Spatial Quantification and Mathematical Modelling of Tissue Development

Saber Dini

January 30, 2018

Thesis submitted for the degree of
Doctor of Philosophy
in
Applied Mathematics
at The University of Adelaide
Faculty of Engineering, Computer and Mathematical Sciences
School of Mathematical Sciences

THE UNIVERSITY
of ADELAIDE

Contents

Signed Statement

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Signed: ..... Date: 7/11/2017

Acknowledgements

Moving to Australia and undertaking a PhD at The University of Adelaide created an outstanding chapter of my life. I have gained so many invaluable experiences during these years which have positively impacted my life. I would like to take this opportunity to thank those who accompanied me throughout this journey, and helped me complete this thesis.

First of all, I am appreciative to my dear wife, Safa, and my family who encouraged me to continue my education, and constantly supported me throughout my PhD.

I would like to express my gratitude to my supervisors, Dr. Edward Green and A/Prof. Benjamin Binder, for their continuous encouragement and support, without which I could not complete this thesis. Having such kind and considerate supervisors was a great privilege to me.

My sincere thanks goes to Dr. Edward Green for offering me a scholarship that enabled me to undertake my PhD at The University of Adelaide.

I am also grateful to my closest friend, Maha Mansor, whose company during my PhD was an absolute pleasure.

I would like to thank Prof. Nigel Bean for his valuable advice during my PhD. He also made a significant contribution to the spatial quantification parts of my thesis.

I acknowledge that I was covered by an ARC Grant-Funded PhD Scholarship and a University of Adelaide Full Fee Scholarship for three years. The School of Mathematical Sciences also provided me with a scholarship for two months. I am grateful to A/Prof. Gary Glonek, our head of school, for granting this scholarship to me.

Dedication

To my beautiful wife, Safa, who has warmed my heart throughout this long journey.

Abstract

In this thesis, we study biological tissue development, during which cells organise themselves into structures which perform a specific function. Understanding how particular types of mechanisms lead to the emergence of various cell patterns in tissues is the main motivation of this research. Quantifying the tissue patterns is a first step towards understanding which mechanisms are at work in particular experiments. For this purpose, we develop pair-correlation functions (PCFs) which quantify how a spatial distribution of cells deviates from complete spatial randomness over specified directions. We evaluate the usefulness of PCFs for studying the three-dimensional organisation of cells in tumour spheroids and show that the PCFs robustly reveal information about their spatial structure. In particular, we demonstrate that the boundary that separates the necrotic and viable zones in the tumour spheroids can be detected using the PCF with a high degree of accuracy.

We then turn to development of mathematical models to investigate the types of patterns that can arise from simple hypothesised interactions between cells. We begin in Chapter 3 by developing an on-lattice agent-based model (ABM) to investigate tumour spheroid growth using two different culture methods: suspension culture, and culture within a microgel. Our results suggest that stratifying the seeded cells into multiple layers and also reducing cell death are the key effects of the microgel that enable it to produce more uniformly-sized spheroids. In Chapter 4, we extend the ABM to study systems with two interacting species. A huge variety of aggregation patterns can arise in these systems, depending upon the underlying attractive-repulsive mechanisms. More specifically, we show that the run-and chase mechanism can produce a striped pattern, similar to that observed on the skin of zebrafish.

Finally, we develop a non-local continuous model, approximating the mean behaviour of the ABM. This provides a connection between the cell-level and population-level models of tissue development. A linear stability analysis of the continuous model allows us to investigate parameter regimes

that produce striped patterns. Importantly, we also point out the disparities that may arise between the behaviours of the continuous and discrete models, which highlights the importance of considering the underlying biological constraints in using the continuous approximated models. In particular, we show that the derivation of the approximate continuum model from the ABM introduces terms representing cell-size effects. These terms can lead to the emergence of stripes in cases where they would not be predicted in the similar continuum model of Painter *et al.* (2015), which does not include these terms.

The combination of spatial quantification and mathematical modelling (using both continuous and discrete methods) developed in this work helps us to gain a better understanding of tissue development. Our approach provides a novel means to investigate the underpinning mechanisms of tissue development by combining model simulations with analysis of biological and synthetic data using the pair-correlation functions.

Chapter 1

Introduction

1.1 Tissue development

Biological tissue as Dorland puts it is "an aggregation of similarly specialised cells united in the performance of a particular function" [40]. Tissues form every part of our body, and can be categorised into five groups: epithelial (*e.g.* skin), connective (*e.g.* bone), muscular, nervous, 'wandering corpuscles of the blood', and lymph [102]. The study of tissues first received serious attention around 1800, when the French anatomist Marie Franois Xavier Bichat (1771 - 1802) identified 21 types of tissues commonly found in the human body [135]. Tissues were considered as the basic building blocks of the organs, until 1830, when their constituent cells were observed using microscopes [179]. Since then, tremendous progress has been made in understanding tissues, especially since the 1990s.

Development and repair of these various tissues are vital and complex biological processes involving precisely controlled sequences of cell movement and proliferation. Such events occur, for example, during morphogenesis (a biological process during which the shape of an organism is formed) [69], scar formation and wound healing [104]. When they go awry, pathological states can occur, *e.g.* tumour growth.

Tissue development is a complex and multi-faceted process. In a developing tissue, different cell populations interact with each other, the extracellular matrix (ECM) within which they reside, and chemical signals, in intricate ways. Our aims in this thesis are i) quantifying the types of tissue patterns observed in experiments, and ii) understanding how particular types of cell interaction create various patterns and forms of tissues. There are two main types of signals that produce the final pattern of a tissue: chemical and mechanical. Chemical molecules, secreted from cells, can stimulate them to

migrate, proliferate, or differentiate. For example, in liver tissue development, hepatocyte nuclear factor 6 mediates the migration of hepatoblasts, precursor cells of liver. Other chemicals, such as fibroblast growth factor, are necessary for differentiation of hepatoblasts into hepatocytes or cholangiocytes [152]. Thus, chemicals play a crucial role in the tissue development process by regulating different cellular behaviours.

Cell-cell and cell-ECM mechanical forces also affect cellular migration, proliferation and differentiation. Cells can generate large traction forces on other cells or on their substratum, the ECM, when they move [120]. The mechanical forces exerted by cells lead to the creation of ECM deformations that affect movement and proliferation of cells [127]. In addition, mechanical tension affects cellular differentiation, for example it can lead to myofibroblast differentiation from its precursor cells [180]. Adhesion between cells or between cells and ECM also affects movement of cells and consequently the final tissue pattern [129]. Mechanical interactions play important roles in many embryogenesis processes, such as limb morphogenesis, where they create the pattern of cell condensations that finally turn into cartilage in a developing limb. A limb pattern is created by aggregation of chondrocyte cells, which occurs when the traction force of chondrocytes overcome the elastic resistance in the ECM in a developing limb bud [120]. Therefore, these interactions have a variety of significant impacts on tissue development.

Technological advancements since the 1980s have given rise to a new field: tissue engineering [94], which aims to grow functioning tissues and organs *in vitro* that are meant to work as 'spare parts' [5, 9, 173]. Achieving this ambitious goal would put an end to the difficulties faced by the patients seeking organ transplant, which is hindered by the lack of suitable donor organs [123], rejection by the immune system [149], *etc.* Thorough understanding of the mechanisms underpinning tissue development is essential to make progress in this area. Although we are still far from being able to grow fully functioning replacement organs, being able to grow tissues *in vitro* for research purposes is extremely useful for gaining better understanding of tissue development processes.

One particular area which has received immense attention is the growth of tumour tissue [74, 116, 161]. These *in vitro* cultured tumours (tumour spheroids) are used to study how tissue development goes awry in tumorigenesis (Fig. 1.1a shows an image of a tumour spheroid [30]). They are also used to mimic real tumours for drug screening experiments [74, 98]. The proliferative ability of tumour-derived cells, and the fact that tumour spheroids do not have a complex structure of many different cell types, as is frequently the case in normal tissues, make them well-suited to in vitro culture.

More recently, the development of more sophisticated culture techniques

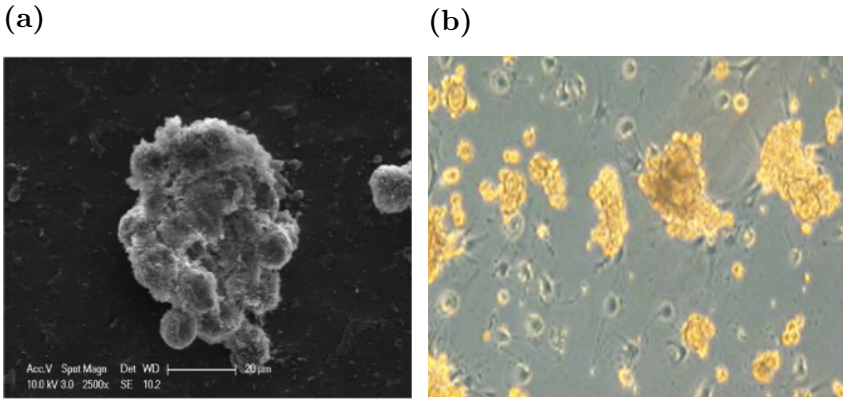


Figure 1.1: Cell aggregation examples. (a) Tumour spheroid (comprised of Hela cells) after 7 days of seeding [30]. (b) Hepatocyte-stellate cells co-cultured for producing spheroids that are fundamental parts of liver (reprinted from Thomas *et al.* (2005), with permission from Eur. Cells Mater).

has led to a new type of *in vitro* tissue culture, known as organoids, that can closely recreate *in vivo* tissues [148]. They are produced by extracting cells from a specific tissue of a person (or animal), and culturing them in an appropriate medium to obtain a miniature version of an organ [93]. For example, organoids derived from intestinal tissues have been frequently used as mini-guts for studying colorectal cancer [146, 147]. Self-organisation of cells into *in vivo*-like structures during organoid formation has drawn immense attention to provide insight into this complex process. Of particular interest is to produce a reliable environment in which organoids are produced with the desired structure. This requires clear understanding of how different cell types might interact with each other and produce organoid forms [6, 20].

Cellular aggregation is a major focus of this study, as it plays a pivotal role in tissue development processes, especially in tissue engineering applications. Culturing cells as aggregates has a number of advantages over traditional mono-layer cultures, including higher proliferation, lower cell death, *etc.* [2]. Huge efforts have been put into producing *in vitro* culture media that promotes sustainable cell aggregations [62, 73, 164]. For example, Thomas *et al.* (2005) demonstrate that co-culture of stellate and hepatocyte cells (fundamental cells of liver tissue) enhances aggregate (spheroid) formation, which in turn prevents cell death and improves tissue function, as opposed to the mono-culture of hepatocytes. Fig. 1.1b depicts an image of a hepatocyte-stellate co-culture, containing several aggregates.

In this thesis, we aim to develop statistical tools to quantify the spatial patterns produced by the arrangement of cells during tissue development. This allows us to describe precisely how cells are arranged in different tissues. We then turn to the development of mathematical models that explore different types of cell behaviour to further our understanding of the mechanisms that underpin the tissue development process. By coupling our statistical tools and mathematical models, we are able to gain insight into the characteristic tissue patterns that arise from particular combinations of cell interactions.

In the following, we briefly review the spatial quantification techniques and mathematical modelling methods that we will use and build upon in this thesis.

1.2 Spatial quantification

We can recognise some features in the spatial arrangement of cells in tissues by simple visual inspection of experimental images. For example, we might observe clustering or segregation in a distribution of cells, simply by looking at microscopic images. However, our ability to discern patterns may be impeded by the intrinsic randomness underlying biological processes. The stochasticity in each experiment presents a challenge in interpreting the patterns in the experimental images. Furthermore, the subjectivity of these interpretations casts doubt on their accuracy and robustness. In order to overcome these issues, we require statistical tools to draw out the main features from each spatial pattern, and, in turn allow precise comparisons to be made amongst a whole ensemble of experiments. For the same reasons, the patterns produced from realisations of stochastic mathematical models require statistical tools to be interpreted.

In this work, we consider the random distribution of the position of cells in a tissue as a spatial point process. A spatial point process is a stochastic process in which objects such as stars, trees or cells are distributed randomly in a domain [77]. Different spatial patterns may be associated with particular underlying mechanisms of interaction between the objects. For example, when the points are distributed uniformly at random, it may indicate that the objects are non-interacting. Similarly, patterns showing clustering or space-filling can indicate attraction or repulsion between the objects, respectively. Therefore, quantifying the patterns can provide information about the underlying mechanisms. Moreover, it allows precise comparison between patterns, which may help us to determine whether they are from similar mechanisms or not.

Testing for the complete spatial randomness (CSR) of point processes is important, because it investigates the first basic assumption about the distributed points: do they influence each other's positions? [37]. Consequently, deviations from CSR suggest that further investigation of the mechanisms underlying the development of the observed pattern are needed, which can lead to important new insights. Statistical tools can be used to test the CSR on samples of a point process (*i.e.* point patterns) and characterise spatial features (*e.g.* clustering, segregation) by indicating the extent of deviations from CSR.

A huge variety of statistical tools have been developed for analysing point processes and, in particular, testing CSR. We mention only a few of them here- more can be found in [27, 37, 77] and others. The index-of-dispersion test introduced by Sachs (1984) tests for CSR by discretising the space into a number of quadrats and evaluating the mean and variance of the number of points per quadrat [145]. In this method, an index close to unity for a point pattern is associated with CSR, and indices higher or lower than unity indicate clustering or segregation, respectively. The Greig-Smith method is similar, but it also involves counting the number of points in the neighbouring quadrats [66]. The density of points is a key feature in these types of methods. However, the distance between points could also be utilised. Statistical tools that deal with distance are usually called second-order characteristics [77]. Ripley's K-function can indicate deviations from homogeneity by comparing the K-function for a point pattern with that at the CSR [142]. It is computed from the expected number of points within a radius of a typical point. The L-function is a variant of this method, and offers an easier interpretation [8].

Pair correlation functions (PCFs) are popular statistical tools for analysing point patterns and testing for CSR as they encapsulate considerable information about the point pattern and are easy to interpret. A PCF is a second-order summary characteristic that describes the frequency of distances between pairs of points [77]. PCFs can be found in different contexts under a variety of names such as two-point correlation function [85, 105], pair correlation density [36], or radial distribution function [25].

Various methods of estimating PCFs for point processes in different geometries have been considered [124, 142, 159, 160]. PCFs have been used in analysing various kinds of point processes in different fields such as astronomy [85, 105, 134] and forestry [17, 66, 92, 136]. They are also widely used in the field of cellular biology. Riedel *et al.* (2005) use the PCF and triplet distribution functions in analysing the spacing between the hexagonal vortices created by sperm cells. They showed that self-organisation of sperms into hexagonal order can be mediated hydrodynamically, without a need for chemical signals [141]. Mattfeldt *et al.* (1993) uses the PCF to compare the

distribution of normal prostatic epithelial tissue components with cancerous ones in carcinomatous tissue. They found a partial loss in the epithelial interaction in the carcinomatous tissue [106]. Fozard *et al.* (2011) applied the PCF and quadrat histograms to snapshots of individual-based model simulations, and demonstrated that these statistical measures can be used to characterise patterns that emerge in the distribution of differentiating cells cultured *in vitro* from stem cells [52]. Other applications of PCFs in biology can be found in [1, 14, 169].

The PCFs can be used for estimating parameters in mathematical models of tissue development. In particular, they have been shown to be suitable summary statistics in Approximate Bayesian Computation (ABC), a parameter estimation method [83, 144, 166], as they reveal rich information about the spatial characteristics of patterns. The summary statistics are used in this method to sift out the model simulations that best match experimental observations. For instance, Johnston *et al.* (2014) uses a PCF as a summary statistic for use in their ABC method to provide estimates for proliferation and diffusion rates of cells spreading in scratch assay experiments*.

In this thesis, we develop new PCFs to quantify cell aggregation patterns. Specifically, the PCFs are applied to three-dimensional point patterns of cancer cells distributed in tumour spheroids. Later on, similar PCFs for lattice-based patterns are proposed to study the influences of different mechanisms on the formation of various tissue patterns.

1.3 Mathematical modelling

Mathematical models provide a useful framework to explain the complex effects of different interactions on the final pattern of a tissue. They can also provide a means to validate or disprove experimental hypotheses. If a model is developed based on an experimental hypothesis, and the results of the model do not demonstrate the observed behaviours, it suggests that the hypothesis may need to be reconsidered. Mathematical models of tissue development are helpful not only in the field of embryogenesis [127], but also in tissue engineering [96], wound healing [151], cancer [56] and many other similar areas.

Two main approaches can be taken to model tissue development: macroscopic (also known as population-level) and microscopic (individual-level) approaches. In the following, we briefly introduce each of the approaches.

*Scratch assays are *in vitro* experiments that are designed to study cell migration occurring particularly in wound healing [97].

1.3.1 Continuum models

The macroscopic approach often describes the evolution of a biological system in terms of continuous variables, such as the density of cells or chemical concentrations. These variables evolve according to partial differential equations (PDEs); spatio-temporal differential equations that describe the reactions and interactions in the system, such as cell motion, proliferation, chemical production / degradation, *etc.* They can provide insights into the average behaviours of cells. In some cases, analytical solutions might be obtained for the PDEs, but in general, numerical methods like Finite Difference Method (FDM) or Finite Element Method (FEM) must be used [139]. A major benefit of using PDEs is their amenability to analyses, such as linear stability analysis, weakly non-linear analysis, travelling wave analysis, *etc.* Using these analyses help to bypass the need to solve the PDEs for a wide range of values of parameters in order to investigate the behaviours of a system.

Turing's seminal work on pattern formation [172] has influenced numerous studies in this area of research. Turing proposed a reaction-diffusion PDE model to explain the process of morphogenesis, in a system of interacting chemical substances (morphogens), which are hypothesised to control cell behaviour. He showed how simple physical laws are sufficient to explain pattern formation phenomenon. In the model developed by Turing, cells have a passive role: their behaviour is assumed to be dictated by the chemical concentrations existing in their environment, and, in fact, the cells are not represented in the model equations. However, in later models cells have an active role in creating patterns, by migrating, producing chemicals, exerting forces, *etc.* [120]. Another influential work is the PDE model introduced by Keller and Segel [88], developed initially to describe slime mould aggregates. In their model, single cells are chemotactically attracted to each other, which leads to the formation of multicellular aggregation. They show that pattern formation could occur as a result of an instability in an initially homogeneous system, by means of a PDEs model which generally can be formulated as

$$\begin{aligned} \frac{\partial \rho}{\partial t} &= \underbrace{-\nabla.(\chi(a)\rho\nabla a)}_{\text{chemotaxis}} + \underbrace{D_\rho \nabla^2 \rho}_{\text{random motion}}, \\ \frac{\partial a}{\partial t} &= \underbrace{f(\rho, a)}_{\text{checmial production / degradation}} + \underbrace{D\nabla^2 a}_{\text{diffusion}}, \end{aligned}$$

where ρ is the population density of cells that migrate in response to the concentration of a chemo-attractant / repellent a. The strength of chemotaxis is determined by $\chi(a)$, and f consists of the rates of turnover, secretion,

etc. of a. D_ρ and D are diffusion coefficients of the cells and the chemical, respectively.

PDE models have been extensively used to study aspects of tissue developments [43, 63, 64, 127, 129]. The book by Murray [120] gives a reasonably comprehensive overview of tissue pattern formation. However, in most cases considered in this book, the models contain only one cell type. More recently, the behaviour of different interacting cell populations in tissue development received attention [64, 128, 130]. For example, Green *et al.* (2010) investigate the creation of nascent aggregates in a system where some cells (hepatic stellate cells) can pull or push other cells (hepatocytes) with their long cellular processes. In a similar work, Painter *et al.* (2015) proposed a model that allows the investigation of the interaction of pigment cells (melanophores and xanthophores) that produce the stripy patterns on the skin of zebrafish.

A modelling framework that has recently received significant attention is non-local mathematical modelling [109]. These models have widely been used for modelling biological processes [64, 108, 130], since various types of cell interactions can be readily embedded into them, and, also, they are amenable to analyses such as linear stability analysis, weakly non-linear analysis, *etc.* [23, 49, 76, 95, 167]. These models are often formulated as integro-partial differential equations, Eqn. (1.3.1), where the integral part describes the non-local effect of the neighbouring factors, *e.g.* number of cells, chemical concentrations, *etc.* on the motion of cells. Non-local models typically take the form

$$\frac{\partial \rho}{\partial t} = -\nabla.\left(\nabla K(\boldsymbol{x}) \star f\right) + D_\rho \nabla^2 \rho, \tag{1.3.1}$$

where ρ is cell density and $\star$ denotes the convolution, given by

$$K(\boldsymbol{x}) \star f = \int_{-\infty}^{\infty} K(\boldsymbol{x} - \boldsymbol{x}') f(\boldsymbol{x}') \mathrm{d}\boldsymbol{x}',$$

where $K(\boldsymbol{x})$, commonly called interaction kernel, applies the effect of environmental factors (*e.g.* number of neighbouring cells) on the migration of cells, and D_ρ is the diffusion coefficient.

Nevertheless, a major disadvantage of continuum models is the lack of individual cell-level details. Particularly, estimating the population-level parameters of PDEs is not a straightforward procedure, since most experimental data are provided on the cell-level. Moreover, in certain cases, such as when there are relatively few cells, relating cell densities in a PDE to discrete cells observed in experiments might be problematic. Also, PDE models usually suffer from a lack of stochasticity, predicting the average behaviour of a given system; thus, they cannot incorporate the variabilities originating from the randomness existing in real biological systems.

1.3.2 Discrete models

Discrete mathematical models have increasingly been used to study real-world systems. They are seen to provide a natural framework for modelling situations where that objects of interest are discrete entities, which has led to their being applied in a variety of fields, such as biology, physics, social sciences and economics. Various types of discrete mathematical models can be found in the literature with different names. However, they all have one property in common: discrete objects (cells here) are the main elements of interest in these models. Recent advances in computer technologies have provided incredible power to simulate these models easily on a computer. Discrete models are also known as agent-based models (ABMs), or individual-based models (IBMs), in the literature, as they encompass a wide range of different discrete models, introduced in the following.

Discrete models can be divided into two main categories: lattice-free and lattice-based models, based on the spatial domain on which the cells reside. Variants of lattice-based models include: cellular automata [46] and the cellular Potts model [103]; two well-known types of lattice-free models are particle-based models [16] and vertex-based models [121]. In the lattice-based models, the shape and position of cells are restricted, *i.e.* cells can only occupy lattice sites with predetermined shapes, *e.g.* square, hexagon, *etc.* Conversely, the cells in lattice-free models can take any position (particle-based models) or shape (vertex-based models) on the domain. Producing realisations of these discrete models can be carried out by various methods. For example, for the models that involve stochasticity, Monte Carlo algorithms are often used, whilst for the deterministic models, differential equations describe the evolution of the position of cells, mechanical energy of the system, *etc.* are to be solved.

Particle-based models are one of the lattice-free models frequently used to study biological phenomena [16, 51, 108, 110]. These models formulate the time-evolution of the particles' (cells) positions in a continuous spatial domain, using Newton's second law. Then, usually by neglecting the inertial effects (due to low Reynolds number), the velocity of each particle is found as the superposition of the (often pairwise) forces imposed by the neighbouring cells, which is commonly modelled through 'potential' functions. Therefore, the state of the system over time is essentially determined by solving many ordinary differential equations for the cells positions. Randomness could be included in these models by adding *e.g.* white noise terms to the velocity of the particles [16].

In the ABMs, cells (also chemicals [170], *etc.*) are represented by agents that may proliferate, move, die, *etc.* according to some predefined rates or

rules. As a result, the ABMs can incorporate a wider variety of biological actions and reactions, compared to the particle-based models. In this thesis, we use lattice-based ABMs to model the cell aggregations occurring in tissue development.

Numerous studies can be found that use ABMs for analysing tissue development [12, 155, 1, 174]. For example, Drasdo and Loeffler (2001) propose an ABM for one-layer growing tissues that could represent epithelium. They exclusively study pattern formation in the folding of intestinal crypts and formation of blastula. In another work, Binder *et al.* (2012) discuss how different spatial patterns of green fluorescent protein labelled enteric neural crest cells (ENC) occur among unlabelled ENCs in enteric neural system growth. They suggest that the percentage of initially labelled neural crest cells and the ratio of cell proliferation to cell motility are the key factors that influence the spatial patterns in avian embryos. In a later study, Simpson *et al.* (2013) estimated parameters of a model describing the diffusion and proliferation of cells in a circular barrier assay. Based on the fact that the probability of proliferation is much smaller than the probability of movement of the cells, they used the method of averaging the occupancy of sites in the ABM to derive the cell motility parameter and intrinsic cell proliferation rate. They thus obtain appropriate parameters to investigate the features of moving fronts of cells in tissue development, wound healing and cancer metastasis problems.

Nevertheless, simulation of ABMs can be quite time consuming. For example, the randomness in the activities of agents leads to different outcomes from each simulation. Hence, we need to consider suitably averaged data by running enough simulations to filter out the random effects and extract the important features of a system. The advantage of this property is the similarity in randomness to biological experiments and the fact that we can obtain measures of the variability of certain features of the solution as well as the average behaviour. However, it comes at the cost of running many possibly computationally intensive simulations. In addition, analytical techniques cannot be used to determine how model parameters influence the produced results, as opposed to the continuum models. Hence, to study the behaviour of an ABM, we might need to simulate the model over a wide range of parameter space, which is potentially a computational challenge.

1.4 Thesis overview

We begin our study in Chapter 2 with the development of a new PCF that is used to quantify the spatial distribution of cells in tumour spheroids. Our

new PCF is an appropriate quantifying tool when we are interested in studying spatial patterns over specified directions, *e.g.* horizontal / vertical direction in scratch assays [11], angular direction in yeast colonies [13], *etc.* By applying our method on point patterns of cells from tumour spheroids, we show that the PCF enables us to provide reliable estimates for the necrotic zone boundary of the spheroids, which is of great importance to biologists. The results are then verified against measurements made by a skilled human technician, and compared with some commonly used computational clustering algorithms.

Having examined the static distribution of cells in tumour spheroids, we then turn to study the growth of these *in vitro* tissues. The culture medium in which the tumour spheroids are grown can have a significant impact on their morphology. In Chapter 3, we develop an ABM to provide insight into how two particular culture methods, namely suspension culture and culture within a microgel, affect the size distribution of the spheroids. This is of immense importance in their mass production for drug screening experiments and other applications. After successfully reproducing the results of biological experiments using the ABM, we assess the effect of factors, such as cell motility, death, *etc.* on the size distribution of the produced tumour spheroids, and also, suggest strategies for future experiments. Importantly, we show that reducing the effective initial density of seeded cells in microgel cultures is the main factor contributing to the production of uniformly-sized clusters.

Most tissues contain more than one cell type, hence, we develop a two-species ABM in Chapter 4, which allows us to include various kinds of between-type (heterotypic), as well as same-type (homotypic), cellular interactions. Using the ABM, we simulate generic systems driven by attractive-repulsive interactions. Examples of two types of scenario include hepatocyte-stellate cell aggregates, striped patterns in zebrafish and predator-prey systems. We vary the homotypic and heterotypic interactions between the species and demonstrate how they change the emerging patterns. As a case study, we use the model to show that, contrary to what is suggested in some earlier work [130, 182], the so-called run-and-chase mechanism is capable of producing the striped pattern when the range of attraction is higher than that of repulsion. To analyse the cell distributions produced by the ABM, we develop PCFs suitable for lattice-based patterns. Similar to the PCFs developed in Chapter 2, these PCFs reflect spatial characteristics of a pattern by showing deviations (aggregations / segregations) of the pattern from CSR over projected directions; two-species systems are investigated in contrast to Chapter 2. We show that the PCF is a powerful tool that can distinguish spatial features in patterns, when this is impossible by visual inspection. Also,

we demonstrate that homotypic and heterotypic PCFs must be used simultaneously, since different mechanisms can have similar homotypic / heterotypic PCFs.

A disadvantage of using the discrete ABMs is that, unlike continuum models, they are not amenable to analytical techniques that provide valuable means to study the behaviour of a system. In Chapter 5, we discuss the derivation of a non-local continuum model from a stochastic discrete model. The aim is to provide a means of predicting the emerging patterns in the simulations of discrete models. The run-and-chase mechanism, in particular, is investigated, and its ability to produce striped patterns, similar to those observed on the skin of zebrafish, is discussed. Importantly, we show cases where the continuous model accurately predict the average behaviour of the stochastic model, and conversely, cases where disagreement between their behaviours is so large such that they produce completely different patterns. Moreover, we outline the importance of considering the discrete nature of cells in our derivations. Hence, beside reinforcing the advantages of using the continuous models for pattern prediction purposes, we sound a note of caution that these models may not accurately represent the behaviour of discrete models for every parameter regime.

Finally, in Chapter 6, we summarise how this thesis can contribute to improving our understanding of tissue development by demonstrating ways to combine spatial quantification of cell distributions and mathematical modelling of processes underlying tissue development. We highlight our main results, and discuss directions for future work.

1.4.1 Published and submitted works

The material in Chapter 2 has been published as [38]:
S. Dini, B. J. Binder, S. C. Fischer, C. Mattheyer, A. Schmitz, E. H. K. Stelzer, N. G. Bean, and J. E. F. Green. *Identifying the necrotic zone boundary in tumour spheroids with pair-correlation functions.* Journal of The Royal Society Interface, 13(123):1-20, 2016.

Similarly, the material presented in Chapter 3 has appeared as [30]:
X. Cui, S. Dini, S. Dai, J. Bi, B. J. Binder, J. E. F. Green, and H. Zhang. *A mechanistic study on tumour spheroid formation in thermosensitive hydrogels: experiments and mathematical modelling.* RSC Adv., 6(77):73282-73291, 2016.

The material in Chapter 4 has been published as [39]:
S. Dini, B. J. Binder, J. E. F. Green, *Understanding interactions between populations: individual based modelling and quantification using pair correlation functions.* Journal of Theoretical Biology, 439: 50-64, 2018.

Chapter 2

Quantification of Spatial Distributions of Cells in Tumour Spheroids

2.1 Introduction

Tumour spheroids are *in vitro* cell aggregates grown from a smaller number of cells initially placed in a non-adhesive environment* [74, 116, 161]. They provide a way to study cancer cell behaviours and interactions in a well-controlled environment, whilst mimicking the *in vivo* arrangement of cells more closely than monolayer cultures. Importantly, tumour spheroids are used in drug testing assays [74, 98], and assessments of the effectiveness of anti-tumour treatments often rely on quantitative measures of the cell distribution within the spheroid [15, 176].

In a sufficiently large spheroid, only the cells in the outermost cell layers can receive enough oxygen and other nutrients to proliferate. This region is termed the proliferative zone, whilst deeper inside the spheroid, quiescent and necrotic zones are formed [161]. Cells in the quiescent zone remain viable but do not proliferate, whilst the innermost cells die due to an accumulation of toxic waste products and a lack of oxygen and other nutrients, forming the necrotic zone [31]. The schematic diagram of Fig. 2.1 illustrates the necrotic, quiescent and proliferative zones within a tumour spheroid. Identifying and quantifying these three regions is important in the analysis of comparative assays on tumour spheroids [35, 42, 55] and mathematical models of the

*In non-adhesive environments, the surfaces of the culture well / container are coated with a layer, which prevents cells from sticking to the surfaces.

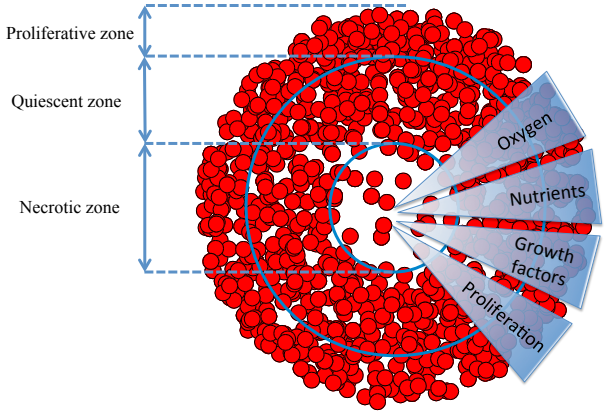


Figure 2.1: Schematic of the necrotic, quiescent and proliferative zones within a tumour spheroid. In the necrotic zone, cell death occurs due to an accumulation of toxic waste products and a lack of oxygen and nutrient supply [31]. The quiescent zone is a region of viable and non-proliferative cells. In the proliferative zone, cells receive enough oxygen and nutrients to proliferate [161].

tumour growth process [4, 24, 68, 156].

In this chapter, we analyse the *in vitro* distribution of cells in nine homotypic tumour spheroids (*i.e.*, spheroids where all the cells are of the same type). The data were obtained using light sheet-based fluorescence microscopy in combination with optical clearing, which allows high-quality three-dimensional images to be produced [157]. Subsequently, the images are processed with a three-dimensional segmentation method to obtain a point cloud representing the cell distribution. The cell distribution gives a point pattern that is subsequently analysed, with the aim being to provide an estimate of the position of the necrotic zone boundary, *i.e.* the distance from the spheroid centre to where the necrotic zone transitions into the quiescent zone.

Standard density-based spatial clustering and data clustering methods (DBSCAN and k-means) are implemented to identify the boundary of the necrotic zone [47, 72, 80]. However, we find that using such existing methods can fail, or produce unreliable results, when the difference in the cell density between the quiescent zone and necrotic zone is small. Therefore, we offer an alternative, statistically based, approach by developing a one-dimensional pair-correlation function to identify the necrotic zone boundary in tumour spheroids.

The pair-correlation function (PCF) is a second-order summary statistic commonly used for analysing point patterns in cell biology [1, 14, 52, 106,

141, 169]. Typically, PCFs describe the relative frequency of Euclidean distances between pairs of data points, indicating the extent of deviations from complete spatial randomness (CSR) [25, 36, 77, 85, 105]. The PCF for a stationary and isotropic point process is defined as

$$g(r) = \frac{\varrho(r)}{\lambda^2}, \tag{2.1.1}$$

where r is usually the Euclidean distance between points, $\varrho(r)$ is the second-order product density (frequency of points separated by a distance r) and λ is the intensity of the point process [77]. When points are distributed uniformly at random (*i.e.* CSR) $\varrho(r) = \lambda^2$ and thus $g(r) = 1$ at all distances. Consequently, aggregation and segregation length-scales correspond to $g(r) > 1$ and $g(r) < 1$, respectively. Therefore, we can quantify spatial features by estimating the PCF for point patterns [142, 124, 159, 160].

However, in some situations the Euclidean distance between points is not the most appropriate distance to study. For example, in a scratch assay used to assess wound healing *in vitro*, the cells move into the wounded region as a front which is approximately a straight line. Binder and Simpson [11] used a one-dimensional pair-correlation function to quantify the spatial patterning of the cells in the Cartesian direction perpendicular to the front in both experimental images and simulations. They also showed that there was no spatial structure in the Cartesian direction parallel to the front. In other situations, the Euclidean distance between points may not be the quantity of interest. For example, Binder *et al.* [13] analysed the angular separation of filaments emanating from two-dimensional images of yeast colonies with a one-dimensional angular pair-correlation function. Similarly, in attempting to identify the necrotic zone boundary in tumour spheroids, we are concerned primarily with variations of cell density in one particular direction (the radial direction).

This then motivates us to formulate a projected one-dimensional pair-correlation function to analyse three-dimensional spatial point patterns with respect to the directions (radial, polar and azimuthal) of the spherical coordinates. In the derivation of this PCF, the (usual) conditions of stationarity and isotropy are relaxed because the projected point processes are in general non-stationary and anisotropic. We examine the accuracy of our method by estimating the PCF for simulated CSR and regular spatial patterns in the spherical coordinate system. In addition, we generate synthetic datasets of cell distributions in tumour spheroids, and demonstrate that the PCF can accurately identify the necrotic zone boundary. This helps with the interpretation and analysis of the PCF results for nine experimental datasets which provide a 'proof of concept' for the usefulness of our approach.

2.2 Mathematical methods

We derive a projected one-dimensional (non-periodic) pair-correlation function to analyse three-dimensional spatial point patterns described in spherical coordinates, (r, θ, ϕ). The aim is to use a sample of N data points to estimate the PCF of the underlying point process. The sample, or point pattern, is a finite subset of three-dimensional space that we can define by

$$S = \{\mathbf{b}_i = (r_i, \theta_i, \phi_i) \mid \ 0 \leq r_i \leq R, \ \ 0 \leq \theta_i \leq \pi, \ \ 0 \leq \phi_i < 2\pi, \\ i = 1, 2, \cdots, N\},$$

where $\mathbf{b}_i$ is the position vector of the ith sample point, and $R = \max(r_i)$ for $i = 1, 2, \cdots, N$. Without loss of generality, we consider the projection, Q, of S onto the interval $[0, L]$, to obtain the projected point pattern

$$S_a = \{a_i = Q(\mathbf{b}_i) \mid \ a_i \in [0, L], \ \ i = 1, 2, \cdots, N\},$$

where a_i is the projected position of the ith sample point. Note that for the radial, polar and azimuthal projections, we have $a_i = r_i$ and $L = R$, $a_i = \theta_i$ and $L = \pi$, and $a_i = \phi_i$ and $L = 2\pi$, respectively. The approach now taken in deriving the PCF is based on normalising the proportion of pairs of projected data points, $G(\Delta a)$, that are separated by a distance Δa in S_a. The normalisation is with respect to the probability of observing such pairs in the projection of the Poisson process, $\overline{G}(\Delta a)$. Due to the discrete nature of points, $G(\Delta a)$ has to be estimated using the average over an interval (numerator of $G(\Delta a)$ in Eqn. (2.2.1)). The quantities $G(\Delta a)$ and $\overline{G}(\Delta a)$ are analogous to the numerator and denominator in Eqn. (2.1.1).

To evaluate $G(\Delta a)$, we introduce the bandwidth h, and obtain the expression

$$G(\Delta a) = \frac{\dfrac{1}{h}\displaystyle\sum_{i=1}^{N}\sum_{j>i}^{N}\mathbf{1}_{(0,h]}(|a_j - a_i| - \Delta a)}{\binom{N}{2}}, \tag{2.2.1}$$

where

$$\mathbf{1}_{(0,h]}(x) = \begin{cases} 1 & \text{if } \ x \in (0, h] \\ \\ 0 & \text{otherwise.} \end{cases}$$

Note that the denominator in Eqn. (2.2.1) accounts for the total number of all possible combinations of pairs of data points.

To evaluate $\overline{G}(\Delta a)$, we first consider the homogeneous Poisson process, $\mathcal{P}$, which is synonymous with CSR. Depending on the form of the projection operator, Q, the projection of the Poisson process, $\mathcal{P}_a$, can be non-stationary and anisotropic (*e.g.* radial projection). However, since $\mathcal{P}$ is the Poisson process, the probability density function f_a for the projected points of $\mathcal{P}_a$ is known. The projected points, $\mathcal{P}_a$, can be considered as samples drawn from a random variable A with probability density function f_a.

In order to find the probability of having a pair of points that are separated by a distance Δa in $\mathcal{P}_a$, we can use the density at a location $a \in [0, L]$, and the cumulative distribution of the points in the intervals $(a + \Delta a, a + \Delta a + h]$ and $[a - \Delta a - h, a - \Delta a)$. Integrating over the interval $[0, L]$ then gives

$$\begin{aligned} \overline{G}(\Delta a) &= \frac{1}{h}\int_0^{L-\Delta a} f_a(a)\ P(a + \Delta a < A < a + \Delta a + h)\ \mathrm{d}a \\ &+ \frac{1}{h}\int_{\Delta a}^{L} f_a(a)\ P(a - \Delta a - h < A < a - \Delta a)\ \mathrm{d}a. \end{aligned} \tag{2.2.2}$$

Since the probability of finding a pair of points a distance Δa apart does not depend on the order of counting the possible pairs (left-wise or right-wise) it can be shown that the two integrals in Eqn. (2.2.2) are equal. When written in terms of the probability density function, Eqn. (2.2.2) becomes

$$\overline{G}(\Delta a) = \frac{2}{h}\int_0^{L-\Delta a} f_a(a) \int_{a+\Delta a}^{a+\Delta a+h} f_a(t)\mathrm{d}t\ \mathrm{d}a. \tag{2.2.3}$$

Using Eqns. (2.2.1) and (2.2.3), the estimate of the non-periodic PCF is then given by

$$g(\Delta a) = \frac{G(\Delta a)}{\overline{G}(\Delta a)}, \tag{2.2.4}$$

where the probability density function, f_a, for each of the three projections is

$$f_r(r) = \begin{cases} \dfrac{3r^2}{R^3} & \text{for } r \in [0, R] \\ 0 & \text{otherwise,} \end{cases} \tag{2.2.5}$$

$$f_\theta(\theta) = \begin{cases} \dfrac{\sin(\theta)}{2} & \text{for } \theta \in [0, \pi] \\ 0 & \text{otherwise,} \end{cases}$$

$$f_\phi(\phi) = \begin{cases} \dfrac{1}{2\pi} & \text{for } \phi \in [0, 2\pi) \\ 0 & \text{otherwise.} \end{cases}$$

Note that depending on the point pattern under study, suitable probability density functions should be selected to identify spatial characteristics over directions of interest. Hence, the above probability density functions are not necessarily the only options to analyse point patterns.

2.2.1 CSR and regular spatial patterns

The method is validated by evaluating the non-periodic PCF, Eqn. (2.2.4), for simulated CSR and regular spatial patterns within a sphere of radius R. The point patterns are shown in the panels of the top row of Fig. 2.2. From left to right they are: (i) CSR pattern, (ii) segregated clusters of points in spherical shells, (iii) segregated clusters of points that are locally aggregated around prescribed angles of ϕ, and (iv) segregated clusters of points in conical shells. The three panels directly below each test pattern in Fig. 2.2 correspond to the PCF evaluation of the radial, azimuthal and polar projections from 1000 simulations.

At short and intermediate length-scales the results for the simulated CSR pattern, in the first column of Fig. 2.2, indicate that there is no spatial structure (in any direction) as $g \approx 1$. However, we see that there is significant deviation from unity in the PCF signals at large distances. This deviation at large distances appears to be inconsistent with our formulation of the PCF, as we might expect the signal to be close to unity at all distances. The explanation for this disparity is the division of small numbers in Eqn. (2.2.4), where the observed frequency and expected frequency of pairs of points at large distances are both small. The results show that the non-periodic PCF is a reliable predictor of CSR at short and intermediate length-scales.

We now consider the regular spatial patterns. They were chosen such that for each spatial pattern there is only spatial structure expected in one of the three corresponding projected patterns. For example, we discuss the results for the points distributed in spherical shells, in the second column of Fig. 2.2. As expected, the PCF indicates that there is no spatial structure for the azimuthal and polar projections (Figs. 2.2(g) and 2.2(h)). For the radial projection (Fig. 2.2(f)), we see a series of five decreasing peaks for increasing values of Δr in the PCF. The highest peak at $\Delta r = 0$ corresponds to pairs of points within each of the five spherical shells (*e.g.* pairs of red points, pairs of green points, etc). The smallest peak at $\Delta r = 200$ corresponds to pair of points with one point belonging to the innermost spherical shell (red)

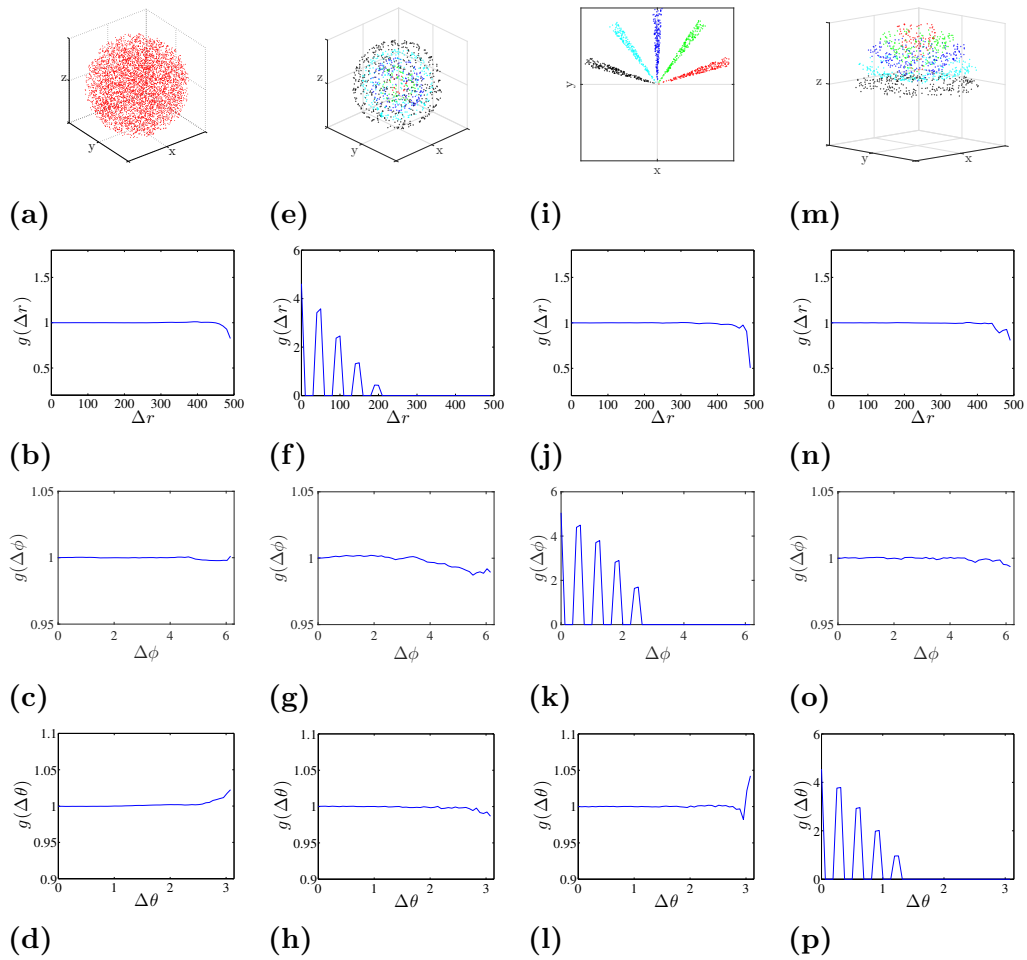


Figure 2.2: CSR and regular point patterns, $h = L/50$. Top row: (a) CSR pattern, (e) segregated clusters of points in spherical shells, (i) segregated clusters of points that are locally aggregated around prescribed angles of ϕ, and (m) segregated clusters of points in conical shells. In the CSR pattern, $N = 5000$ and $R = 500$. In all of the regular patterns, $N = 1000$. The centres of clusters are equally spaced with distance (e) $R/10$, (i) $\pi/5$, (m) $\pi/10$, and the points are distributed uniformly in each cluster in an interval of size (e) $R/100$, (i) $\pi/50$, (m) $\pi/50$. The remaining rows are for the averaged non-periodic PCF from 1000 simulated patterns. Second row: radial projection. Third row: azimuthal projection. Bottom row: polar projection.

and the other to the outermost spherical shell (black). Further information about the spatial patterning can be easily obtained from the signal (*e.g.* the distance between the peaks in the signal is a measure of distance between the spherical shells). A similar discussion holds for the two sets of results in the last two columns of Fig. 2.2. This demonstrates the application of the PCF to quantify spatial structures in point patterns.

2.2.2 Periodic pair-correlation function

In our analysis of the simulated CSR point pattern of Fig. 2.2, we observed that the non-periodic PCF, Eqn. (2.2.4), had significant deviations from unity at large distances– due to the low frequency of occurrence of large separations. This can lead to the incorrect assessment that a point pattern has spatial structure at large distances. This problem of the division of small numbers at large distances can be overcome by defining a periodic PCF.

Following the work of Agnew *et al.* [1] (and others), we re-define the distance between two points in Eqn. (2.2.1) as

$$||a_j - a_i|| = \min(|a_j - a_i|, L - |a_j - a_i|) \quad \text{for} \quad i \neq j \in \{1, \cdots, N\}, \quad (2.2.6)$$

with the consequence that

$$0 \leq ||a_j - a_i|| \leq \frac{L}{2}.$$

Using Eqns. (2.2.1) and (2.2.6), the proportion of pairs of data points, $G_p(\Delta a)$, separated by a distance Δa is given by

$$G_p(\Delta a) = \frac{\dfrac{1}{h}\displaystyle\sum_{i=1}^{N}\sum_{j>i}^{N}\mathbf{1}_{(0,h]}(||a_j - a_i|| - \Delta a)}{\binom{N}{2}}. \quad (2.2.7)$$

This implies that the number of pairs of points separated by a distance $L-\Delta a$ will be added to the number of pairs of points separated by a distance Δa (for $\Delta a \leq L/2$). Therefore, we also need to re-formulate the normalisation term for the periodic PCF, which yields

$$\begin{aligned} \overline{G}_p(\Delta a) &= \frac{2}{h}\int_0^{L-\Delta a} f_a(a) \int_{a+\Delta a}^{a+\Delta a+h} f_a(t)\mathrm{d}t \ \mathrm{d}a \\ &+ \frac{2}{h}\int_0^{\Delta a} f_a(a) \int_{a+L-\Delta a-h}^{a+L-\Delta a} f_a(t)\mathrm{d}t \ \mathrm{d}a. \end{aligned} \quad (2.2.8)$$

Equations (2.2.7) and (2.2.8) then give the periodic PCF

$$g_p(\Delta a) = \frac{G_p(\Delta a)}{\overline{G}_p(\Delta a)}. \tag{2.2.9}$$

When evaluating the periodic PCF, Eqn. (2.2.9), for the simulated CSR pattern in Fig. 2.2, we observed a reduction in the deviations from unity at large distances - *e.g.* compare the blue curves in the second row (PCF) with those in the third row (periodic PCF) of Fig. 2.4.

For the azimuthal projection the periodic distance that separates two points has a clear physical interpretation. It is simply the acute angle that separates the two points a_i and a_j ($i \neq j$). However, the physical interpretation of the distance at which departures from unity occur in the periodic PCF for the radial and polar projections are, in isolation, unclear. This is because the periodic PCF cannot distinguish the distances $L - \Delta a$ and Δa. Therefore, we evaluate both the non-periodic PCF and periodic PCF for synthetic (§ 2.3) and experimental datasets of cell distributions within a tumour spheroid (§ 2.4).

2.3 Synthetic tumour spheroids

We first analyse synthetic tumour spheroid datasets, to help with interpreting the PCFs for the nine experimental datasets. A spatial model is used to generate point patterns with two zones of uniform density. For each synthetic dataset a total of N points are distributed within a sphere of radius R. The points are distributed under the assumption that there are two zones of uniform cell density, partitioned by a radial distance, $r = B$, representing the necrotic zone boundary. The inner zone, $r < B$, is the necrotic zone of the synthetic tumour. The outer zone, $B < r < R$, corresponds to a viable zone (*i.e.* the quiescent and proliferative zones together). To ensure that the viable zone has a larger cell density than the necrotic zone, we distribute uniformly at random $N - \Delta N$ points throughout the entire spatial domain, along with an additional ΔN points in the viable zone only. There is no spatial structure in the azimuthal and polar projections of the synthetic datasets. Therefore, we analyse the radial projection of the synthetic datasets with the non-periodic and periodic PCF.

As the point process is known for the synthetic tumour spheroids, we can derive analytical expressions for the non-periodic and periodic PCF. The

probability density function of distribution of the $N - \Delta N$ points is given by

$$f_1(r) = \begin{cases} \dfrac{3r^2}{R^3} & \text{for } r \in [0, R] \\ 0 & \text{otherwise,} \end{cases}$$

and for the ΔN points

$$f_2(r) = \begin{cases} \dfrac{3r^2}{R^3 - B^3} & \text{for } r \in [B, R] \\ 0 & \text{otherwise.} \end{cases}$$

The total number of pairs of points can be divided into three parts: (1) pairs that the $N - \Delta N$ points in $[0, R]$ generate with each other, (2) pairs that the ΔN points in $[B, R]$ generate with each other, and (3) pairs that the $N - \Delta N$ points in $[0, R]$ generate with the ΔN points in $[B, R]$. Using Eqns. (2.2.2) and (2.2.3), this can be expressed as

$$\begin{aligned} g(\Delta r) &= g_1(\Delta r) + g_2(\Delta r) + g_3(\Delta r) \\ &= \frac{G_1(\Delta r) + G_2(\Delta r) + G_3(\Delta r)}{\overline{G}(\Delta r)}. \end{aligned}$$

The expected proportion of pairs for part (1) is

$$G_1(\Delta r) = \frac{\binom{N-\Delta N}{2}}{\binom{N}{2}} \frac{2}{h} \int_0^{R-\Delta r} f_1(r) \int_{r+\Delta r}^{r+\Delta r+h} f_1(t) \mathrm{d}t \ \mathrm{d}r,$$

which yields

$$g_1(\Delta r) = \frac{G_1(\Delta r)}{\overline{G}(\Delta r)} = \frac{(N - \Delta N)(N - \Delta N - 1)}{N(N-1)}.$$

For part (2) we have

$$G_2(\Delta r) = \frac{\binom{\Delta N}{2}}{\binom{N}{2}} \frac{2}{h} \int_B^{R-\Delta r} f_2(r) \int_{r+\Delta r}^{r+\Delta r+h} f_2(t) \mathrm{d}t \ \mathrm{d}r,$$

and

$$g_2(\Delta r) = \frac{G_2(\Delta r)}{\overline{G}(\Delta r)}.$$

For part (3) the order of counting possible pairs is important, unlike that for the integral of Eqn. (2.2.3). Therefore, we split G_3 into two parts, G_{3r} and G_{3l}, which gives

$$G_{3r}(\Delta r) = \frac{(N-\Delta N)\Delta N}{\binom{N}{2}} \frac{1}{h} \int_B^{R-\Delta r} f_2(r) \int_{r+\Delta r}^{r+\Delta r+h} f_1(t)\mathrm{d}t \; \mathrm{d}r, \quad (2.3.1)$$
$$0 \leq \Delta r \leq R - B,$$

and

$$G_{3l}(\Delta r) = \begin{cases} \dfrac{(N-\Delta N)\Delta N}{\binom{N}{2}} \dfrac{1}{h} \displaystyle\int_B^{R} f_2(r) \int_{r-\Delta r-h}^{r-\Delta r} f_1(t)\mathrm{d}t \; \mathrm{d}r, & 0 \leq \Delta r < B \\ \dfrac{(N-\Delta N)\Delta N}{\binom{N}{2}} \dfrac{1}{h} \displaystyle\int_{\Delta r}^{R} f_2(r) \int_{r-\Delta r-h}^{r-\Delta r} f_1(t)\mathrm{d}t \; \mathrm{d}r, & B \leq \Delta r \leq R. \end{cases}$$

Therefore,

$$G_3(\Delta r) = G_{3r}(\Delta r) + G_{3l}(\Delta r),$$

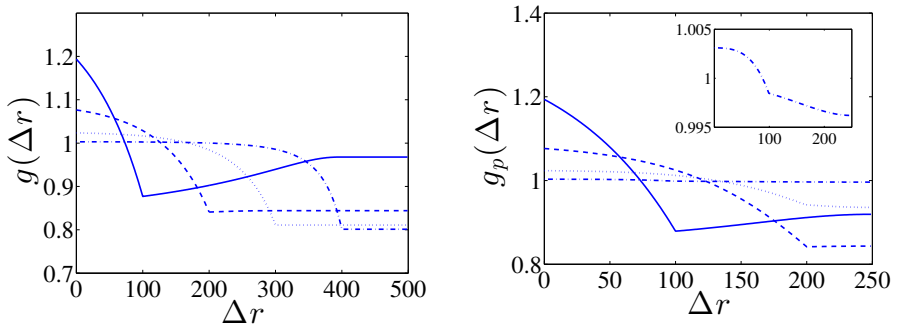
and

$$g_3(\Delta r) = \frac{G_3(\Delta r)}{\overline{G}(\Delta r)}.$$

A similar analysis holds for the periodic PCF, but with the number of pairs a distance $R - \Delta r$ apart being added to the number of pairs a distance Δr apart (*e.g.* see Eqn. (2.2.6)).

Results for four values of the necrotic boundary are shown in Fig. 2.3. We observe that the distance at which a slope discontinuity occurs in the non-periodic PCF (Fig. 2.3a), $\Delta r = s$, uniquely identifies the width of the viable zone, $W = s$. This means that the necrotic boundary, $B = R - W$, can be identified from the non-periodic PCF. But the slope discontinuity in the periodic PCF (at $\Delta r = s$ in Fig. 2.3b) identifies either the width of the viable zone, $W = s$, or the necrotic boundary, $B = s$, with the non-uniqueness being due to periodicity, as discussed in §2.2.2. Therefore, we cannot uniquely determine the necrotic zone boundary from the distance at which the slope discontinuity occurs in the periodic PCF.

We now examine estimates of the non-periodic and periodic PCF from samples of 1000 synthetic tumour spheroids. The central slice from synthetic tumour spheroids (equivalent to the cross-section images of real tumour spheroids) for four values of the necrotic boundary are shown in the first row of Fig. 2.4, where the difference in density between the necrotic and viable zones is visually undetectable (*i.e.* visually indistinguishable from CSR spatial patterning). In the panels of the second and third row, directly



(a) **(b)**

Figure 2.3: Analytic PCFs for synthetic tumour spheroids, $N = 5000, \Delta N = 1000, R = 500$ and $h = R/50$. Solid curves, $B = 400$. Dashed curves, $B = 300$. Dotted curves, $B = 200$. Dashed-dotted curves, $B = 100$. (a) Non-period PCF. (b) Periodic PCF. The inner graph in (b) is for $B = 100$ and it illustrates the small variations of g_p in this case.

below each test pattern, are the corresponding average non-periodic and periodic PCFs (solid curves). The upper and lower broken curves are the 95% confidence intervals, which are the 97.5 and 2.5 percentiles of the 1000 simulations. The arrows identify *critical points* in the estimates of the PCFs, similar to that seen for the points of slope discontinuity in the continuous PCFs (Fig. 2.3).

At short to moderate distances we find a comparable amount of (small) variability in both the non-periodic and periodic PCFs (broken curves, second and third rows, Fig. 2.4). However, at moderate to large distances there is a much greater variability in the non-periodic PCF than that of the periodic PCF. This implies that we have greater confidence inferring salient features of the periodic PCF at large distances. It is therefore advantageous to examine both PCFs together when assessing the spatial pattern of tumours.

To assess the PCFs of the synthetic tumour spheroids it is useful to consider statistical significance of the results, *i.e.* our confidence in distinguishing the results from the CSR state (null case). This is done by presenting the 2.5 and 97.5 percentiles of the PCFs of simulated CSR point patterns (the grey regions in the second and third rows of Fig. 2.4). In each case, 1000 CSR point patterns are generated with the same number of points as the synthetic spheroid. Then, a point pattern's PCF signal found within the grey region could be interpreted as not distinctive from the CSR point process. Therefore, this makes it difficult to estimate the critical points of the PCFs with just one point pattern (or a small sample size) for $B = 100$ in Figs. 2.4(b)

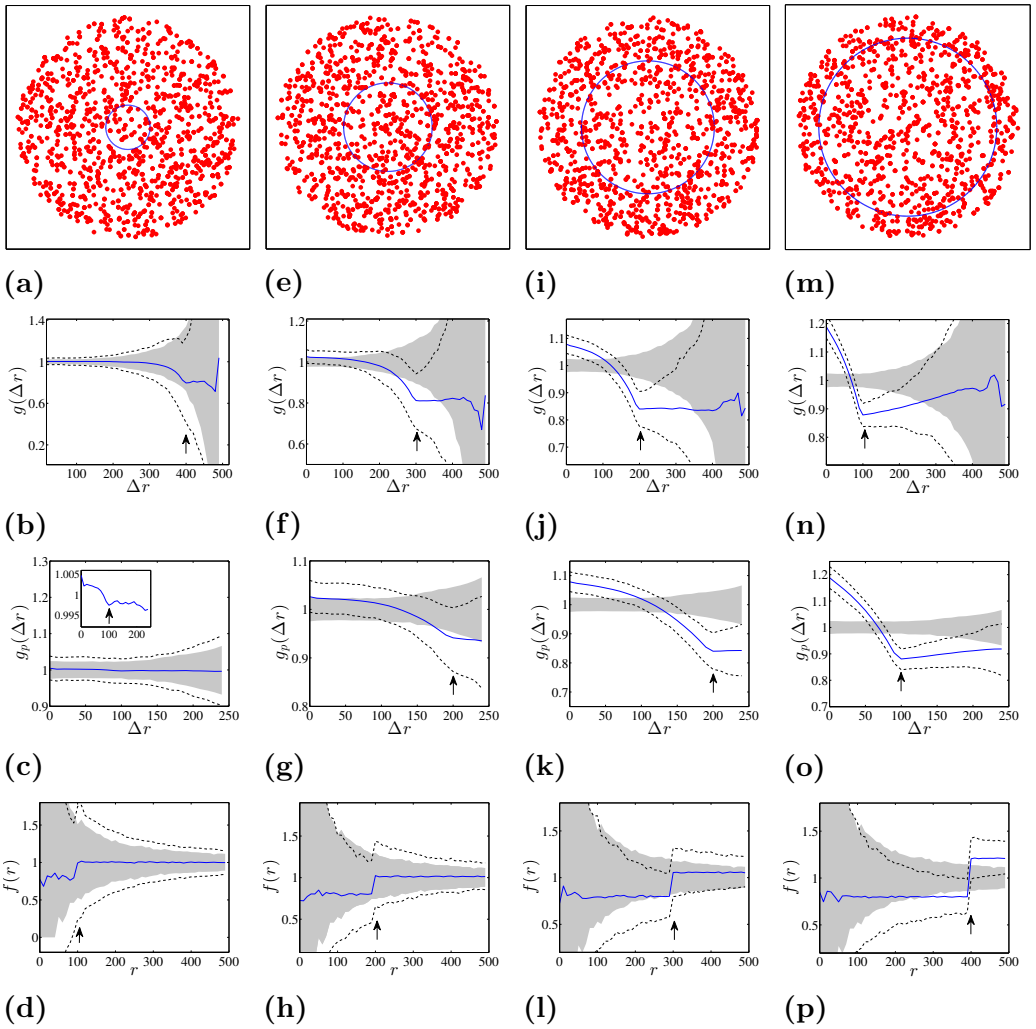


Figure 2.4: Synthetic tumour spheroid point patterns, $N = 5000$, $\Delta N = 1000$, $R = 500$ and $h = R/50$. Columns: left to right $B = \{100, 200, 300, 400\}$. Top row: Central slice of synthetic tumour spheroid. The remaining rows are for statistics from 1000 realisations. Second row: Average non-periodic PCF, solid curves. Third row: Average periodic PCF, solid curves. Fourth row: Average normalised density, solid curves. The upper and lower broken curves are for the 97.5 and 2.5 percentiles, and the arrows are for the known distances of the critical points. The shaded region is for the 97.5 and 2.5 percentiles of 1000 CSR point patterns.

and (c), where the broken curves essentially bound the grey CSR regions. In contrast, the two broken curves at the critical points in Figs. 2.4(j), (n), (k) and (o) both lie below the grey CSR region. This indicates that it is possible to provide an estimate of the necrotic boundary from just one point pattern in the cases when $B = 300$ and $B = 400$. We note that this contrast is mainly due to the difference in density between the necrotic and viable zones, rather than the increase in the necrotic boundary (see Appendix A.1, for a fixed value of $B = 200$ and varying ΔN).

The statistical significance of the PCF results is further examined by comparison to those for the normalised density [13]. This first-order statistic is derived by considering the proportion of points at a distance r from the origin

$$F(r) = \frac{\dfrac{1}{h}\displaystyle\sum_{i=1}^{N} \mathbf{1}_{(0,h]}(r_i - r)}{N}. \tag{2.3.2}$$

Equation (2.3.2) is normalised by the radial projection of a Poisson process

$$\overline{F}(r) = \frac{1}{h}\int_{r}^{r+h} \frac{3r^2}{R^3}\ \mathrm{d}t \ = \ \frac{(r+h)^3 - r^3}{hR^3},$$

to obtain the normalised density

$$f(r) = \frac{F(r)}{\overline{F}(r)}. \tag{2.3.3}$$

In the bottom row of Fig. 2.4, three of the four critical points lie within the grey CSR region and the broken curves essentially bound the grey CSR region. This illustrates that it is difficult to provide an estimate of the necrotic zone boundary from just one point pattern (or small sample size) with the normalised density, demonstrating a greater confidence in the statistical significance of the second-order PCF method.

2.4 Real tumour spheroids

We have shown that the PCF method can provide a reliable estimate of the (known) radial distance of the necrotic zone boundary, B, from the centre of a synthetic tumour spheroid. However, in practice, there are two important distinctions to consider when examining real tumour spheroids.

The first distinction is that real tumour spheroids can be ellipsoidal in shape, for example, see Fig 2.5. Therefore, we consider an ellipsoidal spatial

domain centred at the origin of a Cartesian coordinate system and aligned with the axes,

$$\frac{x^2}{X^2} + \frac{y^2}{Y^2} + \frac{z^2}{Z^2} \leq 1,$$

where X, Y, Z are the lengths of the three semi-principle axes. We now project, or map the points to the interval $[0, 1]$ via

$$a = \sqrt{\frac{x^2}{X^2} + \frac{y^2}{Y^2} + \frac{z^2}{Z^2}}, \quad \text{with} \quad 0 \leq a \leq 1. \tag{2.4.1}$$

The previous PCF analysis for the radial distance of spherical synthetic tumours holds for this non-dimensional problem, with $r = a$ and $R = 1$ in Eqn. (2.2.5). Therefore, we can estimate the dimensionless necrotic zone boundary, $\tilde{B} \leq 1$, for this projection of the data points. The inverse mapping then provides estimates for the lengths of the three semi-principle axes of the ellipsoidal necrotic zone boundary

$$B_X = \tilde{B}X, \quad B_Y = \tilde{B}Y \quad \text{and} \quad B_Z = \tilde{B}Z. \tag{2.4.2}$$

The second distinction is that the necrotic zone boundary in a real tumour spheroid is unknown, but it can be estimated by a human expert. Without prior knowledge of the PCF estimates, human expert estimates for the necrotic zone boundary in nine tumour spheroids were obtained (see Table 2.2). In addition to this, we automate the data collection process and subsequent evaluation of the PCF estimates of the necrotic zone boundary for each tumour (see §2.4.1 and §2.4.2). Together, this allows for an unbiased comparison between the human expert and PCF estimates of the necrotic zone boundary for each tumour (see Tables 2.2 and 2.3).

2.4.1 Experiments and data collection

The experiments from which our data were derived were performed in Buchmann Institute for Molecular Life Sciences (BMLS) at Goethe Universität Frankfurt am Main by C. Mattheyer. Full details of the experimental techniques are given in [38].

Nine homotypic tumour spheroids were formed from the breast cancer cell line T47D by the liquid overlay method [22]. An initial number, N_s, of seed cells (see Table 2.1) together with medium were placed in a convex well, which is coated with a non-adhesive layer. Mature spheroids were obtained after being cultured for 12 days in the incubator, each with N_e number of cells (see Table 2.1). The spheroids were then removed from the wells, fixed, stained with Draq 5 to label the nuclei, and optically cleared. Imaging of the

spheroids was performed with a Digital scanned laser light-sheet fluorescence microscope [89]. They were categorised, by visual inspection, into three groups: (I) spheroids with no visible necrotic core, (II) spheroids with a small necrotic core and (III) spheroids with a large necrotic core. A central slice of the image stack of one tumour spheroid from each of the three groups is shown in Fig. 2.5, with central slices from the remaining six tumour spheroids presented in Fig. A.2 of the Appendix A.2.

The point patterns, or raw data sets are the positions of the nuclei of the cells. Detecting the cell nuclei in the three-dimensional images of the spheroids is achieved by applying a custom multi-scale Laplacian of Gaussian (LoG) detection algorithm [99]. Further details of the detection method can be found in the [38].

Subsets of the raw data points are obtained by removing data points associated with the irregularities of the surface of the tumour spheroids. For each data subset a spatial domain is defined by an ellipsoid centred at the origin with the three semi-principal axes aligned with the Cartesian axes. This is done systematically. (1) Find the smallest convex set of points (i.e. the convex hull) that contain all the raw data points [34]. (2) Fit an ellipsoid to the convex hull, using a linear least squares algorithm [143], to obtain initial estimates of the lengths of the three semi-principal axes, X^*, Y^* and Z^*, with $X^* \leq Y^*$. (3) The origin is chosen as the centre of mass in each data set. The MATLAB function #pca is used to find three orthogonal directions (principle components) in which each data set has the largest variances. The point pattern is then rotated so that the three principle axes of each data set coincide with the Cartesian coordinate system [84]. (4) The data points are projected onto the interval $[0, 1]$ using Eqn. (2.4.1), with $X = X^*$, $Y = Y^*$ and $Z = Z^*$. The global maximum of $F(a)$, given by Eqn (2.3.2), provides an estimate of the non-dimensional distance, $a = \tilde{A}$, associated with the surface irregularities of each tumour spheroid (see Fig. 2.6). (5) Data points with $a > \tilde{A}$ are removed from each raw data set and the inverse map, Eqn. (2.4.2), provides the lengths of the three semi-principal axes, $X = \tilde{A}X^*$, $Y = \tilde{A}Y^*$ and $Z = \tilde{A}Z^*$, (with $X \leq Y$) for the ellipsoidal spatial domains (see Table 2.1).

The subsets of N data points for each of the nine tumours are subsequently analysed with the PCF methods.

2.4.2 Spatial analysis

The nine subsets of data points, with ellipsoidal spatial domains, are mapped to the interval $[0, 1]$ using Eqn. (2.4.1). The non-periodic and periodic PCFs (blue curves) are shown in Fig. 2.5, and Fig. A.2 of the Appendix A.2.

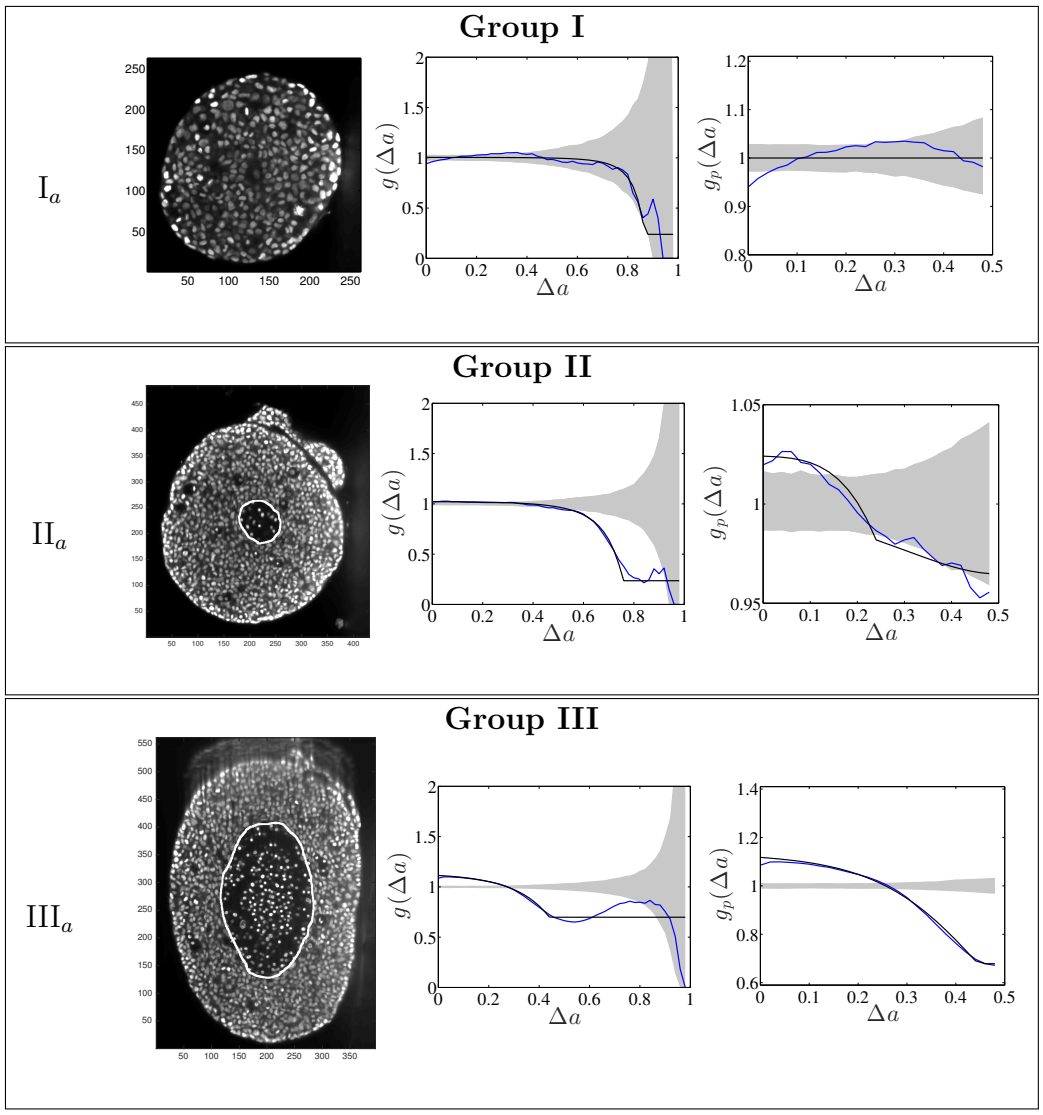
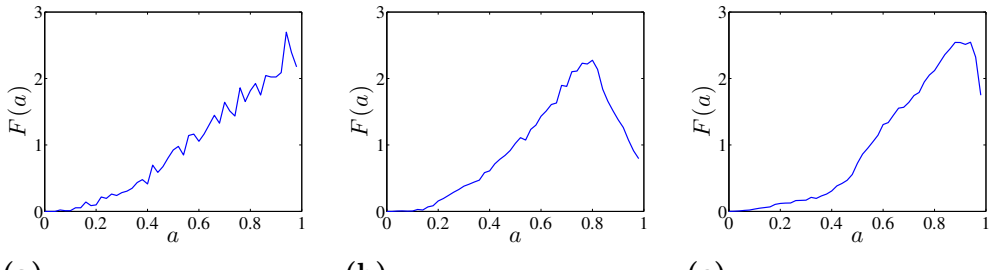


Figure 2.5: Central image slice and corresponding point pattern analysis for one tumour spheroid from each of the three groups. The rows top to bottom correspond to I_a, II_a and III_a respectively. The first column is the central slice of the tumour spheroids. The unit of the length in the images is a μm. The white curves superimposed on the central image slices outline the necrotic zone identified by a human assessor. The second and third columns are the non-periodic PCF and periodic PCF respectively, $h = 0.02$. The blue curves are for the PCF statistics. The black curves are the analytic PCFs best-fit to the statistics (blue curves), using a non-linear least squares method. The shaded region is for the 97.5 and 2.5 percentiles of 1000 CSR point patterns.

	Group I			Group II			Group III		
	I_a	I_b	I_c	II_a	II_b	II_c	III_a	III_b	III_c
X	93	102	114	173	173	145	169	190	169
Y	115	107	149	190	176	201	250	196	233
Z	110	88	133	145	159	178	218	165	180
N_s	500	1000	2000	5000	5000	5000	10000	10000	10000
N_e	4597	3983	10334	25806	23739	25943	36732	32916	35200
N	3607	3127	7900	16356	15593	17864	30555	21226	24729

Table 2.1: Data for nine tumour spheroids. Group I: no visible necrotic core. Group II: small necrotic core. Group III: large necrotic core. X, Y, Z are the lengths of the three semi-principle axes in μm for each ellipsoidal spatial domain. N_s is the number of initially seeded cells for each tumour spheroid. N_e is the total number of cells in the raw data set for each mature tumour spheroid. N is the number of cells in each data subset (i.e. ellipsoidal spatial domain) used in the spatial analysis.



(a) **(b)** **(c)**

Figure 2.6: Removal of data points associated with the irregularities of the surface for three tumour spheroids. Points with $a > \tilde{A}$ are removed from the raw datasets. (a) Tumour spheroid I_a, $\tilde{A} = 0.94$. (b) Tumour spheroid II_a, $\tilde{A} = 0.80$. (c) Tumour spheroid III_a, $\tilde{A} = 0.94$.

The PCFs for the real tumour spheroids can be examined in a similar way to that of the synthetic tumour spheroids, as discussed in § 2.3. However, unlike the synthetic tumour spheroids, the non-dimensional necrotic zone boundary, $\tilde{B}$, is unknown in the real tumour spheroids. Additionally, it is difficult to provide an estimate of the necrotic zone boundary by visual examination of the PCFs alone. To overcome this difficulty, and to automate the PCF estimates of necrotic zone boundary, we fit the analytical PCFs (see §2.3) to the statistical PCFs. A non-linear least squares method with two parameters ΔN and $\tilde{B}$ is used to find the best fit (e.g. see black curves in Fig. 2.5). The point at which there is slope discontinuity in the fitted PCF is taken as the critical point used in determining the estimates for the necrotic zone boundary, $\tilde{B}$. Equation (2.4.2) gives estimates for the lengths, B_X, B_Y and B_Z, of the three semi-principle axes of the ellipsoidal necrotic zone boundary in each tumour spheroid. The two lengths B_X and B_Y (with $B_X \leq B_Y$) for each tumour spheroid are recorded in the second and third row of Table 2.2.

		Group I			Group II			Group III		
		I_a	I_b	I_c	II_a	II_b	II_c	III_a	III_b	III_c
Human	B_X	-	-	-	36	36	34	81	83	93
	B_Y	-	-	-	43	40	53	138	90	128
Non-periodic PCF	B_X	12	6	23	42	46	34	94	92	96
	B_Y	15	6	30	46	47	46	140	94	132
Periodic PCF	B_X	-	-	52	42	49	30	94	101	95
	B_Y	-	-	69	46	50	42	139	104	131
DBSCAN	B_X	84	-	-	43	47	33	77	74	80
	B_Y	103	-	-	47	48	45	114	77	110

Table 2.2: Necrotic zone boundary estimates in μm for the nine tumour spheroids. The shaded PCF estimates are for fitted PCFs (and therefore critical points) that reside within the 97.5 and 2.5 percentiles of 1000 CSR point patterns (e.g. see first row of results for spheroid I_a in Fig. 2.5).

The shaded PCF estimates in Table 2.2 are for fitted PCFs (and therefore critical points) that reside within the CSR shaded regions (e.g. see black curves for spheroid I_a in Fig. 2.5). This means that we can reject these estimates as they may be considered as not distinctive from the CSR point process. This is the case for all three Group I tumour spheroids with no visible necrotic cores, and since the Group I tumour spheroids are the smallest in size it is reasonable to infer that the innermost cells still have sufficient nutrient and oxygen to remain viable [65].

We compare the non-periodic and periodic PCF estimates for B_X and B_Y

		Group II			Group III			Average
		II_a	II_b	II_c	III_a	III_b	III_c	
Non-periodic PCF	B_X	15.9%	26.9%	0.1%	16.5%	10.7%	3.2%	12.2%
	B_Y	7.6%	18.2%	-12.5%	1.4%	4.7%	3.5%	8%
Periodic PCF	B_X	15.9%	35.2%	-11.8%	16.5%	21.6%	2.1%	17.2%
	B_Y	7.6%	25.8%	-20.1%	0.6%	15.9%	2.7%	12.1%
DBSCAN	B_X	18.1%	30.7%	-4.2%	-4.1%	-10.6%	-13.8%	13.6%
	B_Y	10.0%	21.2%	-14.1%	-17.1%	-14.7%	-13.4%	15.1%

Table 2.3: Percentage difference in computed estimates relative to the human estimates, for Group II and III tumour spheroids. The smallest (absolute) percentage differences are highlighted in blue. The rightmost column shows the average (absolute) percentage difference for each method.

to those of a human. The human estimates are based on visual examination of a central image slice of each tumour, where white curves are superimposed onto the images to identify the necrotic core boundary in each tumour spheroid (e.g. see central image slices in Fig. 2.5). Fitting an ellipse to the white curves in each image then provides human estimates for B_X and B_Y, which are shown in the first row of Table 2.2.

Recorded in Table 2.3 (first and second rows) are the percentage difference in the PCF estimates relative to the human estimates for the Group II and III tumour spheroids, with visible necrotic cores. The averaged results (rightmost column) show that the non-periodic PCF estimates have the smallest (absolute) percentage difference, when compared to those of the human. A similar result is found when calculating the overall mean squared error (MSE) for each PCF method; Non-periodic PCF MSE=45 μm^2 and Periodic PCF MSE=96 μm^2 . We believe that the main difference between the PCF and human estimates can be attributed to the fact that the human estimates are based on a single two-dimensional central slice of each spheroid, whereas the PCF estimates are based on the three-dimensional point pattern data of each spheroid. Other contributing factors in the percentage difference are likely to include the processing method of the raw data and the spatial model used for the estimation of the the critical point in the PCFs. We also note that there appears to be a positive bias in the estimates (21 out of 24 in Table 2.3), and further investigation of this is left to future research.

To conclude the analysis, we implement an existing method commonly used to evaluate spatial clustering, using the three-dimensional point pattern data of each spheroid. The density-based spatial clustering of applications with noise (DBSCAN) algorithm classifies points in high-density regions (e.g. viable zone) as a cluster [47]. Points that are in low-density regions (e.g. necrotic zone) are classified as outliers. A subset of outliers for each tumour

spheroid is used to calculate the DBSCAN estimates shown in the bottom row of Table 2.2 (see Appendix A.3). The difference in the DBSCAN and human estimates is comparable to the difference in the PCF and human estimates (see bottom row of Table 2.3), with an overall MSE=138 μm^2. The results demonstrate that the PCF method is a potentially useful alternative to existing standard clustering methods in providing estimates for the necrotic zone boundary in tumour spheroids.

2.5 Discussion

We have derived one-dimensional non-periodic and periodic PCFs for the radial, azimuthal, and polar projections of a point pattern. This is different from the usual Euclidean distance based PCF [25, 36, 77, 85, 105]. Analysis of spatial structure with one-dimensional PCFs is useful in situations where the Euclidean distance is not the main quantity of interest, as is the case in identifying the necrotic zone boundary in tumour spheroids.

We analysed synthetic tumour spheroids (*i.e.* simulated datasets) with our PCFs to illustrate the methodology for identifying the necrotic zone boundary. It was found that a critical point (at which there is a slope discontinuity) in the non-periodic PCFs corresponds to the width of the viable zone in synthetic tumour spheroids, and this allows us to identify the necrotic zone boundary. However, for small sample sizes, when the non-periodic PCF is used, the critical point can be obscured by noise (at large distances). Using the periodic PCF, we reduced the noise in the PCFs, but with the drawback that the critical point corresponds to either the width of the viable zone, or the necrotic zone boundary itself. However, by examining both the non-periodic and periodic PCFs, we are able to provide a more statistically significant (*i.e.* it lies farther from CSR) estimate of the necrotic zone boundary than that obtained using the normalised cell density.

The PCF method was modified to provide estimates for the lengths of the three semi-principle axes of an ellipsoidal shaped necrotic zone boundary, and then applied to three-dimensional point patterns of nine experimental tumour spheroids. The PCF estimates were compared to those of a human and the DBSCAN method [47]. The primary difference in the computed and human estimates was attributed to the human estimates being based only on a two-dimensional slice (image) of each tumour spheroid. The average percentage differences of the PCF and DBSCAN method were comparable (see rightmost column in Table 2.3), and this demonstrates that the PCF method potentially has merit as an alternative to the existing DBSCAN method.

The focus of this study has been to provide estimates for the radial distance(s) of the necrotic core boundary from the centre of a tumour spheroid. These estimates provide a simple metric to classify and categorise tumour spheroids, which has potential application to high-throughput comparative assays [35, 42, 55]. For example, the PCF method could be used to investigate population-level variability in the size of the necrotic zone by using a larger sample of mature tumour spheroids from the same cell-line, grown from the same number of seeded cells. This would allow one to study differences in nutrient consumption between cell types, or the effects of different cell culture methods, or drug treatments. The automatic evaluation of PCF estimates has clear advantage over manual human estimates in the assessment of such high-throughput comparative assays.

The PCF method also has two main advantages over the DBSCAN method. The first being that the DBSCAN method is semi-automatic, requiring human input for each tumour spheroid analysed. We acknowledge that a visual check might be needed after using the least-squares fitting, since this method can lead to undesirable estimates depending on the initial condition, *etc.* However, in this work, we did not observe unsatisfactory fitting results. The second advantage is that only the PCF method provides reliable estimates for point pattern data that is visually indistinguishable from the CSR point process (see Appendix A.3).

Although this work is concerned with homotypic spheroids there are approaches that aim at more complex spheroids, including different cell types and heterogeneities in the microenvironment. Our data analysis can be readily applied to data sets from such complex spheroids. Furthermore, adjusting the segementation method would allow for the extraction of cell position information from histological stains and enable the analysis of sections of cancer patient tumours. Therefore, the PCF method could potentially have a role in diagnostic testing and personalised cancer treatment.

More generally, and in addition to the estimates for the necrotic zone boundary, it is important to understand that the PCFs can provide multiscale spatial information on tumour spheroids (e.g. Fig. 2). Previous studies have shown that the PCF is a close to a sufficient summary statistic, essentially capturing all the spatial information in a given point pattern [11, 83]. Therefore, our PCF method could be implemented in combination with inferencing algorithms such as approximate Bayesian computation [83], which require close to sufficient summary statistics, to parameterise tumour growth models for specific cell types and culture conditions [4, 24, 68, 156].

A further improvement of our quantification method is to analyse the density of triplets of points, which might enable us to reveal more information from the distribution of cells in the tumour spheroids. It has been shown that

studying third-order structure using the density of triplets is essential in some patterns, where PCF and first-order density functions are unable to detect some subtle characteristics of patterns [86]. We leave this as a future study.

Chapter 3

Agent-based Models of Tumour Spheroid Growth in Different Media

3.1 Introduction

In the previous chapter, we developed new methods to study the spatial structure of tumour spheroids grown *in vitro*. In this chapter, we are again concerned with tumour spheroids, but now we aim to understand how they grow under different culture conditions. This requires us to model the growth process, rather than simply quantifying experimental data. In particular, we are interested in understanding how culture conditions can influence the shape and size of the spheroids during their growth. These features are extremely important in anti-cancer drug screening. The aim is to develop culture techniques that produce large, uniformly-sized spheroids. Large variations in size and the presence of irregularly shaped spheroids can reduce the usefulness of the cultured spheroids for research studies.

Culture of the cells within a thermosensitive hydrogel, P(NIPAM-AA), has been proposed as a new method which overcomes some of the disadvantages of traditional suspension culture on a non-adhesive surface. Each of these culture techniques has distinct characteristics, as we explain in the following sections, which influence the growth of the tumour spheroids.

The biological experiments show that microgel culture can produce more uniformly-sized tumour spheroids than the conventional suspension culture method. This is extremely desirable for biologists, since they require a large number of identical tumour spheroids to ensure that they repeat an experiment, *e.g.* drug screening experiment, under same conditions. In addition,

it is observed that cells are more evenly distributed throughout the microgel (compared to suspension culture, where they tend to congregate at the bottom of the well), cell aggregate formation is much slower due to network resistance and the cell death rate is much smaller in comparison with the suspension culture. Having observed the experimental results, we hypothesise that the ability of the microgel to produce a more homogeneous distribution of the cells in the vertical direction, and lower the cell death are the key factors that result in a more uniform size-distribution of spheroids. In order to investigate this hypothesis, we develop an agent-based model (ABM), incorporating the distinctive features of the two culture methods.

ABMs have previously been widely used to investigate various biological phenomena [46, 165], including tumour growth [81, 132, 138]. However the study of tumour spheroid formation progression within the microgel scaffold has not been conducted yet. By changing some of the parameters in our model, we are able to take into account the different environments experienced by cells in microgel and suspension cultures, and investigate their effects on spheroid formation.

3.2 Development of agent-based model

The ABM consists of a lattice, where each site on the lattice can either be empty or occupied by a cell. At the beginning of a simulation, a specified number of cells are placed at random on the lattice. The number and position of the cells are updated at discrete intervals of time. During each timestep, cells are selected in random order without replacement* (*i.e.* each cell is selected once), and act according to rules for cell behaviour (including cell movement, proliferation and death) that are described in detail below. Our model uses a two-dimensional lattice, as the experimental data derives from two-dimensional images. This has the further advantage of reducing the computational cost compared to a three-dimensional model. The method by which we take into account the three-dimensional nature of the experiments is explained in detail below.

3.2.1 Cell movement

We let the probability that a cell will attempt to move during a time step be P_m. We assume this probability will depend upon the cell's environment. We take a higher value of P_m to represent cells in suspension, and a lower value of

*This method is known as Random Sequential Update (RSU)

P_m for cells in microgels, representing some resistance to movement (such as cell-ECM adhesion, and high viscosity of microgel solution). Cell movement is assumed to involve two component behaviours: unbiased random motion (in which a cell will attempt to move to one of the four neighbouring lattice sites with equal probability) and biased motion (where cells will preferentially move towards other cells). In our model, cells attempt to move according to the biased motion rule with probability P_b (hence the probability of moving according to the unbiased random motion rule is $1 - P_b$). Thus the tendency of cells to aggregation is represented by the probability of biased motion, P_b: for cell types that are strongly inclined to create clusters after seeding, P_b will be close to one, whilst for those that do not tend to aggregate, P_b is close to zero [1]. P_b and P_m are chosen based on the hypothesised characteristics of the cells in the medium before the start of each simulation.

When a cell attempts to move using the biased motion rule, the direction in which it moves is determined as follows. For each of the four directions, we calculate the following probability:

$$P_v(k) = \frac{v(k)}{\sum_{i=1}^{4} v(i)} \quad \text{for} \quad k = 1, 2, 3, 4,$$

where $v(k)$ is the number of cells at the right ($k = 1$), left ($k = 2$), up ($k = 3$) and down ($k = 4$) direction of a cell within its range of attraction (l) (see Fig. 3.1). Note that the $P_v(k)$ sum to unity. We then subdivide the interval $[0, 1]$ into four sub-intervals: $[0, P_v(1)], (P_v(1), P_v(1) + P_v(2)], (P_v(1) + P_v(2), 1 - P_v(4)), (1 - P_v(4), 1]$, and draw a random number uniformly distributed on $[0, 1]$. If the number chosen lies in the first interval, the cell attempts to move right, if in the second, it attempts to move left, *etc.*

Area exclusion is accounted for in the cell movement rules [10]: two or more cells cannot occupy a site at a time. Therefore, at any point a cell attempts to move to an already occupied site, the movement is aborted.

Mombach and Glazier (1996) suggest that in Brownian motion a cell moves 1/6 of its diameter in 30 minutes. Therefore, in 3 hours each cell moves one site in the lattice when $P_m = 1$. The probability of biased motion is set to $P_b = 0.9$ with the range of attraction, $l = 3$ for both types of culture. Thus, it is very likely that the cells move towards each other when they are very close to each other (*e.g.* close enough that some parts of the cell may make physical contact). Long range attractions are not considered here as the chemotactic signals are assumed to be negligible.

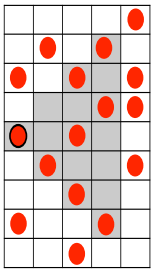


Figure 3.1: A sample distribution of cells is depicted. For example, the red cell marked out with the black border senses 7 cells at its right side, $v(k = 1) = 7$, where $l = 3$ is the range of the attraction range, illustrated by the shaded sites.

3.2.2 Cell proliferation

The probability of proliferation, P_p, determines the rate of proliferation at each timestep. When proliferation occurs for a cell, the parent cell keeps its position and the daughter cell occupies one of the four adjacent sites. Area exclusion is accounted for in the proliferation rules as well, *i.e.* if a cell already occupies the chosen site for the daughter, the proliferation event is aborted.

We follow the model of Qi *et al.* (1993) who take into account the effects of nutrient depletion as cancer cells proliferate by making the probability of proliferation dependent upon the total number of cells. We define the two-dimensional cell density, ρ, as the area fraction occupied by the cells, given by

$$\rho(t) = \frac{N(t)B}{A},$$

where $N(t)$ is the number of cells at time t, B is the cell area ($15^2 \mu m^2$) and A is the area of the lattice in μm^2. The area of a well in the experiments is around 20mm^2. Thus, the length of each side of the well is $a = \sqrt{A} \approx 4.5$mm. Therefore, a lattice representing the wells has a length of $L = a/0.015 = 300$ sites. As the proliferation rate is assumed to decrease when cell density increases, due to lack of nutrients, we define the probability of proliferation as

$$P_p = k\left(1 - \frac{\rho(t)}{C}\right), \tag{3.2.1}$$

where C is carrying capacity and k is growth rate. Qi *et al.* (1993) suggest $0.26\ \mathrm{d}^{-1} < k < 0.48\ \mathrm{d}^{-1}$ (d denotes day). From the experimental data shown in [30], cells have approximately the same rate of proliferation in suspension and microgel: we hence use the same probability of proliferation for both

types of culture with $k = 0.48$ d^{-1} $= 0.06$ (3h)$^{-1}$ (recall that 3 hours is the length of our timestep, as this is approximately the time taken to move one cell diameter). C is estimated to be 0.6 by processing the images of the experiments to obtain the maximum area fraction that may occur.

3.2.3 Cell death

We assume that cells need to adhere to another cell or a surface to survive and proliferate [2]. As the cells in suspension culture are prevented from adhering to the substrate, we assume that they are likely to die after one day if they do not adhere to other cells [61, 178]. In the model, we assume a cell is adhered to another cell if there is a least one cell in the four squares adjacent to it. However, in microgel culture, the cells can survive and proliferate by adhering to the microgel. We thus considered two different death probabilities, P_d, in simulations of suspension and microgel experiments. In suspension simulations, P_d for an isolated cell is low during the first day of culture, increasing rapidly to unity thereafter. For microgel simulations, for the sake of simplicity, P_d is taken to be zero.

Thus, the probability of death for isolated cells in suspension and microgel culture respectively can be defined as follows

$$P_d(n_t) = \begin{cases} \dfrac{1}{1+e^{-100(n_t-8)}} & \text{suspension,} \\ 0 & \text{microgel,} \end{cases} \tag{3.2.2}$$

where n_t is the number of consecutive timesteps (each timestep represents 3h) during which a cell is not adhered to another cell. This equation gives a rapid increase in the death probability after $n_t = 8$ (1 day) for suspension culture, whilst $P_d = 0$ for microgel cultures for $0 \leq n_t \leq 168$ timesteps (168 timesteps equals 21 days, which is the period of experiments). Note that the functional form of P_d in Eqn. (3.2.2) is simply chosen to reproduce the assumed qualitative behaviour of the cells in the two different culture environments described above. In the case of suspension culture, other functions that increase rapidly to unity after a period of 1 day would be expected to produce similar results.

The ABM was run at a timestep of $T_s = 3h$. The initial population was set to $\rho_0 = \rho(0) = 0.05$ in suspension and $\rho_0 = 0.005$ in microgel to give the best fit to the experiments. The reason for this difference is that in suspension most cells settle down and interact in a layer close to the substrate, whereas the cells in microgels lie in multiple layers. Note that ρ_0 represents the effective initial population of cells in our 2D ABM, *i.e.* the cells that

interact with each other within a layer, not the total population of cells in a well. This leads to a considerably higher effective ρ_0 for the suspension culture case where the cells are mainly within one layer, compared to the microgel, where they are distributed more evenly throughout the gel.

The number of lattice sites at each side was scaled down to half: 150 sites (*i.e.* lattice size of 150×150 sites), similar to the size of the images of the experiments. Moreover, a periodic boundary condition was applied in the simulations. This means that when a cell moves out of the domain at one side, it will re-enter from the opposite side. Applying periodic boundary conditions mitigates the effect of boundaries on the distribution of cells [1], since there is no physical boundary present at the edge of the experimental images with which we aim to compare our results.

The size of the clusters (cluster diameter) was calculated as follows. Firstly, the areas of the clusters were computed, using the function #bwarea in MATLAB. This function gives an estimate of the area, n, created from adjacent connected pixels, using the algorithm explained in [137]. Multiplying the area by 15^2 (we assume that each cell is a square with the size of a lattice site), gives the area of a cluster in μm^2 - note that the cells are considered squares with the size of a lattice site. The cluster diameter, D, was then calculated as the diameter of a circular cluster of equivalent area *i.e.* :

$$D = 30\sqrt{\frac{n}{\pi}}.$$

3.3 Results

3.3.1 Experimental data

The experiments were performed by Dr. Xiaolin Cui under the supervision of Dr. Hu Zhang in the School of Chemical Engineering, University of Adelaide. Full details of the techniques used can be found in [30]. Briefly, HeLa cells were cultured in the two growth media, *i.e.* microgel and suspension. Fig. 3.2 shows selected images of the experiments (the spheroids are viewed as clusters in the images) after 7 (first row), 14 (second row) and 21 (third row) days. As the first column of this figure shows, the growth speed of the spheroids in microgel is lower than that in suspension (see the second column of this figure). However, the spheroids have more regular and uniform shapes in the microgel. This can also be inferred from the histograms of Fig. 3.3 which were obtained by processing the two-dimensional images using a software package [30]. The histograms show the average frequency of the cluster diameters at the mentioned days for 12 parallel conducted experiments. The graphs show

a narrower size distribution for microgel, meaning that microgel produces a more uniform size distribution compared to suspension. However, this comes with the downside of having the most frequent spheroids (the ones that will possibly be used for drug screening) with a smaller size, about 95 μm, at day 21 (see Fig. 3.3e).

3.3.2 Simulation results

Now, we turn to simulating spheroid growth for microgel and suspension cultures, and compare the results with the ones illustrated earlier in Figs. 3.2 and Figs. 3.3 for the experiments. Fig. 3.4 shows the pattern and distribution of clusters in suspension and microgel simulations. Here, we consider extreme cases for P_m of is 1 for suspension and 0 for microgels respectively. Thus, the cells are very motile in suspension and the cells in microgel do not move at all. Note that in the histograms, clusters with $n < 7$ are not included, since they are too small to be considered as spheroids.

Comparing the histograms of Fig. 3.4 with Figs. 3.2 and 3.3 shows that the distributions of cluster size at days 7, 14 and 21 generated by simulations are very similar to those of the biological experiments, implying that the simulation results are in good agreement with the experimental results.

At day 21, the average cluster size is $\bar{D} \approx 186$ μm and the standard deviation of the cluster size is $\delta \approx 107$ μm for simulated suspension cultures, while $\bar{D} \approx 133$ μm and $\delta \approx 54$ μm for simulated microgel cultures. Hence, the distribution of clusters is more uniform in the microgel with a lower standard deviation and the average cluster size is smaller as well.

The main differences between the simulated suspension and microgel experiments are in the cells motion, death and the initial cell density - the latter affects proliferation rate as defined by Eqn. (3.2.1). We aim to understand how each of them affects the distribution of cluster size, and so, parameter-sweeping tests were carried out in which one of the parameters is varied while the others are kept constant. We swept the parameters in a physically plausible region, *i.e.* where their values are within a range that is consistent with the physical properties of the medium. Small variations about the previous values of P_m and ρ_0 (see the caption of Fig. 3.5) are analysed.

The parameter sweeping was done for 150×150 lattice with $\rho_0 = \{0.005,$ $0.01,$ $0.015,$ $0.02\}$ and $P_m = \{0, 0.005, 0.01\}$ in the microgel and $\rho_0 = \{0.035,$ $0.04,$ $0.045,$ $0.05\}$ and $P_m = \{0.8, 0.9, 1\}$ in suspension. Fig. 3.5 shows that $\bar{D}$ and δ increase with ρ_0. Thus our model suggests that using a higher initial cell density, ρ_0, leads to formation of bigger clusters, which is desirable, but this has the unwanted effect of reducing the uniformity of cluster size.

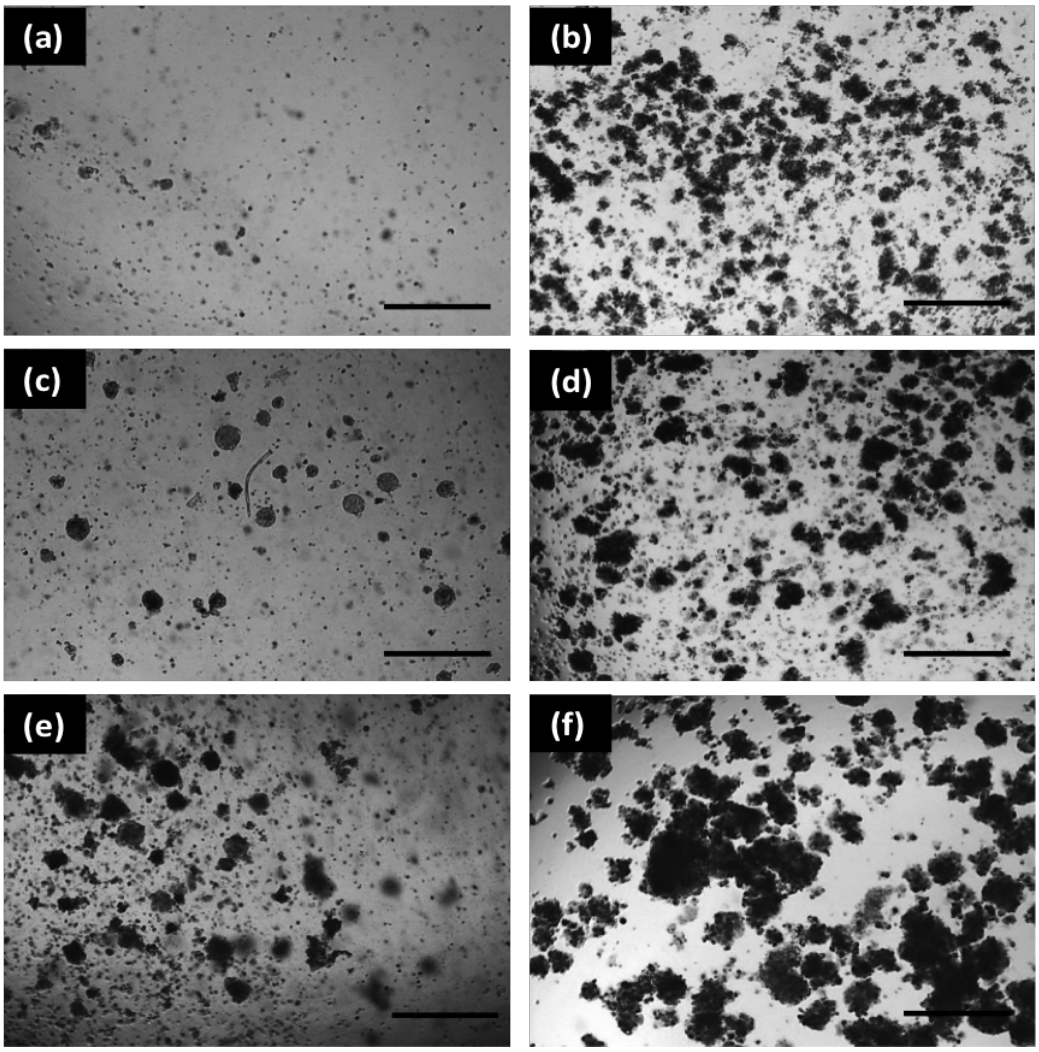


Figure 3.2: Experimental results: HeLa cells within (a,c,e) and without microgel (b,d,f) at different culture days. The cells within microgel are released (*i.e.* extracted from the surrounding microgel) at room temperature. Scale bar is 500m. (a),(c),(e) is HeLa cells within microgels culture after 7 days, 14 days, 21 days respectively. (b), (d), (f) is HeLa cells in suspension culture after 7 days, 14 days and 21 days respectively. (Images reprinted from [30])

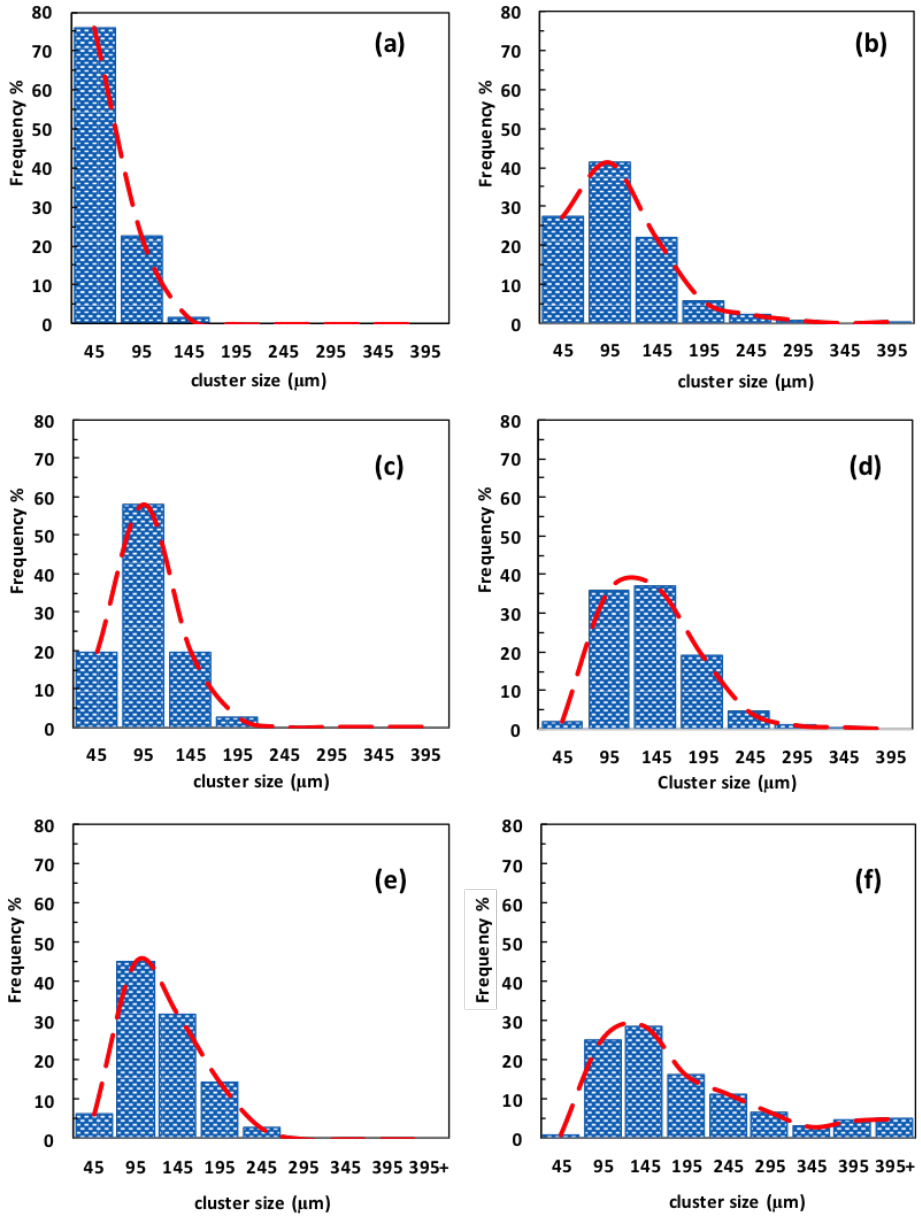


Figure 3.3: Experimental results: HeLa cells spheroids size distribution. (a),(c),(e) is HeLa cells within microgel cultures after 7 days, 14 days, 21 days respectively. (b), (d), (f) is HeLa cells in suspension culture after 7 days, 14 days and 21 days respectively. (Images reprinted from [30]).

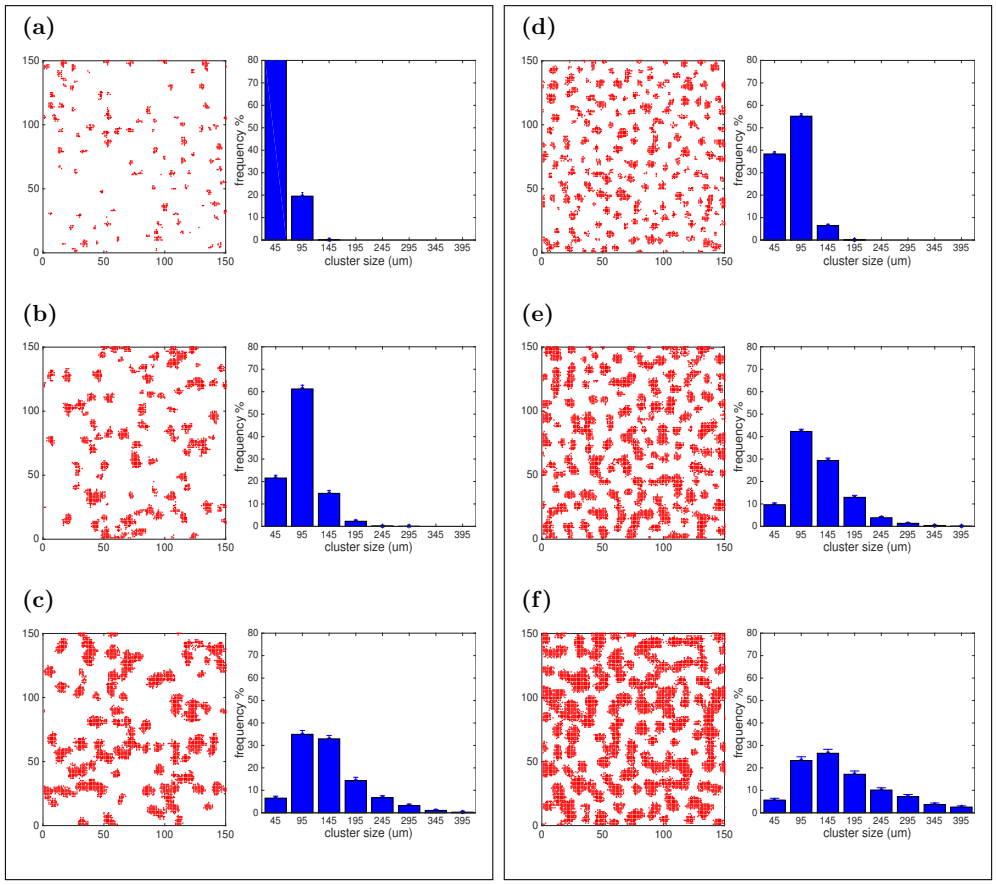


Figure 3.4: Simulation results: Patterns of spheroid formation and distribution of cluster size in microgel (left box) and suspension (right box) simulations. (a,d) day 7, (b,e) day 14 and (c,f) day 21. The average with the %95 confidence intervals of t-distribution are depicted as error bars for 50 simulations. In suspension: $P_m = 1$ and $\rho_0 = 0.05$. In microgel: $P_m = 0$ and $\rho(0) = 0.005$. The lattice has side of $L = 150$ sites, the range of attraction is $l = 3$, the probability of biased movement is $P_b = 0.9$, the proliferation constant is $k = 0.06\text{d}^{-1}$, the carrying capacity is $C = 0.6$ and the death probability is defined in Eqn. (3.2.2). The dead cells are not shown in the images.

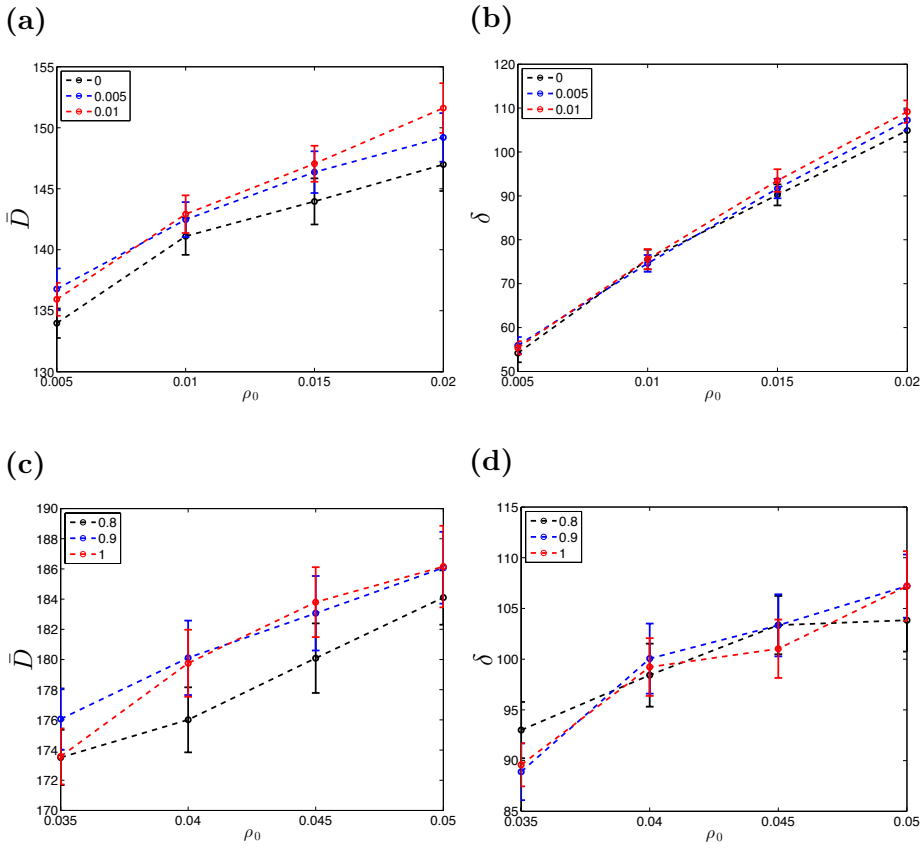


Figure 3.5: Parameter sweeping test. Standard deviation, δ, and average, $\bar{D}$, of cluster size are depicted. (a,b) microgel: $\rho_0 = \{0.005, 0.01, 0.015, 0.02\}$. $P_m = 0$ black, $P_m = 0.005$ blue and $P_m = 0.01$ red. (c,d) suspension: $\rho_0 = \{0.035, 0.04, 0.045, 0.05\}$. $P_m = 0.8$ black, $P_m = 0.9$ blue, $P_m = 1$ red. The points in the graphs are computed average over 50 simulations and the error bars are %95 confidence intervals of t-distribution. The values of the other parameters are the same as in Fig. 3.4.

Hence the choice of ρ_0 would involve a trade-off between cluster size and size-variability.

As illustrated in Figs. 3.5b and 3.5d, our model predicts that the rate of increase of δ with ρ_0 for microgel culture is greater than for suspension culture. Thus, increasing the initial population would have a more deleterious effect on the uniformity of the clusters in microgels compared to suspension culture.

Sweeping the values of P_m in a wider range, we determined that motility of cells can strongly affect $\bar{D}$ and δ. Fig. 3.6 shows that increasing P_m leads to the formation of bigger clusters. The reason is that, when P_m is high enough, randomly moving cells can find bigger clusters nearby and attach to them. In addition, the highly motile cells in small clusters are more likely to find bigger clusters and attach to them. Therefore, increasing P_m can reduce the number of small clusters and increase the number of larger ones.

3.4 Discussion

We developed an ABM to explore the reasons for the different size distributions observed in spheroids grown by two different culture methods. In the model, the cells behave according to rules for movement, proliferation and death. This work was motivated by the biological experiments, as described in [30]. We tried to keep the model as simple as possible in order to focus on the effect of the major parameters on the spheroid formation. The ABM was developed in two-dimensions, since this facilitates comparison with the two-dimensional experimental images, and also reduces the computation time required.

Our ABM was successfully able to reproduce the experimental results for spheroid formation rate and size distribution, for both microgel and suspension cultures. Our results are thus consistent with the main differences between cells in the two different cultures being in their proliferation and death rates, and their effective initial density. Note that proliferation rate depends on cell density, as described by Eqn. (3.2.1), thus, it is significantly different in suspension and microgel. Hence, the more uniform size distribution of spheroids produced by microgel culture could be due to its ability to separate the cells in multiple layers, reducing the effective initial density. However, even when the number of cells is low, the microgel provides a substrate for cells to survive and proliferate, so spheroids can still be produced. By contrast it is not possible to reduce the initial population of cells significantly in suspension cultures, since the cells would die out and spheroid formation would not occur.

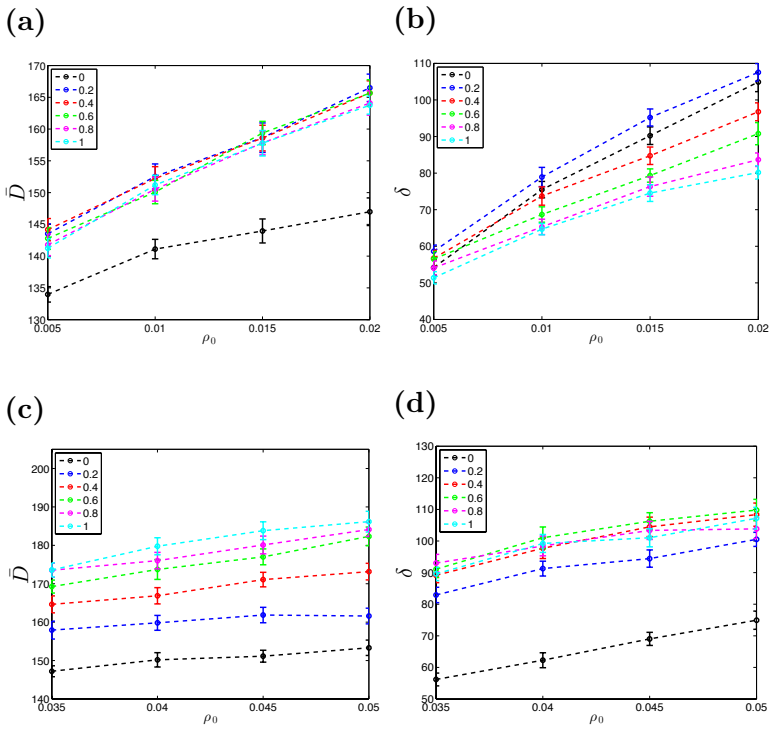


Figure 3.6: Parameter sweeping with a same range of P_m for the culture media. Standard deviation, δ, and average cluster size, $\bar{D}$ are depicted for (a,b) microgel and (c,d) suspension. The curves are for $P_m = \{0,\ 0.2, 0.4,\ 0.6, 0.8\ , 0.1\}$ with the colours as indicated in the legends of the graphs. The other parameters are the same as for the simulations of Fig. 3.4.

The model predicts that the initial cell density plays a crucial role in determining the features of the formed spheroids. For the parameter ranges we considered, higher initial densities led to larger spheroids, but at the cost of introducing greater variability in spheroid size. This effect was predicted to be more pronounced for microgel cultures than suspension cultures. Further experiments will be required to test these predictions.

Our methods for quantifying the experimental and simulation data in this chapter were less sophisticated than those used in Chapter 2. This was due to the limitations of the available experimental data, which consist of two-dimensional images. Specifically, we were not able to extract the three-dimensional cell positions, and thus, we focused on analysing only the size distribution of clusters (tumour spheroids). However, if data on cell positions were available, as in the previous chapter, we could calculate the PCF for the results, which would in turn, allow us to make more exacting comparisons between the simulations and the experiments. Moreover, this would improve our ability to estimate the parameters of the ABM based on the experimental data.

Chapter 4

Agent-based Modelling and Quantification of Pattern Formation in Two Interacting Species

4.1 Introduction

In the previous chapter, we saw how cell aggregates can form as a result of attractive interactions between cells, in combination with proliferation. More generally, there are numerous instances in nature where individuals interact so as to produce a large-scale pattern, such as in the migration of flocks of birds, swarming of bees, foraging in ants, or cells in a developing tissue, *etc.* One longstanding question is to understand how the variety of patterns emerge from relatively simple underlying interactions between the individuals [7]. Commonly, the types of interactions considered are restricted to attraction and repulsion, arising as responses to a range of external stimuli such as those received by the sensory organs of animals (*e.g.* seeing a predator, smelling food resources, *etc.* [131]), or, on the cellular level, resulting from chemical gradients and / or mechanical forces [29, 164]. Patterns such as swarms or cell aggregates may be produced where only a single species is present. However, when individuals of multiple species are present, the potential variety of patterns is greatly increased, as a result of variations in the degree of intermixing or segregation of the species. Three examples, where in each case two different types of individuals are present, are shown in Fig. 4.1. Fig. 4.1a shows a co-culture of stellate cells and hepatocytes (two types of liver cell), where stellate cells pull the hepatocytes into aggregated

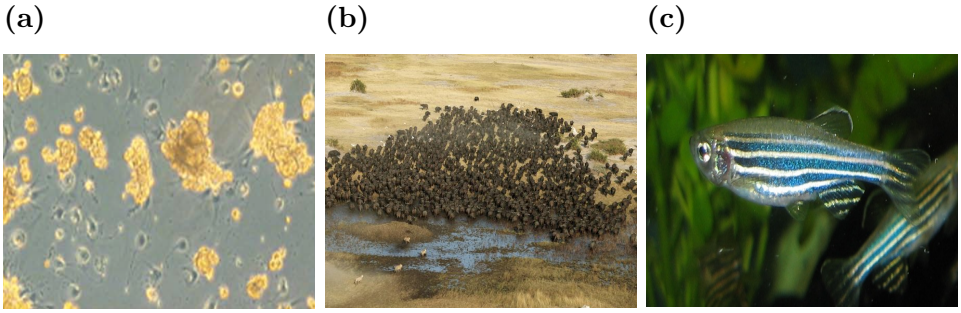


Figure 4.1: Examples of multi-species spatial patterns. (a) Hepatocyte-stellate cells co-cultured to produce spheroids as fundamental part of liver (reprinted from Thomas *et al.* (2005), with permission from Eur. Cells Mater). (b) Lions hunting buffalo on the Duba Plains (image by Beverly Joubert, beverlyjoubert.com). (c) Pattern formation on the skin of a zebrafish (image from Wikipedia).

spheroids [164]; Fig. 4.1b shows three lions chasing a buffalo herd, where the buffalo try to escape and the herd pattern changes accordingly; and Fig. 4.1c shows an image of a stripe pattern on the skin of a zebrafish formed by two differently-coloured types of cell (xanthophores and melanophores) [78, 79, 133].

Mathematical models have made significant contributions to our understanding of how different types of inter-individual interactions can lead to different large-scale patterns. Well known examples include the reaction-diffusion model of morphogenesis, introduced by Turing (1952) (which was later applied to animal coat patterns by Murray [117, 118, 119]), and the chemotaxis model presented by Keller and Segel (1970) which produced new insights into the mechanisms underpinning the formation of cell aggregates in dictyostelium (and many other cell types). In many mathematical models, the underlying interactions between the individuals are idealised as being combinations of attraction and repulsion [109]. The interplay between these opposing forces influences the pattern. For example, Mogilner *et al.* (2003) shows that short-ranged repulsion combined with long-range attraction is necessary to produce cohesive and well-spaced groups in a population of a single species (similar works can be found in [28, 50, 168]). At the cellular level, a variety of factors which would produce these types of interactions, such as chemo- attractants / repellents [100], traction forces [64], volume exclusion [10], *etc.*, have been investigated. For predator-prey interactions, a model is proposed by Chen and Kolokolnikov (2014), where different behaviours

depending on the strength of attraction between the prey and predator are discussed.

The modelling approaches used to study pattern formation can broadly be divided into continuum and agent-based. Continuum models are formulated in terms of the densities of the species of interest, functions of space and time which obey systems of partial differential equations (PDEs). They have the advantage that analytical techniques, such as linear stability analysis, can be used to understand aspects of their behaviour over the entire parameter space, which can help to give insight into the mechanisms underpinning particular phenomena. Examples would include the non-local (integro-PDE) models introduced by Mogilner and Edelstein-Keshet (1999) for the study of swarms. In this kind of model, movement of an individual is influenced by superposition of the forces exerted by the surrounding individuals. These forces are represented by a convolution integral in this continuum framework [64, 70, 130]. Although the continuum models provide good population-level information on pattern formation, because individuals are discrete, it is difficult to compare their predictions with experiments on anything other than a qualitative level [64].

In contrast to continuum models, agent-based models (ABMs) represent each individual explicitly. ABMs are more realistic in this sense, as they allow us to consider individual level behaviour, which usually includes an element of stochastic behaviour (*e.g.* [30, 108, 110, 125]). However, this comes at higher computational cost. In this work we develop an ABM with two species of agents to study the interactions between the individuals. The model developed here builds upon the earlier work for the one-species case described in [1, 30].

The fact that stochasticity can play a role in the distribution of the individuals presents a challenge. Each realisation of the process (*i.e.* , simulation of the model) will be different, owing to differences in the (random) initial conditions, or the randomness inherent in the interactions (even when the interaction rules are fixed). We need to be able to identify which features of the spatial pattern are robust (*i.e.*, recur in many realisations of the process). As a first step, this requires suitable statistical tools to quantify the spatial distributions of individuals so that we can compare the patterns observed in different experimental or simulated datasets. Pair correlation functions (PCFs) are popular candidates for statistical analysis of individuals patterns [37, 77]. PCFs quantify spatial patterns by showing deviations from complete spatial randomness (CSR), where the individuals are distributed uniformly at random. They describe the frequency of pairs of individuals separated by a certain distance, relative to what is expected for CSR [11]. PCFs are increasingly being used to quantify cell distributions [13, 38, 52, 114, 169].

Agnew *et al.* (2014) suggest that the PCF can be used to distinguish between the spatial patterns that arise from cell aggregate formation due to cell proliferation, and those where the aggregates form as a result of cell-cell attraction (*e.g.*, due to chemotaxis).

Here, we follow the approach presented by [11] for quantifying single-species volume exclusion processes. We extend their method to develop a PCF for multi-species volume exclusion processes on a two-dimensional Cartesian lattice. The method is based on normalising the frequency of pairs by finding their expected values for the uniform distribution [11]. Periodic continuation is applied in order to eliminate boundary effects [77]. We then compute the PCF for realisations of our ABM simulated using different sets of parameter values (strengths and ranges of attraction and repulsion). The PCF allows us to characterise various spatial features from the patterns, such as intensity of clustering, spacing between the clusters, *etc.* We demonstrate how the parameter values in the model can be related to the PCF calculated from the model output (averaged over a number of realisations). This suggests ways in which the models can be fitted to data [83, 144, 166].

4.2 Model and quantification method

4.2.1 Multi-species agent-based model

In this chapter, we simulate the interactions between individuals (which can be cells, animals, *etc.*) using an ABM. The model used here is an extension of that developed in [1, 30]. Agents move on a discrete two-dimensional lattice (x, y), of dimensions $X \times Y$. Motivated by applications in cell biology, we consider the domain to be periodic, thus, any agent that leave the lattice from one side, re-enters again from the opposite side. We introduce this assumption since in cell biology experiments, only a small 'window' is observed (corresponding to the field of view of the microscope), which generally does not include the edge of the well in which the cells are cultured. We consider that, on average, a cell moving outside the field of view is likely to be replaced by another moving into it, which is approximated by adopting periodic boundary conditions. Such boundary conditions eliminate edge effects, allowing us to focus on the effects of different types of inter-individual interactions. In this work, we restrict our attention to the case where there are only two types of agents.

Agents move on the lattice according to two motion rules: biased motion (with probability P_b) and unbiased random motion (with probability $1 - P_b$). In unbiased random motion, an agent attempts to move to one of its four

neighbouring sites with equal probability. In biased motion, the probability of moving to each of the neighbouring sites is affected by the fact that an agent can sense the number of other agents within a certain range. The probabilities are calculated based on the rules of attraction and repulsion which are assumed to operate between agents of the specific types involved. Note that each site can only be occupied by only one agent at a time; if a agent attempts to move into an occupied site, the move is aborted. The model thus takes account of volume exclusion ([10, 154]).

In the case of biased motion, we need to define and calculate the probability of movement to each of the four neighbouring lattice sites. We denote this probability by P_k, where k indicates the direction of attempted movement: $k = 1$ - right, $k = 2$ - left, $k = 3$ - up, and $k = 4$ - down, where $\sum P_k = 1$. We assume that agents of type m are attracted (repelled) by agents of type p within a range α_{mp} (γ_{mp}), where $m, p \in \{1, 2\}$. Fig. 4.2 shows how the agents sense their neighbours within a certain range and in a specific direction. We take the bias probability to depend only upon the number of agents of each type within the relevant attraction / repulsion ranges in each of the four directions.

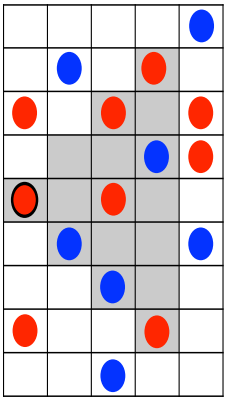


Figure 4.2: Directional neighbourhood of an agent. A sample two-species pattern is depicted with the red ('1') and blue ('2') colours. For example, the red agent marked out with the black border has its range of attraction shown by the shaded sites, $\alpha_{11} = \alpha_{12} = 3$.

The probability of moving in each of the four directions, P_k, is calculated from a four-dimensional directional bias vector, $\boldsymbol{v}$. We define this for each agent at each timestep as

$$\boldsymbol{v} = \boldsymbol{A}\boldsymbol{a} + \boldsymbol{R}\boldsymbol{r},$$

where $\boldsymbol{A}$ ($\boldsymbol{R}$) is a matrix whose entries are the number of neighbouring agents

within the attraction (repulsion) range and $\boldsymbol{a}$ ($\boldsymbol{r}$) is a two-dimensional weight of attraction (repulsion) vector. These, in turn, are defined as follows. For an agent of type m (where $m = 1$ or 2), $\boldsymbol{A}$ is given by

$$\boldsymbol{A} = \begin{bmatrix} E_{mm}(\alpha_{mm}) & E_{mp}(\alpha_{mp}) \\ W_{mm}(\alpha_{mm}) & W_{mp}(\alpha_{mp}) \\ N_{mm}(\alpha_{mm}) & N_{mp}(\alpha_{mp}) \\ S_{mm}(\alpha_{mm}) & S_{mp}(\alpha_{mp}) \end{bmatrix}, \qquad \text{for} \quad m, p \in \{1, 2\},\ m \neq p,$$

where $E_{mm}(\alpha_{mm})$ is the number of agents of type m to the right of the agent of type m, within the range determined by α_{mm} (*e.g.* in Fig. 4.2 $E_{11}(\alpha_{11}) = 4, E_{12}(\alpha_{12}) = 3$, where $\alpha_{11} = \alpha_{12} = 3$). Similarly, $W_{mm}(\alpha_{mm})$ is the number of agents to the left, $N_{mm}(\alpha_{mm})$ is the number of agents above and $S_{mm}(\alpha_{mm})$ is the number of agents below. Likewise, $E_{mp}(\alpha_{mp})$, $W_{mp}(\alpha_{mp})$, $N_{mp}(\alpha_{mp})$ and $S_{mp}(\alpha_{mp})$ account for the heterotypic neighbouring agents. The weight of attraction vector is defined by

$$\boldsymbol{a} = \begin{bmatrix} a_{mm} \\ a_{mp} \end{bmatrix},$$

where a_{mm} and a_{mp} are the weights of attraction of agents of type m to agents of the same type and of different type, respectively. Noting that repulsion will produce biases is the opposite direction compared to attraction, for an agent of type m, $\boldsymbol{R}$ and $\boldsymbol{r}$ are similarly defined to be:

$$\boldsymbol{R} = \begin{bmatrix} W_{mm}(\gamma_{mm}) & W_{mp}(\gamma_{mp}) \\ E_{mm}(\gamma_{mm}) & E_{mp}(\gamma_{mp}) \\ S_{mm}(\gamma_{mm}) & S_{mp}(\gamma_{mp}) \\ N_{mm}(\gamma_{mm}) & N_{mp}(\gamma_{mp}) \end{bmatrix}, \qquad \text{for} \quad m, p \in \{1, 2\},\ m \neq p,$$

and

$$\boldsymbol{r} = \begin{bmatrix} r_{mm} \\ r_{mp} \end{bmatrix},$$

where r_{mm} is the weight of homotypic (same type) repulsion for agents of type m and r_{mp} is the weight of heterotypic (different type) repulsion between agents of type m and p.

Finally, we calculate the probability of moving in each of the four directions, P_k, by normalising $\boldsymbol{v}$:

$$P_k = \frac{v_k}{\sum_{q=1}^{4} v_q}, \quad \text{for} \quad k = 1, 2, 3, 4. \tag{4.2.1}$$

Thus, the direction in which the agent will tend to move (reflected by P_k) is determined by the interplay of homotypic and heterotypic attraction and repulsion influences.

Note that there are alternative ways to implement the biased movement of cells. For example, instead of presetting the value of P_b, we can use the magnitude of $\boldsymbol{v}$, which depends on the number of surrounding cells, to determine P_b at each timestep.

We simulate the model as follows. During a discrete time-step, each agent is selected in random sequential order (without replacement), and is given the opportunity to move either according to the unbiased or biased motion rule. For an agent located at (x, y), the nature of the motion is determined by generating a random number r_1 from a uniform distribution with support $[0, 1]$. If $r_1 \in [0, P_b)$ the agent attempts to move according to the biased motion rule; otherwise, the motion is unbiased and it attempts to move with equal probability to one of its four neighbouring sites $(x \pm 1, y \pm 1)$. In the case of biased motion, we generate a second random number r_2 from a uniform distribution with support $[0, 1]$, and calculate the values of P_k using Eqn. (4.2.1). The agent moves to $(x + 1, y)$ if $r_2 \in [0, P_1)$, $(x - 1, y)$ if $r_2 \in (P_1, P_1 + P_2]$, $(x, y + 1)$ if $r_2 \in (P_1 + P_2, 1 - P_4)$ and $(x, y - 1)$ if $r_2 \in (1 - P_4, 1]$. Note that if the agent attempts to move to a site that is occupied by another agent, then that move is aborted, and one aborted movement is counted. These processes are repeated until the spatial distribution of agents evolves to a quasi-steady state, which is determined by examining the evolution of the number of aborted moves in the simulation [1]. As shown by Agnew *et al.* (2014), in systems that produce steady patterns, the number of aborted movements increases initially, and then plateaus after a certain time. Based on this, we terminated simulations well after the number of aborted movements plateaus.

4.2.2 Pair correlation function

We use PCFs as a means to characterise the spatial patterns generated by the ABM simulations. PCFs characterise spatial patterns by showing deviations from CSR (where agents are distributed uniformly at random) [77]. The deviations characterise different spatial features such as the lengthscales of aggregation (or segregation) which are affected by the mechanisms underlying the interactions between individuals. Computing the PCF for a sufficiently large number of realisations of the same experiment can thus potentially provide us with information about these mechanisms.

Binder and Simpson (2013) presented a method to compute the PCF for single-species distributions on lattice. In their method, the normalisation

factor, which is the expected value of the number of pairs separated by a certain distance in CSR, is computed for distances in the x and y direction (not the Euclidean distance). Then, the number of pairs occurring in the pattern of interest is divided by the corresponding normalisation factor to give the PCF. Here, we use the same approach and derive the relevant formulae for computing the PCF on periodic domains populated with two (or potentially more) species of agents. Note that the restriction to periodic domains is motivated by the application to cell biology experiments, as explained in §4.2.1 (we assume our domain represents a 'window' onto a larger domain in which the interactions occur). A further advantage of making this assumption is that PCFs estimated for patterns on finite domains have significant errors for large distances, whereas periodic continuation helps to mitigate this problem [77]. (Henceforth a reference to a PCF in this chapter should be taken to imply one derived assuming periodic boundary conditions.) Our PCFs are defined for homotypic and heterotypic pairs. Indicating the two types of agent by '1' and '2', we denote the PCFs for the 1-1, 2-2 and 1-2 pairs g_{11}, g_{22} and g_{12} respectively. In addition, since the methods for x and y distances are identical, for brevity we present only the formulae for the x direction.

The PCF in the x direction can be defined as

$$g_{mp}(i) = \frac{c_{mp}(i)}{\hat{c}_{mp}(i)}, \quad \text{for} \quad m, p \in \{1, 2\}, \tag{4.2.2}$$

where $c_{mp}(i)$ is the number of pairs a distance i apart and $\hat{c}_{mp}(i)$ is the normalization factor. $c_{mp}(i)$ for any pattern is obtained by counting the number of pairs with distance i in the x direction and it can be mathematically formulated as

$$c_{mp}(i) = \sum_{l=1}^{N_m} \sum_{j=J}^{N_p} \mathbf{1}_i(|x_l^{(m)} - x_j^{(p)}|), \quad J = \begin{cases} l+1, & m = p, \\ 1, & m \neq p, \end{cases} \tag{4.2.3}$$

where $x_l^{(m)}$ is the x-coordinate of the l^{th} cell of type m. N_m and N_p are the populations of the species m and p respectively and $\mathbf{1}_i(x)$ is the indicator function defined as

$$\mathbf{1}_i(x) = \begin{cases} 1, & x = i, \\ 0, & \text{otherwise.} \end{cases}$$

The expected number of pairs a distance i apart for CSR can be written in terms of the probability of observing the pairs, given by

$$\hat{c}_{mp}(i) = M_{mp}\, P_{mp}(i),$$

where M_{mp} is the total number of m-p pairs and $P_{mp}(i)$ is the probability that an m-p pair, distanced i apart, occurs in the lattice. M_{mp} is given by

$$M_{mp} = \begin{cases} \dfrac{N_m(N_m-1)}{2}, & \text{if} \quad m = p, \\ N_m N_p, & \text{if} \quad m \neq p. \end{cases}$$

The difference in M_{mp} for m-m and m-p is due to the permutation of heterotypic pairs.

In a lattice of width X and height Y, $P_{mp}(i)$ can be found by the ratio of the number of pairs of lattice sites distance i apart, $\hat{d}_{mp}(i)$, to the total possible pairs, given by

$$P_{mp}(i) = \begin{cases} \dfrac{\hat{d}_{mp}(i)}{\frac{1}{2}XY(XY-1)}, & \text{if} \quad m = p, \\ \dfrac{\hat{d}_{mp}(i)}{XY(XY-1)}, & \text{if} \quad m \neq p. \end{cases} \tag{4.2.4}$$

All of the possible pairs that may be observed in a lattice of size $X = 6$ and $Y = 2$ are shown in Fig. 4.3, for the sake of illustration. As this figure shows, the number of possible pairs of lattice sites distance i apart in one row is

$$\hat{r}_{mm}(i) = X, \quad \text{for} \quad i = 1, ..., \frac{X}{2} - 1.$$

Here, we assumed X and Y are even, however, the method is the same for odd X and Y for distances $i = 1, ..., (X-1)/2$.

Note that if the two agents were of different types, each possible pair shown in Fig. 4.3 should be counted twice due to the possible permutation between them. Thus, for m-p pairs the counts should be multiplied by two, *i.e.* $\hat{d}_{mp} = 2\hat{d}_{mm}$, $m \neq p$.

Then, the possible number of pairs that can occur when there are multiple rows is given by

$$\hat{d}_{mm}(i) = \hat{r}_{mm}(i)Y + 2\hat{r}_{mm}(i)\binom{Y}{2} = Y^2 X, \quad \text{for} \quad i = 1, ..., \frac{X}{2} - 1,$$

where the first term at the RHS of the above equation accounts for the Y cases where the paired points are in one row (Figs. 4.3a and 4.3b), and the second term accounts for the paired points in different rows (Figs. 4.3c and 4.3d).

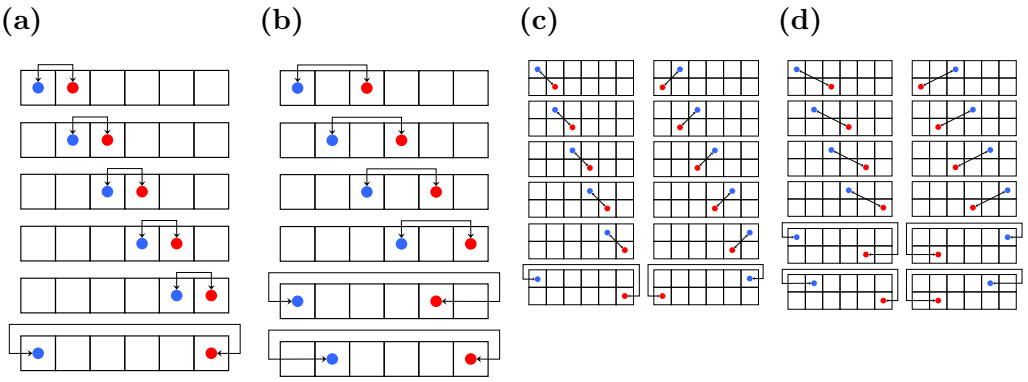


Figure 4.3: All possible combinations of pairs of sites in a lattice with $X = 6$ and $Y = 2$. The arrows depict the distances which can be direct or periodic. The pairs of unit distance apart in (a) one row and (c) two rows. The pairs of distance two apart in (b) one row and (d) two rows.

Finally, substituting $\hat{d}_{mm}(i)$ into Eqn. (4.2.4), yields the normalisation factor:

$$\hat{c}_{mp}(i) = \begin{cases} \dfrac{N_m(N_m-1)Y}{XY-1}, & \text{if} \quad m = p, \\ \dfrac{2N_mN_p\,Y}{XY-1}, & \text{if} \quad m \neq p. \end{cases} \tag{4.2.5}$$

Thus, by computing $c_{mp}(i)$ defined in Eqn. (4.2.3), the normalising factor $\hat{c}_{mp}(i)$ in Eqn. (4.2.5) and plugging them into Eqn. (4.2.2), we are able to evaluate the PCF for a given pattern.

For a spatial domain that is populated uniformly at random the expected value of the PCF is unity at all distances. When the PCF is greater than unity at some distance we have aggregation, and when the PCF is less than unity at some distance we have segregation. In the following section, we use the PCF to identify homotypic and heterotypic aggregation / segregation length-scales in three examples of two-species spatial patterns.

4.2.3 Illustrative two-species spatial patterns

We begin by evaluating the PCFs in the x direction for the patterns shown in the first column of Fig. 4.4 (illustrations of the PCF for a variety of one-species patterns can be found in [11] for comparison). Understanding how information about the pattern is encoded in the PCF for these examples will help to guide our interpretation of the PCF in the more complex situations which are presented in §4.3. We follow the practice found in earlier papers [1,

11] and use linear interpolation to represent the discrete PCFs as continuous curves (columns 2-4 in Fig. 4.4).

The first example we consider is where the two species are distributed uniformly at random, as is illustrated in the first row of Fig. 4.4. As expected, we observe that the heterotypic and homotypic PCFs have a small-amplitude oscillation around unity, indicating that there is no heterotypic or homotypic spatial structure in the pattern. This provides a check on the derivation of the PCFs, which were normalised with respect to CSR.

In the second example, individuals of the same type are distributed in two single-species clusters (red and blue squares in second row Fig. 4.4). The heterotypic PCF is equal to zero for $0 < i \leq 30$ (as the two clusters are 30 units apart), and then increases linearly for $30 < i \leq 50$ (as each cluster is 20 units in length) to a maximum value at $i = 50$ (the distance between the two centres of the clusters). Therefore, there is heterotypic segregation at short to intermediate distances and heterotypic correlation at large distances. In contrast, both homotypic PCFs indicate short to intermediate scale aggregation only, for $0 < i \leq 20$.

For the third and final example, the two-species are distributed uniformly at random within two clusters and the heterotypic and homotypic PCFs are all the same (third row of Fig. 4.4). Additionally, the interpretation of the PCFs is similar to that of the PCF for the single-species clustering pattern in Figure 4 of [11]. Subsequently, the heterotypic, homotypic and overall population aggregation / segregation length-scales are also all the same. We observe short scale aggregation ($0 < i \leq 20$), intermediate scale segregation ($20 < i \leq 30$), and long scale aggregation ($30 < i \leq 50$).

4.3 Results

Having described our modelling approach and quantification method, we are now in a position to investigate the types of patterns produced by different combinations of inter-individual interactions, and the extent to which different pattern-forming mechanisms can be distinguished by analysis of the resulting pattern using the PCF. For each set of results in Figs. 4.5–4.8, we evaluate the average PCFs in the x direction for $N = 60$ simulations that have evolved to a quasi-steady state from initial conditions where both species were distributed uniformly at random throughout the domain. As in Agnew *et al.* (2014), we use the number of aborted movements as the measure of reaching the quasi-steady state. The layout of Figs. 4.5–4.8 is similar to that of Fig. 4.4, with the additional broken curves representing the 95% confidence intervals of the t-distribution for the average PCFs (solid curves).

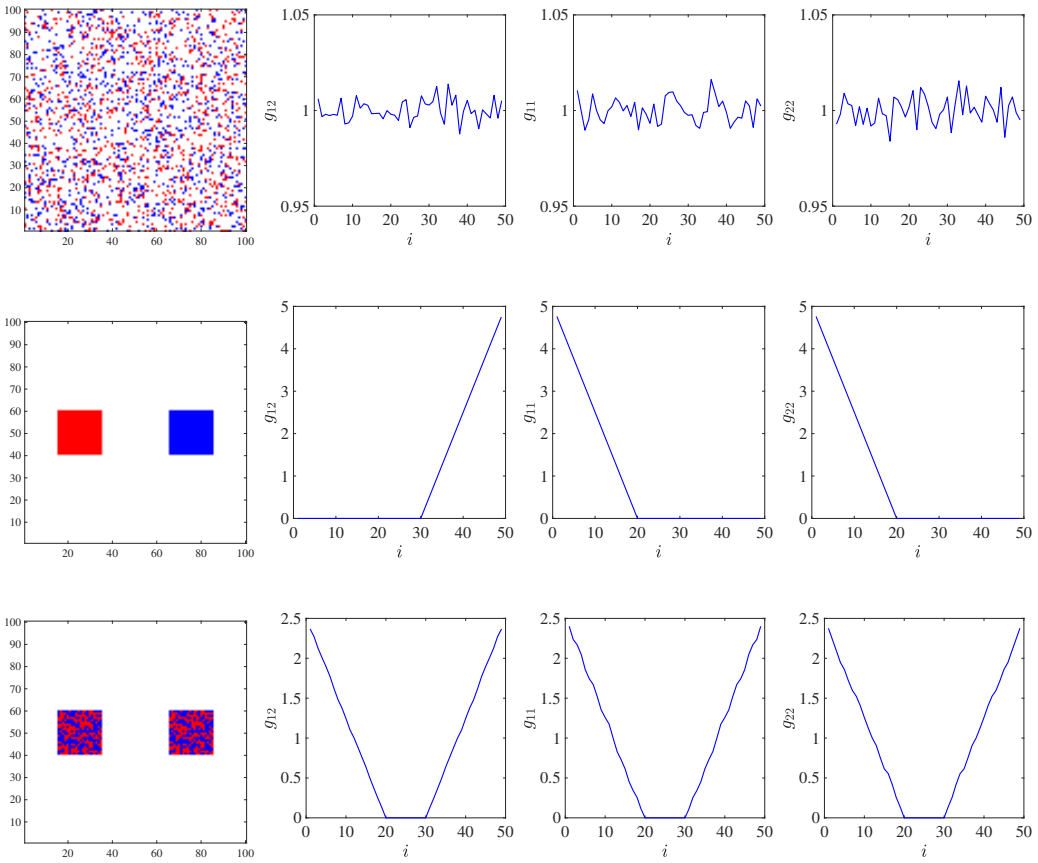


Figure 4.4: Three two-species patterns and their PCFs in the x direction, $N_1 = N_2 = 400$. First column of panels: From top to bottom, the agents are distributed uniformly at random, in two segregated clusters and in two intermixed clusters. Second column of panels: Heterotypic PCFs. Third column of panels: Homotypic PCFs for the red species. Fourth column of panels: Homotypic PCFs for the blue species.

In all the simulations, we choose a high value of the probability of biased motion, $P_b = 0.8$, which allows us to focus our study on cell interactions effects, rather than the effect of unbiased motion. Reducing P_b does not significantly change the spatial characteristics of the produced patterns, but can lead to *e.g.* emergence of less compact clusters. Also, the ranges of attraction and repulsion are set to $\alpha_{mp} = \gamma_{mp} = 5$, further concentrating the majority of our investigation on the effect of varying the strengths of attraction and repulsion, a_{mp} and r_{mp}. This restriction is relaxed in the last set of results (Fig. 4.9), where we study the effect of varying the ranges of attraction / repulsion in the system. Similar sets of results are found for the average PCFs in the y direction, as the interactions and initial conditions make no difference between the x and y directions. We consider two generic types of inter-species interactions: mutually attractive (or repulsive) and attractive-repulsive.

4.3.1 Mutually attractive or repulsive inter-species interactions

The first case we consider is where the inter-species interactions are either mutually attractive or repulsive (*i.e.*, type 1 agents are attracted to (repelled by) type 2 agents, and type 2 agents are attracted to (repelled by) type 1 agents). This is motivated by observations of co-cultured hepatocytes and stellate cells *in vitro*, which produce cell aggregates (see Fig.4.1a). The experiments undertaken by Thomas *et al.* (2005) indicate that in co-culture, cluster formation is enhanced compared to hepatocyte only culture (see the images and videos in [164]). A continuum model of this process was developed by Green *et al.* (2010). They assumed attractive interactions between hepatocytes, and mutual attraction between hepatocytes and stellates. (Short range repulsion between cells of all types was also included in the model, to represent the effect of overcrowding at high cell densities.) Depending upon the relative strength of hepatocyte-hepatocyte and hepatocyte-stellate interaction, the model found that cells formed clusters within which the two types were either segregated, partially-segregated, or intermixed. We will consider similar scenarios here using our model, which has the advantage over continuum models of representing the cells as discrete entities, and including the effects of crowding more naturally (since only once agent can occupy each lattice site).

Heterotypic interactions only

We begin by analysing a system governed by heterotypic interactions only (so $a_{11} = a_{22} = r_{11} = r_{11} = 0$) and where the strengths of attraction and re-

pulsion are the same for both species $a_{12} = a_{21}$, $r_{12} = r_{21}$. Thus, interactions between agents of the same type are governed simply by volume exclusion (as only one agent can occupy a lattice site). The strengths of heterotypic attraction and repulsion are then varied, and the results are presented in Fig. 4.5.

When attraction is much stronger than repulsion clusters are formed in which the two species intermix, as shown in the first row of Fig. 4.5. The PCF for this type of pattern can be interpreted by considering the illustrative example in the last row of Fig. 4.4 and the one-species patterns found in [1, 11]. Similar to the results in the last row of Fig. 4.4, the heterotypic and homotypic PCFs are almost identical and therefore the heterotypic, homotypic and overall population (*i.e.* ignoring agents types) aggregation / segregation length-scales are also all the same. The values of the PCFs above unity at short distances indicate short scale aggregation (*i.e.* the clusters in the first row of Fig. 4.5), with the minimum in the PCFs at $i \approx 10$ providing a quantitative measure of the average spacing between the clusters, or the scale of segregation. The fact that the three PCFs are the same shows that there are no significant differences in the distributions of the two species, indicating they are intermixed uniformly at random within each cluster. At intermediate and large distances, for $i > 20$, the PCFs are close to unity, implying that the centres of the clusters are distributed uniformly at random, as observed for the one-species patterns in [1, 11].

When repulsion is much stronger than attraction clusters are again produced, but in this case the species are now segregated, so that the red and blue agents are no longer intermixed (see fourth row of Fig. 4.5). The clusters are also no longer compact and roughly circular, but elongated, eccentric shapes. However, the fundamental characteristics of the PCFs are similar to those for the illustrative example in the second row of Fig. 4.4. Where heterotypic aggregation / segregation occurs, we observe homotypic segregation / aggregation. In other words, the heterotypic and homotypic PCFs are in anti-phase. For example, at the distance $i \approx 10$ there is maxima in the heterotypic PCF and a minima in both the homotypic PCFs (fourth row of Fig. 4.5).

Unlike the case of strong attraction where there is heterotypic aggregation at short-distances (first row of Fig. 4.5), in the case of strong repulsion we observe short-scale heterotypic segregation (fourth row of Fig. 4.5). Furthermore, for strong repulsion we observe heterotypic and homotypic aggregation / segregation at large distances, which is in contrast to the lack of spatial structure at large distances when compared to that of strong attraction.

The above discussion shows that the PCFs can distinguish and characterise either strong attraction or strong repulsion heterotypic interactions

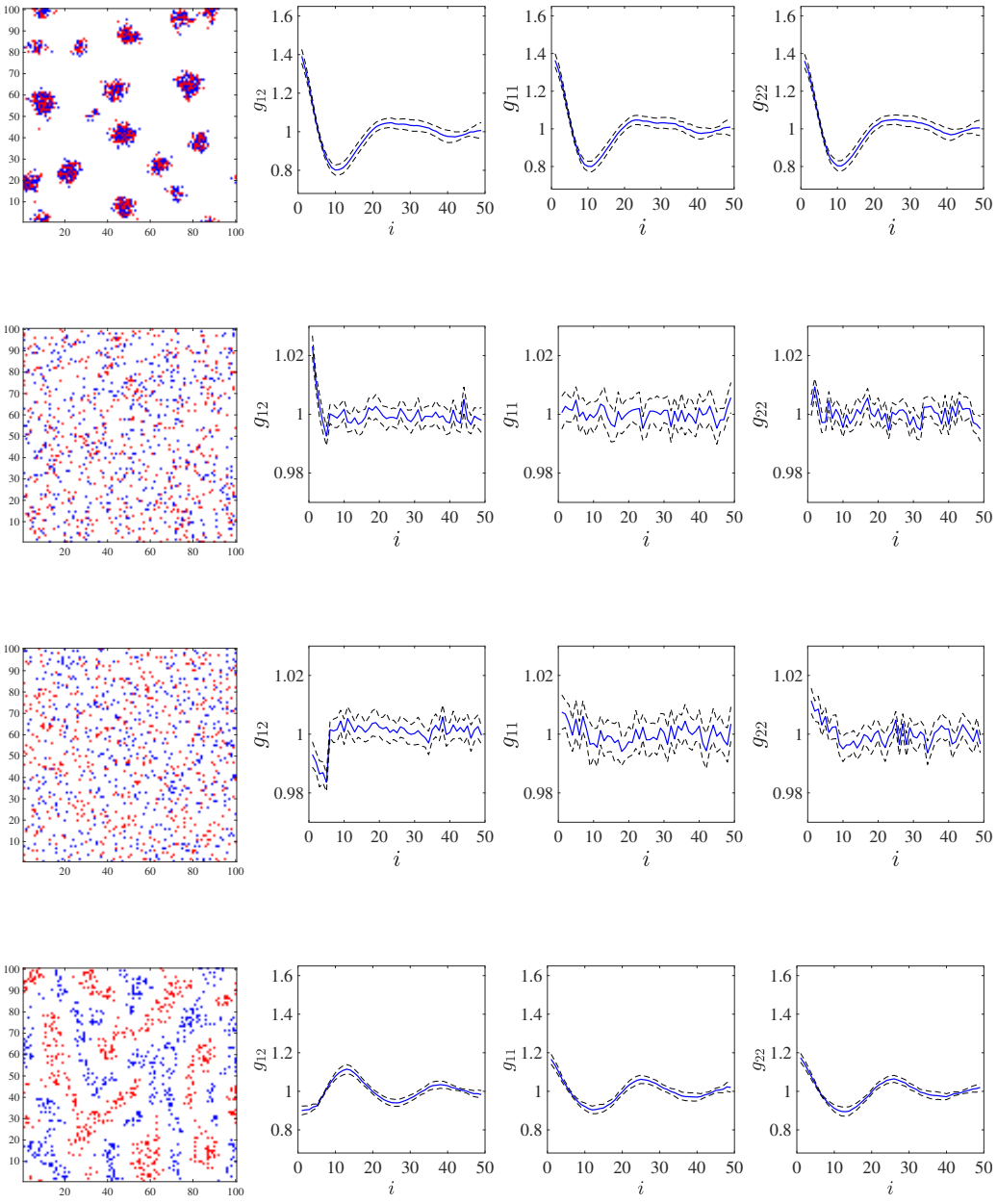


Figure 4.5: Mutually attractive or repulsive species with heterotypic interactions only, $a_{11} = a_{22} = r_{11} = r_{11} = 0$ and $N_1 = N_2 = 500$. First row: Strong heterotypic attraction, $a_{12} = a_{21} = 5$ and $r_{12} = r_{21} = 1$. Second and third row: Weak heterotypic attraction and repulsion. Second row: $a_{12} = a_{21} = 1.2$ and $r_{12} = r_{21} = 1$. Third row: $a_{12} = a_{21} = 1$ and $r_{12} = r_{21} = 1.2$. Fourth row: Strong heterotypic repulsion, $a_{12} = a_{21} = 1$ and $r_{12} = r_{21} = 5$.

that produce visible spatial clustering. We now consider the case of weak attraction and weak repulsion that generate patterns with no visible spatial structure (see second and third row of Fig. 4.5). In the case of weak heterotypic interactions, the deviations of the maximum and minimum values of the PCFs from unity are at least an order of magnitude smaller than those for the case of strong heterotypic interactions, indicting that the patterns are close to CSR. However, even with a relatively small number of simulations ($N = 60$), the heterotypic PCFs correctly identify the weak, short scale behaviour. For example, in the second row of Fig. 4.5 the heterotypic attraction ($a_{12} = a_{21} = 1.2$) is slightly stronger than the heterotypic repulsion ($r_{12} = r_{21} = 1$) and we observe weak short scale heterotypic aggregation, similar to that found for strong attraction in the first row of Fig. 4.5.

These few examples clearly illustrate the potential usefulness of the PCF for identifying the presence of attractive (or repulsive) heterotypic interactions, even when no pattern is distinguishable by eye. However, they also sound a note of caution in using the PCF to infer the mechanism of pattern formation. The fact that the homotypic PCFs are above unity at short distances in the first and fourth rows of Fig. 4.5 is not indicative of homotypic attraction; it occurs because agents of the same type are forced together by strong attractive, or repulsive heterotypic interactions. Thus, if we wish to try to make inferences about the mechanism of pattern formation from the PCFs, we need to consider the heterotypic and homotypic PCFs simultaneously, and may require information about the homotypic interactions (*e.g.* from an experiment where only one species is present).

Homotypic interactions only

We now investigate the patterns which can occur with only homotypic interactions between the two species (so $a_{12} = a_{21} = r_{12} = r_{21} = 0$), with the same strengths of attraction and repulsion for each species $a_{11} = a_{22}$ and $r_{11} = r_{22}$. The results for strong homotypic attraction and strong homotypic repulsion are presented in the first and second row of Fig. 4.6.

For strong attraction, we observe short scale heterotypic segregation, short scale homotypic aggregation, intermediate scale heterotypic aggregation and intermediate scale homotypic segregation (first row of Fig. 4.6). We remark that the short scale heterotypic segregation is not indicative of heterotypic repulsion (as $r_{12} = r_{21} = 0$), and is instead the result of the strong homotypic attraction which pulls species of the same type together. We also see that the heterotypic PCF is quite similar to that of the system with only heterotypic repulsion (see fourth row of Fig. 4.5). However, at intermediate and large distances, the homotypic PCFs in the system with only homotypic

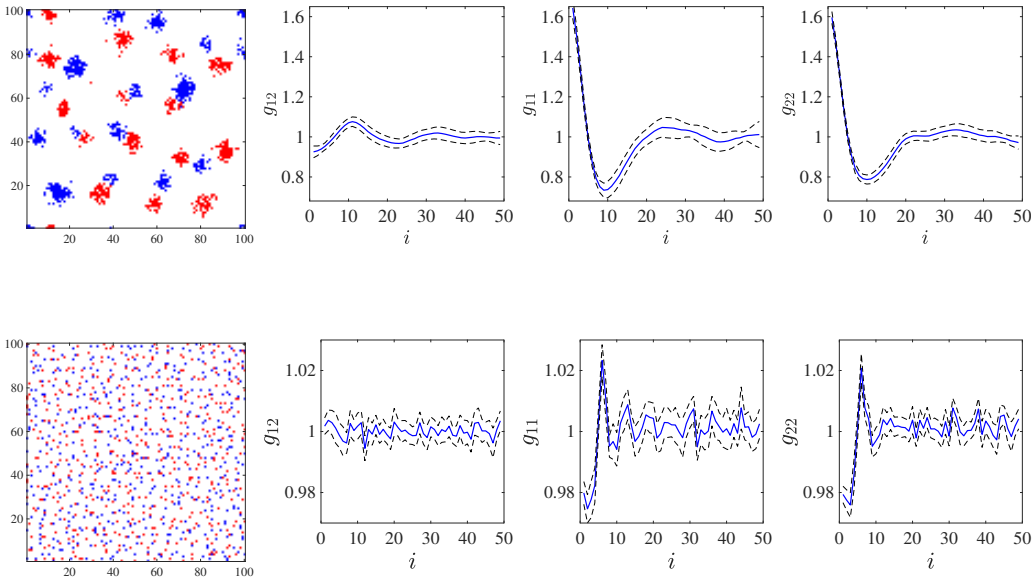


Figure 4.6: Mutually attractive or repulsive species with homotypic interactions only, $a_{12} = a_{21} = r_{21} = r_{12} = 0$ and $N_1 = N_2 = 500$. First row: Strong homotypic attraction, $a_{11} = a_{22} = 5$ and $r_{11} = r_{22} = 1$. Second row: Strong homotypic repulsion, $a_{11} = a_{22} = 1$ and $r_{11} = r_{22} = 5$.

interactions are quite different from those of the system with only heterotypic interactions. This is because the compact clusters are distributed uniformly at random in the first row of Fig. 4.6 (as shown by the fact that the PCFs are close to unity at large distances), whereas the elongated clusters are distributed in a segregated pattern in the fourth row of Fig. 4.5 (as the PCFs show noticeable fluctuations about unity at large distances). Hence, as mentioned earlier, this warrants the use of the homotypic PCFs in identifying the dominant underlying interactions along with the heterotypic PCF.

Strong repulsion produces patterns with almost no visible spatial structure (second row of Fig. 4.6). The heterotypic PCF shows that the interspecies distribution is close to CSR, whilst the homotypic PCFs reveal weak short scale homotypic segregation, due to the dominance of homotypic repulsion.

Heterotypic and homotypic interactions

We now consider an example of when both homotypic and heterotypic interactions are present in the system, and focus our attention on weak homotypic

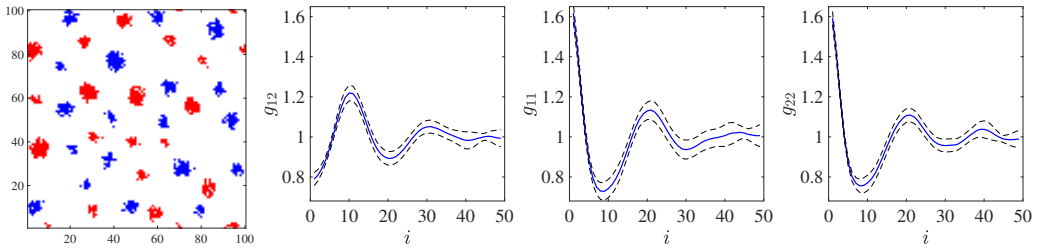


Figure 4.7: Mutually attractive or repulsive species with weak homotypic attraction and strong heterotypic repulsion, $a_{11} = a_{22} = 1$, $r_{12} = r_{21} = 5$, $a_{12} = a_{21} = r_{11} = r_{22} = 0$ and $N_1 = N_2 = 500$.

attraction and strong heterotypic repulsion (Fig. 4.7). When compared to the elongated and segregated clusters in the fourth row of Fig. 4.5 (*i.e.* , strong heterotypic repulsion only), we see that the effect of weak homotypic attraction forms more compact clusters in Fig. 4.7. At a glance, these compact clusters appear similar to those in the first row of Fig. 4.6 (*i.e.* strong homotypic attraction only). However, upon closer inspection, the compact clusters in Fig. 4.7 are not distributed uniformly at random throughout the domain, as is the case for those in the first row of Fig. 4.6. The PCF analysis is consistent with these observations, and further demonstrates the usefulness of the PCFs to quantify multi-species spatial patterns.

4.3.2 Attractive-repulsive inter-species interactions

We now turn to the case where the heterotypic interactions produce opposite effects on the two species - *e.g.*, species 1 is attracted to species 2, whilst species 2 is repelled by species 1. This type of interaction is termed *run-and-chase*. We consider two examples of run-and-chase interactions that can produce visible clusters of one of the species (*e.g.* prey) and two-species striped patterns (*e.g.* zebrafish skin).

4.3.3 Predator-prey system

A common situation where the run-and-chase mechanism pertains is a predator-prey system where predators (species 1) are attracted to prey (species 2), which try to flee from them (Fig. 4.1b). Various swarming patterns may emerge [26]. Living in a swarm provides a variety of advantages for animals, such as enhanced defensive capabilities, vigilance, foraging, mating success,

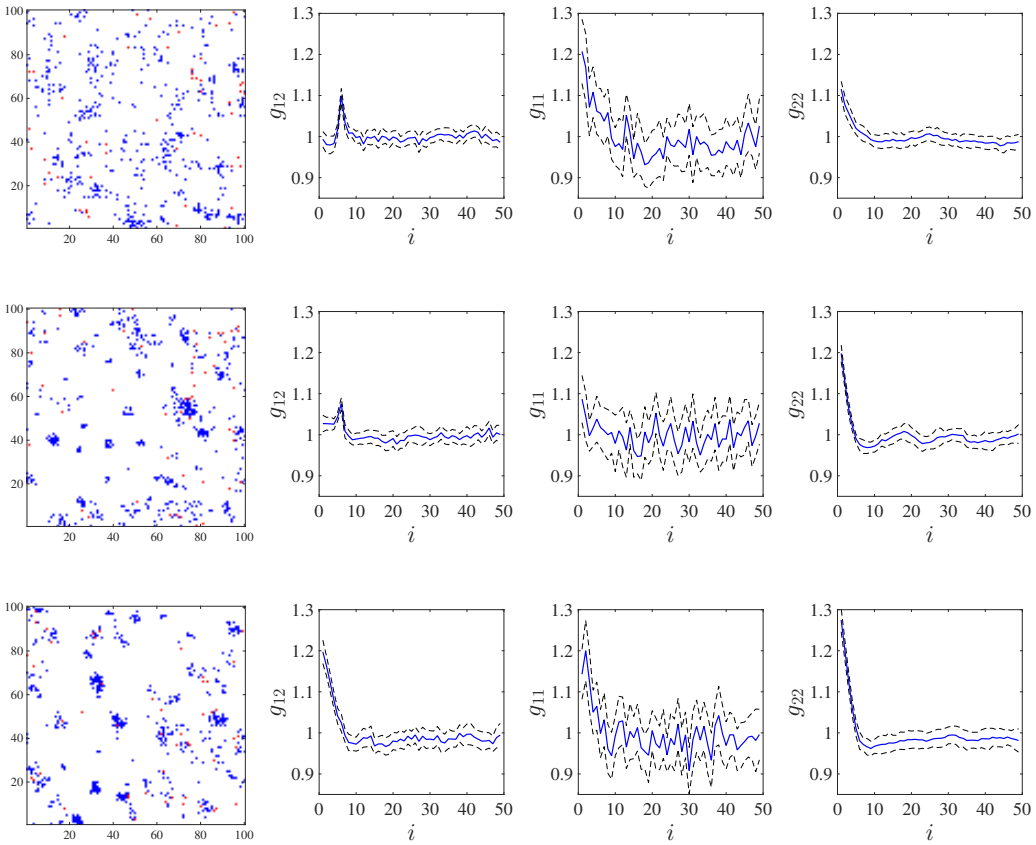


Figure 4.8: Strong attractive-repulsive interactions with no homotypic repulsion and no predator homotypic attraction, $a_{12} = r_{21} = 50$, $r_{11} = r_{22} = r_{12} = a_{21} = a_{11} = 0$, $N_1 = 50$ and $N_2 = 500$. Predators (red) are attracted to preys (blue). The effect of varying the prey homotypic attraction is shown. First row, $a_{22} = 0$ Second row, $a_{22} = 5$. Third row, $a_{22} = 10$.

etc. [91, 131], but it can as well have detrimental effects, which are discussed in the following.

The dynamical behaviour of predator-prey systems have been extensively studied [3, 26, 185]. However, the analysis of spatial patterns generated with discrete models of this system has received much less attention. Therefore, we set the parameters of our ABM to mimic a predator-prey system and apply our PCF to the resulting patterns. Obviously, our model does not incorporate all of the underlying mechanisms found in specific predator-prey systems. The aim here then is to investigate the generic behaviour of attractive / repulsive interactions found in all predator-prey systems.

We assume that a smaller group of predators chase a larger group of prey, with strong attractive (predator)-repulsive (prey) heterotypic interactions. We also assume homotypic attraction between prey, to provide group cohesion. All other potential interactions are neglected. We choose much higher heterotypic strengths compared to homotypic ones, based on the idea that for the prey avoiding the predators is of greater importance than staying near to other members of the same species.

The assumption of homotypic attraction between the fleeing prey when attacked by predators is often justified by the 'selfish herd' hypothesis [113, 115]. The hypothesis states that each prey tries to distance itself from the predators by attempting to remain close to the centre of the herd, which would make other individuals more exposed. We use our PCF to investigate the influence of the prey's homotypic attraction on the patterns produced. The results can then be used to discuss the extent to which different behaviours contribute to the success / failure of the prey / predators [126].

The effect of increasing the prey homotypic attraction is shown in Fig. 4.8 (top-bottom rows). Although the homotypic PCFs are not remarkably different in different experiments, the effect of increasing the prey's homotypic attraction produces a qualitative and quantitative change in the heterotypic PCFs. As shown in the first row of Fig. 4.8, for $a_{22} = 0$, the maxima of the heterotypic PCF at $i \approx 5$ indicates that large number of prey and predators are 5 sites apart, which can be interpreted as the prey successfully keeping their distance from the predators. But as the homotypic attraction of the prey increases, this maximum at $g_{12}(5)$ decays, while $g_{12}(1)$ becomes dominant (see second row of Fig. 4.8). Fig. 4.8 illustrates the dominance of $g_{12}(1)$ for $a_{22} = 10$. This indicates proximity of predators and prey, which is likely to be associated with success for the predators.

These results support the notion that living in a swarm can increase the risk of predation for the prey [131]. Importantly, the detrimental effect of homotypic attraction in prey is quantified by the computed PCFs. Our results suggest that homotypic attraction may save individuals by placing

others closer to predators, but it works at the expense of increased overall predation. This is shown by the increase in $g_{12}(1)$ with increasing the strength of homotypic attraction, which indicates the increased averaged proximity of predators to the preys.

The quantification approach taken here can be readily applied to more sophisticated predator-prey models that incorporate greater levels of biological detail. We believe this may help to produce insights into the optimum strategies for prey, which minimise the risk of predation.

4.3.4 Zebrafish stripes

The zebrafish is a popular model organism for the study of pattern formation in animals [71, 78, 79, 107, 133]. In particular, we consider the development of the striped pattern on its skin, as is illustrated in Fig. 4.1c. The two species of interest here are melanophores and xanthophores, two types of pigment cell (dark and light coloured, respectively) which are involved in creating the skin patterns (for simplicity, other cell types such as iridophores [53] are neglected here).

Pattern formation in zebrafish has been analysed using various mathematical models, both continuum [54, 122, 183] and discrete [19, 112, 177]. Recent studies suggest that *run-and-chase* is the major underlying mechanism for stripe formation [78, 184]. It has been observed that melanophores migrate away from the chasing xanthophores when the melanophores contact the dendrites of the xanthophores [78]. We implement a generic run-and-chase scenario in our ABM, and investigate possible patterns and their spatial characteristics. In this work, we concentrate on the effect of cells interactions on stripe formation. Thus, other factors that may impact on this process, such as domain growth, are not considered here (see *e.g.* [177] for a more complex model of stripe formation in zebrafish).

The different patterns generated by varying the heterotypic ranges of attraction and repulsion are illustrated by Fig. 4.9a where the homotypic interactions are deactivated ($a_{11} = a_{22} = r_{11} = r_{22} = 0$). In these simulations, the escaping agents of type 2 (blue) are being chased by agents of type 1 (red). We vary the range that chasers are attracted to runners, α_{12}, and the range that runners are repulsed by chasers, γ_{21}. Fig. 4.9b shows the associated heterotypic PCFs, g_{12}. As Fig. 4.9a indicates, stripes (which can randomly be horizontal or vertical) are formed when $\alpha_{12} \in \{10, 15\}$ and $\gamma_{21} \in \{1, 5\}$, without any homotypic interactions. This implies that a simple run-and-chase system (without homotypic interactions) is able to produce a striped pattern. The highly deviated g_{12} for $\alpha_{12} \in \{10, 15\}$ and $\gamma_{21} \in \{1, 5\}$ confirms that these generated patterns are quite distinct from CSR. In addition, the

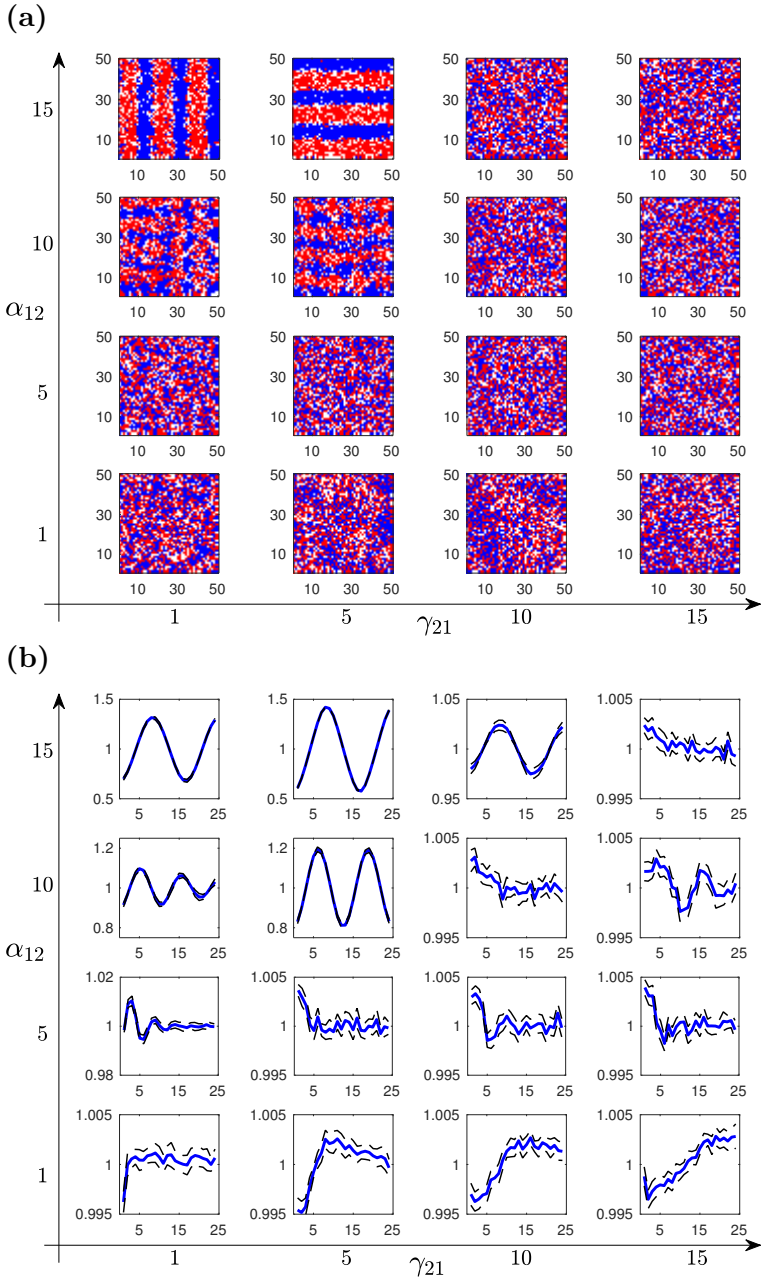


Figure 4.9: Attractive-repulsive interactions with no homotypic interactions, $a_{12} = r_{21} = 1$, $a_{21} = r_{12} = 0$, $a_{11} = a_{22} = r_{11} = r_{22} = 0$, and $N_1 = N_2 = 1000$. The effect of varying the heterotypic ranges attraction / repulsion is shown, $\alpha_{21} = \gamma_{21} \in \{1, 5, 10, 15\}$. (a) Typical spatial patterns. (b) Average heterotypic PCF, g_{12}.

values of g_{12} below unity at short distances demonstrate the segregation of the two types.

The emergence of only horizontal / vertical stripes implies that there is an inherent anisotropy in the model, as otherwise, diagonal stripes would have been likely to emerge. This may be due to our considered method of implementing the cell motion, *i.e.* the cells can either move right, left, up or down during a timestep, rather than moving diagonally. However, to be able to exactly determine the source of this anisotropy, we require to change the model and carry out further simulations accordingly. This is left as future work.

The PCFs for $\alpha_{12} \in \{10, 15\}$ and $\gamma_{21} \in \{1, 5\}$ show that g_{12} magnitude is higher in $\gamma_{21} = 5$ than that in $\gamma_{21} = 1$. This means that increasing γ_{21} can give more pronounced stripes. However, g_{12} for $\alpha_{12} \in \{10, 15\}$ and $\gamma_{21} \in \{10, 15\}$ lie very close to unity, implying that the stripes are significantly distorted. To sum up, we deduce that increasing γ_{21} can help to form more distinct stripes, provided it is kept significantly lower than α_{12}. Generally, this is equivalent to short range heterotypic repulsion and long range of heterotypic attraction, which is consistent with what is observed in *in vitro* experiments [78]. Moreover, the results indicate that the number of stripes, which is related to the number of maxima / minima of g_{12}, depends only on α_{12} in the studied cases here. Hence, it appears that repulsion can disrupt stripe formation, but it does not affect the number of stripes.

The patterns and PCFs in Figs. 4.9a and 4.9b show that stripes are not formed when $\alpha_{12} \leq \gamma_{12}$. However, as noted earlier, when the range of attraction is larger than the range of repulsion the model is capable of producing the stripes without the need to include homotypic interactions. Painter *et al.* (2015) studied the run-and-chase mechanism by means of a one-dimensional continuous model, assuming equal cell-scale ranges of heterotypic interactions between the cells. Based on their results, they suggest that homotypic attraction is necessary for the production and maintenance of the stripes. Woolley *et al.* (2014) posit a similar argument that run-and-chase does not lead to formation of persistent stripes, using a lattice-free individual-based model. Our results confirm that when the heterotypic interactions have equal ranges, stripe formation does not occur. However, if the range of attraction is longer than that of repulsion, stripes can be produced without the need for homotypic interactions.

4.4 Discussion

In this chapter, we have developed a generic ABM for two interacting species. The interactions between individuals can be attractive or repulsive in any combination, and the ranges of interactions can be varied for each type of interaction. The model enables us to investigate how different inter-individual interactions can generate a variety of distinctive, large-scale patterns. However, the main novelty of our approach is to combine the agent-based modelling with a method of quantifying the resulting patterns, which provides a means by which model results and experimental observations can be compared, and the underlying inter-individual interactions identified. We have extended the periodic PCF introduced in [1] to allow us to quantify the multi-species spatial patterns produced by different combinations of homotypic and heterotypic interactions, and have shown how to interpret the PCF and obtain information about the underlying interactions.

We began by considering the case where the heterotypic interactions between the two species were either both attractive or both repulsive. For the case of strong heterotypic attraction (in the absence of homotypic interactions), compact aggregates form in which both types of individuals are intermixed, which is consistent with the results of [64], who considered interactions between two types of liver cell cultured *in vitro*. Conversely, when homotypic attraction dominates over heterotypic, compact aggregates form in which the two types are segregated. Interestingly, and perhaps less intuitively, we observe the formation of similar compact, segregated aggregates when the dominant interaction is heterotypic repulsion, combined with weaker homotypic attraction. This is because the strong repulsion 'sorts' the agents into regions where only one type is present, and the weak homotypic attraction then causes them to coalesce into clumps. Although both situations give rise to aggregates of the same composition, the difference in the interactions can be distinguished by our PCF. Similar to the results of Agnew *et al.* (2014), our PCF is also able to identify weak interactions, even if the resulting pattern is indistinguishable by eye from the CSR state.

We then turned our attention to the case of attractive-repulsive heterotypic interactions - the run-and-chase scenario. We considered two biological examples of this situation: predator-prey interactions, and melanophore-xanthophore interactions in zebrafish. In the first of these, we were able to demonstrate using the PCF that increasing the strength of homotypic attraction between the prey would lead to closer proximity of predators and prey on average, which would be to the advantage of the predators. We believe this is due to the fact that the homotypic attractions help to maintain a compact,

coherent group of prey, which is more strongly attractive to the predators than a more dispersed group would be. Similar to some of the earlier cases, this result is not obvious from visual comparison of the patterns (see Fig. 4.8), and so provides another example of the potential usefulness of the PCF in quantifying the spatial distributions of individuals.

Run-and-chase has also been proposed as the mechanism giving rise to stripe formation in zebrafish [78, 184], and we investigated this possibility using our model. We varied the ranges of attraction (xanthophores towards melanophores) and repulsion (melanophores towards xanthophores) to determine which values would lead to the initiation of stripes. Long range attraction and short range repulsion was found to be able to produce stripes. Importantly, we found that homotypic attraction is not necessary for stripe formation, in contrast to what has been suggested by Painter *et al.* (2015) and Woolley *et al.* (2014).

The ABM used here is an idealised, simple model, intended to cover a wide range of situations driven by homotypic-heterotypic attraction-repulsion forces. One can specialise this model for a specific system by including more underlying mechanisms. Some examples include incorporating lattice growth to mimic the growth in the size of the zebrafish during its development [177], or adding mechanisms, such as predator confusion [90], active defending [91], *etc.* to the predator-prey model to gain insight into how different swarming behaviours may be advantageous in these species.

Our PCF can also be used for the purpose of parameter estimation of models. Recent work on parameter inference for simpler (one population) ABMs suggest that it is a good candidate for being used as a summary statistic in approximate Bayesian computation methods [83, 144, 166]. However, we note that due to the higher complexity of our model (which includes attraction and repulsion strengths and lengthscales for each combination of interactions), extending the previous work will be a considerable undertaking.

Chapter 5

Non-local continuum approximations of agent-based models of interacting populations

5.1 Introduction

In the previous chapter, we studied pattern formation in two interacting populations using an agent-based model. By varying the homotypic and heterotypic interactions between cells, we simulated various mechanisms that influence pattern formation. We studied the *run-and-chase* mechanism, in particular, and examined the capability of this mechanism in influencing the stripes generation on the skin of zebrafish.

In this chapter, we take a more analytical approach in studying tissue pattern formation, by using a continuous approximation of an individual-based model. These models permit analytic methods, such as linear stability analysis, to be used, which helps us to give important insights into pattern formation. As a result, we can predict the macro-scale behaviour of interacting cells in a systematic way.

In contrast to the previous chapter, we focus purely on hypothesised run-and-chase interactions here. This mechanism has recently been recognised important in influencing cell sorting [140, 150, 162, 163], and formation of the stripes on the skin of zebrafish [78, 184]. The horizontal stripes on the skin of zebrafish are mainly comprised of two pigment cells: Melanophores (black cells in Fig. 5.1b) and xanthophores (yellow cells in Fig. 5.1b). Recent experiments suggest that melanophores migrate away from xanthophores,

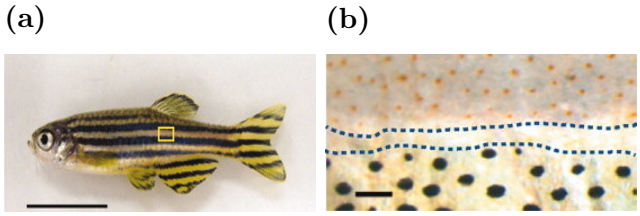


Figure 5.1: (a) Zebrafish (*Danio rerio*). Scale bar, 1 cm. (b) Enlarged view of the stripes. Scale bar, 100 μm. The black and yellow cells are melanophores and xanthophores respectively. The image is reprinted from Inaba *et al.* (2012) with permission from Science.

when the melanophores are in contact with dendrites of xanthophores (membrane to membrane contact) [78]. (The effect of other cell types, including iridophores [53], on stripe formation is ignored here for simplicity.)

Collective behaviour of individuals is mathematically well studied in animals [18, 28, 175] and cells [64, 130], but run-and-chase has received less attention. A recent work by Evers *et al.* (2016) studies self-organisation of two-species particles in a deterministic lattice-free system (a similar two-species model is [101]). In a similar work, Painter *et al.* (2015) introduces a non-local continuous model, in which they investigate the pigmentation patterns in zebrafish as a case study. Volkening and Sandstede (2015) study stripes formation exclusively in zebrafish, using an individual-based lattice-free model with a growing domain.

Stochastic individual-based models enable us to investigate the influences of cell-scale interactions on the emergent patterns. In addition, taking the inherent stochasticites into account, produces more realistic results, comparable to biological experiments. However, utilizing these models may present a challenge of computationally intensive simulations, where parameter sweeping approaches to study pattern formation can become intractable. To overcome this hurdle, continuous approximations of the stochastic models can be employed, which are amenable to analytical analyses, such as linear analysis, weakly nonlinear analysis *etc.* In this study, we seek to develop a non-local continuous model describing the population-level average behaviours of the stochastic model.

Non-local mathematical models are suitable candidates to model run-and-chase. They have widely been of interest, since different types of interactions can be readily embedded in them, and various analyses can be applied upon them [23, 49, 76, 167]. An advantage of the non-local models is that both mechanical and chemical interactions can be implemented by choosing suit-

able interaction kernels [95, 110]. Deterministic non-local models are often represented as an integro-partial differential equations of the form of Eqn. (1.3.1).

In this chapter, the run-and-chase mechanism is investigated using a stochastic model [75, 158], which is simulated on a one-dimensional lattice using direct Gillespie algorithm [58]; the methods are extended to two-dimensional lattices, but due to the high computational cost of the simulations, we focus on a one-dimensional lattice. Then, following the work of Hackett-Jones *et al.* (2012), we use a mean-field type approximation to derive a continuous non-local model in the form of Eqn. (1.3.1) (the equivalent off-lattice continuous model approximations can be found in [16, 67, 108]). By applying linear stability analysis to the continuous model, we try to predict the behaviour of the stochastic model. Particularly, we study formation of striped patterns, similar to the ones produced on the skin of zebrafish (see Fig. 5.1a).

5.2 Stochastic Model

We use a lattice-based stochastic model that has two cell types: (1) chaser cells which pursue the (2) escaping runner cells. The model is based on the one in Chapter 4, however, we relax the constraint that only one cell can occupy each lattice site (*i.e.* we neglect volume exclusion effects). Clearly, considering volume exclusion would make the model more biologically realistic, but it would make the subsequent analysis more complicated.

We present the case of a one-dimensional lattice in this chapter, and restrict our analyses and simulations to the one-dimensional model, because simulations of the two-dimensional model are computationally expensive. However, the method is easily extendable to higher dimensions, as explained in Appendix B.3 for a 2D domain.

The cells are distributed on a discrete lattice of size W, which has lattice sites of equal size δ. They change their position in the lattice according to motion events, which depend on environmental factors. Cell proliferation and death are neglected. Each cell can move to the right or left adjacent sites or remain at its current site. The run-and-chase mechanism is implemented by inducing attractive-repulsive forces between the cells, so that the chasers are attracted towards the runners, which are repelled by the chasers. The method is essentially the same for both cell types, thus, for brevity we develop the formulae for the chasers and then adapt them for the runners.

We define the number of chaser and runner cells at site i at time t by $U_i(t)$ and $V_i(t)$ respectively for $i = 1, \ldots, W$ and $t \in \mathbb{R}^+$. We assume that

the rates of stepping right and left depend on the number of neighbouring cells within a range of interaction. For the sake of simplicity, and following the work of previous chapters, we do not take into account the effect of the distance between agents as long as they are within the range of interaction. (However, there exist other alternatives in the literature, which incorporate this effect, see *e.g.* [110]). Based on this, the rates of moving right, $a_i^+(t)$, and left, $a_i^-(t)$, at time t are defined by

$$a_i^+(t) = d_u + \frac{A}{N_v} \sum_{a=1}^{\alpha} V_{i+a}(t), \tag{5.2.1a}$$

$$a_i^-(t) = d_u + \frac{A}{N_v} \sum_{a=-\alpha}^{-1} V_{i+a}(t). \tag{5.2.1b}$$

where V_i is the number of cells in site i, d_u is an unbiased motion constant, N_v is the total number of runner cells on the lattice, α is the range of attraction ($1 \leq \alpha \leq W/2$), and A is the strength of attraction. The lattice is chosen to be periodic, similar to the lattices used in the previous chapters. Thus, we define

$$V_j = \begin{cases} V_{\{j \bmod W\}} & \text{if} \quad j \in \{-W/2+1, \ldots, 3W/2\} \backslash \{0, W\}, \\ V_W & \text{if} \quad j \in \{0, W\}. \end{cases}$$

Note that this periodic definition holds for all of the quantities that depend on position, *i.e.* the rates and the number of cells. Thus, in simulating the model, if a cell exits the lattice from one side, it reappears on the opposite side.

Similarly, the rates of stepping right or left can be defined for the runners. Taking into account that the runners move away from the chasers, we define the rates by

$$r_i^+(t) = d_v + \frac{R}{N_u} \sum_{a=-\gamma}^{-1} U_{i+a}(t),$$

$$r_i^-(t) = d_v + \frac{R}{N_u} \sum_{a=1}^{\gamma} U_{i+a}(t),$$

where R and γ are the strength and range of repulsion respectively ($1 \leq \gamma \leq W/2$), and N_u is the total number of chaser cells on the lattice.

We assume that right and left movements occur according to a Poisson process with intensities of $a_i^+(t)$ and $a_i^-(t)$ respectively. Thus, the probability

that one right step occurs during an infinitesimal time span, $\mathrm{d}t$, for a chaser cell in a site i at a time t is given by

$$(a_i^+(t)\mathrm{d}t)^1 \frac{e^{-a_i^+(t)\mathrm{d}t}}{1!} \approx a_i^+(t)\mathrm{d}t,$$

when $a_i^+(t)\mathrm{d}t << 1$. Similar conditions hold for the rate of stepping left.

When there are $U_i(t)$ cells in site i at time t, the number of right-stepping events that take place during $\mathrm{d}t$ has a binomial distribution. Thus, the probability that exactly one right step occurs is

$$\binom{U_i(t)}{1}(a_i^+(t)\mathrm{d}t)^1(1-a_i^+(t)\mathrm{d}t)^{U_i(t)-1} \approx U_i(t)a_i^+(t)\mathrm{d}t.$$

Similarly, the probability of occurring one left step in site i is $U_i(t)a_i^-(t)\mathrm{d}t$.

5.2.1 Simulation algorithms

We can simulate the stochastic model using the fixed time-step algorithm introduced in Chapter 3 and 4, which is summarised as follows. We choose an small time-interval $\mathrm{d}t$, and generate a random number s from a uniform distribution over $[0, 1]$ for each cell. Then, we determine the behaviour of a cell at site i during the time-step $\mathrm{d}t$ as follows. If $s \in [0, a_i^+\mathrm{d}t]$, the cell steps right to site $i+1$, if $s \in [a_i^+\mathrm{d}t, a_i^+\mathrm{d}t + a_i^-\mathrm{d}t]$ it steps left to site $i-1$, and if $s \in [a_i^+\mathrm{d}t + a_i^-\mathrm{d}t, 1]$ it stays at site i [1, 30]. Note that this event selection process can be run over the sites (instead of cells), by multiplying the probabilities by U_i, *e.g.* the first interval would be $[0, U_i a_i^+\mathrm{d}t]$ and so on.

However, in §5.2 we assumed that $a_i^+\mathrm{d}t$ and $a_i^-\mathrm{d}t$ are small. This requires a small time-step, which can make the simulations of the model using the algorithm computationally expensive. However, because $1 - a_i^+\mathrm{d}t - a_i^-\mathrm{d}t$ is large (as is $1 - U_i a_i^+\mathrm{d}t - U_i a_i^-\mathrm{d}t$), no movement occurs in most of the time steps [45]. Importantly, defining $a_i^+\mathrm{d}t$ and $a_i^-\mathrm{d}t$ using heaviside functions (*i.e.* $a_i^+\mathrm{d}t = H(a_i^+\mathrm{d}t - a_i^-\mathrm{d}t)$ and $a_i^-\mathrm{d}t = H(a_i^-\mathrm{d}t - a_i^+\mathrm{d}t)$, where H is the Heaviside function) to speed up the simulations, as done by Hackett-Jones *et al.* (2012), can also violate the foregoing assumptions made to obtain the master equation (explained in Appendix B.1).

The direct Gillespie algorithm helps to overcome this problem by estimating the time of the next event that takes place [57, 58]. Essentially, this algorithm skips over the time-steps that cells stay at the original site. This algorithm significantly speeds up the simulations and permits the simulation of a stochastic system with complete accuracy, *i.e.* no approximations are considered. The simulations in the following sections are all generated using direct Gillespie algorithm.

However, since in the direct Gillespie algorithm only one of the cells (or sites) can undergo motion during a time-step, using this method for simulating a system can still be very time-consuming, particularly for finer lattices, when we keep the cell density constant. In order to overcome this issue, approximate simulation algorithms, such as τ-leaping Gillespie algorithm, can be used. This method is introduced in Appendix B.2, and, for the sake of illustration, results of a single simulation are shown.

5.3 Continuous approximation of the stochastic model

Before deriving the continuous approximation of our stochastic model, we must first determine the evolution of the distribution of cells (a continuous-time Markov chain). Since the probability of occurrence of a single event is small, we assume the probability of concurrent events is negligible. Additionally, we assume that the number of cells at each site is independent of that at other sites, known as the mean-field approximation [181]. This implies the independence of the rates $a_j^+(t), a_j^-(t)$ and the number of cells $U_j(t)$, as the rates depend only on the number of cells in the adjacent sites within the interaction range. Then, following the calculations detailed in Appendix B.1, we obtain the master equation describing the time-evolution of the probability of finding a specific number of cells in site i and time t. Then, using the master equation, we find the evolution of the average density of cells, *i.e.* $\langle u_i(t)\rangle = \langle U_i(t)\rangle/N_u$, over many realisations of the stochastic model, which is

$$\frac{\mathrm{d}\langle u_i\rangle}{\mathrm{d}t} = \langle a_{i-1}^+\rangle\langle u_{i-1}\rangle + \langle a_{i+1}^-\rangle\langle u_{i+1}\rangle - \langle a_i^+ + a_i^-\rangle\langle u_i\rangle. \tag{5.3.1}$$

A similar equation is found for the average density of the runners, $\langle v_i(t)\rangle = \langle V_i(t)\rangle/N_v$.

More accurate descriptions of the evolution of the distributions can be found using other closure techniques [33, 82, 153] instead of the mean-field approximation. However, since these approximations often include the dynamics of joint probabilities, undertaking analyses on the continuous approximation of such models would be extremely difficult. Hence, we do not pursue such extensions here

We now derive a continuous approximation of the Eqn. (5.3.1). Assuming that the number of cells is large, the discrete average density of cells at site

i, $\langle u_i(t)\rangle$, can be approximated by a continuous variable $u(x,t)$, where

$$\langle u_i(t)\rangle = \int_{x_i-\frac{\delta}{2}}^{x_i+\frac{\delta}{2}} u(x,t)\mathrm{d}x \approx \delta u(x_i,t), \tag{5.3.2}$$

and $x_i = i\delta$ (with δ assumed to be small). We then rename x_i to x and substitute the continuous variables for the discrete ones in Eqn. (5.3.1). In order to simplify the calculations, we shift the lattice to set the continuous domain over $[-X/2, X/2]$, where $X = \delta W$ is the size of the continuous domain. The continuous approximation of Eqn. (5.3.1) is then given by

$$\begin{aligned}\frac{\partial u(x,t)}{\partial t} &= a^+(x-\delta,t)u(x-\delta,t) + a^-(x+\delta,t)u(x+\delta,t)\\ &\quad -[a^+(x,t)+a^-(x,t)]u(x,t),\end{aligned} \tag{5.3.3}$$

where $a^+(x,t)$ and $a^-(x,t)$ are the continuous approximations of the right- and left- stepping rates (defined in Eqns. (5.2.1a) and (5.2.1b)), respectively. By Taylor series expansion of the terms in the RHS of Eqn. (5.3.3), we then find

$$\frac{\partial u(x,t)}{\partial t} = -\delta\frac{\partial u(x,t)[a^+(x,t)-a^-(x,t)]}{\partial x} + \delta^2 d_u\frac{\partial^2 u(x,t)}{\partial x^2} + \mathcal{O}(\delta^3),$$

where we neglect terms of order $\mathcal{O}(\delta^3)$ and higher. Then, following the work of Hackett-Jones *et al.* (2012), $a^+(x,t) - a^-(x,t)$ can be written in the following form:

$$a^+(x,t) - a^-(x,t) = K(x,l) \star v(x,t) = \int_\Omega K(x-s,l)v(s,t)\ \mathrm{d}s, \tag{5.3.4}$$

where $\Omega = [-X/2, X/2]$, and $K(x,l)$ over $x \in [-X/2, X/2]$ is given by

$$K(x,l) = \begin{cases} -1 & \text{for} \quad x \in [\delta/2, (l+0.5)\delta], \\ 1 & \text{for} \quad x \in [-(l+0.5)\delta, -\delta/2], \\ 0 & \text{otherwise,} \end{cases}$$

where l is the range of interaction, *i.e.* α or γ. By periodic extension of the kernel, similar to the models in Chapters 3 and 4, over $x \in [-X, X]$, we have $K(x \pm X, l) = K(x,l)$. Then the attraction and repulsion kernels are defined as $K_a(x) = A\delta K(x,\alpha)$ and $K_r(x) = -R\delta K(x,\gamma)$, respectively. Fig. 5.2 shows the attraction and repulsion kernels for arbitrarily selected ranges and strengths of interactions.

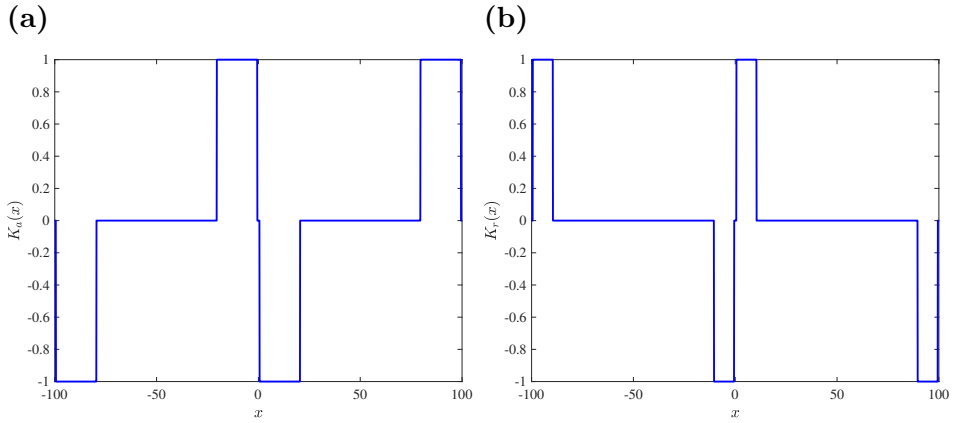


Figure 5.2: Interaction kernels for $X = 100, A = R = 1$ and $\delta = 1$. (a) Attraction kernel with the range of $\alpha = 20$. (b) Repulsion kernel with the range of $\gamma = 10$.

Finally, the PDEs approximating the stochastic system are found to be:

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\left[(K_a \star v)\ u\right] - D_u \frac{\partial^2 u}{\partial x^2} = 0, \tag{5.3.5a}$$

$$\frac{\partial v}{\partial t} + \frac{\partial}{\partial x}\left[(K_r \star u)\ v\right] - D_v \frac{\partial^2 v}{\partial x^2} = 0, \tag{5.3.5b}$$

where $D_u = \delta^2 d_u$, $D_v = \delta^2 d_v$.

5.3.1 Non-dimensionalisation

We now non-dimensionalise Eqns. (5.3.5a) and (5.3.5b) by introducing

$$u = u_0\tilde{u}, v = v_0\tilde{v}, x = L\tilde{x}, t = T\tilde{t}, \tag{5.3.6}$$

where tildes denote dimensionless variables, and we choose

$$u_0 = v_0 = 1/W\delta, L = \gamma\delta, T = W/R. \tag{5.3.7}$$

For non-dimensionalising the kernels, we define $K_a = K_a^*\tilde{K}_a$ and $K_r = K_r^*\tilde{K}_r$, where $K_a^* = 1/Tv_0$ and $K_r^* = 1/Tu_0$. Then, we obtain the non-dimensional model as

$$\frac{\partial \tilde{u}}{\partial \tilde{t}} + \tilde{Q}\frac{\partial}{\partial \tilde{x}}\left[\left(\tilde{K}(\tilde{x},\alpha) \star \tilde{v}\right)\tilde{u}\right] - \tilde{D}_u \frac{\partial^2 \tilde{u}}{\partial \tilde{x}^2} = 0,$$

$$\frac{\partial \tilde{v}}{\partial \tilde{t}} - \frac{\partial}{\partial \tilde{x}}\left[\left(\tilde{K}(\tilde{x},\gamma) \star \tilde{u}\right)\tilde{v}\right] - \tilde{D}_v \frac{\partial^2 \tilde{v}}{\partial \tilde{x}^2} = 0,$$

where $\tilde{Q} = Av_0/Ru_0$, $\tilde{D}_u = d_u W/R\gamma^2$ and $\tilde{D}_v = d_v W/R\gamma^2$, and the non-dimensional kernel for $\tilde{x} \in [-\tilde{X}/2, \tilde{X}/2]$ is given by

$$\tilde{K}(\tilde{x}, l) = \begin{cases} -1 & \text{for} \quad \tilde{x} \in [0.5/\gamma, l/\gamma + 0.5/\gamma], \\ 1 & \text{for} \quad \tilde{x} \in [-l/\gamma - 0.5/\gamma, -0.5/\gamma], \\ 0 & \text{otherwise}, \end{cases} \tag{5.3.9}$$

which is then extended over $\tilde{x} \in [-\tilde{X}, \tilde{X}]$ to be periodic, *i.e.* $\tilde{K}(\tilde{x} \pm \tilde{X}, l) = \tilde{K}(\tilde{x}, l)$. For simplicity of notation, we then drop the tildes for all of the quantities:

$$\frac{\partial u}{\partial t} + Q\frac{\partial}{\partial x}\left[(K(x,\alpha) \star v)\, u\right] - D_u \frac{\partial^2 u}{\partial x^2} = 0, \tag{5.3.10a}$$

$$\frac{\partial v}{\partial t} - \frac{\partial}{\partial x}\left[(K(x,\gamma) \star u)\, v\right] - D_v \frac{\partial^2 v}{\partial x^2} = 0. \tag{5.3.10b}$$

Of note, continuum models of this form, using a variety of alternative kernels have been studied. These frequently incorporate a (exponential, linear, *etc.*) decay in the strength of interaction with distance [44, 109, 130]. For instance, Painter *et al.* (2015) studied linear stability of a non-local model for three kernels, and showed that similar stability regions are found for the different kernels. A significant difference between the kernels derived here and those used by Painter *et al.* (2015) is the 'offset' $(-0.5\gamma, 0.5\gamma)$, in Eqn. (5.3.9). This offset, the length of which is equal to a lattice site, emerges in the continuum model derivation due to the fact that in the discrete model, cells do not sense the number of cells at their current site. In the following, we discuss how this difference can significantly influence the stability of the system.

5.3.2 Numerical methods for solving the continuous model

COMSOL Multiphysics software was used to solve the continuous model, using a finite element method. For domain discretisation, we used different number of grids, and observed no significant difference in the results. For the time-stepping of the solver, a backwards differentiation formula (BDF) method is used.

5.4 Pursuit example - run-and-chase with two groups

We begin with a simple example, where the two cell types are initially aggregated in well separated clusters, as shown in Fig. 5.3a. The aim of this experiment is to check whether the proposed run-and-chase mechanism produces the expected pursuit behaviour. The selected parameter values of the stochastic model are $d_u = d_v = 1, A = R = 5, \alpha = \gamma = 30, W = 100$ and $N_u = N_v = 18$, which correspond to $Q = 1, D_u = D_v = 0.0222$ and $X = 100/30$ of the non-dimensional continuous model defined in Eqns. 5.3.10. In order to be able to compare the results of the stochastic model with the continuum approximations, we have plotted the results in terms of the dimensional variables. In addition, we scale the densities in the continuous model to obtain the number of cells at each site, using $U(x,t) = N_u u(x,t)$ and $V(x,t) = N_v v(x,t)$. This approach is taken for showing the solutions of the continuous model henceforth. In cases where we initially choose the non-dimensional parameters, *e.g.* in linear stability analysis, we use Eqns. (5.3.6) and (5.3.7) to find the corresponding parameters of the stochastic model and dimensional continuous model.

The bars in Fig. 5.3 represent the average number of cells over 500 simulations of the stochastic model, and the broken curves are obtained from the numerical solution of the dimensional continuous model, Eqns. (5.3.5). The red and blue colours correspond to the chaser and runner cells, respectively. As the figure illustrates, the initial accumulation of cells is dispersed at time $t = 3$ due to the unbiased random motion (equivalently, diffusion terms in the PDEs). The groups of runners and chasers move to the right, as the runners are migrating away from the pursuing chasers. Subsequently, at time $t = 9$, the cells have travelled further to the right end of the domain; they keep moving in the same direction, re-entering the domain at the far left due to the periodic lattice. Comparison of the results of the continuous and stochastic models show an excellent qualitative agreement, although quantitatively, the continuous model over-estimates the maxima of the amplitudes of the aggregations of the runners and chasers.

The results clearly demonstrate the expected pursuit behaviour in the model, where the cells move according to the attraction-repulsion forces described in §5.2. In the following sections, we discuss cases were patterns emerge from an initial perturbation about a spatially uniform steady state.

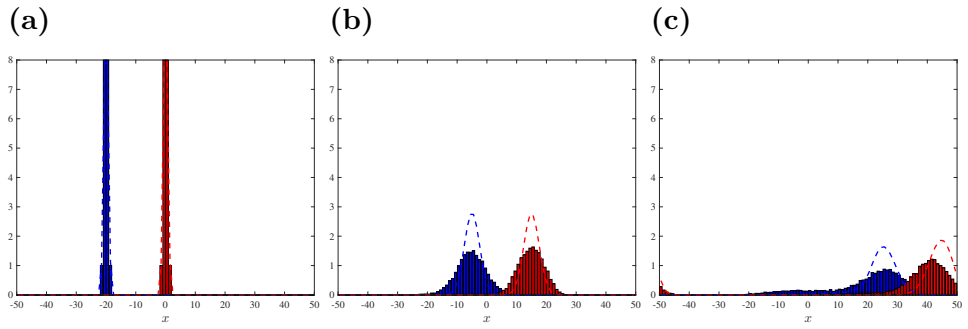


Figure 5.3: A pursuit run-and-chase example. (a) The chaser (blue) and runner (red) cells are initially accumulated around $x = -20$ and $x = 0$ respectively. The distribution of cells at (b) $t = 3$ and (c) $t = 9$. The bars show the average number of cells over 500 simulations and the broken curves are the solutions of the dimensional continuous model, Eqns. (5.3.5). The selected parameter values are $d_u = d_v = 1, A = R = 5, N_u = N_v = 18, W = 100, \alpha = \gamma = 30$ for the stochastic model, which correspond to $Q = 1, D_u = D_v = 0.0222, X = 100/30$ of the non-dimensional continuous model, Eqns. (5.3.10).

5.5 Linear stability analysis of the continuum model

Linear stability analysis provides a means to determine in which parameter regimes a system might produce a spatial pattern. From a biological perspective, it enables us to predict whether an initially unpatterned tissue may lead to a patterned one for a specific set of parameter values.

In this analysis, we evaluate how constituent modes (inherent components of a pattern that has certain wavenumbers) of a small perturbation about equilibrium evolve over time [48, 120, 130]. Therefore, we will be able to find the mode with the highest growth rate, which eventually becomes the dominant element of the pattern that emerges. However, linear stability analysis is only valid close to equilibrium; thus, it can not be used to interpret the behaviours of a system far from equilibrium.

To carry out the linear stability analysis, we first linearise the non-dimensional model about the equilibrium by substituting

$$u = 1 + \hat{u},$$
$$v = 1 + \hat{v},$$

into Eqn. (5.3.10) where $\hat{u}$ and $\hat{v}$ are the perturbations. Then, we neglect

the non-linear terms by assuming $\hat{u} << 1$ and $\hat{v} << 1$, which yields the linearised equations for the perturbations:

$$\frac{\partial \hat{u}}{\partial t} + Q\frac{\partial}{\partial x}\left[(K(x,\alpha) \star \hat{v})\right] - D_u\frac{\partial^2 \hat{u}}{\partial x^2} = 0,$$
$$\frac{\partial \hat{v}}{\partial t} - \frac{\partial}{\partial x}\left[(K(x,\gamma) \star \hat{u})\right] - D_v\frac{\partial^2 \hat{v}}{\partial x^2} = 0.$$

We seek the solutions of the form

$$\begin{pmatrix}\hat{u}\\ \hat{v}\end{pmatrix} \sim \begin{pmatrix}\epsilon_u\\ \epsilon_v\end{pmatrix} e^{iqx+\lambda t}, \tag{5.5.2}$$

where ϵ_u and ϵ_v are the amplitudes of the perturbations ($\epsilon_u^2 + \epsilon_v^2 = 1$), q is the wavenumber and λ is the growth rate. Substituting Eqn. (5.5.2) into the linearised model gives

$$\begin{pmatrix}\lambda + D_u q^2 & iQq\hat{K}(q,\alpha)\\ -iq\hat{K}(q,\gamma) & \lambda + D_v q^2\end{pmatrix}\begin{pmatrix}\epsilon_u\\ \epsilon_v\end{pmatrix} = 0. \tag{5.5.3}$$

For non-trivial solutions of this system, we require

$$\begin{vmatrix}\lambda + D_u q^2 & iQq\hat{K}(q,\alpha)\\ -iq\hat{K}(q,\gamma) & \lambda + D_v q^2\end{vmatrix} = 0,$$

which finally yields the dispersion relation

$$\lambda^2 + (D_u + D_v)q^2\lambda + D_u D_v q^4 - Q\hat{K}(q,\alpha)\hat{K}(q,\gamma)q^2 = 0, \tag{5.5.4}$$

where $\hat{K}(q,l)$ is the Fourier transform of the kernel, Eqn. (5.3.9), given by

$$\begin{aligned}\hat{K}(q,l) &= \mathscr{F}\{K\} = \int_{-\infty}^{\infty} K(x,l)e^{-iqx}\mathrm{d}x\\ &= \hat{K}(q,l) = \frac{-2i}{q}\left[\cos\left(\frac{l+0.5}{\gamma}q\right) - \cos\left(\frac{0.5}{\gamma}q\right)\right]\\ &= \frac{4i}{q}\sin(\frac{l+1}{2\gamma}q)\sin(\frac{l}{2\gamma}q).\end{aligned} \tag{5.5.5}$$

Hence, from Eqn. (5.5.4), λ is given by

$$\lambda = -0.5(D_u+D_v)q^2 \pm 0.5\sqrt{(D_u+D_v)^2q^4 - 4D_uD_vq^4 + 4Q\hat{K}(q,\alpha)\hat{K}(q,\gamma)q^2}. \tag{5.5.6}$$

We now impose the periodic boundary conditions. The domain of the non-dimensional continuous model is $x \in [-X/2, X/2]$. Therefore, the following conditions must be satisfied on the boundaries

$$\frac{\partial u}{\partial x}|_{x=\frac{-X}{2}} = \frac{\partial u}{\partial x}|_{x=\frac{X}{2}}, \quad \frac{\partial v}{\partial x}|_{x=\frac{X}{2}} = \frac{\partial v}{\partial x}|_{x=\frac{-X}{2}},$$

implying

$$\sin(\frac{qX}{2}) = 0,$$

that leads to

$$q = q(n) = \frac{2n\pi}{X} \quad \text{for} \quad n \in \mathbb{Z}.$$

Therefore, for each mode n we can find $\lambda(n)$ which determines the exponential growth of the corresponding term via Eqn. (5.5.2). In particular, we are interested to find the unstable modes, which have $\Re(\lambda(n)) > 0$, as these modes grow over time and, potentially, form the final patterns. Using Eqn. (5.5.6), the unstable modes are the ones that satisfy

$$q^4 + 16\frac{Q}{D_u D_v}\sin(\frac{\alpha+1}{2\gamma}q)\sin(\frac{\alpha}{2\gamma}q)\sin(\frac{\gamma+1}{2\gamma}q)\sin(\frac{1}{2}q) < 0,$$

which in terms of n is

$$n^4 + H(n) < 0, \tag{5.5.7}$$

where $H(n)$ is

$$H(n) = \frac{ARW^2}{d_u d_v \pi^4}\sin[\frac{(\alpha+1)\pi}{W}n]\sin[\frac{\alpha\pi}{W}n]\sin[\frac{(\gamma+1)\pi}{W}n]\sin[\frac{\gamma\pi}{W}n]. \tag{5.5.8}$$

From Eqn. (5.5.7), we see that a necessary condition for a mode, n, to be unstable is that $H(n)$ must be negative. In addition, $S = AR/d_u d_v$ must be large enough such that $|H(n)|$ would be greater than n^4, indicating that the strength of biased motion must be large enough to overcome the dissipative effect of the unbiased ones (*i.e.* cell diffusion).

Note that the $\alpha + 1$ and $\gamma + 1$ in the arguments of the sine functions in Eqn. (5.5.8) arise from the summation (integral) of the number (density) of neighbouring cells in the stochastic model (continuous model), see Eqns. (5.2.1a) and (5.3.4); also see the $0.5/\gamma$ terms in the kernels of Eqn. (5.3.9). This produces an interesting effect on the stability of the system. If we had not considered this term, similar to many studies using non-local continuous models [64, 130], $H(n)$ would be always positive, and the system would be stable for any values of α and γ. This implies that a system driven by the run-and-chase mechanism with such underlying interactions can not yield

any pattern from an initial perturbation. Hence, in those models, other interactions, *e.g.* homotypic attraction as suggested by Painter *et al.* (2015), are required to produce patterns. It also demonstrates the significant effect that considering cell area (volume) exclusion in a model can have on the results. Hence, this helps to explain the differences between the results we obtained in Chapter 4 and the literature.

We use the linear stability analysis to investigate how the varying parameters in the model affect stripe formation. We are interested in finding the most unstable mode m, in which $\Re(\lambda(n = m))$ has its largest value, $n \in \mathbb{Z}$. This mode will tend to dominate other modes and hence determine the shape of the spatial pattern. In addition, for patterns where the two cell types are segregated (as in zebrafish), we require the cell distributions to be anti-phase, meaning that $\epsilon_u(m)$ and $\epsilon_v(m)$ must have opposite signs, *i.e.* $\epsilon_u(m)/\epsilon_v(m) < 0$. From Eqn. (5.5.3) we find

$$\begin{aligned} \epsilon_u(q)/\epsilon_v(q) &= \frac{4Q\sin(\frac{\alpha+1}{2\gamma}q)\sin(\frac{\alpha}{2\gamma}q)}{\lambda + D_u q^2} \\ &= \frac{-(\lambda + D_v q^2)}{4\sin(\frac{\gamma+1}{2\gamma}q)\sin(\frac{1}{2}q)}. \end{aligned} \tag{5.5.9}$$

This equation shows that the positive quantities Q, D_u and D_v do not impact the sign of $\epsilon_u(q)/\epsilon_v(q)$, implying that α and γ are the only parameters that affect whether the of distributions of the runners and chasers are in-phase or anti-phase.

Fig. 5.4 shows the stability status of the system and the phase of the distributions of the species for the dominant unstable mode m, when varying the ranges of attraction, α, and repulsion, γ. The grey colour represent the pairs of α and γ for which the system is stable, *i.e.* no pattern emerges from initial perturbations. The rest of the pairs are unstable; the red and blue regions represent anti-phase (segregated clusters, associated with stripes) and in-phase patterns (intermixed cell aggregates), respectively. We also examine the effect of increasing $S = AR/d_u d_v$, the ratio of the strengths of biased motions to those of unbiased ones, on the results (Fig. 5.4a: $S = 0.25 \times 10^4$, Fig. 5.4b: $S = 1 \times 10^4$ and Fig. 5.4c: $S = 4 \times 10^4$). The figures show that the number of stable cases reduces by increasing S, as we expected considering Eqn. (5.5.7). Furthermore, there is a region around $\alpha = \gamma$ in each graph, throughout which the system is stable, and this is shrinking as S increases (from left to right in the figure). This is in consistent with the finding of Painter *et al.* (2015) that stripes cannot be formed when the ranges of attraction and repulsion are equal.

For the sake of illustration, we study the case where mode $m = 3$ is the

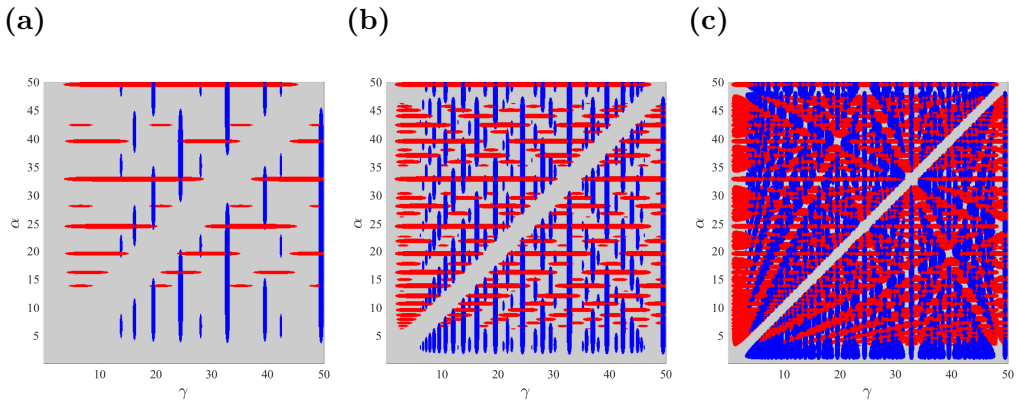


Figure 5.4: The stability state of the system when sweeping the ranges of attraction, α, and repulsion, γ, for $X = 100/16$. (a) $S = 0.25 \times 10^4$; (b) $S = 1 \times 10^4$; (c) $S = 4 \times 10^4$. The colour codes are as follows: grey: stable, blue: in-phase unstable and red: anti-phase unstable.

dominant unstable mode. We choose $\alpha = 33$ and $\gamma = 16$ which enables us to produce striped pattern (red regions in Fig. 5.4). We then assess the effect of the parameters Q and $D = D_v/D_u$ on the evolution of the dominant unstable mode m. Note that $Q = A/R$ (as we assumed $u_0 = v_0$) represents the ratio of the strengths of interactions and $D = d_v/d_u$ the ratio of diffusion constants. Fig. 5.5 illustrates how $m, \Re(\lambda(m)), |\epsilon_u(m)|/|\epsilon_v(m)|$ vary with D and Q. Figs. 5.5a and 5.5b show that m and $\Re(\lambda(m))$ essentially decrease with D, implying that setting $d_v > d_u$, *i.e.* runners being more diffusive than the chasers, leads to a more stable system and emergence of smaller unstable modes. Comparing the three curves in Figs. 5.5a and 5.5b (blue: $Q = 1$, black: $Q = 5$ and red: $Q = 10$) indicate that having a stronger strength of attraction than that of repulsion makes the system more unstable. This is demonstrated by having higher dominant unstable modes, m, and higher growth rates, $\Re(\lambda)$, when we increase Q.

Note that the emergence of large unstable wavenumbers (*e.g.* see the red and blue curves in Fig. 5.5a at small D) might be impeded in the discrete model due to the relatively small lattice size ($W = 100$). This is because large wavenumbers change significantly over one lattice site, which deteriorates the accuracy of continuous approximations. Therefore, the sets of parameter values that would be expected to provide a good agreement between the stochastic and continuous models are the ones that initiate the emergence of $m = 3$, as the other unstable modes (*e.g.* $m = 9, 15$) are relatively large given the coarseness of the lattice.

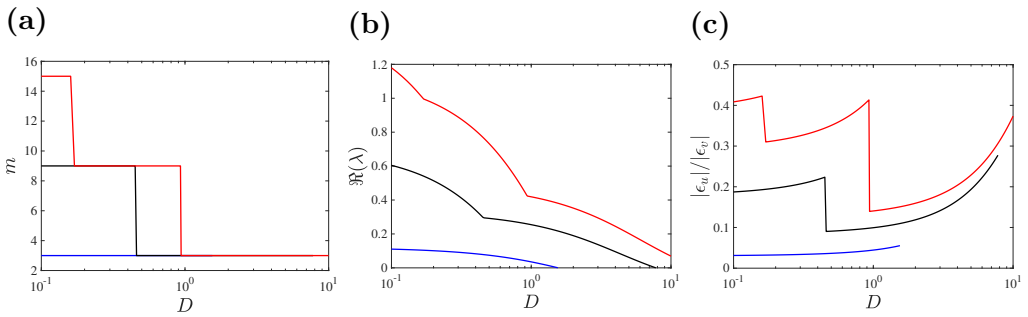


Figure 5.5: The behaviour of the system close to equilibrium when varying the ratio of unbiased diffusion constants, $D = d_v/d_u$, and the ratio of strengths of interactions, $Q = A/R$ (we assumed $u_0 = v_0$). (a) the dominant unstable mode (m); (b) the associated growth rate, $\Re(\lambda(m))$; (c) the ratio of the amplitudes of perturbations $|\epsilon_u(m)|/|\epsilon_v(m)|$. The colours correspond to blue: $Q = 0.5$, black: $Q = 5$ and red: $Q = 10$. Also, $X = 100/16, \alpha = 33$ and $\gamma = 16$ for all of the graphs.

In Fig. 5.5c, we examine the ratio of the amplitudes of the stripes, $|\epsilon_u(m)|/|\epsilon_v(m)|$, as it can significantly affect the generated patterns. The figure shows a non-monotonic behaviour of $|\epsilon_u(m)|/|\epsilon_v(m)|$ with respect to D (see Eqn. (5.5.9)). However, it essentially increases with Q. Note that for some values of Q and D we have significantly small $|\epsilon_u(m)|/|\epsilon_v(m)|$, *e.g.* see the blue curve in Fig. 5.5c. This implies that the emerging amplitude of the unstable mode in u would be significantly smaller than that for v when choosing the corresponding selected parameter values. Therefore, in order to have a reasonably large $|\epsilon_u(m)|/|\epsilon_v(m)|$ for a small mode like $m = 3$, we require large Q and D (for instance, $D > 1$ and $Q = 10$). However, this would place us in a parameter regime where we would expect the agreement between the stochastic and continuous models to be less strong. This is because the high strength of biased motion significantly violates the mean-field assumptions, which results in considerable inaccuracies in our approximations. We shall consider this point in more detail in the following section.

5.6 Simulations of stripe formation - comparison of stochastic and continuum models

We now study the formation of stripes in the stochastic model. We are interested in patterns that are similar to zebrafish stripes, where there are regions with an abundance of melanophores and few xanthophores, alternating with regions having the opposite proportions of cells. Emergence of this type of striped pattern corresponds to the instability of a mode, where the two distributions are anti-phase, *i.e.* negatively correlated. In particular, we consider the hypotheses posited by Painter *et al.* (2015), who suggested that without homotypic interactions, stripes would not emerge from a uniform distribution, and, furthermore, that initial striped patterns would not persist. Their results were obtained for the case where the ranges of attraction and repulsion are equal, and close to the lengthscale of a cell. In a similar work using an individual-based model, Woolley *et al.* (2014) also suggested that the run-and-chase does not yield persistent stripes.

Using the analyses represented in Figs. 5.4 and 5.5, we choose $Q = 1, D_u = D_v = 0.0156, \alpha = 33, \gamma = 16$ and $X = 100/16$ to initiate the emergence of mode $m = 3$ as the dominant mode, *i.e.* the mode with the highest growth rate; the corresponding parameter values of the stochastic model are $\alpha = 33, \gamma = 16, W = 100, d_u = d_v = 0.04, A = R = 1$ when choosing $T = 100, \delta = 1$ and $L = \gamma\delta = 16$. We choose $N_u = N_v = 10^4$ for all of the following results. For the selected set of parameter values, we depict $\Re(\lambda(n)), \Im(\lambda(n)), |\epsilon_u(n)|, |\epsilon_v(n)|$ in Fig. 5.6; the red cross in these figures indicate mode $m = 3$. In the following, we discuss how the quantities depicted in this figure help us to understand and justify the behaviour of the stochastic model.

We begin with the initial condition where the two cell types are already distributed in the striped pattern (anti-phase distribution). As Fig. 5.7a illustrates, the initial distribution is the perturbation of mode 3 with amplitude of 10 about 100 cells. The red and blue curves correspond to the number of runners, V, and chasers, U, respectively. The solid curves in this figure denote the average results of the stochastic model over 500 simulations, *i.e.* $\langle U \rangle$ and $\langle V \rangle$. The broken curves denote the numerical solution of the continuous model, described by Eqns. (5.3.5). Figs. 5.7b-5.7d show that the amplitude of the variations in $\langle V \rangle$ grows with time, unlike $\langle U \rangle$ which initially decays until $t \approx 25$ and steadies afterwards. As can be seen in the Fig. 5.7d, the amplitude of $\langle U \rangle$ has become significantly smaller than that

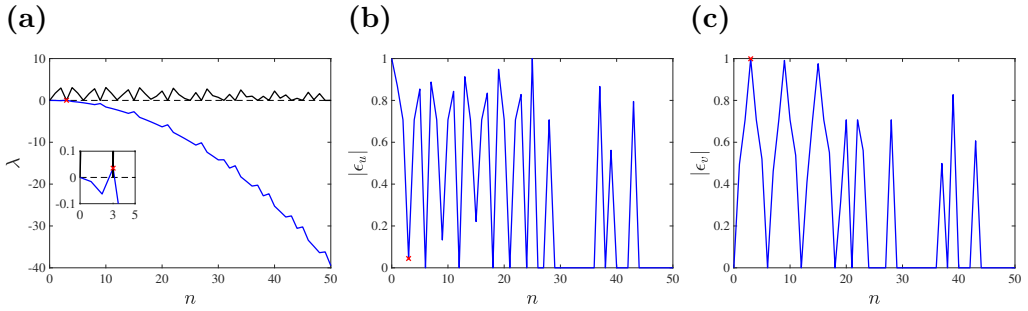


Figure 5.6: Evolution of constituent modes of a perturbation about equilibrium, studied by (a) $\Re(\lambda)$ (blue) and $\Im(\lambda)$ (black), (b) $|\epsilon_u|$ and (c) $|\epsilon_v|$. The red cross denotes the dominant unstable mode ($\Re(\lambda) > 0$), which is $m = 3$ here for the selected values of the parameters: $X = 100/16, Q = 1, D_u = D_v = 0.0156, \alpha = 33$ and $\gamma = 16$.

of $\langle V \rangle$ at time $t = 100$. This is in consistent with the prediction of the linear stability analysis for the continuous model that the ratio of the emergent amplitude of u to that of v is small for the chosen parameter values ($|\epsilon_u(3)|/|\epsilon_v(3)| \approx 0.04$). As a result, it takes a longer time for $\langle U \rangle$, compared to $\langle V \rangle$, to become noticeably large.

Comparing the results of the stochastic model with the continuous model indicates a good agreement between them until $t = 50$, but as Fig. 5.7d shows, there is a significant difference between the results at $t = 100$. This is due to the growing establishment of the striped pattern for $t > 50$, which consequently increases the dependence of the number of cells at adjacent sites. Hence, as time goes by, the validity of the mean-field assumption is weakening and the disagreement between the results of the models is intensifying. The following results demonstrate a similar disagreement at longer times between the stochastic and continuous models.

Of note, the imaginary part of λ for the dominant unstable mode is quite small, such that we do not observe oscillatory solution, travelling wave, *etc.* in both the stochastic model and its continuous approximation. Therefore, the pattern shaped by the unstable mode continues to be the striped pattern over time; to be specific, the amplitude of the unstable mode in the continuous model (broken curves in Fig. 5.7) grows increasingly as $t \to \infty$, and that in the stochastic model (solid curves in Fig. 5.7) persists, but stops growing after a certain time.

We now turn to the case where the cells have the same initial distributions containing only mode 3, as shown in Fig. 5.8a ($\langle U \rangle$ is not visible in the graph). In this case, $\langle V \rangle$ initially decays with time and then grows in the

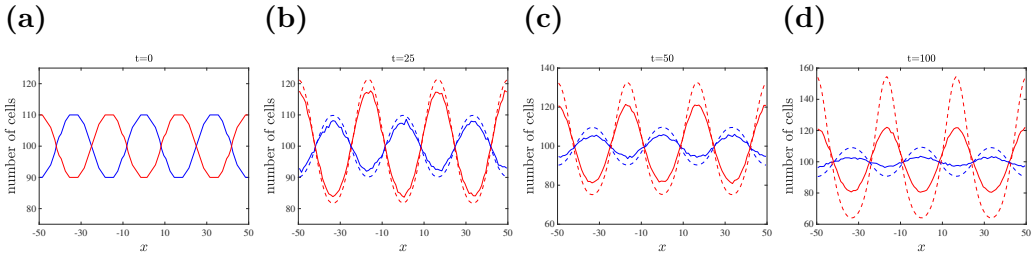


Figure 5.7: Persistence of the stripes, when the cells are initially distributed in a striped pattern of mode $m = 3$. The graphs from left to right show the evolution of the pattern over time for (a) $t = 0$, (b) $t = 25$, (c) $t = 50$ and (d) $t = 100$. The parameter values of the stochastic model are: $\alpha = 33$, $\gamma = 16, W = 100, A = R = 1$ and $d_u = d_v = 0.04$, corresponding to $Q = 1, D_u = D_v = 0.0156$ and $X = 100/16$, selected using the linear stability analysis. The red and blue curves correspond to the runners and chasers. The solid curves show the average results of the stochastic model, $\langle V \rangle$ and $\langle U \rangle$, over 500 simulations, and the broken curves show the numerical solution of the continuous model, as described by Eqns. (5.3.5).

opposite direction, indicating the inclination of the system to produce anti-phase distributions (see Figs. 5.8b-5.8d), consistent with the predictions of the linear stability analysis. The long-term behaviour of the stochastic and continuous models is similar to those shown in Fig. 5.7: as $t \to \infty$, the amplitude of solution of the continuous model grows incessantly, unlike the average pattern of the stochastic model, which stops growing after $t \approx 100$.

Therefore, the mechanism is able to produce the stripes from equal initial distributions. Similar to the results of the initially striped pattern, shown in Fig. 5.7, the amplitude of $\langle V \rangle$ is growing significantly larger than that of $\langle U \rangle$, verifying the small $|\epsilon_u(3)|/|\epsilon_v(3)|$. Similar to Fig. 5.8, the broken curves show the results of the continuous model. Comparing the results of the models shows a similar escalation of disagreement between the results at longer times, as explained earlier.

The stochastic model is thus able to produce stripes without homotypic interactions, from a perturbation containing only the dominant unstable mode found by the linear stability analysis. This is in contrast to the propositions made in [130, 182] that the run-and-chase mechanism alone does not lead to the formation or persistence of stripes. In addition, the results suggest that the linear stability analysis of the continuum approximation can be a useful predictor of the behaviour of the stochastic model.

To check the extent of agreement between the linear stability analysis and

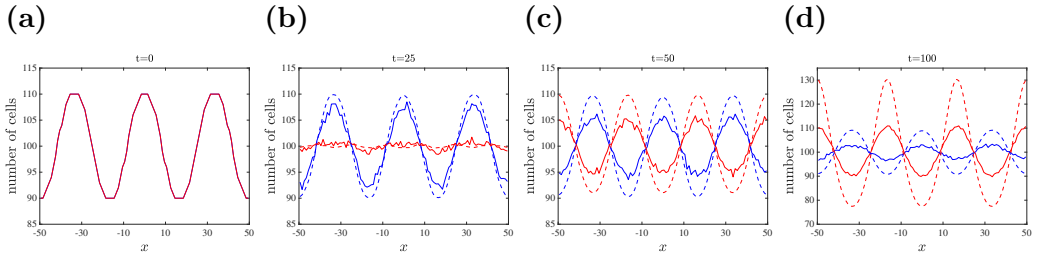


Figure 5.8: Stripe formation from (a) equally distributed cells containing only a single mode 3 perturbation. The values of the parameters can be found in the caption of Fig. 5.7. The patterns are shown at times (b) $t = 25$, (c) $t = 50$ and (d) $t = 100$. The red and blue curves correspond to the runners and chasers. The solid curves show the average results of the stochastic model, $\langle V \rangle$ and $\langle U \rangle$, over 500 simulations, and the broken curves show the numerical solution of the continuous model, as described by Eqns. (5.3.5).

the results of the stochastic model, we simulate the same system (the values of the parameters can be found in the caption of Fig. 5.7) for another single-mode perturbation, namely $n = 6$, which is a stable mode (see Fig. 5.6). As can be seen in Figs. 5.9a-5.9d, the perturbation decays over time, indicating that the linear stability analysis has correctly predicted the stability of this mode in the stochastic model, at least in certain regimes.

For the selected set of parameter values, mode $n = 1$ exhibits oscillatory behaviour, as $\Im(\lambda(1)) \approx 1.72$ and $\Re(\lambda(1)) \approx -0.016$ (see Fig. 5.6). Here, we study the evolution of this mode when it is the only perturbed mode in the initial distribution, depicted in Fig. 5.10a. The parameter values are not changed (see the caption of Fig. 5.7), which make $m = 3$ the dominant unstable mode. Figs. 5.10b-5.10d demonstrate the oscillating behaviour of the system as predicted by the linear stability analysis.

Having observed the good agreement between the behaviour of the stochastic model and the linear stability analysis for single-mode perturbation experiments, we simulate the system for a more realistic initial perturbation: uniform distributions (shown in Fig. 5.11a). The values of the parameters are the same as in Figs. 5.7-5.10 and 500 simulations are generated. Thus, according to the linear stability analysis, we expect that mode 3 would dominate other modes and form the striped pattern. The results show that mode 3 can be discerned (to some degree) in $\langle V \rangle$ (the red curve) at times $t = 50$ (Fig. 5.11b) and $t = 100$ (Fig. 5.11c), but not in $\langle U \rangle$ (the blue curve). Comparing the results of the stochastic model at these times with the numerical solution of the PDEs (see Figs. 5.11f and 5.11g) indicate significant differ-

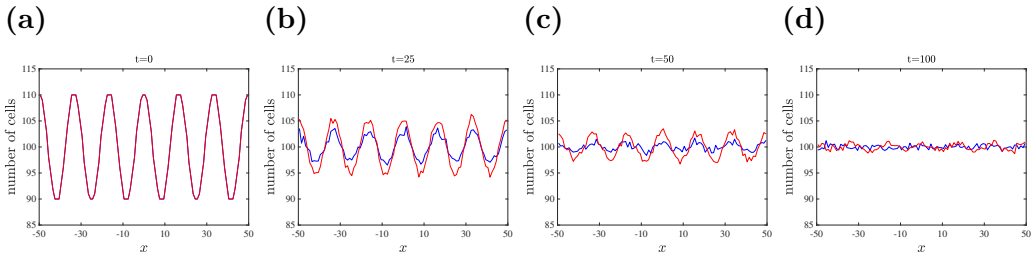


Figure 5.9: Simulation of the stochastic model for (a) equal initial distributions containing only a stable mode (n=6) as the perturbation. (b-d) the patterns at times $t = \{25, 50, 100\}$. The values of the parameters can be found in the caption of Fig. 5.7. The red and blue curves show the average number of runners, $\langle V \rangle$, and chasers, $\langle U \rangle$, over 500 simulations respectively.

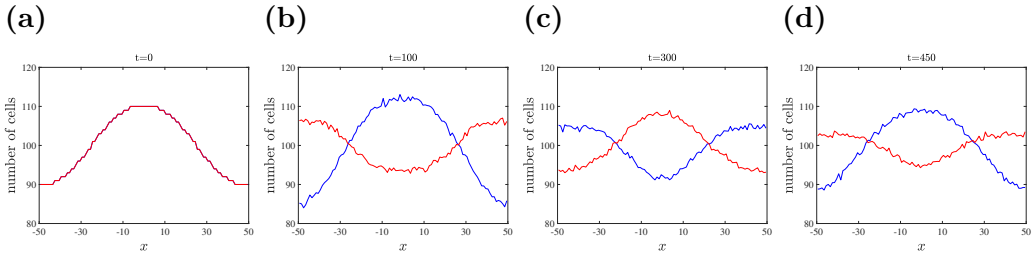


Figure 5.10: Simulating the stochastic model with (a) the initial distribution containing only the stable mode $n = 1$, which has a $\Im(\lambda(1)) \approx 1.72$ and $\Re(\lambda(1)) \approx -0.016$. The patterns at times (a) $t = 0$, (b) $t = 100$, (c) $t = 300$ and (d) $t = 450$ are shown. The parameter values of the stochastic model are: $\alpha = 33$, $\gamma = 16, W = 100, A = R = 1$ and $d_u = d_v = 0.04$. The red and blue curves show the average number of runners, $\langle V \rangle$, and chasers, $\langle U \rangle$, over 500 simulations respectively.

ences. In particular, some stable modes, *e.g.* mode 6, seem to have fairly large amplitudes in the PDE solutions (see Fig. 5.11g), in contrast to the results of the stochastic model where small modes are dominant (*e.g.* see the large amplitude of mode 1 in Fig. 5.11c). Fig. 5.11d shows the distributions of the two cell types at time $t = 500$, where mode 3 is not recognisable anymore in either of the distributions. However, as Fig. 5.11h shows, the striped pattern has clearly emerged in the PDE solutions - the results at longer times show clear segregation of the two species (data not shown).

In order to determine the magnitude of the constituent modes of the patterns, we evaluate the average absolute value of the Fourier Transform (FT) of the patterns, $F(n)$, which is computed by

$$F(n) = \frac{1}{M} \sum_{i=1}^{M} |\mathcal{F}_i\{\mathbf{f}\}|,$$

where $\mathcal{F}_i$ is the Fourier transform of the i^{th} out of $M = 500$ realisation of the model, and $\mathbf{f}$, *e.g.* for chasers, at a certain time is $(U_1 - N_u/W, \ldots, U_W - N_u/W)$; a similar formula holds for finding the FT of the runners. Fig. 5.11e shows a notable peak at $n = 3$ of the FT of $\langle V \rangle$, verifying the dominance of mode 3 in the distribution of the runners, unlike the FT of $\langle U \rangle$. The remarkable peaks at $n = 3$ of the FTs of the PDEs solution indicate that mode $m = 3$ has clearly emerged in the results of the continuous model.

We can infer from the results that the dominant unstable mode found by the linear stability analysis emerges in $\langle V \rangle$, but fails to emerge in $\langle U \rangle$, when the initial distributions are uniform, as opposed to the emergence / persistence of this mode for initial distributions that contain single-mode perturbations (see Figs. 5.7 and 5.8). This leads us to investigate the differences in the evolution of the distributions of the two species about equilibrium. As discussed before, $|\epsilon_u(3)|$ is significantly smaller than $|\epsilon_v(3)|$ for the selected parameter values - in fact, $|\epsilon_u(3)|$ is smaller than most of the other modes, as shown in Fig. 5.6. Observing that the attenuation of mode 3 in u is significantly stronger than that of v suggests that the significantly smaller value of $|\epsilon_u(3)|$ has led to the failure of the striped pattern to form.

In order to test this idea, we select a set of parameter values that provides a larger $|\epsilon_u(3)|$, to test if it can produce a larger amplitude of u, and, in turn, lead to stripe formation from initial uniform distributions. According to the results shown in Fig. 5.5 (also see Eqn. (5.5.9)), we must have larger Q and D to have a higher $|\epsilon_u(3)|/|\epsilon_v(3)|$. Thus, we choose $Q = 4, D = 5$ ($D_v = 0.78, D_u = 0.0156$), while the rest of the parameter values are unchanged: $\alpha = 33, \gamma = 16, X = 100/16$. This new set of parameter values yields $|\epsilon_u(3)| \approx 0.18$ (about five times higher than the $|\epsilon_u(3)| = 0.04$ for the former

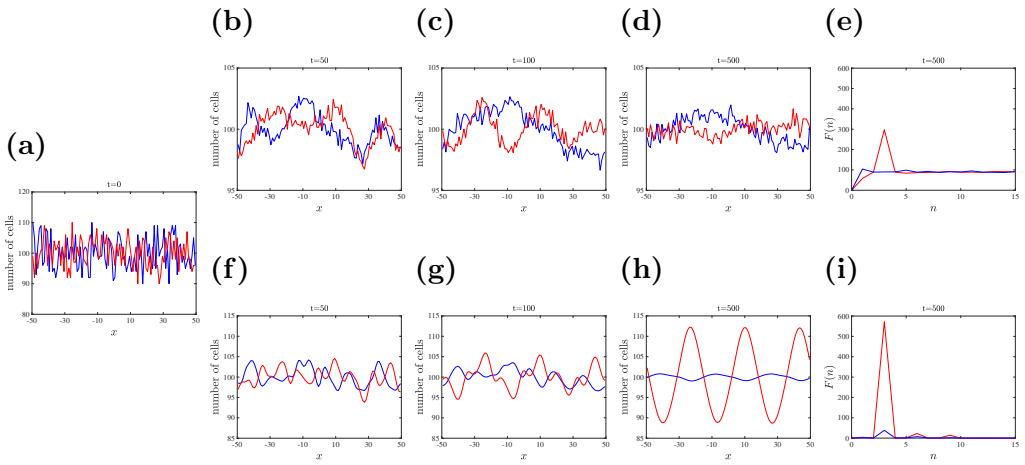


Figure 5.11: Simulating the model with (a) uniform initial distributions. The patterns produced by the stochastic model, averaged over 500 simulations, at times (b) $t = 50$, (c) $t = 100$, (d) $t = 500$. (e) the average FT of the patterns, $F(n)$, produced by the stochastic model at $t = 500$. (f-h) the numerical solutions of the dimensional continuous model, Eqns. (5.3.5), at same times as (b-e). (i) FT of the solution of the continuous model at $t = 500$. The parameter values can be found in the caption of Fig. 5.7.

set of parameter values); the growth rate for the previous set of parameter values was $\Re(\lambda(3)) = 0.035$, which is changed to $\Re(\lambda(3)) = 0.028$ for the new ones. Thus, due to the higher $|\epsilon_u(3)|$ we expect that mode 3 in $\langle U \rangle$ will emerge with a higher amplitude, compared to the previous cases.

Fig. 5.12 shows the results for the same single-mode perturbation initial distributions as those shown in Fig. 5.8a; the parameter values of the stochastic model associated with the ones for the non-dimensional continuous model are $A = 4, R = 1, d_u = 0.04, d_v = 0.2, \alpha = 33, \gamma = 16, W = 100$. As Figs. 5.12b-5.12c show, mode $m = 3$ has significantly attenuated in the distribution of both of the species, contradicting the predictions of the linear stability analysis. The numerical solution of the PDEs, shown in Figs. 5.12d-5.12f, illustrates the behaviour of the continuous model for the selected set of parameters. As can be seen in these graphs, striped pattern has clearly emerged, as predicted by the linear stability analysis. We try to explain the reason for this significant disagreement.

We investigate the behaviour of the stochastic model with higher strengths of attraction, A, and repulsion, R; the initial cell distributions are the same as those shown in Fig. 5.8a. The columns of Fig. 5.13, from left to right, correspond to $A = R = 2$, $A = R = 3$ and $A = R = 4$. Fig. 5.13a shows that mode 3 has almost disappeared in $\langle U \rangle$, while it still has a large amplitude in $\langle V \rangle$ when $A = R = 2$. Comparing this figure with Fig. 5.8 shows that the disagreement between the results of the stochastic and continuous models significantly increased. The attenuation of mode 3 in the distribution of runners (red curves) has been intensified as we increase A and R, see Figs. 5.13b and 5.13c where $A = R = 3$ and $A = R = 4$, respectively. Figs. 5.13d-5.13f verify the reduction in the magnitude of $F(3)$, indicating that mode 3 is attenuating in $\langle V \rangle$, as well as the total disappearance of this mode in $\langle U \rangle$ for all of the cases.

Our results imply that increasing the strengths of biased motion has a clear negative impact on the agreement between the stochastic and continuous models. However, this behaviour appears to be consistent with the findings of Davies *et al.* (2014). In using the mean-field approximation to derive the continuum approximation of the stochastic model, we assumed that the numbers of cells at each site is independent of that at the others. Yet, a large biased motion rate will tend to produce a significant correlation between the numbers of cells at adjacent sites. As a result, it seems impossible to increase $\epsilon_u(m)$ for a small mode like $m = 3$ by increasing $Q = A/R$ and $D = d_v/d_u$ without falling into regimes of high biased motion, and hence poor agreement between the two models.

The stochastic model was simulated with a wide range of different parameter values that produce higher $|\epsilon_u(3)|$. However, a similar breakdown

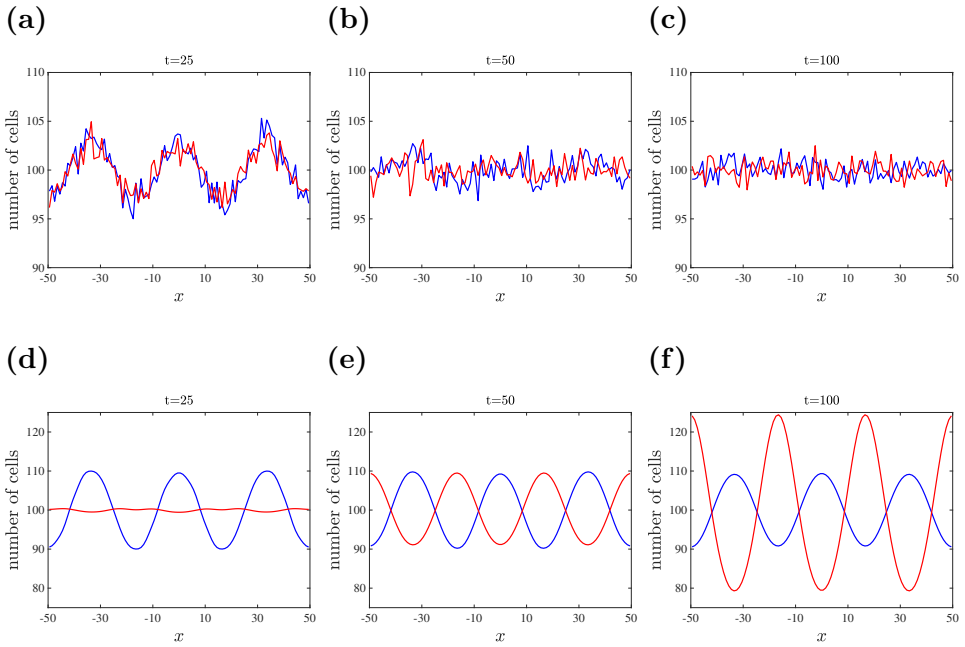


Figure 5.12: Simulating the stochastic model for $A = 4, R = 1, d_u = 0.04, d_v = 0.2, \alpha = 33, \gamma = 16, W = 100$, selected using the linear stability analysis of the non-dimensional continuous model with parameter values of $Q = 4, D = 5$ ($D_v = 0.78, D_u = 0.0156$). The first and second rows show the results of (a-c) the stochastic model, averaged over 500 simulations, and (d-f) the dimensional continuous model, respectively, at times $t = \{25, 50, 100\}$ from left to right. The initial distributions are depicted in Fig. 5.8a

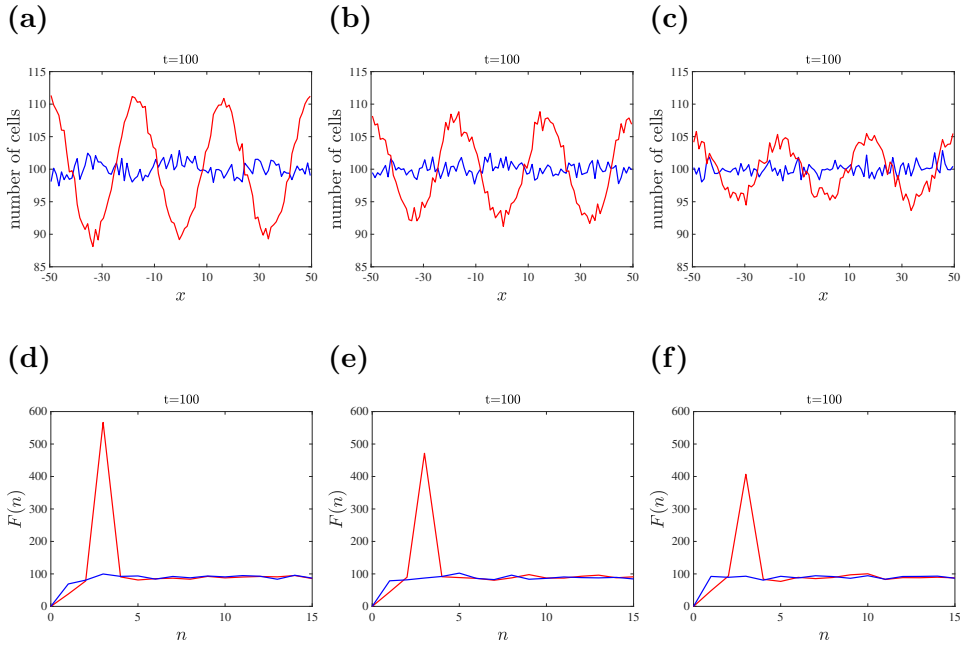


Figure 5.13: Simulating the stochastic model when varying the strength of attraction and repulsion. The initial cell distributions are the same as those shown in Fig. 5.8a, and the results show averages over 500 simulations. (a) $A = R = 2$, (b) $A = R = 3$, (c) $A = R = 4$. The average FTs, $F(n)$, corresponding to the patterns in (a-c). The values of the rest of parameters are: $d_u = 0.04, d_v = 0.04, \alpha = 33, \gamma = 16, W = 100$.

of stripe formation was observed in each case (data not shown). Note that we cannot significantly reduce R and increase d_v, since the system would become stable (as outlined in §5.5, instability of a mode requires $AR/d_u d_v$ to be sufficiently large (see Eqn. (5.5.7)). Having a larger A and smaller d_u also failed to result in a good agreement between the two models due to the high ratio of biased motion rate to the unbiased one. Furthermore, simulating the model with uniformly distributed cells as the initial condition for these new parameter values demonstrated a significantly higher attenuation of mode $m = 3$ (data not shown), compared to the results shown in Fig. 5.11d.

Fig. 5.5 shows that higher dominant unstable modes, *e.g.* $m = 9$, can potentially lead to patterns with higher $|\epsilon_u(m)|$. Thus, we select $Q = 1, D_u = D_v = 0.004$ which correspond to $A = R = 4$ (other parameter values can be found in the caption of Fig. 5.14), which initiates mode $m = 9$ as the dominant mode with the highest $\Re(\lambda)$; the initial distributions are as depicted in Fig. 5.14a. The results at time $t = 100$ (see Fig. 5.14b) indicates that

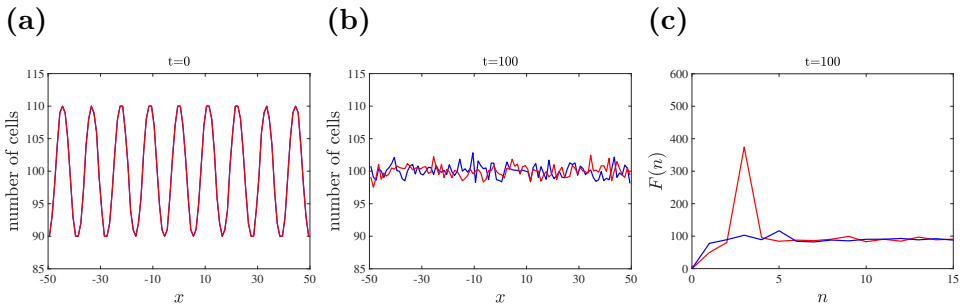


Figure 5.14: Simulation of the stochastic model with $d_u = d_v = 0.04, A = R = 5$ for a lattice of size $W = 100$. The parameter values of the non-dimensional continuous model are: $Q = 1, D_u = D_v = 0.004, X = 100/16$. The linear stability analysis predicts mode $m = 9$ as the dominant mode in the pattern for the selected set of parameter values. (a) the initial distributions. (b) the distribution of the cells at time $t = 100$. (c) the average FTs of the distributions. The results are averaged over 500 simulations.

the model fails to initiate the emergence of large modes like $m = 9$ with the lattice size of $W = 100$, even in $\langle V \rangle$; this is verified by the average FTs shown in Fig. 5.14c.

In order to lower the inaccuracies originating from the discretisation level of the domain (see Eqn. (5.3.2)), we must simulate the model on larger lattice sites. This may enable us to have larger strengths of interactions (higher biased motion rates), and also larger dominant unstable modes like $m = 9$. However, larger lattices make the computations intractable using the direct Gillespie algorithm. To solve this problem, we can choose approximate simulation algorithms, such as τ-leaping Gillespie algorithm. We use a variant of this method introduced by Cao *et al.* (2006) and simulate the system with $W = 1000$ for the sake of illustration in Appendix B.2. The results show that using this method allows us to initiate the emergence of dominant modes when we have larger strengths of interactions. In addition, the unstable mode $m = 9$ (see Fig. B.2) can emerge as the dominant mode in this larger lattice. We leave further analysis of the models when using larger lattices to future work.

5.7 Discussion

In this chapter, we have investigated pattern formation driven by a run-and-chase mechanism, where cells of type 1 are attracted to cells of type 2, whilst

cells of type 2 are repelled by those of type 1. Run-and-chase has recently received attention for being involved in various biological processes, such as stripe formation on the skin of zebrafish [78, 184], and collective migration of neural crest cells [162]. Using stochastic discrete and deterministic continuous models developed in this chapter, we studied striped pattern formation in systems driven by this mechanism.

We developed a stochastic discrete model where cells move on a one-dimensional lattice according to the motion rates determined by the attractive / repulsive interactions between the two species. Gillespie algorithms [57, 60] were used to simulate the model, as using the simulation algorithms introduced in Chapters 3 and 4 proved to be very time-consuming due to small motion rates. Using a mean-field type approximation, we derived a continuous non-local model approximating the average behaviour of the stochastic model. This consists of a pair of integro-PDEs of the form of Eqn. (1.3.1), a framework which has been frequently used to describe collective behaviour of cells [64, 108, 109, 130, 167].

In the derivation of the continuous model, a cell-size offset arose in the kernels, due to the way we include the effect of neighbouring cells to find the biased motion rates (see Eqns. (5.2.1a) and (5.3.4)). To be specific, we do not include the centre site in the interaction kernel, *i.e.* interaction does not start at distance zero. We showed that this can significantly affect the stability analysis of the system such that without it the system would always be stable. This indicates the importance of considering the underlying discrete nature of cells, as it can impact our analysis, and, in turn, our interpretations of the system of interest. Similarly, in Chapter 4, we showed that the ABMs, which incorporate cell area exclusion [10], can produce striped patterns. Hence, we may conclude that the emergence of striped pattern in the model developed in this chapter and the ABM in Chapter 4 can be due to cell-size effects (although included / modelled differently in each chapter), rather than *e.g.* the presence of homotypic interactions, as suggested by Painter *et al.* (2015).

Linear stability analysis was applied to the continuous model to assess the behaviour of the system when it is excited with a small initial perturbation about the equilibrium. Using the analysis, we identified how the parameters may influence the patterns produced in the stochastic model.

To produce striped patterns, we selected a set of parameter values that was predicted to give rise to anti-phase distributions of the species (similar to the stripes of zebrafish), and simulated the stochastic model accordingly. The results for small rates of biased motion demonstrated that the linear stability analysis can accurately predict stability of a mode, phase of the distribution of the two species, and also the emergent amplitudes of the variations in each of the species, when the initial distributions contained only single-mode

perturbation about equilibrium. Also, the results showed that persistence and emergence of striped patterns are possible in a system driven by the run-and-chase mechanism, contradicting what is suggested in [130, 182].

In the simulations with uniform initial distributions, the dominant mode appeared in the distribution of the runners, but failed to appear in that of the chasers. Using the linear stability analysis, we selected another set of parameters that would lead to a larger amplitude of variations in the dominant mode of the distribution of the chasers and potentially produce the stripes. However, this led to severe attenuation of the dominant mode, contradicting the prediction of the linear stability analysis. Using further analyses and simulations, we demonstrated that the desirable set of parameter values that yield large amplitudes of variations in the distribution of both of the species requires significant biased diffusion rates. This, however, reduces the agreement between the discrete stochastic and continuous models due to violation of the mean-field approximation.

In summary, we conclude that the behaviour of the stochastic model agrees well with the continuous model when the linear stability analysis predicts stability. This allows us to narrow down our search in the parameter space by excluding the set of parameter values corresponding to stability, where no pattern formation will arise. However, when instability is predicted, poor agreement between the models could occur. For the unstable cases where the rate of biased motion is not significantly larger than the unbiased motion, the agreement is reasonably good, but, on increasing the rate of biased motion, the agreement drastically declines. This is because high strengths of attraction/repulsion (producing biased motion) result in strong correlations between the numbers of cells at neighbouring sites, violating the mean-field assumption, in the derivation of the continuous model.

Our work reinforces the usefulness of continuous models in helping to understand the mechanisms of pattern formation, but sounds a note of caution that the patterns arising from the continuous models may not accurately reflect those observed in nature, where individual-level and stochastic effects may be important. In particular, we showed that for the non-local continuous models, incorporating cell-level features in the model can considerably influence our analysis, and consequently our interpretations of a system.

We observed that the stochastic model fails to give rise to the emergence of large-wavenumber modes. To resolve this failure, larger lattices must be used. However, this can increase the computational cost of using the direct Gillespie algorithm such that the simulations would be intractable. To solve this problem, approximate simulation algorithms, such as τ-leaping Gillespie algorithm [60], could be employed. We implemented this method and simulated a system for the sake of illustration, as can be found in Appendix B.2.

However, further investigations are left as possible future work.

Extending the methods to two-dimensional (or three-dimensional) domains would provide more realistic results. We developed the stochastic model for two-dimensional lattices and derived the corresponding continuous approximations (see Appendix B.3). However, simulations of two-dimensional lattices are extremely time-consuming, even with approximate simulation algorithms, thus, this is also left for future work.

More accurate approximation techniques, such as Kirkwood superposition approximation (KSA) [33, 82, 108, 153], can be employed to obtain continuous models with higher fidelity to the average behaviour of the stochastic model. However, using these methods could lead to more complicated continuum models, which are much less analytically tractable. Alternatively, spatial moment dynamics modelling approaches [14] can be used to study the collective behaviour of individuals in stochastic discrete models. These approaches can provide more accurate description of the population-level behaviour of a system, as the dependence between the number of cells at distinct lattice sites (positions in particle-based models) is taken into account. Formulating the moments dynamics also provides valuable insights into how spatial statistics, like PCF (second moment), for a system might evolve over time.

Chapter 6

Conclusions

The aim of this thesis is to improve our understanding of tissue development by i) quantifying the distribution of cells in tissues, and ii) modelling cell interactions that influence formation of tissue patterns. Towards this end, we have introduced a method of quantifying the spatial arrangement of cells in tissues using pair-correlation functions (PCFs), and developed mathematical models to simulate the underpinning mechanisms influencing cell aggregation in tissue development. In the following, we summarise the contents of each chapter and suggest potential future work.

In Chapter 2 of this thesis, we developed a PCF to quantify distribution of cells in tissues. The PCF empowers us to detect deviations in distribution of cells from complete spatial randomness (CSR) over specified directions of interest. As a result, we can identify aggregation or segregation lengthscales in the patterns. We applied the PCF to tumour spheroids for automatic identification of the necrotic zone boundary in these *in vitro* grown cancerous tissues, which is crucial in the assessment of different treatment strategies. We showed that this can be difficult especially when the difference in cell density between the necrotic and viable zones of a tumour spheroid is small. By detecting the deviations from complete spatial randomness (CSR state) over one-dimensional directions, the PCFs enabled us to provide quantitative estimates of the radial distance of the necrotic zone boundary from the centre of a tumour spheroid. We validated our approach on synthetic tumour spheroids in which the position of the necrotic zone boundary was known *a priori*. It was then applied to nine real tumour spheroids imaged with light sheet-based fluorescence microscopy. Comparing the PCFs estimates of the necrotic zone boundary with those of a human expert showed a reasonable agreement, and outperformance of other computational methods, *i.e.* DBSCAN and K-means. An obvious extension would be to apply this method

on a larger number of tumour spheroids to obtain population-level data about their spatial features. The method might also be refined by considering more complicated spatial models for the structure of the spheroids. For instance, one can consider a gradual change for modelling the transition between the zones (*e.g.* represented by a sigmoid function).

Quantification methods can help us identify the differences in the structure of tissues, but they do not provide us with explanation of how these differences arise. Mathematical models can be used to investigate the underlying mechanisms that result in these differences. In Chapter 3, we assessed tumour spheroid growth under two different culture protocols, namely microgel and suspension cultures. The experiments conducted by our collaborators [30] showed that the spheroids cultured in microgel are more uniform in size than the ones cultured in suspension, and we were interested in explaining these experimental observations by means of mathematical modelling. Thus, we developed an ABM in which cells can move, proliferate and die, and adjusted the parameters for each culture medium to simulate their distinct properties. The results showed that the ability of the microgel to separate cells into multiple layers and hence reduce the effective initial density is one factor contributing to generation of more uniform spheroids. Another is the ability of the microgel in reducing cell death due to lack of (cell-cell or cell-substrate) adhesion in small cell aggregates (or even isolated single cells). Extending the ABM to a three-dimensional model is an obvious avenue for future work, as the biological processes are three-dimensional problems. Clearly, in that case, we would need data on the three-dimensional locations of cells, as for the tumour spheroids described in Chapter 2. Provided such data were available, more sophisticated methods, *e.g.* PCFs, could be used to accurately compare the results of simulations and experiments.

Although the cells are of the same type in tumour spheroids grown *in vitro*, this is not usually the case for *in vivo* tissues, which can contain many different cell types. Thus, in Chapter 4, we extended the ABM developed in Chapter 3 to a two-species model. We demonstrated that a variety of different pattens can emerge by varying the homotypic / heterotypic attractive-repulsive inter-individual interactions. We specifically simulated some archetypal mechanisms involving attraction and repulsion that govern the interactions between hepatocyte-stellate cells (liver cells), predator-prey systems and melanophore-xanthophore cells (zebrafish pigment cells). Then, we developed a PCF for two-species patterns to analyse the heterotypic features in spatial distributions, *i.e.* features in distribution of each cell type relative to that of other type, along with the homotypic ones. We showed the importance of analysing the homotypic and heterotypic PCFs simultaneously, as distinct patterns might have similar homotypic (heterotypic) PCFs

but different heterotypic (homotypic) PCFs. The run-and-chase mechanism was shown to be a potential mechanism for producing the stripes on the skin of zebrafish, which is consistent with some biological experiments [78, 184]. This is in contrast to the earlier findings from mathematical models [130, 182] suggesting that this mechanism is incapable of producing the striped pattern.

A disadvantage of the ABMs used in Chapters 3 and 4 is that without simulating them over a wide range of parameter space, it is not possible to determine under what conditions a tissue pattern will emerge from uniform initial cell distributions. When the parameter space is large, this is computationally expensive and time consuming. Therefore, in Chapter 5, we considered the connections between a two-species discrete stochastic model and a nonlocal continuous model derived from it to approximate its average behaviour. We showed that when the strength of interactions between the two species is not significantly higher than the rate of unbiased motion, linear stability analysis of the continuous model is able to accurately predict the regions of parameter space where patterns form in the stochastic model. This is because high strength of attraction / repulsion creates a strong correlation between the numbers of cells at adjacent sites, violating the mean-field assumption. Conversely, increasing the unbiased diffusion reduces the correlation and improves the agreement between the stochastic and continuous models. Therefore, we must be aware of the caveats of using this kind of approximated models to analyse biological systems, as in certain cases, they may produce patterns that are considerably different from those of their discrete counterparts, which take the discrete nature of cells into account.

As a potential avenue for furthering this research, we could explore more accurate continuous approximation techniques, such as the Kirkwood superposition approximation (KSA) [33, 82, 153]. These techniques allow us to develop continuous models with higher fidelity to the average behaviour of stochastic models, though at the cost of being less analytically tractable. Spatial moment dynamics modelling [14] can also be used to study cell aggregation in tissue development. These approaches do not use the mean-field assumption, thus, enable them to provide more accurate description of the population-level behaviour of a system, by taking the dependence between the number of cells at distinct lattice sites (positions in particle-based models) into account. Formulating the moment dynamics also provides valuable insights into how spatial statistics, like the PCF (second moment), for a system might evolve over time.

In this thesis, we have demonstrated how combining new quantification and modelling approaches provides a means to gain more comprehensive insights into tissue development, by analysing the biological data along with simulating the underpinning mechanisms. The next step in extending this

research would be to connect our quantification and modelling approaches to estimate our model parameters based on experimental data, using parameter estimation methods such as Approximate Bayesian Computation [83, 144, 166]. We can use our PCFs as summary statistics in these methods to estimate values (distributions) for the parameters in our models, *e.g.* strength / range of attraction / repulsion. Then, we will have biologically realistic models that provide us with a means to accurately evaluate various tissue development hypotheses.

Ultimately, we could use the approximate continuous models to investigate how variation of parameters about their original values (estimated by the aforementioned methods) may influence tissue pattern formation. This would be extremely useful for tissue engineering studies as it would enable us to suggest new strategies to optimise culture methods based on the model predictions.

Appendix A

Quantification of three-dimensional tumour spheroids: further details of methods and data

This appendix provides additional details on the data we used for our analysis and the computational methods with which we compared our PCF.

A.1 The effect of density differences on the estimate of the necrotic zone boundary

Fig. A.1 shows synthetic tumour spheroid data results for given values of the necrotic zone boundary $B = 200$ and total cell number $N = 5000$. From left to right in the columns of Fig. A.1 the value of ΔN is increasing, or the difference in density between the necrotic and viable zone is increasing, and we observe similar changes in the PCFs and normalised density to that of Fig. 2.4 (from left to right). This demonstrates that it is primarily the increase in the difference of density (between the necrotic and viable zones) that produces more reliable necrotic zone boundary estimates.

A.2 Additional tumour data

Fig. 2.5 in §2.4.1 of the thesis shows only three tumour spheroids we used in our analysis. Fig. A.2 includes the other six tumour spheroids that were

used in our PCF quantification method.

A.3 Implementation of DBSCAN

DBSCAN [72] classifies points as either outliers or as a compact cluster (see blue and red points for the synthetic tumour spheroids in Fig. A.3). When the difference in density between necrotic and viable zones is visually distinguishable the majority of outliers are located within the necrotic zone (see Fig. A.3(a)). However, as the contrast in density between the two zones diminishes (see Figs. A.3(a)–(d)) there is an increase in the proportion of outliers found in the viable zone. Therefore, we post-process the DBSCAN results to obtain accurate estimates of the necrotic zone boundary. This is a two stage process.

First, we remove outliers with a projected or mapped value of $a > v$, to exclude outliers from the analysis that lie very close to the outer boundary of the viable zone. Second, we percentile rank the remaining mapped values of the outliers and remove projected outlier points above the pth percentile, to remove outliers residing within the viable zone that are a considerable distance away from the necrotic core. It is important to recognise that these two stages are semi-automatic and we choose $v = 0.9$ and $p = 95$ based upon visual inspection of outliers that can be clearly identified to reside within the necrotic core in Fig. A.3. The remaining outliers are used for the DBSCAN estimate of the necrotic zone boundary.

The DBSCAN non-dimensional necrotic zone boundary estimate is taken as the maximum value of a in the set of remaining outliers. A comparison between DBSCAN and PCF estimates is shown in Table A.1 for the synthetic tumours in Figs. A.3(a)–(d). The DBSCAN and PCF methods provide accurate estimates for the synthetic tumour spheroids in Figs. A.3(a)–(c). However, only the PCF method provides accurate estimates for the synthetic tumour spheroid in Fig. A.3(d), where the point pattern is visually indistinguishable from the CSR point process.

In addition to DBSCAN, we also implemented MATLAB's k-means standard clustering algorithm [80]. In a similar way to DBSCAN, the algorithm can be used to partition points into two sets (*i.e.* $k = 2$ clusters). However, the two sets contained points belonging to both the necrotic and viable zones that we were not able to post-process as effectively as we did for DBSCAN. This led to inaccurate estimates (not shown) for all four synthetic tumour spheroids shown in Fig. A.3.

	$\Delta N = 8000$	$\Delta N = 6000$	$\Delta N = 4000$	$\Delta N = 2000$
Non-periodic PCF	0.62	0.61	0.59	0.57
Periodic PCF	0.61	0.60	0.61	0.59
DBSCAN	0.62	0.57	0.61	0.88

Table A.1: Estimates for the known necrotic zone boundary, $\tilde{B} = 0.6$, for the synthetic ellipsoids shown in Fig. A.3.

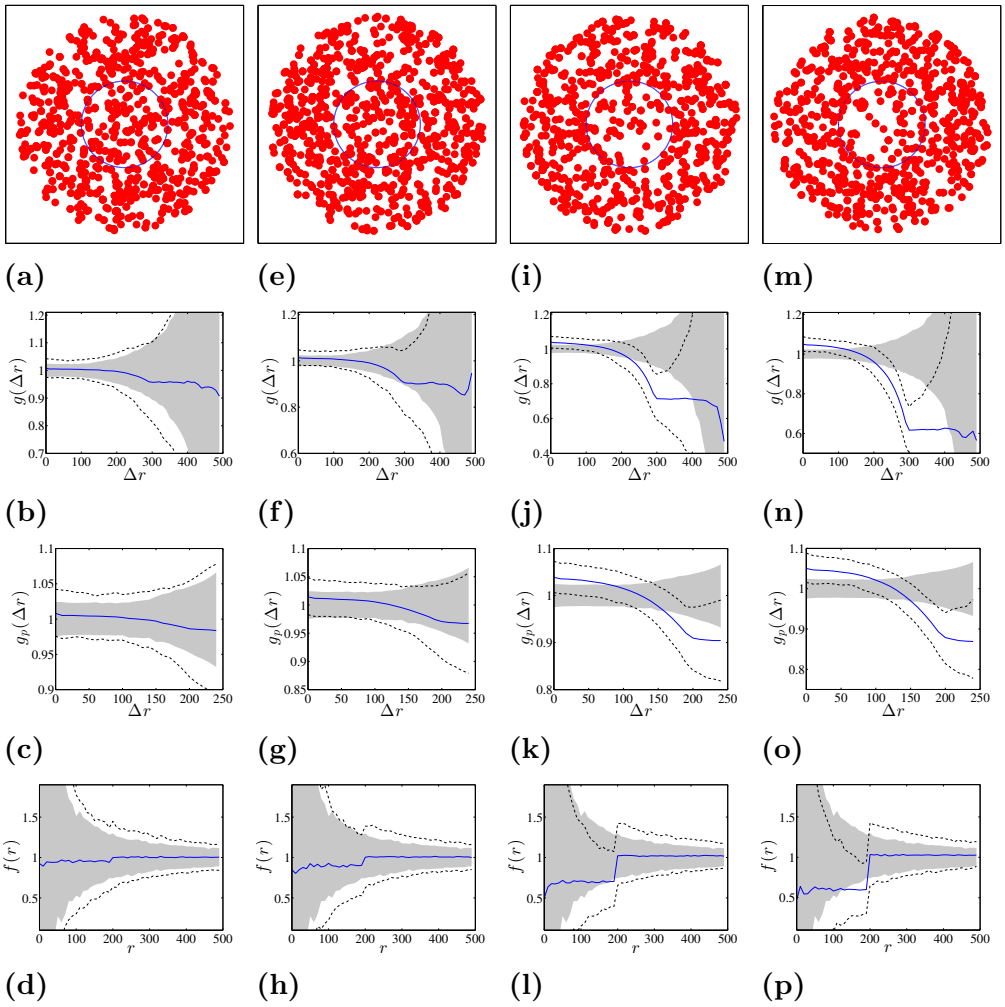


Figure A.1: Synthetic tumour spheroid point patterns, $N = 5000$, $B = 200$, $R = 500$ and $h = R/50$. Columns: left to right $\Delta N = \{250, 500, 1500, 2000\}$. Top row: Central slice of synthetic tumour. The remaining rows are for statistics from 1000 realisations. Second row: Average non-periodic PCF, solid curves. Third row: Average periodic PCF, solid curves. Fourth row: Average normalised density, solid curves. The upper and lower broken curves are for the 97.5 and 2.5 percentiles, and the shaded region is for the 97.5 and 2.5 percentiles of 1000 CSR point patterns.

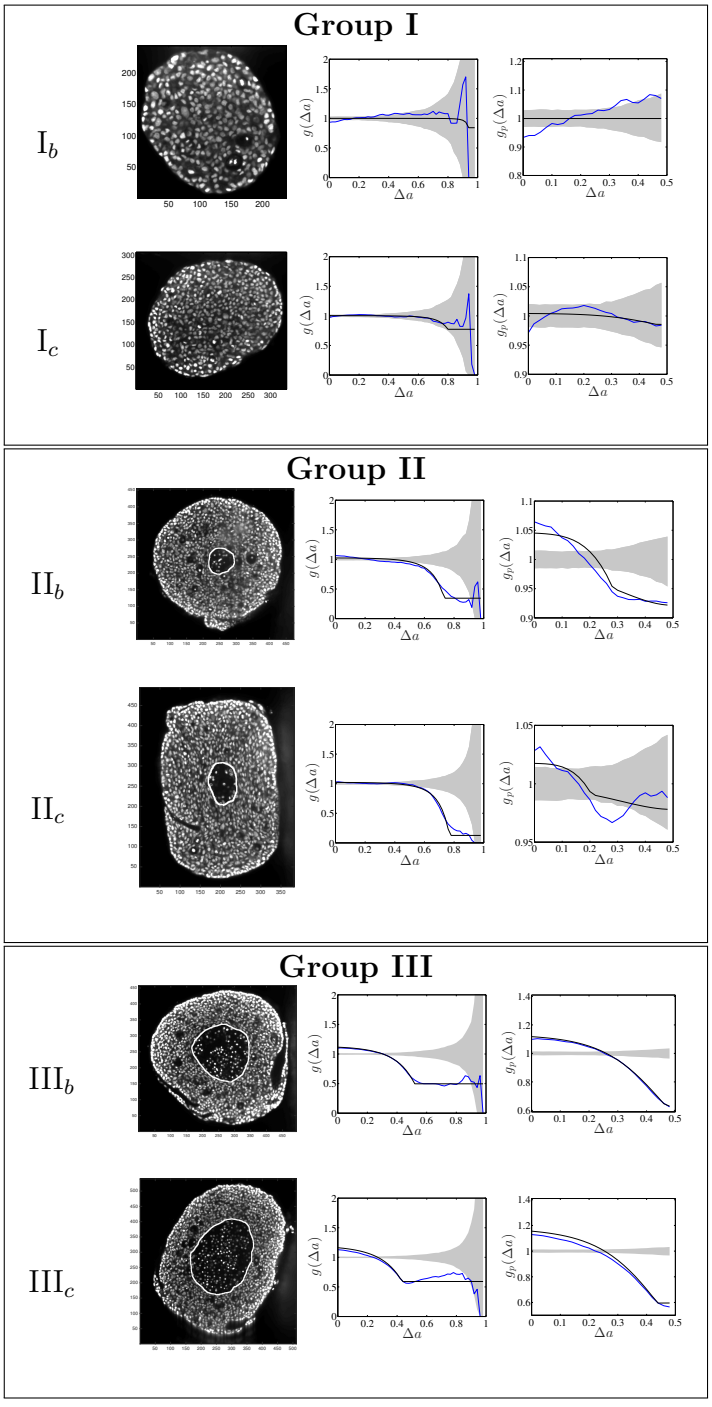
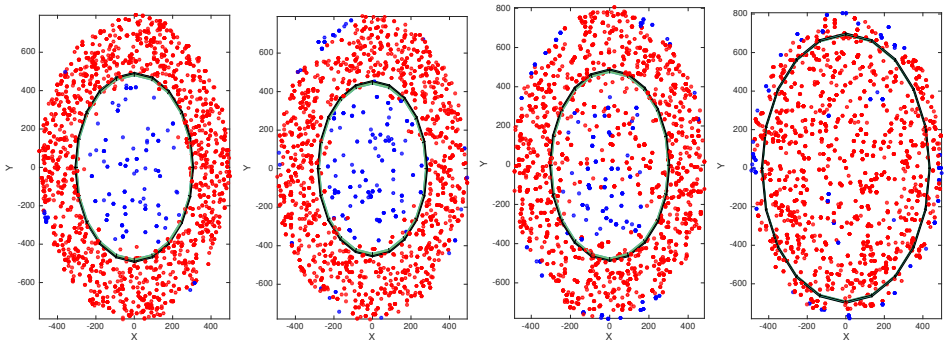


Figure A.2: Central image slice and corresponding point pattern analysis for the remaining six tumour spheroids. See the caption of Fig. 2.5 for further information.



(a) **(b)** **(c)** **(d)**

Figure A.3: Estimating the known necrotic zone boundary, $\tilde{B} = 0.6$, in synthetic ellipsoids with DBSCAN. (a)–(d) Central slices of four ellipsoids, $N = 10000$, $X = 500$, $Y = 800$, and $Z = 600$. (a) $\Delta N = 8000$, (b) $\Delta N = 6000$, (c) $\Delta N = 4000$ and (d) $\Delta N = 2000$. The red and blue points are classified as a cluster and set of outliers, respectively. The curves are for the estimates of the necrotic zone boundary.

Appendix B

Non-local continuous model approximations

B.1 Master equations derivation

The details of derivation of Eqn. (5.3.1) in §5.2, are explained here. We define $P_t(M_i)$ to be the probability that M_i cells exist at time t at site i, *i.e.* $P_t(M_i) = \Pr\{U_i(t) = M_i\}$. Then, we wish to find $P_{t+\mathrm{d}t}(M_i)$, which in turn yields equation(s) describing the evolution of the distribution of the cells. This can be done by enumerating the events that may occur during a time interval $\mathrm{d}t$ and lead to M_i cells being at site i, as our model is a Markov process. The number of cells in site i can only change if the site i or one of its adjacent sites undergo a motion during $\mathrm{d}t$, since we assume that $\mathrm{d}t$ is so small such that no more than one event can occur in a site during this infinitesimal time span. Thus, we define the joint probability $P_t(\mathbf{M}_i)$= $P_t(M_{i-1}, M_i, M_{i+1}) = \Pr\{\mathbf{U}_i(t) = \mathbf{M}_i\}$ where $\mathbf{M}_i = (M_{i-1}, M_i, M_{i+1})$ and $\mathbf{U}_i(t) = (U_{i-1}(t), U_i(t), U_{i+1}(t))$. The aim is to find the marginal probability $P_{t+\mathrm{d}t}(M_i)$ defined by

$$P_{t+\mathrm{d}t}(M_i) = \sum_{\mathbf{M}'_i} T(M_i|\mathbf{M}'_i) P_t(\mathbf{M}'_i)$$

where $T(M_i|\mathbf{M}'_i)$ is the transition probability that leads to $U_i(t + \mathrm{d}t) = M_i$, conditioned to $\mathbf{U}_i(t) = \mathbf{M}'_i$. As outlined in §5.2, we neglect the probabilities of occurring more than one right / left motion events during $\mathrm{d}t$ at a site. By

finding all of the relevant cases, we get

$$
\begin{aligned}
P_{t+\mathrm{d}t}(M_i) = & \\
& \sum_{M_{i-1},M_{i+1}} P_t(M_{i-1}, M_i - 1, M_{i+1}) q^+_{i-1}(M_{i-1}) \left[1 - q^-_{i+1}(M_{i+1})\right] \\
& \qquad \left[1 - q^+_i(M_i - 1) - q^-_i(M_i - 1)\right] \\
& + \sum_{M_{i-1},M_{i+1}} P_t(M_{i-1}, M_i - 1, M_{i+1}) q^-_{i+1}(M_{i+1}) \left[1 - q^+_{i-1}(M_{i-1})\right] \\
& \qquad \left[1 - q^+_i(M_i - 1) - q^-_i(M_i - 1)\right] \\
& + \sum_{M_{i-1},M_{i+1}} P_t(M_{i-1}, M_i, M_{i+1}) \left[1 - q^+_{i-1}(M_{i-1})\right] \left[1 - q^-_{i+1}(M_{i+1})\right] \\
& \qquad \left[1 - q^+_i(M_i) - q^-_i(M_i)\right] \\
& + \sum_{M_{i-1},M_{i+1}} P_t(M_{i-1}, M_i + 1, M_{i+1}) \left[q^+_i(M_i + 1) + q^-_i(M_i + 1)\right] \\
& \qquad \left[1 - q^+_{i-1}(M_{i-1})\right] \left[1 - q^-_{i+1}(M_{i+1})\right]
\end{aligned}
$$

where $q^+_j(M) = M a^+_j(t)\mathrm{d}t$ and $q^-_j(M) = M a^-_j(t)\mathrm{d}t$. Similar equation holds for the runners, but we only show the method for the chasers for brevity.

Here, as explained in §5.2, we use the mean-field approximation, *i.e.* $P(M_{i-1}, M_i, M_{i+1}) = P(M_{i-1})P(M_i)P(M_{i+1})$, to be able to simplify this equation. Then, by neglecting the terms of order $\mathcal{O}(\mathrm{d}t^2)$ in Eqn. (B.1), which is equivalent to neglecting concurrent motion events in different sites, we obtain

$$
\begin{aligned}
P_{t+\mathrm{d}t}(M_i) &= P_t(M_i - 1) \sum_{M_{i-1}} \left(M_{i-1} a^+_{i-1}\mathrm{d}t\right) P_t(M_{i-1}) \\
&+ P_t(M_i - 1) \sum_{M_{i+1}} \left(M_{i+1} a^-_{i+1}\mathrm{d}t\right) P_t(M_{i+1}) \\
&- P_t(M_i) \sum_{M_{i-1}} \left(M_{i-1} a^+_{i-1}\mathrm{d}t\right) P_t(M_{i-1}) \\
&- P_t(M_i) \sum_{M_{i+1}} \left(M_{i+1} a^-_{i+1}\mathrm{d}t\right) P_t(M_{i+1}) \\
&- P_t(M_i) \left(M_i a^+_i \mathrm{d}t + M_i a^-_i \mathrm{d}t\right) \\
&+ P_t(M_i + 1)(M_i + 1)(a^+_i + a^-_i)\mathrm{d}t + P_t(M_i)
\end{aligned}
$$

which simplifies to

$$\begin{aligned} P_{t+\mathrm{d}t}(M_i) &= P_t(M_i-1)\langle M_{i-1}a_{i-1}^+\rangle \mathrm{d}t + P_t(M_i-1)\langle M_{i+1}a_{i+1}^-\rangle \mathrm{d}t \\ &- P_t(M_i)\langle M_{i-1}a_{i-1}^+\rangle \mathrm{d}t - P_t(M_i)\langle M_{i+1}a_{i+1}^-\rangle \mathrm{d}t \\ &- P_t(M_i)M_i a_i^+ - P_t(M_i)M_i a_i^- \\ &+ (M_i+1)P_t(M_i+1)\left(a_i^+ + a_i^-\right)\mathrm{d}t \\ &+ P_t(M_i). \end{aligned} \tag{B.1.1}$$

We define

$$\langle U_i(t)\rangle = \sum_{M_i} M_i P_t(M_i).$$

Then on multiplying Eqn. (B.1.1) by M_i and summing over all possible values, we find

$$\begin{aligned} \langle U_i(t+\mathrm{d}t)\rangle &= \left[\langle U_{i-1}a_{i-1}^+\rangle + \langle U_{i+1}a_{i+1}^-\rangle\right]\left[\langle U_i\rangle + 1\right]\mathrm{d}t \\ &- \langle U_i\rangle\left[\langle U_{i-1}a_{i-1}^+ + U_{i+1}a_{i+1}^-\rangle\right]\mathrm{d}t \\ &- \sum_{M_i} M_i^2 P_t(M_i)\left[a_i^+ + a_i^-\right]\mathrm{d}t \\ &+ \sum_{M_i} M_i(M_i-1)P_t(M_i)(a_i^+ + a_i^-)\mathrm{d}t \\ &+ \langle U_i\rangle \end{aligned}$$

which yields

$$\begin{aligned} \frac{\mathrm{d}\langle U_i\rangle}{\mathrm{d}t} \approx \frac{\langle U_i(t+\mathrm{d}t)\rangle - \langle U_i(t)\rangle}{\mathrm{d}t} = \\ \langle a_{i-1}^+(t)U_{i-1}(t)\rangle + \langle a_{i+1}^-(t)U_{i+1}(t)\rangle \\ -\langle\left[a_i^+(t) + a_i^-(t)\right]U_i(t)\rangle. \end{aligned} \tag{B.1.2}$$

A consequence of the earlier assumption of the independence of the number of cells of different sites is that the motion rates at a typical site i is independent of the number of cells at this site. Thus, for example, in the first term on RHS of Eqn. (B.1.2) we will have

$$\langle a_i^+(t)U_i(t)\rangle = \langle a_i^+(t)\rangle\langle U_i(t)\rangle.$$

By making similar assumptions for the other terms at the RHS of Eqn. (B.1.2), we get

$$\frac{\mathrm{d}\langle U_i(t)\rangle}{\mathrm{d}t} = \langle a_{i-1}^+(t)\rangle\langle U_{i-1}(t)\rangle + \langle a_{i+1}^-(t)\rangle\langle U_{i+1}(t)\rangle - \langle a_i^+(t) + a_i^-(t)\rangle\langle U_i(t)\rangle,$$

which is then changed to the equation, Eqn. (5.3.1), by dividing both sides by N_u and introducing the average cell density as $\langle u_i(t)\rangle = \langle U_i(t)\rangle/N_u$. A similar equation is found for $\langle v_i(t)\rangle$ with the corresponding left and right motion rates.

B.2 Larger lattices

As outlined in §5.2.1, in order to have fast enough simulations for large lattices, we must use an approximate simulation algorithm, such as τ-leaping Gillespie algorithm [59, 60, 21], which is significantly faster than the direct Gillespie algorithm. In the τ-leaping Gillespie algorithm, we find the number of motion events that may occur over all sites, during a time-interval τ, known as the time leap. As mentioned in §5.2, the motion events occur according to a Poisson process with the defined rates. Hence, the number of cells in site i at time $t + \mathrm{d}t$ is given by

$$U_i(t+\tau) = U_i(t) + \mathcal{P}(U_{i-1}a^+_{i-1}\tau) + \mathcal{P}(U_{i+1}a^-_{i+1}\tau) - \mathcal{P}(U_i a^+_i\tau) - \mathcal{P}(U_i a^-_i\tau) \tag{B.2.1}$$

where $\mathcal{P}(\Lambda)$ denotes a sample from a Poisson distribution with mean Λ.

We choose a suitable value for the time-leap, τ, by balancing these conditions: first, it must not be so small that no significant speed-up is gained; second, it must not be so large that it produces significant inaccuracies in the simulations (by violating the assumptions considered in §5.2 and Appendix B.1).

The extent to which τ can enlarge is determined by bounding the difference of *propensity* function during time span τ [60]. Propensity function of a specific motion event indicates the inclination of a site to undergo that event [45]. We define the propensity function of stepping right at site i at time t as $G^+_i(t) = U_i(t)a^+_i(t)$, and that of stepping left as $G^-_i(t) = U_i(t)a^-_i(t)$. Therefore, to bound the propensity functions from above, a selected τ must satisfy (for brevity, we develop the method only for stepping right propensities of U)

$$\left|\frac{\Delta G^+_i}{G_i}\right| < \epsilon, \tag{B.2.2}$$

where $\Delta G^+_i = G^+_i(t + \mathrm{d}t) - G^+_i(t)$ and $\epsilon << 1$. Here, we follow the approach introduced by Cao *et al.* (2006) to find the suitable τ. Using Eqn. (5.2.1a), the propensity function of stepping right is given by (dropping time arguments)

$$G^+_i = U_i\left[d_u + \frac{A}{N_v}\sum_{a=1}^{\alpha} V_{i+a}\right],$$

which gives

$$\begin{aligned}\Delta G_i^+ &= d_u \Delta U_i + \frac{A}{N_v} U_i \sum_{a=1}^{\alpha} \Delta V_{i+a} + \Delta U_i \sum_{a=1}^{\alpha} V_{i+a} \\ &= \frac{A}{N_v} U_i \sum_{a=1}^{\alpha} \Delta V_{i+a} + \Delta U_i \left(d_u + \frac{A}{N_v} \sum_{a=1}^{\alpha} V_{i+a} \right).\end{aligned}$$

Thus, (B.2.2) becomes

$$|\frac{\Delta G_i^+}{G_i^+}| = |\frac{\sum_{a=1}^{\alpha} \Delta V_{i+a}}{\sum_{a=1}^{\alpha} (V_{i+a} + N_v d_u / A\alpha)} + \frac{\Delta U_i}{U_i}| < \epsilon. \tag{B.2.3}$$

Therefore, by choosing $|\Delta U_i|/U_i < \epsilon/2$ and

$$|\frac{\sum_{a=1}^{\alpha} \Delta V_i}{\sum_{a=1}^{\alpha} (V_i + N_v d_u / A\alpha)}| < \frac{\epsilon}{2},$$

the condition (B.2.3) will be satisfied. If we bound $|\Delta V_i|$ by $|\Delta V_i| < \frac{\epsilon}{2}(V_i + N_v d_u / A\alpha)$, then, we have

$$|\frac{\sum_{a=1}^{\alpha} \Delta V_i}{\sum_{a=1}^{\alpha} (V_i + N_v d_u / A\alpha)}| < \frac{\sum_{m=1}^{\alpha} \frac{\epsilon}{2}(V_i + N_v d_u / A\alpha)}{\sum_{m=1}^{\alpha} (V_i + N_v d_u / A\alpha)} = \frac{\epsilon}{2}.$$

which consequently yields

$$|\frac{\Delta G_i^+}{G_i}| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Using a similar method for limiting the propensity functions of V, we find

$$|\frac{\Delta V_i}{V_i}| < \frac{\epsilon}{2},$$
$$|\frac{\Delta U_i}{(U_i + N_u d_v / R\gamma)}| < \frac{\epsilon}{2},$$

which eventually yields the following conditions

$$|\frac{\Delta U_i}{U_i}| < \frac{\epsilon}{2}, \tag{B.2.4a}$$
$$|\frac{\Delta V_i}{V_i}| < \frac{\epsilon}{2}. \tag{B.2.4b}$$

Then, as suggested by [21], bounding the mean, μ_i, and standard deviation, σ_i^2, of the *expected* differences, ΔU_i and ΔV_i, leads to the satisfaction of conditions (B.2.4). We find these quantities for $j = i \in [1, W]$

$$\mu_j = -U_i[a_i^+(t) + a_i^-(t)]\tau + a_{i-1}^+(t)U_{i-1}\tau + a_{i+1}^-(t)U_{i+1}\tau, \quad \text{(B.2.5a)}$$
$$\sigma_j^2 = U_i[a_i^+(t) + a_i^-(t)]\tau + a_{i-1}^+(t)U_{i-1}\tau + a_{i+1}^-(t)U_{i+1}\tau, \quad \text{(B.2.5b)}$$

and for $j = (i + W) \in [W + 1, 2W]$

$$\mu_j = -V_i[r_i^+(t) + r_i^-(t)]\tau + r_{i-1}^+(t)V_{i-1}\tau + a_{i+1}^-(t)V_{i+1}\tau, \quad \text{(B.2.6a)}$$
$$\sigma_j^2 = V_i[a_i^+(t) + a_i^-(t)]\tau + a_{i-1}^+(t)V_{i-1}\tau + a_{i+1}^-(t)V_{i+1}\tau, \quad \text{(B.2.6b)}$$

taking into account that U and V are periodic (see Eqn. (5.2.2)).

Now, we define $X_j \in [U, V]$ for $j = 1, \cdots, 2W$, where $[U, V]$ is the horizontal concatenation of the two vectors U and V. We bound μ_j and σ_j^2 by $|\mu_j| < \epsilon X_j/2$ and $\sigma_j^2 < (\epsilon X_j/2)^2$, which finally yields the condition for selecting a suitable leap τ in time:

$$\tau = \min_j \left\{ \frac{\max\{\epsilon X_j/2, 1\}}{|\hat{\mu_j}|}, \frac{\max\{\epsilon X_j/2, 1\}^2}{\hat{\sigma_j}^2} \right\}, \quad \text{(B.2.7)}$$

where $\hat{\mu_j}$ and $\hat{\sigma_j}^2$ are determined from Eqns. (B.2.5) and (B.2.6) by $\hat{\mu_j} = \mu_j/\tau$ and $\hat{\sigma_j}^2 = \sigma_j^2/\tau$. The max functions in Eqn. (B.2.7) are used to ensure that the selected τ is big enough such that the minimum expected difference is one at each site [21].

Having found τ, we substitute it into Eqn. (B.2.1) and run the τ-leaping Gillespie algorithm. We have chosen $\epsilon = 0.1$ for our simulations, as it provides reasonable simulation times and does not induce high inaccuracies - we repeated some of the simulations for significantly lower values of ϵ and obtained fairly similar results.

Using the linear stability analysis, outlined in §5.5, we select a set of parameter values that makes $m = 9$ the unstable mode with the highest growth rate, $\Re(\lambda)$; the values of the parameters are: $\alpha = 111, \gamma = 87, Q = 1, D_u = D_v = 6.6e^{-5}, X = 1000/16, \delta = 0.1$ and $T = 100$. Fig. B.1a indicates that $\Re(\lambda)$ decays with n, implying that higher modes are more stable except for the four modes $n = 9, n = 18, n = 27, n = 36$ and $n = 45$ where $\Re(\lambda)$ spikes. Fig. B.1b shows that at these five modes, $|\epsilon_u|$ have quite small values - note that the global minimum has occurred in $|\epsilon_u(9)|$. Conversely, $|\epsilon_v|$ is large at the five modes, shown in Fig. B.1c, meaning that mode 9 in $\langle V \rangle$ would emerge with such a larger amplitude than $\langle U \rangle$ in these modes.

We now simulate the stochastic model on a finer lattice, $W = 1000$, and examine the emergence of mode $m = 9$; this mode failed to emerge for lattice

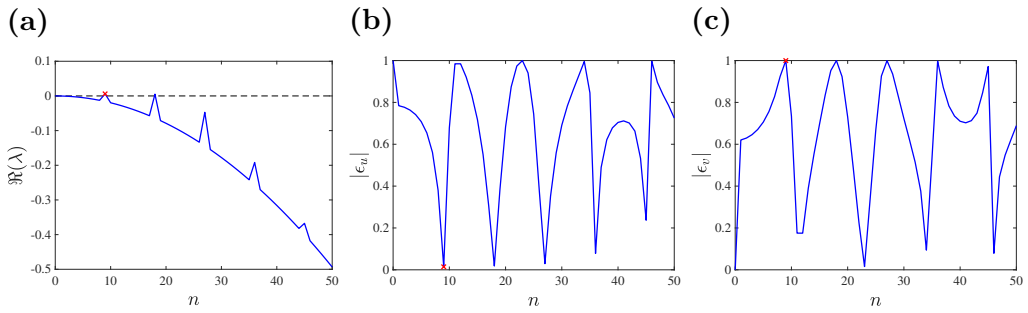


Figure B.1: The behaviour of the system about equilibrium: (a) $\Re(\lambda)$, (b) $|\epsilon_u|$ and (c) $|\epsilon_v|$ versus modes n. The red cross show the unstable mode with highest growth rate, which is $m = 9$ here for the selected values of the parameters: $\alpha = 111, \gamma = 87, Q = 1, D_u = D_v = 6.6e^{-5}, X = 1000/16, \delta = 0.1$ and $T = 100$.

size of $W = 100$, see Fig. 5.14. Using the τ-leaping Gillespie algorithm, we generate 500 simulations of the stochastic model with the parameter values of $A = R = 1$ and $d_u = d_v = 0.05$ corresponding to the earlier mentioned parameter values of the continuous model. Similar to §5.6, we consider single-mode perturbation for the initial distributions. As Fig. B.2 shows, the initial distributions of $\langle U \rangle$ and $\langle V \rangle$ are equal and contains only mode 9. Then, similar to the results for $W = 100$, shown in Fig. 5.8, $\langle V \rangle$ starts to decay first and then grows in the opposite direction, forming the striped pattern.

Therefore, as the results show, finer lattices must be used when we aim to initiate the emergence of a relatively large unstable wavenumber. As we showed here, mode $m = 9$ failed to emerge in a lattice of size $W = 100$, but successfully emerged in the one with $W = 1000$. Running simulations on lattices with such a big size has become feasible only with the implemented approximate simulation algorithm, τ-leaping Gillespie algorithm. Simulations of the model with uniform initial distributions are left as future work.

B.3 Two-dimensional lattices

In this section, we extend the continuous approximations undertaken in §5.2 to two-dimensional domains. We have a $W \times H$ two-dimensional lattice, where W and H are the number of columns and rows of the lattice made up of sites with equal area δ^2; thus, $X = \delta W$ is the width and $Y = \delta H$ is the height of the continuous domain. The number of chaser cells at i^{th} column and j^{th} row, *i.e.* site (i, j), at time t is defined by $U_{i,j}(t)$ and the number of

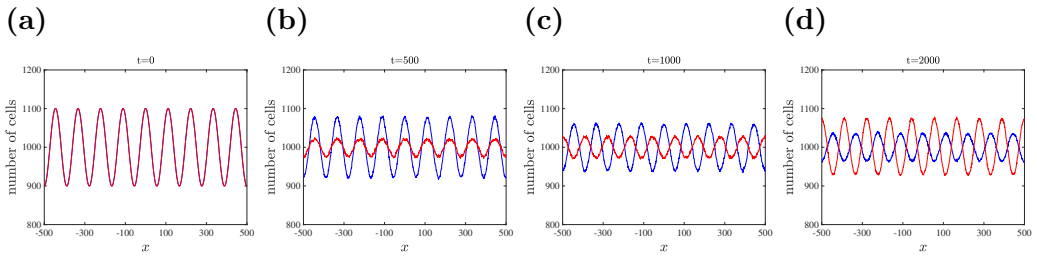


Figure B.2: Simulation of the stochastic model for (a) equal initial distributions containing only mode 9 as the perturbation. (b-d) the patterns at times $t = \{500, 1000, 2000\}$. The values of the parameters are $\alpha = 111, \gamma = 87$, $A = R = 1$, $d_u = d_v = 0.05$, $W = 1000$. The red and blue curves show the average number of runners, $\langle V \rangle$, and chasers, $\langle U \rangle$, respectively, over 500 simulations.

runner cells at the same site and time is $V_{i,j}(t)$. Similar to the method in Chapter 4, we define the rate of stepping right as

$$a_{i,j}^{(+,\cdot)}(t) = d_u + \frac{A}{N_v} \sum_{a=1}^{\alpha} \sum_{b=-a}^{a} V_{i+a,j+b}(t). \tag{B.3.1}$$

Similarly the rate of stepping left is define by

$$a_{i,j}^{(-,\cdot)}(t) = d_u + \frac{A}{N_v} \sum_{a=-\alpha}^{-1} \sum_{b=a}^{-a} V_{i+a,j+b}(t). \tag{B.3.2}$$

Then, as the cells can move to the up and down sites, we define the rates of stepping up and down by $O_{i,j}$ and $D_{i,j}$ similar to Eqns. (B.3.1) and (B.3.2) by

$$a_{i,j}^{(\cdot,+)}(t) = d_u + \frac{A}{N_v} \sum_{b=1}^{\alpha} \sum_{a=-b}^{b} V_{i+a,j+b}(t),$$

$$a_{i,j}^{(\cdot,-)}(t) = d_u + \frac{A}{N_v} \sum_{b=-\alpha}^{-1} \sum_{a=b}^{-b} V_{i+a,j+b}(t).$$

Note that the cells can ideally move to the corner adjacent sites, *i.e.* the sites at up-right, up-left, and so on, but as concurrence of two motion events are required to reach these sites, we consider them unlikely to happen. Similar formulae are used for the rates of the runners but not shown here for brevity.

By considering the same assumptions as in §5.2, and introducing cell density $u_{i,j}(t) = U_{i,j}(t)/N_u$, the equation describing the evolution of the distribution of cells over time for the two-dimensional lattice can be found as

$$\begin{aligned}\frac{\mathrm{d}\langle u_{i,j}(t)\rangle}{\mathrm{d}t} &= \langle a^{(+,\cdot)}_{i-1,j}(t)\rangle\langle u_{i-1,j}(t)\rangle + \langle L_{i+1,j}(t)\rangle\langle u_{i+1,j}(t)\rangle \\ &- \langle R_{i,j}(t) + L_{i,j}(t)\rangle\langle u_{i,j}(t)\rangle \\ &+ \langle O_{i,j-1}(t)\rangle\langle u_{i,j-1}(t)\rangle + \langle D_{i,j+1}(t)\rangle\langle u_{i+1,j}(t)\rangle \\ &- \langle O_{i,j}(t) + D_{i,j}(t)\rangle\langle u_{i,j}(t)\rangle.\end{aligned} \tag{B.3.3}$$

Then, by assuming that the lattice sites are very small and the number of cells approaches infinity, we introduce the continuous variable $u(x, y, t)$ by

$$u_{i,j}(t) = \int_{x_i-\frac{\delta}{2}}^{x_i+\frac{\delta}{2}} \int_{y_j-\frac{\delta}{2}}^{y_j+\frac{\delta}{2}} u(x, y, t)\ \mathrm{d}y\mathrm{d}x \approx \delta^2 u(x_i, y_j, t), \tag{B.3.4}$$

in which

$$x_i = i\delta, y_j = j\delta, \quad (i, j) \in \{1, 2, \cdots, W\} \times \{1, 2, \cdots, H\},$$

where we then rename x_i and y_i to x and y, respectively, centre the lattice on $(x, y) = (0, 0)$, and substitute the discrete variables in Eqn. (B.3.3) with the continuous ones. Using the Taylor series expansion of the variables and ignoring the terms of order $\mathcal{O}(\delta^3)$ and higher, we get

$$\frac{\partial u}{\partial t} = -\delta\left(\frac{\partial u(R-L)}{\partial x} + \frac{\partial u(U-D)}{\partial y}\right) + \delta^2 d_u\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right), \tag{B.3.5}$$

where we do not show the space and time arguments, *i.e.* x, y and t, of the variables for the ease of notation. The second term in the RHS of Eqn. (B.3.5) describes unbiased diffusion in a two-dimensional domain. The first term can be represented as a non-local advection term, as explained in §5.3, which consequently yields the final continuous model

$$\frac{\partial u}{\partial t} + \nabla.\left(\{\mathbf{K}_a \star v\}\ u\right) - D_u\nabla^2 u = 0, \tag{B.3.6a}$$

$$\frac{\partial v}{\partial t} + \nabla.\left(\{\mathbf{K}_r \star u\}\ v\right) - D_v\nabla^2 v = 0, \tag{B.3.6b}$$

where $D_u = \delta^2 d_u$, $D_v = \delta^2 d_v$, $\mathbf{K}_a(x, y) = A\delta\mathbf{K}(x, y, \alpha)$, $\mathbf{K}_r(x) = -R\delta\mathbf{K}(x, y, \gamma)$

where $\mathbf{K}(x,y,l)$ over $x \in [-X/2, X/2]$ and $y \in [-Y/2, Y/2]$ is given by

$$\mathbf{K}(x,y,l) = \begin{cases} \begin{pmatrix} -1 \\ 0 \end{pmatrix} & \text{for} \quad x \in [\delta/2, (l+0.5)\delta], y \in [-(l+0.5)\delta, (l+0.5)\delta], \\ \begin{pmatrix} 1 \\ 0 \end{pmatrix} & \text{for} \quad x \in [\delta/2, (l+0.5)\delta], y \in [-(l+0.5)\delta, (l+0.5)\delta], \\ \begin{pmatrix} 0 \\ -1 \end{pmatrix} & \text{for} \quad x \in [\delta/2, (l+0.5)\delta], y \in [-(l+0.5)\delta, (l+0.5)\delta], \\ \begin{pmatrix} 0 \\ 1 \end{pmatrix} & \text{for} \quad x \in [\delta/2, (l+0.5)\delta], y \in [-(l+0.5)\delta, (l+0.5)\delta], \\ 0 & \text{otherwise}, \end{cases} \tag{B.3.7}$$

which is then extended to be periodic over $x \in [-X, X]$ and $y \in [-Y, Y]$, *i.e.* $\mathbf{K}(x \pm X, y, l) = \mathbf{K}(x, y \pm Y, l) = \mathbf{K}(x \pm X, y \pm Y, l) = \mathbf{K}(x, y, l)$.

We then non-dimensionalise the PDEs described in Eqns. (B.3.6a) and (B.3.6b) with the same non-dimensionalising variables as in §5.3, and obtain the following model (dropping tildes)

$$\frac{\partial u}{\partial t} + Q\nabla.\left(\{\mathbf{K} \star v\}\, u\right) - D_u \nabla^2 u = 0, \tag{B.3.8a}$$

$$\frac{\partial v}{\partial t} - \nabla.\left(\{\mathbf{K} \star u\}\, v\right) - D_v \nabla^2 v = 0, \tag{B.3.8b}$$

where the non-dimensional kernel is given by

$$\mathbf{K}(x,y,l) = \begin{cases} \begin{pmatrix} -1 \\ 0 \end{pmatrix} & \text{for} \quad x \in [0.5, (l+0.5)]/\gamma, y \in [-(l+0.5), (l+0.5)]/\gamma, \\ \begin{pmatrix} 1 \\ 0 \end{pmatrix} & \text{for} \quad x \in [0.5, (l+0.5)]/\gamma, y \in [-(l+0.5), (l+0.5)]/\gamma, \\ \begin{pmatrix} 0 \\ -1 \end{pmatrix} & \text{for} \quad x \in [0.5, (l+0.5)]/\gamma, y \in [-(l+0.5), (l+0.5)]/\gamma, \\ \begin{pmatrix} 0 \\ 1 \end{pmatrix} & \text{for} \quad x \in [0.5, (l+0.5)]/\gamma, y \in [-(l+0.5), (l+0.5)]/\gamma, \\ 0 & \text{otherwise}, \end{cases} \tag{B.3.9}$$

which is extended periodically as explained for Eqn. (B.3.7).

We then assess the behaviour of the system about the equilibrium using the linear stability analysis. Similar to the presented calculations in §5.5, we find the stability condition for the model in two-dimensional domain, given by

$$h(n,p) = \mathbf{q}^4 - \frac{Q}{D_u D_v}(\hat{\mathbf{K}} \cdot \mathbf{q})(\hat{\mathbf{K}} \cdot \mathbf{q}) < 0,$$

where $\mathbf{q} = [q_x \; q_y]^T$; q_x and q_y are the wavenumbers of x and y directions respectively; $\hat{\mathbf{K}}$ is the Fourier transform of the kernel, Eqn. (B.3.9). By defining $\hat{\mathbf{K}} = [\hat{K}_x \; \hat{K}_y]^T$, we find

$$\hat{K}_x(q_x, q_y, l) = F(q_x, q_y, \frac{l+0.5}{\gamma}) - F(q_x, q_y, \frac{0.5}{\gamma}),$$
$$\hat{K}_y(q_x, q_y, l) = F(q_y, q_x, \frac{l+0.5}{\gamma}) - F(q_y, q_x, \frac{0.5}{\gamma}),$$

where $F(q_1, q_2, z)$ is given by

$$F(q_1, q_2, z) = \frac{-2i}{q_2(q_1+q_2)} \sin[(q_1+q_2)z] + \frac{2i}{q_2(q_1-q_2)} \sin[(q_1-q_2)z].$$

Finally, by imposing the boundary conditions, we get

$$q_x = q_x(n) = \frac{2n\pi}{X}, \quad q_y = q_y(p) = \frac{2p\pi}{Y}, \quad \text{for} \quad n, p \in \mathbb{Z}.$$

For the sake of illustration, we picked a set of parameter values, $Q = 1, D_u = D_v = 0.011, X = Y = 20/3, T = 2, L = 3\delta$, that initiates the emergence of mode (3,3), *i.e.* we expect to see spotty patterns with wa. As explained by Murray (2008), we note that horizontal (vertical) stripes can be formed if we consider $H > W$ ($W > H$). These stripes are in the opposite direction of what emerges in zebrafish (in zebrafish $W > H$ and the stripes are horizontal). Here, we restrict our attention to the ability of run-and-chase in producing anti-phase distributions, similar to [130], many underlying influential biological factors such as domain growth [177] are not considered here for simplicity that could lead to significant differences. Thus, in this work, we do not concern ourselves with the direction of the stripes.

The cells are initially distributed uniformly on the domain, as shown in Figs. B.3a and B.3b. As Figs. (c-f) show, the dominant mode (3,3) has appeared in the numerical solution of the PDEs and the distribution of the two species are anti-phase, as predicted by the linear stability analysis (see the details of the used numerical solver in §5.3.2). Simulating the stochastic model on a two-dimensional lattice will provide an opportunity to investigate the extent to which the results of the two models agree, similar to the analysis undertaken in §5.5. This is left for a future work.

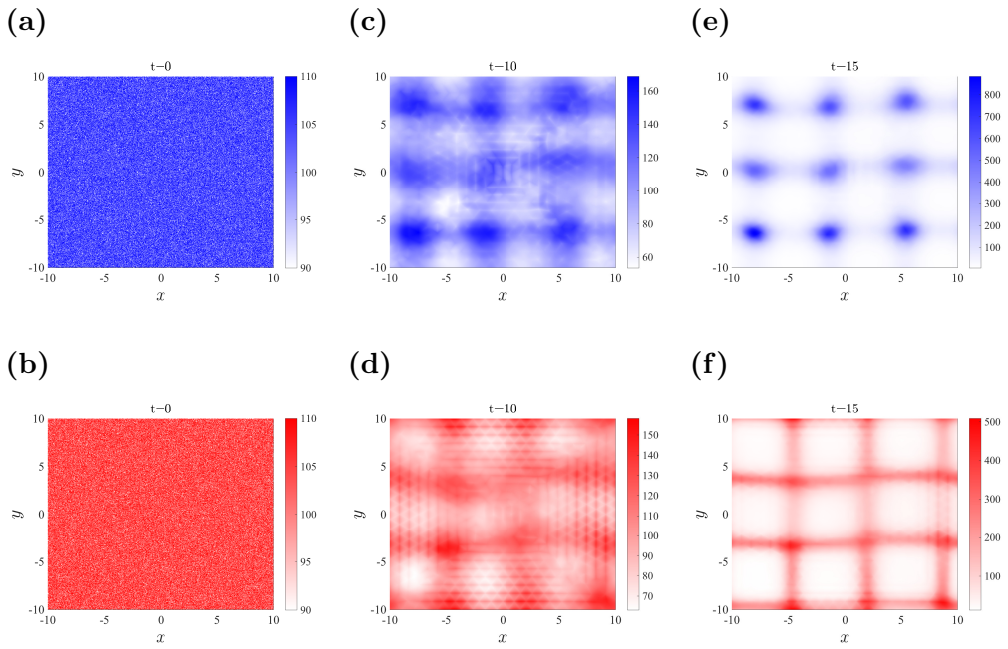


Figure B.3: Numerical solution of the continuous model with a two-dimensional domain. The first and second rows illustrate the solutions of U and V respectively. The columns correspond to the solutions at times $t = \{0, 10, 15\}$ from left to right. The parameter values are: $Q = 1, D_u = D_v = 0.011, X = Y = 20/3, T = 2, L = 3\delta, \delta = 1$.

Bibliography

[1] 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: 022705, 2014.

[2] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell*. New York: Garland Science, fourth edition, 2002.

[3] L. Angelani. Collective Predation and Escape Strategies. *Physical Review Letters*, 109(11):118104, 2012.

[4] R. Araujo and D. McElwain. A history of the study of solid tumour growth: the contribution of mathematical modelling. *Bulletin of Mathematical Biology*, 66(5):1039–1091, 2004.

[5] S. F. Badylak and R. M. Nerem. Progress in tissue engineering and regenerative medicine. *Proceedings of the National Academy of Sciences of the United States of America*, 107(8):3285–6, 2010.

[6] V. Balbi, E. Kuhl, and P. Ciarletta. Morphoelastic control of gastrointestinal organogenesis: Theoretical predictions and numerical insights. *Journal of the Mechanics and Physics of Solids*, 78:493–510, 2015.

[7] P. Ball. Pattern formation in nature: Physical constraints and self-organising characteristics. *Architectural Design*, 82(2):22–27, 2012.

[8] J. Besag. Contribution to the discussion of dr. ripleys paper. *Journal of the Royal Statistical Society: Series B*, 39(2):193–195, 1977.

[9] P. Bianco and P. G. Robey. Stem cells in tissue engineering. *Nature*, 414(6859):118–121, 2001.

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

[11] 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.

[12] 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):474490, 2012.

[13] B. J. Binder, J. F. Sundstrom, J. M. Gardner, V. Jiranek, and S. G. Oliver. Quantifying two-dimensional filamentous and invasive growth spatial patterns in yeast colonies. *PLoS Computational Biology*, 11(2): e1004070–e1004070, 2015.

[14] R. N. Binny, M. J. Plank, and A. James. Spatial moment dynamics for collective cell movement incorporating a neighbour-dependent directional bias. *Journal of The Royal Society Interface*, 12(106):20150228, 2015.

[15] S. Blacher, C. Erpicum, B. Lenoir, J. Paupert, G. Moraes, S. Ormenese, E. Bullinger, and A. Noel. Cell Invasion in the Spheroid Sprouting Assay: A Spatial Organisation Analysis Adaptable to Cell Behaviour. *PLoS ONE*, 9(5):e97019, 2014.

[16] M. Bodnar and J. J. L. Velazquez. Derivation of macroscopic equations for individual cell-based models: a formal approach. *Mathematical Methods in the Applied Sciences*, 28(15):1757–1779, 2005.

[17] C. Brown, D. Burslem, J. Illian, L. Bao, W. Brockelman, M. Cao, L. Chang, H. Dattaraja, S. Davies, C. Gunatilleke, *et al.* Multispecies coexistence of trees in tropical forests: spatial signals of topographic niche differentiation increase with environmental heterogeneity. *Proceedings of the Royal Society B: Biological Sciences*, 280(1764): 20130502, 2013.

[18] J. Buhl, D. J. T. Sumpter, I. D. Couzin, J. J. Hale, E. Despland, E. R. Miller, and S. J. Simpson. From disorder to order in marching locusts. *Science (New York, N.Y.)*, 312(5778):1402–1406, 2006.

[19] D. Bullara and Y. De Decker. Pigment cell movement is not required for generation of Turing patterns in zebrafish skin. *Nature communications*, 6(May):6971, 2015.

[20] P. Buske, J. Przybilla, M. Loeffler, N. Sachs, T. Sato, H. Clevers, and J. Galle. On the biomechanics of stem cell niche formation in the gut - modelling growing organoids. *FEBS Journal*, 279(18):3475–3487, 2012.

[21] Y. Cao, D. T. Gillespie, and L. R. Petzold. Efficient step size selection for the tau-leaping simulation method. *The Journal of Chemical Physics*, 124(4):044109, 2006.

[22] J. Carlsson and J. M. Yuhas. Liquid-Overlay Culture of Cellular Spheroids. In H. Acker, J. Carlsson, R. Durand, and R. M. Sutherland, editors, *Spheroids in Cancer Research: Methods and Perspectives*, pages 1–23. Springer Berlin Heidelberg, Berlin, Heidelberg, 1984.

[23] J. A. Carrillo, Y.-P. Choi, and S. Pérez. A review on attractive-repulsive hydrodynamics for consensus in collective behavior. 2016.

[24] J. J. Casciari, S. V. Sotirchos, and R. M. Sutherland. Mathematical modelling of microenvironment and growth in EMT6/Ro multicellular tumour spheroids. *Cell Proliferation*, 25(1):1–22, 1992.

[25] D. Chandler and J. K. Percus. Introduction to Modern Statistical Mechanics. *Physics Today*, 41(12):114, 1988.

[26] Y. Chen and T. Kolokolnikov. A minimal model of predator-swarm dynamics. *Journal of the Royal Society Interface*, I(1):20131208, 2014.

[27] S. N. Chiu, D. Stoyan, W. S. Kendall, and J. Mecke. *Stochastic Geometry and Its Applications*. John Wiley & Sons, third edition, 2013.

[28] I. D. Couzin, J. Krause, R. James, G. D. Ruxton, and N. R. Franks. Collective memory and spatial sorting in animal groups. *Journal of Theoretical Biology*, 218(1):1–11, 2002.

[29] S. C. Cowin. How Is a Tissue Built? *Journal of Biomechanical Engineering*, 122(6):553, 2000.

[30] X. Cui, S. Dini, S. Dai, J. Bi, B. J. Binder, J. E. F. Green, and H. Zhang. A mechanistic study on tumour spheroid formation in thermosensitive hydrogels: experiments and mathematical modelling. *RSC Adv.*, 6(77):73282–73291, 2016.

[31] S. H. Dairkee, G. Deng, M. R. Stampfer, F. M. Waldman, and H. S. Smith. Selective Cell Culture of Primary Breast Carcinoma. *Cancer Research*, 55(12):2516–2519, 1995.

[32] K. Davies, J. Green, N. Bean, B. Binder, and J. Ross. On the derivation of approximations to cellular automata models and the assumption of independence. *Mathematical Biosciences*, 253:63–71, 2014.

[33] K. J. Davies. *On the derivation and application of closure approximations of cellular automata models*. PhD thesis, 2016.

[34] M. De Berg, M. Van Kreveld, M. Overmars, and O. C. Schwarzkopf. *Computational Geometry*. Springer, 2000.

[35] T. Deisboeck, M. Berens, A. Kansal, S. Torquato, A. Stemmer-Rachamimov, and E. Chiocca. Pattern of self-organization in tumour systems: complex growth dynamics in a novel brain tumour spheroid model. *Cell Proliferation*, 34(2):115–134, 2001.

[36] U. Dieckmann, R. Law, and J. A. Metz. *The geometry of ecological interactions: simplifying spatial complexity*. Cambridge University Press, 2000.

[37] P. J. Diggle *et al. Statistical analysis of spatial point patterns*. Academic Press, 1983.

[38] S. Dini, B. J. Binder, S. C. Fischer, C. Mattheyer, A. Schmitz, E. H. K. Stelzer, N. G. Bean, and J. E. F. Green. Identifying the necrotic zone boundary in tumour spheroids with pair-correlation functions. *Journal of The Royal Society Interface*, 13(123):1–20, 2016.

[39] S. Dini, B. J. Binder, and J. E. F. Green. Understanding interactions between populations: Individual based modelling and quantification using pair correlation functions. *Journal of Theoretical Biology*, 439: 50–64, 2018.

[40] W. A. N. Dorland. *Dorland's Illustrated Medical Dictionary*. Philadelphia, Pa. : Elsevier/Saunders, c2012., 32nd edition, 2012.

[41] D. Drasdo and M. Loeffler. Individual-based models to growth and folding in one-layered tissues: intestinal crypts and early development. *Nonlinear Analysis: Theory, Methods & Applications*, 47(1):245–256, 2001.

[42] I. Dufau, C. Frongia, F. Sicard, L. Dedieu, P. Cordelier, F. Ausseil, B. Ducommun, and A. Valette. Multicellular tumor spheroid model to evaluate spatio-temporal dynamics effect of chemotherapeutics: application to the gemcitabine/CHK1 inhibitor combination in pancreatic cancer. *BMC Cancer*, 12(1):15, 2012.

[43] R. J. Dyson, J. E. F. Green, J. P. Whiteley, and H. M. Byrne. An investigation of the influence of extracellular matrix anisotropy and cellmatrix interactions on tissue architecture. *Journal of Mathematical Biology*, 72(7):1775–1809, 2016.

[44] R. Eftimie, G. de Vries, and M. A. Lewis. Complex spatial group patterns result from different animal communication mechanisms. *Proceedings of the National Academy of Sciences of the United States of America*, 104(17):6974–9, 2007.

[45] R. Erban, J. Chapman, and P. Maini. A practical guide to stochastic simulations of reaction-diffusion processes. *arXiv preprint arXiv:0704.1908*, page 35, 2007.

[46] G. B. Ermentrout and L. Edelstein-Keshet. Cellular Automata Approaches to Biological Modeling. *Journal of Theoretical Biology*, 160 (1):97–133, 1993.

[47] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD)*, volume 96, pages 226–231, 1996.

[48] J. H. Evers, R. C. Fetecau, and T. Kolokolnikov. Equilibria for an aggregation model with two species. *arXiv preprint arXiv:1612.08074*, 2016.

[49] R. Fetecau and Y. Huang. Equilibria of biological aggregations with nonlocal repulsiveattractive interactions. *Physica D: Nonlinear Phenomena*, 260:49–64, 2013.

[50] R. C. Fetecau, Y. Huang, and T. Kolokolnikov. Swarm dynamics and equilibria for a nonlocal aggregation model. *Nonlinearity*, 24(10):2681–2716, 2011.

[51] G. Flierl, D. Grünbaum, S. Levins, and D. Olson. From Individuals to Aggregations: the Interplay between Behavior and Physics. *Journal of Theoretical Biology*, 196(4):397–454, 1999.

[52] J. A. Fozard, G. R. Kirkham, L. D. Buttery, J. R. King, O. E. Jensen, and H. M. Byrne. Techniques for analysing pattern formation in populations of stem cells and their progeny. *BMC Bioinformatics*, 12(1): 396, 2011.

[53] H. G. Frohnhöfer, J. Krauss, H.-M. Maischein, and C. Nüssslein-Volhard. Iridophores and their interactions with other chromatophores are required for stripe formation in zebrafish. *Development (Cambridge, England)*, 140(14):2997–3007, 2013.

[54] E. A. Gaffney and S. Seirin Lee. The sensitivity of turing self-organization to biological feedback delays: 2D models of fish pigmentation. *Mathematical Medicine and Biology*, 32(1):56–78, 2015.

[55] S. Garg, S. Fischer, E. Schuman, and E. Stelzer. Lateral assembly of n-cadherin drives tissue integrity by stabilizing adherens junctions. *Journal of The Royal Society Interface*, 12(104):20141055, 2015.

[56] A. Gerisch and M. Chaplain. Mathematical modelling of cancer cell invasion of tissue: Local and non-local models and the effect of adhesion. *Journal of Theoretical Biology*, 250(4):684–704, 2008.

[57] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434, 1976.

[58] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[59] D. T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, 115(4):1716–1733, 2001.

[60] D. T. Gillespie and L. R. Petzold. Improved leap-size selection for accelerated stochastic simulation. *The Journal of Chemical Physics*, 119(16):8229, 2003.

[61] A. P. Gilmore. Anoikis. *Cell Death and Differentiation*, 12:1473–1477, 2005.

[62] P. Godoy, N. J. Hewitt, U. Albrecht, M. E. Andersen, N. Ansari, S. Bhattacharya, J. G. Bode, J. Bolleyn, C. Borner, J. Böttger, A. Braeuning, R. A. Budinsky, B. Burkhardt, N. R. Cameron, G. Camussi, C.-S. Cho, Y.-J. Choi, J. Craig Rowlands, U. Dahmen, G. Damm,

O. Dirsch, M. T. Donato, J. Dong, S. Dooley, D. Drasdo, R. Eakins, K. S. Ferreira, V. Fonsato, J. Fraczek, R. Gebhardt, A. Gibson, M. Glanemann, C. E. P. Goldring, M. J. Gómez-Lechón, G. M. M. Groothuis, L. Gustavsson, C. Guyot, D. Hallifax, S. Hammad, A. Hayward, D. Häussinger, C. Hellerbrand, P. Hewitt, S. Hoehme, H.-G. Holzhütter, J. B. Houston, J. Hrach, K. Ito, H. Jaeschke, V. Keitel, J. M. Kelm, B. Kevin Park, C. Kordes, G. A. Kullak-Ublick, E. L. LeCluyse, P. Lu, J. Luebke-Wheeler, A. Lutz, D. J. Maltman, M. Matz-Soja, P. McMullen, I. Merfort, S. Messner, C. Meyer, J. Mwinyi, D. J. Naisbitt, A. K. Nussler, P. Olinga, F. Pampaloni, J. Pi, L. Pluta, S. A. Przyborski, A. Ramachandran, V. Rogiers, C. Rowe, C. Schelcher, K. Schmich, M. Schwarz, B. Singh, E. H. K. Stelzer, B. Stieger, R. Stöber, Y. Sugiyama, C. Tetta, W. E. Thasler, T. Vanhaecke, M. Vinken, T. S. Weiss, A. Widera, C. G. Woods, J. J. Xu, K. M. Yarborough, and J. G. Hengstler. Recent advances in 2D and 3D in vitro systems using primary hepatocytes, alternative hepatocyte sources and non-parenchymal liver cells and their use in investigating mechanisms of hepatotoxicity, cell signaling and ADME. *Archives of Toxicology*, 87(8):1315–1530, 2013.

[63] J. E. F. Green, S. L. Waters, K. M. Shakesheff, and H. M. Byrne. A mathematical model of liver cell aggregation in vitro. *Bulletin of Mathematical Biology*, 71(4):906930, 2009.

[64] J. E. F. Green, S. L. Waters, J. P. Whiteley, L. Edelstein-Keshet, K. M. Shakesheff, and H. M. Byrne. Non-local models for the formation of hepatocyte-stellate cell aggregates. *Journal of Theoretical Biology*, 267 (1):106–120, 2010.

[65] H. P. Greenspan. Models for the Growth of a Solid Tumor by Diffusion. *Studies in Applied Mathematics*, 51(4):317–340, 1972.

[66] P. Greig-Smith. *Quantitative Plant Ecology*. University of California Press, third edition, 1983.

[67] R. Grima and T. J. Newman. Accurate discretization of advection-diffusion equations. *Physical Review E*, 70(3):036703, 2004.

[68] D. R. Grimes, C. Kelly, K. Bloch, and M. Partridge. A method for estimating the oxygen consumption rate in multicellular tumour spheroids. *Journal of The Royal Society Interface*, 11(92):20131124, 2014.

[69] B. M. Gumbiner. Cell adhesion: The molecular basis of tissue architecture and morphogenesis. *Cell*, 84(3):345 – 357, 1996.

[70] E. J. Hackett-Jones, K. A. Landman, and K. Fellner. Aggregation patterns from nonlocal interactions: Discrete stochastic and continuum modeling. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 85(4), 2012.

[71] H. Hamada, M. Watanabe, H. E. Lau, T. Nishida, T. Hasegawa, D. M. Parichy, and S. Kondo. Involvement of Delta/Notch signaling in zebrafish adult pigment stripe patterning. *Development (Cambridge, England)*, 141(2):318–24, 2014.

[72] J. A. Hartigan and M. A. Wong. Algorithm AS 136: A K-Means Clustering Algorithm. *Applied Statistics*, 28(1):100–108, 1979.

[73] J. Hendriks, J. Riesle, and C. A. van Blitterswijk. Co-culture in cartilage tissue engineering. *Journal of Tissue Engineering and Regenerative Medicine*, 1(3):170–178, 2007.

[74] F. Hirschhaeuser, H. Menne, C. Dittfeld, J. West, W. Mueller-Klieser, and L. A. Kunz-Schughart. Multicellular tumor spheroids: an underestimated tool is catching up again. *Journal of Biotechnology*, 148(1): 3–15, 2010.

[75] D. Horstmann, K. J. Painter, and H. G. Othmer. Aggregation under local reinforcement: From lattice to continuum. *European Journal of Applied Mathematics*, 15(5):545–576, oct 2004.

[76] B. D. Hughes and K. Fellner. Continuum models of cohesive stochastic swarms: The effect of motility on aggregation patterns. *Physica D: Nonlinear Phenomena*, 260:26–48, 2013.

[77] J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan. *Statistical Analysis and Modelling of Spatial Point Patterns*, volume 70. John Wiley & Sons, 2008.

[78] M. Inaba, H. Yamanaka, and S. Kondo. Pigment pattern formation by contact-dependent depolarization. *Science*, 335(6069):677, 2012.

[79] S. Inoue, S. Kondo, D. M. Parichy, and M. Watanabe. Tetraspanin 3c requirement for pigment cell interactions and boundary formation in zebrafish adult pigment stripes. *Pigment Cell and Melanoma Research*, 27(2):190–200, 2014.

[80] A. K. Jain and R. C. Dubes. *Algorithms for Clustering Data*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1988.

[81] Y. Jiao and S. Torquato. Emergent Behaviors from a Cellular Automaton Model for Invasive Tumor Growth in Heterogeneous Microenvironments. *PLoS Computational Biology*, 7(12):e1002314, 2011.

[82] S. T. Johnston, M. J. Simpson, and R. E. Baker. Mean-field descriptions of collective migration with strong adhesion. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 85(5), 2012.

[83] S. T. Johnston, M. J. Simpson, D. 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.

[84] I. Jolliffe. *Principal Component Analysis*. John Wiley & Sons, Ltd, 2014.

[85] B. J. Jones, V. J. Martínez, E. Saar, and V. Trimble. Scaling laws in the distribution of galaxies. *Reviews of Modern Physics*, 76(4):1211, 2005.

[86] C. Kaito, U. Dieckmann, A. Sasaki, and F. Takasu. Beyond pairs: Definition and interpretation of third-order structure in spatial point patterns. *Journal of Theoretical Biology*, 372:22–38, 2015.

[87] E. F. Keller and L. A. Segel. Initiation of slime mold aggregation viewed as an instability. *Journal of Theoretical Biology*, 26(3):399–415, 1970.

[88] E. F. Keller and L. A. Segel. Initiation of slime mold aggregation viewed as an instability. *Journal of Theoretical Biology*, 26(3):399415, 1970.

[89] P. J. Keller, A. D. Schmidt, J. Wittbrodt, and E. H. Stelzer. Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy. *Science*, 322(5904):1065–1069, 2008.

[90] D. C. Krakauer. Groups confuse predators by exploiting perceptual bottlenecks: a connectionist model of the confusion effect. *Behavioral Ecology and Sociobiology*, 36(6):421–429, 1995.

[91] J. Krause and G. D. Ruxton. *Living in groups*. Oxford University Press, 2002.

[92] G. Lan, S. Getzin, T. Wiegand, Y. Hu, G. Xie, H. Zhu, and M. Cao. Spatial distribution and interspecific associations of tree species in a tropical seasonal rain forest of china. *PloS ONE*, 7(9):e46074, 2012.

[93] M. A. Lancaster and J. A. Knoblich. Organogenesis in a dish: Modeling development and disease using organoid technologies. *Science*, 345 (6194), 2014.

[94] R. Lanza, R. Langer, and J. P. Vacanti. *Principles of Tissue Engineering*. Academic press, third edition, 2011.

[95] C. T. Lee, M. F. Hoopes, J. Diehl, W. Gilliland, G. Huxel, E. V. Leaver, K. S. McCann, J. Umbanhowar, and A. Mogilner. Non-local concepts and models in biology. *Journal of Theoretical biology*, 210(2):201–219, 2001.

[96] G. Lemon, J. R. King, H. M. Byrne, O. E. Jensen, and K. M. Shakesheff. Mathematical modelling of engineered tissue growth using a multiphase porous flow mixture theory. *Journal of Mathematical Biology*, 52(5):571–594, 2006.

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

[98] D. Loessner, K. S. Stok, M. P. Lutolf, D. W. Hutmacher, J. A. Clements, and S. C. Rizzi. Bioengineered 3d platform to explore cell–ecm interactions and drug resistance of epithelial ovarian cancer cells. *Biomaterials*, 31(32):8494–8506, 2010.

[99] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[100] M. Luca, A. Chavez-Ross, L. Edelstein-Keshet, and A. Mogilner. Chemotactic signaling, microglia, and Alzheimer's disease senile plaques: Is there a connection? *Bulletin of Mathematical Biology*, 65(4):693–730, 2003.

[101] A. Mackey, T. Kolokolnikov, and A. L. Bertozzi. Two-species particle aggregation and stability of co-dimension one solutions. *Discrete and Continuous Dynamical Systems - Series B*, 19(5):1411–1436, 2014.

[102] H. Marcovitch. *Black's Medical Dictionary*. London : A. & C. Black, 41 edition, 2005.

[103] A. F. M. Marée, V. A. Grieneisen, and P. Hogeweg. The Cellular Potts Model and Biophysical Properties of Cells, Tissues and Morphogenesis. In *Single-Cell-Based Models in Biology and Medicine*, pages 107–136. Birkhäuser Basel, Basel, 2007.

[104] P. Martin. Wound healing–aiming for perfect skin regeneration. *Science (New York, N.Y.)*, 276(5309):75–81, 1997.

[105] V. J. Martínez and E. Saar. *Statistics of the Galaxy Distribution*. CRC Press, 2010.

[106] T. Mattfeldt, H.-W. Gottfried, H. Frey, and U. Vogel. Second-order stereology of prostatic adenocarcinoma and normal prostatic tissue. *Acta Stereologica*, 1993.

[107] S. K. McMenamin, E. J. Bain, A. E. McCann, L. B. Patterson, D. S. Eom, Z. P. Waller, J. C. Hamill, J. A. Kuhlman, J. S. Eisen, and D. M. Parichy. Thyroid hormone–dependent adult pigment cell lineage and pattern in zebrafish. *Science*, 345(6202):1358–1361, 2014.

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

[109] A. Mogilner and L. Edelstein-Keshet. A non-local model for a swarm. *Journal of Mathematical Biology*, 38(6):534–570, 1999.

[110] A. Mogilner, L. Edelstein-Keshet, L. Bent, and A. Spiros. Mutual interactions, potentials, and individual distance in a social aggregation. *Journal of Mathematical Biology*, 47(4):353–389, 2003.

[111] J. C. M. Mombach and J. A. Glazier. Single Cell Motion in Aggregates of Embryonic Cells. *Physical Review Letters*, 76(16):3032–3035, 1996.

[112] J. Moreira and A. Deutsch. Pigment pattern formation in zebrafish during late larval stages: A model based on local interactions. *Developmental Dynamics*, 232(1):33–42, 2005.

[113] L. J. Morrell, G. D. Ruxton, and R. James. Spatial positioning in the selfish herd. *Behavioral Ecology*, 22(1):16–22, 2011.

[114] R. L. Mort, R. J. H. Ross, K. J. Hainey, O. J. 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.

[115] T. L. Morton, J. W. Haefner, V. Nugala, R. D. Decino, and L. Mendes. The Selfish Herd Revisited: Do Simple Movement Rules Reduce Relative Predation Risk? *Journal of Theoretical Biology*, 167(1):73–79, 1994.

[116] W. Mueller-Klieser. Multicellular spheroids. *Journal of Cancer Research and Clinical Oncology*, 113(2):101–122, 1987.

[117] J. Murray. A Pre-pattern formation mechanism for animal coat markings. *Journal of Theoretical Biology*, 88(1):161–199, 1981.

[118] J. D. Murray. A Pattern Formation Mechanism and Its Application to Mammalian Coat Markings. pages 360–399. Springer, Berlin, Heidelberg, 1980.

[119] J. D. Murray. On Pattern Formation Mechanisms for Lepidopteran Wing Patterns and Mammalian Coat Markings. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 295(1078), 1981.

[120] J. D. Murray. *Mathematical Biology II - Spatial Models and Biomedical Applications*. Springer-Verlag New York, 2008.

[121] T. Nagai and H. Honda. A dynamic cell model for the formation of epithelial tissues. *Philosophical Magazine Part B*, 81(7):699–719, 2001.

[122] A. Nakamasu, G. Takahashi, A. Kanbe, and S. Kondo. Interactions between zebrafish pigment cells responsible for the generation of Turing patterns. *Proceedings of the National Academy of Sciences of the United States of America*, 106(21):8429–34, 2009.

[123] R. M. Nerem. Tissue engineering: confronting the transplantation crisis. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 214(1):95–99, 2000.

[124] J. Ohser and F. Mücklich. *Statistical Analysis of Microstructures in Materials Science*. Wiley Chichester, 2000.

[125] R. Olfati-Saber. Flocking for Multi-Agent Dynamic Systems:Algorithms and Theory. *IEEE Transactions on Automatic Control*, 51(3):1–20, 2006.

[126] R. S. Olson, D. B. Knoester, and C. Adami. Evolution of Swarming Behavior Is Shaped by How Predators Attack. *Artificial Life*, 22(3): 299–318, 2016.

[127] G. F. Oster, J. D. Murray, and A. K. Harris. Mechanical aspects of mesenchymal morphogenesis. *Journal of Embryology and Experimental Morphology*, 78(1):83125, 1983.

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

[129] K. J. Painter, N. J. Armstrong, and J. A. Sherratt. The impact of adhesion on cellular invasion processes in cancer and development. *Journal of Theoretical Biology*, 264(3):10571067, 2010.

[130] K. J. Painter, J. M. Bloomfield, J. A. Sherratt, and A. Gerisch. A Nonlocal Model for Contact Attraction and Repulsion in Heterogeneous Cell Populations. *Bulletin of Mathematical Biology*, 77(6):1132–1165, 2015.

[131] J. K. Parrish and L. Edelstein-Keshet. Complexity, pattern, and evolutionary trade-offs in animal aggregation. *Science*, 284(5411):99–101, 1999.

[132] A. A. Patel, E. T. Gawlinski, S. K. Lemieux, and R. A. Gatenby. A Cellular Automaton Model of Early Tumor Growth and Invasion: The Effects of Native Tissue Vascularity and Increased Anaerobic Tumor Metabolism. *Journal of Theoretical Biology*, 213(3):315–331, 2001.

[133] L. B. Patterson, E. J. Bain, D. M. Parichy, C. Nüsslein-Volhard, and D. M. Parichy. Pigment cell interactions and differential xanthophore recruitment underlying zebrafish stripe reiteration and Danio pattern evolution. *Nature Communications*, 5:5299, 2014.

[134] P. J. E. Peebles. *The large-scale structure of the universe*. Princeton university press, 1980.

[135] J. M. Pérez García. [THE FRENCH ANATOMICAL XAVIER BICHAT (1771 - 1802). FUNDATOR OF DOCTRINE OF TISSUES. BIOGRAPHICAL MEMORY BEFORE HIS GRAVE]. *Anales de la Real Academia Nacional de Medicina*, 131(1):257–71; discussion 272, 2014.

[136] A. Pommerening. Approaches to quantifying forest structures. *Forestry*, 75(3):305–324, 2002.

[137] W. K. Prat. *Digital Image Processing*. John Wiley & Sons, Inc., 1991.

[138] A.-S. Qi, X. Zheng, C.-Y. Du, and B.-S. An. A Cellular Automaton Model of Cancerous Growth. *Journal of Theoretical Biology*, 161(1): 1–12, 1993.

[139] J. N. Reddy. *An Introduction to The Finite Element Method*, volume 2. McGraw-Hill New York, 1993.

[140] G. Reig, E. Pulgar, and M. L. Concha. Cell migration: from tissue culture to embryos. *Development*, 141(10), 2014.

[141] I. H. Riedel, K. Kruse, and J. Howard. A self-organized vortex array of hydrodynamically entrained sperm cells. *Science*, 309(5732):300–303, 2005.

[142] B. D. Ripley. The second-order analysis of stationary point processes. *Journal of Applied Probability*, pages 255–266, 1976.

[143] P. L. Rosin. A note on the least squares fitting of ellipses. *Pattern Recognition Letters*, 14(10):799–808, 1993.

[144] R. J. H. Ross, R. E. Baker, A. Parker, M. J. Ford, R. L. Mort, and C. A. Yates. Using approximate Bayesian computation to quantify cellcell adhesion parameters in a cell migratory process. *npj Systems Biology and Applications*, 3(1):9, 2017.

[145] L. Sachs. *Applied statistics*. Springer, 1984.

[146] N. Sachs and H. Clevers. Organoid cultures for the analysis of cancer phenotypes. *Current Opinion in Genetics & Development*, 24:68–73, 2014.

[147] T. Sato and H. Clevers. Growing Self-Organizing Mini-Guts from a Single Intestinal Stem Cell: Mechanism and Applications. *Science*, 340(6137), 2013.

[148] T. Sato, R. G. Vries, H. J. Snippert, M. van de Wetering, N. Barker, D. E. Stange, J. H. van Es, A. Abo, P. Kujala, P. J. Peters, and H. Clevers. Single Lgr5 stem cells build cryptvillus structures in vitro without a mesenchymal niche. *Nature*, 459(7244):262–265, 2009.

[149] M. H. Sayegh and L. A. Turka. The role of t-cell costimulatory activation pathways in transplant rejection. *New England Journal of Medicine*, 338(25):1813–1821, 1998.

[150] E. Scarpa and R. Mayor. Collective cell migration in development. *The Journal of Cell Biology*, 212(2), 2016.

[151] J. A. Sherratt and J. D. Murray. Mathematical analysis of a basic model for epidermal wound healing. *Journal of Mathematical Biology*, 29(5):389–404, 1991.

[152] K. Si-Tayeb, F. P. Lemaigre, and S. A. Duncan. Organogenesis and development of the liver. *Developmental cell*, 18(2):175189, 2010.

[153] 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.

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

[155] M. J. Simpson, K. K. Treloar, B. J. Binder, P. Haridas, K. J. Manton, D. I. Leavesley, D. 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.

[156] A. M. Stein, T. Demuth, D. Mobley, M. Berens, and L. M. Sander. A mathematical model of glioblastoma tumor spheroid invasion in a three-dimensional in vitro experiment. *Biophysical Journal*, 92(1):356–365, 2007.

[157] E. H. K. Stelzer. Light-sheet fluorescence microscopy for quantitative biology. *Nature Methods*, 12(1):23–26, 2015.

[158] A. Stevens and H. G. Othmer. Aggregation, Blowup, and Collapse: The ABC's of Taxis in Reinforced Random Walks. *SIAM Journal on Applied Mathematics*, 57(4):1044–1081, 1997.

[159] D. Stoyan and H. Stoyan. Estimating pair correlation functions of planar cluster processes. *Biometrical Journal*, 38(3):259–271, 1996.

[160] D. Stoyan and H. Stoyan. Improving Ratio Estimators of Second Order Point Process Characteristics. *Scandinavian Journal of Statistics*, 27 (4):641–656, 2000.

[161] R. M. Sutherland. Cell and environment interactions in tumor microregions: the multicell spheroid model. *Science*, 240(4849):177–184, 1988.

[162] A. Szabó and R. Mayor. Cell traction in collective cell migration and morphogenesis: The chase and run mechanism. *Cell Adhesion & Migration*, 9(5):380–383, 2015.

[163] E. Theveneau, B. Steventon, E. Scarpa, S. Garcia, X. Trepat, A. Streit, and R. Mayor. Chase-and-run between adjacent cell populations promotes directional collective migration. *Nature Cell Biology*, 15(7):763–772, 2013.

[164] R. J. Thomas, A. Bennett, B. Thomson, and K. M. Shakesheff. Hepatic stellate cells on poly (DL-lactic acid) surfaces control the formation of 3D hepatocyte co-culture aggregates in vitro. *European cells & materials*, 11:1626, 2005.

[165] T. Toffoli and N. Margolus. *Cellular Automata Machines: a New Environment for Modeling*. MIT Press, 1987.

[166] T. Toni, D. Welch, N. Strelkowa, A. Ipsen, M. P. Stumpf, and C. A. Yates. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of The Royal Society Interface*, 6(31):187–202, 2009.

[167] C. M. Topaz, A. L. Bertozzi, and M. A. Lewis. A nonlocal continuum model for biological aggregation. *Bulletin of Mathematical Biology*, 68 (7):1601–1623, 2006.

[168] C. M. Topaz, A. J. Bernoff, S. Logan, and W. Toolson. A model for rolling swarms of locusts. In *European Physical Journal: Special Topics*, volume 157, pages 93–109, 2008.

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

[170] H. Tronnolone, J. M. Gardner, J. F. Sundstrom, V. Jiranek, S. G. Oliver, and B. J. Binder. Quantifying the dominant growth mechanisms of dimorphic yeast using a lattice-based model. *Journal of The Royal Society Interface*, 14(134), 2017.

[171] A. M. Turing. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 237(641):37–72, 1952.

[172] A. M. Turing. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 237(641):3772, 1952.

[173] J. P. Vacanti and C. A. Vacanti. The History and Scope of Tissue Engineering. In *Principles of Tissue Engineering*, pages 3–8. Elsevier, 2014.

[174] R. F. M. van Oers, E. G. Rens, D. J. LaValley, C. A. Reinhart-King, and R. M. H. Merks. Mechanical Cell-Matrix Feedback Explains Pairwise and Collective Endothelial Cell Behavior In Vitro. *PLoS Computational Biology*, 10(8):e1003774, 2014.

[175] T. Vicsek and A. Zafeiris. Collective motion. *Physics Reports*, 517(3): 71–140, 2012.

[176] M. Vinci, S. Gowan, F. Boxall, L. Patterson, M. Zimmermann, C. Lomas, M. Mendiola, D. Hardisson, S. A. Eccles, *et al.* Advances in establishment and analysis of three-dimensional tumor spheroid-based functional assays for target validation and drug evaluation. *BMC Biology*, 10(1):29, 2012.

[177] A. Volkening and B. Sandstede. Modelling stripe formation in zebrafish: an agent-based approach. *Journal of The Royal Society Interface*, 12 (112):20150812, 2015.

[178] Q. Wei, V. Hariharan, H. Huang, J. Wu, and Z. Guo. Cell-Cell Contact Preserves Cell Viability via Plakoglobin. *PLoS ONE*, 6(10):e27064, 2011.

[179] D. Wilson. Tissue. *The Lancet*, 374(9684):109, 2009.

[180] P. J. Wipff, D. B. Rifkin, J.-J. Meister, and B. Hinz. Myofibroblast contraction activates latent TGF-1 from the extracellular matrix. *The Journal of Cell Biology*, 179(6):13111323, 2007.

[181] S. Wolfram. Statistical mechanics of cellular automata. *Reviews of Modern Physics*, 55(3):601–644, 1983.

[182] T. E. Woolley, P. K. Maini, and E. A. Gaffney. Is pigment cell pattern formation in zebrafish a game of cops and robbers? *Pigment Cell & Melanoma Research*, 27(5):686–687, 2014.

[183] M. Yamaguchi, E. Yoshimoto, and S. Kondo. Pattern regulation in the stripe of zebrafish suggests an underlying dynamic and autonomous mechanism. *Proceedings of the National Academy of Sciences*, 104(12): 4790–4793, 2007.

[184] H. Yamanaka and S. Kondo. In vitro analysis suggests that difference in cell movement during direct interaction can generate various pigment patterns in vivo. *Proceedings of the National Academy of Sciences*, 111 (5):1867–1872, 2014.

[185] V. Zhdankin and J. C. Sprott. Simple predator-prey swarming model. *Physical Review E*, 82(5):056209, 2010.