first-order linear PDE, which can then be solved using the method of characteristics [41]. However, this approach is only valid for systems *without* agent proliferation, and so is something we do not pursue further.

We now consider the evolution of the pairwise density functions. As previously discussed we can rewrite the pairwise density functions $\rho_2^L(A_j, A_k; t)$ as $\rho_2^L(A_j, A_{j+r}; t)$ for $r = 0, \dots, L-1$ (if $j+r > L$ the pairwise density function is instead $\rho_2^L(A_j, A_{j+r-L}; t)$ as the domain boundary is periodic). The evolution of the pairwise density functions on a growing domain can be written as

$$\begin{aligned} \sum_{j=1}^L \rho_2^L(A_j, A_{j+r}; t + \delta t) &= (1 - \delta t P_g L) \sum_{j=1}^L \rho_2^L(A_j, A_{j+r}; t) \\ &\quad + \delta t P_g \sum_{j=1}^{L-1} (r-1) \rho_2^{L-1}(A_j, A_{j+r-1}; t) \\ &\quad + \delta t P_g \sum_{j=1}^{L-1} (L-1-r) \rho_2^{L-1}(A_j, A_{j+r}; t) + \mathcal{O}(\delta t^2). \end{aligned} \quad (4.7)$$

The LHS of Eq. (4.7) is the sum of pairwise density functions for lattice sites a distance r apart on a domain of length L at $[t + \delta t)$. The RHS of Eq. (4.7) corresponds to: (i) the sum of pairwise density functions for lattice sites a distance r apart on a domain of length L at time t multiplied by the probability that no growth event occurs in $[t, t + \delta t)$; (ii) the sum of pairwise density functions for lattice sites a distance $r - 1$ apart on a domain of length $L - 1$ multiplied by the probability that a domain growth event occurs in the interval between these sites (i.e. in one of $r - 1$ sites), thereby moving two lattice sites to a distance r apart on a domain of length L at time $[t + \delta t)$; (iii) the sum of pairwise density functions for lattice sites a distance r apart on a domain of length $L - 1$ multiplied by the probability that a growth event occurs in one of the $L - 1 - r$ sites in which the distance r between the two lattice sites is not affected, therefore moving two lattice sites a distance r apart on a domain of length $L - 1$ into a domain of length L at $[t + \delta t)$.

To simplify Eq. (4.7) we can rewrite $\rho_2^L(A_j, A_{j+r}; t)$ as $\rho_2^L(r; t)$, as in Eq. (4.3). Equation (4.7) then becomes

$$\begin{aligned}
\sum_{j=1}^L \rho_2^L(r; t + \delta t) &= (1 - \delta t P_g L) \sum_{j=1}^L \rho_2^L(r; t) \\
&\quad + \delta t P_g \sum_{j=1}^{L-1} (r - 1) \rho_2^{L-1}(r - 1; t) \\
&\quad + \delta t P_g \sum_{j=1}^{L-1} (L - 1 - r) \rho_2^{L-1}(r; t) + \mathcal{O}(\delta t^2). \tag{4.8}
\end{aligned}$$

Equation (4.8) can be simplified and rewritten as

$$\begin{aligned}
L\rho_2^L(r; t + \delta t) &= (1 - \delta t P_g L) L\rho_2^L(r; t) \\
&\quad + \delta t P_g (L - 1)(r - 1) \rho_2^{L-1}(r - 1; t) \\
&\quad + \delta t P_g (L - 1)(L - 1 - r) \rho_2^{L-1}(r; t) + \mathcal{O}(\delta t^2).
\end{aligned} \tag{4.9}$$

Finally, if we rearrange Eq. (4.9) and take the limit as $\delta t \rightarrow 0$ we arrive at

$$\begin{aligned}
\frac{d\rho_2^L(r; t)}{dt} &= -P_g L \rho_2^L(r; t) \\
&\quad + P_g \left(\frac{L - 1}{L} \right) (r - 1) \rho_2^{L-1}(r - 1; t) \\
&\quad + P_g \left(\frac{L - 1}{L} \right) (L - 1 - r) \rho_2^{L-1}(r; t).
\end{aligned} \tag{4.10}$$

Equation (4.10) describes the evolution of all pairwise density functions for distances $r = 1, \dots, L - 1$ on a growing domain. Similar to Eq. (4.6), Eq. (4.10) constitutes a system of coupled ODEs for the evolution of the pairwise density functions at a distance r for each domain length L .

4.2.3 Agent motility and proliferation

We now demonstrate how to include the effects of agent motility and proliferation in Eqs. (4.6) and (4.10). The addition of agent proliferation will introduce three-point density terms into Eq. (4.10), that is, the probability of three sites having a given occupancy. Similarly, an expression for three-point distribution terms would include four-point distribution terms and so on. Therefore, it is necessary to close the system of equations. To do so we will employ the Kirkwood superposition approximation (KSA) in due course, which has been used successfully in scenarios in which the domain does not grow [18, 21, 22]. It is important to acknowledge that other closure methods exist, such as those that fall under the Ornstein-Zernike category, or

the BBGKY hierarchy, of which the KSA is a member [91]. Both approaches have advantages and disadvantages, and their success is often evaluated in comparison with stochastic simulations, as is the case in this chapter for the KSA. How to choose the most appropriate closure relation is not well defined, and typically aesthetic concerns are prioritised. We implement the KSA as it has been implemented in models of cell proliferation and migration previously, such as those described in this chapter [18, 21, 22]. However, a more rigorous study into other types of closure that could be employed in these models is an important topic for further research.

Individual density functions

On a growing domain the individual density functions for motile and proliferative agents evolve according to

$$\begin{aligned}
\sum_{j=1}^L \rho_1^L(A_j; t + \delta t) &= (1 - \delta t P_g L) \sum_{j=1}^L \rho_1^L(A_j; t) + \delta t P_g \sum_{j=1}^{L-1} (L-1) \rho_1^{L-1}(A_j, t) \\
&\quad + \frac{\delta t P_p}{2} \sum_{j=1}^L \left[\rho_2^L(A_{j-1}, 0_j) + \rho_2^L(0_j, A_{j+1}) \right] \\
&\quad + \frac{\delta t P_m}{2} \sum_{j=1}^L \left[\rho_2^L(A_{j-1}, 0_j) - \rho_2^L(0_{j-1}, A_j) \right. \\
&\quad \left. + \rho_2^L(A_{j+1}, 0_j) - \rho_2^L(0_{j+1}, A_j) \right] + \mathcal{O}(\delta t^2). \tag{4.11}
\end{aligned}$$

As we assume translational invariance we have $\rho_2^L(A_{j-1}, 0_j) = \rho_2^L(0_{j-1}, A_j)$ and $c^L(t) = \rho_1^L(A_j, t) = \rho_1^L(A_k, t)$. Furthermore, if we recognise that

$$\rho_1^L(A_j; t) = \rho_2^L(A_j, A_{j+1}; t) + \rho_2^L(A_j, 0_{j+1}; t), \tag{4.12}$$

that is, by conservation arguments, an individual density function can be written as the sum of two pairwise density functions. Then we can use Eq. (4.12) and simplify

Eq. (4.11) to obtain

$$\begin{aligned} c^L(A_j; t + \delta t) &= (1 - \delta t P_g L) c^L(A_j; t) + \delta t P_g \left(\frac{L-1}{L} \right) (L-1) c^{L-1}(A_j, t) \\ &\quad + \delta t P_p [c^L(A_j, t) - \rho_2^L(1, t)] + \mathcal{O}(\delta t^2). \end{aligned} \quad (4.13)$$

If we rearrange Eq. (4.13) and take the limit as $\delta t \rightarrow 0$ we arrive at

$$\frac{dc^L(t)}{dt} = -P_g L c^L(t) + P_g \left(\frac{L-1}{L} \right) (L-1) c^{L-1}(t) + P_p [c^L(t) - \rho_2^L(1, t)]. \quad (4.14)$$

Equation (4.14) describes the evolution of the individual density functions for a motile, proliferative agent population on a growing domain. It can be seen that the addition of agent motility does not affect the evolution of individual density functions given by Eq. (4.14) (i.e. P_m is not present in Eq. (4.14)) [18].

Pairwise density functions

On a growing domain the pairwise density functions for motile and proliferative agents for the distance $r = 1$, such that the lattice sites are nearest neighbours, evolve

according to

$$\begin{aligned}
\sum_{j=1}^L \rho_2^L(A_j, A_{j+1}; t + \delta t) &= (1 - \delta t P_g L) \sum_{j=1}^L \rho_2^L(A_j, A_{j+1}; t) \\
&+ \delta t P_g \sum_{j=1}^{L-1} (L-2) \rho_2^{L-1}(A_j, A_{j+1}; t) \\
&+ \delta t \frac{P_m}{2} \sum_{j=1}^L \left[\rho_3^L(A_{j-1}, 0_j, A_{j+1}; t) - \rho_3^L(0_{j-1}, A_j, A_{j+1}; t) \right. \\
&\quad \left. + \rho_3^L(A_j, 0_{j+1}, A_{j+2}; t) - \rho_3^L(A_j, A_{j+1}, 0_{j+2}; t) \right] \\
&+ \delta t \frac{P_p}{2} \sum_{j=1}^L \left[\rho_3^L(A_{j-1}, 0_j, A_{j+1}; t) + \rho_2^L(0_j, A_{j+1}; t) \right. \\
&\quad \left. + \rho_3^L(A_j, 0_{j+1}, A_{j+2}; t) + \rho_2^L(A_j, 0_{j+1}; t) \right] + \mathcal{O}(\delta t^2). \quad (4.15)
\end{aligned}$$

We now assume translational invariance and recognise, by conservation arguments, that

$$\rho_2^L(A_j, A_{j+1}; t) = \rho_3^L(A_j, A_{j+1}, 0_{j+2}; t) + \rho_3^L(A_j, A_{j+1}, A_{j+2}; t). \quad (4.16)$$

Equation (4.16) demonstrates how a pairwise density function can be written as the sum of two three-point density functions. We can use Eqs. (4.12) and (4.16) in Eq.

(4.15) and obtain

$$\begin{aligned}
\sum_{j=1}^L \rho_2^L(A_j, A_{j+1}; t + \delta t) &= (1 - \delta t P_g L) \sum_{j=1}^L \rho_2^L(A_j, A_{j+1}; t) \\
&+ \delta t P_g \sum_{j=1}^{L-1} (L-2) \rho_2^{L-1}(A_j, A_{j+1}; t) \\
&+ \delta t P_m \sum_{j=1}^L [\rho_2^L(A_{j-1}, A_{j+1}; t) - \rho_2^L(A_j, A_{j+1}; t)] \\
&+ \delta t P_p \sum_{j=1}^L [\rho_3^L(A_{j-1}, 0_j, A_{j+1}; t) + c^L(A_j; t) - \rho_2^L(A_j, A_{j+1}; t)] \\
&+ \mathcal{O}(\delta t^2).
\end{aligned} \tag{4.17}$$

We now implement the KSA [18], which allows us to write three-point density functions in terms of pairwise density functions, and so close our system of equations Eqs. (4.14) and (4.17). The KSA for three-point density functions is

$$\rho_3^L(A_{j-1}, 0_j, A_{j+1}; t) = \frac{\rho_2^L(A_{j-1}, 0_j; t) \rho_2^L(0_j, A_{j+1}; t) \rho_2^L(A_{j-1}, A_{j+1}; t)}{\rho_1^L(A_{j-1}; t) \rho_1^L(0_j; t) \rho_1^L(A_{j+1}; t)}. \tag{4.18}$$

If we substitute Eq. (4.18) into Eq. (4.17) we obtain

$$\begin{aligned}
\sum_{j=1}^L \rho_2^L(A_j, A_{j+1}; t + \delta t) &= (1 - \delta t P_g L) \sum_{j=1}^L \rho_2^L(A_j, A_{j+1}; t) \\
&+ \delta t P_g \sum_{j=1}^{L-1} (L-2) \rho_2^{L-1}(A_j, A_{j+1}; t) \\
&+ \delta t P_m \sum_{j=1}^L [\rho_2^L(A_{j-1}, A_{j+1}; t) - \rho_2^L(A_j, A_{j+1}; t)] \\
&+ \delta t P_p \sum_{j=1}^L \left[\frac{\rho_2^L(A_{j-1}, 0_j; t) \rho_2^L(0_j, A_{j+1}; t) \rho_2^L(A_{j-1}, A_{j+1}; t)}{\rho_1^L(A_{j-1}; t) \rho_1^L(0_j; t) \rho_1^L(A_{j+1}; t)} \right. \\
&\quad \left. + c^L(A_j; t) - \rho_2^L(A_j, A_{j+1}; t) \right] + \mathcal{O}(\delta t^2).
\end{aligned} \tag{4.19}$$

Eq. (4.19) can be simplified to obtain

$$\begin{aligned}
\rho_2^L(1; t + \delta t) &= (1 - \delta t P_g L) \rho_2^L(1; t) + \delta t P_g \left(\frac{L-1}{L} \right) (L-2) \rho_2^{L-1}(1; t) \\
&\quad + \delta t P_m [\rho_2^L(2; t) - \rho_2^L(1; t)] \\
&\quad + \delta t P_p \left[\frac{(c^L(A_j; t) - \rho_2^L(1; t))^2 \rho_2^L(2; t)}{c^L(t)^2 (1 - c^L(t))} + c^L(A_j; t) - \rho_2^L(A_j, A_{j+1}; t) \right] \\
&\quad + \mathcal{O}(\delta t^2).
\end{aligned} \tag{4.20}$$

If we take the limit as $\delta t \rightarrow 0$ Eq. (4.20) becomes

$$\begin{aligned}
\frac{d\rho_2^L(1; t)}{dt} &= -P_g L \rho_2^L(1; t) + P_g \left(\frac{L-1}{L} \right) (L-2) \rho_2^{L-1}(1; t) \\
&\quad + P_m [\rho_2^L(2; t) - \rho_2^L(1; t)] \\
&\quad + P_p \left[\frac{(c^L(A_j; t) - \rho_2^L(1; t))^2 \rho_2^L(2; t)}{c^L(t)^2 (1 - c^L(t))} + c^L(A_j; t) - \rho_2^L(A_j, A_{j+1}; t) \right].
\end{aligned} \tag{4.21}$$

Equation (4.21) describes how the pairwise density function for a distance of $r = 1$ (lattice sites that are nearest neighbours) for motile and proliferative agents evolves on an exponentially growing domain.

For distances $1 < r < L$ the pairwise density functions for motile and proliferative

agents evolve according to

$$\begin{aligned}
\sum_{j=1}^L \rho_2^L(A_j, A_{j+r}; t + \delta t) &= (1 - \delta t P_g L) \sum_{j=1}^L \rho_2^L(A_j, A_{j+r}; t) \\
&+ \delta t P_g \sum_{j=1}^{L-1} (r-1) \rho_2^{L-1}(A_j, A_{j+r-1}; t) \\
&+ \delta t P_g \sum_{j=1}^{L-1} (L-1-r) \rho_2^{L-1}(A_j, A_{j+r}; t) \\
&+ \delta t \frac{P_m}{2} \sum_{j=1}^L \left[\rho_3^L(A_{j-1}, 0_j, A_{j+r}; t) + \rho_3^L(0_j, A_{j+1}, A_{j+r}; t) \right. \\
&+ \rho_3^L(A_j, A_{j-r-1}, 0_{j+r}; t) + \rho_3^L(A_j, 0_{j+r}, A_{j+r+1}; t) \\
&- \rho_3^L(0_{j-1}, A_j, A_{j+r}; t) - \rho_3^L(A_j, 0_{j+1}, A_{j+r}; t) \\
&\left. - \rho_3^L(A_j, 0_{j-r-1}, A_{j+r}; t) - \rho_3^L(A_j, A_{j+r}, 0_{j+r+1}; t) \right] \\
&+ \delta t \frac{P_p}{2} \sum_{j=1}^L \left[\rho_3^L(A_{j-1}, 0_j, A_{j+r}; t) + \rho_3^L(0_j, A_{j+1}, A_{j+r}; t) \right. \\
&\left. + \rho_3^L(A_j, A_{j+r-1}, 0_{j+r}; t) + \rho_3^L(A_j, 0_{j+r}, A_{j+r+1}; t) \right] + \mathcal{O}(\delta t^2).
\end{aligned} \tag{4.22}$$

If we follow a similar procedure as that presented for Eq. (4.21), from Eq. (4.22) we

obtain

$$\begin{aligned}
\frac{d\rho_2^L(r; t)}{dt} = & -P_g L \rho_2^L(r; t) \\
& + P_g \left(\frac{L-1}{L} \right) (r-1) \rho_2^{L-1}(r-1; t) \\
& + P_g \left(\frac{L-1}{L} \right) (L-1-r) \rho_2^{L-1}(r; t) \\
& + P_m [\rho_2^L(r+1; t) - 2\rho_2^L(r; t) + \rho_2^L(r-1; t)] \\
& + \frac{P_p}{(c^L(t))^2(1-c^L(t))} [(c^L(t) - \rho_2^L(1; t))(\rho_2^L(r+1; t))(c^L(t) - \rho_2^L(r; t)) \\
& + (c^L(t) - \rho_2^L(1; t))(\rho_2^L(r-1; t))(c^L(t) - \rho_2^L(r; t))]. \tag{4.23}
\end{aligned}$$

As can be seen from Eqs. (4.21) and (4.23), the inclusion of agent proliferation and the use of the KSA means that individual density functions are now present in the equations for the evolution of the pairwise density functions, which is not the case without agent proliferation (Eq. (4.10)). Similarly, the inclusion of agent proliferation means the evolution of the individual density functions now depend on the pairwise density functions, as can be seen in Eq. (4.14). It is important to note that Eqs. (4.21) and (4.23) are not exact due to the use of the KSA [18].

4.3 Results

We initially present results for motile agents on a growing domain in which domain growth is exponential. We will present results including agent proliferation in due course. The state of lattice site j in the i th realisation of the one-dimensional IBM is described by variable σ_{j_i} (i.e. occupied by an agent (A) or unoccupied (0)). This means the normalised average agent density for the i th realisation in the one-

dimensional IBM is

$$d_i^L(t) = \frac{1}{L} \sum_{j=1}^L \mathbb{1}_A\{\sigma_{ji}\}. \quad (4.24)$$

Here $\mathbb{1}_A$ is the indicator function for the occupancy of a lattice site (i.e. 1 if an agent occupies lattice site j , and 0 if it does not). If we take the ensemble average of Eq. (4.24) over many repeats we obtain the average density. Given our initial conditions, the expected (average) probability of two specified sites a distance r apart being occupied can be calculated exactly as

$$\frac{\binom{\text{number of agents}}{2}}{\binom{\text{domain length}}{2}}, \quad (4.25)$$

that is, the number of agents on the domain choose two, divided by the number of lattice sites choose two. For an initial condition in which all lattice sites are occupied this is simply unity, and for an initial condition containing only one agent it is undefined (because at least two agents are required for a pairwise density). Expression (4.25) provides the means of calculating the initial values for Eqs. (4.21) and (4.23). The initial condition for the individual density functions, Eq. (4.14), is simply the initial density at the initial domain length L_0 , and zero for all other lengths (i.e. for $L > L_0$). That is

$$c^{L_0}(0) = \frac{\text{number of agents}}{L_0}, \quad (4.26)$$

and

$$c^L(0) = 0, \quad L > L_0. \quad (4.27)$$

We calculate the pairwise counts between agents from IBM simulations in the following way: at each time-point every distance between an agent and all other agents is recorded. For a one-dimensional domain with periodic boundaries there are two distances between a pair of agents, of which we record the shorter distance. This count, defined as $s(r, L, t)$, is then normalised so that

$$p(r, L, t) = \frac{s(r, L, t)}{L}. \quad (4.28)$$

To make comparisons between ensemble averages from our discrete model and numerical solutions of the system of ODEs straightforward we only solve the system of ODEs given by Eqs. (4.21) and (4.23) for distances

$$r \leq \left\lfloor \frac{L}{2} \right\rfloor, \quad (4.29)$$

where $\lfloor \cdot \rfloor$ represents the floor function. Implementing Eq. (4.29) makes the results directly comparable to much of the previously published work on density functions, whereby the shortest distance between two sites is taken as the distance between these two sites [44, 45, 88]. In cases where $\rho_2^{L-1}(r; t)$ does not exist due to the condition imposed by Eq. (4.29), for instance when the domain length is even, we set $\rho_2^{L-1}(r; t)$ to $\rho_2^{L-1}(r-1; t)$. This is simply the anti-clockwise distance between the same two sites. These two sites must have the same joint probability of occupancy independent of the direction taken to measure the distance between them. In cases where $\rho_2^L(r+1; t)$ does not exist due to the condition imposed by Eq. (4.29), it is set to $\rho_2^L(r-1; t)$ if the domain length L is even (again, by considering the symmetry of the domain), and $\rho_2^L(r; t)$ if the domain length L is odd.

Finally, given we are modelling domain growth it is necessary to truncate our state space, and so we truncate the derived ODE model at approximately three times the

expected final domain length. As the domain grows exponentially the expected final domain length is simply

$$L(t_f) = L_0 e^{P_g t_f}, \quad (4.30)$$

where t_f is the end time-point of the simulation. The truncation length is represented by L_{trunc} in the figures. Small alterations to L_{trunc} had no noticeable effect on the results presented here. We solve Eqs. (4.14), (4.21) and (4.23) using MATLAB's `ode15s`, with an absolute error tolerance of 10^{-12} .

4.3.1 Motile agents on a growing domain

In Fig. 4.2 the evolution of the individual density functions can be seen for two different simulations with different parameters. In Fig. 4.2 (a) results are shown for a domain with an initial length of ten lattice sites and ten agents. In Fig. 4.2 (b) results are shown for a domain with an initial length of thirty lattice sites and fifteen agents. We see an excellent agreement between the numerical solution of Eq. (4.14) and the ensemble averages from the IBM, Eq. (4.24), for all domain lengths for both simulations ((a) LPE = 0.00, (b) 0.00). This is to be expected as Eqs. (4.14) and (4.21), (4.23) are all exact when $P_p = 0$. The probability of a site being occupied when the domain is a given length increases and then decreases for all domain lengths (apart from the initial domain length L_0). This is because initially all domains are of length L_0 (see Eq. (4.27)).

In Fig. 4.3 the evolution of the pairwise densities, given by Eqs. (4.21) and (4.23), can be seen. In the top row of Fig. 4.3 the evolution for distances $r = 1, 6$ and 10 is displayed (panels (a)-(c) correspond to Fig. 4.2 (a)). We see an excellent agreement between the ensemble average of the IBM and the numerical solution of Eqs. (4.21) and (4.23) for all values of r as both time and domain length increases ((a) 0.00, (b)

0.02, (c) 0.15). In the bottom row of Fig. 4.3 the evolution for distances $r = 1, 16$ and 30 corresponding to Fig. 4.2 (b) is displayed. Again we see an excellent agreement for all values of r ((d) 0.00, (e) 0.01, (f) 0.16). It is important to note that the ensemble averages become increasingly noisy for larger values of r . This is because the number of samples in the ensemble average from the IBM decreases for each count, Eq. (4.28), as the variance in the domain length increases.

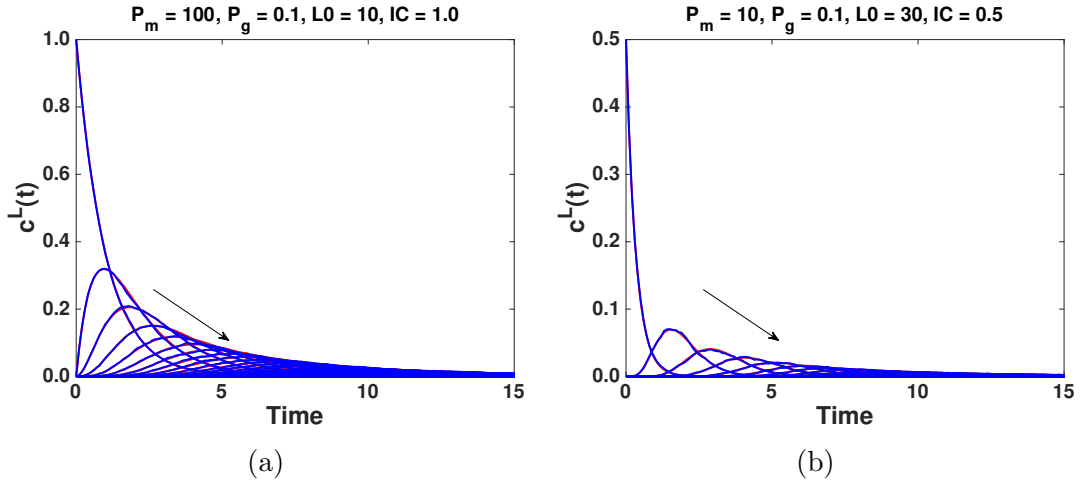


Figure 4.2: The evolution of the individual densities, $c^L(t)$, for two different sets of parameters and initial conditions. L_0 is the initial domain length, and IC is the initial density of agents. Numerical solution of Eq. (4.14) (red) and the corresponding ensemble averages from the IBM given by Eq. (4.24) (blue). It can be seen that the blue curves lie on top of the red curves, which indicates an excellent agreement between the numerical solution of Eq. (4.14) and the ensemble averages from the IBM. In (a) domain lengths $L_0, L_0 + 1, \dots, L_{trunc} = 135$ are plotted, in (b) domain lengths $L_0, L_0 + 5, \dots, L_{trunc} = 405$ are plotted. Ensemble averages are taken from 10000 repeats. Increasing domain length is down and to the right in all panels, as indicated by the arrows.

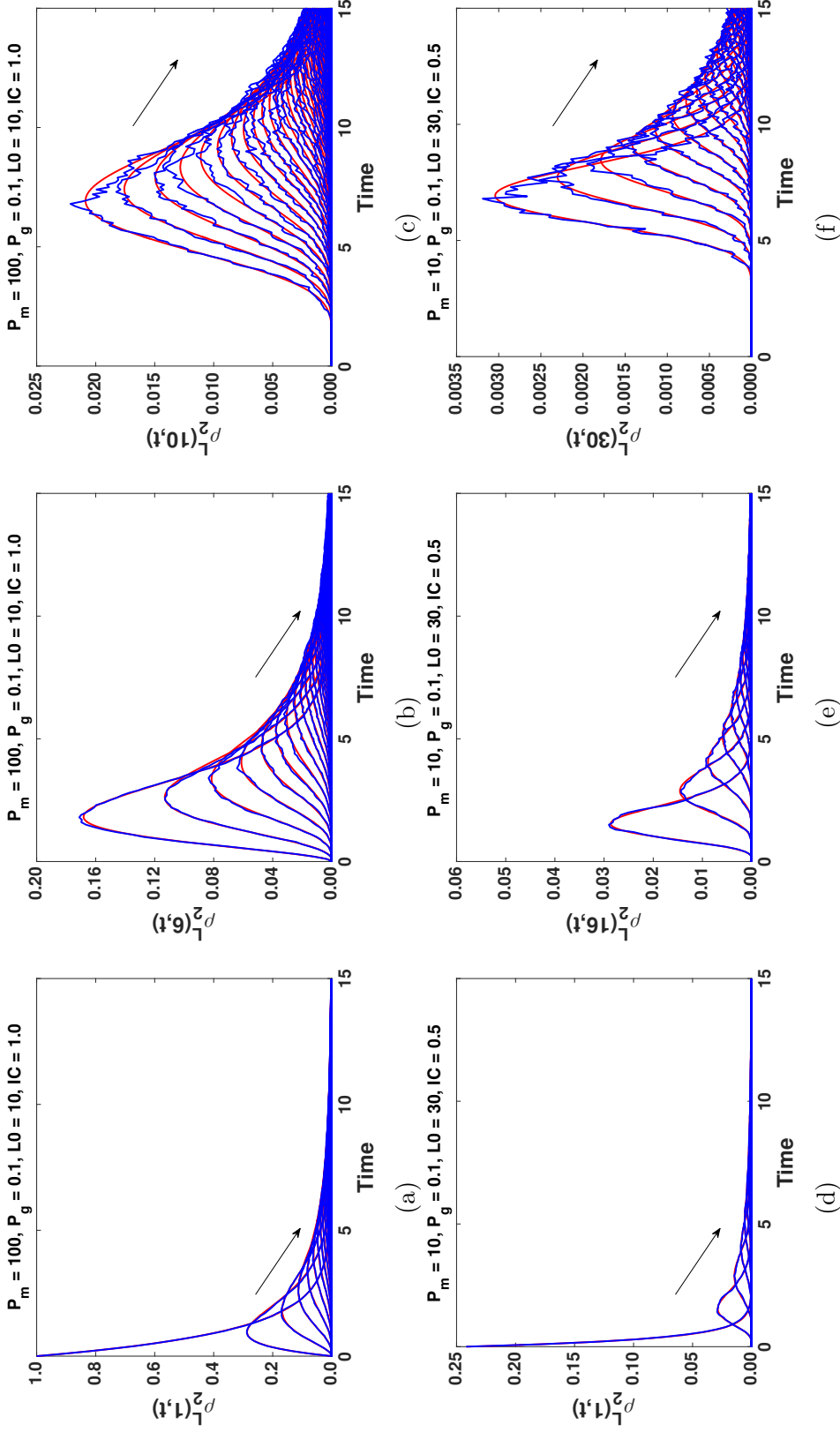


Figure 4.3: The evolution of the pairwise densities, $\rho_L^2(r, t)$, for two different sets of parameters and initial conditions (a)-(c) and (d)-(f) (panels (a)-(c) correspond to Fig. 4.2 (a) and (d)-(f) correspond to Fig. 4.2 (b)). Numerical solution of Eqs. (4.21) and (4.23) (red), and the corresponding ensemble averages from the IBM given by Eq. (4.28) (blue). Ensemble averages are taken from 10000 repeats. Top row: (a) The evolution of the pairwise density function at distance $r = 1$ for the domain length $L_0, L_0 + 1, \dots, L_{trunc} = 135$. Increasing domain length is down and to the right in all panels, as indicated by the arrows. (b) The evolution of the pairwise density function at distance $r = 6$. (c) The evolution of the pairwise density function at distance $r = 10$. Bottom row: (d) The evolution of the pairwise density function at distance $r = 1$ for the domain length $L_0, L_0 + 5, \dots, L_{trunc} = 405$. (e) The evolution of the pairwise density function at distance $r = 16$. (f) The evolution of the pairwise density function at distance $r = 30$.

4.3.2 Motile, proliferative agents on a growing domain

We now present results for motile, *proliferative* agents on a growing domain. As before the initial condition in the IBM is achieved by populating a certain number of sites uniformly at random. We initialise the individual density functions corresponding to the initial domain length in the same manner as Eq. (4.26). However, to avoid a singularity in Eqs. (4.21) and (4.23) we follow Johnston et al. [20] and initialise all other individual densities (for each domain length) with a small parameter $\epsilon = 10^{-20}$ [20]. That is,

$$c^L(0) = \epsilon, \quad L > L_0. \quad (4.31)$$

This small parameter was selected by iteratively reducing its size until the error tolerance threshold of the ODE numerical solver was compromised. This approach meant that the effect of introducing the small parameter ϵ into the equations was minimised.

In Fig. 4.4 the evolution of the individual density functions can be seen for two simulations with different parameters. In Fig. 4.4 (a) we see the evolution of the individual densities for a domain with an initial length of ten lattice sites and two agents. As $P_p \neq 0$ Eqs. (4.21) and (4.23) are not exact due to the use of the KSA. Despite this, we see that a good agreement is achieved between the numerical solution of Eq. (4.14) and the ensemble averages from the IBM for a low rate of agent proliferation (0.11). However, as we increase agent proliferation, relative to agent motility, the approximation by the numerical solution of Eq. (4.14) of the ensemble averages from the IBM becomes less accurate, as can be seen in Fig. 4.4 (b) (0.68). This is because as agent proliferation is increased spatial correlations become more prevalent in the IBM [18, 21]. The error associated with the KSA also grows monotonically as time evolves, as can be seen in Fig. 4.4 (b) where the evolution of the individual

density functions for a domain with an initial length of $L_0 = 30$ lattice sites and fifteen agents is displayed. This error arises for two reasons: i) the use of the KSA; and ii) the initialisation of the one-point distributions with a small parameter ϵ to avoid a singularity.

In Fig. 4.5 the evolution of the pairwise densities can be seen. In the top row of Fig. 4.5, which corresponds to Fig. 4.4 (a), the evolution for distances $r = 1, 6$ and 10 is displayed. We see a good agreement between numerical solution of Eqs. (4.21) and (4.23) and the ensemble averages from the IBM for $r = 1, 6$ and 10 ((a) 0.15, (b) 0.20, (c) 0.29). In the bottom row of Fig. 4.5, which corresponds to Fig. 4.4 (b), the evolution for distances $r = 1, 16$ and 30 is displayed. We initially see a good agreement for $r = 1$ and 16 ((d) 0.24, (e) 0.33, (f) 0.25). This agreement begins to break down as time evolves for the two reasons discussed above.

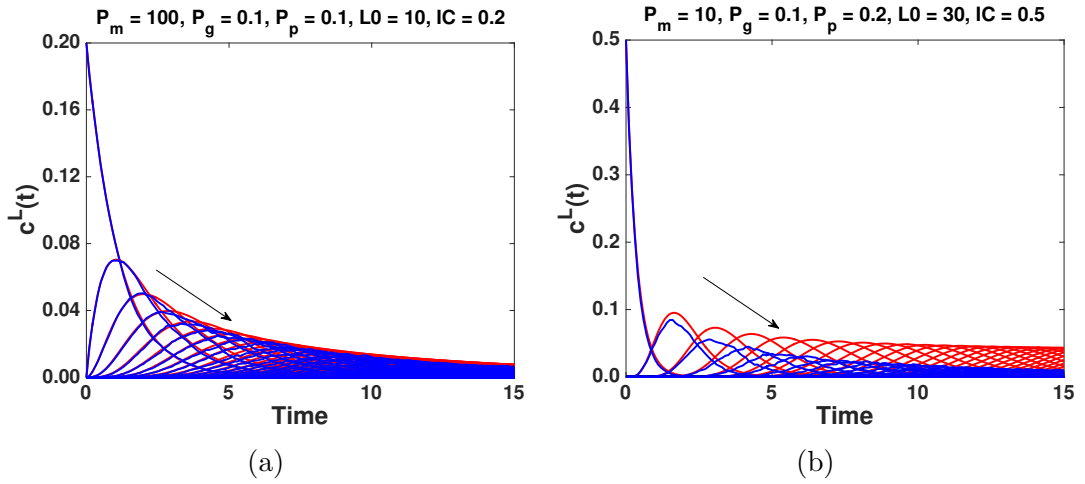


Figure 4.4: The evolution of the individual densities, $c^L(t)$, with agent proliferation for two different sets of parameters and initial conditions. Numerical solution of Eq. (4.14) (red) and the corresponding ensemble averages from the IBM given by Eq. (4.24) (blue). In (a) domain lengths $L_0, L_0 + 1, \dots, L_{trunc} = 135$ are plotted, in (b) domain lengths $L_0, L_0 + 5, \dots, L_{trunc} = 405$ are plotted. Ensemble averages are taken from 10000 repeats. Increasing domain length is down and to the right in all panels as indicated by the arrows.

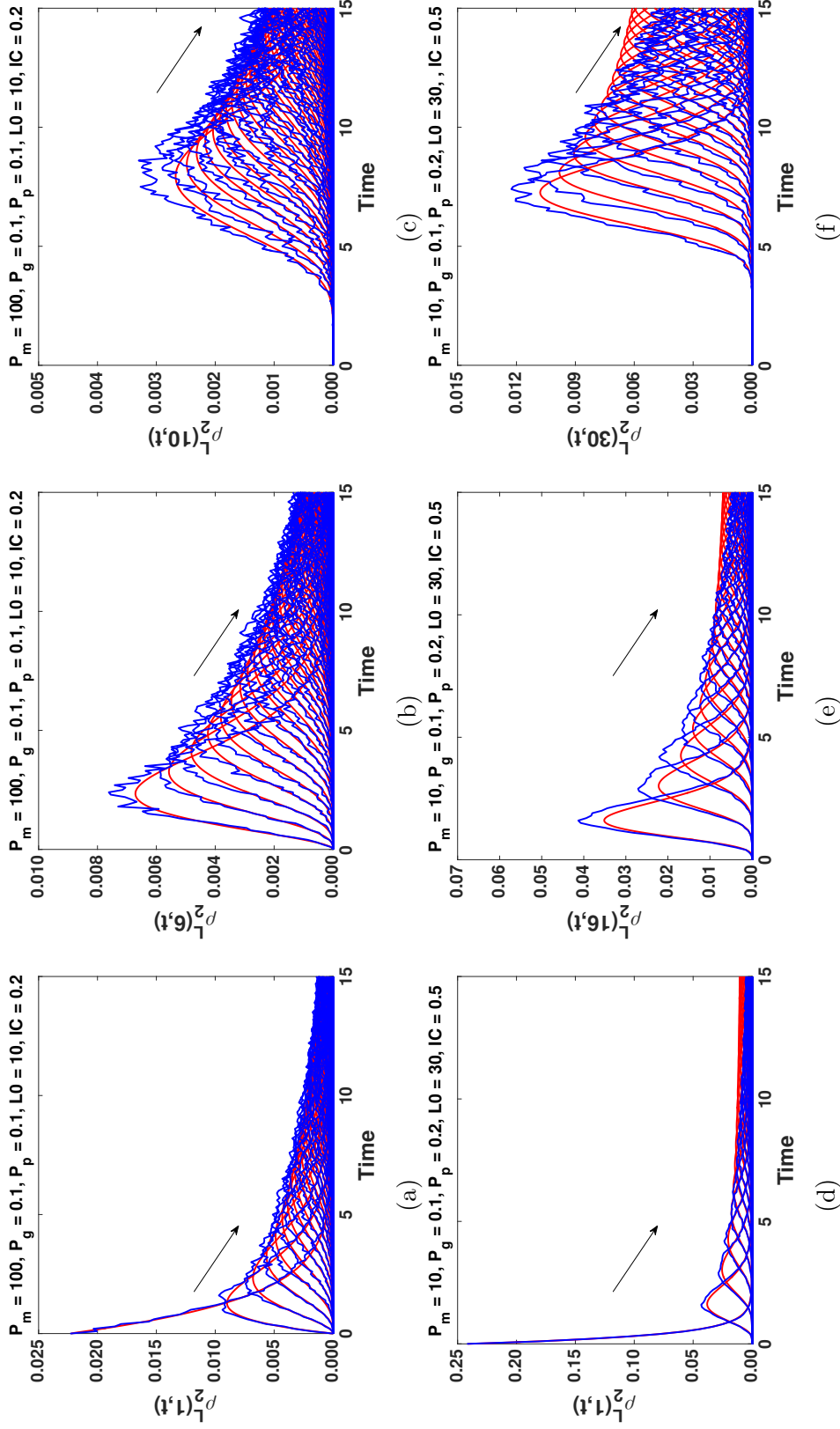


Figure 4.5: The evolution of the pairwise densities, $\rho_L^z(r, t)$, with agent proliferation for two different sets of parameters and initial conditions (a)-(c) and (d)-(f) (panels (a)-(c) correspond to Fig. 4.4 (a) and (d)-(f) correspond to Fig. 4.2 (b)). Comparison of the numerical solution of Eqs. (4.21) and (4.23) (red) and the corresponding ensemble averages from the IBM given by Eq. (4.28) (blue). Ensemble averages are taken from 10000 repeats. Top row: (a) The evolution of the pairwise density function at distance $r = 1$ for the domain length $L_0, L_0 + 1, \dots, L_{trunc} = 135$, (b) at distance $r = 6$, and (c) at distance $r = 10$. Bottom row: (d) The evolution of the pairwise density function at distance $r = 1$ for the domain length $L_0, L_0 + 5, \dots, L_{trunc} = 405$, (e) at distance $r = 16$, and (f) at distance $r = 30$. Increasing domain length is down and to the right in all panels as indicated by the arrows.

4.4 Integrating domain growth into correlation functions

Our current framework, Eqs. (4.14), (4.21) and (4.23) results in a system of individual density functions given by Eq. (4.14) which, while mathematically sound, is not easy to relate to biological systems and the experimental data associated with them, whereby the evolution of a macroscopic density on a growing domain is generally only measured with respect to time [2, 3, 10, 36]. Therefore, to mitigate these issues, we present a method to reduce the ODE system describing the evolution of the individual density functions, Eq. (4.14), into a single equation describing the evolution of the macroscopic density on a growing domain. To do this we introduce ‘correlation functions’, and integrate the effect of domain growth into pre-existing work on the calculation of spatial correlations between agents [18–23]. To derive this simplified model we will assume all domains are of length L , where L is the mean length of the domain in the stochastic model.

To reduce the ODE system describing the evolution of the individual density functions, Eq. (4.14), into a single equation describing the evolution of the macroscopic density on a growing domain we use the following heuristic approximations

$$c^L(t) \approx \frac{(L-1)}{L} c^{L-1}(t), \quad (4.32)$$

for the individual density functions, and

$$\rho_2^L(r; t) \approx \frac{(L-1)}{L} \rho_2^{L-1}(r; t), \quad (4.33)$$

for the pairwise density functions. This is a previously published approximation [17] and implies that domain growth ‘dilutes’ both the agent density and the pairwise densities. To demonstrate the validity of Eqs. (4.32) and (4.33) we measure the

relative error

$$\text{RE} = \frac{\frac{(L-1)}{L}c^{L-1}(t) - c^L(t)}{c^L(t)}, \quad (4.34)$$

of approximation Eq. (4.32).

In Fig. 4.6 the error caused by approximation Eq. (4.32) for the individual density functions for two different simulations can be seen. The relative error is initially high, due to all of the domains initially being the same length (Eqs. (4.26) and (4.31)), but decreases towards zero as the expected domain length increases.

If we substitute approximation Eq. (4.33) into Eqs. (4.21) and (4.23), for $r = 1$ we obtain

$$\begin{aligned} \frac{d\rho_2(1;t)}{dt} &= -2P_g\rho_2(1;t) \\ &+ P_m[\rho_2(2;t) - \rho_2(1;t)] \\ &+ \frac{P_p}{(c(t))^2(1-c(t))} [(c(t) - \rho_2(1;t))^2(\rho_2(2;t))] \\ &+ P_p[c(t) - \rho_2(1;t)], \end{aligned} \quad (4.35)$$

and for $1 < r < L$

$$\begin{aligned} \frac{d\rho_2(r;t)}{dt} &= P_g(r-1)\rho_2(r-1;t) - P_g(r+1)\rho_2(r;t) \\ &+ P_m[\rho_2(r+1;t) - 2\rho_2(r;t) + \rho_2(r-1;t)] \\ &+ \frac{P_p}{(c(t))^2(1-c(t))} [(c(t) - \rho_2(1;t))(\rho_2(r+1;t))(c(t) - \rho_2(r;t)) \\ &+ (c(t) - \rho_2(1;t))(\rho_2(r-1;t))(c(t) - \rho_2(r;t))], \end{aligned} \quad (4.36)$$

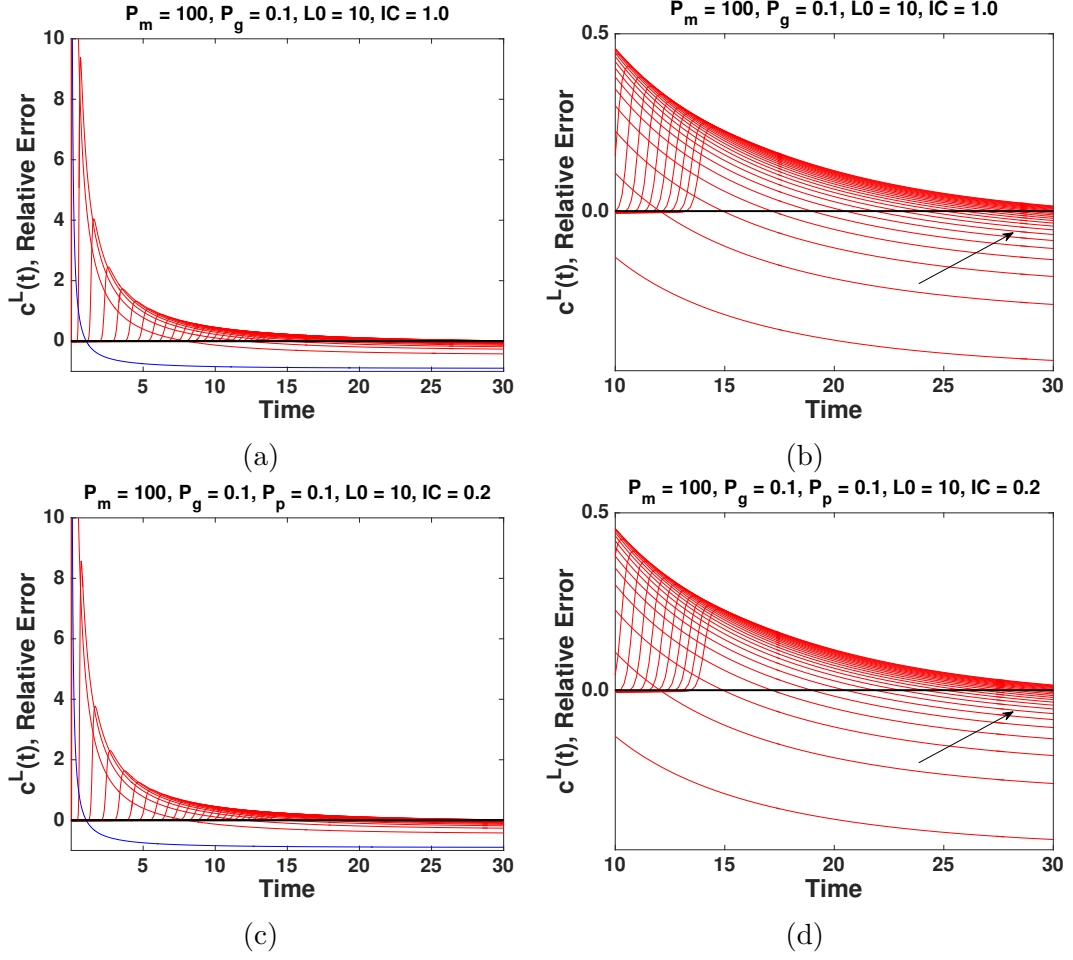


Figure 4.6: The relative error, calculated with Eq. (4.34), in the individual densities due to approximations Eq. (4.32) and (4.33) for different domain lengths. The domain length increases in distances of 10Δ starting from the initial domain length L_0 (indicated by the blue line). In panels (a) and (c) increasing domain length is up and to the right of the blue line. Panels (b) and (d) are magnified aspects of panels (a) and (c), and increasing domain length is indicated by the black arrows. We can see that the relative error approaches zero as the domain length increases for long times. The black line indicates a relative error of zero.

whereby all pairwise density functions are now for the same domain length, L , and so the superscript indicating domain length has been dropped. Similarly, when approximation Eq. (4.32) is applied to Eq. (4.14) it becomes

$$\frac{dc(t)}{dt} = -P_g c(t) + P_p [c(t) - \rho_2(1; t)]. \quad (4.37)$$

This is a single equation, albeit coupled to a pairwise density function, describing the evolution of the macroscopic density on a growing domain that includes the effect of spatial correlations between agents created by agent proliferation.

4.4.1 Correlation function

We finally introduce correlation functions. We define the correlation function [86, 87] as

$$F(r, t) := \frac{\rho_2(r, t)}{c(t)^2}. \quad (4.38)$$

The correlation function is a useful variable to improve the visualisation of results. This is because $F(r, t) \equiv 1$ indicates that the occupancy of two lattice sites a given distance apart is independent. It has therefore been used in the study of spatial correlations in models of cell motility and proliferation [18–23, 44].

Eq. (4.38) can now be substituted into Eqs. (4.35) and (4.36) to describe the evolution of spatial correlations between agents on a growing domain. This results in Eq. (4.35) becoming

$$\begin{aligned} \frac{dF(1, t)}{dt} = & P_m[F(2, t) - F(1, t)] - 2P_p F(1, t) [1 - c(t)F(1, t)] \\ & + \frac{P_p}{c(t)} [1 - c(t)F(1, t)] + \frac{P_p}{1 - c(t)} F(1, t) [1 - c(t)F(1, t)]^2 \\ & - 2P_g F(1, t), \end{aligned} \quad (4.39)$$

and Eq. (4.36) becoming

$$\begin{aligned} \frac{dF(r, t)}{dt} = & P_m[F(r-1, t) - 2F(r, t) + F(r+1, t)] - 2P_p F(r, t)[1 - c(t)F(1, t)] \\ & + \frac{P_p}{1 - c(t)}[1 - c(t)F(1, t)][1 - c(t)F(r, t)][F(r-1, t) + F(r+1, t)] \\ & - P_g(r+1)F(r, t) + P_g(r-1)F(r-1, t). \end{aligned} \quad (4.40)$$

If we rewrite Eq. (4.37) in terms of Eq. (4.38) we obtain

$$\frac{dc(t)}{dt} = -P_g c(t) + P_p c(t)[1 - F(1, t)c(t)]. \quad (4.41)$$

Eq. (4.41) is a hybrid (corrected) mean-field correlation ODE model (referred to from now on as the *corrected* ODE model), whereby the effect of agent proliferation on spatial correlations between agents is included in an equation describing the evolution of the agent density on a growing domain.

4.4.2 Results for correlation ordinary differential equation model

We now compare simulations for motile, proliferative agents on a growing domain with the numerical solutions of Eqs. (4.39)-(4.41). As before the initial condition in the IBM is achieved by populating a certain number of sites uniformly at random. The initial length of the domain in all simulations is one hundred lattice sites. We also compare ensemble averages from the IBM with a (*uncorrected*) MFA

$$\frac{dc(t)}{dt} = P_p c(t)[1 - c(t)] - P_g c(t), \quad (4.42)$$

in which it has been assumed that $F(1, t) = 1$, i.e. the effect of spatial correlations is insignificant. This means we compare ensemble averages from the IBM with both the MFA (Eq. (4.42)) and the corrected ODE model (Eq. (4.41)).

To solve Eqs. (4.39)-(4.41) numerically we use an implicit Euler scheme. For all numerical solutions presented here $\delta t = 0.1$, and our correlation truncation value is $r = 50$. By this it is meant we set all correlation functions at and beyond this distance equal to one. That is,

$$F(r, t) = 1, \quad \forall r \geq 50. \quad (4.43)$$

Equation (4.43) stipulates that at long distances between sites, relative to the domain size, we would not expect significant correlations. Similarly, our initial condition is that all distances are initially uncorrelated, that is

$$F(r, 0) = 1. \quad (4.44)$$

Initially, we provide results from simulations that involve only domain growth i.e. no agent proliferation or movement. In Fig. 4.7 (b) we see the effects of exponential domain growth on spatial correlations between agents in the IBM. Domain growth does not create spatial correlations between agents that are initially distributed uniformly at random (when averaged over many repeats). This is indicated by the correlation function remaining at a value of one, and is also captured by the numerical solutions of Eqs. (4.39) and (4.40), as can be seen in Fig. 4.7 (c) and (f). As the position of the agents remains uncorrelated throughout the course of the simulation, the evolution of the macroscopic density in the IBM, mean-field and corrected ODE models all match exactly as can be seen in Fig. 4.7 (a) and (d) ((a) 0.00, (d) 0.00).

We now turn our attention to simulations that include agent proliferation and motility. As we increase P_p relative to P_g the approximation of the spatial correlations in the IBM by the corrected ODE model becomes less accurate, as can be seen by comparison of Fig. 4.8 (b) and (c), and Fig. 4.8 (e) and (f). This is due to both

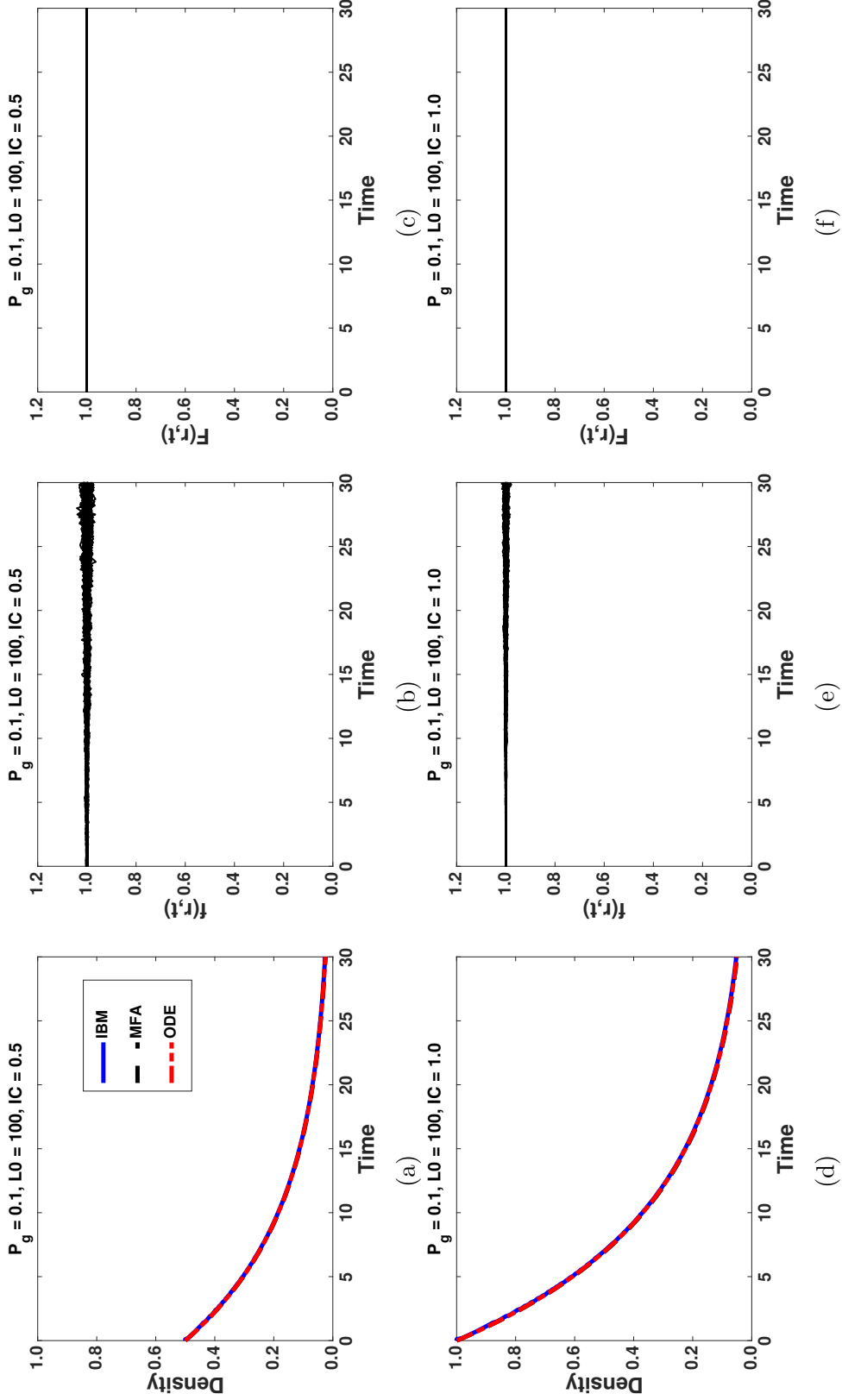


Figure 4.7: Panels (a) and (d) demonstrate the evolution of the macroscopic agent density as given by Eqs. (4.24), (4.41) and (4.42). Panels (b) and (e) demonstrate the evolution of the pairwise correlations in the IBM. The correlation function is plotted for increasing distance from Δ to 50Δ in steps of Δ in panels (b) and (e). Panels (c) and (f) demonstrate the pairwise correlations from the numerical solution of Eqs. (4.39) and (4.40). The distance, r , increases from Δ to 50Δ in steps of Δ . The ensemble averages are taken from 100 repeats of the IBM.

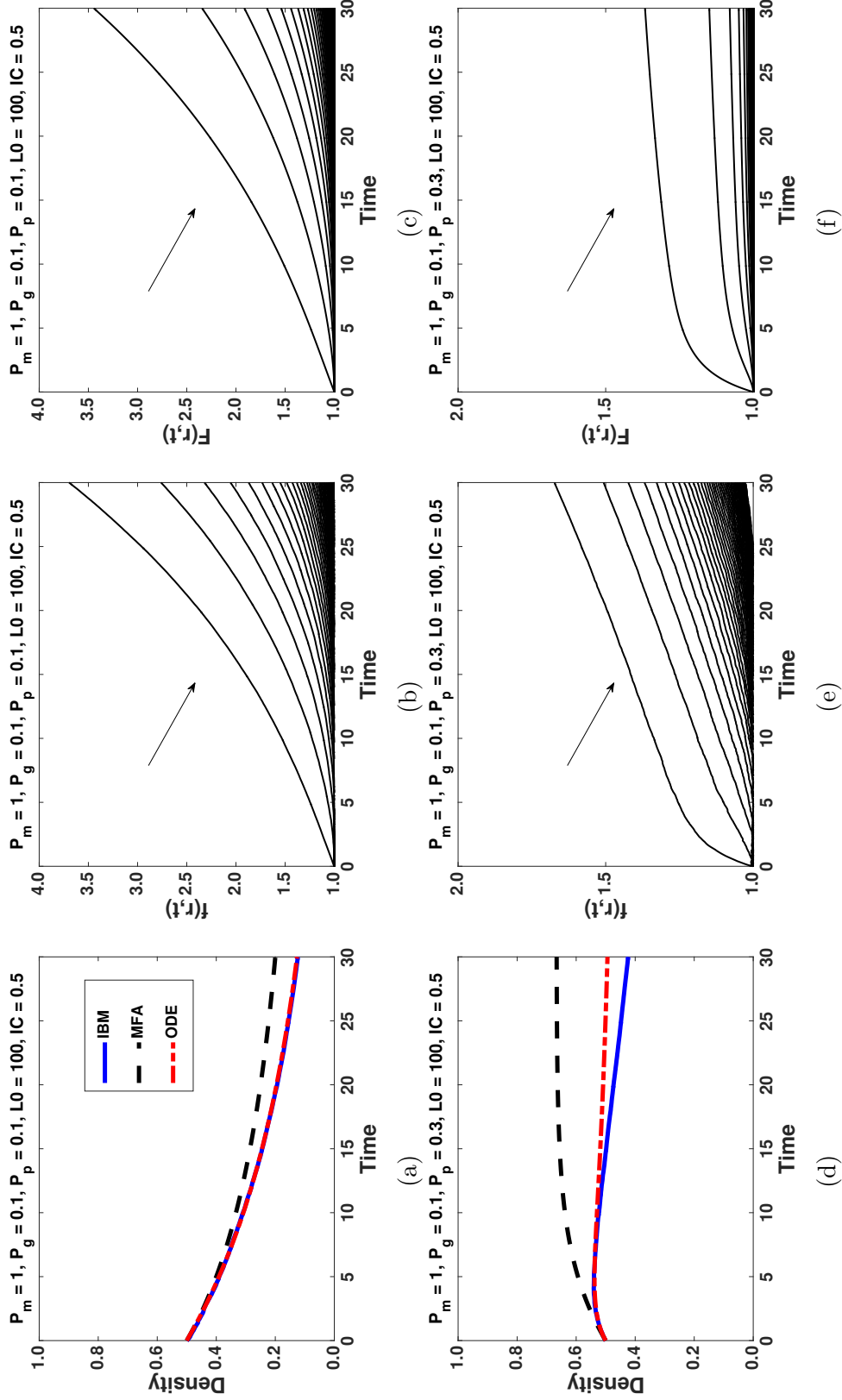


Figure 4.8: Panels (a) and (d) demonstrate the evolution of the macroscopic agent density with agent proliferation as given by Eqs. (4.24), (4.41) and (4.42). Panels (b) and (e) demonstrate the evolution of the pairwise correlations in the IBM. The correlation function is plotted for increasing distance from Δ to 50Δ in steps of Δ in the direction indicated by the arrow in panels (b), (c), (e) and (f). Panels (c) and (f) demonstrate the pairwise correlations at different distances from the numerical solution of Eqs. (4.39) and (4.40). The distance, r , increases from Δ to 50Δ in steps of Δ . The ensemble averages are taken from 100 repeats of the IBM.

the KSA approximation and the use of the closure Eqs. (4.32) and (4.33), and leads to the approximation of the evolution of the agent density in the IBM becoming less accurate as can be seen in Fig. 4.8 (a) and (d) ((a) 0.00, (d) 0.14). However, in both scenarios the corrected ODE model more accurately approximates the evolution of the agent density in the IBM than the standard mean-field model given by Eq. (4.42).

4.5 Summary

In this chapter we have presented methods to include the effects of domain growth on the evolution of the individual and pairwise agent density functions in a continuum approximation of a one-dimensional IBM. Spatial correlations have been shown to play an important role in cell migration and tumour development, with both of these processes often associated with domain growth [1–3, 10, 89]. For systems without agent proliferation, comparisons between averaged IBM results and our corrected ODE model are excellent, as can be seen in Figs. 4.2 and 4.3. These results show that uniform domain growth, somewhat intuitively, reduces spatial correlations between agents. The introduction of agent proliferation into the correlation ODE model leads to some inaccuracies. However, there is still good agreement between the averaged IBM results and our corrected ODE model in parameter regimes with low proliferation rates, as can be seen in Figs. 4.4 and 4.5. These inaccuracies are largely due to the use of the KSA approximation, and to a lesser extent the small parameter ϵ introduced into the model [18, 20].

We then integrated our correlation ODE framework into the existing literature on calculating the evolution of pairwise spatial correlations, namely, correcting MFAs for the evolution of the agent density in the IBM [18–23]. This required heuristic approximations [18], Eqs. (4.32) and (4.33), that allowed us to derive a single equation describing the evolution of the macroscopic density in the IBM on a growing domain.

This meant we could place the results naturally alongside the measurement of cell density on growing domains in experimental data [1–3, 36, 44, 45, 88].

The results presented here are extendable to non-periodic boundary conditions and systems where we do not assume translational invariance. In addition, the framework presented could also be extended to higher-dimensional models of cell motility and proliferation. It is also important to reiterate that there are alternative approaches to moment closure than those presented in this chapter. Both closures we implement (KSA and domain length closure) could possibly be replaced by closures that would result in more accurate correlation ODE models. Different moment closures may also lead to less cumbersome equations than those presented here. This could save time and allow a more rigorous analysis of these equations' properties. Therefore, the revisiting of the closure methods employed in deriving the equations presented in this chapter is a topic of future research that could prove fruitful.

.....

In the next chapter we extend the methods presented in Chapter 4 to two-dimensional models of cell motility and proliferation. This is important as to realistically describe many cell migratory processes modelling them in two or three spatial dimensions is crucial.

Chapter 5

Modelling spatial correlations on a growing domain in two dimensions

Overview

In this chapter we present methods to include the effects of domain growth on the evolution of spatial correlations in a continuum approximation of a two-dimensional lattice-based model of cell motility and proliferation. We show that, depending on the way in which domain growth is implemented, different steady-state densities are predicted for an agent population. Furthermore, we demonstrate that the way in which domain growth is implemented can result in the evolution of the agent density depending on the size of the domain. Continuum approximations that ignore spatial correlations cannot capture these behaviours, while those that account for spatial correlations can.

The majority of the work in this chapter appears in the paper: R. J. H. Ross, R. E. Baker and C. A. Yates. “How domain growth is implemented determines the long term behaviour of a cell population through its effect on spatial correlations”. *Physical Review E*, 94(1):012408, 2016. R. J. H. Ross performed all the computational and mathematical modelling presented in this chapter.

5.1 Introduction

In this chapter we examine how domain growth affects the evolution of spatial correlations between agents in IBMs in a two-dimensional model, and importantly, how the way in which domain growth is implemented affects spatial correlations differently. To do so we use an agent-based, discrete random-walk model on a two-dimensional square lattice with volume-exclusion. We reasoned that a standard MFA would not be able to capture the behaviours exhibited by the IBM, and that this standard MFA would require correction by the inclusion of spatial correlations in the form of a correlation ODE model to accurately approximate the IBM results. This last point has been shown previously in scenarios without growth [18].

The outline of this chapter is as follows. To begin with we introduce the IBM and growth mechanisms in Section 5.2. We then define our individual and pairwise density functions, and derive a system of ODEs describing the evolution of the individual and pairwise density functions with respect to time on a growing domain. To test the accuracy of these ODE models we compare them with ensemble averages for the evolution of the macroscopic density from the IBM in Section 5.3. We demonstrate that the standard MFA is not able to accurately capture the effects of domain growth in the IBM, whereas the ODE models that include the effect of spatial correlations can. In Section 5.4 we summarise the findings presented in this chapter.

5.2 Model

In this section we introduce the IBM and the domain growth mechanisms we employ throughout this chapter. We then derive equations describing the evolution of the individual and pairwise density functions in the IBM for both growth mechanisms.

5.2.1 IBM and domain growth mechanisms

The IBM we use in this chapter is similar to that used in Chapter 2. However, for clarity we briefly define the IBM again. The IBM is simulated on a two-dimensional square lattice with lattice spacing Δ [46] and size $N_x(t)$ by $N_y(t)$, where $N_x(t)$ is the number of lattice sites in a row and $N_y(t)$ is the number of sites in a column. All simulations are performed with periodic boundary conditions, and throughout this chapter the lattice spacing, Δ , is equal to one. For notational convenience we shall now write $N_x(t)$ as N_x , and $N_y(t)$ as N_y .

In the IBM each agent is initially assigned to a lattice site, from which it can move or proliferate into an adjacent site. If an agent attempts to move into a site that is already occupied, the movement event is aborted. Similarly, if an agent attempts to proliferate into a site that is already occupied, the proliferation event is aborted. Time is evolved continuously, in accordance with the Gillespie algorithm [90], such that movement, proliferation and growth events are modelled as exponentially distributed ‘reaction events’ in a Markov chain. Attempted agent movement or proliferation events occur with rates P_m or P_p per unit time, respectively. For example, $P_m \delta t$ is the probability of an agent attempting to move in the next infinitesimally small time interval δt . Throughout this chapter the initial agent distribution for all simulations of the IBM is achieved by populating lattice sites uniformly at random until the required density is achieved.

Both growth mechanisms we employ are stochastic growth mechanisms [24]: the insertion of new lattice sites into the domain occurs with positive rate constants P_{gx} and P_{gy} per unit time, for growth in the x (horizontal) and y (vertical) directions, respectively. That is, each lattice site undergoes growth events in the x direction with rate P_{gx} . This means the total rate of domain growth in the x direction is $P_{gx} N_x$. In

growth mechanism 1 (GM1) when a ‘growth event’ occurs along the x axis (horizontal axis in Fig. 5.1), one new column of sites is added at a position selected uniformly at random. In growth mechanism 2 (GM2) when a ‘growth event’ occurs along the x axis (horizontal axis in Fig. 5.1), for each row, one new site is added in a column which is selected uniformly at random¹. Importantly, when a growth event occurs, the site selected for division is moved one spacing in the positive horizontal/vertical direction along with its contents (i.e. an agent or no agent, an agent is symbolised by a black circle in Fig. 5.1). The new lattice site is empty, and the contents of all other lattice sites remain unaffected. Growth in the y direction is implemented in an analogous manner to the x direction for both growth mechanisms. We chose these growth mechanisms as they are significantly different to each other, and both may have biological relevance [92–94]. For example, biological growth similar to GM1 can occur when adjacent cells become synchronous in their cell cycles (that is, the underlying cells that form the domain). Furthermore, both of these growth mechanisms can be used to implement any form of isotropic growth in our IBM, and are adaptable to three spatial dimensions [24]. Finally, it is important to stress that both growth mechanisms give rise to the same overall growth rate when implemented with the same growth rate parameters.

5.2.2 Individual density functions

We define the individual density functions, $\rho_1^{N_x \times N_y}(\mathbf{m}; t)$, as the probability that site $\mathbf{m}$ is occupied by an agent at time t on a domain of size $N_x \times N_y$, where $\mathbf{m}$ is the vector (i, j) , with i indexing the row number of a lattice site, and j indexing the column number of a lattice site. For instance, $(2, 3)$ would be the lattice site situated in the second row and the third column of the lattice.

¹This is the same domain growth mechanism as implemented in Chapters 2 and 3.

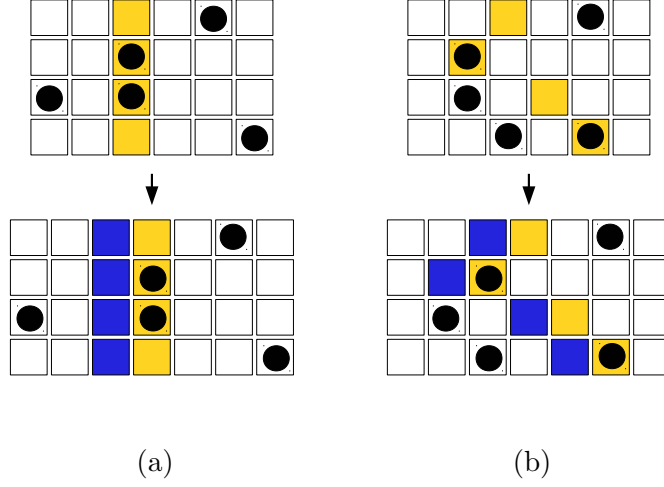


Figure 5.1: Before and after a growth event for both (a) GM1 and (b) GM2, in which the growth event is along the x -axis (horizontal direction) for a two-dimensional lattice. In each row the yellow site has been chosen to undergo a growth event. Following this the yellow site is moved to the right with its contents, for instance an agent (represented by a black cell). The blue sites are the new lattice sites and are always initially empty. The contents of all the other sites remain unaffected, although in some cases their neighbouring sites will change.

The following derivation for the individual density functions is the same for GM1 and GM2. The derivation also follows a similar pattern to the work presented in Chapter 4. The sum of the individual density functions on a domain of size $N_x \times N_y$ at $[t + \delta t)$ can be written in terms of the individual density functions at time t :

$$\begin{aligned}
\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \rho_1^{N_x \times N_y}(\mathbf{m}; t + \delta t) &= (1 - \delta t P_{g_x} N_x - \delta t P_{g_y} N_y) \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \rho_1^{N_x \times N_y}(\mathbf{m}; t) \\
&+ \delta t P_{g_x} \sum_{i=1}^{N_x-1} \sum_{j=1}^{N_y} (N_x - 1) \rho_1^{(N_x-1) \times N_y}(\mathbf{m}; t) \\
&+ \delta t P_{g_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y-1} (N_y - 1) \rho_1^{N_x \times (N_y-1)}(\mathbf{m}; t) + \mathcal{O}(\delta t^2).
\end{aligned} \tag{5.1}$$

The terms on the RHS of Eq. (5.1) correspond to the following events: i) the probability that no growth event occurs in $[t, t + \delta t)$; ii) the probability that a growth

event occurs in the horizontal (x) direction in $[t, t + \delta t)$; and iii) the probability that a growth event occurs in the vertical (y) direction in $[t, t + \delta t)$. Due to our initial conditions in the IBM we can assume translational invariance for the probability of an agent occupying a site in this derivation, by this we mean

$$\rho_1^{N_x \times N_y}(\mathbf{m}; t) = \rho_1^{N_x \times N_y}(\mathbf{n}; t), \quad \forall \mathbf{m}, \mathbf{n}, \quad (5.2)$$

where $\mathbf{n}$ indexes any other site on the domain².

Therefore we can rewrite Eq. (5.1) as

$$\begin{aligned} (N_x)(N_y)\rho_1^{N_x \times N_y}(\mathbf{m}; t + \delta t) &= (1 - \delta t P_{g_x} N_x - \delta t P_{g_y} N_y)(N_x)(N_y)\rho_1^{N_x \times N_y}(\mathbf{m}; t) \\ &\quad + \delta t P_{g_x} (N_x - 1)(N_y)(N_x - 1)\rho_1^{(N_x - 1) \times N_y}(\mathbf{m}; t) \\ &\quad + \delta t P_{g_y} (N_x)(N_y - 1)(N_y - 1)\rho_1^{N_x \times (N_y - 1)}(\mathbf{m}; t) + \mathcal{O}(\delta t^2). \end{aligned} \quad (5.3)$$

We can simplify Eq. (5.3) in the following manner

$$\begin{aligned} \rho_1^{N_x \times N_y}(\mathbf{m}; t + \delta t) &= (1 - \delta t P_{g_x} N_x - \delta t P_{g_y} N_y)\rho_1^{N_x \times N_y}(\mathbf{m}; t) \\ &\quad + \delta t P_{g_x} \left(\frac{N_x - 1}{N_x} \right) (N_x - 1)\rho_1^{(N_x - 1) \times N_y}(\mathbf{m}; t) \\ &\quad + \delta t P_{g_y} \left(\frac{N_y - 1}{N_y} \right) (N_y - 1)\rho_1^{N_x \times (N_y - 1)}(\mathbf{m}; t) + \mathcal{O}(\delta t^2). \end{aligned} \quad (5.4)$$

²We can assume translational invariance throughout this chapter because the initial agent density for all simulations in the IBM are achieved by populating lattice sites uniformly at random until the required density is achieved.

If we now rearrange Eq. (5.4) and take the limit as $\delta t \rightarrow 0$ we obtain

$$\begin{aligned} \frac{d\rho_1^{N_x \times N_y}(\mathbf{m}; t)}{dt} = & -(P_{g_x} N_x + P_{g_y} N_y) \rho_1^{N_x \times N_y}(\mathbf{m}; t) \\ & + P_{g_x} \left(\frac{N_x - 1}{N_x} \right) (N_x - 1) \rho_1^{(N_x - 1) \times N_y}(\mathbf{m}; t) \\ & + P_{g_y} \left(\frac{N_y - 1}{N_y} \right) (N_y - 1) \rho_1^{N_x \times (N_y - 1)}(\mathbf{m}; t). \end{aligned} \quad (5.5)$$

If we make the approximation $\rho_1^{N_x \times N_y}(\mathbf{m}; t) \approx ((N_y - 1)/N_y) \rho_1^{N_x \times (N_y - 1)}(\mathbf{m}; t)$ and $\rho_1^{N_x \times N_y}(\mathbf{m}; t) \approx ((N_x - 1)/N_x) \rho_1^{(N_x - 1) \times N_y}(\mathbf{m}; t)$, Eq. (5.5) can be written as

$$\begin{aligned} \frac{d\rho_1^{N_x \times N_y}(\mathbf{m}; t)}{dt} = & -(P_{g_x} N_x + P_{g_y} N_y) \rho_1^{N_x \times N_y}(\mathbf{m}; t) \\ & + P_{g_x} (N_x - 1) \rho_1^{N_x \times N_y}(\mathbf{m}; t) \\ & + P_{g_y} (N_y - 1) \rho_1^{N_x \times N_y}(\mathbf{m}; t). \end{aligned} \quad (5.6)$$

This approximation has been previously published [17], and reasonably implies that domain growth ‘dilutes’ the agent density. It is the same approximation that we used in Chapter 4. Finally, we simplify Eq. (5.6) to obtain

$$\frac{d\rho_1^{N_x \times N_y}(\mathbf{m}; t)}{dt} = -(P_{g_x} + P_{g_y}) \rho_1^{N_x \times N_y}(\mathbf{m}; t). \quad (5.7)$$

Equation (5.7) describes how exponential domain growth affects agent density. The method presented here is also extendable to other forms of isotropic domain growth, as demonstrated in Chapter 6. It is important to note that Eq. (5.7) describes how *exponential* domain growth affects the evolution of individual density functions because we have defined P_{g_x} and P_{g_y} as constants. In the course of the following derivation it will be useful to write the pairwise density functions in terms of the displacement between sites, therefore we shall rewrite the individual density functions

as

$$\rho_1^{N_x \times N_y}(\mathbf{m}; t) = c(t). \quad (5.8)$$

If we substitute Eq. (5.8) into Eq. (5.7) we obtain

$$\frac{dc(t)}{dt} = -(P_{gx} + P_{gy})c(t). \quad (5.9)$$

Eq. (5.9) is a MFA describing the effect of domain growth on the evolution agent density [78].

5.2.3 Pairwise density functions

We now derive the terms necessary to include the effect of two-dimensional exponential domain growth in the evolution of the pairwise density functions. Importantly, GM1 and GM2 affect the evolution of the pairwise density functions differently. Figure 5.2 displays two configurations of two agents, which we will term colinear and diagonal.

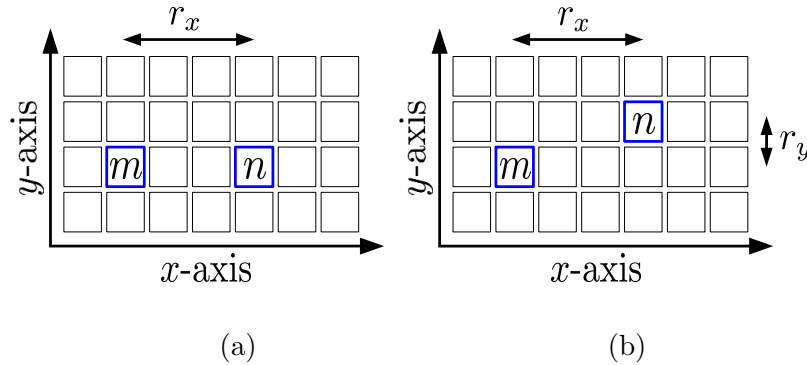


Figure 5.2: The two types of configuration of lattice sites: (a) colinear and (b) diagonal. The lattice sites in question are labelled $\mathbf{m}$ and $\mathbf{n}$ and bordered in blue (thick border). In (a), two colinear lattice sites share the same row but not the same column (or vice versa). In (b), two lattice sites are diagonal, meaning they do not share the same row or column. In both (a) and (b) r_x is the distance between two lattice sites in the horizontal direction, and r_y is the distance between two lattice sites in the vertical direction. In (b) this means $r_x = 3$ and $r_y = 1$.

The horizontal and vertical distances between sites is measured from their centres, as displayed in Fig. 5.2 (b), where r_x is the distance between two lattice sites in the horizontal direction and r_y is the distance between two lattice sites in the vertical direction.

5.2.3.1 Growth mechanism 1

We now derive equations for the evolution of the pairwise density functions for GM1 domain growth. We define the pairwise density functions, $\rho_2^{N_x \times N_y}(\mathbf{m}, \mathbf{n}; t)$, as the probability sites $\mathbf{m}$ and $\mathbf{n}$ are occupied at time t on a domain of size $N_x \times N_y$ (where $\mathbf{m} \neq \mathbf{n}$). We now rewrite $\rho_2^{N_x \times N_y}(\mathbf{m}, \mathbf{n}; t)$ in terms of a displacement vector between lattice sites, that is

$$\rho_2^{N_x \times N_y}(\mathbf{m}, \mathbf{n}; t) = \rho_2^{N_x \times N_y}(\mathbf{m}, \mathbf{m} + (r_x, r_y); t), \quad (5.10)$$

where $\mathbf{m}$ is (i, j) , with i indexing the row number of a lattice site, j indexing the column number of a lattice site, and (r_x, r_y) is a vector that represents a fixed displacement between $\mathbf{m}$ and $\mathbf{n}$, as defined in Fig. 5.2.

As the initial conditions in the IBM are, on average, spatially uniform we are able to assume translational invariance for the probability of two sites a given displacement apart being occupied. This means the pairwise density function can be written solely as a function of the displacement between two lattice sites, (r_x, r_y) . Therefore, we will further simplify our notation to obtain

$$\rho_2^{N_x \times N_y}(\mathbf{m}, \mathbf{m} + (r_x, r_y); t) = \rho_2^{N_x \times N_y}(r_x, r_y; t). \quad (5.11)$$

This ‘abuse’ of notation, whereby $\rho_2^{N_x \times N_y}$ is rewritten as a function of the displacement between two lattice sites as opposed to the lattice sites themselves, will prove useful

in the following derivation.

Colinear component

We begin with the colinear component of the equations for the evolution of the pairwise density functions, that is, the scenario in which the lattice sites in question share the same column or row, as depicted in Fig. 5.2 (a). For agents colinear in the horizontal direction, that is, $r_y = 0$, the evolution of the pairwise density functions for GM1 is

$$\begin{aligned}
\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \rho_2^{N_x \times N_y}(r_x, 0; t + \delta t) &= (1 - \delta t P_{g_x} N_x - \delta t P_{g_y} N_y) \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \rho_2^{N_x \times N_y}(r_x, 0; t) \\
&+ \delta t P_{g_x} \sum_{i=1}^{N_x-1} \sum_{j=1}^{N_y} (r_x - 1) \rho_2^{(N_x-1) \times N_y}(r_x - 1, 0; t) \\
&+ \delta t P_{g_x} \sum_{i=1}^{N_x-1} \sum_{j=1}^{N_y} (N_x - 1 - r_x) \rho_2^{(N_x-1) \times N_y}(r_x, 0; t) \\
&+ \delta t P_{g_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y-1} (N_y - 1) \rho_2^{N_x \times (N_y-1)}(r_x, 0; t) + \mathcal{O}(\delta t^2).
\end{aligned} \tag{5.12}$$

The terms on the RHS represent: i) the probability that no growth event occurs in $[t, t + \delta t)$; ii) the probability that a growth event occurs in the horizontal direction between agents $(r_x - 1, 0)$ apart, moving them $(r_x, 0)$ apart on a domain of size $N_x \times N_y$ at $[t + \delta t)$; iii) the probability that a growth event occurs in the horizontal direction at a site that is not in between agents $(r_x, 0)$ apart, meaning that they remain $(r_x, 0)$ apart but now on a domain of size $N_x \times N_y$ at time $[t + \delta t)$; and iv) the probability that a growth event occurs in the vertical direction (as the sites are horizontally colinear in this a GM1 growth event cannot change the distance between them).

Similarly, the evolution of the colinear component for GM1 for agents colinear in

the vertical direction, that is, $r_x = 0$, is

$$\begin{aligned}
 \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \rho_2^{N_x \times N_y}(0, r_y; t + \delta t) &= (1 - \delta t P_{g_x} N_x - \delta t P_{g_y} N_y) \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \rho_2^{N_x \times N_y}(0, r_y; t) \\
 &+ \delta t P_{g_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y-1} (r_y - 1) \rho_2^{N_x \times (N_y-1)}(0, r_y - 1; t) \\
 &+ \delta t P_{g_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y-1} (N_y - 1 - r_y) \rho_2^{N_x \times (N_y-1)}(0, r_y; t) \\
 &+ \delta t P_{g_x} \sum_{i=1}^{N_x-1} \sum_{j=1}^{N_y} (N_x - 1) \rho_2^{(N_x-1) \times N_y}(0, r_y; t) + \mathcal{O}(\delta t^2).
 \end{aligned} \tag{5.13}$$

As in the case of the individual density functions, we can simplify Eq. (5.12) to obtain

$$\begin{aligned}
 \rho_2^{N_x \times N_y}(r_x, 0; t + \delta t) &= (1 - \delta t P_{g_x} N_x - \delta t P_{g_y} N_y) \rho_2^{N_x \times N_y}(r_x, 0; t) \\
 &+ \delta t P_{g_x} \left(\frac{N_x - 1}{N_x} \right) (r_x - 1) \rho_2^{(N_x-1) \times N_y}(r_x - 1, 0; t) \\
 &+ \delta t P_{g_x} \left(\frac{N_x - 1}{N_x} \right) (N_x - 1 - r_x) \rho_2^{(N_x-1) \times N_y}(r_x, 0; t) \\
 &+ \delta t P_{g_y} \left(\frac{N_y - 1}{N_y} \right) (N_y - 1) \rho_2^{N_x \times (N_y-1)}(r_x, 0; t) + \mathcal{O}(\delta t^2).
 \end{aligned} \tag{5.14}$$

If we apply the approximations $((N_x - 1)/N_x) \rho_2^{(N_x-1) \times N_y} \approx \rho_2^{N_x \times N_y}$ and

$((N_y - 1)/N_y)\rho_2^{N_x \times (N_y - 1)} \approx \rho_2^{N_x \times N_y}$ ³, Eq. (5.14) can be rewritten as

$$\begin{aligned} \rho_2^{N_x \times N_y}(r_x, 0; t + \delta t) &= (1 - \delta t P_{g_x} N_x - \delta t P_{g_y} N_y) \rho_2^{N_x \times N_y}(r_x, 0; t) \\ &\quad + \delta t P_{g_x}(r_x - 1) \rho_2^{N_x \times N_y}(r_x - 1, 0; t) \\ &\quad + \delta t P_{g_x}(N_x - 1 - r_x) \rho_2^{N_x \times N_y}(r_x, 0; t) \\ &\quad + \delta t P_{g_y}(N_y - 1) \rho_2^{N_x \times N_y}(r_x, 0; t) + \mathcal{O}(\delta t^2). \end{aligned} \quad (5.15)$$

Rearranging Eq. (5.15) and taking the limit as $\delta t \rightarrow 0$ we finish with

$$\begin{aligned} \frac{d\rho_2^{N_x \times N_y}(r_x, 0; t)}{dt} &= P_{g_x}(r_x - 1) \rho_2^{N_x \times N_y}(r_x - 1, 0; t) \\ &\quad - P_{g_x}(r_x + 1) \rho_2^{N_x \times N_y}(r_x, 0; t) \\ &\quad - P_{g_y} \rho_2^{N_x \times N_y}(r_x, 0; t). \end{aligned} \quad (5.16)$$

The equivalent equation for sites colinear in the vertical direction (see Eq. (5.13)) is

$$\begin{aligned} \frac{d\rho_2^{N_x \times N_y}(0, r_y; t)}{dt} &= P_{g_y}(r_y - 1) \rho_2^{N_x \times N_y}(0, r_y - 1; t) \\ &\quad - P_{g_y}(r_y + 1) \rho_2^{N_x \times N_y}(0, r_y; t) \\ &\quad - P_{g_x} \rho_2^{N_x \times N_y}(0, r_y; t). \end{aligned} \quad (5.17)$$

Equations (5.16) and (5.17) describe the temporal evolution of the colinear components of the pairwise density functions for GM1.

³This approximation sensibly implies that domain growth ‘dilutes’ pairwise agent densities.

Diagonal component

For the diagonal component for GM1, that is, $r_x, r_y \neq 0$, we have, by similar reasoning,

$$\begin{aligned}
\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \rho_2^{N_x \times N_y}(r_x, r_y; t + \delta t) &= (1 - \delta t P_{g_x} N_x - \delta t P_{g_y} N_y) \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \rho_2^{N_x \times N_y}(r_x, r_y; t) \\
&+ \delta t P_{g_x} \sum_{i=1}^{N_x-1} \sum_{j=1}^{N_y} (r_x - 1) \rho_2^{(N_x-1) \times N_y}(r_x - 1, r_y; t) \\
&+ \delta t P_{g_x} \sum_{i=1}^{N_x-1} \sum_{j=1}^{N_y} (N_x - 1 - r_x) \rho_2^{(N_x-1) \times N_y}(r_x, r_y; t) \\
&+ \delta t P_{g_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y-1} (r_y - 1) \rho_2^{N_x \times (N_y-1)}(r_x, r_y - 1; t) \\
&+ \delta t P_{g_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y-1} (N_y - 1 - r_y) \rho_2^{N_x \times (N_y-1)}(r_x, r_y; t).
\end{aligned} \tag{5.18}$$

If we follow the same procedure as for the colinear component we obtain

$$\begin{aligned}
\frac{d\rho_2^{N_x \times N_y}(r_x, r_y; t)}{dt} &= P_{g_x}(r_x - 1) \rho_2^{N_x \times N_y}(r_x - 1, r_y; t) \\
&- P_{g_x}(r_x + 1) \rho_2^{N_x \times N_y}(r_x, r_y; t) \\
&+ P_{g_y}(r_y - 1) \rho_2^{N_x \times N_y}(r_x, r_y - 1; t) \\
&- P_{g_y}(r_y + 1) \rho_2^{N_x \times N_y}(r_x, r_y; t).
\end{aligned} \tag{5.19}$$

5.2.3.2 Growth mechanism 2

It is simpler to derive the evolution of the pairwise density functions for GM2 for a domain of size $(N_x + 1) \times (N_y + 1)$ ⁴. For GM2 it is also necessary to introduce some new notation. The distance of the lattice site $\mathbf{m}$ from the left-hand and bottom

⁴This avoids having $(N_x - 1)$ in the denominator of a number of fractions.

boundary of the domain is m_i and m_j , respectively, i.e. the position of $\mathbf{m}$, and the distance of the lattice site $\mathbf{n}$ from the left-hand and bottom boundary of the domain is n_i and n_j , respectively, i.e. the position of $\mathbf{n}$. How n_i and n_j are measured for lattice site $\mathbf{n}$ is displayed in Fig. 5.3.

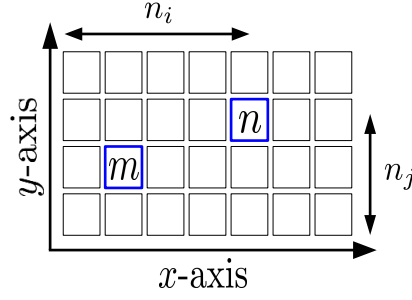


Figure 5.3: Measuring the distance of a lattice site from the boundary. For lattice site $\mathbf{n}$ the distance from the left-hand boundary of the domain is n_i , and the distance from the bottom boundary is n_j . For lattice site $\mathbf{m}$, m_i and m_j are measured in an analogous manner.

Growth mechanism 2 diagonal component

The evolution of the pairwise density functions for the diagonal component of GM2 is described by the following equation

$$\begin{aligned}
& \sum_{i=1}^{N_x+1} \sum_{j=1}^{N_y+1} \rho_2^{(N_x+1) \times (N_y+1)}(\mathbf{m}, \mathbf{m} + (r_x, r_y); t + \delta t) = \\
& (1 - \delta t P_{g_x}(N_x + 1) - \delta t P_{g_y}(N_y + 1)) \sum_{i=1}^{N_x+1} \sum_{j=1}^{N_y+1} \rho_2^{(N_x+1) \times (N_y+1)}(\mathbf{m}, \mathbf{m} + (r_x, r_y); t) \\
& + \delta t P_{g_y}(N_y) \sum_{i=1}^{N_x+1} \sum_{j=1}^{N_y} \left(\frac{m_j n_j}{(N_y)(N_y)} \right) \rho_2^{(N_x+1) \times N_y}(\mathbf{m}, \mathbf{m} + (r_x, r_y); t) \\
& + \delta t P_{g_y}(N_y) \sum_{i=1}^{N_x+1} \sum_{j=1}^{N_y} \left(\frac{(N_y - m_j)(N_y - n_j)}{(N_y)(N_y)} \right) \rho_2^{(N_x+1) \times N_y}(\mathbf{m}, \mathbf{m} + (r_x, r_y); t) \\
& + \delta t P_{g_y}(N_y) \sum_{i=1}^{N_x+1} \sum_{j=1}^{N_y} \left(\frac{(N_y - m_j)(n_j - 1)}{(N_y)(N_y)} \right) \rho_2^{(N_x+1) \times N_y}(\mathbf{m}, \mathbf{m} + (r_x, r_y - 1); t) \\
& + \delta t P_{g_y}(N_y) \sum_{i=1}^{N_x+1} \sum_{j=1}^{N_y} \left(\frac{(m_j - 1)(N_y - n_j)}{(N_y)(N_y)} \right) \rho_2^{(N_x+1) \times N_y}(\mathbf{m}, \mathbf{m} + (r_x, r_y + 1); t) \\
& + \delta t P_{g_x}(N_x) \sum_{i=1}^{N_x} \sum_{j=1}^{N_y+1} \left(\frac{m_i n_i}{(N_x)(N_x)} \right) \rho_2^{N_x \times (N_y+1)}(\mathbf{m}, \mathbf{m} + (r_x, r_y); t) \\
& + \delta t P_{g_x}(N_x) \sum_{i=1}^{N_x} \sum_{j=1}^{N_y+1} \left(\frac{(N_x - m_i)(N_x - n_i)}{(N_x)(N_x)} \right) \rho_2^{N_x \times (N_y+1)}(\mathbf{m}, \mathbf{m} + (r_x, r_y); t) \\
& + \delta t P_{g_x}(N_x) \sum_{i=1}^{N_x} \sum_{j=1}^{N_y+1} \left(\frac{(N_x - m_i)(n_i - 1)}{(N_x)(N_x)} \right) \rho_2^{N_x \times (N_y+1)}(\mathbf{m}, \mathbf{m} + (r_x - 1, r_y); t) \\
& + \delta t P_{g_x}(N_x) \sum_{i=1}^{N_x} \sum_{j=1}^{N_y+1} \left(\frac{(m_i - 1)(N_x - n_i)}{(N_x)(N_x)} \right) \rho_2^{N_x \times (N_y+1)}(\mathbf{m}, \mathbf{m} + (r_x + 1, r_y); t).
\end{aligned} \tag{5.20}$$

The terms on the RHS of Eq. (5.20) represent: i) the probability that no growth event occurs in $[t, t + \delta t)$; ii) the probability that a growth event occurs in the vertical direction and both agents are moved by it; iii) the probability that a growth event occurs in the vertical direction and both agents are not moved by it; iv) the probability that a growth event occurs in the vertical direction and $\mathbf{m} + (r_x, r_y - 1)$ moves and $\mathbf{m}$ does not move (we can write this in terms of a displacement vector: $(r_x, r_y - 1)$); and v)

the probability that a growth event occurs in the vertical direction and $\mathbf{m}+(r_x, -1)$ moves and $\mathbf{m}+(r_x, r_y)$ does not move. The rest of the terms are the equivalent for growth in the horizontal direction. First, we assume translational invariance and simplify Eq. (5.20) to obtain

$$\begin{aligned}
& \sum_{i=1}^{N_x+1} \sum_{j=1}^{N_y+1} \rho_2^{(N_x+1) \times (N_y+1)}(r_x, r_y; t + \delta t) = \\
& (1 - \delta t P_{g_x}(N_x + 1) - \delta t P_{g_y}(N_y + 1)) \sum_{i=1}^{N_x+1} \sum_{j=1}^{N_y+1} \rho_2^{(N_x+1) \times (N_y+1)}(r_x, r_y; t) \\
& + \delta t P_{g_y} \sum_{i=1}^{N_x+1} \sum_{j=1}^{N_y} \left(N_y - n_j - m_j + \frac{2m_j n_j}{N_y} \right) \rho_2^{(N_x+1) \times N_y}(r_x, r_y; t) \\
& + \delta t P_{g_y} \sum_{i=1}^{N_x+1} \sum_{j=1}^{N_y} \left(\frac{(N_y - m_j)(n_j - 1)}{(N_y)} \right) \rho_2^{(N_x+1) \times N_y}(r_x, r_y - 1; t) \\
& + \delta t P_{g_y} \sum_{i=1}^{N_x+1} \sum_{j=1}^{N_y} \left(\frac{(m_j - 1)(N_y - n_j)}{(N_y)} \right) \rho_2^{(N_x+1) \times N_y}(r_x, r_y + 1; t) \\
& + \delta t P_{g_x} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y+1} \left(N_x - n_i - m_i + \frac{2m_i n_i}{N_x} \right) \rho_2^{N_x \times (N_y+1)}(r_x, r_y; t) \\
& + \delta t P_{g_x} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y+1} \left(\frac{(N_x - m_i)(n_i - 1)}{(N_x)} \right) \rho_2^{N_x \times (N_y+1)}(r_x - 1, r_y; t) \\
& + \delta t P_{g_x} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y+1} \left(\frac{(m_i - 1)(N_x - n_i)}{(N_x)} \right) \rho_2^{N_x \times (N_y+1)}(r_x + 1, r_y; t).
\end{aligned} \tag{5.21}$$

We rewrite Eq. (5.21) using $\tilde{n}_x$, $\tilde{m}_x$ and $m_x \tilde{n}_x$, which are constants defined as

$$\tilde{n}_x = \frac{1}{(N_x)(N_y + 1)} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y+1} n_i, \tag{5.22}$$

$$\tilde{m}_x = \frac{1}{(N_x)(N_y + 1)} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y+1} m_i, \tag{5.23}$$

and

$$\begin{aligned}
m_x \tilde{n}_x &= \frac{1}{(N_x)(N_y + 1)} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y+1} m_i n_i = \\
&= \frac{1}{(N_x)(N_y + 1)} \sum_{j=1}^{N_y+1} \left[\sum_{i=1}^{N_x-r_x} (m_i)(m_i + r_x) + \sum_{i=N_x-r_x+1}^{N_x} (m_i)(m_i - (N_x - r_x)) \right] \\
&= \frac{1}{(N_x)(N_y + 1)} \sum_{j=1}^{N_y+1} \left[\sum_{i=1}^{N_x} (m_i)(m_i + r_x) - \sum_{i=N_x-r_x+1}^{N_x} N_x m_i \right] \\
&= \frac{1}{(N_x)(N_y + 1)} \sum_{j=1}^{N_y+1} \left[\sum_{i=1}^{N_x} (m_i)(m_i + r_x) + N_x \left(\sum_{i=1}^{N_x-r_x} m_i - \sum_{i=1}^{N_x} m_i \right) \right]. \quad (5.24)
\end{aligned}$$

Eqs. (5.22)-(5.24) can be evaluated directly to give

$$\tilde{n}_x = \frac{(N_x + 1)}{2}, \quad (5.25)$$

$$\tilde{m}_x = \frac{(N_x + 1)}{2}, \quad (5.26)$$

and

$$m_x \tilde{n}_x = N_x \left(\frac{N_x}{3} + \frac{1}{2} \right) + \frac{r_x}{2} (r_x - N_x) + \frac{1}{6}. \quad (5.27)$$

We substitute Eqs. (5.22)-(5.24) into Eq. (5.21) to obtain

$$\begin{aligned}
& (N_x + 1)(N_y + 1)\rho_2^{(N_x+1) \times (N_y+1)}(r_x, r_y; t + \delta t) = \\
& (1 - \delta t P_{g_x}(N_x + 1) - \delta t P_{g_y}(N_y + 1))(N_x + 1)(N_y + 1)\rho_2^{(N_x+1) \times (N_y+1)}(r_x, r_y; t) \\
& + \delta t P_{g_y}(N_x + 1)(N_y) \left(N_y - \tilde{n}_y - \tilde{m}_y + \frac{2m_y \tilde{n}_y}{N_y} \right) \rho_2^{(N_x+1) \times N_y}(r_x, r_y; t) \\
& + \delta t P_{g_y}(N_x + 1)(N_y) \left(\tilde{n}_y - 1 - \frac{m_y \tilde{n}_y}{N_y} + \frac{\tilde{m}_y}{N_y} \right) \rho_2^{(N_x+1) \times N_y}(r_x, r_y - 1; t) \\
& + \delta t P_{g_y}(N_x + 1)(N_y) \left(\tilde{m}_y - 1 - \frac{m_y \tilde{n}_y}{N_y} + \frac{\tilde{n}_y}{N_y} \right) \rho_2^{(N_x+1) \times N_y}(r_x, r_y + 1; t) \\
& + \delta t P_{g_x}(N_x)(N_y + 1) \left(N_x - \tilde{n}_x - \tilde{m}_x + \frac{2m_x \tilde{n}_x}{N_x} \right) \rho_2^{N_x \times (N_y+1)}(r_x, r_y; t) \\
& + \delta t P_{g_x}(N_x)(N_y + 1) \left(\tilde{n}_x - 1 - \frac{m_x \tilde{n}_x}{N_x} + \frac{\tilde{m}_x}{N_x} \right) \rho_2^{N_x \times (N_y+1)}(r_x - 1, r_y; t) \\
& + \delta t P_{g_x}(N_x)(N_y + 1) \left(\tilde{m}_x - 1 - \frac{m_x \tilde{n}_x}{N_x} + \frac{\tilde{n}_x}{N_x} \right) \rho_2^{N_x \times (N_y+1)}(r_x + 1, r_y; t).
\end{aligned} \tag{5.28}$$

That Eqs. (5.21) and (5.28) are equivalent can be shown by substituting Eqs. (5.22)-(5.24) into Eq. (5.28).

If we now apply the approximation

$$\rho_2^{(N_x+1) \times (N_y+1)}(r_x, r_y; t) \approx \left(\frac{N_y}{N_y + 1} \right) \rho_2^{(N_x+1) \times N_y}(r_x, r_y; t), \tag{5.29}$$

to Eq. (5.28) we obtain

$$\begin{aligned}
\rho_2^{(N_x+1) \times (N_y+1)}(r_x, r_y; t + \delta t) = & \\
& (1 - \delta t P_{g_x}(N_x + 1) - \delta t P_{g_y}(N_y + 1)) \rho_2^{(N_x+1) \times (N_y+1)}(r_x, r_y; t) \\
& + \delta t P_{g_y} \left(N_y - \tilde{n}_y - \tilde{m}_y + \frac{2m_y \tilde{n}_y}{N_y} \right) \rho_2^{(N_x+1) \times (N_y+1)}(r_x, r_y; t) \\
& + \delta t P_{g_y} \left(\tilde{n}_y - 1 - \frac{m_y \tilde{n}_y}{N_y} + \frac{\tilde{m}_y}{N_y} \right) \rho_2^{(N_x+1) \times (N_y+1)}(r_x, r_y - 1; t) \\
& + \delta t P_{g_y} \left(\tilde{m}_y - 1 - \frac{m_y \tilde{n}_y}{N_y} + \frac{\tilde{n}_y}{N_y} \right) \rho_2^{(N_x+1) \times (N_y+1)}(r_x, r_y + 1; t) \\
& + \delta t P_{g_x} \left(N_x - \tilde{n}_x - \tilde{m}_x + \frac{2m_x \tilde{n}_x}{N_x} \right) \rho_2^{(N_x+1) \times (N_y+1)}(r_x, r_y; t) \\
& + \delta t P_{g_x} \left(\tilde{n}_x - 1 - \frac{m_x \tilde{n}_x}{N_x} + \frac{\tilde{m}_x}{N_x} \right) \rho_2^{(N_x+1) \times (N_y+1)}(r_x - 1, r_y; t) \\
& + \delta t P_{g_x} \left(\tilde{m}_x - 1 - \frac{m_x \tilde{n}_x}{N_x} + \frac{\tilde{n}_x}{N_x} \right) \rho_2^{(N_x+1) \times (N_y+1)}(r_x + 1, r_y; t).
\end{aligned} \tag{5.30}$$

If we rearrange Eq. (5.30) and take the limit as $\delta t \rightarrow 0$ we arrive at

$$\begin{aligned}
\frac{d\rho_2^{(N_x+1) \times (N_y+1)}(r_x, r_y; t)}{dt} = & \\
& - (P_{g_x}(N_x + 1) + P_{g_y}(N_y + 1)) \rho_2^{(N_x+1) \times (N_y+1)}(r_x, r_y; t) \\
& + P_{g_y} \left(N_y - \tilde{n}_y - \tilde{m}_y + \frac{2m_y \tilde{n}_y}{N_y} \right) \rho_2^{(N_x+1) \times (N_y+1)}(r_x, r_y; t) \\
& + P_{g_y} \left(\tilde{n}_y - 1 - \frac{m_y \tilde{n}_y}{N_y} + \frac{\tilde{m}_y}{N_y} \right) \rho_2^{(N_x+1) \times (N_y+1)}(r_x, r_y - 1; t) \\
& + P_{g_y} \left(\tilde{m}_y - 1 - \frac{m_y \tilde{n}_y}{N_y} + \frac{\tilde{n}_y}{N_y} \right) \rho_2^{(N_x+1) \times (N_y+1)}(r_x, r_y + 1; t) \\
& + P_{g_x} \left(N_x - \tilde{n}_x - \tilde{m}_x + \frac{2m_x \tilde{n}_x}{N_x} \right) \rho_2^{(N_x+1) \times (N_y+1)}(r_x, r_y; t) \\
& + P_{g_x} \left(\tilde{n}_x - 1 - \frac{m_x \tilde{n}_x}{N_x} + \frac{\tilde{m}_x}{N_x} \right) \rho_2^{(N_x+1) \times (N_y+1)}(r_x - 1, r_y; t) \\
& + P_{g_x} \left(\tilde{m}_x - 1 - \frac{m_x \tilde{n}_x}{N_x} + \frac{\tilde{n}_x}{N_x} \right) \rho_2^{(N_x+1) \times (N_y+1)}(r_x + 1, r_y; t),
\end{aligned} \tag{5.31}$$

which we can simplify further to obtain

$$\begin{aligned}
\frac{d\rho_2^{(N_x+1) \times (N_y+1)}(r_x, r_y; t)}{dt} = & \\
& P_{g_x} \left(-1 + \left(\frac{2m_x \tilde{n}_x}{N_x} \right) - \tilde{n}_x - \tilde{m}_x \right) \rho_2^{(N_x+1) \times (N_y+1)}(r_x, r_y; t) \\
& + P_{g_y} \left(-1 + \left(\frac{2m_y \tilde{n}_y}{N_y} \right) - \tilde{n}_y - \tilde{m}_y \right) \rho_2^{(N_x+1) \times (N_y+1)}(r_x, r_y; t) \\
& + P_{g_y} \left(\tilde{n}_y - 1 - \frac{m_y \tilde{n}_y}{N_y} + \frac{\tilde{m}_y}{N_y} \right) \rho_2^{(N_x+1) \times (N_y+1)}(r_x, r_y - 1; t) \\
& + P_{g_y} \left(\tilde{m}_y - 1 - \frac{m_y \tilde{n}_y}{N_y} + \frac{\tilde{n}_y}{N_y} \right) \rho_2^{(N_x+1) \times (N_y+1)}(r_x, r_y + 1; t) \\
& + P_{g_x} \left(\tilde{n}_x - 1 - \frac{m_x \tilde{n}_x}{N_x} + \frac{\tilde{m}_x}{N_x} \right) \rho_2^{(N_x+1) \times (N_y+1)}(r_x - 1, r_y; t) \\
& + P_{g_x} \left(\tilde{m}_x - 1 - \frac{m_x \tilde{n}_x}{N_x} + \frac{\tilde{n}_x}{N_x} \right) \rho_2^{(N_x+1) \times (N_y+1)}(r_x + 1, r_y; t).
\end{aligned} \tag{5.32}$$

If we now evaluate Eqs. (5.22)-(5.24) with Eqs. (5.25)-(5.27) we have

$$\begin{aligned}
\frac{d\rho_2^{(N_x+1) \times (N_y+1)}(r_x, r_y; t)}{dt} = & \\
& P_{g_x} \left(-1 - \frac{N_x}{3} + r_x \left(\frac{r_x}{N_x} - 1 \right) + \frac{1}{3N_x} \right) \rho_2^{(N_x+1) \times (N_y+1)}(r_x, r_y; t) \\
& + P_{g_y} \left(-1 - \frac{N_y}{3} + r_y \left(\frac{r_y}{N_y} - 1 \right) + \frac{1}{3N_y} \right) \rho_2^{(N_x+1) \times (N_y+1)}(r_x, r_y; t) \\
& + P_{g_y} \left(-\frac{1}{2} + \frac{N_y}{6} - \frac{r_y}{2} \left(\frac{r_y}{N_y} - 1 \right) + \frac{1}{3N_y} \right) \rho_2^{(N_x+1) \times (N_y+1)}(r_x, r_y - 1; t) \\
& + P_{g_y} \left(-\frac{1}{2} + \frac{N_y}{6} - \frac{r_y}{2} \left(\frac{r_y}{N_y} - 1 \right) + \frac{1}{3N_y} \right) \rho_2^{(N_x+1) \times (N_y+1)}(r_x, r_y + 1; t) \\
& + P_{g_x} \left(-\frac{1}{2} + \frac{N_x}{6} - \frac{r_x}{2} \left(\frac{r_x}{N_x} - 1 \right) + \frac{1}{3N_x} \right) \rho_2^{(N_x+1) \times (N_y+1)}(r_x - 1, r_y; t) \\
& + P_{g_x} \left(-\frac{1}{2} + \frac{N_x}{6} - \frac{r_x}{2} \left(\frac{r_x}{N_x} - 1 \right) + \frac{1}{3N_x} \right) \rho_2^{(N_x+1) \times (N_y+1)}(r_x + 1, r_y; t).
\end{aligned} \tag{5.33}$$

Growth mechanism 2 colinear component

For the colinear component we have (assuming that lattice sites are horizontally colinear, i.e. $r_y = 0$)

$$\begin{aligned}
& \sum_{i=1}^{N_x+1} \sum_{j=1}^{N_y+1} \rho_2^{(N_x+1) \times (N_y+1)}(\mathbf{m}, \mathbf{m} + (r_x, 0); t + \delta t) = \\
& (1 - \delta t P_{g_x}(N_x + 1) - \delta t P_{g_y}(N_y + 1)) \sum_{i=1}^{N_x+1} \sum_{j=1}^{N_y+1} \rho_2^{(N_x+1) \times (N_y+1)}(\mathbf{m}, \mathbf{m} + (r_x, 0); t) \\
& + \delta t P_{g_y}(N_y) \sum_{i=1}^{N_x+1} \sum_{j=1}^{N_y} \left(\frac{m_j n_j}{(N_y)(N_y)} \right) \rho_2^{(N_x+1) \times N_y}(\mathbf{m}, \mathbf{m} + (r_x, 0); t) \\
& + \delta t P_{g_y}(N_y) \sum_{i=1}^{N_x+1} \sum_{j=1}^{N_y} \left(\frac{(N_y - n_j)(N_y - m_j)}{(N_y)(N_y)} \right) \rho_2^{(N_x+1) \times N_y}(\mathbf{m}, \mathbf{m} + (r_x, 0); t) \\
& + \delta t P_{g_y}(N_y) \sum_{i=1}^{N_x+1} \sum_{j=1}^{N_y} \left(\frac{(N_y - m_j)(n_j - 1)}{(N_y)(N_y)} \right) \rho_2^{(N_x+1) \times N_y}(\mathbf{m}, \mathbf{m} + (r_x, -1); t) \\
& + \delta t P_{g_y}(N_y) \sum_{i=1}^{N_x+1} \sum_{j=1}^{N_y} \left(\frac{(N_y - n_j)(m_j - 1)}{(N_y)(N_y)} \right) \rho_2^{(N_x+1) \times N_y}(\mathbf{m}, \mathbf{m} + (r_x, -1); t) \\
& + \delta t P_{g_x}(N_x) \sum_{i=1}^{N_x} \sum_{j=1}^{N_y+1} \left(\frac{(r_x - 1)}{(N_x)} \right) \rho_2^{N_x \times (N_y+1)}(\mathbf{m}, \mathbf{m} + (r_x - 1, 0); t) \\
& + \delta t P_{g_x}(N_x) \sum_{i=1}^{N_x} \sum_{j=1}^{N_y+1} \left(\frac{(N_x - r_x)}{(N_x)} \right) \rho_2^{N_x \times (N_y+1)}(\mathbf{m}, \mathbf{m} + (r_x, 0); t). \tag{5.34}
\end{aligned}$$

The terms on the RHS of Eq. (5.34) represent: i) the probability that no growth event occurs in $[t, t + \delta t]$; ii) the probability that a growth event occurs in the vertical direction and both agents are moved by it; iii) the probability that a growth event occurs in the vertical direction and both agents are not moved by it; iv) the probability that a growth event occurs in the vertical direction and $\mathbf{m} + (r_x, -1)$ moves and $\mathbf{m}$ does not move; and v) the probability that a growth event occurs in the vertical direction and $\mathbf{m} + (r_x, 0)$ does not move and $\mathbf{m} + (0, -1)$ does move. The rest of the terms are for growth in the horizontal direction. We rewrite Eq. (5.34) using Eqs. (5.22)-(5.24) to

obtain

$$\begin{aligned}
(N_x + 1)(N_y + 1)\rho_2^{(N_x+1) \times (N_y+1)}(r_x, 0; t + \delta t) = & \\
(1 - \delta t P_{g_x}(N_x + 1) - \delta t P_{g_y}(N_y + 1))(N_x + 1)(N_y + 1)\rho_2^{(N_x+1) \times (N_y+1)}(r_x, 0; t) & \\
+ \delta t P_{g_y}(N_x + 1)(N_y) \left(\frac{m_y \tilde{n}_y}{N_y} \right) \rho_2^{(N_x+1) \times N_y}(r_x, 0; t) & \\
+ \delta t P_{g_y}(N_x + 1)(N_y) \left(N_y - \tilde{n}_y - \tilde{m}_y + \frac{m_y \tilde{n}_y}{N_y} \right) \rho_2^{(N_x+1) \times N_y}(r_x, 0; t) & \\
+ \delta t P_{g_y}(N_x + 1)(N_y) \left(\tilde{n}_y - 1 - \frac{m_y \tilde{n}_y}{N_y} + \frac{\tilde{m}_y}{N_y} \right) \rho_2^{(N_x+1) \times N_y}(r_x, -1; t) & \\
+ \delta t P_{g_y}(N_x + 1)(N_y) \left(\tilde{m}_y - 1 - \frac{m_y \tilde{n}_y}{N_y} + \frac{\tilde{n}_y}{N_y} \right) \rho_2^{(N_x+1) \times N_y}(r_x, -1; t) & \\
+ \delta t P_{g_x}(N_x)(N_y + 1)(r_x - 1)\rho_2^{N_x \times (N_y+1)}(r_x - 1, 0; t) & \\
+ \delta t P_{g_x}(N_x)(N_y + 1)(N_x - r_x)\rho_2^{N_x \times (N_y+1)}(r_x, 0; t). & \quad (5.35)
\end{aligned}$$

We now make the approximation Eq. (5.29) to arrive at

$$\begin{aligned}
\rho_2^{(N_x+1) \times (N_y+1)}(r_x, 0; t + \delta t) = & (1 - \delta t P_{g_x}(N_x + 1) - \delta t P_{g_y}(N_y + 1))\rho_2^{(N_x+1) \times (N_y+1)}(r_x, 0; t) \\
& + \delta t P_{g_y} \left(\frac{m_y \tilde{n}_y}{N_y} \right) \rho_2^{(N_x+1) \times (N_y+1)}(r_x, 0; t) \\
& + \delta t P_{g_y} \left(N_y - \tilde{n}_y - \tilde{m}_y + \frac{m_y \tilde{n}_y}{N_y} \right) \rho_2^{(N_x+1) \times (N_y+1)}(r_x, 0; t) \\
& + \delta t P_{g_y} \left(\tilde{n}_y - 1 - \frac{m_y \tilde{n}_y}{N_y} + \frac{\tilde{m}_y}{N_y} \right) \rho_2^{(N_x+1) \times (N_y+1)}(r_x, -1; t) \\
& + \delta t P_{g_y} \left(\tilde{m}_y - 1 - \frac{m_y \tilde{n}_y}{N_y} + \frac{\tilde{n}_y}{N_y} \right) \rho_2^{(N_x+1) \times (N_y+1)}(r_x, -1; t) \\
& + \delta t P_{g_x}(r_x - 1)\rho_2^{(N_x+1) \times (N_y+1)}(r_x - 1, 0; t) \\
& + \delta t P_{g_x}(N_x - r_x)\rho_2^{(N_x+1) \times (N_y+1)}(r_x, 0; t), \quad (5.36)
\end{aligned}$$

which we can simplify to obtain

$$\begin{aligned}
\frac{d\rho_2^{(N_x+1) \times (N_y+1)}(r_x, 0; t)}{dt} = & -P_{g_x}(r_x + 1)\rho_2^{(N_x+1) \times (N_y+1)}(r_x, 0; t) \\
& + P_{g_x}(r_x - 1)\rho_2^{(N_x+1) \times (N_y+1)}(r_x - 1, 0; t) \\
& + P_{g_y}\left(-1 + \left(\frac{2m_y\tilde{n}_y}{N_y}\right) - \tilde{n}_y - \tilde{m}_y\right)\rho_2^{(N_x+1) \times (N_y+1)}(r_x, 0; t) \\
& + P_{g_y}\left(\tilde{n}_y - 1 - \frac{m_y\tilde{n}_y}{N_y} + \frac{\tilde{m}_y}{N_y}\right)\rho_2^{(N_x+1) \times (N_y+1)}(r_x, -1; t) \\
& + P_{g_y}\left(\tilde{m}_y - 1 - \frac{m_y\tilde{n}_y}{N_y} + \frac{\tilde{n}_y}{N_y}\right)\rho_2^{(N_x+1) \times (N_y+1)}(r_x, -1; t).
\end{aligned} \tag{5.37}$$

If we evaluate Eqs. (5.22)-(5.24) with (5.25)-(5.27) we have

$$\begin{aligned}
\frac{d\rho_2^{(N_x+1) \times (N_y+1)}(r_x, 0; t)}{dt} = & -P_{g_x}(r_x + 1)\rho_2^{(N_x+1) \times (N_y+1)}(r_x, 0; t) \\
& + P_{g_x}(r_x - 1)\rho_2^{(N_x+1) \times (N_y+1)}(r_x - 1, 0; t) \\
& + P_{g_y}\left(-1 - \frac{N_y}{3} + \frac{1}{3N_y}\right)\rho_2^{(N_x+1) \times (N_y+1)}(r_x, 0; t) \\
& + P_{g_y}\left(-1 + \frac{N_y}{3} + \frac{2}{3N_y}\right)\rho_2^{(N_x+1) \times (N_y+1)}(r_x, -1; t).
\end{aligned} \tag{5.38}$$

Therefore, the evolution equation for the colinear component (horizontally colinear)

for GM2 is

$$\begin{aligned}
\frac{d\rho_2^{(N_x+1)\times(N_y+1)}(r_x, 0; t)}{dt} = & -P_{g_x}(r_x + 1)\rho_2^{(N_x+1)\times(N_y+1)}(r_x, 0; t) \\
& + P_{g_x}(r_x - 1)\rho_2^{(N_x+1)\times(N_y+1)}(r_x - 1, 0; t) \\
& + P_{g_y}\left(-1 - \frac{N_y}{3} + \frac{1}{3N_y}\right)\rho_2^{(N_x+1)\times(N_y+1)}(r_x, 0; t) \\
& + P_{g_y}\left(-1 + \frac{N_y}{3} + \frac{2}{3N_y}\right)\rho_2^{(N_x+1)\times(N_y+1)}(r_x, -1; t).
\end{aligned} \tag{5.39}$$

An analogous equation exists for the vertically colinear component for GM2.

If we compare Eqs. (5.16) and (5.19), and Eqs. (5.33) and (5.39) it is apparent that the length of the domain influences the evolution of the pairwise density functions in the case of GM2, but not in GM1. Therefore, we expect the evolution of the pairwise density functions to be affected by the initial domain size in the case of GM2, but not GM1.

5.3 Results

We initially present results for immotile, non-proliferative agents on a growing domain, in which domain growth is exponential. We simulate the scenario in which all lattice sites are occupied by an agent in the IBM on a lattice of size 100 by 100 sites. This means the initial conditions for Eqs. (5.9), (5.16), (5.19), (5.33) and (5.39) are

$$c(0) = 1, \tag{5.40}$$

and

$$\rho_2^{N_x \times N_y}(r_x, r_y; 0) = 1, \quad \forall r_x < N_x, r_y < N_y. \quad (5.41)$$

As previously stated, N_x and N_y describe the number of lattice sites in the horizontal and vertical directions, respectively. However, as results from the IBM are ensemble averages we substitute N_x and N_y for their continuum analogues $L_x(t)$ and $L_y(t)$, respectively. For exponential domain growth $L_x(t)$ evolves according to

$$\frac{dL_x(t)}{dt} = P_{g_x} L_x(t). \quad (5.42)$$

The evolution of $L_y(t)$ is equivalent. This substitution of $N_x(t)$ and $N_y(t)$ with $L_x(t)$ and $L_y(t)$ avoids jump discontinuities in the numerical solutions of Eqs. (5.16), (5.19), (5.33) and (5.39), which are not present in the averaged IBM results. To solve our equations numerically we use MATLAB's `ode15s`, with an absolute error tolerance of 10^{-12} .

5.3.1 Agents on a growing domain

In Fig. 5.4 the numerical solutions of Eqs. (5.16), (5.19), (5.33) and (5.39) are shown for the evolution of the pairwise density functions on a growing domain. In Fig. 5.4 (a) we plot the evolution of the pairwise density functions for GM1, and see the anticipated difference in the evolution of the colinear and diagonal components. In the absence of agent motility and proliferation the evolution of the pairwise density functions for all colinear distances (i.e. $\Delta, 2\Delta, 3\Delta, \dots$) is the same for GM1, this is also the case for the diagonal component. Conversely, in Fig. 5.4 (b) we plot the evolution of the pairwise density functions for GM2, and in this case in the evolution of all the colinear and diagonal components is the same.

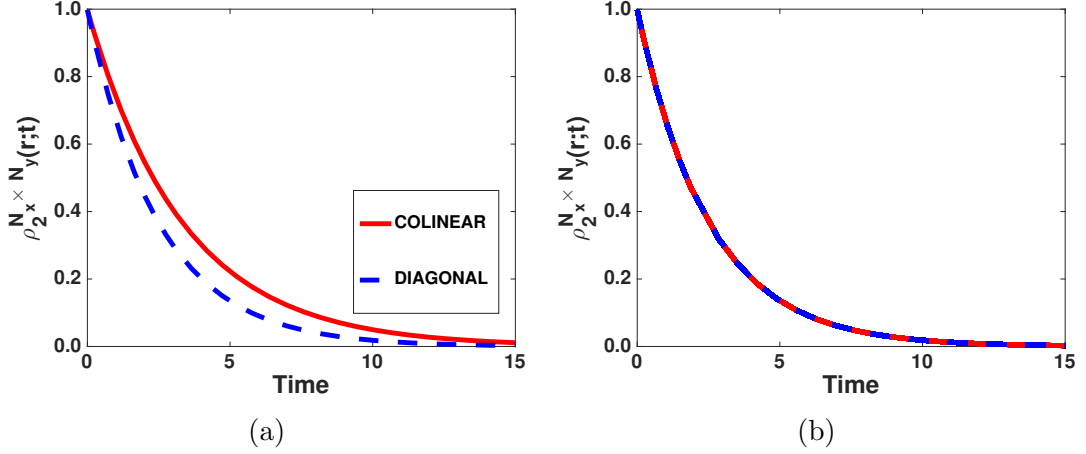


Figure 5.4: (Colour online): The evolution of the colinear and diagonal components of the pairwise density functions for (a) GM1 and (b) GM2. $P_{g_x} = P_{g_y} = 0.1$.

5.3.2 Agent motility, proliferation and death

To include the effect of agent motility and proliferation in Eqs. (5.9), (5.16), (5.19), (5.33) and (5.39) we follow a similar procedure as presented in Chapter 4, and as such we present the detailed derivation in Appendix B. If we include agent motility and proliferation in Eq. (5.9) we obtain

$$\frac{dc(t)}{dt} = P_p \left(c(t) - \rho_2^{N_x \times N_y}(1, 0; t) \right) - (P_{g_x} + P_{g_y})c(t). \quad (5.43)$$

As can be seen from Eq. (5.43), the inclusion of agent proliferation means that pairwise density functions are now present in the equations for the evolution of the individual density functions, which is not the case without agent proliferation (Eq. (5.9)). We present the detailed derivation with the inclusion of agent motility and proliferation in the pairwise density functions in Appendix B.

Correlation functions

To make our results comparable with the large amount of research already completed in this area we display the results in terms of the correlation function [18, 19, 21–

23, 78, 84–87]. The correlation function [86, 87] is simply defined as

$$F(r_x, r_y; t) := \frac{\rho_2^{N_x \times N_y}(r_x, r_y; t)}{c(t)^2}. \quad (5.44)$$

As discussed in Chapter 4, the correlation function is a useful variable to work with as $F(r_x, r_y; t) \equiv 1$ means that the occupancy of two lattice sites of a given displacement is independent. If we substitute Eq. (5.44) into Eq. (5.43) we obtain

$$\frac{dc(t)}{dt} = P_p c(t) \left(1 - F(1, 0; t) c(t) \right) - (P_{g_x} + P_{g_y}) c(t). \quad (5.45)$$

The standard MFA assumes $F(1, 0; t) = 1$, that is, the effect of spatial correlations is insignificant, and so Eq. (5.45) becomes

$$\frac{dc(t)}{dt} = P_p c(t) (1 - c(t)) - (P_{g_x} + P_{g_y}) c(t). \quad (5.46)$$

For all IBM simulations we use initial uniform random seeding of density 0.05. The lattice state of site $\mathbf{m}$ of the two-dimensional IBM is described by variable $\sigma_{\mathbf{m}}$ (i.e. occupied by an agent or unoccupied). This means the normalised average agent density for the two-dimensional domain is

$$c(t) = \frac{1}{N_x(t)N_y(t)} \sum_{\mathbf{m}}^{N_x(t)N_y(t)} \mathbb{1}_A\{\sigma_{\mathbf{m}}\}. \quad (5.47)$$

Here $\mathbb{1}_A$ is the indicator function for species A (i.e. 1 if species A occupies site $\mathbf{m}$, and 0 if it does not). As previously discussed, the initial condition in the IBM is achieved by populating a certain number of sites uniformly at random until the initial density is achieved, the initial condition for the correlation function is therefore

$$F(r_x, r_y; 0) = 1, \quad \forall r_x < N_x, r_y < N_y. \quad (5.48)$$

That is, at all distances there are initially no spatial correlations between agents. We also rescale time to allow for ease of comparison between simulations with different parameters:

$$\bar{t} = \left(P_p - P_d - (P_{g_x} + P_{g_y}) \right) t. \quad (5.49)$$

The rescaling of time is necessary to allow the easy comparison of simulations with different parameters. As before we solve our equations numerically with MATLAB's `ode15s` with an absolute error tolerance of 10^{-12} .

5.3.3 Agent motility and proliferation

In Fig. 5.5 the effects of domain growth via GM1 on the evolution of the macroscopic agent density can be seen ((a) LPE = 0.04, (b) 0.09)⁵. We see that the inclusion of GM1 renders the steady-state density predicted by the standard MFA, Eq. (5.46), to be inaccurate. The steady-state density calculated from an ensemble average from the IBM is lower than the MFA predicts, which the correlation ODE model (Eq. (5.45)) is able to capture. The rate at which the agent density increases is also reduced in the IBM due to spatial correlations, and the correlations ODE model better approximates this than the MFA. One may expect domain growth to decrease spatial correlations by ‘breaking-up’ existing agent clusters which, in turn, would allow greater agent movement and cause the MFA to be more accurate. However, the nature of GM1 means that it does not break up agents effectively, only potentially ‘freeing-up’ one lattice site for two agents in each row/column with every growth event (see Fig. 5.1). GM1 also does not cause agents to make new neighbouring agents, with agents and their neighbours moving synchronously in a growth event (except at the position of growth where two columns/rows will be separated). Therefore,

⁵As in Chapter 4 in Chapter 5 the LPE is calculated between the correlations ODE model and the IBM results.

the spatial correlations associated with agent proliferation (in conjunction with the dilution of agent density due to domain growth) lower the steady-state density of the agents more than the displacement of agents by domain growth counteracts the spatial correlations established by agent proliferation.

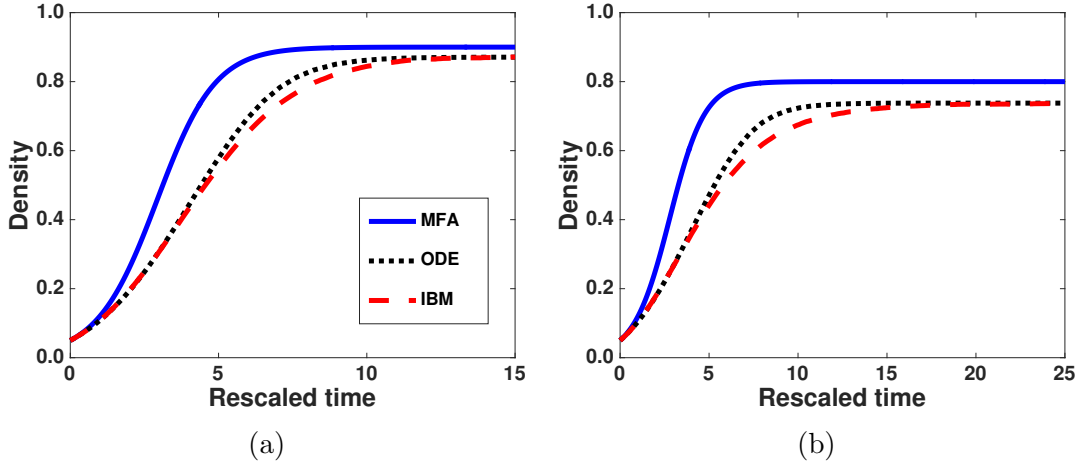


Figure 5.5: For GM1 including pairwise correlations in a correlation ODE model provides a more accurate approximation of the evolution of the macroscopic density in the IBM than the MFA. (a) $P_m = 1$, $P_p = 1$, $P_{g_x} = 0.05$, $P_{g_y} = 0.05$, (b) $P_m = 1$, $P_p = 1$, $P_{g_x} = 0.1$, $P_{g_y} = 0.1$.

In Fig. 5.6 the effect of GM1 on spatial correlations in both the correlation ODE model and the IBM can be seen. Importantly, we see that as the growth rate is increased the spatial correlations augment. From Fig. 5.5 this result is to be expected, as domain growth lowers the steady-state density but is not effective at reducing spatial correlations. A similar result has previously been observed in the case of agent death [18]. In addition, at the steady-state density the spatial correlations do not return to $F = 1$ for the plotted distances. This explains why the MFA incorrectly predicts the agent steady-state density, and shows why the correlation ODE model is able to correctly approximate the steady-state density of the IBM. This is because the MFA is only accurate when $F = 1$, which is not the case in the IBM in this scenario with GM1.

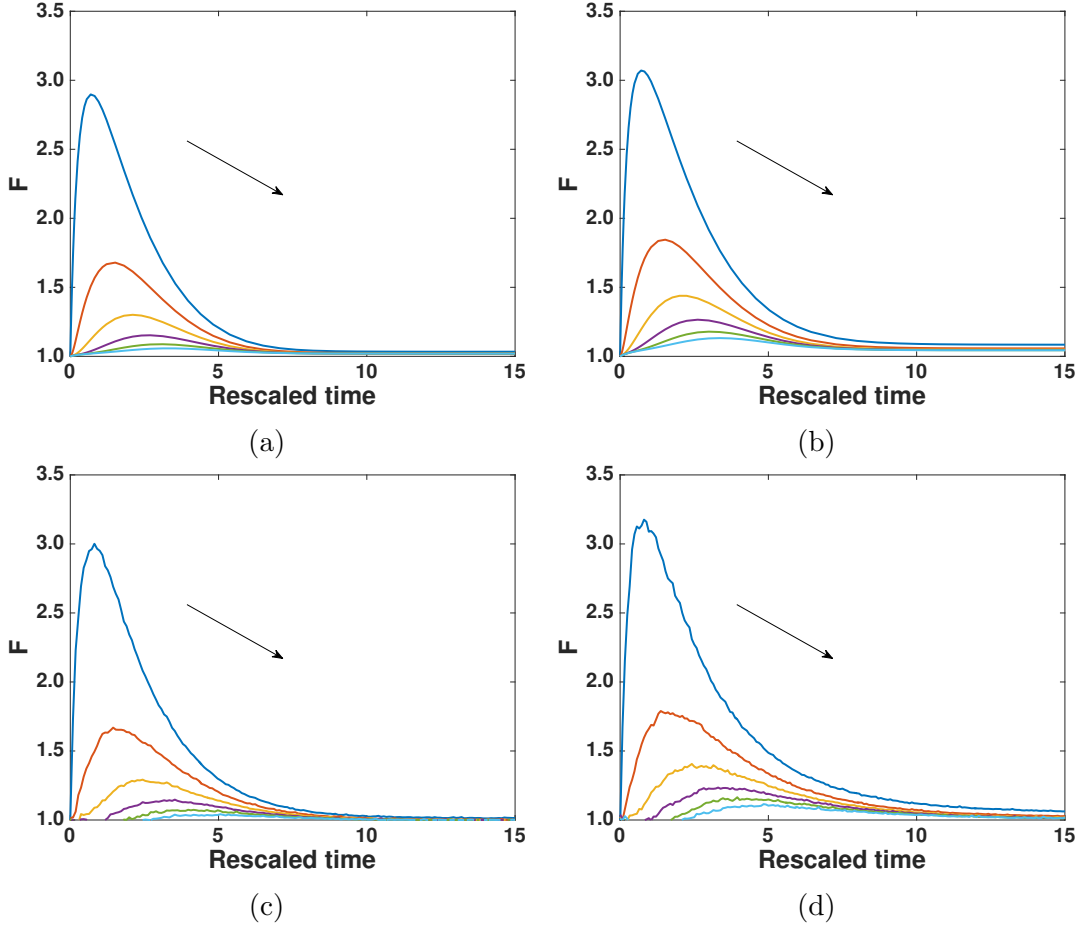


Figure 5.6: Increasing the growth rates, P_{g_x} and P_{g_y} , causes the pairwise correlations to augment for GM1 in the correlation ODE model (compare (a) with (b)), and in the IBM (compare (c) and (d)). In the case of GM1 the pairwise correlations do not return to $F = 1$. In (a)-(d) the distance increases from Δ to 6Δ in steps of Δ . Increasing distance is from top to bottom as indicated by the arrows. The parameters for (a) and (c) are $P_m = 1, P_p = 1, P_{g_x} = 0.05, P_{g_y} = 0.05$, and for (b) and (d) are $P_m = 1, P_p = 1, P_{g_x} = 0.1, P_{g_y} = 0.1$.

In Fig. 5.7 the effect of domain growth via GM2 on agent density can be seen ((a) 0.01, (b) 0.00). In this case we see that with GM2 the steady-state density predicted by the MFA is accurate. This is because GM2 is more effective at breaking up agent clusters than GM1, potentially ‘freeing-up’ one lattice site for four agents in each row/column with every growth event. This counteracts the effect of the spatial correlations created by agent proliferation. GM2 also causes agents to change neighbouring agents in the IBM, and so further reduces the spatial correlations associated with agent proliferation. Despite this we see that the correlations ODE model still more accurately predicts the rate at which the agent density evolves, while also accurately approximating the steady-state density in the IBM.

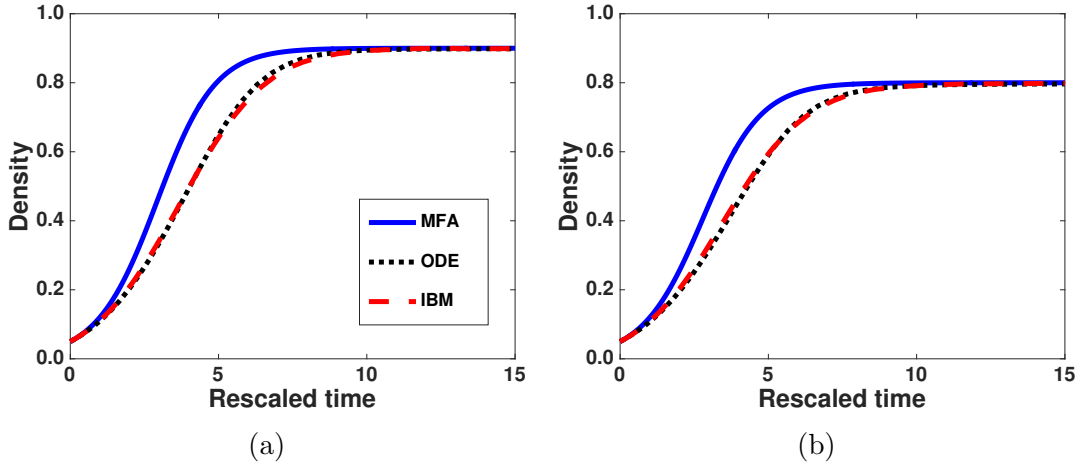


Figure 5.7: For GM2 including pairwise correlations in a correlation ODE model provides a more accurate approximation of the results of the evolution of the macroscopic density in the IBM. (a) $P_m = 1$, $P_p = 1$, $P_{gx} = 0.05$, $P_{gy} = 0.05$, (b) $P_m = 1$, $P_p = 1$, $P_{gx} = 0.1$, $P_{gy} = 0.1$.

In Fig. 5.8 the effect of GM2 on spatial correlations can be seen in both the correlations ODE model and calculated directly from the IBM. In contrast to GM1 we see that as the growth rate is increased the spatial correlations decrease. We also see that at the steady-state density $F = 1$, this provides the reason as to why the agent steady-state predicted by the MFA is correct.

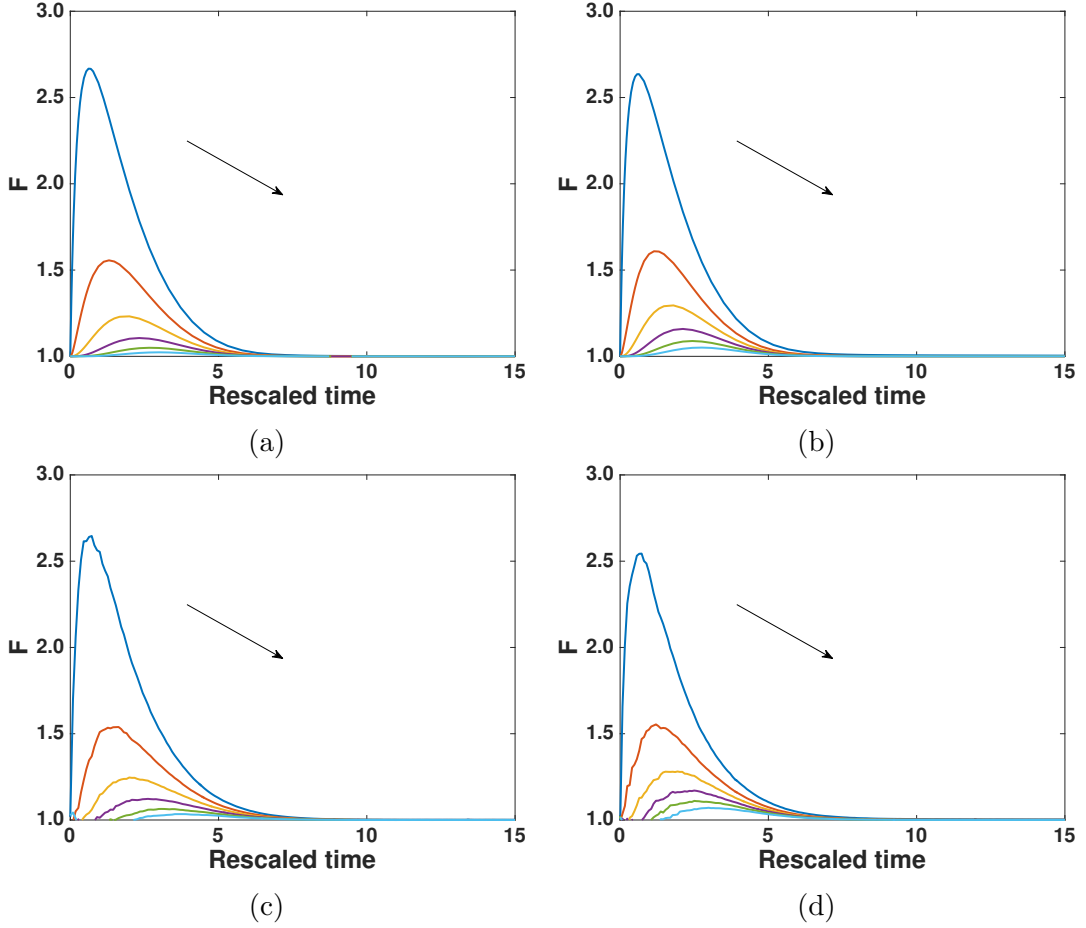


Figure 5.8: For GM2 increasing the growth rates, P_{g_x} and P_{g_y} , causes the pairwise correlations to decrease in the correlation ODE model (compare (a) with (b)), and in the IBM (compare (c) and (d)). In the case of GM2 the pairwise correlations return to $F = 1$. The distance increases from Δ to 6Δ in steps of Δ . Increasing distance is from top to bottom as indicated by the arrows. The parameters for (a) and (c) are $P_m = 1$, $P_p = 1$, $P_{g_x} = 0.05$, $P_{g_y} = 0.05$, and in (b) and (d) are $P_m = 1$, $P_p = 1$, $P_{g_x} = 0.1$, $P_{g_y} = 0.1$.

In Fig. 5.9 we confirm that the size of the domain influences the evolution of the agent density in the case of GM2, but not in the case of GM1. In the case of GM2, as the initial domain size is increased evolution of the macroscopic agent density is accelerated. This is because colinear spatial correlations, established by agent proliferation, are broken down proportional to the domain size in GM2 (Eq. (5.39)).

Finally, in Figs. 5.11 and 5.10 we see the initial domain size also impacts the evolution of the spatial correlations for GM2, but not for GM1. In the case of GM2 as the initial domain size is increased the spatial correlations between agents are reduced.

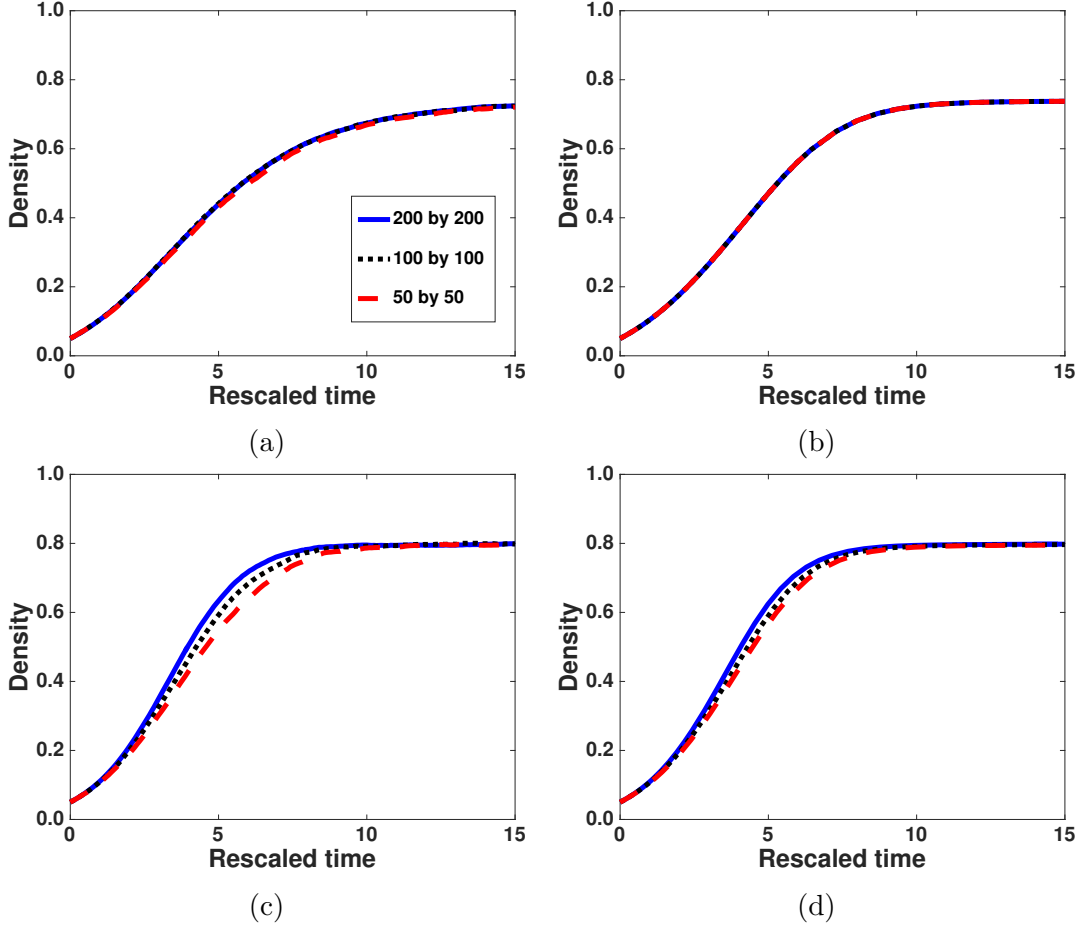


Figure 5.9: Altering the initial size of the domain does not cause the evolution of the agent density to change for GM1, but does for GM2. As the initial size of the domain is increased, from 50 by 50 to 200 by 200 lattice sites, the evolution of the agent density to steady-state remains the same in GM1 as can be seen in both (a) IBM and (b) correlation ODE models. In the case of GM2, altering the initial domain size does cause the evolution of the agent density to change. As the initial size of the domain is increased, from 50 by 50 to 200 by 200 lattice sites, the evolution of the agent density accelerates in GM2, as can be seen in both (c) IBM and (d) correlation ODE models. $P_m = 1$, $P_p = 1$, $P_{g_x} = 0.1$, $P_{g_y} = 0.1$.

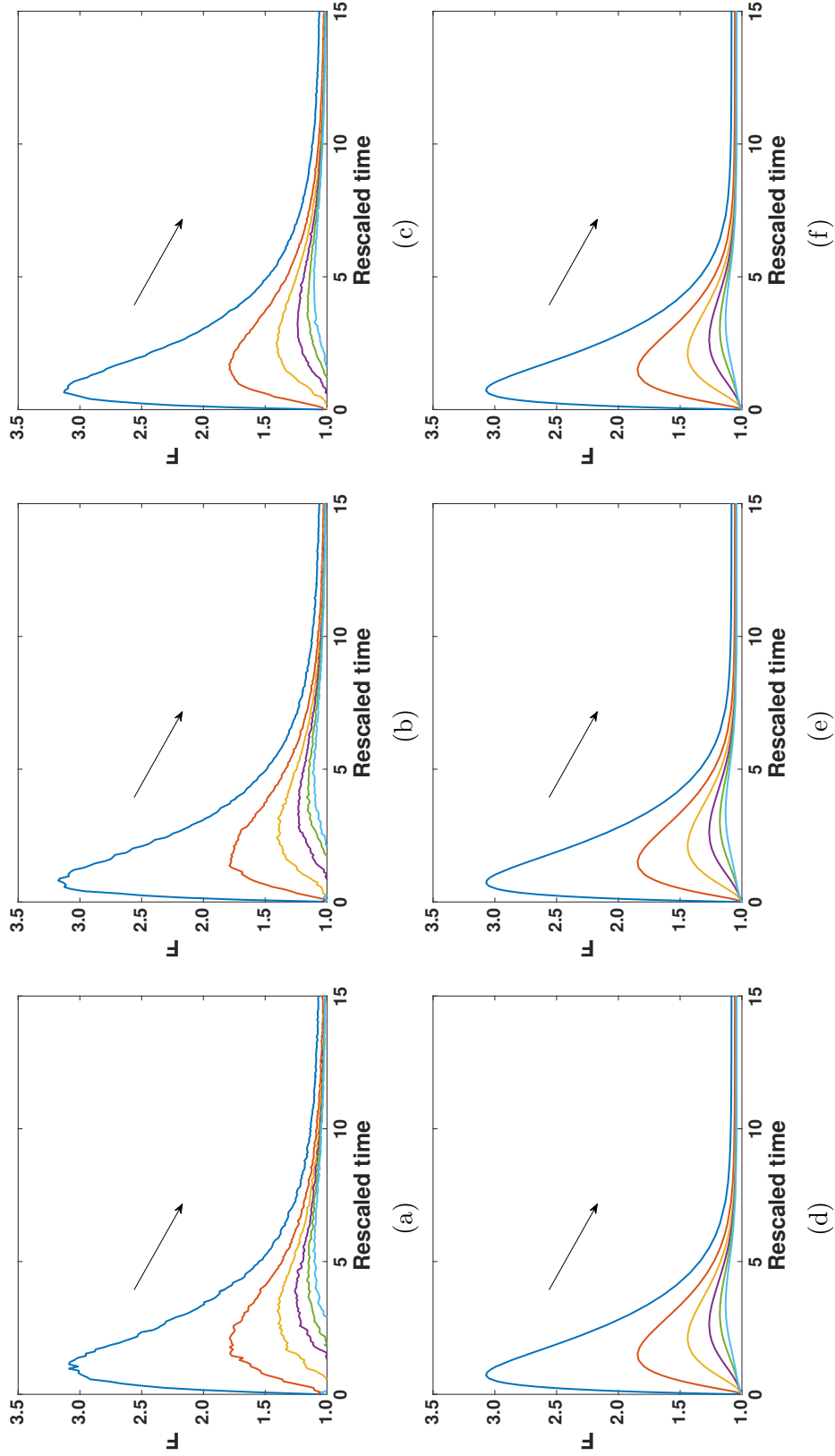


Figure 5.10: Altering the initial size of the domain does not change the evolution of the pairwise correlations in the case of GM1. IBM: (a) 50 by 50, (b) 100 by 100, (c) 200 by 200. Correlation ODE model: (d) 50 by 50, (e) 100 by 100, (f) 200 by 200. Increasing distance is from top to bottom as indicated by the arrows. $P_m = 1$, $P_p = 1$, $P_{gx} = 0.1$, $P_{gy} = 0.1$.

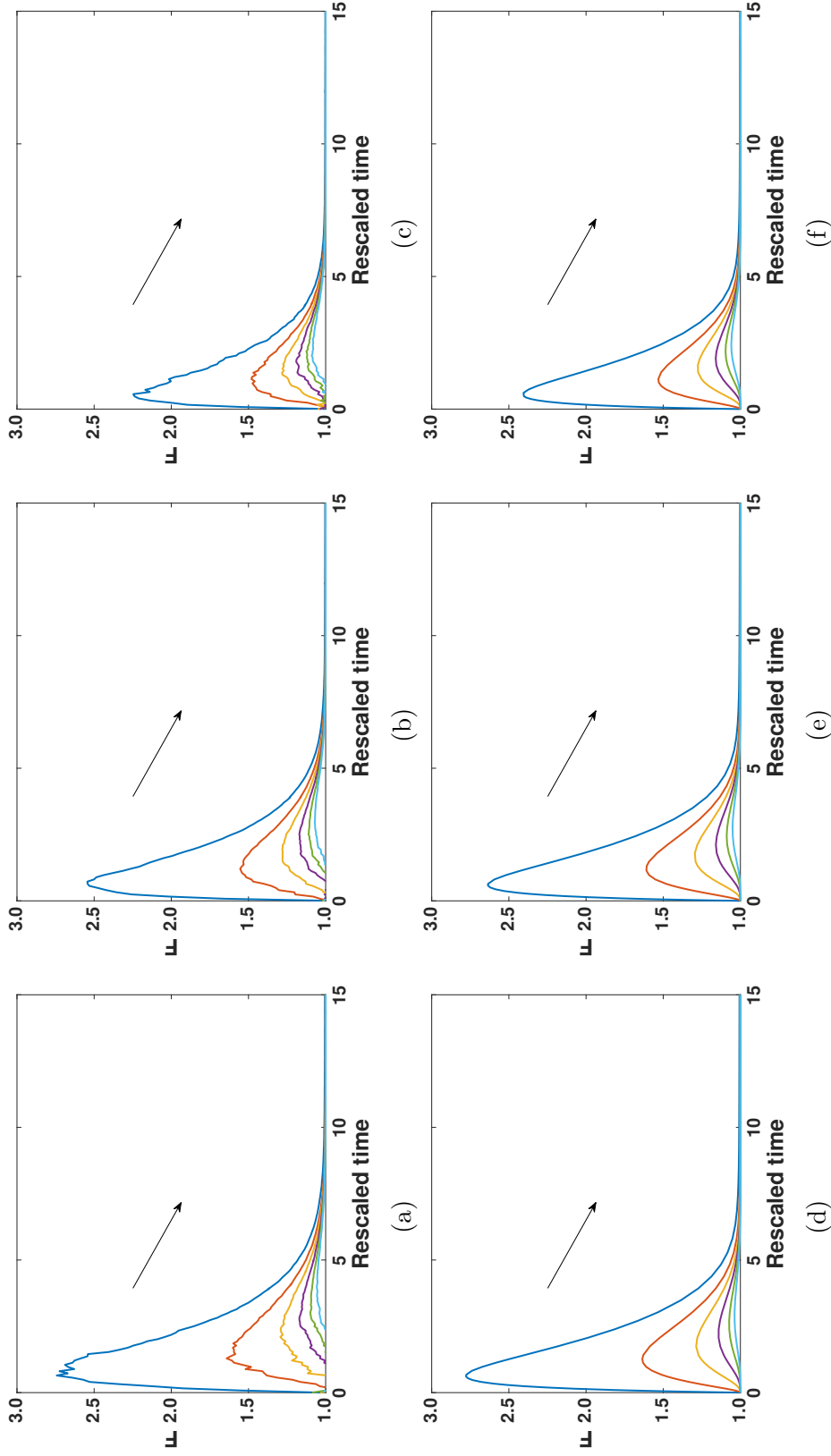


Figure 5.11: Altering the initial size of the domain changes the evolution of the pairwise correlations in the case of GM2. In the case of GM2 the correlations are reduced as the initial domain size becomes larger. IBM: (a) 50 by 50, (b) 100 by 100, (c) 200 by 200. Correlation ODE model: (d) 50 by 50, (e) 100 by 100, (f) 200 by 200. Increasing distance is from top to bottom as indicated by the arrows. $P_m = 1$, $P_p = 1$, $P_{gx} = 0.1$, $P_{gy} = 0.1$.

5.4 Summary

The importance of accurately including the effect of domain growth in models of biological phenomena is becoming increasingly apparent [1, 7, 17, 80, 95–97]. Many of the fundamental developmental processes that determine mammalian morphology are either characterised by growth, or occur during it, and so it is likely that growth plays an important role in determining and coordinating cell population behaviour [10, 98].

In this chapter we have studied two different implementations of domain growth. We chose two different, yet potentially biologically relevant, growth mechanisms [92–94] to highlight how understanding the form of domain growth is crucial. These growth mechanisms originated from thinking about the different ways in which new sites could be added to the IBM domain, and the consequences this may have on agent density. As has been shown in this chapter, GM1 is an implementation of domain growth that conserves colinear correlations in comparison to diagonal correlations, and GM2 is one that reduces colinear correlations as a rate proportional to the size of the domain. This difference between GM1 and GM2 is why the evolution of the agent density for GM2 depends on the initial size of the domain and GM1 does not.

It is important to acknowledge that we have only looked at scenarios where $P_p = P_m$. This case represents a very high rate of cell proliferation relative to motility in a biological context. However, we chose these parameters to highlight the accuracy of our approximate correlation ODE model. In situations where $P_p < P_m$ the correlation ODE model approximates the IBM results even better, as has been shown previously [18, 22]. We have also only looked at the case when $P_{g_x} = P_{g_y}$ in order to reduce the computational complexity of solving the model equations, however, the results presented here can be naturally extended to cases where $P_{g_x} \neq P_{g_y}$.

We wish to stress that the work presented in this chapter is an initial study on the effect of domain growth on spatial correlations between agents and, as such, we have chosen the simplest modelling framework for this. Therefore, an important consideration is whether the work presented here could be extended to other, perhaps more realistic models, such as an off-lattice IBM. Finally, the development of different closure methods as discussed in the summary of Chapter 4 is of equal importance to the work presented in this chapter and the next.

.....

In the next chapter we extend the methods presented in this chapter to an IBM containing two species. Heterogeneous cell populations are widespread in biological systems, and therefore it is important to examine the effects of domain growth on heterogeneous cell populations.

Chapter 6

Modelling spatial correlations on a growing domain in two dimensions with two species

Overview

In this chapter we study the effect of domain growth on spatial correlations in agent populations containing multiple species. This is important as heterogeneous cell populations are numerous during the embryonic development of many species [98]. In the previous chapter we have shown that the long term behaviour of an agent population depends on the way in which domain growth is implemented. We extend this work to show that, depending on the way in which domain growth is implemented, different species dominate in multispecies simulations. Continuum approximations of the lattice-based model that ignore spatial correlations cannot capture this behaviour, while those that explicitly account for spatial correlations can. The results presented here show that the precise mechanism of domain growth can determine the long term behaviour of multispecies populations, and in certain circumstances, establish patterning.

The majority of the work in this chapter appears in the paper: R. J. H. Ross, C. A. Yates and R. E. Baker. “Pattern formation and variable species densities are induced by domain growth”. *Physical Review E* (under review), 2016. R. J. H.

Ross performed all the computational and mathematical modelling presented in this chapter.

6.1 Introduction

Heterogeneous cell populations are widespread throughout biology. Obvious examples in humans include the immune system [99], the brain [100], and the heart [101]. Tumours are often composed of cells that are not phenotypically identical, an important factor that reduces the efficacy of many drug treatments [102]. Cranial neural crest stem cells, a subset of a migratory cell population that give rise to a diverse lineage, exhibit ‘leader’ or ‘follower’ phenotypes during their collective cell migration [1–3]. Similarly melanoblasts, another neural crest stem cell subset, and keratinocytes, simultaneously populate the dorsal lateral epithelium during embryonic development [10]. Importantly, many of the aforementioned examples of heterogeneous cell populations are associated with growing tissues, either during embryonic development [1–3, 10, 100], or in pathological scenarios [101, 102]. Therefore, it is important to be able to include domain growth in models of cell populations containing multiple species where spatial structure plays a significant role [1–3, 10, 21, 26].

In the previous chapter we have shown that the way in which domain growth is implemented in an IBM can alter the population-level behaviour of agents [13, 14]. Therefore, we hypothesised that the way in which domain growth is implemented in a simulation with multiple agent species could change the dominant agent species. We also reasoned that a standard MFA would be insufficient to capture the behaviours exhibited by the IBM. Therefore, as in Chapter 5, the MFA would require correction by the inclusion of spatial correlations in the form of a system of ODEs to accurately approximate the IBM results. This has previously been shown in scenarios without growth. For example, Markham et al. [21] demonstrated the necessity of including the

effect of spatial correlations in continuum models to accurately predict the dominant species in a multispecies context.

The outline of this chapter is as follows. We briefly introduce a two-dimensional IBM containing multiple species in Section 6.2.1. We then define the individual and pairwise density functions, and derive a system of equations (referred to as the correlation ODE model, as in Chapters 4 and 5) describing the evolution of the individual and pairwise density functions for multiple species on a growing domain in Sections 6.2.2-6.2.3. In Section 6.3 we test the accuracy of the correlation ODE model by comparing it with the standard MFA and IBM models for multispecies simulations. We demonstrate that the precise details of the implementation of domain growth can affect agent population fates; a species that dominates in one growth regime might not dominate in the other. We also demonstrate that the MFA is unable to accurately capture the effects of domain growth in the IBM, whereas the correlation ODE models that include the effect of spatial correlations can. Finally, we examine some biologically motivated examples of non-uniform domain growth in our IBM in Section 6.3.2, and show that non-uniform domain growth can cause patterning in multispecies agent populations. We conclude in Section 6.4 with a discussion of our results.

6.2 Model

The IBM we employ in this chapter is the same IBM as in Chapter 5, except that it contains two agent ‘species’. Both growth mechanisms we employ in this chapter are the same as those in Chapter 5 (Fig. 5.1).

6.2.1 IBM containing two species

Throughout this chapter we only present examples with two species, and these species are referred to as A and B . However, all the results presented in this chapter are eas-

ily extendable to scenarios containing more than two species [21]. Attempted agent movement or proliferation events occur with rates P_m^I or P_p^I per unit time, respectively, where I denotes the species type. For example, $P_m^A \delta t$ is the probability of an individual agent of species A attempting to move in the next infinitesimally small time interval δt . Death events occur with rate P_d^I per unit time and result in the removal of an agent from the lattice.

To begin with in this chapter we employ (on average) homogenous initial conditions in our IBM. That is, our initial distribution for both species is achieved by populating a certain number of sites uniformly at random. An occupied site is indicated by A or B , and an unoccupied site is indicated by 0. This means the normalised average agent density for species A on the two-dimensional domain is

$$c_A(t) = \frac{1}{N_x N_y} \sum_{\mathbf{m}}^{N_x N_y} \mathbb{1}_A\{\mathbf{m}\}. \quad (6.1)$$

Here $\mathbb{1}_A$ is the indicator function for species A (i.e. 1 if species A occupies site $\mathbf{m}$, and 0 if it does not). An analogous equation holds for species B .

6.2.2 Individual density functions

The derivation of the evolution equations for the individual density functions is the same as that given in Chapter 5. Therefore, we only state the final equations in this chapter. We define the individual density functions, $\rho_A^{N_x \times N_y}(\mathbf{m}; t)$, as the probability that site $\mathbf{m}$ is occupied by an agent of species A at time t on a domain of size $N_x \times N_y$, where $\mathbf{m}$ is the vector (i, j) , with i indexing the row number of a lattice site, and j indexing the column number of a lattice site. Similarly, $\rho_B^{N_x \times N_y}(\mathbf{m}; t)$ is defined as the probability that site $\mathbf{m}$ is occupied by an agent of species B at time t on a domain of size $N_x \times N_y$.

$$\frac{d\rho_A^{N_x \times N_y}(\mathbf{m}; t)}{dt} = -(P_{g_x} + P_{g_y})\rho_A^{N_x \times N_y}(\mathbf{m}; t). \quad (6.2)$$

Equation (6.2) describes how exponential domain growth affects the evolution of the individual density functions for species A . It is straightforward to derive equations for linear and logistic domain growth analogous to Eq. (6.2)¹.

As in Chapter 5 we will write the pairwise density functions in terms of the displacements between sites, and so to clarify our notation we shall rewrite the individual density functions as

$$\rho_A^{N_x \times N_y}(\mathbf{m}; t) = c_A^{N_x \times N_y}(t) = c_A(t). \quad (6.3)$$

If we substitute Eq. (6.3) into Eq. (6.2) we obtain

$$\frac{dc_A(t)}{dt} = -(P_{g_x} + P_{g_y})c_A(t), \quad (6.4)$$

as expected.

6.2.3 Pairwise density functions

We define the *auto*-correlation pairwise density functions, $\rho_{A,A}^{N_x \times N_y}(\mathbf{m}, \mathbf{n}; t)$, as the probability sites $\mathbf{m}$ and $\mathbf{n}$ are both occupied by species A at time t on a domain of size $N_x \times N_y$ (where $\mathbf{m} \neq \mathbf{n}$). Similarly, the auto-correlation pairwise density function $\rho_{B,B}^{N_x \times N_y}(\mathbf{m}, \mathbf{n}; t)$ is defined as the probability sites $\mathbf{m}$ and $\mathbf{n}$ are both occupied by species B at time t on a domain of size $N_x \times N_y$. The *cross*-correlation pairwise density function $\rho_{A,B}^{N_x \times N_y}(\mathbf{m}, \mathbf{n}; t)$ is the probability sites $\mathbf{m}$ and $\mathbf{n}$ are occupied by

¹The simplest way to do this is to rescale P_{g_x} and P_{g_y} in Eq. (6.2). For instance, the rescaled growth parameter for linear growth in the x direction, $\bar{P}_{g_x}$, is $\bar{P}_{g_x} = (N_x(0)/N_x(t))P_{g_x}$.

species A and B^2 , respectively, at time t on a domain of size $N_x \times N_y$. We now rewrite the pairwise density functions in terms of the displacement vector between lattice sites, that is

$$\rho_{A,A}^{N_x \times N_y}(\mathbf{m}, \mathbf{n}; t) = \rho_{A,A}^{N_x \times N_y}(\mathbf{m}, \mathbf{m} + (r_x, r_y); t). \quad (6.5)$$

As the initial conditions in the IBM are, on average, spatially uniform we are able to assume translational invariance for the probability of two sites a given displacement apart being occupied. This means that the pairwise density function can be written as a function of the displacement between two lattice sites, (r_x, r_y) . Therefore, we will further simplify our notation and write

$$\rho_{A,A}^{N_x \times N_y}(\mathbf{m}, \mathbf{m} + (r_x, r_y); t) = \rho_{A,A}^{N_x \times N_y}(r_x, r_y; t). \quad (6.6)$$

6.2.3.1 Agent motility, proliferation and death

The inclusion of agent motility and proliferation in Eq. (6.4) is similar to the derivation presented in Chapter 5 [18, 21, 22], therefore we refer the reader to Appendix C for the derivation and simply state the result for the individual density functions in the main text. The evolution of the individual density function for a motile and proliferating species A on a growing domain is

$$\frac{dc_A(t)}{dt} = P_p^A \left(c_A(t) - \rho_{A,A}^{N_x \times N_y}(1, 0; t) - \rho_{A,B}^{N_x \times N_y}(1, 0; t) \right) - P_d^A c_A(t) - (P_{g_x} + P_{g_y}) c_A(t). \quad (6.7)$$

An analogous equation exists for species B . As can be seen from Eq. (6.7), the inclusion of agent proliferation means that pairwise density functions are now present

²The ordering of A and B in the subscript of the cross-correlation pairwise density function does not matter in this chapter as we assume translational invariance of the individual and pairwise density functions throughout.

in the equations for the evolution of the individual density functions, which is not the case without agent proliferation (Eq. (6.4)).

6.2.3.2 Growth mechanism 1

We now present the equations for the evolution of the pairwise density functions for GM1 domain growth. These equations have the same derivation as those presented in Chapter 5, and so we only present the final result.

Colinear component

For agents colinear in the horizontal direction, that is, $r_y = 0$, the evolution of the auto-correlation pairwise density functions for species A with GM1 is

$$\begin{aligned} \frac{d\rho_{A,A}^{N_x \times N_y}(r_x, 0; t)}{dt} &= P_{g_x}(r_x - 1)\rho_{A,A}^{N_x \times N_y}(r_x - 1, 0; t) \\ &\quad - P_{g_x}(r_x + 1)\rho_{A,A}^{N_x \times N_y}(r_x, 0; t) \\ &\quad - P_{g_y}\rho_{A,A}^{N_x \times N_y}(r_x, 0; t). \end{aligned} \quad (6.8)$$

An analogous equation exists for sites colinear in the vertical direction for GM1³.

Diagonal component

For the diagonal component, that is, $r_x, r_y \neq 0$, we have⁴

$$\begin{aligned} \frac{d\rho_{A,A}^{N_x \times N_y}(r_x, r_y; t)}{dt} &= P_{g_x}(r_x - 1)\rho_{A,A}^{N_x \times N_y}(r_x - 1, r_y; t) \\ &\quad - P_{g_x}(r_x + 1)\rho_{A,A}^{N_x \times N_y}(r_x, r_y; t) \\ &\quad + P_{g_y}(r_y - 1)\rho_{A,A}^{N_x \times N_y}(r_x, r_y - 1; t) \\ &\quad - P_{g_y}(r_y + 1)\rho_{A,A}^{N_x \times N_y}(r_x, r_y; t). \end{aligned} \quad (6.9)$$

³These are Eqs. (5.16) and (5.17) in Chapter 5.

⁴This is Eq. (5.18) in Chapter 5.

6.2.3.3 Growth mechanism 2

We now present the equations for the evolution of the pairwise density functions for GM2 domain growth. Like GM1 these equations have the same derivation as those presented in Chapter 5, and so we only present the final result. The evolution equation for the colinear component (horizontally colinear) for GM2 is⁵

$$\begin{aligned}
\frac{d\rho_{A,A}^{(N_x+1)\times(N_y+1)}(r_x, 0; t)}{dt} = & -P_{g_x}(r_x + 1)\rho_{A,A}^{(N_x+1)\times(N_y+1)}(r_x, 0; t) \\
& + P_{g_x}(r_x - 1)\rho_{A,A}^{(N_x+1)\times(N_y+1)}(r_x - 1, 0; t) \\
& + P_{g_y}\left(-1 - \frac{N_y}{3} + \frac{1}{3N_y}\right)\rho_{A,A}^{(N_x+1)\times(N_y+1)}(r_x, 0; t) \\
& + P_{g_y}\left(-1 + \frac{N_y}{3} + \frac{2}{3N_y}\right)\rho_{A,A}^{(N_x+1)\times(N_y+1)}(r_x, -1; t).
\end{aligned}
\tag{6.10}$$

⁵This is Eq. (5.33) in Chapter 5.

An analogous equation exists for the vertically colinear component for GM2. The equation for the evolution of the diagonal component for GM2 is⁶

$$\begin{aligned}
\frac{d\rho_{A,A}^{(N_x+1)\times(N_y+1)}(r_x, r_y; t)}{dt} = & \\
& P_{g_x} \left(-1 - \frac{N_x}{3} + r_x \left(\frac{r_x}{N_x} - 1 \right) + \frac{1}{3N_x} \right) \rho_{A,A}^{(N_x+1)\times(N_y+1)}(r_x, r_y; t) \\
& + P_{g_y} \left(-1 - \frac{N_y}{3} + r_y \left(\frac{r_y}{N_y} - 1 \right) + \frac{1}{3N_y} \right) \rho_{A,A}^{(N_x+1)\times(N_y+1)}(r_x, r_y; t) \\
& + P_{g_y} \left(-\frac{1}{2} + \frac{N_y}{6} - \frac{r_y}{2} \left(\frac{r_y}{N_y} - 1 \right) + \frac{1}{3N_y} \right) \rho_{A,A}^{(N_x+1)\times(N_y+1)}(r_x, r_y - 1; t) \\
& + P_{g_y} \left(-\frac{1}{2} + \frac{N_y}{6} - \frac{r_y}{2} \left(\frac{r_y}{N_y} - 1 \right) + \frac{1}{3N_y} \right) \rho_{A,A}^{(N_x+1)\times(N_y+1)}(r_x, r_y + 1; t) \\
& + P_{g_x} \left(-\frac{1}{2} + \frac{N_x}{6} - \frac{r_x}{2} \left(\frac{r_x}{N_x} - 1 \right) + \frac{1}{3N_x} \right) \rho_{A,A}^{(N_x+1)\times(N_y+1)}(r_x - 1, r_y; t) \\
& + P_{g_x} \left(-\frac{1}{2} + \frac{N_x}{6} - \frac{r_x}{2} \left(\frac{r_x}{N_x} - 1 \right) + \frac{1}{3N_x} \right) \rho_{A,A}^{(N_x+1)\times(N_y+1)}(r_x + 1, r_y; t).
\end{aligned} \tag{6.11}$$

6.3 Results

As in Chapters 4 and 5 we present our results in terms of correlation functions [78, 84–87] in order to simplify the visualisation of results, and allow the results presented here to be easily related to other research in this field [18, 19, 21–23]. The correlation function is defined as

$$F_{A,A}(r_x, r_y; t) := \frac{\rho_{A,A}^{N_x \times N_y}(r_x, r_y; t)}{c_A(t)^2}, \tag{6.12}$$

and is simply a measure of the degree to which the occupancies of two lattice sites by species A are independent of one another. Analogous correlation functions exist for auto-correlations in species B and for the cross-correlations between species A and

⁶This is Eq. (5.39) in Chapter 5.

B :

$$F_{B,B}(r_x, r_y; t) := \frac{\rho_{B,B}^{N_x \times N_y}(r_x, r_y; t)}{c_B(t)^2}, \quad (6.13)$$

and

$$F_{A,B}(r_x, r_y; t) := \frac{\rho_{A,B}^{N_x \times N_y}(r_x, r_y; t)}{c_A(t)c_B(t)}. \quad (6.14)$$

If we substitute Eqs. (6.12) and (6.14) into Eq. (6.7) we obtain

$$\begin{aligned} \frac{dc_A(t)}{dt} = & P_p^A c_A(t) \left(1 - F_{A,A}(1, 0; t) c_A(t) - F_{A,B}(1, 0; t) c_B(t) \right) \\ & - P_d^A c_A(t) - (P_{g_x} + P_{g_y}) c_A(t). \end{aligned} \quad (6.15)$$

We refer to Eq. (6.15) as the correlation ODE model. The standard MFA assumes $F_{A,A}(r_x, r_y; t) = F_{A,B}(r_x, r_y; t) = 1$, that is, spatial correlations between agents are insignificant, and so Eq. (6.15) becomes

$$\frac{dc_A(t)}{dt} = P_p^A c_A(t) (1 - c_A(t) - c_B(t)) - P_d^A c_A(t) - (P_{g_x} + P_{g_y}) c_A(t). \quad (6.16)$$

As discussed in Chapters 4 and 5, Eq. (6.16) is relevant as it represents the standard MFA often used to model the evolution of the macroscopic density of a cell population [12, 34, 37]. However, in certain scenarios the standard MFA has been shown to be inadequate [18], especially when the spatial structure of a cell population is known to be important. As such, we will compare Eqs. (6.15) and (6.16) in the results section.

For our IBM simulations we use a regular square lattice of initial size 100 by 100 lattice sites. The boundary conditions are periodic, and we have an initial uniform random seeding of density 0.05 for each species A and B (so the total agent density

is 0.1). By an initial uniform random seeding it is meant that, on average, the initial conditions of the IBM are spatially uniform for both species. All results presented from the IBM are ensemble averages taken from 500 repeats. To solve Eqs. (6.7)-(6.11) we use MATLAB's `ode15s`, with an absolute error tolerance of 10^{-12} .

Our initial condition for all simulations entails that site occupancies at all pairwise distances are initially uncorrelated, that is,

$$F_{A,A}(r_x, r_y; 0) = F_{B,B}(r_x, r_y; 0) = F_{A,B}(r_x, r_y; 0) = 1. \quad (6.17)$$

Initially we study the effect of exponential domain growth for both GM1 and GM2 on agent density and spatial correlations in the IBM and correlation ODE model. As previously stated, N_x and N_y are integers that describe the number of lattice sites in the horizontal and vertical directions, respectively. However, as results from the IBM are ensemble averages we replace N_x and N_y with their continuum analogues $L_x(t)$ and $L_y(t)$, respectively. This substitution of N_x and N_y with $L_x(t)$ and $L_y(t)$ avoids jump discontinuities in the numerical solutions of Eqs. (6.10)-(6.11), which are not present in the averaged IBM results. For exponential domain growth in the IBM $L_x(t)$ can be shown to evolve according to

$$\frac{dL_x(t)}{dt} = P_{g_x} L_x(t). \quad (6.18)$$

An analogous equation exists for $L_y(t)$.

We also study the effect of linear and logistic domain growth (for both GM1 and GM2) on agent density and spatial correlations in the IBM and correlation ODE model. For simulations with logistic domain growth the individual density function

for species A evolves according to

$$\begin{aligned} \frac{dc_A}{dt} = & P_p^A c_A \left(1 - F_{A,A}(1, 0; t) c_A(t) - F_{A,B}(1, 0; t) c_B(t) \right) - P_d^A c_A(t) \\ & - (P_{g_x} + P_{g_y}) \left(1 - \frac{L_x(t) L_y(t)}{K^2} \right) c_A(t), \end{aligned} \quad (6.19)$$

where K^2 is the carrying capacity and $L_x(t)$ evolves according to

$$\frac{dL_x(t)}{dt} = P_{g_x} L_x(t) \left(1 - \frac{L_x(t) L_y(t)}{K^2} \right). \quad (6.20)$$

An analogous equation exists for $L_y(t)$, and for all simulations presented here $K = 300$. Finally, the individual density function for species A with linear domain growth evolves according to

$$\begin{aligned} \frac{dc_A}{dt} = & P_p^A c_A (1 - F_{A,A}(1, 0; t) c_A(t) - F_{A,B}(1, 0; t) c_B(t)) - P_d^A c_A(t) \\ & - \left(\frac{P_{g_x} L_x(0)}{L_x(t)} + \frac{P_{g_y} L_y(0)}{L_y(t)} \right) c_A(t), \end{aligned} \quad (6.21)$$

where $L_x(t)$ evolves according to

$$\frac{dL_x(t)}{dt} = P_{g_x} L_x(0), \quad (6.22)$$

that is, linear growth. As in Chapter 5 we also rescale time to allow for ease of comparison between simulations with different parameters:

$$\bar{t} = \left(P_p - P_d - (P_{g_x} + P_{g_y}) \right) t. \quad (6.23)$$

6.3.1 Uniform domain growth

In Fig. 6.1 we recapitulate results from [21] ((a) $LPE = 0.00$, (b) 0.05)⁷. We see that in the case of an IBM with a non-growing domain and two species, a more motile, slower proliferating species (species A) can dominate a less motile, faster proliferating species (species B) in a specific parameter regime. This is the case on a non-growing domain without agent death, as can be seen in Fig. 6.1 (a), and the dominance of species A is augmented with agent death as evident in Fig. 6.1 (b), whereby species B eventually goes extinct. Importantly, the standard MFA (Eq. (6.16)) is not able to accurately approximate the IBM results in either case, whereas the correlations ODE model is able to. The reason as to why species A dominates over species B is because of the following: although species B has a higher proliferation rate, its motility is much lower than species A . This means the spatial correlations species B establishes (with itself) by its own proliferation causes future proliferation events to be aborted by those same agents. This does not happen to species A , to the same extent, as its higher motility rate means the spatial correlations established by its own proliferation are broken down more quickly.

Figure 6.2 (a) shows that domain growth implemented via GM1 has a similar effect on agent density as agent death in Fig. 6.1 (b), allowing species A to dominate (although species B does not become extinct in this case as there is no agent death) (0.08). In Fig. 6.2 (b) we can see that domain growth implemented via GM2 has the opposite effect, and enables species B to dominate (0.04). This is because GM2 breaks up colinear correlations (correlations between agents that share the same row or column) at a rate proportional to the size of the domain. This means species B , which is more affected by colinear correlations due to its higher proliferation and

⁷As in Chapters 4 and 5 in Chapter 6 the LPE is calculated between the correlations ODE model and the IBM results. The largest LPE out of species A and B is presented.

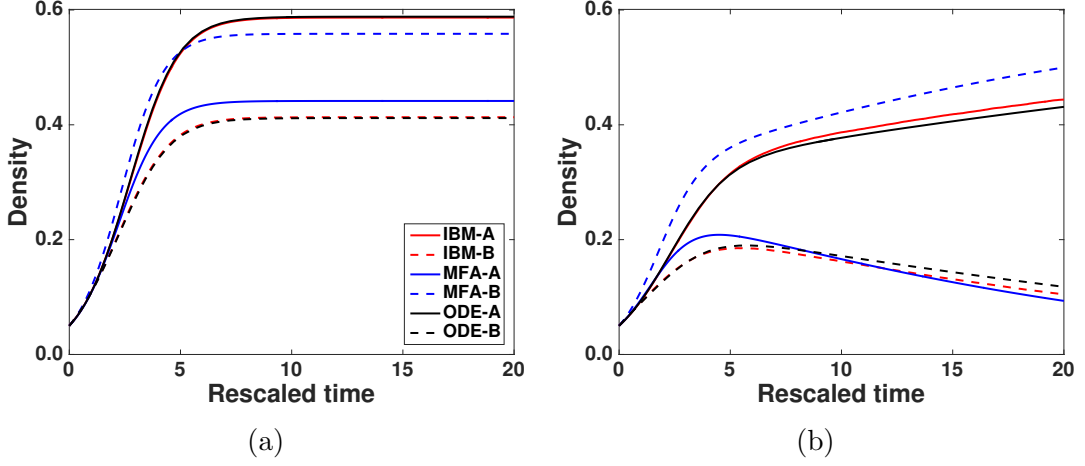


Figure 6.1: Including the effects of pairwise correlations renders the correlations ODE model (Eq. (6.15)) able to accurately approximate the averaged results from the IBM, whereas the standard MFA (Eq. (6.16)) cannot. In (a) the parameters are $P_m^A = 20$, $P_p^A = 0.9$, $P_d^A = 0$, $P_m^B = 1$, $P_p^B = 1$ and $P_d^B = 0$. In (b) the parameters are $P_m^A = 20$, $P_p^A = 0.9$, $P_d^A = 0.4$, $P_m^B = 1$, $P_p^B = 1$ and $P_d^B = 0.4$.

lower motility rates compared to species A , increases in ‘fitness’ as the domain grows. Importantly, the standard MFA is not able to capture the GM1 results, whereas the correlations ODE model is. In the GM2 scenario, the MFA ultimately predicts the correct dominant species, but the correlations ODE model more accurately predicts the temporal evolution of the system to $\bar{t} = 20$.

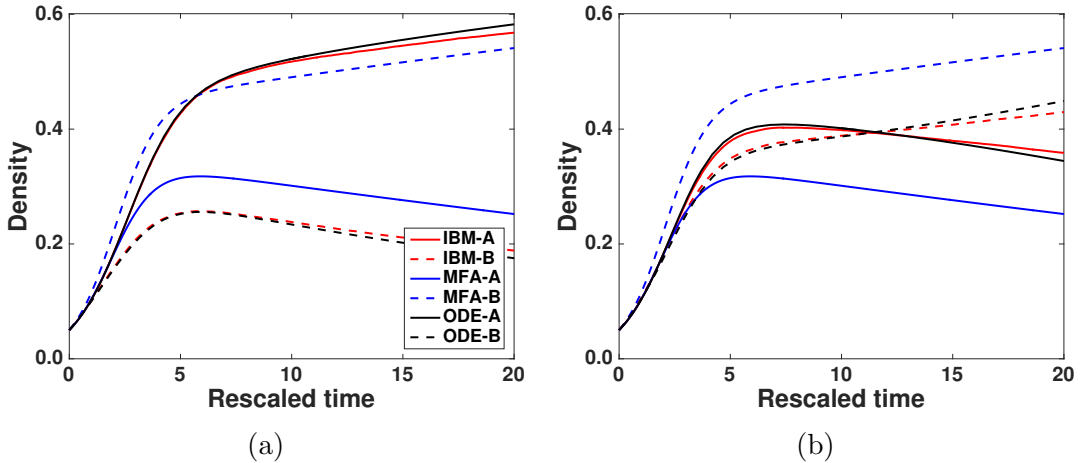


Figure 6.2: Including the effects of pairwise correlations renders the correlations ODE model (Eq. (6.15)) able to accurately approximate the averaged results from the IBM, whereas the standard MFA (Eq. (6.16)) cannot. The parameters for both panels (a) GM1 and (b) GM2 are $P_m^A = 20$, $P_p^A = 0.9$, $P_m^B = 1$, $P_p^B = 1$, $P_{gx} = 0.1$ and $P_{gy} = 0.1$.

In Fig. 6.3 it can be seen that the inclusion of domain growth has a different effect on spatial correlations depending on the type of growth mechanism implemented. We see in Fig. 6.3 a good agreement between the spatial correlations computed from the correlation ODE model and those calculated from the IBM. For GM1 in Fig. 6.3 (a) the species A auto-correlations decrease as species A begins to dominate (long term behaviour). This is because as species A begins to dominate its density becomes increasingly spatially uniform. In Fig. 6.3 (b) we see that species B auto-correlations increase as species B becomes less spatially uniform, due to proliferation. Meanwhile in Fig. 6.3 (c) we see that the cross-correlations decrease as species A begins to dominate.

With GM2 we see that domain growth has a different effect on spatial correlations, as shown in Fig. 6.4 (d)-(e). Again we see a good agreement between the spatial correlations predicted by the correlation ODE model and those calculated from the IBM. It can be seen that with GM2 the auto-correlations for both species A and B decrease as the domain grows. In Fig. 6.4 (f) we see that the cross-correlations between agents decrease with increasing distance, and note that this is because GM2 breaks up spatial correlations between agents more effectively than GM1.

In Fig. 6.5 we see that the initial size of the domain influences the evolution of the agent density in the case of GM2, but not in the case of GM1 with exponential domain growth ((a) 0.08, (b) 0.08, (c) 0.08, (d) 0.04, (e) 0.04, (f) 0.08). In the case of GM2, as the initial domain size is increased the evolution of the macroscopic agent densities is accelerated, i.e. species B begins to dominate at an earlier time. This is because colinear spatial correlations established by agent proliferation, which affect species B more significantly than species A , are broken down at a rate proportional to the domain size in GM2. This means the relative ‘fitness’ of species B increases

as the domain grows.

Figure 6.6 (a)-(b) shows the results from the same two-species scenario with linear domain growth ((a) 0.04, (b) 0.03). As before we see that with GM1 species A dominates. The correlation ODE model is able to capture this behaviour while the standard MFA is not. In the case of GM2 we see that species A initially dominates, but as the domain grows species B begins to increase relative to A (and the density of species B will exceed the density of species A at a later time). Finally, Fig. 6.6 (c) and (d) shows the results from the same two-species scenario with logistic domain growth ((c) 0.04, (d) 0.02). In this case we see that species A dominates with both GM1 and GM2. This is because the domain stops growing when the domain size carrying capacity is reached.

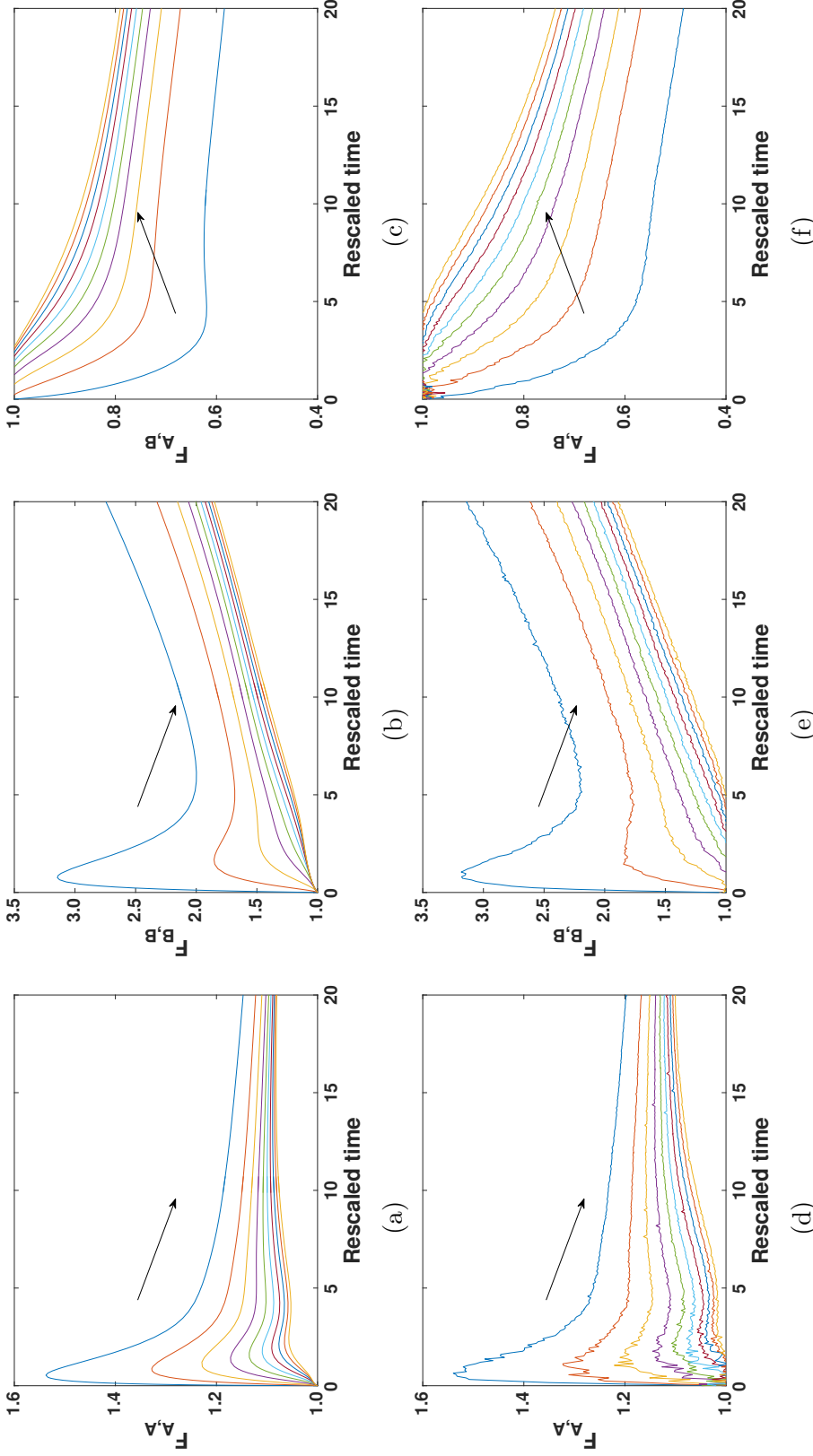


Figure 6.3: (a)-(c) The evolution of the correlation functions for $F_{A,A}$, $F_{B,B}$ and $F_{A,B}$ with exponential growth via GM1. The distance plotted increases from Δ to 10Δ in steps of Δ . Increasing distance is from top to bottom in (a) and (b), i.e. blue line $= \Delta$, red line $= 2\Delta$, and so on. In (c) distance increases from bottom to top. In (d)-(f) the correlation functions for $F_{A,A}$, $F_{B,B}$ and $F_{A,B}$ are calculated from ensemble averages from the IBM. In (a)-(b) and (d)-(e) the auto-correlation functions are greater than or equal to unity, indicating that lattice sites are more likely to be occupied by a particular species if there are others nearby. The lower movement rate of species B leads to it having larger auto-correlations, and these auto-correlations grow as species A begins to dominate. In contrast, the auto-correlations of species A decrease as it begins to dominate. The parameters for all panels are $P_m^A = 20$, $P_p^A = 0.9$, $P_m^B = 1$, $P_p^B = 1$, $P_{ge} = 0.1$ and $P_{gy} = 0.1$.

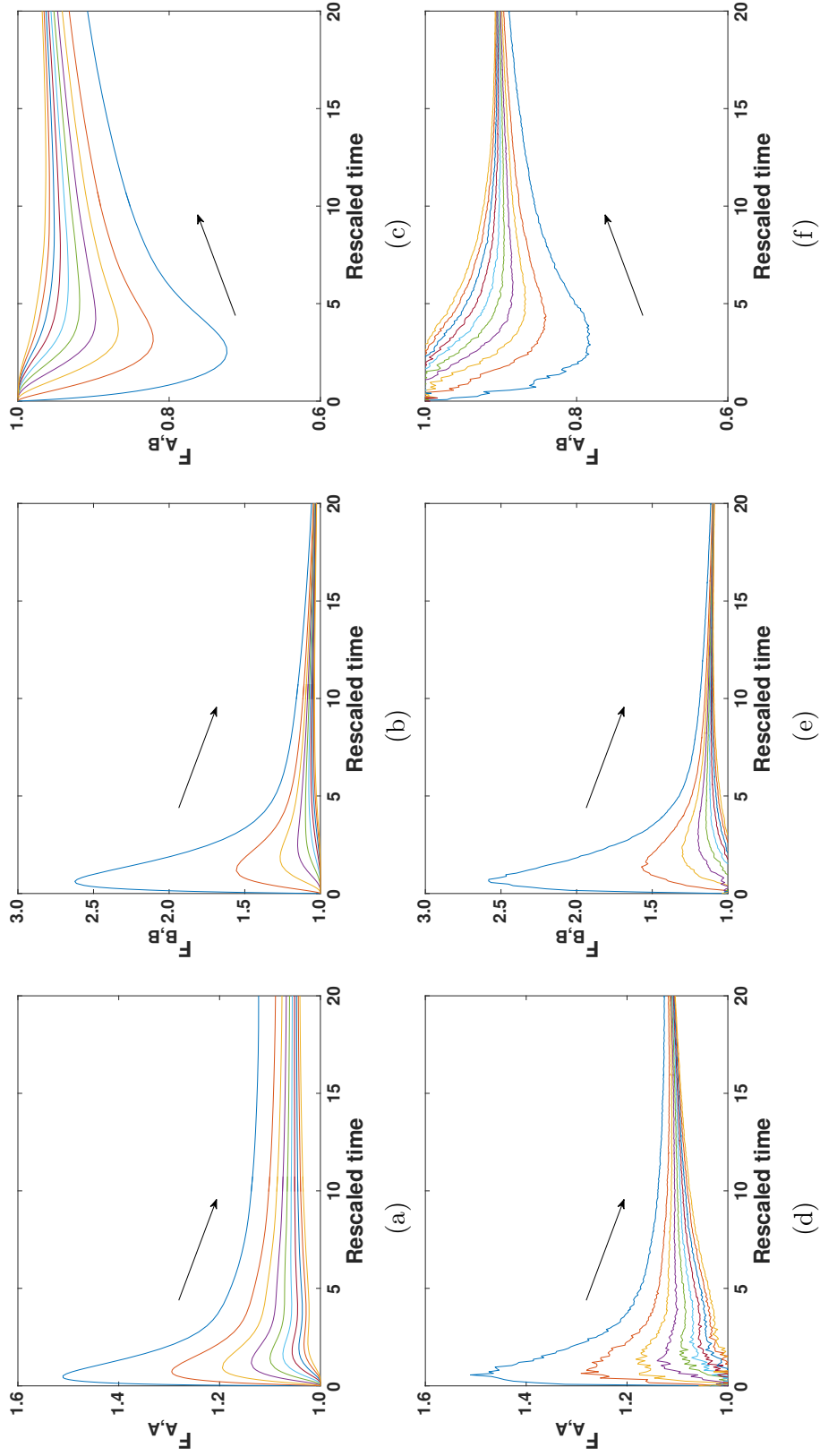


Figure 6.4: (a)-(c) The evolution of the correlation functions for $F_{A,A}$, $F_{B,B}$ and $F_{A,B}$ with exponential growth via GM2. In (d)-(f) the correlation functions for $F_{A,A}$, $F_{B,B}$ and $F_{A,B}$ are calculated from ensemble averages from the IBM. See Fig. 6.3 for further details. The lower movement rate of species B leads to species B having larger auto-correlations, and these auto-correlations decrease as species B begins to dominate. The parameters for all panels are $P_m^A = 20$, $P_p^A = 0.9$, $P_m^B = 1$, $P_p^B = 1$, $P_{gx} = 0.1$ and $P_{gy} = 0.1$.

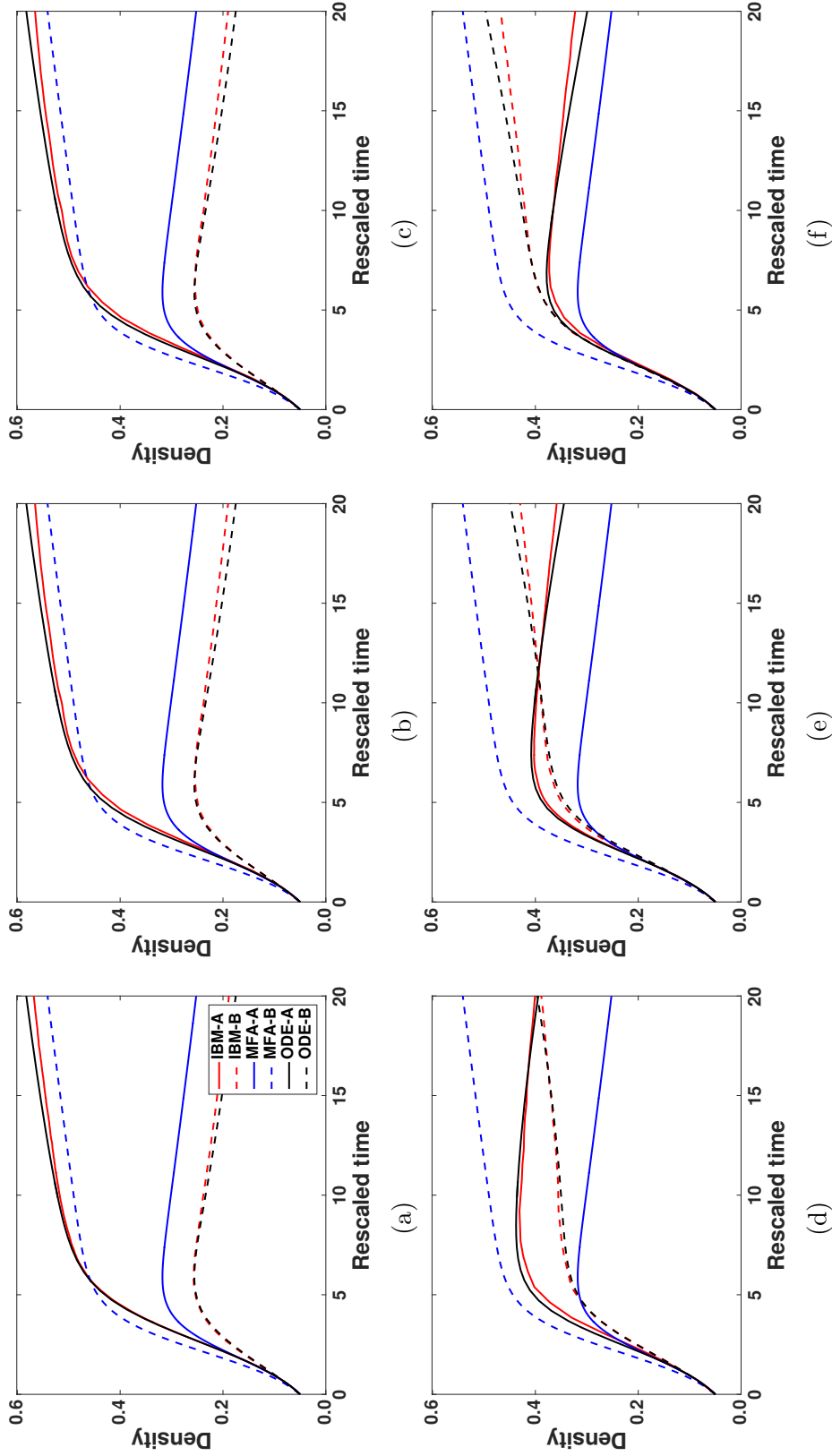


Figure 6.5: Altering the initial domain size causes the evolution of both species to change in GM2 but not in GM1. As the initial size of the domain is increased the dominance of species B in GM2 occurs earlier. GM1: (a) 50 by 50, (b) 100 by 100, (c) 200 by 200. GM2: (d) 50 by 50, (e) 100 by 100, (f) 200 by 200. The parameters for all panels are $P_m^A = 20$, $P_p^A = 0.9$, $P_m^B = 1$, $P_p^B = 1$, $P_{gy} = 0.1$ and $P_{gw} = 0.1$.

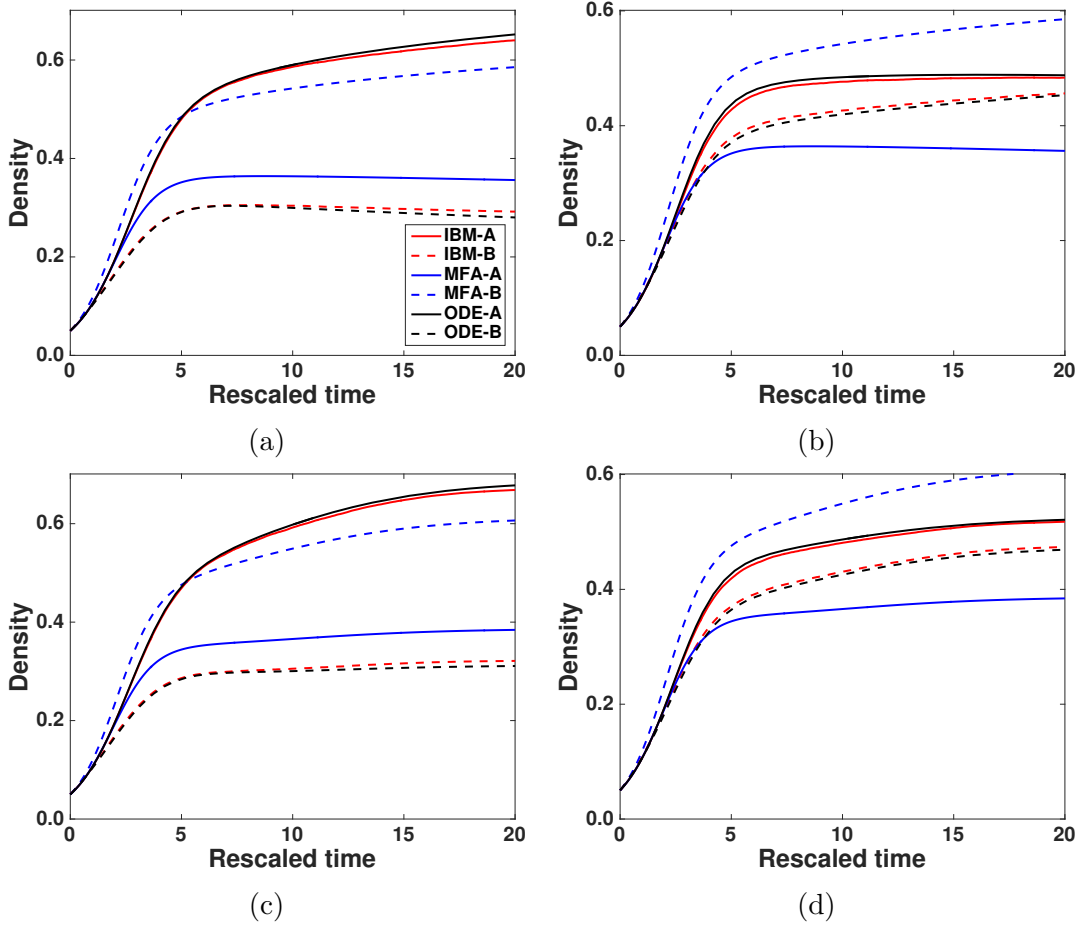


Figure 6.6: Including the effects of pairwise correlations renders the correlations ODE model (Eqs. (6.19) and (6.21)) able to accurately approximate the averaged results from the IBM, whereas the standard MFA cannot. (a) GM1 linear domain growth, (b) GM2 linear domain growth, (c) GM1 logistic domain growth, (d) GM2 logistic domain growth. The parameters for all panels are $P_m^A = 20$, $P_p^A = 0.9$, $P_m^B = 1$, $P_p^B = 1$, $P_{gx} = 0.1$ and $P_{gy} = 0.1$.

6.3.2 Non-uniform domain growth

To conclude our results we study some biologically motivated examples of non-uniform domain growth. In these examples we only present results from the IBM. However, we do so aware of the differing effects that GM1 and GM2 have on the evolution of agent density due to the previous work we have carried out in this thesis. We hypothesise that, given the same motility and proliferation parameters for species A and B as in Fig. 6.2, in a two-species scenario non-uniform domain growth could enable species B to dominate in a faster growing region of the domain, while species A could dominate in a slower growing (or non-growing) region of the domain. This means non-uniform domain growth could lead to pattern formation in simulations containing two species.

We choose two ‘canonical’ examples of domain growth that serve to represent idealised versions of known growth mechanisms in biological systems. The first example we term ‘apical’. We use apical to mean domain growth localised to one end of the domain. This type of growth has been observed in root growth and embryonic limb development [103, 104]. For our apical example domain growth is uniform in the vertical direction (we look at both exponential and linear domain growth), while in the horizontal direction growth is restricted to the right-most half of the x -axis, as displayed in Fig. 6.7 (a). This half of the domain is updated throughout the simulation as the IBM domain grows. That is, this half of the domain is recalculated as the IBM grows to include new lattice sites.

The second example we term ‘enteric’. Enteric growth, that is, intestinal growth, is associated with different regions of the intestine growing at different rates [24]. For our enteric example domain growth is again uniform in the vertical direction, while in the horizontal direction ninety percent of the growth events are restricted to the middle third of the x -axis, as displayed in Fig. 6.7 (b) (this ‘third’ of the domain is

updated throughout the simulation as the IBM domain grows). The other ten percent of growth events are distributed uniformly amongst the two remaining regions.

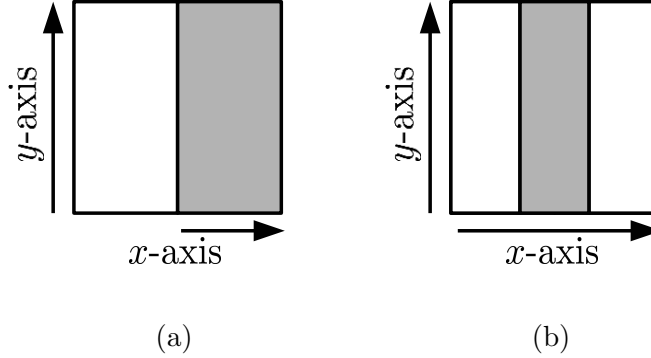


Figure 6.7: The non-uniform domain growth examples. For apical growth (a) all growth events in the horizontal direction occur in the region indicated by the grey zone, this zone is fifty percent of the lattice in the x -axis, and its position is updated throughout the simulation. For enteric growth (b) ninety percent of growth events in the horizontal direction occur in the region indicated by the grey zone with the other ten percent distributed equally between the two other regions, this zone is middle third of the lattice in the x -axis, and its position is updated throughout the simulation. For both examples growth in the vertical direction (y -axis) is uniform.

For both of these growth mechanisms we implement no-flux boundary conditions in the x direction, and periodic boundary conditions in the y direction. With these boundary conditions the IBM domain can be thought of as a cylinder, and could therefore represent a growing root or the developing intestine [24]. We use no-flux boundaries to augment the differences between GM1 and GM2 on the density of agents in apical growth [103, 104]. We only present results for linear and exponential domain growth in a two-species scenario, and in all simulations the domain grows to a horizontal length of 1500 lattice sites in the x -axis before the simulation is terminated. Both agent species are, on average, initially placed uniformly at random at densities of 0.05 (giving a total initial agent density of 0.1). All figures presented in this section are column averages taken from 1000 IBM repeats.

In Fig. 6.8 (a) and (b) density profiles for exponential apical growth are shown.

In Fig. 6.8 (a) GM1 is implemented and this enables species A to dominate across the domain. However, in Fig. 6.8 (b) GM2 is implemented and this enables species A and B to have similar densities in the middle region (grey zone, Fig. 6.7 (b)) of the domain. Similarly, in Fig. 6.8 (c) and (d), density profiles for linear apical growth are

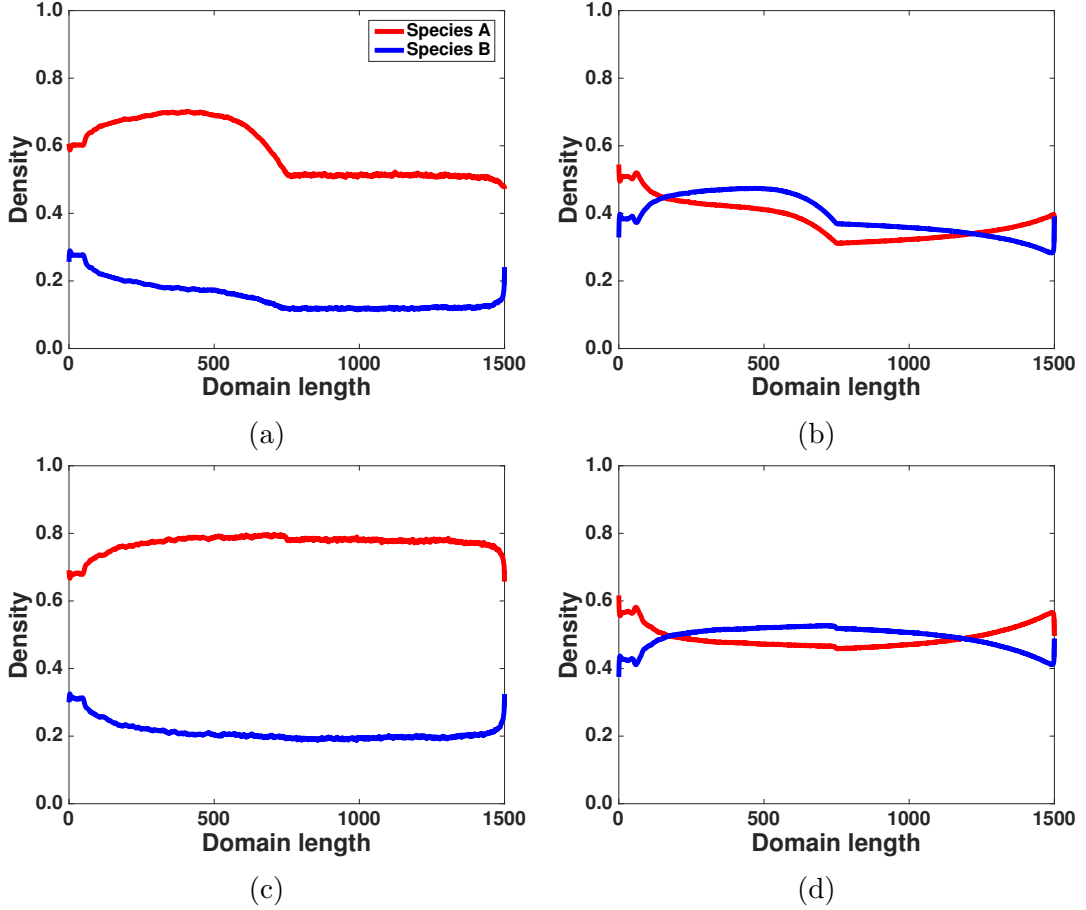


Figure 6.8: Apical domain growth. (a) GM1 exponential domain growth, (b) GM2 exponential domain growth, (c) GM1 linear domain growth, (d) GM2 linear domain growth. The parameters for all panels are $P_m^A = 20$, $P_p^A = 0.9$, $P_m^B = 1$, $P_p^B = 1$, $P_{gx} = 0.1$ and $P_{gy} = 0.1$.

shown. In Fig. 6.8 (c) GM1 is implemented and, much like exponential growth, this results in species A dominating across the domain. In contrast, in Fig. 6.8 (d) GM2 is implemented and this results in species A and B having similar densities in the middle region of the growing domain. This result suggests that apical growth can determine the dominant species at the interface of two differently growing regions, and

could have interesting implications in biological systems that undergo apical growth [103, 104].

In Fig. 6.9 (a) and (b) density profiles for exponential enteric growth are shown. When GM1 is implemented species *A* dominates across the domain, although the density of species *A* is reduced in the region of high growth (see Fig. 6.9 (a)). However, in Fig. 6.9 (b) GM2 is implemented and causes species *B* to have a higher density in the middle region of the domain. In Fig. 6.9 (c) and (d) the density

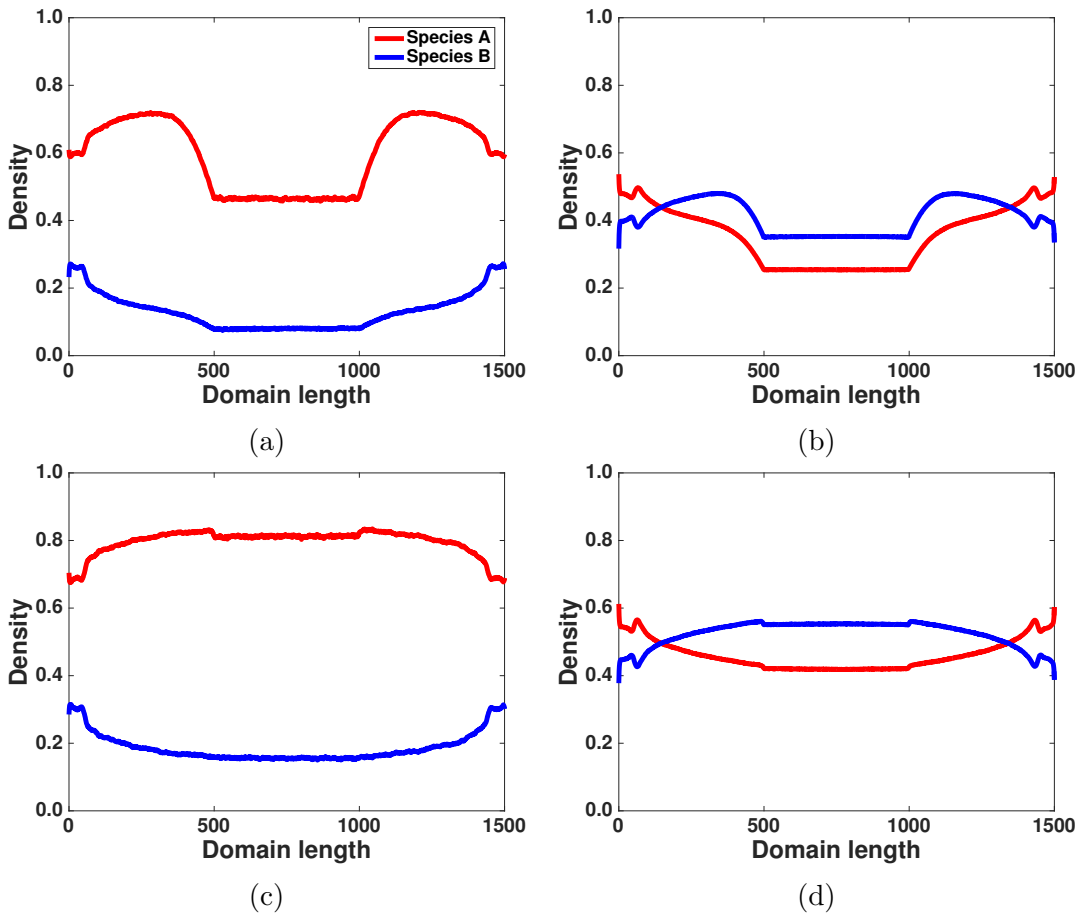


Figure 6.9: Enteric domain growth. (a) GM1 exponential domain growth, (b) GM2 exponential domain growth, (c) GM1 linear domain growth, (d) GM2 linear domain growth. The parameters for all panels are $P_m^A = 20$, $P_p^A = 0.9$, $P_m^B = 1$, $P_p^B = 1$, $P_{gx} = 0.1$ and $P_{gy} = 0.1$.

profiles for linear enteric growth are shown. As before, when GM1 is implemented

this enables species A to dominate across the domain (see Fig. 6.9 (c)). However, in Fig. 6.9 (d) GM2 is implemented and this causes species B to have a slightly higher density in the middle region of the domain. Finally, in both Fig. 6.8 (b) and (d) and Fig. 6.9 (b) and (d) small ‘oscillations’ in the agent density can be observed near the boundary. These oscillations only occur in the presence of GM2, but the reason for their appearance is not clear.

In conclusion, with initial conditions that are, on average, spatially uniform for two species non-uniform growth can establish patterns in species density across the domain given certain parameter values. The development of these patterns is directly attributable to the effects of GM1 and GM2 on spatial correlations between agents in the IBM.

6.4 Summary

In this chapter we have studied the effect of two growth mechanisms on spatial correlations in agent populations containing multiple species. We chose two different, yet potentially biologically relevant growth mechanisms [92–94], to highlight how understanding the form of the domain growth in biological systems is important. It is important to acknowledge that in reality it is unlikely that domain growth in biological systems is captured by algorithms as simple as GM1 and GM2. However, it stands to reason that more realistic growth mechanisms may exist that exhibit similar effects on spatial correlations as GM1 and GM2.

Our key finding is that the specific type of growth mechanism can influence the dominant species, as shown in Figs. 6.2, 6.5 and 6.6. Under certain parameter regimes a more motile, slower proliferating species will dominate under growth mechanism GM1, whereas a less motile, faster proliferating species will dominate under growth

mechanism GM2. This is because GM2 breaks down colinear correlations more effectively than GM1, and so benefits the faster proliferating species. Interestingly, this result suggests that the way in which a domain grows could play a role in determining cell population fates in biological systems.

To conclude our results section we studied some biologically motivated examples of non-uniform domain growth. We found that we were able to establish simple patterns (Fig. 6.9), and that these patterns changed depending on the way domain growth was implemented. This shows that non-uniform growth can establish agent patterning on a domain, which is an intriguing result.

In this chapter all models studied are two-dimensional. The correlation ODE model has been derived for three dimensions on a non-growing domain [22], and so this is an obvious extension to the work presented here. In addition, we have also only considered the case when $P_{g_x} = P_{g_y}$. We did this to reduce the complexity of the equations, however, the results presented here could be extended to cases where $P_{g_x} \neq P_{g_y}$.

A final consideration is whether the work presented here could be extended to other types of model, such as an off-lattice IBM whereby agents can occupy any position in space (while taking volume exclusion into account, if necessary). Research has been directed towards including spatial structure in continuum approximations of off-lattice IBMs [50, 51, 105]. The effect of the domain growth mechanisms on the evolution of the agent density in an off-lattice IBM, as with the work presented here, will depend on how effectively the growth mechanisms break up spatial correlations established by agent proliferation.

.....

In the next chapter we analyse an experiment commonly used to study cell migration,

and examine how some of the techniques presented in this thesis (especially those presented in Chapter 3) can be used to improve the design of this experiment.

Chapter 7

Model parameterisation for a simple wound healing experiment

Overview

In this chapter we implement approximate Bayesian computational (ABC) methods to improve the design of a wound-healing assay¹ used to quantify cell-cell interactions. It is important to be able to quantify cell-cell interactions, such as adhesion and repulsion, as they have been shown to play an important role in cell migration. Initially, we demonstrate with an IBM of an *ideal* experiment that we are able to identify model parameters for agent motility and adhesion, given we choose appropriate summary statistics. Following this, we replace our IBM of an ideal experiment with an IBM representative of a practically realisable experiment. We demonstrate that, given the current (and commonly used) experimental set-up, model parameters cannot be accurately identified using ABC methods. We compare new experimental designs through simulation, and show more accurate identification of model parameters is possible by expanding the size of the domain upon which the experiment is performed, as opposed to increasing the number of experimental repeats. The results presented in this chapter therefore describe time *and* cost-saving alterations for a commonly performed experiment for identifying cell motility parameters. More-

¹Wound-healing assays are also often referred to as scratch assays.

over, the results presented in this chapter will be of interest to those concerned with performing experiments that allow for the accurate identification of parameters governing cell migratory processes, especially cell migratory processes in which cell-cell adhesion or repulsion are known to play a significant role.

The majority of the work in this chapter appears in the paper: R. J. H. Ross, R. E. Baker, Andrew Parker, M. Ford, R. L. Mort, and C. A. Yates. “Using approximate Bayesian computation to quantify cell-cell adhesion parameters in a cell migratory process”. *npj Systems Biology and Applications* (under review), 2016. R. J. H. Ross performed all the computational and mathematical modelling presented in this chapter. R. L. Mort and M. Ford performed the wet lab work and provided the experimental data.

7.1 Introduction

As described in Chapter 3, cell-cell interactions are known to play an important role in several cell migration processes. For example, multiple different cell-cell interactions, such as cell-cell signalling and cell-cell adhesion [65], have been identified as promoting metastasis in breast cancer. Repulsive interactions mediated via ephrins on the surface of neural crest stem cells are known to coordinate the early stages of melanoblast migration away from the neural tube [52]. More fundamentally, it is hypothesised that the emergence of cell-cell interactions over one billion years ago helped establish the necessary conditions for multicellular organisms [69].

If an IBM is an *effective*² representation of a cell migration process it can be used for a number of purposes. One such purpose for an IBM is to perform *in silico* ex-

²By an effective representation we mean the IBM captures the salient features of the process of interest, and is therefore a viable research tool with which to study the process of interest.

periments to test scientific hypotheses. For instance, in Chapter 2 we used an IBM to demonstrate that a simple mechanism of undirected cell movement and proliferation could account for neural crest stem cell colonisation of the developing epidermis in the embryonic mouse [10]. Other studies involving IBMs have tested hypotheses concerning the influence of matrix stiffness and matrix architecture on cell migration [106], and the mechanism by which cranial NCSCs become ‘leaders’ or ‘followers’ in the embryonic chick to allow their collective migration [1–3].

IBMs can also be used to *identify* parameters in experimental data (with the caveat that the parameters are model-dependent). The reasoning behind using an IBM to identify parameters in experimental data is as follows: if an IBM is an effective representation of an experiment, then the parameter values the IBM requires to reproduce the experimental data may be representative of the parameter values in the biological process that is the focus of the experiment³. For instance, the value of a parameter that describes cell proliferation rate. Even if the parameter values in the parameterised IBM are not representative of the parameter values in the biological process, the parameterised IBM may still be used to make predictions about the process of interest by performing *in silico* experiments, as described above. These predictions can then be tested experimentally.

Alternatively, if the IBM is an effective representation of an experiment (i.e. the experimental data can be reproduced), but the parameters of the IBM are not identifiable, this may suggest the experiment is not well-designed (that is, if the experiment has been designed to estimate parameters). By parameters not being identifiable it is meant very different parameter values in the IBM can reproduce the same experimental data. If this is the case, the IBM can then be used to suggest improvements

³Throughout this chapter we assume that cellular processes such as migration have constant parameter values associated with them.

to the experiment's design, namely by altering the IBM design such that the IBM parameters become identifiable. These alterations can then be applied to the experiment to improve parameter identifiability. For example, a recent study using an IBM has examined the time-points at which data should be collected from an experiment to maximise the identifiability of IBM parameters [9]. Other theoretical work has shown how to maximise the information content of an experiment by choosing an appropriate experimental design [8].

The focus of this chapter is to determine the experimental conditions, and experimental data, required for the accurate identification of cell motility and adhesion parameters in a wound-healing assay. To do so we employ ABC, a probabilistic approach whereby a probability distribution for the parameter(s) of interest is generated, as opposed to a point estimate [107–109]. Although ABC is well-established in some fields, for instance in population genetics [110], its applicability for IBMs representing cell migration is still an area of active research [9, 108]. Recent studies combining ABC and IBMs have been able to identify motility and proliferation rates in cell migratory processes [108], and improve the experimental design of scratch assays [9]. However, as far as we are aware nobody has used ABC methods to examine the experimental conditions, and experimental data, required for the accurate identification of cell motility and adhesion parameters in a wound-healing assay. It is important to acknowledge that in migrating cell populations there are many more factors at play than simply cell motility and adhesion. For instance, the cell cycle and a cell's response to environmental cues may be important factors in a cell's behaviour. However, in agreement with other studies we have purposely simplified our study before attempting to introduce additional complexity to our IBM. This 'bottom-up' approach is common in the literature [9, 108, 111–114], and we further address this point in this chapter's summary.

Other methods to identify parameters in experimental data using IBMs also exist. For instance, a standard approach is to generate point estimates of model parameters that best reproduce statistics of the experimental data in the IBM. For example, the generation of motility and proliferation rates for agents in an IBM representing a biological process. This method of parameter identification was demonstrated in Chapter 2 [10]. This approach, while applicable in some circumstances, often gives no insight into how much uncertainty exists in the parameters chosen, a factor that can be of importance when analysing biological systems. For example, relationships between parameter uncertainty and system robustness are thought to be connected in biological function at a systems level [115].

The outline of this chapter is as follows: in Section 7.2 we introduce the IBM and define the cell-cell interactions we consider. We also outline the method of ABC, and the summary statistics we use to analyse the IBM output. In Section 7.3 we present results and demonstrate that, given an IBM representing an ideal experiment, we are able to identify IBM parameters for agent motility and adhesion. Following this, we replace our IBM representing an ideal experiment with an IBM that simulates a practically realisable experiment. In doing so we show that parameters cannot be successfully identified using ABC given the current experimental design. To improve parameter identifiability we compare different experimental designs, and show that identification of IBM parameters is made more accurate if the size of the domain upon which the experiment is performed is expanded, as opposed to increasing the number of experimental repeats. Experimentally, expanding the size of the domain is equivalent to increasing the field of view of the microscope used to collect the experimental data. For instance, five simulation repeats on a larger domain provides more accurate identification of IBM parameters than 500 simulation repeats on a smaller

domain. In Section 7.4 we discuss the results presented in this chapter.

7.2 Methods

In this section we first introduce the IBM. We then define what we mean by summary statistics and explain ABC and its implementation.

7.2.1 Individual-based model

The IBM is simulated on a two-dimensional square lattice with lattice spacing Δ [46] and size L_x by L_y , where L_x is the number of lattice sites in a row and L_y is the number of sites in a column. Each agent is initially assigned to a lattice site, from which it can move into adjacent sites. If an agent attempts to move into a site that is already occupied by another agent, the movement event is aborted. In the IBM time evolves continuously, in accordance with the Gillespie algorithm [90], such that agent movement events are modelled as exponentially distributed reaction events in a Markov chain. Attempted agent movement events occur with rate P_m per unit time. $P_m \delta t$, therefore, is the probability of an agent attempting to move in the next infinitesimally small time interval δt . A lattice site is denoted by $v = (i, j)$, where i indicates the column number and j the row number. Each lattice site has four adjacent lattice sites (except for those sites situated on nonperiodic boundaries), and so the number of nearest neighbour lattice sites that are occupied by an agent, denoted by n , is $0 \leq n \leq 4$. We denote the set of unoccupied nearest neighbour lattice sites by $\mathcal{U}$.

The IBM domain size for simulations representing ideal experiments is $L_x = 100$ by $L_y = 100$, and the lattice sites indexed by $1 \leq j \leq L_y$ and $1 \leq i \leq 10$, and $1 \leq j \leq L_y$ and $91 \leq i \leq L_x$ are initially occupied by agents. In Fig. 7.1 the initial conditions in the IBM for the ideal experiment can be seen. The initial condition in Fig. 7.1 represents a ‘wound’, in that agents are positioned either side of

an unoccupied region of the lattice, the ‘wound’, that they can migrate into. The agent migration into this space simulates one aspect of the wound-healing process. We refer to this simulation as ideal because the symmetry of the initial conditions may not be possible in a realistic experimental setting. The initial condition is also ideal as it is ‘double-sided’, as opposed to the ‘single-sided’ experiment data that we will later analyse. It has been shown that double-sided initial conditions can provide more information than single-sided initial conditions for some model parameters [9]. For instance, double-sided initial conditions can provide more information when increasing the number of agents in a simulation improves parameter identifiability.

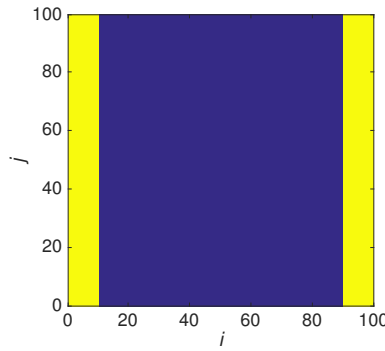


Figure 7.1: The initial condition in the IBM for the ideal experiment. Yellow indicates a site occupied by an agent and blue indicates an empty lattice site.

For the IBM of an ideal experiment all simulations have periodic boundary conditions at the top and bottom of the domain (i.e. for lattice sites indexed by $j = 1$ or $j = L_y$), and no-flux boundary conditions at the left-hand and right-hand boundaries of the domain (i.e. for lattice sites indexed by $i = 1$ or $i = L_x$).

7.2.2 Cell-cell adhesion models

In the IBM cell-cell interactions are simulated by altering the probability of an agent attempting to move, depending on the number of nearest occupied neighbours, n , an agent has. We employ two models to simulate cell-cell interactions in the IBM, one of which has been published before [12, 33]. We define $T(v'|v)$ as the transition

probability of an agent situated at site v , having been selected to move, attempting to move to site v' , where v' indicates one of the nearest neighbour sites of v . Therefore, $T(v'|v)$ is only non-zero if v and v' are nearest neighbours. The transition probability in the first model, which we refer to as model A, is defined as

$$T_A(v'|v) = \frac{1 - n\alpha}{4}, \quad (7.1)$$

where α is the adhesion parameter. The subscript A on the transition probability in Eq. (7.1) indicates that this is the transition probability for model A. If $\alpha > 0$ Eq. (7.1) represents cell-cell adhesion, and if $\alpha < 0$ Eq. (7.1) represents cell-cell repulsion. The transition probabilities stated in Eq. (7.1) must satisfy

$$0 \leq \sum_{v' \in \mathcal{U}} T_A(v'|v) \leq 1. \quad (7.2)$$

Equation (7.2) ensures the probability of an agent, if selected to move, attempting to move to any of its unoccupied nearest neighbour sites never exceeds unity, and so constrains the value α can take. The transition probability in the second model⁴, which we refer to as model B [12, 33], is defined as

$$T_B(v'|v) = \frac{(1 - \alpha)^n}{4}, \quad (7.3)$$

and must satisfy

$$0 \leq \sum_{v' \in \mathcal{U}} T_B(v'|v) \leq 1. \quad (7.4)$$

As in model A if $\alpha > 0$ Eq. (7.3) models cell-cell adhesion, and if $\alpha < 0$ Eq. (7.3) models cell-cell repulsion.

⁴This is the same model for cell-cell adhesion that we implemented in Chapter 3.

Models A and B simulate different forms of cell-cell interaction. In model A the transition probability is a linear function of n . Meanwhile, in model B the transition probability is a nonlinear function of n . Not only may these different forms of cell-cell interaction be relevant for different cell types, but implementing two models of cell-cell interaction allows us to test the robustness of the methods we present in this chapter.

7.2.3 Summary statistics

Summary statistics are lower-dimensional summaries of data that provide a tractable means to compare different sets of data. Summary statistics are important because experimental data is often of high dimensionality, and if we want to use experimental data to efficiently guide computational algorithms we require ways to accurately summarise it. We now define the summary statistics we apply to the IBM output and experimental data. Following this we describe how we utilise these summary statistics to implement ABC.

We initially use three summary statistics to evaluate the IBM output, all of which have been considered previously [111, 116]. The reason we study three summary statistics is to ascertain which summary statistic is most effective for the identification of agent motility and adhesion parameters in the IBM. These summary statistics are now described.

Average horizontal displacement of agents

The average horizontal displacement of all agents, $\bar{i}$, in a given time interval, $[t_i, t_f]$, in the IBM is calculated as

$$\bar{i} = \frac{1}{N} \sum_{k=1}^N |i_{t_i}^k - i_{t_f}^k|, \quad (7.5)$$

where N is the total number of agents in the simulation, $i_{t_i}^k$ is the column position of agent k at time t_i , and $i_{t_f}^k$ is the column position of agent k at time t_f . We only look at the horizontal displacement of agents as this is the direction in which the majority of agent displacement will occur, due to the initial conditions of the IBM (Fig. 7.1). We have shown in Chapter 3 that different cell-cell interactions have different effects on the average displacement of agents in an IBM [12]. As demonstrated in Chapter 3, repulsive (adhesive) interactions between agents tend to increase (decrease) the average displacement of agents, and so the average displacement of agents may be a useful summary statistic for distinguishing between repulsive and adhesive cell-cell interactions in the IBM.

Agent density profile

The agent density profile, $C_t(i)$, at time t in the IBM is calculated as

$$C_t(i) = \frac{1}{L_y} \sum_{j=1}^{L_y} \mathbb{1}\{v\}. \quad (7.6)$$

Here $\mathbb{1}$ is the indicator function for the occupancy of a lattice site v (i.e. 1 if an agent occupies lattice site v , and 0 if it is not occupied by an agent). We have shown in Chapter 3 that different cell-cell interactions have different effects on the agent density profile [12]. For instance, repulsive interactions between agents can create a concave agent density profile, whereas adhesive interactions between agents can

create a convex agent density profile. Therefore, the agent density profile may be an effective summary statistic for distinguishing between repulsive and adhesive cell-cell interactions in the IBM.

Pairwise-correlation function

The final summary statistic we consider is the pairwise-correlation function (PCF). The PCF provides a measure of the spatial clustering between agents in an IBM, and has been used frequently in the analysis of cell migratory processes [10, 44, 45, 111]. The PCF has also been successfully used as a summary statistic for the parameterisation of IBMs of cell migration [108]. We use i_t^k to denote the column position of agent k at time t , i_t^l to denote the column position of agent l at time t , and define $c_t(m)$ to be the number of occupied pairs of lattice sites for each *nonperiodic*⁵ horizontal pair distance $m = 1, \dots, L_x - 1$ at time t . This means $c_t(m)$ is given by

$$c_t(m) = \sum_{k=1}^N \sum_{l=k+1}^N \mathbb{1}\{|i_t^k - i_t^l| = m\}, \quad \forall m = 1, \dots, L_x - 1, \quad (7.7)$$

where $\mathbb{1}$ is the indicator function such that it is equal to 1 if $|i_t^k - i_t^l| = m$, and 0 otherwise. In Eq. (7.7) only the pair distances in the horizontal direction are counted. This is because, given the translational invariance of the initial conditions in the vertical direction of the IBM, the majority of important spatial information will be in the horizontal direction⁶. Binder and Simpson [45] demonstrated that it is necessary to normalise Eq. (7.7) to account for volume exclusion. The normalisation

⁵By nonperiodic it is meant the distance measured between two agents cannot cross the IBM boundary.

⁶This approach is in agreement with previous studies [45], which showed the most relevant information from the PCF summary statistic is perpendicular to the wound axis in a wound-healing assay.

term is

$$\hat{c}_t(m) = L_y^2(L_x - m)\rho\hat{\rho}, \quad \forall m = 1, \dots, L_x - 1, \quad (7.8)$$

where $\rho = N/(L_x L_y)$, and $\hat{\rho} = (N - 1)/(L_x L_y - 1)$. Equation (7.8) describes the expected number of pairs of occupied lattice sites, for each nonperiodic horizontal pair distance, m , in an agent population distributed uniformly at random on the IBM domain. Combining Eqs. (7.7) and (7.8), the PCF is

$$q_t(m) = \frac{c_t(m)}{\hat{c}_t(m)}, \quad (7.9)$$

where $q_t(m)$, the PCF, is a measure of how far $c_t(m)$ departs from describing the expected number of occupied lattice pairs for each horizontal distance of an agent population spatially distributed uniformly at random on the IBM domain.

It is important to briefly explain why we chose these summary statistics and not others, as other summary statistics different to those described here have been used to analyse cell migration [108, 112, 114]. Other summary statistics were initially implemented in this study. However, they were found to add little in terms of novelty and so were excluded from this chapter. Therefore, in this chapter we are not presenting an exhaustive list of summary statistics to analyse cell migration. Instead, we are studying three previously published ones and analysing their effectiveness in parameterising cell adhesion and motility in an IBM of a wound-healing assay.

7.2.4 Approximate Bayesian computation

Here we introduce the ABC algorithm [107]. We define M as a stochastic model that takes parameters Θ and produces data D . This relationship can be written as $D \sim M(\Theta)$. For the IBM presented in this chapter $\Theta = (P_m, \alpha)$, where Θ is sampled

from a prior distribution, π , and so this relationship can be written as $\Theta \sim \pi$. The relationship between π and Θ is often written as $\Theta \sim \pi(\Theta)$, which indicates that a new Θ sampled from the prior distribution may depend on the previous Θ . This relationship will be relevant later on in this chapter, however, initially each Θ sampled from π is independent of the previous Θ .

The identification of IBM parameters in this chapter centres around the following problem: given a stochastic model, M , and data, D , what is the probability density function that describes Θ being the model parameters that produced data D ? More formally, we seek to obtain a posterior distribution, $p(\Theta|D)$, which is the conditional probability of Θ given D (and the model, M).

Typically, to compute the posterior distribution a likelihood function, $L(D|\Theta)$, is required. This is because the likelihood function and posterior distribution are related in the following manner by Bayes' theorem:

$$p(\Theta|D) \propto L(D|\Theta)\pi(\Theta). \quad (7.10)$$

That is, the posterior distribution is proportional to the product of the likelihood function and the prior distribution. Approximate Bayesian computation is a well-known method for estimating posterior distributions of model parameters in scenarios where the likelihood function is *intractable* [107]. By an intractable likelihood function it is meant that the likelihood function is impossible or computationally prohibitive to obtain.

In many cases for ABC, due to the high dimensionality of the data, D , it is necessary to utilise a summary statistic, $S = S(D)$. The summary statistics we employ

in this chapter are of varying dimension. For instance, the agent density profile at time t has L_x data points, whereas the average agent displacement at time t has one data point. Therefore we write $S(D)$ as $S(D)_{r,t}$, where $S(D)_{r,t}$ is the r^{th} data point in the summary statistic at the t^{th} sampling time.

The ABC method proceeds in the following manner: we wish to estimate the posterior distribution of Θ given D . We now simulate the process that created D using model M with parameters Θ , sampled from π , and produce data $\tilde{D}$. We calculate the difference, d , between a summary statistic applied to D and $\tilde{D}$ with

$$d = \sum_{t=1}^T \sum_{r=1}^R |S(D)_{r,t} - S(\tilde{D})_{r,t}|, \quad (7.11)$$

where R is the number of data points in $S(D)$ and T is the number of sampling times. We repeat the above process many times, that is, sample Θ from π , produce $\tilde{D}$, calculate d with Eq. (7.11), and only accept Θ when d is below a user defined certain threshold (alternatively, a predefined number of Θ that minimise d can be accepted). This enables us to generate a distribution for Θ that is an approximation of the posterior distribution, $p(\Theta|D)$, given M . More specific details of the ABC algorithms we implement are introduced when necessary in the text.

7.3 Results

We begin by demonstrating that for an IBM representing an ideal experiment we are able to identify model parameters, given we use the appropriate summary statistics.

7.3.1 Ideal experiment

To ascertain the effectiveness of the chosen summary statistics to identify model parameters, we first attempt to identify Θ from data generated *synthetically*. Synthetic

data is IBM data generated with fixed parameter values, and so can be thought of as a simulation equivalent of experimental data. To generate the synthetic data using the IBM we proceed as follows:

1. We choose parameters Θ to identify. To help clarify this explanation let us make these parameters $\Theta = (P_m, \alpha) = (0.5, 0.1)$ in model A⁷.
2. For model A we perform a simulation of the IBM with $\Theta = (0.5, 0.1)$, generate data, D , and calculate summary statistics, $S(D)$, from the simulation at our time-points of interest. These times are $t = [240, 480, 720]$. We choose these times as they are the times (in minutes) we will later analyse for the simulations of the practically realisable experiment, and correspond to 4 hours, 8 hours and 12 hours into an experiment.
3. We repeat step 2 ten times and calculate the ensemble average for each summary statistic for each individual time-point.

This procedure generates synthetic data for which we will now attempt to identify the parameters. In this chapter we present estimates for $P_m = 0.5$ and $\alpha = 0.1$ for model A, and $P_m = 0.5$ and $\alpha = 0.25$, and $P_m = 0.5$ and $\alpha = -0.1$ for model B. We examined identifying further combinations of values of P_m and α from synthetic data, and what we present here is a representative sample of the combinations we tested.

Throughout this chapter we sample P_m and α for our model from uniform priors. In the case of model A, $P_m \in [0, 1]$ and $\alpha \in [-0.2, 0.25]$, and for model B, $P_m \in [0, 1]$ and $\alpha \in [-0.2, 1.0]$. We stipulate these lower and upper bounds for α for both models A and B to make sure inequalities (7.2) and (7.4) are satisfied.

⁷A value of $P_m = 0.5$, given that the simulation time will later be defined to be in minutes, and the length of a lattice site represents cell length (typically between 10 μm -100 μm), means that the motility of the agents is biologically realistic. The parameter α is dimensionless. The experimental realism of these parameters will be expanded on when we address the simulation of a practically realisable experiment.

We begin by implementing an ABC rejection algorithm since we expect to identify model parameters quickly as we are simulating an ideal experiment. The rejection ABC algorithm proceeds as follows:

1. run 10^4 IBM simulations, in each case using Θ sampled uniformly at random from the prior distributions;
2. compute the distance d as defined in Eq. (7.11), where $t = [240, 480, 720]$;
3. accept the 100 parameter values, Θ , that minimise d .

In Fig. 7.2 the posteriors generated using each of the three summary statistics applied to data from simulations of an ideal experiment are displayed. The most effective summary statistic for identifying the synthetic data parameters is the PCF. The effectiveness of the PCF for parameter identification is evident in the location of the posterior distribution density relative to the red dot (the red dot represents the synthetic data parameter values), and the narrow spread of the posterior distribution density as indicated by the scale bar in Fig. 7.2 (c), (f) and (i). The agent density profile summary statistic performs less well than the PCF for parameter identification, especially for model A (Fig. 7.2 (b)). In the case of the average agent displacement many combinations of P_m and α lead to the same average agent displacement, which results in an extended region of probable parameter values. To some extent this is to be expected, as increasing either P_m or α will have opposing effects on the average agent displacement. This means that using agent displacement as a summary statistic results in parameter identifiability issues in this example.

To quantify the difference between the performance of the different summary statistics we use the Kullback-Leibler divergence, which is a measure of the information gained in moving from the prior distribution to the posterior distribution [117]. The

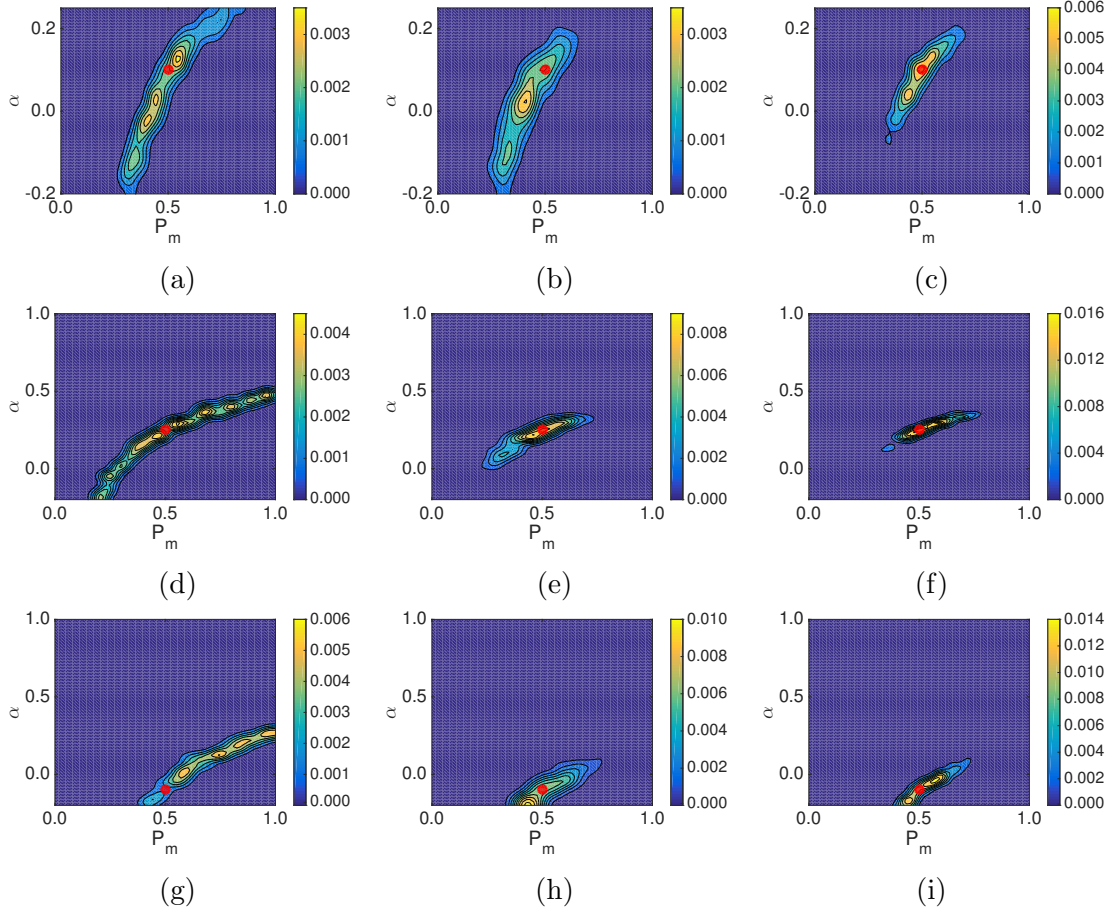


Figure 7.2: (a)-(c) Posterior distributions for model A for an ideal experiment with different summary statistics: (a) average displacement of agents in the horizontal direction; (b) agent density profile; (c) PCF. In all cases the red dot indicates the value of the parameters used to generate the synthetic data, $P_m = 0.5$, $\alpha = 0.1$. As indicated by the colour bar the yellow regions indicate areas of high relative density of the posterior distribution, while the blue regions indicate areas of low relative density of the posterior distribution. (d)-(f) Model B, $P_m = 0.5$, $\alpha = 0.25$: (d) average displacement of agents in the horizontal direction; (e) agent density profile; (f) PCF. (g)-(i) Model B, $P_m = 0.5$, $\alpha = -0.1$: (g) average displacement of agents in the horizontal direction; (h) agent density profile; (i) PCF.

Kullback-Leibler divergence for a discrete probability distribution is defined as follows:

$$D_{KL}(p|\pi) = \sum_l p(\Theta_l|D) \log \left(\frac{p(\Theta_l|D)}{\pi(\Theta_l)} \right), \quad (7.12)$$

where the index l accounts for all possible discretised parameter pairs (i.e. all com-

binations of P_m and α). A larger $D_{KL}(p|\pi)$ value suggests that more information is obtained (the entropy of the distribution is reduced) when moving from the prior distribution to the posterior distribution⁸. We discretise our posterior distribution onto a lattice with 2^6 equally spaced values of P_m and 2^6 equally spaced values of α .

Computing $D_{KL}(p|\pi)$ for all nine plots in Fig. 7.2 gives: (a) 1.77; (b) 1.70; (c) 2.32; and (d) 2.15; (e) 2.57; (f) 3.35; and (g) 2.45; (h) 2.72; (i) 3.27. In tandem with the proximity of the peak of the posterior distribution densities to the red dots in Fig. 7.2 (c), (f) and (i), compared to Fig. 7.2 (a)-(b), (d)-(e) and (g)-(h), this suggests that the PCF summary statistic is more effective for parameter identification than the average agent displacement and agent density profile summary statistics.

7.3.2 Practically realisable experiment

In the previous section we demonstrated that for ideal experimental conditions the PCF summary statistic is best able to identify synthetic data parameters (for an IBM of an ideal experiment), and so now we will only use the PCF summary statistic for parameter identification. Previous work has combined summary statistics to improve parameter identification [108]. However, in this case it makes a negligible improvement to the posterior (results not shown)⁹.

We now replace our IBM that represents an ideal experiment with an IBM that represents an actual experiment, and examine whether synthetic data parameters can

⁸However, this does not necessarily mean the posterior distribution is a more accurate representation of the parameter distribution. It is important to stress that the KL divergence should not be seen as ubiquitously applicable to inference problems of this nature. For instance, the use of the KL divergence as presented here should be used with caution in scenarios whereby ABC is being implemented with an informative prior. In scenarios such as these, other methods to measure the improvement of an inference procedure have been reported [118].

⁹That there is little improvement in parameter identification from combining summary statistics is to be expected. Combining summary statistics is most effective when the posterior distributions are ‘orthogonal’, which is not the case for the posterior distributions created by the summary statistics presented here.

be identified in the IBM. We provide brief details of the experiment here, however, a more detailed description can be found in Appendix D. In Fig. 7.3 a typical initial frame of the experimental data can be seen.

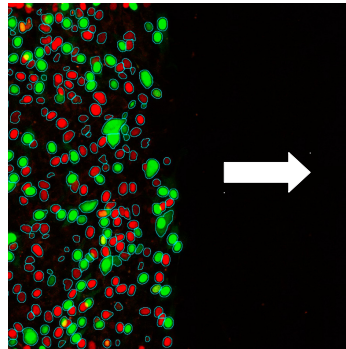


Figure 7.3: Typical initial frame of the experimental data. The cells are positioned such that they will migrate primarily horizontally into the space without cells, this space represents a wound (the direction of migration is indicated by the white arrow). The red and green cells are the same cells, with red and green indicating which phases of the cell cycle cells are in. In this chapter we do not take the cell cycle or cell proliferation into account (figure courtesy of Richard Mort).

In total we have data from five repeats of the experiment. Each data set contains cell track data for every cell for 64 hours imaged at 20 minute intervals. Therefore, we have the information required to apply our summary statistics to the experimental data. More specifically, we have the position of all cells at each time interval so that the PCF may be computed.

One key difference between the ideal and practically realisable experiments is the size of the domain and, because of this, the number of agents in a simulation. As we have data from five experiments we now generate our synthetic data from five repeats of the IBM, using the same procedure as described in Section 3.1.

The experimental images were captured by a microscope with a field of view of $597.24 \mu\text{m}$ by $597.24 \mu\text{m}$. The cell size in the experimental images is consistent with each cell occupying a $26 \mu\text{m}$ by $26 \mu\text{m}$ square lattice site. Given the size of the microscope

Column	Initial density
1 st	0.8261
2 nd	0.7826
3 rd	0.8261
4 th	0.8261
5 th	0.8261
6 th	0.7391
7 th	0.6957
8 th	0.6087
9 th	0.5217
10 th	0.2609
11 th	0.2174
12 th	0.0870
13 th – 23 rd	0

Table 7.1: Average initial condition for the IBM.

field of view this means the IBM domain size is $L_x = 23$ by $L_y = 23$. To generate the initial conditions in the IBM we proceed as follows: to map the position of cells in the experimental images where cell position is a continuous variable, (x, y) , to a discrete lattice site, (i, j) , we use the following formulae

$$i = \left\lceil \frac{x}{\Delta} \right\rceil, \quad j = \left\lceil \frac{y}{\Delta} \right\rceil, \quad (7.13)$$

where $\lceil \cdot \rceil$ denotes the ceiling function and Δ is the lattice site length. No two cells were mapped to the same lattice site¹⁰.

The application of Eq. (7.13) to the initial frames of the five experiments allowed the average initial condition for the IBM to be calculated. These initial conditions are expressed in terms of the average initial density of each column. These average initial column densities can be seen in Table 7.1.

¹⁰If two cells did map to the same lattice site, one of these cells would be placed in the nearest unoccupied lattice site to the original lattice site. If there was more than one nearest unoccupied lattice site, one of these sites would be chosen uniformly at random for the cell to be mapped to.

To generate the initial conditions at the start of each IBM realisation each site in a column receives an agent uniformly at random at a probability equal to the average initial column density of the column the site is in. Therefore, the initial condition in the IBM is generated such that an ensemble average of the initial conditions of many realisations would equal the averaged initial conditions from the experiment.

We also alter the IBM to have flux (nonperiodic) boundary conditions at the left-hand and right-hand boundaries of the domain (i.e. for lattice sites with $j = 1$ or $j = N_y$). The left-most column is kept at or above a constant density throughout the simulation time course. This means following the start of the simulation the density of the left-most column is checked after each agent movement event out of the first column in the IBM. If the left-most column's density is below 0.6, agents are added uniformly at random to empty sites in the left-most column until the density of the first column is greater than 0.6. This mechanism (and density) ensures that the agent density profile in the IBM matches the experimental density profile throughout the simulation. The top and bottom boundaries of the IBM domain remain periodic as cells were seen to move in and out of the microscope field at these boundaries in the experimental images, at an approximately equal rate.

To reduce the computational time of the ABC algorithm we are using we now employ the Metropolis-Hastings algorithm. Details of the implementation of the algorithm are as follows: we define a transition kernel w that proposes Θ' values as a bivariate uniform distribution. The transition kernel ensures $P_m \in [0, 1]$ and $\alpha \in [-0.2, 0.25]$ for the model A, and $P_m \in [0, 1]$ and $\alpha \in [-0.2, 1.0]$ for the model B. The parameter d^* is a constant selected so that approximately one percent of the proposed parameter sets are accepted, the value of which is obtained through trial and error.

To implement Markov Chain Monte ABC (Metropolis-Hastings algorithm) we proceed as follows [107]:

1. if at Θ step to Θ' according to a transition kernel $w(\Theta \rightarrow \Theta')$;
2. simulate $\tilde{D}$ from the model using Θ' and calculate the summary statistic $S(\tilde{D})$ at each sampling point. That is, for each individual $t = [240, 480, 720]$ calculate d :

$$d = \sum_{r=1}^R |S(D)_{r,t} - S(\tilde{D})_{r,t}|, \quad (7.14)$$

If $d > d^*$ (at any t) reject Θ' and return to step 1;

3. calculate

$$h = \min \left(1, \frac{\pi(\Theta')w(\Theta' \rightarrow \Theta)}{\pi(\Theta)w(\Theta \rightarrow \Theta')} \right);$$
4. accept Θ' with probability h ;
5. return to step 1 until 10^6 steps have been attempted.

Initially, we sample Θ randomly from the prior distribution until a parameter set has been accepted (step 4). We do not implement rejection ABC as we expect parameter identification to be less efficient with a more realistic model, and so we implement a correlated Monte Carlo method. Given our model assumptions our implementation of the Metropolis-Hastings algorithm reduces to a Markov chain Monte Carlo method with a correlated outcome [107], of which we attempt 10^6 realisations.

We collect simulation data at three time-points so that the computational time is of practical length (our longest ABC implementations took approximately 192 hours). A value of $P_m = 0.5$, given that the simulation time is in minutes, and the length of

a lattice site is 26 μm , means that the motility of the agents is biologically realistic. To be precise, the agents here are approximately five times faster than cell motility rates previously published [10, 111]¹¹. However, the cells considered in [10, 111] are not thought to exhibit cell-cell adhesion, and so a higher motility rate is sensible as agent movement in our IBM is being reduced in the case of cell-cell adhesion.

In Fig. 7.4 it can be seen that the synthetic data parameters cannot be accurately identified using our ABC method, with the PCF summary statistic, given the current IBM design. This is evident in the location of the red dots relative to the posterior distributions, and the wide spread of the posterior distributions (as indicated by the scale bar in Fig. 7.4). A possible reason why the synthetic data parameters cannot be identified is that the synthetic data does not accurately represent the parameter values used to generate it, making parameter identification infeasible. To examine this possibility we calculated the variance in the PCF synthetic data¹². In Fig. 7.5 (a)-(c) the blue line indicates the variance in the PCF synthetic data for the current simulation design generated from five repeats of the IBM on a domain of size $L_x = 23$ by $L_y = 23$.

If the variance in the summary statistics of the synthetic data precludes accurate identification of model parameters using ABC, a sensible strategy may be to examine methods to reduce the variance in the summary statistics of the synthetic data. Reducing the variance of the summary statistics may mean the synthetic data is a more accurate reflection of the parameter values used to generate it. This may also explain why parameter identification for the ideal experiment was successful, as the variance in the summary statistics of the synthetic data was much smaller than for the practically realisable experiment (data not shown).

¹¹Using the relationship that the diffusion coefficient is equal to $P_m\Delta^2$.

¹²The variance was calculated using MATLAB's in-built `var` function.

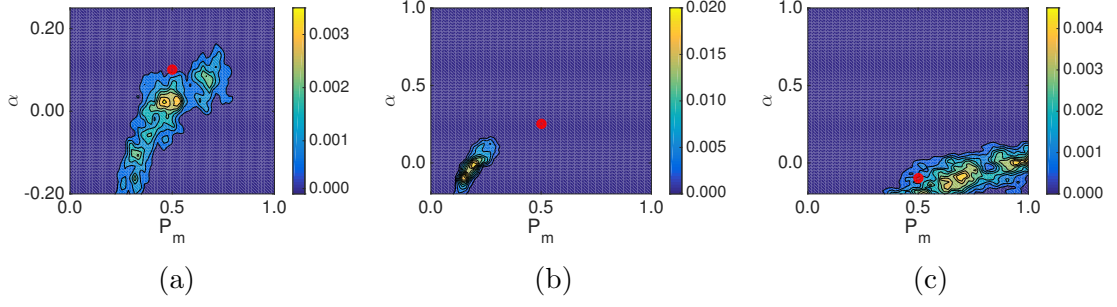


Figure 7.4: Posterior distributions for simulations of the experiment described in Section 2.5 using the PCF as a summary statistic for an IBM of size $L_x = 23$ and $L_y = 23$. The synthetic data is generated from five repeats of the IBM. (a) Model A: $P_m = 0.5$, $\alpha = 0.1$, (b) model B: $P_m = 0.5$, $\alpha = 0.25$, (c) model B: $P_m = 0.5$, $\alpha = -0.1$. In all cases the red dot indicates the value of the parameters used to generate synthetic data.

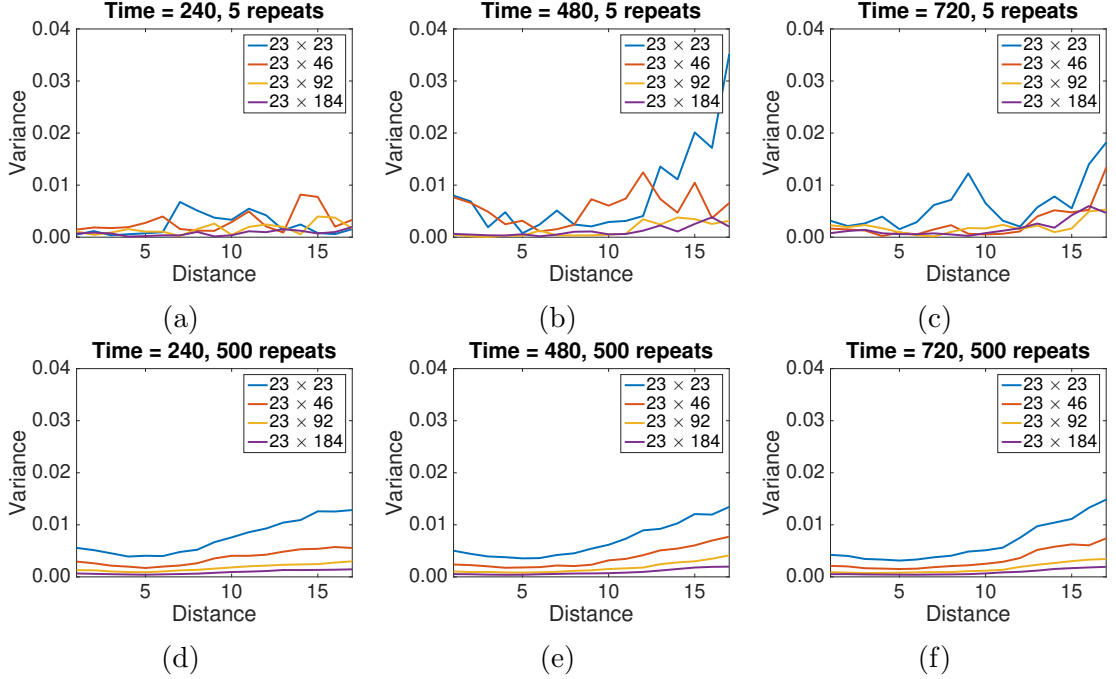


Figure 7.5: The variance in the PCF synthetic data for model B with $P_m = 0.5$, $\alpha = 0.25$ and different IBM domain sizes. Panels (a)-(c) display synthetic data generated from five repeats of the IBM, panels (d)-(f) display synthetic data generated from 500 repeats of the IBM. The domain size is indicated in the legend.

We conjectured that the variance in the summary statistics of the synthetic data could be reduced in two ways:

1. increasing the number of IBM repeats used to generate the synthetic data;
2. increasing the size of the IBM domain while keeping the column density of the initial conditions invariant. An example of this proposed initial condition is given in Fig. 7.6 (b). Importantly, increasing the size of the IBM domain increases the number of agents in the simulation, and can be thought of as equivalent to increasing the field of view of the microscope.

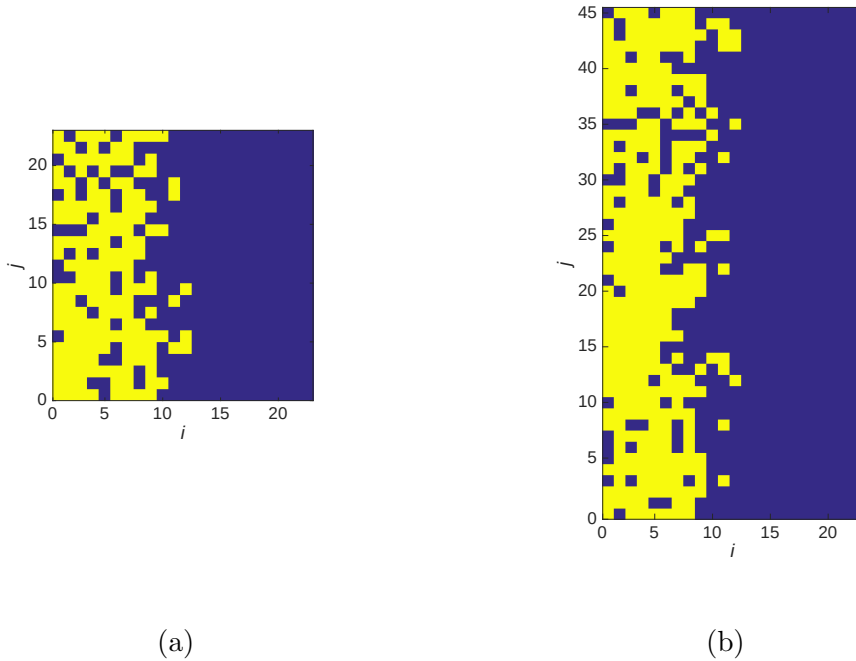


Figure 7.6: Increasing the size of the simulation domain while keeping the initial column densities the same. Panel (b) is twice the size of panel (a), however, the average initial density of each column is the same for both panels (a) and (b).

In Fig. 7.5 the variance in the PCF synthetic data for model B with $P_m = 0.5$ and $\alpha = 0.25$ for different domain sizes and varying numbers of repeats can be seen. It is evident that the variance in the PCF calculated from 500 repeats of a $L_x = 23$ by $L_y = 23$ sized domain (blue line in Fig. 7.5 (d)-(f)) is greater than the variance in the PCF calculated from five repeats of a $L_x = 23$ by $L_y = 184$ sized domain (purple line in Fig. 7.5 (a)-(c)). This can be understood by considering Eq. (7.7): the number of

occupied lattice pairs for each horizontal pair distance used to generate the PCF does not increase linearly with the number of agents. Specifically, the number of occupied lattice pairs for each horizontal pair distance that generates the PCF is proportional to¹³

$$\frac{N(N-1)}{2}. \quad (7.15)$$

Therefore, the identification of parameters in experimental data using the PCF as a summary statistic may be best facilitated by increasing the size of the domain upon which the experiment is performed, rather than increasing the number of repeats of an experiment with a smaller domain. Further variance plots for models A and B for the PCF summary statistic can be found in Appendix D.

It is important to note that it is also the case for the *agent density profile* synthetic data. If generated from 500 repeats of a $L_x = 23$ by $L_y = 23$ sized domain, the agent density profile synthetic data will have greater variance than the agent density profile synthetic data generated from five repeats of a $L_x = 23$ by $L_y = 184$ sized domain (data not shown). In this case the variance reduction is an artefact of the lattice-based model. This is because the density of each column in the IBM can take on a greater range of values between 0 and 1 as the column length is increased, leading to a reduction in variance in the agent density profile synthetic data (especially in the initial conditions of the simulations used to generate the synthetic data). However, as we do not use the agent density profile summary statistic to identify parameters in the current simulation design we do not pursue this matter further.

¹³This is not quite correct as a distance of ‘0’ between agents, that is they share the same column, is not accounted for in Eq. (7.7). To make Eq. (7.15) exact is not trivial as the expected number of agents each agent shares a column with depends on both the column position and simulation time.

7.3.3 Improving the experimental design

We now confirm that more accurate identification of synthetic data parameters occurs by expanding the domain upon which the experiment is performed, as opposed to increasing the number of experimental repeats.

In Fig. 7.7 (a)-(c) we plot the posterior distribution for synthetic data generated from 500 repeats of a $L_x = 23$ by $L_y = 23$ sized domain, while in Fig. 7.7 (d)-(f) we plot the posterior distribution generated from synthetic data generated five repeats of a $L_x = 23$ by $L_y = 184$ sized domain. As predicted, it is apparent that increasing the domain size is more effective for parameter identification than increasing the number of repeats used to generate the synthetic data. This is evident in the location (and narrow spread) of the posterior distribution relative to the red dot, whereby the posterior distribution is closer to the red dot in the case of Fig. 7.7 (d)-(f) compared to Fig. 7.7 (a)-(c). Despite this, the identification of the parameters for repulsive interactions remains somewhat elusive (Fig. 7.7 (f)). A possible reason for this is that the repulsive interaction we present here is a weak one, due to the constraint of Eqs. (7.2) and (7.4), and larger values of $|\alpha|$ are easier to identify as they have a more profound effect on the behaviour of the agent population.

Computing $D_{KL}(p|\pi)$ for all six plots in Fig. 7.7 gives: (a) 2.55; (b) 2.69; (c) 1.53; and (d) 3.69; (e) 2.97; (f) 3.54. In tandem with the proximity of the peak of the posterior distribution densities to the red dots in Fig. 7.7 (d)-(f) compared to Fig. 7.7 (a)-(c), this suggests that generating synthetic data on a larger domain is more effective for improving parameter identification than increasing the number of repeats used to generate the synthetic data.

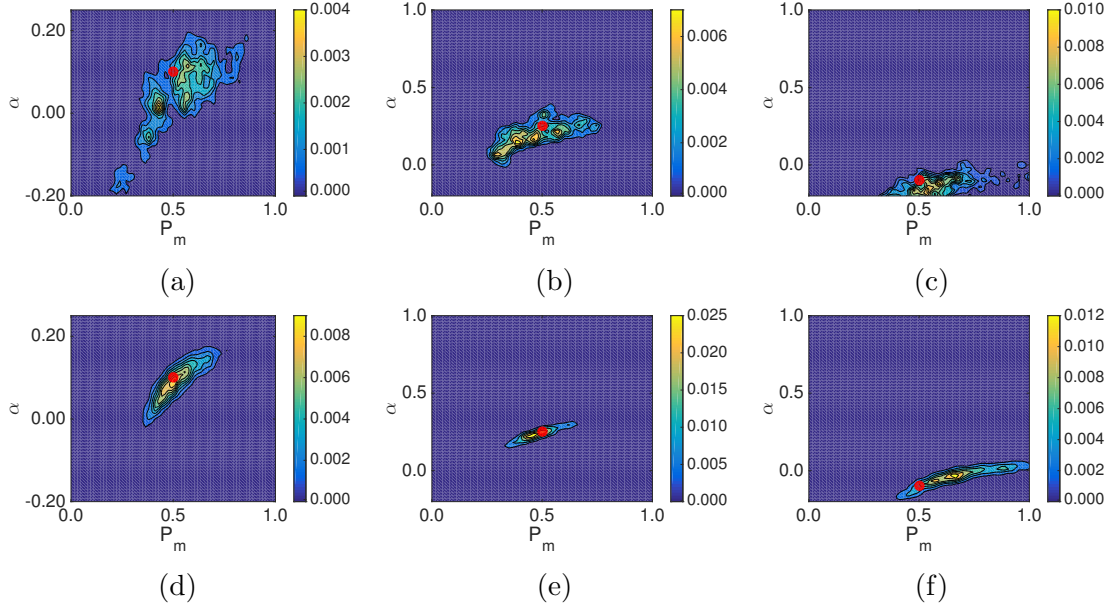


Figure 7.7: (a)-(c) Posterior distributions for simulations of the experiment using the PCF as a summary statistic for an IBM simulated on a domain of size $L_x = 23$ by $L_y = 23$ with synthetic data generated from 500 repeats. (a) Model A: $P_m = 0.5$, $\alpha = 0.1$, (b) model B: $P_m = 0.5$, $\alpha = 0.25$, (c) model B: $P_m = 0.5$, $\alpha = -0.1$. (d)-(f) Posterior distribution plots for simulations of the experiment using the PCF as a summary statistic for an IBM simulated on a domain of size $L_x = 23$ by $L_y = 184$ with synthetic data generated from five repeats. (a) Model A: $P_m = 0.5$, $\alpha = 0.1$, (b) model B: $P_m = 0.5$, $\alpha = 0.25$, (c) model B: $P_m = 0.5$, $\alpha = -0.1$. Further figure information can be found in Fig. 7.4.

7.4 Summary

In this chapter we have presented methods to identify motility and adhesion parameters in an IBM of a wound-healing assay. Our findings suggest that for a commonly performed experiment increasing the size of the experimental domain can be more effective in improving the accuracy of parameter identification, than increasing the number of repeats of the experiment. This is because increasing the size of the domain, which is equivalent to increasing the number of cells in the experiment, more effectively reduces the variance in the synthetic data from which the parameters are identified. The reason for this reduction in variance is explained by Eq. (7.7), where the number of agent pair counts that generate the PCF increases nonlinearly with

the number of agents on the domain. In addition, increasing the size of the experimental domain may make the collection of experimental data less time-consuming, as potentially fewer repeats of the experiment will need to be conducted. For instance, five repeats of the experiment on a larger domain provides more information about parameters than 500 repeats of the experiment on a smaller domain (in the examples we have presented in this chapter). Given this, it would be of great interest to carry out a comprehensive study of all commonly used summary statistics for analysing cell migration, and see how their variance scales with the size of the experimental domain. Not only would that put the work presented in this chapter on a firmer theoretical footing, but it would also provide the means by which to parameterise IBMs of cell migration more accurately.

We also studied implementing the average horizontal displacement of agents and the agent density profile as summary statistics. These were found to be less effective than the PCF in parameter identification. This was especially the case for the averaged agent displacement, whereby a range of adhesion and motility parameters could result in the same average agent displacement. This result suggests that agent displacement may not be a suitable summary statistic for estimating cell motility and adhesion parameters, due to parameter identifiability issues.

The obvious extension to the work presented in this chapter is to experimentally validate the findings. That is, expand the wound-healing experimental domain and demonstrate: i) that the cell migratory process can be effectively described by the model we have presented here; and ii) that the experimental parameters are identifiable with a larger experimental domain. If validated, alterations could be made to the IBM to try and further improve parameter identification, and evidence may be provided that demonstrates which adhesion model, A or B, is more applicable to the

cell type under consideration.

To conclude, the findings presented in this chapter will be of particular interest to those concerned with performing experiments that enable the effective parameterisation of cell migratory processes. In particular, cell migratory processes in which cell-cell adhesion or repulsion are known to play an important role. More generally, we have also suggested time and cost-saving alterations to a commonly performed experiment for identifying cell motility parameters.

Chapter 8

Discussion

Research into the effect of domain growth on many-body processes is not new. In 1927 Georges Lemaître originally proposed the “Cosmic Egg” hypothesis, which is now better known as the Big Bang theory. Work on domain growth also exists in quantum theory [119–121]. For example, examining the effect expansion of the canonical square well has on the wave function [122]. Statistical physics is also concerned with domain expansion, for instance in a Carnot engine, and what effect that has on macroscopic properties such as heat capacitance, energy, entropy, and so on [123]. A well-established collection of work exists on describing the evolution of reaction-diffusion systems on growing domains, especially in the context of Turing patterns [80]. Mathematical models describing tumour growth are prevalent in biological research, as are models describing growth in many other biological tissues [32].

Despite this long and varied research history, research into the effect of domain growth on many-body processes is far from complete. Typically, in many of the examples given above, and often with good reason, domain growth is implemented as a uniform stretch. In this thesis we have studied examples of domain growth where the domain is grown in discrete ‘units’. This has led to a range of interesting results, especially in the case of proliferative agent populations, whereby the method of implementing domain growth had dramatic consequences on the long term behaviour of agent pop-

ulations.

The work presented in this thesis, in conjunction with the biological and theoretical examples provided throughout, demonstrates that accurately describing the behaviour of motile and proliferative cells on growing domains is an important part of developmental biology research. We conclude this thesis by restating the main findings, and discussing a collection of both experimental and theoretical extensions to the work presented in this thesis.

8.1 Conclusions and future work

In Chapter 2 we motivated the use of mathematical and computational modelling in developmental biology research with a study of how melanoblasts colonise the DLE during murine embryonic development [10]. We demonstrated, through the use of an IBM parameterised by experimental data, that a simple mechanism consisting of undirected cell migration and proliferation, in tandem with domain growth, could account for the colonisation of the epidermis by melanoblasts in the embryonic mouse. We also demonstrated with our IBM that the formation of diffuse *LaacZ* patterns and chimeric stripes can arise through a *common* mechanism. Further experiments, using stochastic labelling methods such as brainbow and confetti [64], are required to definitively confirm our findings concerning the formation of pigmentation patterns in mice. Our model, in conjunction with experimental data, also suggests that impaired proliferation in melanoblasts is the cause of bellyspotting in mice rather than impaired migration. We showed that, in the model, colonisation of the DLE by melanoblasts is more sensitive to changes in the melanoblast proliferation rate, as opposed to the melanoblast motility rate. This suggests piebaldism is caused by a defect in melanoblast proliferation and not motility.

In Chapter 3 we extended our modelling approach to examine how cell migration is affected by cell-cell interactions, and included both the effects of domain growth and a chemoattractant [12]. From our studies we found that cell-cell interactions could be inferred from population density characteristics in the presence of a chemoattractant, and that these population density characteristics were conserved on both static and growing domains. Visualising the profile of the cells at the edge of a cell population in the presence of a chemoattractant is experimentally tractable, and so our findings could be experimentally tested. We also studied the expected displacement of a tagged cell in the cell population, and showed that different types of cell-cell interactions give rise to cell trajectories with distinct characteristics. These cell trajectory differences could be used to identify cell-cell interactions, and so is something that could be experimentally explored.

In Chapter 4 we examined how to derive accurate continuum approximations of IBMs with agent proliferation on growing domains. To do so we initially developed methods to include the effects of domain growth on the evolution of spatial correlations between agents in a continuum approximation of a one-dimensional model of cell motility and proliferation [13]. We included the effect of spatial correlations by deriving a system of ODEs that describe the expected evolution of individual and pairwise density functions for agents on a growing domain. We then demonstrated how to simplify this system of ODEs by using an appropriate approximation. This simplification allows domain growth to be included in models describing the evolution of spatial correlations between agents in a tractable manner. Possible extensions to the work presented in Chapter 4 would be to implement no-flux boundary conditions in the model, or non-uniform initial conditions. Mathematical extensions such as these could then be implemented in two-dimensional models. Perhaps a more pressing matter is further research into implementing alternative approaches to moment closure.

Both closures we implement (KSA and domain length closure) could possibly be replaced by closures that would result in more accurate correlation ODE models. It could also lead to developing less cumbersome equations than the current ones, which could save time and allow a more rigorous analysis of their properties. Therefore, the revisiting of the closure methods employed in deriving the equations in Chapters 4 to 6 is a topic of future research that could prove fruitful.

In Chapter 5 we presented methods to include the effects of domain growth on the evolution of spatial correlations between agents in a continuum approximation of a two-dimensional model of cell motility and proliferation [14, 15]. We showed that, depending on the way in which domain growth is implemented, different steady state densities are predicted for otherwise identical agent populations. Furthermore, we demonstrated that the way in which domain growth is implemented can cause the initial size of the domain to affect the evolution of the agent density. In Chapter 6 we extended the work presented in Chapter 5 to show that again, depending on the way in which domain growth is implemented, different agent species dominate in multispecies simulations. Continuum approximations of the IBM that ignore spatial correlations cannot capture any of the aforementioned behaviour, while those that account for spatial correlations can. The results in Chapters 5 and 6 show that the mechanism used to implement domain growth can determine the long term behaviour of agent populations, and in certain circumstances, establish domain patterning. The most obvious extension to the work presented in Chapters 5 and 6 is to try and experimentally validate the findings. Experimental methods have recently been developed that can control the expansion of a gel upon which a cell population is situated, and we are currently in discussions with experimental researchers to use this to test some of the findings we have presented.

In Chapter 7 we developed methods to identify parameters for cell motility and adhesion in a simple wound-healing experiment, which is an experiment commonly used to study cell migration. Building on the results of Chapter 3 we implemented different cell-cell interactions in an IBM simulating a wound-healing experiment, and addressed issues of parameter identifiability and experimental design. We showed that cell motility and adhesion parameters can be successfully retrieved from synthetic data using appropriate summary statistics and ABC. Using our IBM we then improved the experimental design of the wound-healing assay, and showed that identification of simulation parameters is more likely if the domain upon which the experiment is performed is expanded, as opposed to increasing the number of repeats of the experiment. For instance, 500 repeats of the experiment on a standard domain provides less information about simulation parameters than five repeats of the experiment on a domain eight times larger than the standard domain. The most pressing extension to the work presented in Chapter 7 is experimental validation of the findings. That is, expand the wound-healing experimental domain and demonstrate i) the cell migratory process can be effectively described by our IBM; and ii) show the experimental parameters are identifiable with a larger experimental domain. If validated, alterations could be made to the IBM to try and further improve parameter identification, and evidence may be provided that demonstrates which cell-cell adhesion model, A or B, is more applicable to the cells examined in Chapter 7. It would also be of great interest to carry out a comprehensive study of all commonly used summary statistics for analysing cell migration, and see how their variance scales with the size of the experimental domain. Not only would that recast the work presented in Chapter 7 in a more theoretical framework, and so allow for further analysis, but would also provide the means by which to parameterise IBMs of cell migration more accurately.

In this thesis we have presented methods to describe the behaviour of motile, prolifer-

ative and interactive cells on growing domains. Our work has generated many novel hypotheses, which will be of interest to both theoretical and experimental researchers alike.

Appendix A

Appendices for Chapter 3

A.1 Example transition probabilities

An example two-dimensional agent adhesion transition probability for a movement from $(i, j+1) \rightarrow (i, j)$ is given in Eq. (A.1):

$$\begin{aligned} T(i, j|i, j+1)_A = & (1 - \langle C_{i,j} \rangle) \left[\frac{(1 - \rho_x)}{4} \left[(1 - \langle C_{i+2,j} \rangle)(1 - \langle C_{i+1,j-1} \rangle)(1 - \langle C_{i+1,j+1} \rangle) \right] \right. \\ & + \frac{(1 - \rho_x)(1 - \alpha)}{4} \left[[(\langle C_{i+2,j} \rangle)(1 - \langle C_{i+1,j-1} \rangle)(1 - \langle C_{i+1,j+1} \rangle)] \right. \\ & + [(1 - \langle C_{i+2,j} \rangle)(1 - \langle C_{i+1,j-1} \rangle)(\langle C_{i+1,j+1} \rangle) \\ & + (1 - \langle C_{i+2,j} \rangle)(\langle C_{i+1,j-1} \rangle)(1 - \langle C_{i+1,j+1} \rangle)] \\ & + \frac{(1 - \rho_x)(1 - \alpha)^2}{4} \left[[(\langle C_{i+2,j} \rangle)(1 - \langle C_{i+1,j-1} \rangle)(\langle C_{i+1,j+1} \rangle) \right. \\ & + (\langle C_{i+2,j} \rangle)(\langle C_{i+1,j-1} \rangle)(1 - \langle C_{i+1,j+1} \rangle)] \\ & + [(1 - \langle C_{i+2,j} \rangle)(\langle C_{i+1,j-1} \rangle)(\langle C_{i+1,j+1} \rangle)] \\ & \left. \left. + \frac{(1 - \rho_x)(1 - \alpha)^3}{4} [(\langle C_{i+2,j} \rangle)(\langle C_{i+1,j-1} \rangle)(\langle C_{i+1,j+1} \rangle)] \right] \right]. \quad (\text{A.1}) \end{aligned}$$

An example two-dimensional myopic transition probability for a movement from

$(i, j+1) \rightarrow (i, j)$ is given in Eq. (A.2):

$$\begin{aligned}
T(i, j|i, j+1)_M = & (1 - \langle C_{i,j} \rangle) \left[\frac{(1 - \rho_x)}{4} \left[(1 - \langle C_{i+2,j} \rangle)(1 - \langle C_{i+1,j-1} \rangle)(1 - \langle C_{i+1,j+1} \rangle) \right] \right. \\
& + \frac{(1 - \rho_x)}{(3 - \rho_x)} \left[(\langle C_{i+2,j} \rangle)(1 - \langle C_{i+1,j-1} \rangle)(1 - \langle C_{i+1,j+1} \rangle) \right] \\
& + \frac{(1 - \rho_x)}{3} \left[(1 - \langle C_{i+2,j} \rangle)(1 - \langle C_{i+1,j-1} \rangle)(\langle C_{i+1,j+1} \rangle) \right. \\
& + (1 - \langle C_{i+2,j} \rangle)(\langle C_{i+1,j-1} \rangle)(1 - \langle C_{i+1,j+1} \rangle) \left. \right] \\
& + \frac{(1 - \rho_x)}{(2 - \rho_x)} \left[(\langle C_{i+2,j} \rangle)(1 - \langle C_{i+1,j-1} \rangle)(\langle C_{i+1,j+1} \rangle) \right. \\
& + (\langle C_{i+2,j} \rangle)(\langle C_{i+1,j-1} \rangle)(1 - \langle C_{i+1,j+1} \rangle) \left. \right] \\
& + \frac{(1 - \rho_x)}{2} \left[(1 - \langle C_{i+2,j} \rangle)(\langle C_{i+1,j-1} \rangle)(\langle C_{i+1,j+1} \rangle) \right] \\
& \left. + \left[(\langle C_{i+2,j} \rangle)(\langle C_{i+1,j-1} \rangle)(\langle C_{i+1,j+1} \rangle) \right] \right]. \tag{A.2}
\end{aligned}$$

Appendix B

Appendices for Chapter 5

B.1 Individual density functions

To include the effect of agent motility and proliferation in Eqs. (5.9) we follow a similar procedure as presented in Chapter 4. To do so we are required to introduce further notation. We indicate a site unoccupied by an agent in the following manner: $\rho_1^{N_x \times N_y}(\mathbf{m}'; t)$, that is, the joint probability of lattice site $\mathbf{m}$ not being occupied on a domain of size $N_x \times N_y$ at time t . We also introduce the summation,

$$\begin{aligned} \sum_{\mathbf{n}}^{\mathbf{m}} \rho_2^{N_x \times N_y}(\mathbf{m}, \mathbf{n}; t) &= \rho_2^{N_x \times N_y}((i, j), (i+1, j); t) + \rho_2^{N_x \times N_y}((i, j), (i-1, j); t) \\ &\quad + \rho_2^{N_x \times N_y}((i, j), (i, j+1); t) + \rho_2^{N_x \times N_y}((i, j), (i, j-1); t), \end{aligned} \tag{B.1}$$

which indexes over the von Neumann neighbourhood of the site indicated by the upper limit of the sum.

The sum of the individual density functions on a domain of size $N_x \times N_y$ at $[t + \delta t)$

for motile and proliferating agents is

$$\begin{aligned}
 \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \rho_1^{N_x \times N_y}(\mathbf{m}; t + \delta t) &= (1 - \delta t P_{g_x} N_x - \delta t P_{g_y} N_y) \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \rho_1^{N_x \times N_y}(\mathbf{m}; t) \\
 &+ \delta t P_{g_x} \sum_{i=1}^{N_x-1} \sum_{j=1}^{N_y} (N_x - 1) \rho_1^{(N_x-1) \times N_y}(\mathbf{m}; t) \\
 &+ \delta t P_{g_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y-1} (N_y - 1) \rho_1^{N_x \times (N_y-1)}(\mathbf{m}; t) \\
 &+ \delta t \frac{P_m}{4} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[\sum_{\mathbf{n}}^{\mathbf{m}} \left[\rho_2^{N_x \times N_y}(\mathbf{m}', \mathbf{n}; t) - \rho_2^{N_x \times N_y}(\mathbf{m}, \mathbf{n}'; t) \right] \right] \\
 &+ \delta t \frac{P_p}{4} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[\sum_{\mathbf{n}}^{\mathbf{m}} \left[\rho_2^{N_x \times N_y}(\mathbf{m}', \mathbf{n}; t) \right] \right] + \mathcal{O}(\delta t^2).
 \end{aligned} \tag{B.2}$$

As we assume translational invariance, Eq. (B.2) can be simplified to obtain

$$\begin{aligned}
 \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \rho_1^{N_x \times N_y}(\mathbf{m}; t + \delta t) &= (1 - \delta t P_{g_x} N_x - \delta t P_{g_y} N_y) \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \rho_1^{N_x \times N_y}(\mathbf{m}; t) \\
 &+ \delta t P_{g_x} \sum_{i=1}^{N_x-1} \sum_{j=1}^{N_y} (N_x - 1) \rho_1^{(N_x-1) \times N_y}(\mathbf{m}; t) \\
 &+ \delta t P_{g_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y-1} (N_y - 1) \rho_1^{N_x \times (N_y-1)}(\mathbf{m}; t) \\
 &+ \delta t \frac{P_p}{4} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[\sum_{\mathbf{n}}^{\mathbf{m}} \rho_2^{N_x \times N_y}(\mathbf{m}', \mathbf{n}; t) \right] + \mathcal{O}(\delta t^2). \tag{B.3}
 \end{aligned}$$

If we recognise that $\rho_2^{N_x \times N_y}(\mathbf{m}', \mathbf{n}; t) = \rho_1^{N_x \times N_y}(\mathbf{m}; t) - \rho_2^{N_x \times N_y}(1, 0; t)$ we can rewrite

Eq. (B.3) as

$$\begin{aligned}
 \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \rho_1^{N_x \times N_y}(\mathbf{m}; t + \delta t) &= (1 - \delta t P_{g_x} N_x - \delta t P_{g_y} N_y) \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \rho_1^{N_x \times N_y}(\mathbf{m}; t) \\
 &+ \delta t P_{g_x} \sum_{i=1}^{N_x-1} \sum_{j=1}^{N_y} (N_x - 1) \rho_1^{(N_x-1) \times N_y}(\mathbf{m}; t) \\
 &+ \delta t P_{g_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y-1} (N_y - 1) \rho_1^{N_x \times (N_y-1)}(\mathbf{m}; t) \\
 &+ \delta t P_p \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[\rho_1^{N_x \times N_y}(\mathbf{m}; t) - \rho_2^{N_x \times N_y}(1, 0; t) \right] + \mathcal{O}(\delta t^2).
 \end{aligned} \tag{B.4}$$

From Eq. (B.4) we can obtain

$$\frac{dc(t)}{dt} = P_p \left(c(t) - \rho_2^{N_x \times N_y}(1, 0; t) \right) - (P_{g_x} + P_{g_y}) c(t). \tag{B.5}$$

As can be seen from Eq. (B.5) the inclusion of agent proliferation means that pairwise density functions are now present in the equations for the evolution of the individual density functions, which is not the case without agent proliferation (Eq. (5.9)).

B.2 Pairwise density functions

The derivation for the evolution of the pairwise density functions can be found in [18], and the equations are stated in the supplementary material of [18]. However, we also outline the derivation below. The addition of agent motility and proliferation is the same for GM1 and GM2. For agents colinear in the horizontal direction, that is, $r_y = 0$, and $r_x > 1$, the evolution of the pairwise density functions for motile and

proliferative agents with GM1 is

$$\begin{aligned}
 \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \rho_2^{N_x \times N_y}(r_x, 0; t + \delta t) = & \\
 & (1 - \delta t P_{g_x} N_x - \delta t P_{g_y} N_y) \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \rho_2^{N_x \times N_y}(r_x, 0; t) \\
 & + \delta t P_{g_x} \sum_{i=1}^{N_x-1} \sum_{j=1}^{N_y} (r_x - 1) \rho_2^{(N_x-1) \times N_y}(r_x - 1, 0; t) \\
 & + \delta t P_{g_x} \sum_{i=1}^{N_x-1} \sum_{j=1}^{N_y} (N_x - 1 - r_x) \rho_2^{(N_x-1) \times N_y}(r_x, 0; t) \\
 & + \delta t P_{g_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y-1} (N_y - 1) \rho_2^{N_x \times (N_y-1)}(r_x, 0; t) \\
 & + \delta t \frac{P_m}{4} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[\sum_{\mathbf{k}}^{\mathbf{n}} \left[\rho_3^{N_x \times N_y}(\mathbf{m}, \mathbf{n}', \mathbf{k}; t) - \rho_3^{N_x \times N_y}(\mathbf{m}, \mathbf{n}, \mathbf{k}'; t) \right] \right. \\
 & + \left. \sum_{\mathbf{k}}^{\mathbf{m}} \left[\rho_3^{N_x \times N_y}(\mathbf{m}', \mathbf{n}, \mathbf{k}; t) - \rho_3^{N_x \times N_y}(\mathbf{m}, \mathbf{n}, \mathbf{k}'; t) \right] \right] \\
 & + \delta t \frac{P_p}{4} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[\sum_{\mathbf{k}}^{\mathbf{m}} \left[\rho_3^{N_x \times N_y}(\mathbf{m}', \mathbf{n}, \mathbf{k}; t) \right] + \sum_{\mathbf{k}}^{\mathbf{n}} \left[\rho_3^{N_x \times N_y}(\mathbf{m}, \mathbf{n}', \mathbf{k}; t) \right] \right] \\
 & + \mathcal{O}(\delta t^2). \tag{B.6}
 \end{aligned}$$

If $r_x = 1$

$$\begin{aligned}
 & \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \rho_2^{N_x \times N_y}(r_x, 0; t + \delta t) = \\
 & (1 - \delta t P_{g_x} N_x - \delta t P_{g_y} N_y) \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \rho_2^{N_x \times N_y}(r_x, 0; t) \\
 & + \delta t P_{g_x} \sum_{i=1}^{N_x-1} \sum_{j=1}^{N_y} (r_x - 1) \rho_2^{(N_x-1) \times N_y}(r_x - 1, 0; t) \\
 & + \delta t P_{g_x} \sum_{i=1}^{N_x-1} \sum_{j=1}^{N_y} (N_x - 1 - r_x) \rho_2^{(N_x-1) \times N_y}(r_x, 0; t) \\
 & + \delta t P_{g_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y-1} (N_y - 1) \rho_2^{N_x \times (N_y-1)}(r_x, 0; t) \\
 & + \delta t \frac{P_m}{4} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[\sum_{\mathbf{k} \neq \mathbf{m}}^{\mathbf{n}} \left[\rho_3^{N_x \times N_y}(\mathbf{m}, \mathbf{n}', \mathbf{k}; t) - \rho_3^{N_x \times N_y}(\mathbf{m}, \mathbf{n}, \mathbf{k}'; t) \right] \right. \\
 & \left. + \sum_{\mathbf{k} \neq \mathbf{n}}^{\mathbf{m}} \left[\rho_3^{N_x \times N_y}(\mathbf{m}', \mathbf{n}, \mathbf{k}; t) - \rho_3^{N_x \times N_y}(\mathbf{m}, \mathbf{n}, \mathbf{k}'; t) \right] \right] \\
 & + \delta t \frac{P_p}{4} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[\sum_{\mathbf{k} \neq \mathbf{n}}^{\mathbf{m}} \left[\rho_3^{N_x \times N_y}(\mathbf{m}', \mathbf{n}, \mathbf{k}; t) \right] + \rho_2^{N_x \times N_y}(\mathbf{m}', \mathbf{n}; t) \right. \\
 & \left. + \sum_{\mathbf{k} \neq \mathbf{m}}^{\mathbf{n}} \left[\rho_3^{N_x \times N_y}(\mathbf{m}, \mathbf{n}', \mathbf{k}; t) \right] + \rho_2^{N_x \times N_y}(\mathbf{m}, \mathbf{n}'; t) \right] \\
 & + \mathcal{O}(\delta t^2), \tag{B.7}
 \end{aligned}$$

from which we can obtain the necessary equations.

Appendix C

Appendices for Chapter 6

C.1 Individual density functions

We first display how to include agent proliferation, motility and death in the individual density functions. To do so we are required to introduce further notation. We indicate a site unoccupied by an agent in the following manner: $\rho_0^{N_x \times N_y}(\mathbf{m}; t)$, that is, the joint probability of lattice site $\mathbf{m}$ not being occupied on a domain of size $N_x \times N_y$ at time t . We indicate a site unoccupied by an agent in a pairwise density function in the following manner, $\rho_{0,A}^{N_x \times N_y}(\mathbf{m}, \mathbf{n}; t)$, that is, the joint probability of lattice site $\mathbf{m}$ not being occupied, and lattice site $\mathbf{n}$ being occupied by species A on a domain of size $N_x \times N_y$ at time t . We also introduce the summation,

$$\begin{aligned} \sum_{\mathbf{n}} \rho_{A,A}^{N_x \times N_y}(\mathbf{m}, \mathbf{n}; t) &= \rho_{A,A}^{N_x \times N_y}((i, j), (i+1, j); t) + \rho_{A,A}^{N_x \times N_y}((i, j), (i-1, j); t) \\ &\quad + \rho_{A,A}^{N_x \times N_y}((i, j), (i, j+1); t) + \rho_{A,A}^{N_x \times N_y}((i, j), (i, j-1); t), \end{aligned} \tag{C.1}$$

which indexes over the von Neumann neighbourhood of the site indicated as the upper index.

The sum of the individual density functions on a domain of size $N_x \times N_y$ at $[t + \delta t)$

for motile and proliferating agents is

$$\begin{aligned}
 \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \rho_A^{N_x \times N_y}(\mathbf{m}; t + \delta t) &= (1 - \delta t P_{g_x} N_x - \delta t P_{g_y} N_y) \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \rho_A^{N_x \times N_y}(\mathbf{m}; t) \\
 &+ \delta t P_{g_x} \sum_{i=1}^{N_x-1} \sum_{j=1}^{N_y} (N_x - 1) \rho_A^{(N_x-1) \times N_y}(\mathbf{m}; t) \\
 &+ \delta t P_{g_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y-1} (N_y - 1) \rho_A^{N_x \times (N_y-1)}(\mathbf{m}; t) \\
 &+ \delta t \frac{P_m^A}{4} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[\sum_{\mathbf{n}} \left[\rho_{0,A}^{N_x \times N_y}(\mathbf{m}, \mathbf{n}; t) - \rho_{A,0}^{N_x \times N_y}(\mathbf{m}, \mathbf{n}; t) \right] \right] \\
 &+ \delta t \frac{P_p^A}{4} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[\sum_{\mathbf{n}} \left[\rho_{0,A}^{N_x \times N_y}(\mathbf{m}, \mathbf{n}; t) \right] \right] \\
 &- \delta t P_d^A \rho_A^{N_x \times N_y}(\mathbf{m}; t) + \mathcal{O}(\delta t^2). \tag{C.2}
 \end{aligned}$$

As we assume translational invariance Eq. (C.2) can be simplified to obtain

$$\begin{aligned}
 \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \rho_A^{N_x \times N_y}(\mathbf{m}; t + \delta t) &= (1 - \delta t P_{g_x} N_x - \delta t P_{g_y} N_y) \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \rho_A^{N_x \times N_y}(\mathbf{m}; t) \\
 &+ \delta t P_{g_x} \sum_{i=1}^{N_x-1} \sum_{j=1}^{N_y} (N_x - 1) \rho_A^{(N_x-1) \times N_y}(\mathbf{m}; t) \\
 &+ \delta t P_{g_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y-1} (N_y - 1) \rho_A^{N_x \times (N_y-1)}(\mathbf{m}; t) \\
 &+ \delta t \frac{P_p^A}{4} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[\sum_{\mathbf{n}} \rho_{0,A}^{N_x \times N_y}(\mathbf{m}, \mathbf{n}; t) \right] \\
 &- \delta t P_d^A \rho_A^{N_x \times N_y}(\mathbf{m}; t) + \mathcal{O}(\delta t^2). \tag{C.3}
 \end{aligned}$$

If we recognise that $\rho_{0,A}^{N_x \times N_y}(\mathbf{m}, \mathbf{n}; t) = \rho_A^{N_x \times N_y}(\mathbf{m}; t) - \rho_{A,A}^{N_x \times N_y}(1, 0; t) - \rho_{B,A}^{N_x \times N_y}(1, 0; t)$

we can rewrite Eq. (C.3) as

$$\begin{aligned}
 \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \rho_A^{N_x \times N_y}(\mathbf{m}; t + \delta t) = & \\
 & (1 - \delta t P_{g_x} N_x - \delta t P_{g_y} N_y) \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \rho_A^{N_x \times N_y}(\mathbf{m}; t) \\
 & + \delta t P_{g_x} \sum_{i=1}^{N_x-1} \sum_{j=1}^{N_y} (N_x - 1) \rho_A^{(N_x-1) \times N_y}(\mathbf{m}; t) \\
 & + \delta t P_{g_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y-1} (N_y - 1) \rho_A^{N_x \times (N_y-1)}(\mathbf{m}; t) \\
 & + \delta t P_p^A \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[\rho_A^{N_x \times N_y}(\mathbf{m}; t) - \rho_{A,A}^{N_x \times N_y}(1, 0; t) - \rho_{B,A}^{N_x \times N_y}(1, 0; t) \right] \\
 & - \delta t P_d^A \rho_A^{N_x \times N_y}(\mathbf{m}; t) + \mathcal{O}(\delta t^2). \tag{C.4}
 \end{aligned}$$

From Eq. (C.4) we can obtain

$$\frac{dc_A(t)}{dt} = P_p^A (c_A(t) - \rho_{A,A}^{N_x \times N_y}(1, 0; t) - \rho_{A,B}^{N_x \times N_y}(1, 0; t)) - P_d^A c_A(t) - (P_{g_x} + P_{g_y}) c_A(t), \tag{C.5}$$

which is Eq. (6.7).

C.2 Pairwise density functions

The derivation of the pairwise density functions for multispecies without domain growth can be found in [21]. However, we outline the derivation for the auto-correlations below. To do so we require some further notation. We indicate a site being unoccupied by an agent in a triplet density function in the following manner: $\rho_{0,A,A}^{N_x \times N_y}(\mathbf{m}, \mathbf{n}, \mathbf{k}; t)$, that is, the joint probability of lattice site $\mathbf{m}$ not being occupied, lattice site $\mathbf{n}$ being occupied by species A , and lattice site $\mathbf{k}$ being occupied by species A on a domain of size $N_x \times N_y$ at time t . The addition of agent motility and prolifer-

ation is the same for GM1 and GM2. For agents colinear in the horizontal direction, that is, r_y is zero and $r_x > 1$, the evolution of the pairwise density functions for motile and proliferative agents with GM1 is

$$\begin{aligned}
 \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \rho_{A,A}^{N_x \times N_y}(r_x, 0; t + \delta t) = & \\
 & (1 - \delta t P_{g_x} N_x - \delta t P_{g_y} N_y) \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \rho_{A,A}^{N_x \times N_y}(r_x, 0; t) \\
 & + \delta t P_{g_x} \sum_{i=1}^{N_x-1} \sum_{j=1}^{N_y} (r_x - 1) \rho_{A,A}^{(N_x-1) \times N_y}(r_x - 1, 0; t) \\
 & + \delta t P_{g_x} \sum_{i=1}^{N_x-1} \sum_{j=1}^{N_y} (N_x - 1 - r_x) \rho_{A,A}^{(N_x-1) \times N_y}(r_x, 0; t) \\
 & + \delta t P_{g_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y-1} (N_y - 1) \rho_{A,A}^{N_x \times (N_y-1)}(r_x, 0; t) \\
 & + \delta t \frac{P_m^A}{4} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[\sum_{\mathbf{k}}^{\mathbf{n}} \left[\rho_{A,0,A}^{N_x \times N_y}(\mathbf{m}, \mathbf{n}, \mathbf{k}; t) - \rho_{A,A,0}^{N_x \times N_y}(\mathbf{m}, \mathbf{n}, \mathbf{k}; t) \right] \right. \\
 & \left. + \sum_{\mathbf{k}}^{\mathbf{m}} \left[\rho_{0,A,A}^{N_x \times N_y}(\mathbf{m}, \mathbf{n}, \mathbf{k}; t) - \rho_{A,A,0}^{N_x \times N_y}(\mathbf{m}, \mathbf{n}, \mathbf{k}; t) \right] \right] \\
 & + \delta t \frac{P_p^A}{4} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[\sum_{\mathbf{k}}^{\mathbf{m}} \left[\rho_{0,A,A}^{N_x \times N_y}(\mathbf{m}, \mathbf{n}, \mathbf{k}; t) \right] + \sum_{\mathbf{k}}^{\mathbf{n}} \left[\rho_{A,0,A}^{N_x \times N_y}(\mathbf{m}, \mathbf{n}, \mathbf{k}; t) \right] \right] \\
 & - \delta t 2 P_d^A \rho_{A,A}^{N_x \times N_y}(r_x, 0; t) + \mathcal{O}(\delta t^2). \tag{C.6}
 \end{aligned}$$

If $r_x = 1$

$$\begin{aligned}
 & \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \rho_{A,A}^{N_x \times N_y}(r_x, 0; t + \delta t) = \\
 & (1 - \delta t P_{g_x} N_x - \delta t P_{g_y} N_y) \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \rho_{A,A}^{N_x \times N_y}(r_x, 0; t) \\
 & + \delta t P_{g_x} \sum_{i=1}^{N_x-1} \sum_{j=1}^{N_y} (r_x - 1) \rho_{A,A}^{(N_x-1) \times N_y}(r_x - 1, 0; t) \\
 & + \delta t P_{g_x} \sum_{i=1}^{N_x-1} \sum_{j=1}^{N_y} (N_x - 1 - r_x) \rho_{A,A}^{(N_x-1) \times N_y}(r_x, 0; t) \\
 & + \delta t P_{g_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y-1} (N_y - 1) \rho_{A,A}^{N_x \times (N_y-1)}(r_x, 0; t) \\
 & + \delta t \frac{P_m^A}{4} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[\sum_{\mathbf{k} \neq \mathbf{m}}^{\mathbf{n}} \left[\rho_{A,0,A}^{N_x \times N_y}(\mathbf{m}, \mathbf{n}, \mathbf{k}; t) - \rho_{A,A,0}^{N_x \times N_y}(\mathbf{m}, \mathbf{n}, \mathbf{k}; t) \right] \right. \\
 & \left. + \sum_{\mathbf{k} \neq \mathbf{n}}^{\mathbf{m}} \left[\rho_{0,A,A}^{N_x \times N_y}(\mathbf{m}, \mathbf{n}, \mathbf{k}; t) - \rho_{A,A,0}^{N_x \times N_y}(\mathbf{m}, \mathbf{n}, \mathbf{k}; t) \right] \right] \\
 & + \delta t \frac{P_p^A}{4} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[\sum_{\mathbf{k} \neq \mathbf{n}}^{\mathbf{m}} \left[\rho_{0,A,A}^{N_x \times N_y}(\mathbf{m}, \mathbf{n}, \mathbf{k}; t) \right] + \rho_{0,A}^{N_x \times N_y}(\mathbf{m}, \mathbf{n}; t) \right. \\
 & \left. + \sum_{\mathbf{k} \neq \mathbf{m}}^{\mathbf{n}} \left[\rho_{A,0,A}^{N_x \times N_y}(\mathbf{m}, \mathbf{n}, \mathbf{k}; t) \right] + \rho_{A,0}^{N_x \times N_y}(\mathbf{m}, \mathbf{n}; t) \right] \\
 & - \delta t 2 P_d^A \rho_{A,A}^{N_x \times N_y}(r_x, 0; t) + \mathcal{O}(\delta t^2), \tag{C.7}
 \end{aligned}$$

from which we can obtain the necessary equations [21].

Appendix D

Appendices for Chapter 7

D.1 Experimental methods

The details of the experiment we aim to identify cell motility and adhesion parameters from is as follows: Fucci2a 3T3 flp-In cells were maintained in dulbeccos modified eagle medium (DMEM) containing 10% fetal calf serum, 1% Penicillin/Streptomycin and 100µg/ml Hygromycin B. A silicon well (Ibidi) was attached to the surface of a 24 well glass bottomed plate (Greiner bio-one) by surface tension and allowed to attach overnight. Cells were plated within the insert and allowed to attach phenol-red free DMEM (Biochrom) containing 10% fetal calf serum, and 1% Penicillin/Streptomycin. Cells migrating from the leading edge of the cell mass were then imaged with a 20x objective using a Nikon A1R inverted confocal microscope in a heated chamber supplied with 5% CO₂ in air. All image analysis tasks (required to generate the initial conditions for the IBM of a practically realisable experiment) were performed using custom written macros for the Fiji [124] distribution of ImageJ an open source image analysis package based on NIH Image [125]. The cell nucleus of each cell was identified by merging of the green and red channels containing the Fucci signal followed by segmentation. The centre of mass of each object in the segmented image was then determined automatically.

In total we have data from five repeats of the experiment. Each data set contains cell

track data for every cell for 64 hours imaged at 20 minute intervals. Therefore, we have the information required to apply our summary statistics to the experimental data. More specifically, we have the position of all cells at each time interval so that the expected horizontal displacement of cells, cell density profile, and PCF may be computed.

D.2 Further variance plots for models A and B for the PCF summary statistic

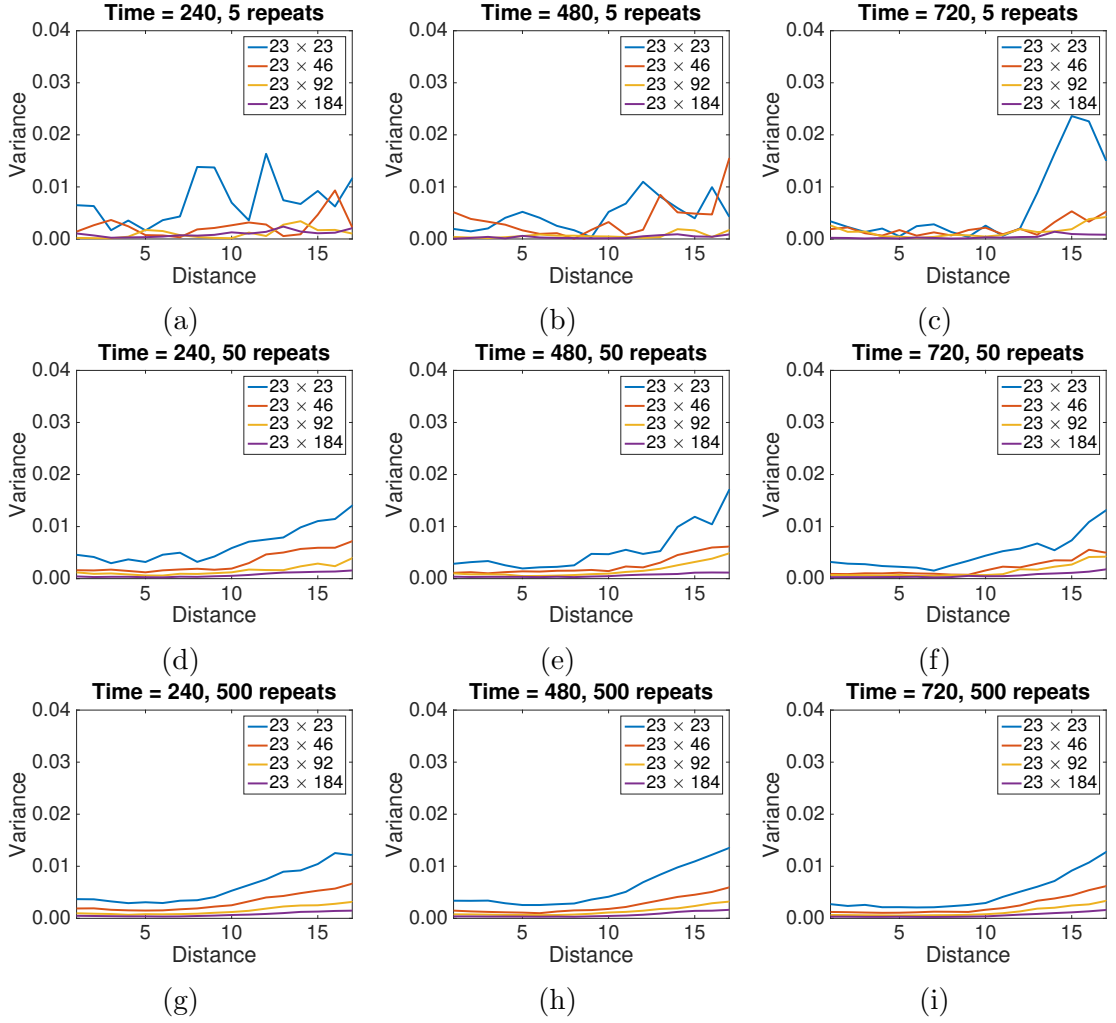


Figure D.1: The variance in PCF synthetic data for model A with $P_m = 0.5$, $\alpha = 0.1$ for different IBM domain sizes. Panels (a)-(c) display synthetic data generated from five repeats of the IBM, panels (d)-(f) display synthetic data generated from 50 repeats of the IBM and panels (g)-(i) display synthetic data generated from 500 repeats of the IBM.

D.2. Further variance plots for models A and B for the PCF summary statistic

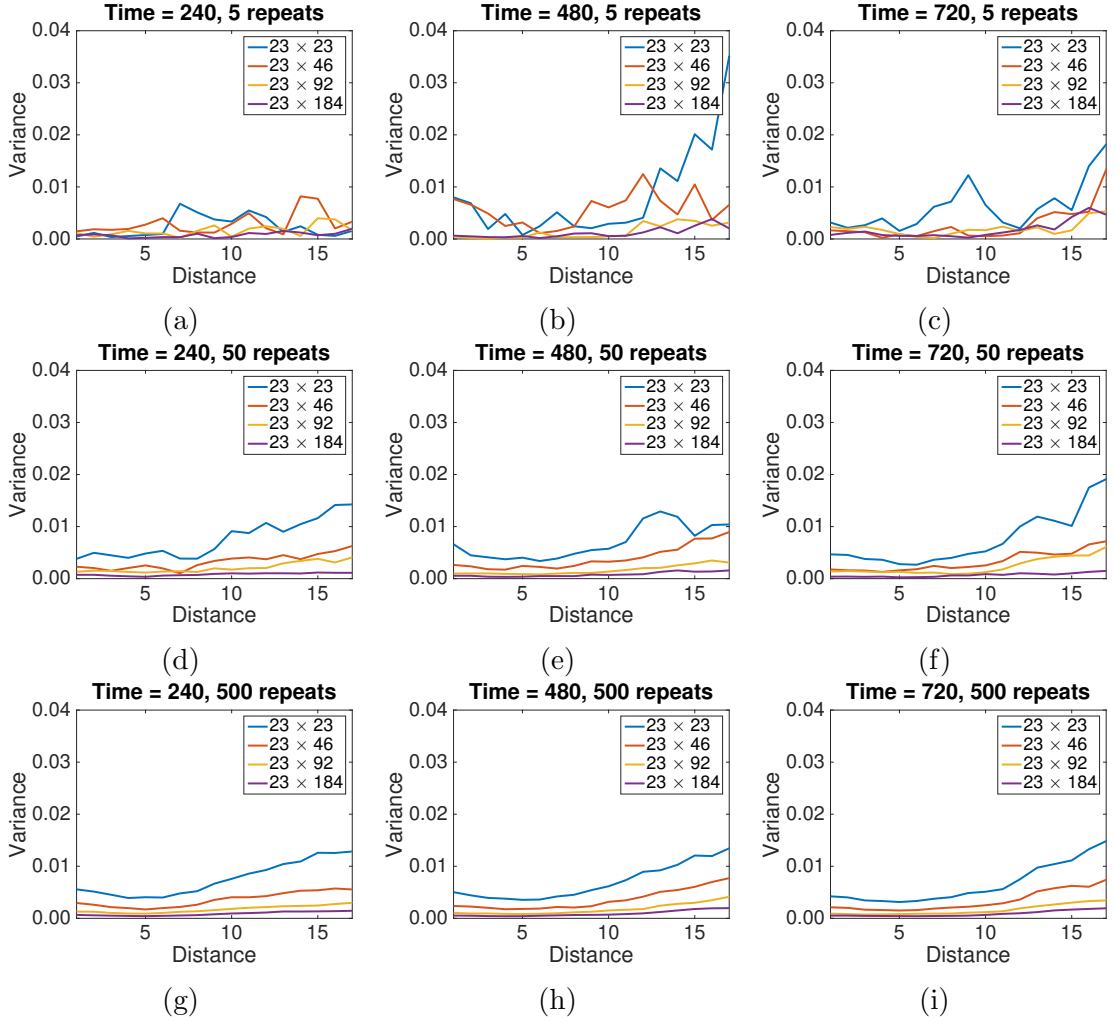


Figure D.2: The variance in the synthetic data for the adhesion model B with $P_m = 0.5$, $\alpha = 0.25$ for different IBM domain sizes. Panels (a)-(c) display synthetic data generated from five repeats of the IBM, panels (d)-(f) display synthetic data generated from 50 repeats of the IBM and panels (g)-(i) display synthetic data generated from 500 repeats of the IBM.

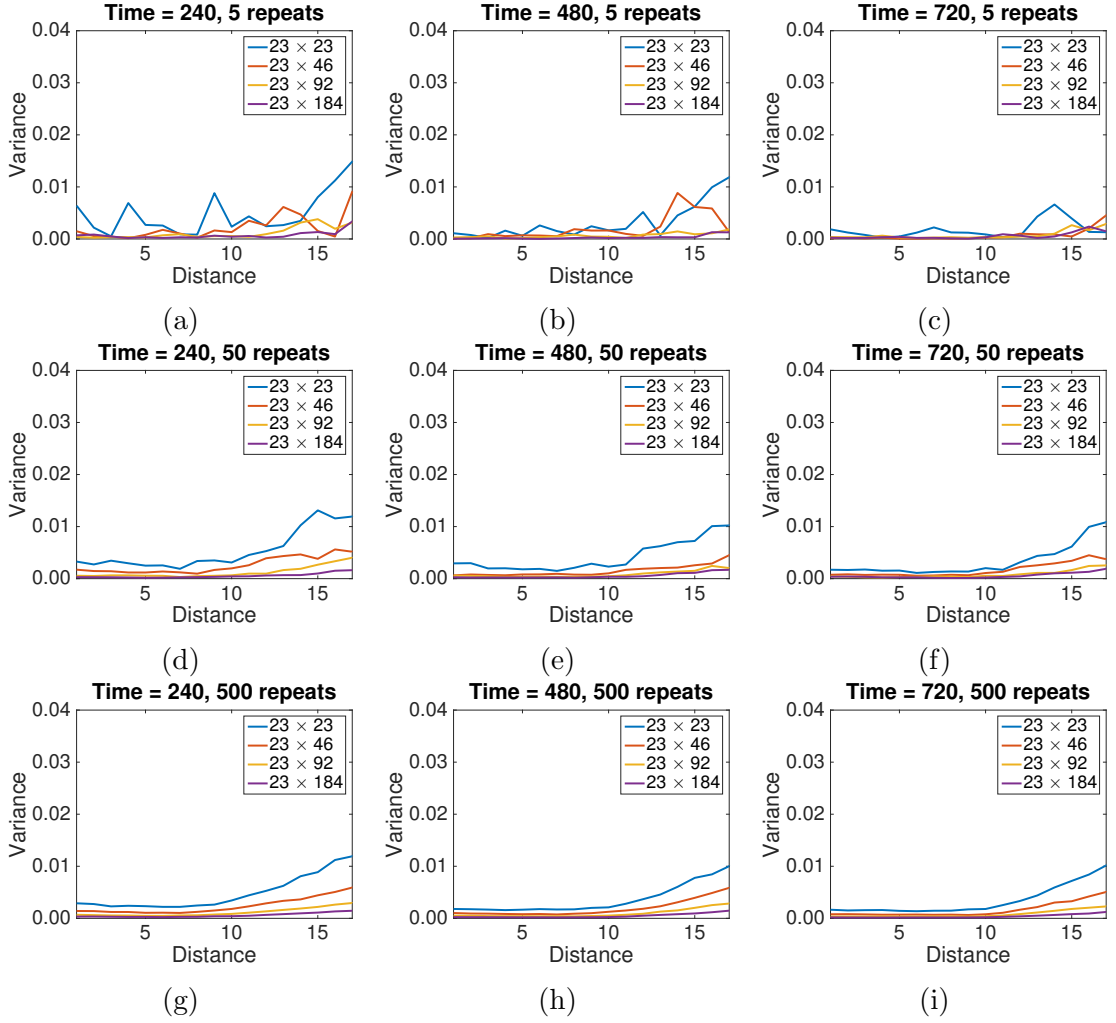


Figure D.3: The variance in PCF synthetic data for model B with $P_m = 0.5$, $\alpha = -0.1$ for different IBM domain sizes. Panels (a)-(c) display synthetic data generated from five repeats of the IBM, panels (d)-(f) display synthetic data generated from 50 repeats of the IBM and panels (g)-(i) display synthetic data generated from 500 repeats of the IBM.

Bibliography

- [1] R. McLennan, L. Dyson, K. W. Prather, J. A. Morrison, R. E. Baker, P. K. Maini, and P. M. Kulesa. Multiscale mechanisms of cell migration during development: theory and experiment. *Development*, 139(16):2935–2944, 2012.
- [2] R. McLennan, L. J. Schumacher, J. A. Morrison, J. M. Teddy, D. A. Ridenour, A. C. Box, C. L. Semerad, H. Li, W. McDowell, D. Kay, P. K. Maini, R. E. Baker, and P. M. Kulesa. Neural crest migration is driven by a few trailblazer cells with a unique molecular signature narrowly confined to the invasive front. *Development*, 142(11):2014–2025, 2015.
- [3] R. McLennan, L. J. Schumacher, J. A. Morrison, J. M. Teddy, D. A. Ridenour, A. C. Box, C. L. Semerad, H. Li, W. McDowell, D. Kay, P. K. Maini, R. E. Baker, and P. M. Kulesa. VEGF signals induce trailblazer cell identity that drives neural crest migration. *Developmental Biology*, 407(1):12–25, 2015.
- [4] B. Wehrle-Haller and J. A. Weston. Soluble and cell-bound forms of steel factor activity play distinct roles in melanocyte precursor dispersal and survival on the lateral neural crest migration pathway. *Development*, 121:731–42, 1995.
- [5] J. A. Weston. Sequential segregation and fate of developmentally restricted intermediate cell populations in the neural crest lineage. *Current Topics in Developmental Biology*, 25:133–53, 1991.
- [6] A. L. Wilkie, S. A. Jordan, and I. J. Jackson. Neural crest progenitors of

- the melanocyte lineage: coat colour patterns revisited. *Development*, 129(14): 3349–3357, 2002.
- [7] A. Aman and T. Piotrowski. Cell migration during morphogenesis. *Developmental Biology*, 341(1):20–33, 2010.
- [8] J. Liepe, S. Filippi, M. Komorowski, and M. P. H. Stumpf. Maximizing the information content of experiments in systems biology. *PLoS Computational Biology*, 9(1):e1002888, 2013.
- [9] S. T. Johnston, J. V. Ross, B. J. Binder, D. L. . McElwain, P. Haridas, and M. J. Simpson. Quantifying the effect of experimental design choices for *in vitro* scratch assays. *Journal of Theoretical Biology*, 400:19–31, 2016.
- [10] R. L. Mort, R. J. H. Ross, K. J. Hailey, O. Harrison, M. A. Keighren, G. Landini, R. E. Baker, K. J. Painter, I. J. Jackson, and C. A. Yates. Reconciling diverse mammalian pigmentation patterns with a fundamental mathematical model. *Nature Communications*, 7(10288), 2016.
- [11] R. J. H. Ross, R. E. Baker, M. Ford, R. L. Mort, and C. A. Yates. Using approximate Bayesian computation to quantify cell-cell adhesion parameters in a cell migratory process. *npj Systems Biology and Applications (under review)*, doi: <http://dx.doi.org/10.1101/068791>, 2016.
- [12] R. J. H. Ross, C. A. Yates, and R. E. Baker. Inference of cell–cell interactions from population density characteristics and cell trajectories on static and growing domains. *Mathematical Biosciences*, 264:108–118, 2015.
- [13] R. J. H. Ross, C. A. Yates, and R. E. Baker. The effect of domain growth on spatial correlations. *Physica A*, 466:334–345, 2016.

- [14] R. J. H. Ross, R. E. Baker, and C.A. Yates. How domain growth is implemented determines the long term behaviour of a cell population through its effect on spatial correlations. *Physical Review E*, 94(1):012408, 2016.
- [15] R. J. H. Ross, C. A. Yates, and R. E. Baker. Pattern formation and variable species densities are induced by domain growth. *Physical Review E (under review)*, doi: <http://dx.doi.org/10.1101/061341>, 2016.
- [16] S. Azaele, A. Maritan, S. J. Cornell, S. Suweis, J. R. Banavar, D. Gabriel, and W. E. Kunin. Towards a unified descriptive theory for spatial ecology: predicting biodiversity patterns across spatial scales. *Methods in Ecology and Evolution*, 6(3):324–332, 2015.
- [17] R. E. Baker, C. A. Yates, and R. Erban. From microscopic to macroscopic descriptions of cell migration on growing domains. *Bulletin of Mathematical Biology*, 72(3):719–762, 2010.
- [18] R. E. Baker and M. J. Simpson. Correcting mean-field approximations for birth-death-movement processes. *Physical Review E*, 82(4):041905, 2010.
- [19] S. T. Johnston, M. J. Simpson, and R. E. Baker. Mean-field descriptions of collective migration with strong adhesion. *Physical Review E*, 85(5):051922, 2012.
- [20] S. T. Johnston, M. J. Simpson, and R. E. Baker. Modelling the movement of interacting cell populations: A moment dynamics approach. *Journal of Theoretical Biology*, 370:81–92, 2015.
- [21] D. C. Markham, M. J. Simpson, P. K. Maini, E. A. Gaffney, and R. E. Baker. Incorporating spatial correlations into multispecies mean-field models. *Physical Review E*, 88(5):052713, 2013.

- [22] D. C. Markham, M. J. Simpson, and R. E. Baker. Simplified method for including spatial correlations in mean-field approximations. *Physical Review E*, 87(6):062702, 2013.
- [23] M. J. Simpson and R. E. Baker. Corrected mean-field models for spatially dependent advection-diffusion-reaction phenomena. *Physical Review E*, 83(5):051922, 2011.
- [24] B. J. Binder, K. A. Landman, M. J. Simpson, M. Mariani, and D. F. Newgreen. Modeling proliferative tissue growth: A general approach and an avian case study. *Physical Review E*, 78(3):031912, 2008.
- [25] B. J. Binder and K. A. Landman. Exclusion processes on a growing domain. *Journal of Theoretical Biology*, 259(3):541–551, 2009.
- [26] B. J. Binder, K. A. Landman, D. F. Newgreen, J. E. Simkin, Y. Takahashi, and D. Zhang. Spatial analysis of multi-species exclusion processes: application to neural crest cell migration in the embryonic gut. *Bulletin of Mathematical Biology*, 74(2):474–90, 2012.
- [27] E. A. Codling, M. J. Plank, and S. Benhamou. Random walk models in biology. *Journal of the Royal Society Interface*, 5(25):813–834, 2008.
- [28] K. J. Davies, J. E. F. Green, N.G. Bean, B. J. Binder, and J. V. Ross. On the derivation of approximations to cellular automata models and the assumptions of independence. *Mathematical Biosciences*, 253:63–71, 2014.
- [29] R. Erban and S. J. Chapman. Reactive boundary conditions for stochastic simulations of reaction-diffusion processes. *Physical Biology*, 4:16–28, 2007.
- [30] R. Erban and S. J. Chapman. Stochastic modelling of reaction-diffusion processes: algorithms for bimolecular reactions. *Physical Biology*, 6:046001, 2009.

- [31] J. D. Hywood, E. J. Hackett-Jones, and K. A. Landman. Modeling biological tissue growth: Discrete to continuum representations. *Physical Review E*, 88(3):032704, 2013.
- [32] G. W. Jones and S. J. Chapman. Modeling growth in biological materials. *SIAM Review*, 54(1):52–118, 2012.
- [33] E. Khain, L. M. Sander, and C. M. Schneider-Mizell. The role of cell-cell adhesion in wound healing. *Journal of Statistical Physics*, 128(1-2):209–218, 2007.
- [34] K. A. Landman and A. E. Fernando. Myopic random walkers and exclusion processes: Single and multispecies. *Physica A*, 390(21-22):3742–3753, 2011.
- [35] A. E. Fernando, K. A. Landman, and M. J. Simpson. Nonlinear diffusion and exclusion processes with contact interactions. *Physical Review E*, 81(1):011903, 2010.
- [36] F. Luciani, D. Champeval, A. Herbette, L. Denat, B. Aylaj, S. Martinozzi, R. Ballotti, R. Kemler, C. R. Goding, F. De Vuyst, L. Larue, and V. Delmas. Biological and mathematical modeling of melanocyte development. *Development*, 138(18):3943–3954, 2011.
- [37] M. J. Simpson, A. Merrifield, K. A. Landman, and B. D. Hughes. Simulating invasion with cellular automata: Connecting cell-scale and population-scale properties. *Physical Review E*, 76:021918, 2007.
- [38] M. J. Simpson, K. A. Landman, and B. D. Hughes. Pathlines in exclusion processes. *Physical Review E*, 79(3):031920, 2009.
- [39] M. J. Simpson, K. A. Landman, and B. D. Hughes. Multi-species simple exclusion processes. *Physica A*, 388(4):399–406, 2009.

- [40] C. A. Yates, R. E. Baker, and P. K. Maini. Going from microscopic to macroscopic on nonuniform growing domains. *Physical Review E*, 86:021921, 2012.
- [41] C. A. Yates. Discrete and continuous models for tissue growth and shrinkage. *Journal of Theoretical Biology*, 350:37–48, 2014.
- [42] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 17(1):10–18, 1976.
- [43] D. Chowdhury, A. Schadschneider, and K. Nishinari. Physics of transport and traffic phenomena in biology: from molecular motors and cells to organisms. *Physics of Life Reviews*, 2(4):318–352, 2005.
- [44] D. J. G. Agnew, J. E. F. Green, T. M. Brown, M. J. Simpson, and B. J. Binder. Distinguishing between mechanisms of cell aggregation using pair-correlation functions. *Journal of Theoretical Biology*, 352:16–23, 2014.
- [45] B. J. Binder and M. J. Simpson. Quantifying spatial structure in experimental observations and agent-based simulations using pair-correlation functions. *Physical Review E*, 88(2):022705, 2013.
- [46] T. M. Liggett. *Stochastic Interacting Systems: Contact, Voter, and Exclusion Processes*. Springer-Verlag, Berlin, 1999.
- [47] S. T. Johnston, M. J. Simpson, and M. J. Plank. Lattice-free descriptions of collective motion with crowding and adhesion. *Physical Review E*, 88:062720, 2013.
- [48] L. Dyson, P. K. Maini, and R. E. Baker. Macroscopic limits of individual-based models for motile cell populations with volume exclusion. *Physical Review E*, 86:031903, 2012.

- [49] L. Dyson and R. E. Baker. The importance of volume exclusion in modelling cellular migration. *Journal of Mathematical Biology*, 71(3):691–711, 2014.
- [50] D. J. Murrell. Local spatial structure and predator-prey dynamics: counter-intuitive effects of prey enrichment. *The American Naturalist*, 166(3):354–367, 2005.
- [51] D. J. Murrell. When does local spatial structure hinder competitive coexistence and reverse competitive hierarchies? *Ecology*, 91(6):1605–1616, 2010.
- [52] A. Santiago and C. A. Erickson. Ephrin-B ligands play a dual role in the control of neural crest cell migration. *Development*, 129(15):3621–3632, 2002.
- [53] S. Nakagawa and M. Takeichi. Neural crest cell-cell adhesion controlled by sequential and subpopulation-specific expression of novel cadherins. *Development*, 121(5):1321–1332, 1995.
- [54] Y. Takahashi, D. Sipp, and H. Enomoto. Tissue interactions in neural crest cell development and disease. *Science*, 341(6148):860–863, 2013.
- [55] A. Thomas and C. Erickson. The making of a melanocyte: the specification of melanoblasts from the neural crest. *Pigment Cell Melanoma Research*, 21(6):598–610, 2008.
- [56] R. Spritz. Piebaldism, waardenburg syndrome, and related disorders of melanocyte development. *Seminars in Cutaneous Medicine and Surgery*, 16(1):15–23, 1997.
- [57] G. Gokalp and B. Hakyemez. Myxoid neurofibromas of the breast: mammographical, sonographical and mri appearances. *The British Journal of Radiology*, 80, 2007.

- [58] J. Khalil, M. Afif, H. Elkacemi, M. Benoulaid, T. Kebdani, and N. Benjaafar. Breast cancer associated with neurofibromatosis type 1: a case series and review of the literature. *Journal of Medical Case Reports*, 9(1):1–4, 2015.
- [59] A. Li, Y. Ma, X. Yu, R. L. Mort, C. R. Lindsay, D. Stevenson, D. Strathdee, R. H. Insall, J. Chernoff, S. B. Snapper, I. J. Jackson, L. Larue, O. J. Sansom, and L. M. Machesky. Rac1 drives melanoblast organization during mouse development by orchestrating pseudopod-driven motility and cell-cycle progression. *Developmental Cell*, 21(4):722–734, 2011.
- [60] D. L. Silver, K. E. Leeds, H. Hwang, E. E. Miller, and W. J. Pavan. The EJC component Magoh regulates proliferation and expansion of neural crest-derived melanocytes. *Developmental Biology*, 375(2):172–181, 2013.
- [61] H. M. Young, C. J. Hearn, P. G. Farlie, A. J. Canty, P. Q. Thomas, and D. F. Newgreen. *GDNF* is a chemoattractant for enteric neural cells. *Developmental Biology*, 229(2):503–516, 2001.
- [62] B. Mintz. Gene control of mammalian pigmentary differentiation. I. Clonal origin of melanocytes. *Proceedings of the National Academy of Sciences*, 58(1):344–351, 1967.
- [63] B. D. Hughes. *Random Walks and Random Environments*, volume 1. Oxford: Clarendon Press, Oxford, UK, 1995.
- [64] J. Livet, T. A. Weissman, H. Kang, R. W. Draft, J. Lu, R. A. Bennis, J. R. Sanes, and J. W. Lichtman. Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature*, 450(7166):56–62, 2007.
- [65] K. J. Cheung and A. J. Ewald. Illuminating breast cancer invasion: diverse roles for cell–cell interactions. *Current Opinion in Cell Biology*, 30:99–111, 2014.

- [66] A. Beaussart, S. El-Kirat-Chatel, R. M. A. Sullan, D. Alsteens, P. Herman, S. Derclaye, and Y. F. Dufrêne. Quantifying the forces guiding microbial cell adhesion using single-cell force spectroscopy. *Nature Protocols*, 9(5):1049–1055, 2014.
- [67] Y. Chen, S. J. Dodd, M. A. Tangrea, M. R. Emmert-Buck, and A. P. Koretsky. Measuring collective cell movement and extracellular matrix interactions using magnetic resonance imaging. *Scientific Reports*, 3(1879), 2013.
- [68] Y. F. Dufrêne and A. E. Pelling. Force nanoscopy of cell mechanics and cell adhesion. *Nanoscale*, 5(10):4094–4104, 2013.
- [69] J. J. Fredberg. Power steering, power brakes, and jamming: Evolution of collective cell-cell interactions. *Physiology*, 29(4):218–219, 2014.
- [70] S. Sivasankar, B. Gumbiner, and D. Leckband. Direct measurements of multiple adhesive alignments and unbinding trajectories between cadherin extracellular domains. *Biophysical Journal*, 80(4):1758–1768, 2001.
- [71] S. Rakshit and S. Sivasankar. Biomechanics of cell adhesion: how force regulates the lifetime of adhesive bonds at the single molecule level. *Physical Chemistry Chemical Physics*, 16(6):2211–2223, 2014.
- [72] M. Ward, C. McCann, M. DeWulf, J. Y. Wu, and Y. Rao. Distinguishing between directional guidance and motility regulation in neuronal migration. *Journal of Neuroscience*, 23(12):5170–5177, 2003.
- [73] A. Sato, A. M. Scholl, E. B. Kuhn, H. A. Stadt, J. R. Decker, K. Pegram, M. R. Hutson, and M. L. Kirby. FGF8 signaling is chemotactic for cardiac neural crest cells. *Developmental Biology*, 354(1):18–30, 2011.

- [74] M. B. Byrne, Y. Kimura, A. Kapoor, Y. He, K. S. Mattam, K. M. Hasan, L. N. Olson, F. Wang, P. J. A. Kenis, and C. V. Rao. Oscillatory behavior of neutrophils under opposing chemoattractant gradients supports a winner-take-all mechanism. *PLoS ONE*, 9(1):e85726, 2014.
- [75] B. E. Richardson and R. Lehmann. Mechanisms guiding primordial germ cell migration: strategies from different organisms. *Nature Reviews Molecular Cell Biology*, 11(1):37–49, 2010.
- [76] T. E. Woolley, R. E. Baker, C. Tickle, P. K. Maini, and M. Towers. Mathematical modelling of digit specification by a sonic hedgehog gradient. *Developmental Dynamics*, 243(2):290–298, 2014.
- [77] M. J. Simpson, K. A. Landman, and B. D. Hughes. Diffusing populations: ghosts or folks. *Australasian Journal of Engineering Education*, 15(2):59–68, 2009.
- [78] D. Hiebeler. Stochastic spatial models: From simulations to mean-field and local structure approximations. *Journal of Theoretical Biology*, 187(3):307–319, 1997.
- [79] G. Schütz and E. Domany. Phase transitions in an exactly soluble one-dimensional exclusion process. *Journal of Statistical Physics*, 72(1-2):277–296, 1993.
- [80] E. J. Crampin, E. A. Gaffney, and P. K. Maini. Reaction and diffusion on growing domains: Scenarios for robust pattern formation. *Bulletin of Mathematical Biology*, 61(6):1093–1120, 1999.
- [81] L. G. Bowden, M. J. Simpson, and R. E. Baker. Design and interpretation of cell trajectory assays. *Journal of the Royal Society Interface*, 10(88):1742–1755, 2013.

- [82] M. J. Simpson and R. E. Baker. Special issue on spatial moment techniques for modelling biological processes. *Bulletin of Mathematical Biology*, 77(4):581–585, 2015.
- [83] W. R. Young, A. J. Roberts, and G. Stuhne. Reproductive pair correlations and the clustering of organisms. *Nature*, 412(6844):328–331, 2001.
- [84] M. A. M. de Aguiar, E. M. Rauch, and Y. Bar-Yam. Invasion and extinction in the mean-field approximation for a spatial host-pathogen model. *Journal of Statistical Physics*, 114(5–6):1417–1451, 2004.
- [85] H. Fuk s and A. T. Lawniczak. Individual-based lattice model for spatial spread of epidemics. *Discrete Dynamics in Nature and Society*, 6(3):191–200, 2001.
- [86] J. Mai, V. N. Kuzovkov, and W. von. Niessen. A theoretical stochastic model for the $A + \frac{1}{2}B \rightarrow 0$ reaction. *Journal of Chemical Physics*, 98(12):10017, 1993.
- [87] J. Mai, V. N. Kuzovkov, and W. von. Niessen. A general stochastic model for the description of surface reaction systems. *Physica A*, 203:298–315, 1994.
- [88] B. J. Binder and M. J. Simpson. Spectral analysis of pair-correlation bandwidth: application to cell biology images. *Royal Society Open Science*, 2(2):140494, 2015.
- [89] B. Waclaw, I. Bozic, M. E. Pittman, R. H. Hruban, B. Vogelstein, and M. A. Nowak. A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. *Nature*, 525(7568):261–264, 2015.
- [90] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 81(25):2340–2361, 1977.
- [91] A. Singer. Maximum entropy formulation of the Kirkwood superposition approximation. *Journal of Chemical Physics*, 121(8):3657–3666, 2004.

- [92] B. J. Thomas, D. A. Gunning, J. Cho, and S. L. Zipursky. Cell cycle progression in the developing *Drosophila* eye: *roughex* encodes a novel protein required for the establishment of G1. *Cell*, 77(7):1003–1014, 1994.
- [93] J. P. Kumar. My what big eyes you have: How the *Drosophila* retina grows. *Developmental Neurobiology*, 71(12):1133–1152, 2011.
- [94] C. J. Barelle, E. A. Bohula, S. J. Kron, D. Wessels, D. R. Soll, A. Schäfer, A. J. P. Brown, and N. A. R. Gow. Asynchronous cell cycle and asymmetric vacuolar inheritance in true hyphae of *Candida albicans*. *Eukaryotic cell*, 2(3):398–410, 2003.
- [95] P. J. Keller. Imaging morphogenesis: Technological advances and biological insights. *Science*, 340(6137):1234168, 2013.
- [96] R. Kennaway, E. Coen, A. Green, and A. Bangham. Generation of diverse biological forms through combinatorial interactions between tissue polarity and growth. *PLoS Computational Biology*, 7(6):e1002071, 2011.
- [97] P. Fried and D. Iber. Dynamic scaling of morphogen gradients on growing domains. *Nature Communications*, 5(5077), 2014.
- [98] L. Wolpert. *Principles of Development*. Oxford University Press, Oxford, 4th edition, 2011.
- [99] R. Satija and A. K. Shalek. Heterogeneity in immune responses: from populations to single cells. *Trends in Immunology*, 35(5):219–229, 2014.
- [100] M. F. Bear, B. W. Connors, and M. A. Paradiso. *Neuroscience: exploring the brain*. Lippincott Williams and Wilkins, Philadelphia, 3rd edition, 2007.
- [101] G. Schram, M. Pourrier, P. Melnyk, and S. Nattel. Differential distribution of

- cardiac ion channel expression as a basis for regional specialization in electrical function. *Circulation Research*, 90(9):939–950, 2002.
- [102] A. Marusyk, V. Almendro, and K. Polyak. Intra-tumour heterogeneity: a looking glass for cancer? *Nature Reviews Cancer*, 12(5):323–334, 2012.
- [103] S. Perilli, R. Di Mambro, and S. Sabatini. Growth and development of the root apical meristem. *Current Opinion in Plant Biology*, 15(1):17–23, 2012.
- [104] J. F. Fallon, A. Lopez, M. A. Ros, M. P. Savage, B. B. Olwin, and B. K. Simandl. FGF-2: apical ectodermal ridge growth signal for chick limb development. *Science*, 264(5155):104–107, 1994.
- [105] R. Law, D. J. Murrell, and U. Dieckmann. Population growth in space and time: spatial logistic equations. *Ecology*, 84(1):252–262, 2003.
- [106] D. K. Schlüter, I. Ramis-Conde, and M. A. J. Chaplain. Computational modeling of single-cell migration: the leading role of extracellular matrix fibers. *Biophysical Journal*, 103(6):1141–1151, 2012.
- [107] P. Marjoram, J. Molitor, V. Plagnol, and S. Tavaré. Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 100(26):15324–15328, 2003.
- [108] S. T. Johnston, M. J. Simpson, and D. L. S. McElwain. 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, 2014.
- [109] E. van der Vaart, M. A. Beaumont, A. S. A. Johnston, and R. M. Sibly. Calibration and evaluation of individual-based models using Approximate Bayesian Computation. *Ecological Modelling*, 312:182–190, 2015.

- [110] M. A. Beaumont, W. Zhang, and D. J. Balding. Approximate Bayesian computation in population genetics. *Genetics*, 162(4):2025–2035, 2002.
- [111] S. T. Johnston, M. J. Simpson, D. L. S. McElwain, B. J. Binder, and J. V. Ross. Interpreting scratch assays using pair density dynamics and approximate Bayesian computation. *Open Biology*, 4(9):140097, 2014.
- [112] B. N. Vo, C. C. Drovandi, A. N. Pettitt, and G. J. Pettet. Melanoma cell colony expansion parameters revealed by approximate Bayesian computation. *PLoS Computational Biology*, 11(12):e1004635, 2015.
- [113] B. N. Vo, C. C. Drovandi, A. N. Pettitt, and M. J. Simpson. Quantifying uncertainty in parameter estimates for stochastic models of collective cell spreading using approximate Bayesian computation. *Mathematical Biosciences*, 263:133–142, 2015.
- [114] P. J. M. Jones, A. Sim, H. B. Taylor, L. Bugeon, M. J. Dallman, B. Pereira, M. P. H. Stumpf, and J. Liepe. Inference of random walk models to describe leukocyte migration. *Physical Biology*, 12(6):066001, 2015.
- [115] H. Kitano. Biological robustness. *Nature Reviews Genetics*, 5(11):826–837, 2004.
- [116] M. J. Simpson, K. K. Treloar, B. J. Binder, P. Haridas, K. J. Manton, D. I. Leavesley, D. L. S. McElwain, and R. E. Baker. Quantifying the roles of cell motility and cell proliferation in a circular barrier assay. *Journal of The Royal Society Interface*, 10(82):20130007, 2013.
- [117] K. P. Burnham and D. R. Anderson. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Berlin, Germany: Springer, 2002.

- [118] M. A. Nunes and D. J. Balding. On optimal selection of summary statistics for approximate Bayesian computation. *Statistical Applications in Genetics and Molecular Biology*, 9(1), 2010.
- [119] E. Fermi. On the origin of the cosmic radiation. *Physical Review Letters*, 75: 1169–1174, 1949.
- [120] S. W. Doescher and M. H. Rice. Infinite square-well potential with a moving wall. *American Journal of Physics*, 37:1246–1249, 1969.
- [121] S. Schmidt, J. G. Muga, and A. Ruschhaupt. Stopping particles of arbitrary velocities with an accelerated wall. *Physical Review A*, 80:023406, 2009.
- [122] S. V. Mousavi. Quantum effective force in an expanding infinite square-well potential and bohmian perspective. *Physica Scripta*, 86(3):035004, 2012.
- [123] F. Reif. *Fundamentals of Statistical and Thermal Physics*. Waveland Press, 2009.
- [124] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona. Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9(7):676–82, 2012.
- [125] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri. NIH image to ImageJ: 25 years of image analysis. *Nature Methods*, 9(7):671–5, 2012.