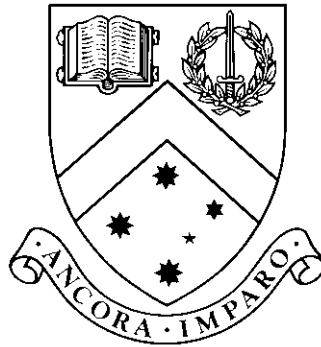


# **A stochastic differential model for cell migration and mutual cell-cell interactions**

by

**Rotem Aharon, BSc, MSc**



## **Thesis**

Submitted by Rotem Aharon

for fulfillment of the Requirements for the Degree of

**Doctor of Philosophy**

Supervisor: Prof. Fima Klebaner

Associate Supervisor: Assoc. Prof. Kais Hamza

**School of Mathematical Sciences  
Monash University**

February, 2016

© The author 2016. Except as provided in the Copyright Act 1968, this thesis may not be reproduced in any form without the written permission of the author.

© Copyright

by

Rotem Aharon

2016

To my family

# **A stochastic differential model for cell migration and mutual cell-cell interactions**

## **Declaration**

I declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institute of tertiary education. Information derived from the published and unpublished work of others has been acknowledged in the text and a list of references is given.



---

Rotem Aharon  
February 29, 2016

# A stochastic differential model for cell migration and mutual cell-cell interactions

Rotem Aharon, BSc, MSc  
Monash University, 2016

Supervisor: Prof. Fima Klebaner  
Associate Supervisor: Assoc. Prof. Kais Hamza

## Abstract

Motivated by two biological questions concerning the way radiation treatment affects cell behaviour and the way interactions between cells control cell segregation and cluster formation, we constructed a mathematical model for cell migration and interactions. Starting from first principles and basic biological assumptions, we arrived at a stochastic differential equation where the drift term accounts for short term interactions between cells and the random term accounts for independent cell motion. Likelihoods of different values for model parameters given particle paths were obtained using Girsanov theorem, and then used to estimate actual parameter values by maximising these likelihoods (MLE). Accuracy of this method was tested by comparing estimated parameter values given computer generated paths to the actual values used to generate these paths, for a few different drift functions. Application of this technique to a data set containing real cell paths which were observed in laboratory experiments studying the effect of radiation treatment on cell migration and interaction unveiled a clear trend in cells' response: radiation dosage of 10Gy was found to increase cell motility by 50% and diminish cell adhesion effectively to zero. An extended version of our model which further accounts for cell births and interactions between different population types was designed to help understand cell segregation and cluster formation regulated by cell membrane proteins called Eph and ephrin. First this helped identify the significant components in controlling the behaviour and dynamics displayed by the biological system, which has countlessly more components over many time and distance scales. Second, when compared against experimental results it was able to replicate both the dynamics and range of cell segregation that was observed in the laboratory by our collaborators<sup>1</sup>.

We thus present here a powerful yet simple model which is both generic and versatile. With only a small number of parameters that can be estimated from data containing cell paths, it holds information regarding independent cell motion and mutual cell-cell interactions, and can reproduce, predict and help analyse dynamics and behaviours observed in laboratory experiments.

---

<sup>1</sup>Work on the Eph-ephrin project was conducted in collaboration with Dr Peter Janes and late Prof. Martin Lackmann who has sadly passed away during our collaborative work. Both from the Department of Biochemistry & Molecular Biology at Monash University.

# Contents

<b>Abstract</b> . . . . .	<b>v</b>
<b>List of Figures</b> . . . . .	<b>ix</b>
<b>List of Notations</b> . . . . .	<b>x</b>
<b>Acknowledgments</b> . . . . .	<b>xi</b>
<b>1 Introduction</b> . . . . .	<b>1</b>
1.1 Background . . . . .	1
1.1.1 Cell motility . . . . .	2
1.1.2 Cell-cell interactions . . . . .	2
1.1.3 Segregation between two cell types . . . . .	3
1.2 Stochastic differential equations, applications and analysis . . . . .	4
1.2.1 Similar approach in other applications . . . . .	4
1.2.2 Parameter estimations in stochastic differential equations . . . . .	4
1.3 In this dissertation... . . . .	5
<b>2 Background</b> . . . . .	<b>7</b>
2.1 Models for cell dynamics in mesoscale biological systems . . . . .	7
2.1.1 Diffusion models . . . . .	7
2.1.2 Agent based models with localised cell-cell interactions . . . . .	8
2.1.3 Potts model . . . . .	9
2.2 Mathematical background, concepts and terminology . . . . .	11
2.2.1 Stochastic differential equations . . . . .	11
2.2.2 Stationary solution . . . . .	11
2.2.3 Parameter estimation methods for stochastic differential equations . . . . .	12
2.2.4 Branching Brownian motion . . . . .	14
<b>3 Model construction and analysis</b> . . . . .	<b>15</b>
3.1 Model formulation . . . . .	15
3.1.1 Model's scope . . . . .	15
3.1.2 Mutual interaction forces . . . . .	16
3.1.3 A specific choice of potential function . . . . .	17
3.1.4 Model equations . . . . .	18
3.2 Model analysis . . . . .	19
3.2.1 Existence and Uniqueness of solutions . . . . .	19
3.2.2 Stationary Solution . . . . .	19
3.2.3 Analysis of the Stationary Solution . . . . .	19

<b>4</b>	<b>Estimation of model parameters from data</b>	<b>23</b>
4.1	Reconstructing $\sigma$ directly from cell paths	23
4.1.1	Obtain $\sigma$ from the variance of the sample of cell paths	24
4.1.2	Obtain $\sigma$ from the variance of the sample of cell displacements	25
4.1.3	Change of time	26
4.2	Maximum Likelihood Estimation of drift term parameters	27
4.2.1	Girsanov theorem: Change of measure and the likelihoods ratio	27
4.2.2	Generalise to higher dimensions	28
4.3	Numerical demonstration	29
4.3.1	Error calculation	30
4.3.2	Linear drift	30
4.3.3	Linear interactions	33
4.3.4	Exponential repulsion and attraction interactions	35
<b>5</b>	<b>Radiation effects migration and interactions</b>	<b>37</b>
5.1	Biological motivation	37
5.2	Results	38
5.2.1	The effect of radiation treatment on cell motility	39
5.2.2	The effect of radiation treatment on cell-cell interactions	39
5.3	Conclusions and future work	42
<b>6</b>	<b>Eph-ephrin regulation of cluster formation</b>	<b>43</b>
6.1	Biological background	43
6.2	Extended model	44
6.2.1	SDE describing the position of all cell centres	45
6.2.2	Dynamics of interactions between cells of different types	45
6.2.3	Birth process	46
6.2.4	Model equations	47
6.3	Comparison of simulated and experimental cell-cell segregation	48
6.3.1	Dynamics of cluster formation	48
6.3.2	Modulation of the Eph-ephrin signalling strength effects cluster formation	49
6.3.3	The formation of tightly-packed islands-in-the-sea Eph cell clusters relies on enhanced cell-cell adhesion within clusters	52
6.4	Discussion	55
<b>7</b>	<b>Summary</b>	<b>57</b>
	<b>Appendix A Proofs</b>	<b>61</b>
A.1	Proof for lemma 4.2.2	61
A.2	Proof for lemma 4.2.4	62
	<b>Appendix B Additional results</b>	<b>65</b>
B.1	Additional numerical demonstration	65
B.1.1	Linear drift	65
B.1.2	Linear (attraction) interactions	67
B.1.3	Exponential (repulsion and attraction) interactions	68
	<b>Appendix C Methods</b>	<b>71</b>
C.1	Methods for chapter 5	71
C.2	Methods for chapter 6	72
C.2.1	Numeric solutions	72
C.2.2	Image analysis	72

C.2.3	Experimental setup . . . . .	72
C.2.4	Statistical analysis of data regarding dynamics of cluster formation .	72
<b>Vita</b>	. . . . .	<b>75</b>



# List of Figures

3.1	Bifurcation diagram over the $l - K$ parameter space. . . . .	21
4.1	Linear drift (well potential) . . . . .	32
4.2	Linear interactions (attraction only) . . . . .	34
4.3	Exponential interactions (repulsion and attraction forces) . . . . .	36
5.1	$\sigma^2$ values obtained from experimental data show radiation treatment effects cell motility . . . . .	40
5.2	Cell paths obtained experimentally . . . . .	41
5.3	Drift term parameters estimated from experimental data show radiation treatment effects cell-cell interaction . . . . .	41
6.1	Time-lapse imaging and simulation of Eph/ephrin-driven cell-cell segrega- tion (Aharon et al., 2014). . . . .	50
6.2	Distribution of cell cluster sizes during segregation (Aharon et al., 2014) . .	51
6.3	Eph expression level and signalling capacity regulate cell segregation: com- parison of experimental versus modelling outcomes (Aharon et al., 2014). .	53
6.4	Increased cell-cell adhesion within one cell population is required for the formation of tightly packed cell clusters (Aharon et al., 2014). . . . .	54
B.1	Linear drift, computer generated sample paths . . . . .	65
B.2	Linear drift, error estimations . . . . .	66
B.3	Linear interactions (attraction only), computer generated sample paths . .	67
B.4	Linear interactions (attraction only), error estimations . . . . .	67
B.5	Exponential interactions (repulsion and attraction), computer generated sample paths . . . . .	68
B.6	Exponential interactions (repulsion and attraction), error estimations . . . .	69
C.1	Cluster size distribution. . . . .	74

# List of Notations

## Model variables

$\mathbf{q}(t)$  Positions of all cells at time  $t$

$\mathbf{q}_i(t)$  Positions of the  $i$ 'th cells at time  $t$

$N(t)$  Number of cells at time  $t$

$N^{eph}(t)$  Number of eph expressing cells at time  $t$

$N^{ephrin}(t)$  Number of ephrin expressing cells at time  $t$

## Model parameters

$\sigma$  'Free' diffusion rate

$a$  Attraction length scale

$r$  Repulsion length scale

$l = a/r$  Relative length scale

$A$  Attraction strength (or amplitude)

$R$  Repulsion strength (or amplitude)

$K = R/A$  Relative strength (or amplitude)

$\lambda$  Proliferation rate

$C$  Eph-ephrin regulated attenuation constant

# Acknowledgments

I would like to thank my supervisors Professor Fima Klebaner and Associate Professor Kais Hamza for all their support and helpful guidance along the way. I would also like to thank Professor Antony Burgess for opening the door to the world of biological applications through providing invaluable feedback and helping with establishing important collaborations. I would like to acknowledge the great contribution late Professor Martin Lackmann had to our collaborative project, his passing was a really tragic loss. From Martin's laboratory, Dr Peter Janes is also acknowledged, for many helpful discussions and hours of work on the biological aspect of our joint work (chapter 6). I thank Dr Jeffrey Crosbie and Dr Camden Lo for providing the experimental data we analyse in chapter 5. I acknowledge the constructive comments of my reviewers, Prof. Yoni Nazarathy and Prof. Zinoviy Landsman who helped make this a better document. Last on the academic note I would like to acknowledge all the help I received from my husband Dr Gideon Kowadlo, from the details of proof reading and programming to late night discussions of high level concepts and ideas.

On a personal note, I would like to thank Gideon again, for his patience, understanding and endless support on all levels. You are the best partner for this journey called life, I love you. I would like to thank my family here in Melbourne for their help which made it possible for me to write up this thesis. And to my family in far away Israel, who's love and support traveled across the ocean and helped keeping me focused and determined. Last of all I would like to thank my beautiful little boy Stav, for being the adorable gorgeous boy that you are, and for teaching me a whole new level of joy.

Rotem Aharon

*Monash University*  
*February 2016*



# Chapter 1

## Introduction

### 1.1 Background

The human body is made of trillions of cells (Bianconi et al., 2013). These cells have many different roles and functionalities. Cell characteristics can vary dramatically between different cell types; blood cells for example can essentially flow like a fluid and transfer oxygen or carbon-dioxide from place to place, while skin cells stay put and form strong yet supple structures to make the skin tissue. These very different cells have a lot of commonalities, they all have a nucleus, a membrane, and a cytoskeleton (an inner cell skeleton) mitochondria, genetic matter, and more. They can all move when on their own, and can all interact with their environment and with neighbouring cells. They all metabolise, may proliferate (this can depend on the cell type etc') and eventually die. The different functionalities can often be achieved by different levels of manifestations of the same characteristic: For example, Blood cells do not adhere to their neighbour, while skin cells do to a very high degree, which orchestrates the difference between the fluid like form of blood and the stable form of the skin (Alberts et al., 2008).

When biologists approach the investigation of a certain biological aspect or mechanism, they would attempt to look at it first in as much isolation as possible. Therefore, when trying to understand mechanisms involved in the behaviour and reaction of a population of cells to changes in its environment or to a different population of cells, a laboratory experiment aimed to study these scales would often start in a two dimensional sample of same type cells that are introduced with the 'trigger' under investigation. Depending on the time scale relevant for the phenomena under investigation, the treated sample is placed under the microscope either for a (semi) continuous observation (i.e. an image taken every 20 or 30 minutes), or at the end of the experiment, to document the end result only (i.e. a single image taken after 24 or 48 hours).

The biological questions we explore in this dissertation deal with such samples. More particularly, in the first application a homogeneous cell population was introduced with a radiation treatment of varying levels to study the effect radiation treatment has on the way cells move and interact with each other (chapter 5). In the second application, a homogenous population of randomly scattered (Eph-expressing) cells was introduced with another population of randomly scattered (ephrin-expressing) cells, to study the way Eph-ephrin interactions regulate the segregation between the two populations and the formation of homogenous clusters over time, while cell birth in both populations increases cell densities to full confluence (chapter 6).

The mathematical model presented here is carefully designed to handle the scope of these systems, and to capture the relevant mechanisms that control their dynamics. In what follows we present the relevant cellular mechanisms that take place, and go through common approaches in their mathematical modelling. A more detailed survey of models

that deal with systems of the same scope as ours is given in chapter 2, where the different models are compared, contrasted and weighted with respect to the functionalities we need included in our model, and the particular questions we want answered.

### 1.1.1 Cell motility

Most cells in the human body can move around by crawling over the surface they are on. This happens essentially in three stages: Protrusion, in which actin rich structures are pushed out at the front of the cell. Attachment, in which the actin cytoskeleton connects to the surface the cell is on, and traction, in which the cell essentially pulls itself forward to ‘catch up’ with the protruding structure. Cell motility is essential for healthy embryonic development, where cells need to move from their place of origin to different sites around the developing body. Further, in the adult, cell motility has an essential part in the functioning of macrophages in a healthy immune system, in the process of bone renewal, and in tissue repair. In cancer, cell motility plays a significant role in metastases, where cells can leave the tumour they originated in, crawl into adjacent blood or lymph vessels and then spread to other sites (Alberts et al., 2008).

Mathematical and computational approaches to cell motility consider different aspects of cell motility that impact on cell motion, including cytoplasm dynamics (Alt and Dembo, 1999), the growth of actin filaments (Bindschadler et al., 2004), and the distribution of adhesion proteins in the cell membrane (Dickinson and Tranquillo, 1993) or, more recently, integrating the different aspects (Gracheva and Othmer, 2004; Shao et al., 2012). For a more comprehensive review of mathematical models for cell motility see for example Mogilner (2009).

When considering migration distances that greatly exceed the length of a single cell, and time scales that are longer than the cell’s ‘persistence’ time, the Mean Square Displacement (MSD) of a single cell is linear in time (Dickinson and Tranquillo, 1993). Thus, when considering these scales, it is sufficient to represent a single cell’s motion by a random walk in the discrete case, or by a Brownian motion in the continuous case, see for example Ionides et al. (2004). Accordingly, when considering large cell populations, averaging techniques can be used to derive the diffusion equation. Diffusion equations are used for example to describe cell migration on growing domains (Baker et al., 2010) and cancerous cell invasion (Gatenby and Gawlinski, 1996).

Cell motion can be directed by external signals, where, for example the direction in which the actin filaments grow is dictated by the concentration of an external chemical (chemotaxis), see e.g. Alberts et al. (2008). Cell migration induced by chemical gradient was one of the first cellular phenomena to be subjected to mathematical modelling using diffusion equations, this enabled understanding of the phenomenal ability of cells that are smaller than the size of spatial fluctuations in the concentration of the chemical that leads them to move up its global gradient (Keller and Segel, 1971). In SDE type models, chemotaxis can be introduced through the introduction of the spatial gradient of the chemical (or other) signal through the drift term (Stokes et al., 1991; Ionides et al., 2004).

### 1.1.2 Cell-cell interactions

A cell can interact with other cells in its environment through the secretion of chemical signals to its environment, which can then trigger different mechanisms in the neighbouring cells, such as directed motion (mentioned above), apoptosis, cell division etc’ (Alberts et al., 2008). This mechanism can have long distant effects, and its role in different biological systems is of great interest for research through mathematical models, for example in wound healing (Maini et al., 2002), or in cancer invasion (Gatenby and Gawlinski, 1996; Smallbone et al., 2005).

A cell can also have short distant interactions with its neighbour. These are essentially contact interactions, in which a physical contact or bond is made between corresponding proteins embedded in both cells' membranes. Depending on the proteins and the interactions between them, this leads to a downstream of signals within the cells which can result in the formation of adhesion connections between the two cells, or in cell retraction and separation. Short term interactions within mesoscale cell populations stand in the core of this research work. Current approaches to mathematical modelling of this are mentioned below, and an extended survey is given in section 2.1, where these are compared and contrasted in the light of the biological questions we want answered.

Models for mutual cell-cell interactions within large cell populations commonly explore an 'agent-based' approach assuming each 'agent' (cell) follows a set of rules, which allow it to perform a random walk exclusion process (where each cell moves randomly on a grid but cannot move to a site already occupied by another cell), and interact with its neighbours. For simple contact interactions between agents (interactions only with immediate neighbours), this detailed description given at the 'microscopic' level, was shown to give rise to a non-linear diffusion at the macroscopic level (Deroulers et al., 2009). The continuous approximation can then be used to give predictions regarding behaviour of very large population sizes, for which there are too many to run the agent based computational model. Fernando et al. (2010) determined the diffusion functions arising for different choices of interaction functions, and their dependency on the chosen lattice and the way the contact neighbourhood is defined. When introducing more general interactions (not only between immediate neighbours, but also imposing some spatial structure, such as aggregation or chain formation), it was shown that while a computerised simulation of the agent based model does indeed give rise to aggregation or chain formation, the continuous approximation of the model may fail to capture those (Simpson et al., 2010).

### 1.1.3 Segregation between two cell types

When contact interactions occur between cells that come from two different cell populations, it can lead for example to segregation between the two populations. Differential adhesion, a mechanism in which a particular cell has different affinity levels to surface adhesion with different types of cells, was shown to lead to cell segregation and cluster formation (Steinberg, 2007). Further, short distance interactions regulated by Eph-ephrin interactions, a specific type of proteins embedded in the cell's membrane, were shown to regulate cell positioning, tissues and organ patterning (Pasquale, 2010; Lackmann and Boyd, 2008). A simple demonstration of this can be observed in laboratory experiments where two types of cells are cultured together on the surface of a two dimensional substrate. Initially they are all intermingled, scattered randomly on the two dimensional surface. With the progression of time, cell migration along with cell-cell interactions result in segregation between the two populations as homogenous cell clusters are formed.

A computational model (the cellular Potts model) for modelling cell dynamics in a two dimensional sample, developed by Glazier and Graner (1993), explored use of cell surface energy as a function of its interactions with neighbouring cells and area constraints, to predict a configuration which minimises the energy of the entire system. Different relative surface energies were found to correspond with different levels of cell sorting.

The cellular Potts model was further used to model cell positioning and segregation along the intestinal crypt epithelium, guided by the assumption that differential ephrinB/EphB activation can be modelled as a form of differential adhesion (Wong et al., 2010). While the predictions of the Potts model often match observations of the biological system they model well, the dynamics or evolution of the system as it approaches the predicted state may be harder to address in them. This is due to the fact that incorporation

of the physical constraints on cell migration under increasing cell density or full confluence conditions is not straight forwards in this setup.

A more recent approach that simulates both adhesive and migratory cell behaviours (Taylor et al., 2011), allowed modelling of the time course of the segregation of cells differing in their expression of cadherin adhesion proteins, which was within 50% agreement of empirical data (Taylor et al., 2012). However, this model was not capable to reproduce the significantly faster rate of cell segregation due to Eph-ephrin interactions, suggesting that mechanisms additional or alternative to differential adhesion contribute to Eph-ephrin mediated cell segregation (Taylor et al., 2012).

## 1.2 Stochastic differential equations, applications and analysis

Stochastic differential equations (SDE's) represent a concise formulation of a random process that follows an evolution law in which the mean follows a deterministic function (the drift function), and the evolution of the covariance of the random variable is defined by the diffusion matrix. A definition is given in section 2.2.1, or see Karlin and Taylor (1981); Bhattacharya and Waymire (1990); Klebaner (2011) to name a few textbooks.

SDE's, as one would expect, are often used to model systems that have a random element in them. For example, pricing of options in the stock market (Black and Scholes, 1973), or physical systems that have thermal noise.

### 1.2.1 Similar approach in other applications

Brillinger (2003) suggested using Brownian motion to model animal behaviour and include the interactions between individuals and their environment through the drift term, and then used this type of model to analyse elk behaviour from experimental data to identify 'attracting spots' in the elks' environment. Ionides et al. (2004) suggested it can be done for modelling cell behaviour, but to our knowledge, while environmental components (like chemical attractants) were incorporated through the drift term before (Stokes et al., 1991), cell-cell interactions have not been included in such models so far. We incorporate this approach and add the interactions between individual cells through the drift term by using functions often used in modelling of swarm behaviour.

This general approach might resemble the one applied in models for molecular dynamics. In particular, the general form of our model's equation is known in the chemical-physics literature as the over-damped Langevin equation, which is obtained under the assumptions that the drag force is large and particle mass is small, similar to what can be assumed in the biological applications of our model (see section 3.1.1 for more detail). Parameter estimation methods applied there often have to deal with more global parameters than the ones our methods provides estimations for, since in the molecular level, observations of individual particle paths over time are not as straightforward. In molecular dynamics, however, much more common conditions are the no friction or the fluctuation-dissipation assumption, for which the steady state of both systems can be represented by the Gibbs canonical ensemble (Espanol and Warren, 1995).

### 1.2.2 Parameter estimations in stochastic differential equations

Parameter estimation methods in stochastic differential equations make a wide field of ongoing research. See for example Rao (1999); Kutoyants (2013), or some relevant definitions in section 2.2.3. Particularly for the type of equation obtained in our model, an interesting discussion comparing parameter estimation procedures which fit to the empirical measure



generated by the data, with estimation procedures which fit diffusion processes to data by applying the maximum likelihood principle to the path-space density of the desired model equation can be found in Pokern et al. (2009).

Furthermore, experimental data is always made of discrete observations, which requires adaptations to the theoretical estimators or the formulations of entirely new estimators, see a review in Nielsen et al. (2000). Indeed, both the sampling rate and the integration rate should be considered, and some estimation techniques can be computationally expensive (Timmer, 2000). However, the likelihood function for our model can be obtained explicitly. This enables us to perform a discretised version of the maximum likelihood estimation procedure and obtain relevant estimations of parameter values from cell paths observed in laboratory experiments.

### 1.3 In this dissertation...

We develop a mathematical model to simulate cell migration and mutual interactions, which when applied to single cell type is kept analytic (as opposed to the common approach of employing an agent based inclusion process described above). This allows an analysis of its solution and formulation of an analytic method for estimation of model parameters. Further, it can handle very low population densities (where the distribution of cells is very sparse) up to high population densities (where cell-cell interactions become significant) seamlessly.

The position of the centre of each cell is described by a Stochastic Differential Equation (SDE) in which the independent random part of cell motion is described by an independent Brownian Motion (BM). The interactions between cells are incorporated into the drift term, which is a vector of deterministic functions that depend on the distance between the cells. Our model analysis shows weak existence and the weak uniqueness property of the solution. We find the stationary solution when it exists, and discuss the conditions on the interaction type for it to predict the formation of a stable cluster in which a most likely value for the distance between neighbouring cells exists (chapter 3). We formulate a method for estimation of model parameters from experimental results containing cell paths, utilising basic statistical concepts along with a Maximum Likelihood Estimation (MLE) technique (chapter 4). A demonstration of the use of this method with real experimental data sheds light on the effect radiation treatment has on cell migration and mutual interactions (chapter 5, see biological motivation therein). An extension to our model that further facilitates simulation of a range of interactions between two different cell types is then applied to explore the Eph-ephrin modulation of cluster formation, comparing our model predictions to experimental results (chapter 6, biological background therein).

#### Contributions of this thesis

- The formulation of a mathematical model that describes cell migration and interactions. The model incorporates a unique approach which to our knowledge was not applied to cell behaviour before. An SDE is used to describe the position of all cell centres, where the BM part accounts for independent cell motion, and all cell-cell interactions are incorporated into the drift term.
- Model analysis which verifies weak existence and the weak uniqueness property of the solution, provides its stationary solution, and demarcates the parameter range for which a stable cluster with a most likely value for the distance between neighbouring cells exists.

- A method for estimation of model parameters from experimental results containing cell paths, utilising basic statistical concepts along with a Maximum Likelihood Estimation technique.
- Application of the aforementioned method to experimental data which unveils clear trends in the effect radiation treatment has on cell migration and mutual interactions. It shows cell motility increases with increase in radiation level, and cell-cell interactions diminish as radiation dosage increases.
- An extended version of the model that further facilitates simulation of a range of interactions between two different cell types aimed to study cell segregation and cluster formation regulated by Eph-ephrin interactions. Comparison to experimental results showcases our model's ability to replicate both the dynamics and possible range of cluster formation in the experimental system.

# Chapter 2

## Background

*In this chapter, current approaches to mathematical modelling of mesoscale biological systems are surveyed in the light of the specific questions relevant to the biological applications studied in this dissertation. The fact we required an analytic model which is relatively tractable and accessible to some form of mathematical analysis, can account for complex localised interactions, and be applied to growing population densities (growing from very sparse, through a pure birth process, to very dense cell population), lead us to take our own innovative approach in formulating our model (see chapter 3 to follow). Further in this chapter, background regarding mathematical concepts and techniques employed in this research is presented.*

### 2.1 Common approaches to modelling cell dynamics in mesoscale biological systems

Here we present a more detailed survey of some common mathematical and computational models currently used to help understand, analyse and predict biological systems of the same scope of the applications we study in this work. Namely, spatiotemporal biological systems focused on the mesoscale level<sup>2</sup>, where mutual cell-cell interactions take place within a small to medium cell population (hundreds to thousands of cells) and control the overall dynamics observed within the relevant time scale (24-48 hours).

For each family of models presented below we briefly describe the main mathematical approach, equations and basic assumptions of the model, followed by summary of model applications. Further, we discuss each model's strengths and weaknesses, in particular with respect to their tractability and ability to account for complex, localised cell-cell interactions.

#### 2.1.1 Diffusion models

##### Model description

Diffusion models essentially use the motility coefficient of an individual cell  $D \equiv \frac{\Delta^2}{\Delta t}$ , where  $\Delta$  is the length of a single step and  $\Delta t$  is the average time interval between steps, to describe the evolution of cell density  $C(x, t)$  over time. The simplest representation of the diffusion equation for modelling cell density  $C(x, t)$  in one dimension is:

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x} \left( D \frac{\partial C}{\partial x} \right) \quad (2.1.1)$$

---

<sup>2</sup>This is the intermediate scale between those of the entire tissue and an individual cell.

## Model applications

Diffusion models with the relevant adaptations are widely used to describe a variety of biological applications. For example, the addition of an advection term representing motion that follows a gradient of a chemical makes it suitable to describe cell density of a population following chemotaxis (Keller and Segel, 1971). Its generalisation to two dimension and the association of cell density with cell differentiation (also known as ‘reaction diffusion models’) can give rise to pattern formation (Maini, 2000). When extended further to include two types of cell populations, normal cells and tumour cells, and the presence of a chemical secreted by the tumour cells that cause apoptosis amongst the normal cells, it is applied to generate predictions regarding cancer cell invasion (Gatenby and Gawlinski, 1996).

## Synopsis

Diffusion models are analytic and can sometimes be solved explicitly. They can often be subjected to analysis in order to find steady state solutions, or dependency of the dynamics on model parameters. These models are highly adaptable in handling interactions, at the density level, between different population types or with varying environmental changes (such as chemical agents). This enables the inclusion of population growth due to cell birth, or some sort of ‘populations competition’ (e.g as in the Lotka-Volterra competition) when needed. However, since these models deal with continuous cell densities and not with individual cells, it is not possible to specify the position of any individual cell, and therefore the introduction of spatial based cell-cell interactions within or between different cell populations is not relevant here.

### 2.1.2 Agent based models with localised cell-cell interactions

#### Model description

This type of models assume all cells can move on a pre-defined grid  $\mathbb{Z}^d$ , where each grid location has a predefined number of neighbouring grid sites  $n$ . Each individual cell performs a random walk exclusion process on the grid. That is, at a given time step, a cell chooses to move into one of its neighbouring grid sites at random. If the chosen site is unoccupied, the cell moves into the new site.

Interactions between cells are normally introduced through the construction of the probability profile over the  $n$  adjacent grid sites. For example, if the model lets the cells have positive (also referred to as attractive) interactions between them, a new site that has  $n - 1$  occupied sites around it is assigned a higher probability to be chosen over a site that has  $n - 2$  (or less) occupied neighbouring sites around it, following a particular (user defined, usually application based) probability function.

#### Applications

Agent based models are widely employed to explore a variety of cellular phenomena in the mesoscale realm. Modelling of cell invasion (Deroulers et al., 2009), cluster formation (Simpson et al., 2010), neural crest cell migration, where an off-lattice model was employed (McLennan et al., 2015), are only a few examples.

Continuous approximations of an agent based model with simple contact interactions in two dimensions were shown to match simulation results (Deroulers et al., 2009). Fernando et al. (2010) derived the continuous approximation for different choices of cell-cell interaction functions and different model geometries (i.e lattice type and arrangement of contact neighbourhood), and obtained non linear diffusion functions:  $D = D(C)$  in equation 2.1.1

above. They showed that model geometry, as well as the choice of cell-cell interaction type, give rise to different diffusion functions  $D(C)$  in the continuous approximation.

Employing such an agent based model and its continuous approximation, Bowden et al. (2013) performed a sensitivity analysis which showed that the ability to correctly interpret trajectory data is extremely sensitive to the geometry and timing of the experiment, the degree of motility bias and the number of experimental replicates. The same type of model was used to infer cell-cell interactions from population density characteristics and individual cell trajectories on static and growing domains (Ross et al., 2015). More particularly, they showed the concavity of a single agents' trajectory as well as its expected displacement and standard error of the mean could give an indication whether cell-cell interactions are of repulsive or attractive type when considered over short time scales. To our knowledge, a more exact estimation of model parameters can only be done in the cases where the continuous approximation to the agent based model holds.

## Synopsis

Agent based models are very common in addressing this scale of biological problems. Indeed they present great flexibility in the design of rules the cells shall follow, and generate reliable results in the form of computer simulations. Cell birth can be formulated and introduced in various ways, and so can cell differentiation, resulting in various types of populations and interactions between them. In terms of analysis, these models do not have an analytic description, and can not be solved explicitly. Continuous approximations, for which methods of parameter estimations are available, can be made in some cases. The continuous approximation may not be valid in the case of low densities where the spatial distribution of cells is very sparse, which is often the case in the biological applications we study (see chapters 3-6 to follow), or in the case of complex interactions between and within two different populations as in our second application, regarding Eph-ephrin regulation of cell-cell interactions (chapter 6).

Even when the continuous approximation is available, the main difficulty with this type of model for us is with methods for parameter estimations. Effectively, the choice of model geometry can change model predictions, this is evident through the fact different choices of grid type and neighbour numbers give rise to different diffusion functions in the continuous approximation (Deroulers et al., 2009; Fernando et al., 2010), or more directly in the sensitivity analysis that indicated extreme sensitivity to model geometry (Bowden et al., 2013). In our case, in particular at the low density setups, this choice of model geometry would be arbitrary, and therefore the effect it has on model dynamics is undesirable. It would jeopardise our ability to estimate model parameters from experimental data correctly. Further, we want the difference in interaction type to become apparent in the model through parameter sizes rather than the actual choice of function that should substitute for a particular term. This enables the method of parameter estimation to differentiate between possible types of cell-cell interactions that could occur in the observed biological system.

### 2.1.3 Potts model

#### Model description

When considering the interactions between two or more cell types, differential adhesion often comes into play. The cellular Potts model, first presented in Glazier and Graner (1993) highlights the role of differential adhesion in cell sorting through the incorporation of different adhesion affinity between adjacent cell types into the energy term (Hamiltonian). The model starts from a grid on which each site is labeled by its cell type based on predefined initial conditions. Each cell is made of a large number of adjacent sites on

the grid (generally around 40, need not to be simply connected grid sites), this number is subjected to an elastic constraint which acts to keep each cell size around its fixed target area.

Propagating on the grid in a Monte Carlo manner, the probability to accept a change of a given cell label  $\sigma(i, j)$  to the label of one of its adjacent sites  $\sigma'$  (chosen at random) for temperature  $T > 0$  is

$$P((\sigma(i, j) \rightarrow \sigma'(i, j))) = \begin{cases} \exp(-\Delta\mathcal{H}/kT) & , \quad \Delta\mathcal{H} > 0 \\ 1 & , \quad \Delta\mathcal{H} \leq 0 \end{cases} \quad (2.1.2)$$

and For  $T = 0$ :

$$P((\sigma(i, j) \rightarrow \sigma'(i, j))) = \begin{cases} 0 & , \quad \Delta\mathcal{H} > 0 \\ 0.5 & , \quad \Delta\mathcal{H} = 0 \\ 1 & , \quad \Delta\mathcal{H} < 0 \end{cases} \quad (2.1.3)$$

where  $\Delta\mathcal{H} > 0$  is the difference in energy gained by performing the change.

For more details see Glazier and Graner (1993), where this model was developed and used to predict the configuration which minimises the energy of the entire system. It was found that different surface energies, used to represent a range of cell-cell interactions, give rise to different levels of cell sorting or patterning.

## Applications

Later adjustments to the model include an added negative surface energies, constrained membrane area and a spin flip energy threshold to include description of cells fixed membrane area, their attractive binding and the dissipation of making and breaking membrane contacts (Ouchi et al., 2003). A more efficient algorithm which enables parallelised computing by making it based on a random walk instead of a Monte Carlo method (Chen et al., 2007), and an extension to enable it to address multi scale systems (Scianna and Preziosi, 2012). A continuous approximation to this model is presented in Lushnikov et al. (2008), using mean field approximation and other techniques.

The cellular Potts model was employed to help explore the role of extracellular matrix in glioma invasion (Rubenstein and Kaufman, 2008), and its multi scale version was used to analyse a wound healing essay (Scianna, 2015).

In the context of Eph-ephrin regulation of cell-cell interaction, the cellular Potts model was employed to model cell positioning and segregation along the intestinal crypt epithelium, controlled by differential adhesion due to varying ephrinB/EphB activation, where the differential adhesion was shown to regulate the coordinated migration of cells within the crypt (Wong et al., 2010).

## Synopsis

Potts model was shown to be applicable to many mesoscale biological systems. In particular, it translates the idea of differential adhesion into a working algorithm which can replicate and predict behaviours in relevant biological systems. In the second application of our model (chapter 6), we explore the role of Eph-ephrin regulation of cell segregation and cluster formation, which possibly acts differently to the mechanism behind differential adhesion, see for example Taylor et al. (2012) for a discussion regarding the faster segregation rate in Eph-ephrin regulated cell-cell interaction and possible differences from Cadherin regulated segregation (which is a ‘classical’ differential adhesion driven mechanism). Therefore a model which does not link back directly to differential adhesion is in need.

Further, in terms of analysis, while this model is based on mathematical equations, it does not have an explicit solution and predictions can be obtained only by employing computer simulations. Continuous approximations exist for some cases under limiting assumptions, which do not apply in our case.

## 2.2 Mathematical background, concepts and terminology

### 2.2.1 Stochastic differential equations

Let  $X(t) \in \mathbb{R}^d$ ,  $t_0 < t < T$  be a random system. Its evolution in time is described by the solution to the stochastic differential equation in the Itô sense:

$$dX(t) = \mu(X(t), t)dt + \sigma(X(t), t)dB(t) \quad \text{for } t \geq 0 \quad (2.2.1)$$

with  $B(t)$  an  $m$ -dimensional Brownian motion. The drift function  $\mu : [t_0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  accounts for the evolution in mean of the state variables in the interval  $[t, t + dt)$ . The evolution of the covariance of the states in the same time interval, which is defined in terms of the dispersion matrix  $\sigma : [t_0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times m}$  is described by the diffusion matrix  $\Sigma(X(t), t) = \sigma(X(t), t)\sigma(X(t), t)^T$ .

### 2.2.2 Stationary solution

The transition probability function  $P(y, t, x, s) = \mathbb{P}(X(t) \leq y | X(s) = x)$  that determines the weak solution to the SDE above (equation 2.2.1) is uniquely defined by the probability distribution  $p(y, t, x, s)$ :  $P(y, t, x, s) = \int_{-\infty}^y p(u, t, x, s)du$ .

The probability distribution is a fundamental solution<sup>3</sup> to the backward equation:

$$\frac{\partial u(x, s)}{\partial s} + L_s u(x, s) = 0 \quad (2.2.2)$$

and if  $\sigma(x, t)$  and  $\mu(x, t)$  are bounded and continuous functions such that

$$(A1) \quad \sigma^2(x, t) \geq c > 0,$$

$$(A2) \quad \mu(x, t) \text{ and } \sigma^2(x, t) \text{ satisfy a Hölder condition with respect to } x \text{ and } t, \text{ that is, for all } x, y \in \mathbb{R} \text{ and } s, t > 0$$

$$|\mu(y, t) - \mu(x, s)| + |\sigma^2(y, t) - \sigma^2(x, s)| \leq K(|y - x|^\alpha + |t - s|^\alpha).$$

Then the PDE (2.2.2) has a fundamental solution  $p(y, t, x, s)$ , which is unique, and is strictly positive.

If in addition  $\mu(x, t)$  and  $\sigma(x, t)$  have two partial derivatives with respect to  $x$ , which are bounded and satisfy a Hölder condition with respect to  $x$ , then  $p(y, t, x, s)$  as a function in  $y$  and  $t$ , satisfies the PDE

$$-\frac{\partial p}{\partial t} + \frac{1}{2} \frac{\partial^2}{\partial y^2} (\sigma^2(y, t)p) - \frac{\partial}{\partial y} (\mu(y, t)p) = 0 \quad (2.2.3)$$

---

<sup>3</sup>A fundamental solution of the PDE 2.2.2 is a non-negative function  $p(y, t, x, s)$  with following properties:

- i. it is jointly continuous in  $y, t, x, s$ , twice continuously differentiable in  $x$ , and satisfies equation (2.2.2) with respect to  $s$  and  $x$ .
- ii. for any bounded continuous function  $g(x)$  on  $\mathbb{R}$ , and any  $t > 0$

$$u(x, s) = \int_{\mathbb{R}} g(y)p(y, t, x, s)dy$$

is bounded, satisfies equation (2.2.2) and  $\lim_{s \uparrow t} u(x, s) = g(x)$ , for  $x \in \mathbb{R}$ .

(this equation is known as the Fokker-Planck, or the forward equation).

The stationary distribution  $\pi$ , if it exists and is twice continuously differentiable, should further satisfy  $\frac{\partial \pi}{\partial t} = 0$ , and thus the ODE:

$$\frac{1}{2} \frac{\partial^2}{\partial y^2} (\sigma^2(y) \pi) - \frac{\partial}{\partial y} (\mu(y) \pi) = 0 \quad (2.2.4)$$

If the coefficients  $\mu$  and  $\sigma$  are twice continuously differentiable with second derivatives satisfying a Hölder condition, an invariant density exists if and only if the following two conditions hold

1.  $\int_{-\infty}^{x_0} \exp \left( - \int_{x_0}^x \frac{2\mu(s)}{\sigma^2(s)} ds \right) dx = \int_{x_0}^{\infty} \exp \left( - \int_{x_0}^x \frac{2\mu(s)}{\sigma^2(s)} ds \right) dx = \infty$ ,
2.  $\int_{-\infty}^{\infty} \frac{1}{\sigma^2(x)} \exp \left( \int_{x_0}^x \frac{2\mu(s)}{\sigma^2(s)} ds \right) dx < \infty$ .

Equation 2.2.4 can be solved using the integration factor:  $e^{\left(-\int_a^x \frac{2\mu(y)}{\sigma^2(y)} dy\right)}$ . Its solution is:

$$\pi(x) = \frac{Z}{\sigma^2(x)} \exp \left( \int_{x_0}^x \frac{2\mu(s)}{\sigma^2(s)} ds \right) \quad (2.2.5)$$

where  $Z$  is obtained from  $\int \pi(x) dx = 1$ .

For more detail see for example Klebaner (2011) or Soize (1994).

### 2.2.3 Parameter estimation methods for stochastic differential equations

Assume the drift and the diffusion functions in equation 2.2.1 are known up to the value of some parameter  $\theta$ , and do not have an explicit dependence on time. The observed process is now:

$$dX(t) = \mu(X(t); \theta) dt + \sigma(X(t); \theta) dB(t) \quad \text{for } 0 \leq t \leq T \quad (2.2.6)$$

Below we describe common methods for estimation of the unknown parameter  $\theta \in \Theta$ , where  $\Theta = (\alpha, \beta)$  is the space of possible parameter values. Theoretical considerations regarding these estimators generally deal with proving an estimator is consistent for the specific choice of drift function, its convergence rate, asymptotic distribution etc'; often by employing the law of large numbers and the central limit theorem for stochastic integrals. Further theoretical discussion is out of the scope of this work, for more details see for example Rao (1999); Kutoyants (2013).

#### Maximum likelihood estimation

The likelihood ratio function for the SDE above is:

$$\Lambda(\theta, X(T)) = \exp \left\{ \int_0^T \frac{\mu(X(t); \theta)}{\sigma(X(t); \theta)^2} dX(t) - \frac{1}{2} \int_0^T \frac{\mu(X(t); \theta)^2}{\sigma(X(t); \theta)^2} dt \right\} \quad (2.2.7)$$

and the maximum likelihood estimator (MLE) is defined by:

$$\hat{\theta}_T = \operatorname{argmax}_{\theta \in \Theta} \Lambda(\theta, X(T)) \quad (2.2.8)$$

(Kutoyants, 2013)

We employ a discretised version of this method for estimation of model parameters from data in section 4.2. See a more detailed derivation, example and application therein.



### Bayesian estimation

Suppose that the unknown parameter  $\theta$  is a random variable with the prior density function  $p(\nu)$ ,  $\nu \in \Theta = (\alpha, \beta)$ . Then the bayesian estimator (BE)  $\tilde{\theta}_T$  is defined by:

$$\tilde{\theta}_T = E(\theta|X(T)) = \int_{\alpha}^{\beta} \nu p(\nu|X(T)) d\nu = \frac{\int_{\alpha}^{\beta} \nu p(\nu) \Lambda(\nu, X(T)) d\nu}{\int_{\alpha}^{\beta} p(\nu) \Lambda(\nu, X(T)) d\nu} \quad (2.2.9)$$

with  $\Lambda(\nu, X(T))$  the likelihood ratio defined in equation 2.2.7 above (Kutoyants, 2013).

### Method of moments

Denote  $m(\theta) = E_{\theta}g(\xi)$  and suppose that  $m(\nu)$  is a strictly monotone function. Then the estimator of the method of moments (EMM)  $\bar{\theta}_T$  is defined by the relation:

$$m(\bar{\theta}_T) = \frac{1}{T} \int_0^T g(X(t)) dt \quad (2.2.10)$$

(Kutoyants, 2013)

### Trajectory fitting approach

Introduce the family of stochastic processes:

$$\hat{X}_t(\theta) = X_0 + \int_0^t \mu(X(s); \theta) ds, \quad 0 \leq t \leq T, \quad \theta \in \Theta \quad (2.2.11)$$

The trajectory fitting estimator (TFE) is defined by:

$$\theta_T^* = \arg \inf_{\theta \in \Theta} \int_0^T [X(t) - \hat{X}_t(\theta)]^2 dt \quad (2.2.12)$$

(Kutoyants, 2013)

### Estimates from discrete time series

Parameter estimations from experimental observations are almost always based on sampled paths rather than continuously observed ones. Numerous estimation methods are available, some are discretised versions of continuous estimators, and some are derived particularly for estimations from discrete paths, see for example a review by Nielsen et al. (2000).

When estimating model parameters from discretely observed paths, two time scales should be noted: One is the observation rate (or sampling time scale), and the second is the integration time scale. Indeed, it is not always possible to prove these estimators are consistent or that they have unbiased convergence, and in particular, the discrete version of MLE is often computationally expensive (Timmer, 2000). However, for our choice of model, the likelihood function can be obtained explicitly and therefore it is possible to employ a discretised version of the the maximum likelihood estimation technique to the type of experimental data sets we analyse in chapters 4 and 5, see further discussion, examples and applications therein.

### 2.2.4 Branching Brownian motion

Branching Brownian motion, referred to as BBM hereafter, is a simple model of an evolving population. As the name suggests, individual particles follow Brownian motion (BM) in space, and reproduce following a continuous time Galton Watson branching process.

**The Galton Watson branching process** starts at time zero with a single particle. After an exponential time of parameter  $V$ , this particle splits into  $k$  particles according to some probability distribution  $p$  on  $\mathbb{N}$ . Then each of the new particles splits at independent exponential times independently according to the same branching rule, and so on (Bovier, 2015).

**A formal definition of BBM** is given below, see for example Etheridge (2000):

**Definition 2.2.1.** *Branching Brownian motion has three ingredients:*

1. *Spatial motion:* Each individual moves around independently in  $\mathbb{R}^d$  according to a Brownian motion.
2. *Branching rate,  $V$ :* Each individual has an exponentially distributed lifetime with parameter  $V$ . Meaning, if the particle is alive at time  $t$ , its probability to die in the time interval  $[t, t + \delta t)$  is  $V\delta t + \mathcal{O}(\delta t)$ .
3. *Branching mechanism,  $\Phi$ :* When it dies, an individual leaves (at the location where it died) a random number of offspring with probability generating function  $\Phi(s) = \sum_{k=0}^{\infty} p_k s^k$ . Offspring evolve independently of each other, but conditionally on their time and place of birth.

The distribution of the BBM can be characterised as the solution to a martingale problem, and some analysis can be performed by construction the of the Dawson-Watanabe super-process as a weak limit of the rescaled version of this characterisation (Etheridge, 2000). Other approach can be to study the process through the analysis of its maximal value, see for example Bovier (2015). Further discussion is outside the scope of this work. The interested reader is referred to literature mentioned above, Etheridge (2000), or Bovier (2015) to name a few.

## Chapter 3

# A mathematical model for cell migration and mutual interactions

*A mathematical model was formulated, starting from first principles and some basic simplifying assumptions. It describes cell migration and mutual cell-cell interactions within ‘in vitro’ experimental systems which are much larger than the size of a single cell, and are observed for 24-48 hours. Having system’s scope in mind, we kept it at the minimum complexity level required to truly replicate laboratory experimental results. More precisely, we employ a multidimensional SDE to describe the position of all cell centres and the way it changes over time. The drift term accounts for all interaction between cells and the Brownian Motion term represents internally generated independent cell migration. Weak existence and uniqueness of its solution were confirmed, and a stationary solution under certain conditions was provided. Further analysis identified the region within the parameter space for which our model predicts a stable cluster with a most likely value for the distance between neighbouring cells.*

### 3.1 Model formulation

#### 3.1.1 Model’s scope

This mathematical model is designed specifically to describe ‘in vitro’ experimental systems, in which the response of a (two dimensional) sample of cells to some manipulation on them or their environment is being observed over a long period of time. These systems can present either high, low or a continuously changing cell density, and their length and time scales of interest are much larger than the length and time scales representing the persistence of a single cell.

Internal cellular mechanisms controlling the motion of all individual cells are therefore not included, nor do we employ an ‘averaging approach’ which averages over all cells and treats those as a continuum. Alternatively, the position of the centre of each cell is described by an equation of motion which includes a random space homogenous term to account for internally generated cell motion, and a deterministic representation of all external effects (such as interactions with the environment and with other cells).

For the  $i$ ’th cell, let  $\mathbf{F}_{ij}$  be the force due the interaction with cell  $j$ , and assume the drag force due to interaction with the substrate is proportional to the velocity with constant  $\delta$ . The equation of motion is then:

$$m d\mathbf{V}_i(t) = \left( \sum_{j \neq i} \mathbf{F}_{ij} - \delta \mathbf{V}_i(t) \right) dt + \tilde{\sigma}(t) d\mathbf{W}_i(t) \quad (3.1.1)$$

where  $m$  is the mass of cell  $i$ ,  $\mathbf{V}_i$  is the velocity of its centre,  $\mathbf{q}_i \in \mathcal{D}$  is its position within the bounded<sup>4</sup>, two dimensional domain  $\mathcal{D}$ .  $\mathbf{W}_i$  is an independent random term, and  $\tilde{\sigma}$  is a  $2 \times 2$  diagonal matrix which can depend on time.

Assume the drag force  $-\delta \mathbf{V}_i(t)$  is large and the mass  $m$  is small so  $\frac{m}{\delta} \ll 1$ , and that  $\frac{\tilde{\sigma}}{\delta} = \sigma \sim \mathcal{O}(1)$ . The equation of motion gives:

$$\mathbf{V}_i(t)dt = d\mathbf{q}_i(t) = \sum \mathbf{F}_{ij}dt + \sigma(t)d\mathbf{W}_i(t) \quad (3.1.2)$$

Further assume the random term is a Brownian motion  $\mathbf{W}_i = \mathbf{B}_i$ , and formulate our first assumption:

**Assumption 1** (Individual cell motion). *The motion of a single cell can be described by a stochastic differential equation of the form:  $d\mathbf{q}_i(t) = \sum_{j \neq i} \mathbf{F}_{ij}dt + \sigma(t)d\mathbf{B}_i(t)$ . Where  $\mathbf{q}_i(t) \in \mathcal{D}$  is the location of the centre of the  $i$ th cell at time  $t$ ,  $\mathbf{F}_{ij}$  is the force exerted on cell  $i$  due to the interaction with cell  $j$ , and  $\mathbf{B}_i(t)$  is an independent Brownian Motion. All vectors are of order 2, and  $\sigma(t)$  is represented by a  $2 \times 2$  diagonal matrix.*

### 3.1.2 Mutual interaction forces

The interactions between cells are introduced as forces which depend only on the distance between cell centres:

**Assumption 2** (Radial force). *Assume the force  $\mathbf{F}_{ij}$  between the  $i$ 'th and  $j$ 'th cells located at  $\mathbf{q}_i$  and  $\mathbf{q}_j$  respectively depends only on the distance between them:*

$\mathbf{F}_{ij} = \mathbf{f}(\mathbf{q}_i - \mathbf{q}_j)$ . *If we further assume that  $\mathbf{f}$  is a smooth function, then  $\exists$  a potential function  $u$ , s.t  $\mathbf{f} = -\nabla u$ .*

All forces which take place in the interaction between two cells are divided into two groups: attraction and repulsion forces, and represented by only two functions which are characterised by their amplitudes and length scales.

**Assumption 3** (Repulsion and attraction forces). *For any pair of cells one can write:  $\mathbf{f} = \mathbf{f}^r - \mathbf{f}^a$ . Where  $\mathbf{f}^r$  is the repulsion force and  $\mathbf{f}^a$  is the attraction force. These are characterised by their amplitudes  $R, A$  and their length scales  $r, a$  respectively.*

*We further assume  $\mathbf{f}^r$  and  $\mathbf{f}^a$  are positive, monotonically decreasing function.*

These assumptions are often used in deterministic descriptions of swarm behaviour, see Mogilner et al. (2003) for example. Here, the repulsion force corresponds to surface tension and all other exclusion forces which prevent the cells from occupying the same space. The attraction force corresponds to cell adhesion and all other forces which act to keep cells close together.

To summarise this, the interactions between cells are represented by radial forces (i.e. depend only on the distance between cell centres) and divided into two groups: attraction and repulsion forces  $\mathbf{f}^r$  and  $\mathbf{f}^a$  respectively. Thus

$$\mathbf{F}_{ij} = \mathbf{f}(\mathbf{d}_{ij}) = \mathbf{f}^r(\mathbf{d}_{ij}) - \mathbf{f}^a(\mathbf{d}_{ij}) \quad (3.1.3)$$

with  $\mathbf{d}_{ij} = \mathbf{q}_i - \mathbf{q}_j$ .

---

<sup>4</sup>Note this can be introduced analytically by adding a 'boundary' term to the potential  $U_{\mathcal{D}}$ , which ensures that  $\bar{U}(\mathbf{q}) = U(\mathbf{q}) + U_{\mathcal{D}}(\mathbf{q}) \rightarrow \infty$  as  $|q_i|$  or  $|q_{ij}| \rightarrow L$ , with  $U(\mathbf{q})$  the interaction potential mentioned below, and  $L$  the relevant length scale representing the boundary of the domain. In the program written to simulate our model's predictions this was introduced as periodic boundary conditions (the difference between implementing periodic or reflective boundaries was tested in early stages of the development of this program, it was found insignificant and thereafter periodic boundary conditions were implemented).

### 3.1.3 A specific choice of potential function

Take  $\mathbf{f}$  to be a smooth function, so there exists a potential function  $u$ , s.t  $\mathbf{f} = -\nabla u$ .

Our specific choice for the functional form of the potential in this model is:

$$\begin{aligned} u(\mathbf{d}) &= u^r - u^a, \quad \text{where} \\ u^r(\mathbf{d}) &= rRe^{-|\mathbf{d}|/r} \quad \text{and} \quad u^a(\mathbf{d}) = aAe^{-|\mathbf{d}|/a} \end{aligned} \quad (3.1.4)$$

Where  $|\mathbf{d}|$  is the Euclidean distance,  $R$  and  $A$  set the magnitudes, and  $r$  and  $a$  set the length scales of the repulsion and attraction forces respectively. Subsequently  $\mathbf{f}^r = -\nabla u^r$ , and  $\mathbf{f}^a = -\nabla u^a$ .

More explicitly, the forces are:

$$\begin{aligned} \mathbf{f}^r(\mathbf{q}_i - \mathbf{q}_j) &= Re^{-d_{ij}/r} \left( \frac{q_i^x - q_j^x}{d_{ij}} \cdot \hat{x} + \frac{q_i^y - q_j^y}{d_{ij}} \cdot \hat{y} \right) \\ \mathbf{f}^a(\mathbf{q}_i - \mathbf{q}_j) &= -Ae^{-d_{ij}/a} \left( \frac{q_i^x - q_j^x}{d_{ij}} \cdot \hat{x} + \frac{q_i^y - q_j^y}{d_{ij}} \cdot \hat{y} \right) \end{aligned} \quad (3.1.5)$$

Where  $d_{ij} = |\mathbf{d}_{ij}| = |\mathbf{q}_i - \mathbf{q}_j|$  is the Euclidean distance between the  $i$ 'th and  $j$ 'th cells, and  $(q_i^x, q_i^y)$  are the  $(x, y)$  coordinates of the  $i$ 'th cell.  $\hat{x}$  and  $\hat{y}$  are the unit vectors in the  $x$  and  $y$  directions respectively.

This choice of potential function comes from a group of functions commonly used in describing swarm behaviour (Mogilner et al., 2003). We particularly chose exponential descent to describe the forces between cells in our system, due to the fact that cell-cell interactions are preliminary short distant and drop very rapidly as cells move further apart. While this choice might seem (in the mathematical sense) somewhat arbitrary, the analysis that follows can, to some extent, be adjusted to fit any smooth monotonically decreasing functions. Limiting to a specific choice of a potential function at this stage indeed leads to some loss of generality, but our particular choice of potential function still allows for a versatile representation of possible cell-cell interactions: Different values of the characteristic parameters above will result in either total repulsion, total attraction, equilibrium distance, or other (non biological) interactions between each two cells (see section 3.2.3 for more detail).

### 3.1.4 Model equations

Consolidate assumptions 1-3 above to formulate our model's equations:

$\mathbf{q}(t)$ , the vector holding all positions of all cell centres at time  $t$  is given by:

$$\begin{aligned}
 d\mathbf{q}(t) &= -\nabla U(\mathbf{q}(t))dt + \sigma(t)d\mathbf{B}(t) \\
 \text{where } U &= \frac{1}{2} \sum_{i=1}^N \sum_{j \neq i}^N u_{ij} = \sum_{i=1}^{N-1} \sum_{j=i+1}^N u_{ij} \\
 \text{with } u_{ij} &= rRe^{-\frac{d}{r}} - aAe^{-\frac{d}{a}} \\
 \text{and } d_{ij} &= \sqrt{(q_i^x - q_j^x)^2 + (q_i^y - q_j^y)^2} \equiv d
 \end{aligned} \tag{3.1.6}$$

Note that  $\nabla_i^x u_{ij} = \left(-Re^{-\frac{d_{ij}}{r}} + Ae^{-\frac{d_{ij}}{a}}\right) \left(\frac{q_i^x - q_j^x}{d_{ij}}\right)$ , and the explicit representation of the SDE for the motion of the  $i$ 'th cell in the  $x$  direction gives:

$$\begin{aligned}
 dq_i^x &= -\nabla_i^x U dt + \sigma dB_x^i(t) \\
 dq_i^x &= \sum_{j \neq i}^N \left(Re^{-\frac{d_{ij}}{r}} - Ae^{-\frac{d_{ij}}{a}}\right) \left(\frac{q_i^x - q_j^x}{d_{ij}}\right) dt + \sigma(t)dB_x^i(t)
 \end{aligned} \tag{3.1.7}$$

and similarly for the  $y$  coordinate.

## 3.2 Model analysis

In our model equation, the drift term is  $F(x, t) = -\nabla U(\mathbf{X})$ , and we set  $\sigma(x, t) = \sigma$  to be a constant, which is sufficient to describe the biological systems we explore here (see biological application chapters 5 and 6 for more details).

### 3.2.1 Existence and Uniqueness of solutions

**Lemma 3.2.1.** *The SDE given in equation 3.1.6 above has weak solution and the weak uniqueness property.*

*Proof.* In equation 3.1.6 we limit the diffusion coefficient  $\sigma(t)$  to be global Lipschitz, and the drift term  $-\nabla U(\mathbf{q}(t))$  is bounded. Thus it has a weak solution and the weak uniqueness property. See for example Theorem 37.7 in (Métivier, 1982).  $\square$

### 3.2.2 Stationary Solution

The stable solution to our system can be obtained by solving the Fokker Planck (or forward) equation with  $\frac{\partial \pi}{\partial t} = 0$  (see appendix 2.2.2 or (Klebaner, 2011; Soize, 1994) for more details). For the drift term we have  $\mu = -\nabla U$ , and the diffusion coefficient  $\sigma(y) = \sigma$  is a constant. Indeed the drift term is once and the diffusion coefficient is twice differentiable, and they meet the a Hölder condition (as in 2.2.2). Condition 1 for the existence of the invariant distribution is readily met. As for condition 2, we need to emphasise our model is on a bounded domain ( $\mathcal{D}$ ), so the integration is over a finite interval<sup>5</sup>. Thus condition 2 also holds.

Now  $\int_a^x \frac{2\mu(y)}{\sigma^2(y)} dy = -\frac{2U(x)}{\sigma^2} + \text{const}$  (some meaningless constant), and for the solution we have:

**Lemma 3.2.2** (Invariant probability distribution). *The weak solution to the SDE in equation 3.1.7 is given by the invariant distribution:*

$$\pi(x) = \frac{1}{\tilde{Z}} \exp\left(-\frac{2U(x)}{\sigma^2}\right) \quad (3.2.1)$$

with  $\tilde{Z} = \int_{\mathcal{D}} \exp\left(-\frac{2U(x)}{\sigma^2}\right) dx$ .

### 3.2.3 Analysis of the Stationary Solution

The weak solution to our system is obtained in terms of an invariant probability distribution, implying the probability to find the system in a any given state does not change over time. Therefore, once the system reached its steady state, the most likely state for the system to be in is the one for which the probability distribution obtains a maximal value. More precisely, the most likely state  $\mathbf{q}_m$  is the one that maximises the value of  $\pi(x = \mathbf{q}_m)$  or equivalently, minimises the value of  $U(x = \mathbf{q}_m)$  ( $= -\log(\pi(\mathbf{q}_m))$ ):

$$\mathbf{q}_m = \operatorname{argmin}_{\mathbf{q} \in \mathcal{D}} U(\mathbf{q}) \quad (3.2.2)$$

(if more than one solution exists, either of the solutions can be considered).

Thus we now need to explore the solution to our system based on the analysis of the function  $U$  under different choices of parameter sizes.

---

<sup>5</sup>Alternatively one can use the amended potential  $\bar{U}$  mentioned in footnote 4 before. The weak solution will be  $\bar{\pi}(x) = \frac{1}{\bar{Z}} \exp\left(-\frac{2\bar{U}(x)}{\sigma^2}\right)$  with  $\bar{Z} = \int \exp\left(-\frac{2\bar{U}(x)}{\sigma^2}\right) dx$

Consider our model with rescaled variables:  $\hat{q} = \frac{q}{r}$  and  $\hat{t} = t \frac{A}{r}$ . Away from the boundary, our model equation for a single coordinate (equation 3.1.7) will then be

$$d\hat{q}_i^x(\hat{t}) = \sum_{j \neq i}^N \frac{\hat{q}_i^x - \hat{q}_j^x}{\hat{d}_{ij}} \left( K e^{-\hat{d}_{ij}} - e^{-\frac{\hat{d}_{ij}}{\hat{t}}} \right) d\hat{t} + \hat{\sigma} d\hat{B}_i^x(\hat{t}) \quad (3.2.3)$$

and the scaled model parameters are:  $l = \frac{a}{r}$ ,  $K = \frac{R}{A}$  and  $\hat{\sigma} = \frac{1}{\sqrt{Ar}}\sigma$ .

with  $\hat{d}^{ij} = \frac{d^{ij}}{r}$ , and  $\hat{B}(\hat{t}) = B\left(\hat{t} \frac{r}{A}\right) = \sqrt{\frac{r}{A}} B(\hat{t})$ .

The rescaled potential is

$$\hat{U} = \frac{1}{2} \sum_{i=1}^N \sum_{j \neq i}^N \hat{u}_{ij} = \sum_{i=1}^{N-1} \sum_{j=i+1}^N \hat{u}_{ij} \quad (3.2.4)$$

$$\text{with } \hat{u}_{ij} = K e^{-\hat{d}_{ij}} - l e^{-\frac{\hat{d}_{ij}}{\hat{t}}}$$

and omit the  $\hat{\phantom{x}}$  symbols from here on. This helps consolidate the number of model parameters to the minimum necessary (Lin et al., 1988).

Look at the  $l-K$  parameter space, considering the function  $U$  assuming an equidistance spacing between all neighbouring individuals as suggested by Mogilner et al. (2003). They showed that when all individuals are spaced within an equal distance to all their nearest neighbours,  $U$  can be found analytically and obtains a minimal value if  $l$  and  $K$  are in region 4a, see figure 3.1(a) below. For example, in the one dimensional case, and under the assumption that all individuals are spaced within an equal distance,  $\delta$ , to all their nearest neighbours, the potential  $U_i$  for individual  $i$  was found analytically:

$$U_i = 2 \sum_j^\infty u_{ij}(\delta j) = 2 \left( \frac{-l}{e^{\delta/l} - 1} + \frac{K}{e^\delta - 1} \right) \quad (3.2.5)$$

and could then be analysed for different regions in  $l-K$  space. It was then generalised to two and three dimensions and a variety of cluster geometries (Mogilner et al., 2003).

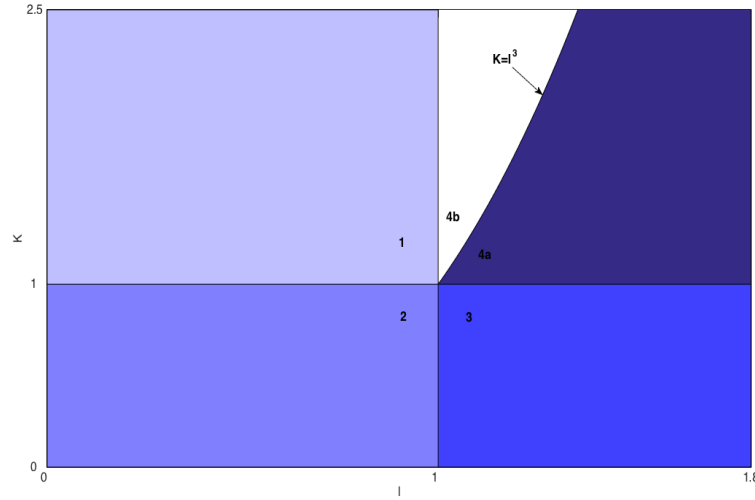
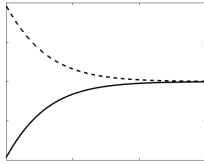
Based on the analysis of the potential between two individuals ( $u_{ij}$ ) and of the total potential ( $U_i$ ) for the 2 dimensional case as mentioned above, we conclude that our stable solution can have different characteristics based on the choice of parameter sizes in the  $l-K$  space, summarised below and in figure 3.1.

In particular, we will have a stable solution with a most likely value for the distance between all neighbouring cells if  $K > 1$ ,  $l > 1$  and  $\gamma l^3 < K$ , where  $\gamma \sim 1$  is a constant for which the actual value is determined by the specific assumptions made regarding the cluster's geometry, as explained in Mogilner et al. (2003). In all other cases the probability density does not obtain a maximal value at a finite distance away from zero and thus does not give a prediction for a most likely value for the distance between neighbouring cells.

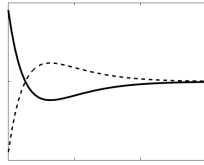
**Region 1**  $K > 1$  and  $l < 1$ . The probability density does not have a maximal value at a finite distance, see figure 3.1(b). This is a case of net repulsion - cells will move away from each other (neither  $u_{ij}$  nor  $U_i$  in the case of equal spacing has a local minima at a finite inter-individual distance).

**Region 2**  $K < 1$  and  $l < 1$ . The probability distribution function does not obtain a maximal value away from zero, figure 3.1(c). All individuals collapse to a point if all mutual distances are closer than a certain distance, and move apart if all mutual distances are greater than some other given distance.

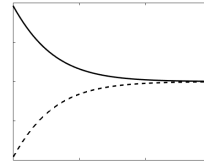


(a) Different regions within the  $l$ - $K$  parameter space.

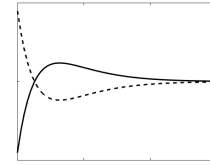
(b) Region 1



(c) Region 2



(d) Region 3



(e) Region 4

Figure 3.1: Bifurcation diagram over the  $l - K$  parameter space. Images (b)-(e) present  $U$  in a dashed line and the probability density  $\propto \log(\pi)$  in solid line. In regions 1 to 3, the probability density which follows  $\log(\pi)$ , does not display a maximal value away from zero, meaning a most likely value between two individuals does not exist. In region 4, a most likely value for the distance between two individuals exists. In the case of many cells, the net potential (sum over all interactions, assuming a solution with equidistant spacing between neighbouring cells) function has been analysed in Mogilner et al. (2003). It has been shown that in region 4a it has no minima other than zero i.e. - a most likely value for the equi-distance spacing does not exist. In region 4b the equal spacing potential was found to have a minimal value at a finite distance (Mogilner et al., 2003), meaning the stable solution in the case of many cells has a most likely value for the equal spacing distance.

**Region 3**  $K < 1$  and  $l > 1$ . Probability density function does not have a maximal value away from zero, this case also fails to lead to an aggregate with a well-defined individual distance, figure 3.1(d). This is a case of net attraction, the functions are not bounded from below when the mutual distances between individuals all tend to zero.

**Region 4**  $K > 1$  and  $l > 1$ .

**N = 2** In zone 4, where  $U_i$  has one minima, one will get a stable equilibrium at a ‘preferred distance’  $r_0$ , between the cells. When translated back into our invariant probability distribution, this means we have one value for the distance between cells which is the most likely value, figure 3.1(e). This distance can be readily calculated:  $r_0 = \frac{ar \log(Ar/Ra)}{2(r-a)}$ , and is used later in the graphic representations of our numeric solutions as the effective radius of all simulated cells.

**N  $\gg$  2**

Similarly, when considering the potential for many particles, it has been shown that it obtains a minima where all cells are positioned within an equal distance,  $\delta$ , from their nearest neighbours as long as  $\gamma l^3 < K$  (where the constant  $\gamma \sim 1$  obtains a specific value depending on cluster’s geometry, e.g. number of nearest neighbours). It has been further shown that no minima exists if  $\gamma l^3 > K$  and thus there is no stable equidistance at this parameter range (Mogilner et al., 2003).

**Region 4a** For many particles: Probability density under the assumption of an equidistant solution obtains a maximal value at the ‘most likely spacing distance’ between neighbouring cells

**Region 4b** For many particles: Probability density under the assumption of an equidistant solution does not obtain a maximal value away from zero.

## Chapter 4

# Estimation of model parameters from experimental data containing cell paths

*A method for estimating model parameters from experimental data containing particle paths was formulated. First, it was shown that under suitable conditions  $\sigma(t)$  can be calculated directly from these paths, and that Dambis Dubins-Schwarz (DDS) theorem can be applied to ‘change time’ so that the model is reduced to have a constant  $\sigma$ , so that  $\sigma(t) = \sigma$ . For the drift term, likelihoods of different parameter values given particle paths were obtained using Girsanov theorem, and then used to estimate actual parameter values by maximising these likelihoods (MLE). Accuracy of this method was tested by comparing estimated parameter values given computer generated paths to the actual values used to generate these paths.*

### 4.1 Reconstructing $\sigma$ directly from cell paths

The SDE describing the displacement of a single cell at time  $t$  given in equation 3.1.7 is a specific case of an Itô process of the form:

$$d\mathbf{q}_i(t) = \mu_i(\mathbf{q}(t), t)dt + \sigma(t)d\mathbf{B}_i(t) \quad \text{with } i = 1, \dots, N \quad (4.1.1)$$

Where as before:  $\mathbf{q}_i(t)$  is the location of the  $i$ 'th particle at time  $t$ ,  $\mu_i$  is its drift term,  $\mathbf{B}_i(t)$  is the  $i$ th particle's independent BM (all are 2-dimensional vectors) and  $\sigma(t)$  is a 2x2 matrix.  $\mathbf{q}(t)$  is a 2xN vector which includes 2 location coordinates for all N particles, and  $\mu$  is a 2xN general function representing the drift term.

Each particle's path can be calculated from Itô equations:

$$\begin{aligned} q_i^x(t) &= q_i^x(0) + \int_0^t \mu_i^x(\mathbf{q}(s), s)ds + \int_0^t \sigma_i^x(s)dB_i^x(s) \\ q_i^y(t) &= q_i^y(0) + \int_0^t \mu_i^y(\mathbf{q}(s), s)ds + \int_0^t \sigma_i^y(s)dB_i^y(s) \quad \text{with } i = 1, \dots, N. \end{aligned} \quad (4.1.2)$$

Where the mean displacement is

$$\begin{aligned} \mathbb{E}_i(q_i^x(t) - q_i^x(0)) &= \mathbb{E} \int_0^t \mu_i^x(\mathbf{q}(s), s)ds \\ \mathbb{E}_i(q_i^y(t) - q_i^y(0)) &= \mathbb{E} \int_0^t \mu_i^y(\mathbf{q}(s), s)ds \end{aligned} \quad (4.1.3)$$

since  $\mathbb{E}_i \left( \int_0^t \sigma_i^x(\mathbf{q}(s), s) dB_i^x(s) \right) = 0$ , and the same holds for the  $y$  coordinate, the quadratic variation (QV) is

$$\begin{aligned} \langle q_i^x, q_i^x \rangle(t) &= \int_0^t \left( \mu_i^x(\mathbf{q}(s), s) \right)^2 ds + \int_0^t (\sigma_i^x(s))^2 dB_i^x(s) \\ \langle q_i^y, q_i^y \rangle(t) &= \int_0^t \left( \mu_i^y(\mathbf{q}(s), s) \right)^2 ds + \int_0^t (\sigma_i^y(s))^2 dB_i^y(s) \end{aligned} \quad (4.1.4)$$

and  $\langle q_i^x, q_i^y \rangle(t) = \langle q_i^x, q_j^x \rangle(t) = \langle q_i^x, q_j^y \rangle(t) = 0$   
see e.g Karlin and Taylor (1981).

**Low density condition:** Importantly, in order to be able to reconstruct  $\sigma(t)$  from known cell paths labeled  $\mathbf{X}(t)$  hereafter, which can either be observed or simulated, it is necessary that these paths were obtained in low cell density conditions, so that all cells  $i \in [1..N]$  are positioned far enough from each other such that  $\mu_i(\mathbf{X}(t)) \approx 0$  for all  $i$  at any given time  $t$ .

When this condition holds, and observed paths are substituted into the equations for the QV, we have:  $\mu_i^x(\mathbf{X}(s), s) ds \rightarrow 0$ ,  $\mu_i^y(\mathbf{X}(s), s) ds \rightarrow 0$ , and:

$$\begin{aligned} \mathbb{E}(\mathbf{X}(0) - \mathbf{X}(t)) &\rightarrow 0 \\ d \langle X_i^x, X_i^x \rangle(t) &= (\sigma^x(t))^2 dt \\ d \langle X_i^y, X_i^y \rangle(t) &= (\sigma^y(t))^2 dt \\ d \langle \mathbf{X}_i, \mathbf{X}_i \rangle(t) &= \sigma \sigma^T(t) dt \\ \langle \mathbf{X}_i, \mathbf{X}_i \rangle(t) &= \int_0^t \sigma \sigma^T(s) ds \end{aligned} \quad (4.1.5)$$

#### 4.1.1 Obtain $\sigma$ from the variance of the sample of cell paths

Under the low density condition described above, all particles follow the exact same evolution rule. I.e  $\sigma_i(t)$  and  $\sigma_j(t)$  are the same for all  $i$  and  $j$ , and  $\mu_i$  or  $\mu_j$  all go to zero for any  $t$ . Thus, equation 4.1.1 is exactly the same for any choice of  $i$ . This means we can think of each observed cell path as a different realisation of the solution to equation 4.1.1 with  $\mu_i \rightarrow 0$ , and the theoretical characteristics of the mean and quadratic variations of the Itô process under the low density condition given in equations 4.1.5-4.1.5 can be associated with the statistical mean and variance calculated from the sample of observed cell paths.

$$\begin{aligned} \text{mean}(X_i(t) - X_i(0)) &= \frac{1}{N} \sum_{i=1}^N X_i(t) - X_i(0) \\ \implies \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N X_i(t) - X_i(0) &\rightarrow 0 \end{aligned} \quad (4.1.6)$$

$$\begin{aligned}
\text{var}(X_i(t) - X_i(0)) &= \frac{1}{N} \sum_{i=1}^N (X_i(t) - X_i(0))^2 \\
&\Rightarrow \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (X_i(t) - X_i(0))^2 \sim \\
&\mathbb{E}((q(t) - q(0))^2) = \langle q, q \rangle(t) = \int_0^t \sigma \sigma^T(s) ds \quad (4.1.7)
\end{aligned}$$

Thus  $\int_0^t \sigma \sigma^T(s) ds$  can be reconstructed from the variance of the sample of all observed cell paths  $\mathbf{X}(t)$ <sup>6</sup>.

#### 4.1.2 Obtain $\sigma$ from the variance of the sample of cell displacements

When handling experimental data, often particle paths are not observed in full. This is mainly due to the tracking algorithm used: It may lose track of a certain cell, or gain track of a cell that was not observed before. Further, cells may leave (or enter) the observed frame throughout the experiment, or simply stop (or start) moving, all can result in ending (or starting) the recording of their paths. Therefore, instead of obtaining  $\mathbf{X}(t)$ , the locations of all particles at all observation times as an input, one may have some fragmented version of it. For this reason we suggest to apply a slightly altered method, described below.

Start from the single equation which describes the displacement of any particle (equation 4.1.1 under the condition of low density):  $dq(t) = \sigma(t)dB(t)$ , but now integrate only over a single time step:

$$\begin{aligned}
q(t_n) &= q(t_{n-1}) + \int_{t_{n-1}}^{t_n} \sigma(s)dB(s) \\
\Delta q(t_n) &= \int_{t_{n-1}}^{t_n} \sigma(s)dB(s) \quad (4.1.8)
\end{aligned}$$

and:

$$\begin{aligned}
d\langle q, q \rangle(s) &= \sigma \sigma^T(s) ds \\
\langle q, q \rangle(t_n - t_{n-1}) &= \int_{t_{n-1}}^{t_n} \sigma \sigma^T(s) ds \\
\Delta \langle q, q \rangle(t_n) &= \int_{t_{n-1}}^{t_n} \sigma \sigma^T(s) ds \quad (4.1.9)
\end{aligned}$$

Note that in this setup, even when the condition of low density is eased, if time intervals  $t_n - t_{n-1} = \delta$  are small, then  $\int_{t_{n-1}}^{t_n} \mu(q(s))ds$  is much smaller than the stochastic integral  $\int_{t_{n-1}}^{t_n} \sigma(q(s))dB(s)$ . Because the first integral is of order  $\delta$  and second is of order  $\sqrt{\delta}$ . Therefore the effect of the drift term  $\mu$  would be negligible in the QV and could be dropped for high cell densities too.

---

<sup>6</sup>Note this can also be related to the mean square displacement,  $\mathbb{E}\left(\frac{(X_i(t) - X_i(0))^2}{t-0}\right)$  often mentioned in literature regarding cell migration (Ionides et al., 2004).

Following the same arguments as before, but this time for the observed displacements defined by:  $\Delta \mathbf{X}(t_n) = \mathbf{X}(t_n) - \mathbf{X}(t_{n-1})$ , rather than the entire paths, we get:

$$\begin{aligned} \text{mean}(\Delta X_i(t_n)) &= \frac{1}{N} \sum_{i=1}^N \Delta X_i(t_n) \\ &\implies \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \Delta X_i(t_n) \sim \mathbb{E}(\Delta q(t_n)) \rightarrow 0 \end{aligned} \quad (4.1.10)$$

$$\begin{aligned} \text{var}(\Delta X_i(t_n)) &= \frac{1}{N} \sum_{i=1}^N \Delta X_i^2(t_n) \\ &\implies \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \Delta X_i^2(t_n) \sim \\ &\mathbb{E}(\Delta q^2(t_n)) = \Delta \langle q, q \rangle(t_n) = \int_{t_{n-1}}^{t_n} \sigma \sigma^T(s) ds \end{aligned} \quad (4.1.11)$$

Clearly,  $\sigma$  can not be obtained in more detail than the experimental observation rate. Alternatively, it can be estimated at the observation times, assuming it is a simple function (which jumps at  $t_n$  and stays constant until  $t_{n+1}$  where it jumps again). The last integral is then:

$$\int_{t_{n-1}}^{t_n} \sigma \sigma^T(s) ds = \sigma \sigma^T(t_n)(t_n - t_{n-1})$$

and the estimate for  $\sigma$  is:

$$\sigma \sigma^T(t_n) \sim \frac{\text{var}(\Delta X_i(t_n))}{\Delta t_n}, \quad (4.1.12)$$

where  $\Delta t_n = t_n - t_{n-1}$ .

$\sigma(t)$  can then be parametrised by extrapolation or by a polynomial (or any other) curve fitting method.

### 4.1.3 Change of time

Recall we assume all Brownian motions are independent of each other and that all cells have the same value for  $\sigma_i^x(t) = \sigma_i^y(t) = \sigma(t)$ . The SDE for a single coordinate is thus:

$$dq_i^x(t) = \mu_i^x(\mathbf{q}(t))dt + \sigma(t)dB_i^x(t) \quad \text{for all } i = 1, \dots, N \quad (4.1.13)$$

One can apply DDS (Dambis, Dubins-Schwarz) theorem for change of time in a similar manner to each coordinate separately and obtain:

**Lemma 4.1.1** (Apply DDS to reduce our model to have a constant  $\sigma$ ). *Let  $q_i^x(t)$  be a solution to 4.1.13 and define  $X_i(t) = q_i(\tau_t)$ . Then  $X_i(t)$  is a weak solution to the SDE:*

$$dX_i^x(t) = \frac{\mu_i^x(\mathbf{X}(t))}{\sigma^2(t)}dt + dB_i^x(t) \quad \text{for all } i = 1, \dots, N \quad (4.1.14)$$

with  $\tau_t = G^{-1}(t)$ ,  $G(t) = \int_0^t \sigma_i^2(s)ds$ , and  $X_i^x(0) = q_i^x(0)$ .  
(and similarly for all coordinates:  $q_i^y(t), q_j^x(t), q_j^y(t)$  )

*Proof.* This is a particular case of theorem 7.41 in Klebaner (2011) with  $g(Y(t)) = \sigma^2(t)$   $\square$

One can apply this lemma in cases where data suggests that  $\sigma(t)$  is not a constant over time, and change to an equivalent model which ensures a constant  $\sigma$ . Taking this step will allow to employ our MLE method (described below) also in cases which have a non constant  $\sigma(t) (\neq \sigma)$ , to find an estimation of the drift term parameters.

## 4.2 Maximum Likelihood Estimation of drift term parameters

In our model, the drift term takes the specific form:  $\bar{\mu}(\mathbf{q}(t), t) = -\nabla U$ . For  $N$  cells in the two dimensional case, it is a vector valued function of length  $2N$ . When  $\sigma(t) = \sigma$  is constant over time, one can use the non dimensional representation of it given in equation 5.2.1. More explicitly, the drift term for a given particle  $i$  in the  $x$  and  $y$  directions is:

$$\begin{aligned}\mu_i^x(\bar{\mathbf{q}}) &= \sum_{j \neq i}^N \left( \bar{R}e^{-\bar{d}_{ij}} - \bar{A}e^{-\bar{d}_{ij}/l} \right) \left( \frac{\bar{q}_i^x - \bar{q}_j^x}{\bar{d}_{ij}} \right) \\ \mu_i^y(\bar{\mathbf{q}}) &= \sum_{j \neq i}^N \left( \bar{R}e^{-\bar{d}_{ij}} - \bar{A}e^{-\bar{d}_{ij}/l} \right) \left( \frac{\bar{q}_i^y - \bar{q}_j^y}{\bar{d}_{ij}} \right)\end{aligned}\tag{4.2.1}$$

with the rescaled variable  $\bar{q} = \frac{q}{r}$ , re-scaled model parameters  $\bar{R} = \frac{rR}{\sigma^2}$ ,  $\bar{A} = \frac{rA}{\sigma^2}$ , and  $l = \frac{a}{r}$ , and rescaled distance  $\bar{d} = \frac{d}{r}$  (further discussed in section 5.2.2).

This definition of  $\bar{\mu}$  gives rise to a ‘family’ of functions, a specific member of it is characterised by setting a specific set of values to the parameters  $\bar{R}, \bar{A}$ , and  $l$ . Let  $\alpha = (\bar{R}_\alpha, \bar{A}_\alpha, l_\alpha)$  be a specific choice of parameter values, and  $\bar{\mu}_\alpha$  is the ‘family’ member which corresponds to this choice.

To best fit our choice of  $\mu_\alpha$  to the experimental data, we use Girsanov theorem for equivalent measures and their likelihood ratio: We change the probability measure from  $\mathbb{P}$  to  $\mathbb{Q}$  so that the process will have a change of drift from  $\mu_{\alpha_0}$  to  $\mu_\alpha$ . The likelihood ratio (given cell paths obtain in experimental data) will obtain its maximum at a certain value of  $\alpha$ . This value will determined the drift term (i.e.  $\mu_\alpha$ ) best fitted to our data.

In this section, our method is stated in detail for a general function  $\bar{\mu}_\alpha(\mathbf{q}(t), t)$ , starting from the original theorem with its generalisation to many dimensions. A few examples of its application to the reconstruction of the drift term from simulated data based on different choices of  $U$  are given along with error calculations. In chapter 5 we go back and focus on the single functional form of  $\mu$  (given in equation 4.2.1 above) which corresponds to the specific potential we found to best describe the biological systems we study.

### 4.2.1 Girsanov theorem: Change of measure and the likelihoods ratio

To demonstrate our method we first look at a 1-dimensional process. Consider the 1-dimensional representation of Girsanov theorem (Girsanov, 1960; Klebaner, 2011):

**Theorem 4.2.1** (Girsanov theorem, 1-dimension). *Let  $B(t)$  be a 1-dimensional  $\mathbb{P}$ -Brownian motion, and  $H(t)$  be a predictable process with  $\int_0^T |H(s)|^2 ds < \infty$ .*

*Let  $Z = -\int_0^t H(s)dB(s) = -(H \cdot B)(t)$  and assume that  $\mathcal{E}(Z)$  is a martingale<sup>7</sup>.*

*Then there is a measure  $\mathbb{Q}$  equivalent to  $\mathbb{P}$ , such that the process:  $W(t)$ , where  $W(t) =$*

<sup>7</sup>A sufficient condition for this is that  $\mathbb{E} \left( e^{\frac{1}{2} \int_0^T |H(s)|^2 ds} \right) < \infty$ .

$B(t) + \int_0^T H(s)ds$  is a  $\mathbb{Q}$ -Brownian motion.

The likelihood ratio is given by  $\Lambda = \frac{d\mathbb{Q}}{d\mathbb{P}} = \mathcal{E}(Z)_T$ .<sup>8</sup>

Now, let  $X(t)$  be a 1-dimensional Itô process, which is defined with a  $\mathbb{P}$ -Brownian motion  $B(t)$ :

$$dX(t) = \mu_\alpha(X(t), t)dt + \sigma dB(t), \quad (4.2.2)$$

where  $B(t)$  is 1-dimensional,  $\mu_\alpha(X(t), t)$  is 1-dimensional, and  $\sigma$  is a scalar. Assume that  $\sigma^{-1}$  exists.

Under change of probability measure from  $\mathbb{P}$  to  $\mathbb{Q}$ , the process  $X(t)$  can be written using  $W(t)$ , a (1-dimensional)  $\mathbb{Q}$ -Brownian motion:

$$dX(t) = \mu_{\alpha_0}(X(t), t)dt + \sigma dW(t) \quad (4.2.3)$$

To simplify this further, choose  $\alpha_0$  such that  $\mu_{\alpha_0} = 0$ , and change notation so  $\mu = \mu_\alpha$ . Obtain the likelihood ratio for the change of measure which eliminates the drift:

**Lemma 4.2.2.** *Let  $X(t)$  be defined by equations 4.2.2 and 4.2.3 above with  $\alpha_0$  s.t.  $\mu_{\alpha_0} = 0$ , and  $\mu_\alpha = \mu$ . Then*

$$\log \Lambda = -(\sigma^{-1})^2 \int_0^t \mu(X(s), s) dX(s) + \frac{(\sigma^{-1})^2}{2} \int_0^t \mu^2(X(s), s) ds \quad (4.2.4)$$

Further, if  $\mu(X(t), t)$  is such that there is a twice differentiable function  $U(X(t), t)$  which satisfies:  $-\nabla U = \mu(X(t), t)$ , then:

$$\log \Lambda = (\sigma^{-1})^2 U(X(t)) - \frac{1}{2} \int_0^t U''(X(s)) ds + \frac{(\sigma^{-1})^2}{2} \int_0^t \mu^2(X(s), s) ds \quad (4.2.5)$$

See proof in appendix A.1.

One then needs to find  $\alpha$  which maximises the likelihood (or the log-likelihood due to monotonicity) given the data:

$$\alpha_e = \operatorname{argmax} \left( \log \Lambda(\alpha | X(t)) \right) \quad (4.2.6)$$

It can be found analytically by differentiation with respect to  $\alpha$  and equating to zero:

$$\frac{d \log \Lambda(\alpha | X(t))}{d\alpha} = 0 \quad (4.2.7)$$

Or by any other method. The drift term best fitted to the data is then  $\mu = \mu_\alpha$ , see for example Klebaner (2011).

## 4.2.2 Generalise to higher dimensions

In order to use the method above for our 2 x N model, now consider Girsanov theorem in higher dimensions (Girsanov, 1960):

**Theorem 4.2.3** (Girsanov theorem,  $d$ -dimensions). *Let  $\bar{B}(t)$  be a  $d$ -dimensional  $\mathbb{P}$ -Brownian motion, and  $\bar{H}(t)$  be a predictable process with  $\int_0^T |\bar{H}(s)|^2 ds < \infty$ .*

*Let  $Z = -\sum_{i=1}^d \int_0^t H^i(s) dB^i(s) = -(\bar{H} \cdot \bar{B})(t)$  and assume that  $\mathcal{E}(Z)$  is a martingale<sup>9</sup>.*

---

<sup>8</sup> $\mathcal{E}(Z)_T \equiv \exp \left( Z_T - \frac{1}{2} [Z, Z]_T \right)$

<sup>9</sup>A sufficient condition for this is that  $\mathbb{E} \left( e^{\frac{1}{2} \int_0^T |\bar{H}(s)|^2 ds} \right) < \infty$ .



Then there is a measure  $\mathbb{Q}$  equivalent to  $\mathbb{P}$ , such that the process:  $\bar{W} = (W^1(t), \dots, W^d(t))$ , where  $W^i(t) = B^i(t) + \int_0^t H^i(s)ds$  is a  $\mathbb{Q}$ -Brownian motion.

The likelihood ratio is given by  $\Lambda = \frac{d\mathbb{Q}}{d\mathbb{P}} = \mathcal{E}(Z)_T$ .<sup>10</sup>

Let  $\bar{X}(t)$  be a  $d$ -dimensional process, with a  $\mathbb{P}$  - Brownian motion  $\bar{B}(t)$ .

$$d\bar{X}(t) = \bar{\mu}_\alpha(\bar{X}(t), t)dt + \sigma d\bar{B}(t) \quad (4.2.8)$$

Where  $\bar{B}(t)$  is a  $d$ -dimensional vectors.  $\bar{\mu}_\alpha(\bar{X}(t), t)$  is a  $d$ -dimensional vector valued function, and  $\sigma$  is a  $d \times d$  matrix. Assume that  $\sigma^{-1}$  exists.

Under change of the probability measure, the process  $\bar{X}(t)$  can be written using a ( $d$ -dimensional)  $\mathbb{Q}$  -Brownian motion  $\bar{W}(t)$ :

$$d\bar{X}(t) = \bar{\mu}_{\alpha_0}(\bar{X}(t), t)dt + \sigma d\bar{W}(t) \quad (4.2.9)$$

This change of drift corresponds to taking  $\bar{H} = \sigma^{-1}(\bar{\mu}_\alpha(\bar{X}(t), t) - \bar{\mu}_{\alpha_0}(\bar{X}(t), t))$  in Girsanov theorem for  $d$ -dimensional diffusions given above (theorem 4.2.3). Further we have:  $H^i = \sum_j (\sigma^{-1})_{ij} (\mu_\alpha^j(\bar{X}(t), t) - \mu_{\alpha_0}^j(\bar{X}(t), t))$ .

Let  $\bar{\mu}_\alpha = \bar{\mu}$  and  $\bar{\mu}_{\alpha_0} = 0$  for simplicity, so  $\bar{H} = \sigma^{-1}\bar{\mu}(\bar{X}(t), t)$  and  $H^i = \sum_j (\sigma^{-1})_{ij} \mu^j(\bar{X}(t), t)$ . Denote  $\Sigma = \sigma^T \sigma$ , and therefore  $\Sigma^T = \Sigma$ , and  $\Sigma^{-1} = \sigma^{-1} \sigma^{-T} = \Sigma^{-T}$ , and obtain:

**Lemma 4.2.4.** Let  $\bar{X}(t)$  be defined by equations 4.2.8 and 4.2.9 above with  $\alpha_0$  s.t.  $\bar{\mu}_{\alpha_0} = 0$ , and  $\bar{\mu}_\alpha = \bar{\mu}$ . Then

$$\log \Lambda = - \int_0^t \bar{\mu}^T(\bar{X}(s), s) \Sigma^{-1} d\bar{X}(s) + \frac{1}{2} \int_0^t \bar{\mu}^T(\bar{X}(s), s) \Sigma^{-1} \bar{\mu}(\bar{X}(s), s) ds \quad (4.2.10)$$

Further, if  $\bar{\mu}(\bar{X}(t), t)$  is such that there is a twice differentiable function  $U(\bar{X}(t), t)$  which satisfies:  $-\nabla U = \bar{\mu}(\bar{X}(t), t)$ , and all the BM's are independent so  $\Sigma = \sigma^2 I$ , then:

$$\log \Lambda = (\sigma^{-1})^2 \left( U(\bar{X}(t)) - U(\bar{X}(0)) - \frac{1}{2} \sigma^2 \int_0^t \nabla^2 U ds + \frac{1}{2} \int_0^t |\bar{\mu}(\bar{X}(s), s)|^2 ds \right) \quad (4.2.11)$$

Proof can be found in appendix A.2

### 4.3 Numerical demonstration

A java code was programed to numerically solve our model's equations for three different choices of potential functions, and a given set of model parameter values, in order to generate numerically simulated cell paths. A different part of the script then analysed these paths based on the method presented in section 4.2 above to estimate the values of model parameters used to generate these paths. A calculation of the error was then performed, based on the difference between actual parameter value (used in generating the paths) and its corresponding estimated value.

This was done using three different choices for the potential function (i.e  $U$ ):

\* Linear drift

\* Linear interactions

---

<sup>10</sup>  $\mathcal{E}(Z)_T \equiv \exp \left( Z_T - \frac{1}{2} [Z, Z]_T \right)$

\* Exponential attraction and repulsion interactions<sup>11</sup>

In each of the cases, the effect of changing the integration time, the integration interval, and the sample interval on the error was demonstrated. As one would expect, smaller integration steps and longer integration times were found to yield the smaller error values. Higher sample rate ( $=1/\text{sample interval}$ ) decreases the error further.

Furthermore, error estimations in the cases of the linear drift and the linear interactions between cells could readily reach their lower limit which are determined by programmatic constraints and adjustable by the user. In the exponential interactions case, errors were found to be further dependent on the spatial arrangement of cell paths, where paths that are too far apart result in larger errors, as the potential diminishes exponentially with distance and parameter sizes can not be retrieved correctly.

#### 4.3.1 Error calculation

Let  $\alpha$  represent the set of parameters characterising the drift term  $\mu_\alpha$ . Depending on the choice of potential function,  $\alpha$  can contain a discrete number of parameters, marked  $\alpha_i$  (for the potential functions demonstrated here,  $i$  is between 1 and 4). Further, set  $\alpha^s$  to represent parameter values used in the simulation to generate cell paths, and let  $\alpha^e$  represent parameter values estimated by using our MLE method (see 4.2).

Then define some error estimators based on the difference between  $\alpha^s$  and  $\alpha^e$ :

**Absolute error** :  $err_i^{abs} = |\alpha_i^s - \alpha_i^e|$ ,

**Relative error** :  $err_i^{rel} = \frac{|\alpha_i^s - \alpha_i^e|}{\alpha_i^s}$ ,

**Errors sum** : sum of relative errors:  $err^{sum} = \sum_i \frac{|\alpha_i^s - \alpha_i^e|}{\alpha_i^s}$

These estimators are calculated for paths obtained using different choices of potential functions below. Note these errors are bounded from below by a pre-defined grid size of the parameter space, since at this stage our algorithm employs a ‘broote force’ method in order to find the arguments which maximise the likelihood ( $argmax(\log \Lambda)$ ), i.e likelihood valued are calculated at a pre-defined (passed in as an input to the code) set of grid points which constitute the parameter space on which argmax is searched for. Of course one can employ a ‘smarter’ method for finding the argmax, such as gradient decent or others, which can increase efficiency, and eliminate dependency on a pre-defined grid. We leave this for future work.

#### 4.3.2 Linear drift

Simulated particle paths were generated by numerically solving the SDE

$$d\mathbf{q}(t) = -\alpha\mathbf{q}(t)dt + \sigma d\mathbf{B}(t) \quad (4.3.1)$$

Here this is a set of  $2 \times N$  independent equations since the linear drift term does not include interactions between particles. While  $\alpha$  here is a  $2 \times N$  vector, all particles are set to follow the same rule, thus  $\alpha_i = \alpha_j$  and  $\sigma_{ii} = \sigma_{jj} = \sigma$ . In particular for this demonstration we set:  $\alpha_i^x = \alpha_j^x = 2$ ,  $\alpha_i^y = \alpha_j^y = 3$ , and  $\sigma = 0.5$ . Figure 4.1(a) below shows simulated particle paths obtained at  $\Delta t = 0.005$ .

---

<sup>11</sup>This is the specific choice for the potential function we use everywhere else in this dissertation

For the calculation of the likelihoods as given in lemma 4.2.4 above, substitute:  
 $U = \sum_{i=1}^N u_i$ , where

$$u_i = \frac{1}{2}\alpha_i^x(q_i^x)^2 + \frac{1}{2}\alpha_i^y(q_i^y)^2 \quad (4.3.2)$$

so  $\nabla_i^x u_i = \alpha_i^x q_i^x$ ,  $\nabla_i^y u_i = \alpha_i^y q_i^y$  and

$$\nabla_i^{x^2} + \nabla_i^{y^2} \equiv \nabla_i^2 u_{ij} = \alpha_i^x + \alpha_i^y \quad (4.3.3)$$

Errors sum ( $err^{sum}$ ) calculated from the difference between estimated parameter values obtained by applying MLE ( $\alpha_e$ ) and actual parameter values used to generate simulated paths ( $\alpha_s$ ) as a function of integration time  $T$ , obtained using different integration intervals ( $\Delta t$ ) are shown in figure 4.1(b), with sample interval = 1. Error estimations generally decrease with decrease in the integration interval, with the estimations from  $\Delta t = 0.01$  and  $\Delta t = 0.005$  pretty close together, both limited from below by the grid size, which at this stage is given as an input by the user, and defines the grid points within the parameter space for which the likelihoods are calculated. Errors sum ( $err^{sum}$ ) obtained using different sample intervals (with  $\Delta t = 0.005$ ) are shown in figure 4.1(c). These error estimators clearly decrease with decrease in sample interval.

Further results including simulated paths with integration intervals ( $\Delta t = 0.1, 0.05, 0.01$ ) and further error calculations including the breakdown of the errors sum to the errors in estimations of the separate parameters (i.e  $err_i^{rel}$  for  $\alpha^x$  and  $\alpha^y$ ) can be seen in appendix B.1.1.

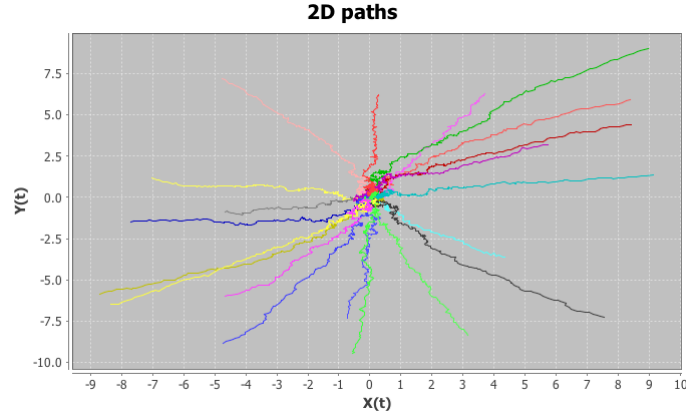
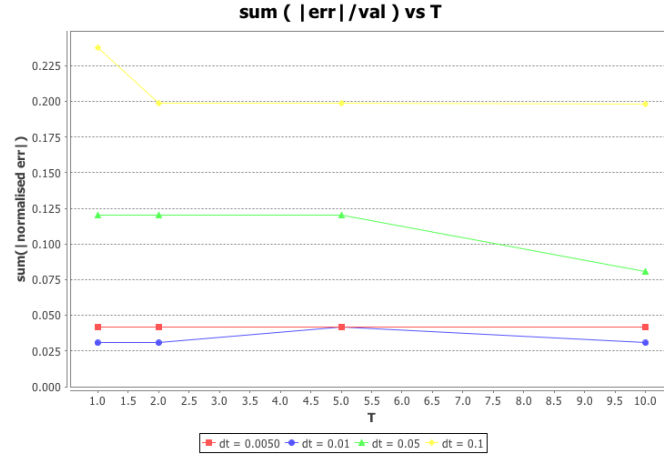
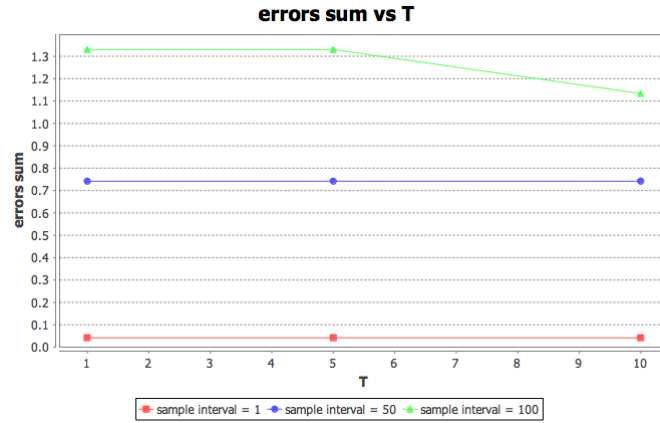
(a) computer generated sample paths,  $\Delta t = 0.005$ (b)  $err^{sum}$  for different  $\Delta t$  values(c)  $err^{sum}$  for different sample intervals at  $\Delta t = 0.005$ 

Figure 4.1: Linear drift (well potential). (a) Computer generated sample paths,  $N = 20$ ,  $\alpha_x = 2$ ,  $\alpha_y = 3$ , and  $\sigma = 0.5$ . Random initial condition was generated from a uniform distribution over a  $20 \times 20$  domain:  $\mathbf{q}(t = 0) \sim U[-10, 10]^{2N}$ . (b) Error estimations obtained for different  $\Delta t$  values (with sample interval=1). (c) Error estimations obtained for different sample intervals at  $\Delta t = 0.005$ .

### 4.3.3 Linear interactions

To verify our method can work to estimate parameter values correctly in the case of linear (attraction) interactions between agents, simulated particle paths were generated by numerically solving the SDE:

$$\begin{aligned} d\mathbf{q}_i &= -\nabla_i U dt + \sigma d\mathbf{B}_t \\ \text{where } U &= \frac{1}{2} \sum_{i=1}^N \sum_{j \neq i}^N u_{ij} \\ \text{with: } u_{ij} &= \frac{1}{2} \alpha d_{ij}^2 \end{aligned}$$

and  $d_{ij} = \sqrt{(q_i^x - q_j^x)^2 + (q_i^y - q_j^y)^2}$  as in chapter 3. Sample simulated paths are presented in figure 4.2(a).

Further, substitute the potential and its derivatives in lemma 4.2.4, to obtain estimations for parameter values.

$$\begin{aligned} \nabla_i^x u_{ij} &= \alpha d \left( \frac{q_i^x - q_j^x}{d} \right) \\ \nabla_i^y u_{ij} &= \alpha d \left( \frac{q_i^y - q_j^y}{d} \right) \\ \nabla_i^2 u_{ij} &= \alpha \end{aligned}$$

Here, since the potential is defined by a single parameter, we have  $err^{sum} = err_0^{rel}$ . Errors calculated for different integration intervals  $\Delta t = 0.05, 0.01, 0.001$  and integration times  $T = 0.5, 1, 2, 5, 7$  are presented in figure 4.2(b). Errors calculated for different sample intervals (1, 50 and 100) are presented in figure 4.2(c). Again, a decrease in error is evident with decrease in integration and in sample intervals. More results can be seen in appendix B.1.2.

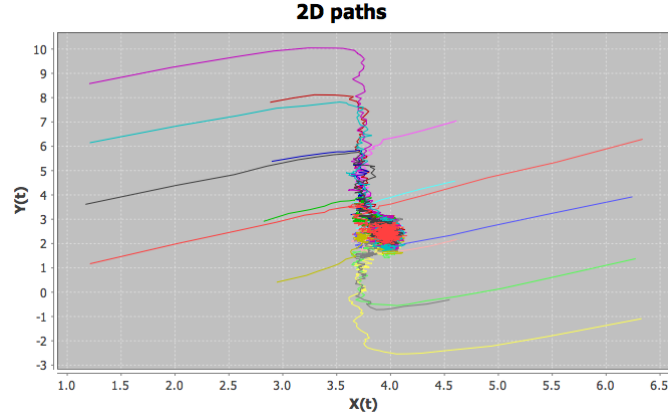
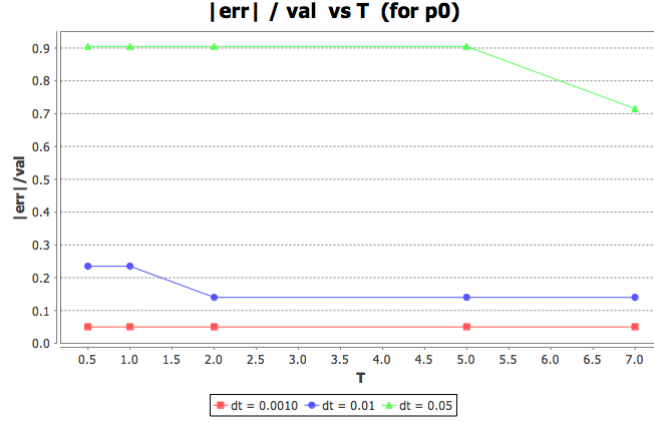
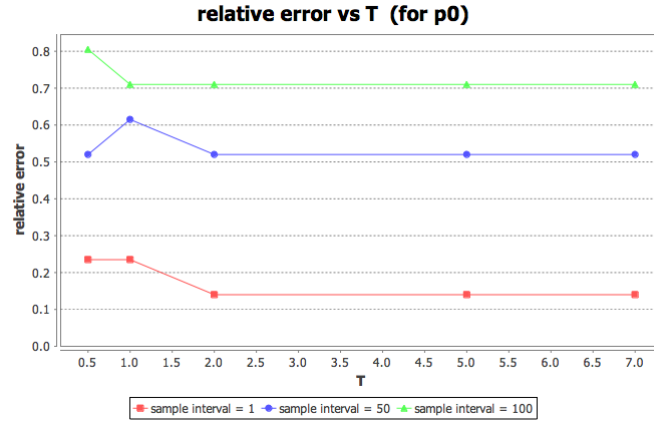
(a) Computer generated sample paths  $\Delta t = 0.01$ (b)  $err^{sum}$  for different  $\Delta t$  values(c)  $err_0^{rel}$  for different sample intervals at  $\Delta t = 0.01$ 

Figure 4.2: Linear interactions, attraction only. (a) Computer generated sample paths.  $N = 16$ ,  $\alpha = 1$ ,  $\sigma = 0.3$ . Initial condition was a grid of  $4 \times 4$  nodes, over an  $8 \times 8$  domain. (b) Error estimations obtained for different  $\Delta t$  values (with sample interval = 1). (c) Error estimations obtained for different sample interval at  $\Delta t = 0.01$ .

#### 4.3.4 Exponential repulsion and attraction interactions

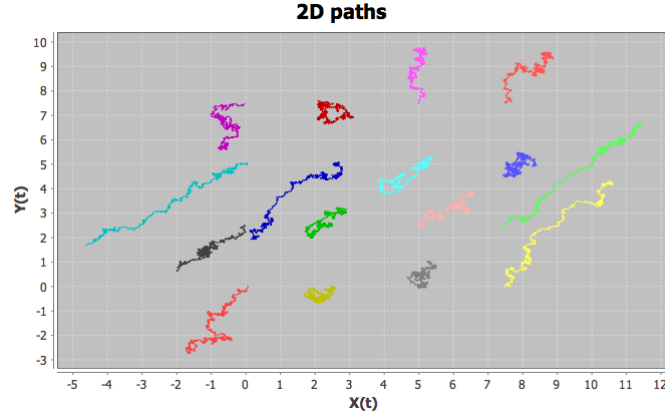
In this section we tested the ability of our method to estimate parameter values for the model we formulate and apply to different biological systems in this dissertation. Simulated particle paths were generated by numerically solving a re-scaled version of our model equations with the drift term introduced in equation 4.2.1 above (more details in chapter 3 and in section 5.2.2). A sample of simulated particle paths can be seen in figure 4.3(a) below, parameter values used were:  $\bar{R} = 25, \bar{A} = 10, l = 1.3$  and  $\sigma = 0.3$  (number of particles  $N=16$ ).

For the calculation of the likelihoods as given in lemma 4.2.4, substitute  $U, u_{ij}$ , and  $\nabla_i^2 u_{ij}$  with:

$$\begin{aligned} \text{where } U &= \frac{1}{2} \sum_{i=1}^N \sum_{j \neq i}^N u_{ij} \\ \text{end here: } u_{ij} &= \bar{R}e^{-\bar{d}} - \bar{A}e^{-\bar{d}/l} \\ \text{so } \nabla_i^x u_{ij} &= \left( -\bar{R}e^{-\bar{d}} + \bar{A}e^{-\bar{d}/l} \right) \left( \frac{\bar{q}_x^i - \bar{q}_x^j}{\bar{d}} \right) \\ \nabla_i^y u_{ij} &= \left( -\bar{R}e^{-\bar{d}} + \bar{A}e^{-\bar{d}/l} \right) \left( \frac{\bar{q}_y^i - \bar{q}_y^j}{\bar{d}} \right) \\ \nabla_i^{x2} + \nabla_i^{y2} \equiv \nabla_i^2 u_{ij} &= \bar{R}e^{-\bar{d}} - \frac{\bar{A}}{l}e^{-\frac{\bar{d}}{l}} \end{aligned}$$

Errors calculated from the difference between estimated parameter values obtained by applying MLE ( $\alpha_e$ ) and actual parameter values used to generate the simulated paths ( $\alpha_s$ ) as a function of integration time  $T$ , obtained using different integration intervals ( $\Delta t$ ) and different sample intervals are shown in figures 4.3(b)-(c) below.

Additional results including the break down to errors estimated separately for the different parameters and sample paths obtained at different integration intervals can be seen in appendix B.1.3.



(a) Cell paths

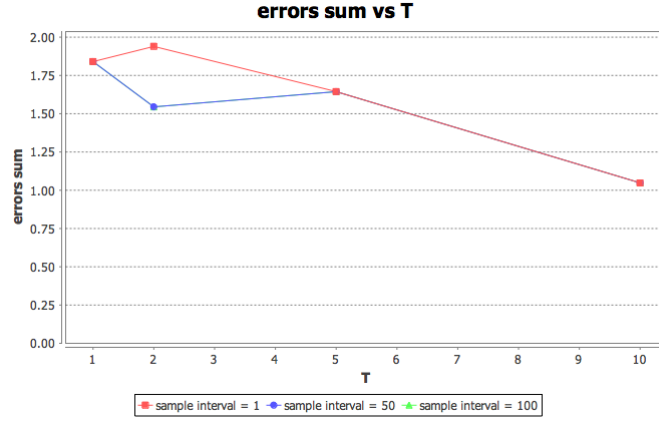
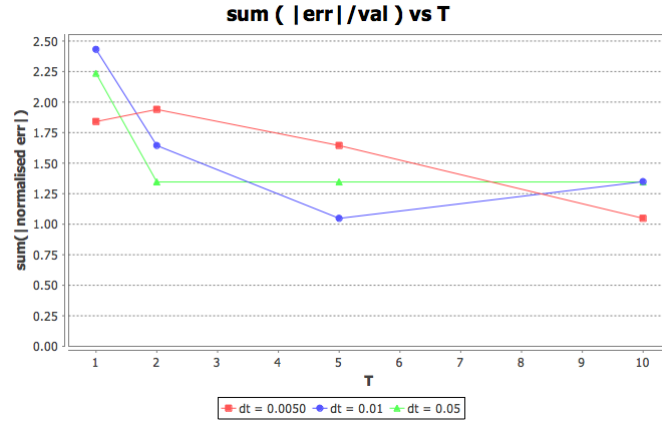
(b)  $err^{sum}$  for different sample intervals at  $\Delta t = 0.005$ (c)  $err^{sum}$  for different  $\Delta t$  values

Figure 4.3: Exponential interactions, repulsion and attraction forces. (a) Computer generated sample paths,  $N = 16$ .  $\bar{R} = 25$ ,  $\bar{A} = 10$ ,  $l = 1.3$  and  $\sigma = 0.3$ . (b) Error estimations obtained for different  $\Delta t$  values (with sample interval = 1). (c) Error estimations obtained for different sample interval at  $\Delta t = 0.005$ .



## Chapter 5

# Model application: The effect of broad beam radiation treatment on cell migration and cell-cell interactions

*The method developed in chapter 4 was applied to analyse the response of normal cells to radiation treatment. Model parameter values were estimated from cell paths which were obtained experimentally by observing normal (i.e not cancerous) fibroblast cells, after being irradiated in vitro, for 48 hours after the treatment. Our analysis showed that cell motility increases dramatically when irradiated by dosages increasing between 0 and 10 Gy, and that interactions between neighbouring cells decrease effectively to zero when irradiated by this dosage.*<sup>12</sup>

### 5.1 Biological motivation

Biological studies of a new radiation treatment, known as Micro Beam Radiation Treatment<sup>13</sup> (MBRT hereafter) which was shown to dramatically decrease cancerous tumour growth, have triggered substantial interest in understanding the way radiation treatment may effect cell motility and mutual interactions in normal and in cancerous cells.

Radiation treatments designed to eliminate cancerous tumours are based on the fact that cancerous cells display a very different response to radiation than that of normal (healthy) cells. The main and probably most well understood difference, which is also the main reason for the success of those treatments is the fact that healthy cells can recover much better and much faster from damage due to radiation than cancerous cells can. The cellular mechanisms responsible for DNA repair are amongst the first ones which get lost in the process of a cell becoming cancerous, and thus leaving it less able to recover from damage due to radiation, which is mostly pronounced as single or double strand breaks in the DNA sequence (Alberts et al., 2008).

The new MBRT studies shed light on another significant, yet not that well studied difference, which was manifested in the fact that in the cancerous tumour, the cells moved

---

<sup>12</sup>The experimental data analysed in this chapter was provided to us by Dr Jeffrey Crosbie from The Royal Women's Hospital, and Dr Camden Lo from The Micro Imaging Centre at Monash University, as a part of a wider project designed to help understand the response of both normal and cancerous cells to radiation treatment.

<sup>13</sup>A new type, still only experimental, of radiation treatment for cancer currently studied in research laboratories around the world. The innovation of it is that it applies a non uniform pattern of radiation, of which the picks may obtain much higher dosage than in the traditional radiation treatment.

much more than in the healthy tissue. As was demonstrated by Crosbie et al. (2010), the healthy irradiated tissue kept its original arrangement and maintained the pattern induced by the initial radiation (i.e alternating lines of radiation damaged cells and non damaged cells) until recovery, while the tumour cells got all intermingled - the tumour initially displayed the same damaged/not damaged pattern as the healthy tissue did, but after 24 hours damaged cells appeared everywhere throughout the tumour, intermingled amongst the non-damaged cells.

There is a vast literature describing the effect of radiation treatment on tumour growth, incorporating a variety of mathematical models mainly to help design the optimal radiation pattern that will result in the most effective elimination of the tumour (Niemierko and Goitein, 1992; Sachs et al., 2001; Rockne et al., 2009). In the past, the effect radiation has on cell motility was recorded mainly in the context of gliomas<sup>14</sup> (Kil et al., 2012). The metastatic potential of carcinoma<sup>15</sup> cancerous cells through their migratory ability was only discussed much more recently (Zheng et al., 2015), and to our knowledge, there are no mathematical models utilised to describe it so far.

While the effect of radiation treatment on cell migration and mutual interactions is clearly a question needs resolving in the context of this new experimental MBRT due to the exploration of increasing radiation dosage levels (Crosbie et al., 2010), finding an answer is also valuable in the case of the widely used traditional treatment to cancer using Broad Beam Radiation (BBR), as it may explain the increased risk of metastasis in some types of cancer after traditional radiation treatment. Here we analyse experimental data that was obtained by observing normal fibroblast cells<sup>16</sup> for 48 hours after being irradiated by BBR in vitro. We apply our method (as described in chapter 4) to find parameter values given cell paths obtained under different radiation dosage, and unveil a clear trend in the effect radiation treatment has on cell migration and mutual interactions amongst these cells.

## 5.2 Results

In this section we present initial results obtained when using the techniques in chapter 4 to estimate parameter values for the model we constructed in chapter 3, given cell paths that were obtained experimentally under different BBR levels (0, 2, 5, and 10 Gy).

Starting from our model equations 3.1.6- 3.1.7, we analysed cell paths obtained under very low cell density conditions at the different radiation levels, to derive estimation of  $\sigma$  for these radiation levels. We found that for the part of the experiment where  $\sigma(t) = \sigma$  is a constant over time (see below),  $\sigma^2$  increased from about 9 to 25  $\mu\text{m}^2/\text{min}$  when radiation levels increases from 0 to 10 Gy.

From cell paths obtained at high cell density (scratch assay<sup>17</sup> for different radiation levels, a value for  $\sigma$  can not be estimated directly from the data. To overcome this, assuming  $\sigma$  is a constant<sup>18</sup>, we consider a rescaled version of our model (equation 5.2.1 below) which allows to obtain an estimations of the drift parameters without an explicit knowledge of the value of  $\sigma$ . When analysing the data from those first 20 hours in which  $\sigma$  was a constant, a trend in parameter values was found:  $K$  and  $l$  decrease as the radiation level

<sup>14</sup>Gliomas are more fluid like tumours, where diffusion models are more often used.

<sup>15</sup>Carcinoma - cancer that develops from epithelial cells. It results in more solid tumours.

<sup>16</sup>L929 - Mouse C3H/An connective tissue cell line.

<sup>17</sup>Cell sample in which cells were grown to full confluence and then ‘scratched’ using a thin laboratory device, to create a gap for which the remaining cells can move in to

<sup>18</sup>Note that even if we do not assume a constant  $\sigma$ , we could apply DDS theory as mentioned in section 4.1.3 and obtain estimations of the parameter values for the updated drift term, see therein.

increases, which implies a decrease in the strength of the interactions between neighbouring cells from having a ‘comfortable distance’ of equilibrium representing cell-cell adhesion at zero Gy to effectively no cell-cell interactions at 10 Gy.

### 5.2.1 The effect of radiation treatment on cell motility

The effect of radiation treatment on cell motility was studied by applying the approach we describe in section 4.1.2 (in particular, equation 4.1.12), which relates  $\sigma$  in our model to the variance of the sample of cell displacements and/or to the mean square displacement (MSD) for data collected at low cell density.

At any given time step  $t_n$ ,  $\Delta X_i(t_n)$  was calculated for each observed cell path  $\mathbf{X}_i(t)$  from:  $\Delta \mathbf{X}_i(t_n) = \mathbf{X}_i(t_n) - \mathbf{X}_i(t_{n-1})$ . The variance over all those  $\Delta X_i(t_n)$ ’s was obtained for every  $t_n$ . The result for this calculation is presented as a function of time for the four radiation levels we studied in figure 5.1(a).

Although our approach can handle  $\sigma$  which is not a constant over time, the literature points to effectively constant MSD at this length and time scales (see model formulation section 3.1, and Dickinson and Tranquillo (1993); Taylor et al. (2012) for further detail). We thus chose to focus on the response to radiation during the first 20 hours after treatment, where the MSD over time can be fitted with linear curves which have close to zero slopes. The fitted curves are plotted on top of the data presented in figure 5.1(a). Curve parameters (constant value of MSD over time) are presented as a function of radiation dosage in figure 5.1(b).

This shows a clear trend which indicates an increase in cell motility due to increase in dosage level of the radiation treatment. As more data will become available to us, error estimations could be made more rigorous.

### 5.2.2 The effect of radiation treatment on cell-cell interactions

Data containing cell paths that were observed at high cell density (scratch assay) was used to estimate parameter values by applying the maximum likelihood estimation (MLE) method we described in section 4.2 (in particular given in lemma 4.2.4) with a rescaled (non-dimensional) version of our model below.

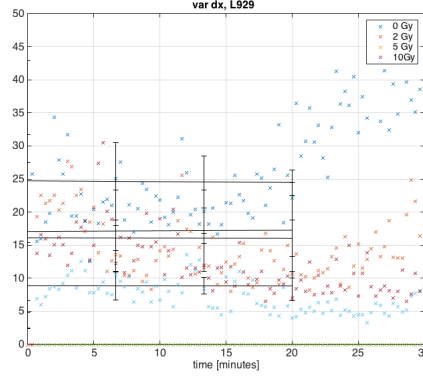
We let  $\sigma$  be a constant over time, as was found in the low density experiment for the first 20 hours, and considered the rescaled model variables  $\bar{q} = \frac{q}{r}$  and  $\bar{t} = t \frac{\sigma^2}{r^2}$ . Our model equation for a single coordinate gives

$$d\bar{q}_i^x = \sum_{j \neq i}^N \left( \bar{R} e^{-\bar{d}} - \bar{A} e^{-\bar{d}/l} \right) \left( \frac{\bar{q}_i^x - \bar{q}_j^x}{\bar{d}} \right) d\bar{t} + d\bar{B}_i \quad (5.2.1)$$

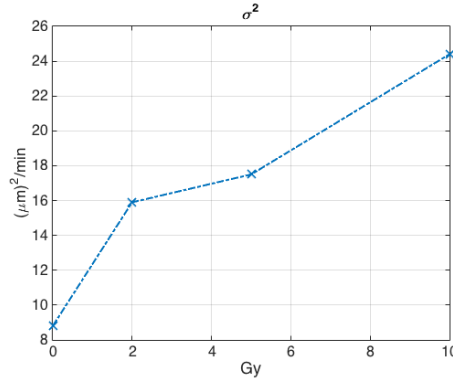
with the rescaled model parameters  $\bar{R} = \frac{rR}{\sigma^2}$ ,  $\bar{A} = \frac{rA}{\sigma^2}$ , and  $l = \frac{a}{r}$ . The rescaled distance  $\bar{d} = \frac{d}{r}$  and rescaled BM  $\bar{B}(\bar{t}) = B(\bar{t} \frac{r^2}{\sigma^2}) = \frac{r}{\sigma} B(t)$ .

This allowed us to perform our estimation despite the fact our method is not capable to retrieve  $\sigma$  value directly from data obtained at high cell density: Instead, the estimation found the non dimensional ratios:  $\bar{R}$ ,  $\bar{A}$ , and  $l$ . System’s dynamics could then be analysed looking at  $K = \frac{\bar{R}}{\bar{A}} = \frac{R}{A}$  and  $l$  as described in section 3.2.3.

Figure 5.2(a) shows an example of all cell paths that were observed between times  $t_{10}$  and  $t_{20}$  at a specific repetition within the scratch assay which had been irradiated by 0 Gy. Then figure 5.2(b) focuses on a group of close neighbours between which the potential does not diminish (recall our potential function goes to zero rapidly as the distance between



(a) Sample variance over time, from data obtained after radiation at four dosage levels (0, 2, 5, 10 Gy). Linear curves fitted to data from the first 20 hours have close to zero slopes.



(b)  $\sigma^2$  values as obtained from curves' parameters, as a function of radiation level.

Figure 5.1:  $\sigma^2$  values obtained from experimental data. Increase in radiation levels increases cell motility.

neighbouring cells increases). Our method was applied to paths of neighbouring cells which are included in such groups, from within data sets that were obtained at the different radiation levels 0, 2, 5, and 10 Gy.

A summary of all estimations obtained for the values of  $K$  and  $l$  as a function of the radiation level is presented in figure 5.3(a). The potential functions drawn with these values, indicating a decrease in cell-cell interactions with the increase of radiation levels can be seen in figure 5.3(b). In particular the potential between two individual cells changes from having one minimum at  $r_0$  at the radiation level of zero Gy, to having no salient points and staying close to zero for the 10 Gy radiation level. Meaning the change from cell cell adhesion at 0 Gy to no significant interaction between neighbouring cells at 10Gy. Based on the analysis in section 3.2.3, our model predictions change from predicting a stationary solution with the equal spacing arrangement obtaining a finite most likely distance between neighbouring cells in the 0 Gy case, to having no particular preferable arrangement in the 10 Gy case.

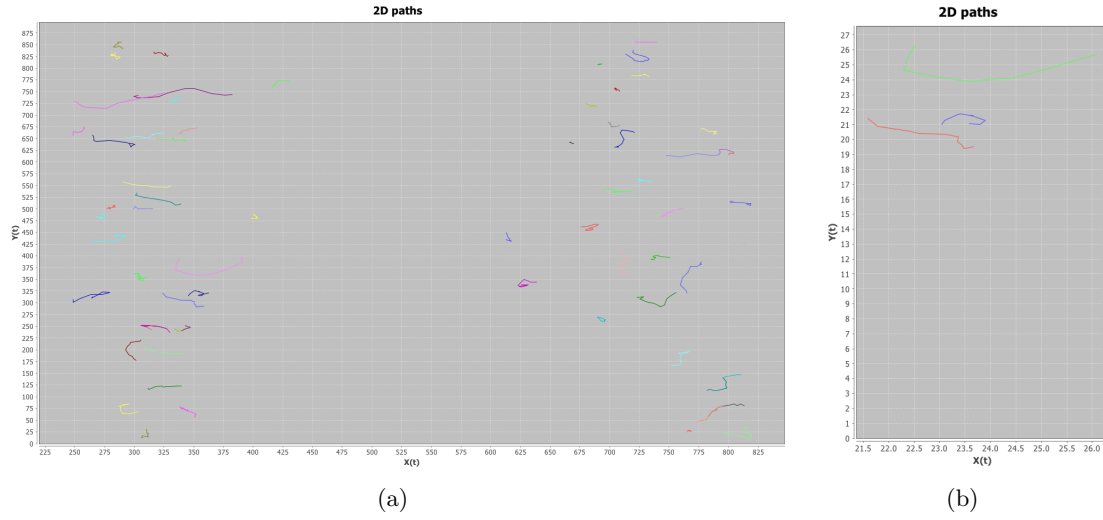


Figure 5.2: Cell paths obtained experimentally. (a) All paths observed from  $t = 10$  to  $t = 20$  (hours). (b) Focus on a group of neighbouring cells to obtain a relevant estimation of model parameters (recall forces between cells decrease rapidly with distance).

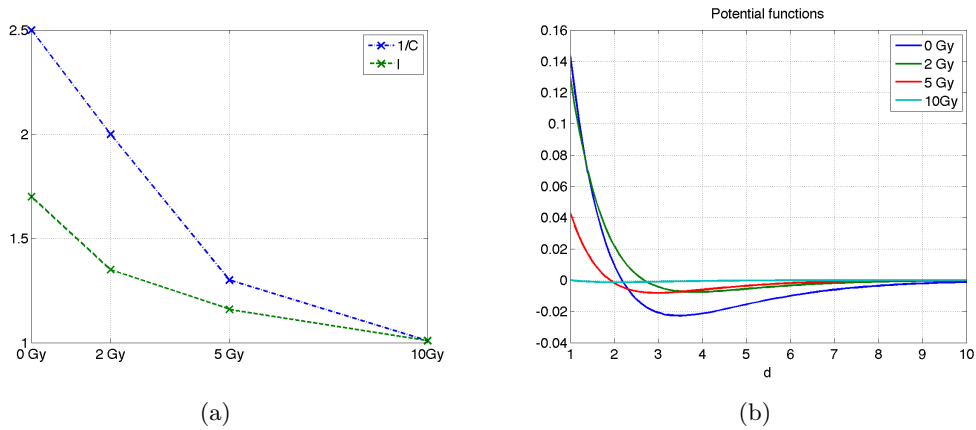


Figure 5.3: Drift term parameters estimated from experimental data. Increase in radiation levels shallows the potential well responsible for neighbouring cells having a 'comfortable equilibrium distance', effectively diminishing cell-cell adhesion.

### 5.3 Conclusions and future work

Studies of an experimental type of radiation treatment to help eliminate cancerous tumours had highlighted the need to deepen the understanding of the effect radiation treatment may have on cell motility and mutual interactions in normal and in cancerous cells (Crosbie et al., 2010). More recently, an increase in tumour metastasis potential, which manifested both as lower cell-cell adhesion and as higher cell motility, was recorded due to increase in radiation dosage (Zheng et al., 2015).

Here we demonstrate how the method formulated in chapter 4 can help analyse experimental data containing cell paths and identify trends in cellular responses to increasing dosage of radiation treatment (0,2,5 and 10 Gy). In particular, we found that cell motility almost doubles when radiated by 10 Gy, and that cell adhesion decreases effectively to zero when radiated at that level.

Although the results presented here are only preliminary in the sense they stem from analysing a limited data set regarding only a single cell type from a single experimental repetition, they already show the strength of our method in studying the way cell behaviour may change under changing conditions. These results further highlight the importance of studying the possible effect radiation treatment may have on cell migration and mutual interactions in more depth and rigour.

A direct continuation of the work presented here would be to analyse data from more experimental repetitions and thus enabling error estimations (e.g. formulating calculations of p-values). Further it would be conducive to look at the response of different cell types (cancerous cells from different types of tumours, other types of normal cells, etc.) to the same radiation treatment investigated here. Another step would be to look at higher BBR dosages, especially at low cell densities, to help understand the tampering in response strength as increasing radiation dosage to approach the physical limitations on cell motility.

In the front of MBRT (micro-beam radiation treatment) studies, as different cells in the same assay are exposed to different radiation levels, once cell paths can be recorded along with the radiation dosage each particular cell was exposed to, one could analyse cells response and compare it to the response patterns collected from BBRT (broad-beam radiation treatment) experiments. Until then, a theoretical extrapolation from data collected in BBR experiments can be valuable.<sup>19</sup>

---

<sup>19</sup>Methods for this chapter can be seen in appendix C.1.

## Chapter 6

# Extended model - The effect of Eph-ephrin regulated cell-cell interaction on cell segregation and cluster formation

*In this second application of our model we explored the way interactions between two different cell populations, which are regulated by the Eph-ephrin signalling pathway, control cell segregation and cluster formation. Our mathematical model was extended to account for two different cell types and the interactions between them, and to also include cell proliferation which we found plays a significant role in controlling the dynamics of this experimental system. The aim of this extension is to enhance the understanding of cell segregation due to Eph-ephrin regulation of cell-cell interactions by highlighting the controlling mechanisms, and studying the inter-relations between them and their effect on cell segregation dynamics and outcome. A designated Java script was written, to numerically solve the extended version of our model's equations and generate computer simulated model predictions. These predictions were tested against results obtained in the laboratory by our collaborators. A statistical analysis of the results suggested a good fit between simulated and experimental result concerning both clustering dynamics and the range of possible segregation due to manipulation of the strength of Eph-ephrin interactions. Furthermore, our model gave rise to the necessity of asymmetry between like-cell interactions in the two different populations in order to generate experimental results that had been observed in the laboratory. This finding triggered additional laboratory and research work, which confirmed this asymmetry is indeed present in the biological system studied.<sup>20</sup>*

### 6.1 Biological background

Eph receptors (Ephs) are the largest subfamily of receptor tyrosine kinases (RTKs). Their cell-cell contact dependent interaction with cell-bound ephrin ligands orchestrates cell positioning, tissues and organ patterning, and controls cell survival during normal and neoplastic development (Pasquale, 2010; Lackmann and Boyd, 2008; Janes et al., 2008). In humans, five GPI cell surface-bound type-A ephrins and three transmembrane type-B ephrins interact promiscuously with nine EphA - and five EphB family members, respectively, initiating receptor clustering, tyrosine phosphorylation and downstream forward

---

<sup>20</sup>Work presented in this chapter was conducted in collaboration with and Dr Peter Janes and Assoc. Prof. Martin Lackmann from the Department of Biochemistry & Molecular Biology at Monash University. It has been published in PLOS ONE (Aharon et al., 2014) with some variations, in memory of late Martin Lackmann, who has sadly passed away during our collaborative work on this project.

signalling into Eph-bearing cells (Lackmann and Boyd, 2008; Davis et al., 1994; Pasquale, 2008). Concurrently, ligated ephrins are drawn into a cluster on the opposite cell surface (Salaita et al., 2010), eliciting a reverse signal (Daar, 2012; Davy and Soriano, 2005), so that mutually dependent consequences from Eph/ephrin interactions are relayed into both cell populations (Egea and Klein, 2007). As a direct consequence of Eph/ephrin signalling, cells undergo changes in their actin cytoskeleton and morphology (Carter et al., 2002; Lawrenson et al., 2002; Wimmer-Kleikamp et al., 2008), their contact to neighbouring cells (Foo et al., 2006; Janes et al., 2005), substrate adhesion, motility, and their viability (Huynh-Do et al., 1999; Miao et al., 1999, 2005; Genander et al., 2009), with downstream effects on cell invasion, tissue boundary formation and on specialised secretory or immune functions, as well as on cell survival, reviewed in Pasquale (2010); Genander and Frisn (2010); Watanabe and Takahashi (2010); Nievergall et al. (2011); Batlle and Wilkinson (2012).

Ephs and ephrins are distinctively less prevalent in adult than in embryonal tissues, but are over-expressed and functionally implicated in a broad range of cancers, including breast, colon, lung and kidney carcinomas, melanomas, sarcomas, neuroblastomas, ovarian and prostate cancers and leukemias (Pasquale, 2010; Watanabe and Takahashi, 2010), and their implication in disease progression is emphasised by a conspicuous frequency of somatic Eph mutations detected in various cancers genome profiles (Watanabe and Takahashi, 2010). Not surprisingly, oncogenic Eph functions often reiterate developmental activities, and depending on tumour type, and disease stage, Eph overexpression can promote or inhibit tumour progression (Genander et al., 2009; Cortina et al., 2007; Noren et al., 2006; Miao et al., 2009), a review can be seen in Genander and Frisn (2010).

Eph/ephrin mediated tissue patterning has been modelled experimentally using isolated zebrafish embryo caps (Mellitzer et al., 1999) and cultured epithelial cell lines (Jorgensen et al., 2009; Nievergall et al., 2010; Poliakov et al., 2008), whereby co-culture of cells expressing either an Eph receptor or corresponding ephrin binding partner(s), leads to cell-cell interactions (Pasquale, 2010) that result in cell-cell segregation or de-adhesion between, and adhesion within, the two cell populations. During developmental patterning, overlapping expression gradients of multiple Ephs and ephrins (Flanagan, 2006; McLaughlin and O’Leary, 2005; Lackmann and Boyd, 2008; Xu et al., 2000) and the integrated signals from all the cell surface Eph receptors competing for available ephrin targets determine the final position and interaction partner of migrating cells (Reber et al., 2004). The complexity of the composition and regulation of such elaborate signalling clusters and emanating pathways is only beginning to be appreciated, but it is clear that the net outcomes depend on cell surface concentration, kinase-signalling capacity, type of co-expressed Ephs and ephrins and their crosstalk with other signaling systems (Nievergall et al., 2011). It is likely that the difficulty of interpreting on cellular, organ and whole organism levels experimental outcomes of manipulated Eph signalling is (at least partially) responsible for the confusing and often contradictory literature describing Eph signalling responses (Pasquale, 2010; Lackmann and Boyd, 2008). Accordingly, there is increasing appreciation that the complexity of this signalling network, which relays input from a large number of parallel cues into a range of fine-tuned cellular responses, will require mathematical modelling to reliably predict signalling outcomes (Salaita et al., 2010; Jorgensen et al., 2009; Miao et al., 1999, 2005; Reber et al., 2004).

## 6.2 Extended model

Our collaborators and others (Janes et al., 2011; Poliakov et al., 2008) have previously used 2D co-culture of ephrin-B1- and EphB2-expressing (HEK293) cells as a model of Eph/ephrin mediated segregation of the two populations. Building on existing knowledge



of Eph/ephrin-mediated cell interactions, where active bi-directional signalling in Eph- and ephrin-expressing cells leads to cell retraction and de-adhesion, overcoming usual cell-cell adhesion, we identified some basic assumptions that we considered essential for orchestrating cell segregation and cluster formation. Our approach is designed to describe the concept underlying Eph/ephrin-mediated sorting of two cell populations by using a small number of generic functions, for which the specific choice of explicit function and parameter sizes can be matched to describe experimental observations.

In the experiment we model, two cell populations are co-cultured within a monitored domain of about  $\sim 500 \times 700 \mu\text{m}$  (representing the microscope field of view). One population expresses green fluorescent protein (GFP)-tagged EphB2, and the other non-tagged ephrin-B1. The total population size grows from a few hundred to thousands within about 48 hours. Given the relative size of a single cell with respect to the size of the entire domain, we can keep all previous assumptions regarding independent cell motion and interactions with the environment and like-type neighbouring cells in place (see chapter 3). We now need to add a term which incorporates the interactions between different type cells into the potential function, and add a birth process to describe the growth of both cell populations due to cell division.

### 6.2.1 SDE describing the position of all cell centres

We start with assumptions 1-3 which are repeated here for clarity, with their formulation and more detailed explanations given previously (see section 3.1). In order to generalise our model to include interactions between cells from different populations, and to account for cell birth, we add two further assumptions (4 and 5) which are explained and formulated in the next sections.

**Assumption 1** (Individual cell motion). *The motion of a single cell can be described by a stochastic differential equation of the form:  $d\mathbf{q}_i(t) = \sum_{j \neq i} \mathbf{F}_{ij} dt + \sigma(t) d\mathbf{B}_i(t)$ . Where  $\mathbf{q}_i(t)$  is the location of the centre of the  $i$ th cell at time  $t$ ,  $\mathbf{F}_{ij}$  is the force exerted on cell  $i$  due to the interaction with cell  $j$ , and  $\mathbf{B}_i(t)$  is an independent Brownian Motion. All vectors are of order 2, and  $\sigma(t)$  is represented by a diagonal matrix.*

**Assumption 2** (Radial force). *Assume the force  $\mathbf{F}_{ij}$  between the  $i$ 'th and  $j$ 'th cells located at  $\mathbf{q}_i$  and  $\mathbf{q}_j$  respectively depends only on the distance between them:  $\mathbf{F}_{ij} = \mathbf{f}(\mathbf{q}_i - \mathbf{q}_j)$ . If we further assume that  $\mathbf{f}$  is a smooth function, then  $\exists$  a potential function  $u$ , s.t  $\mathbf{f} = -\nabla u$ .*

**Assumption 3** (Repulsion and attraction forces). *For any pair of cells one can write:  $\mathbf{f} = \mathbf{f}^r - \mathbf{f}^a$ . Where  $\mathbf{f}^r$  is the repulsion force and  $\mathbf{f}^a$  is the attraction force. These are characterised by their amplitudes  $R, A$  and their length scales  $r, a$  respectively. We further assume  $\mathbf{f}^r$  and  $\mathbf{f}^a$  are positive, monotonically decreasing functions.*

### 6.2.2 Dynamics of interactions between cells of different types

This further assumption is designed to capture the modulation of cell-cell interactions due to Eph-ephrin regulation, which come to play in interactions between the two different subpopulations (Eph and ephrin expressing cells). When an Eph- receptor on a cell membrane binds with an ephrin ligand on an adjacent cell, a signalling pathway within the cell acts to decrease the adhesion between the two cells. This decrease in adhesion can be modulated based on the levels of Eph and ephrin expression within the cell membrane. We introduce a parameter  $C \in [0, 1]$  and can regulate the strength of the Eph-ephrin interruption to the normal adhesion between the different cells by adjusting the value of  $C$ .

**Assumption 4** (Cell segregation due to Eph-ephrin interaction). *When Eph-ephrin interaction takes place, an additional repulsion term is introduced to attenuate the attraction due to adhesion. The attraction term is countered by  $\mathbf{f}^e = +C\mathbf{f}^a$ , where  $C$  is a constant which controls the attenuation level of the attraction force. Now  $\mathbf{f}(\mathbf{q}_i - \mathbf{q}_j) = \mathbf{f}^r(\mathbf{q}_i - \mathbf{q}_j) - (1 - c_{ij})\mathbf{f}^a(\mathbf{q}_i - \mathbf{q}_j)$ . with  $c_{ij} = c_{ji} = 0$  if the cells are of the same type (i.e. no Eph-ephrin interaction), or  $c_{ij} = c_{ji} = C$  if the  $i$ 'th and  $j$ 'th cells are of different types (i.e one is Eph and the other is an ephrin expressing cell).*

The biological analog to it is to assume that once an Eph containing cell meets an ephrin containing cell, the adhesion force between them will be diminished, maximally to zero (when  $C=1$ ). There will only be a repulsion force due to surface tension which does not allow cells to occupy the same space. This indeed gives rise to a form of ‘differential adhesion’, but unlike in other models, the possible change of the effective cell size is accounted for (i.e. two cells which have adhesion between them will normally be closer together than two cells who just had the adhesive bonds between them released due to the Eph-ephrin interaction).

Thus, the specific choice of the potential function is now:

$$u(\mathbf{d}) = rRe^{-|\mathbf{d}|/r} - aAe^{-|\mathbf{d}|/a} + c_{ij}aAe^{-|\mathbf{d}|/a} \quad (6.2.1)$$

where the first term is the potential for the repulsion force between cells and the second term is the potential for the attraction force, with  $R$  and  $A$  setting the magnitudes and  $r$  and  $a$  setting the length scales of the repulsion and attraction respectively. The last term gives rise to the Eph-ephrin regulation: when  $i$  and  $j$  are of the same type it equals zero, and when  $i$  and  $j$  are of different types  $c_{ij} = C$ , so it attenuates the adhesion force.

The list of functions representing the forces given in equation 3.1.5 is completed as follows:

$$\begin{aligned} \mathbf{f}^r(\mathbf{q}_i - \mathbf{q}_j) &= Re^{-d_{ij}/r} \left( \frac{q_i^x - q_j^x}{d_{ij}} \cdot \hat{x} + \frac{q_i^y - q_j^y}{d_{ij}} \cdot \hat{y} \right) \\ \mathbf{f}^a(\mathbf{q}_i - \mathbf{q}_j) &= -Ae^{-d_{ij}/a} \left( \frac{q_i^x - q_j^x}{d_{ij}} \cdot \hat{x} + \frac{q_i^y - q_j^y}{d_{ij}} \cdot \hat{y} \right) \\ \mathbf{f}^e(\mathbf{q}_i - \mathbf{q}_j) &= -c_{ij}\mathbf{f}^a(\mathbf{q}_i - \mathbf{q}_j) \\ &= c_{ij}Ae^{-d_{ij}/a} \left( \frac{q_i^x - q_j^x}{d_{ij}} \cdot \hat{x} + \frac{q_i^y - q_j^y}{d_{ij}} \cdot \hat{y} \right) \end{aligned} \quad (6.2.2)$$

In the computer simulation based on the numeric solution of these we further allow  $\mathbf{f}^r$  and  $\mathbf{f}^a$  to be different between the Eph and ephrin expressing cell populations by setting  $A = A^{eph}$  and  $R = R^{eph}$  to give  $\mathbf{f}^{eph}$ , or  $A = A^{ephrin}$  and  $R = R^{ephrin}$  to give  $\mathbf{f}^{ephrin}$  for the specific cell populations. This lets us explore a much larger set of clustering scenarios.<sup>21</sup>

### 6.2.3 Birth process

In this biological system cell death is insignificant at the relevant time scales, thus we let each population size follow a pure birth process: The initial number of cells is taken to match the experimental initial conditions, and after this, each cell can ‘give birth’ to a daughter cell, following a Poisson process of rate  $\lambda$ . The daughter cell will always be of the same type (i.e. eph or ephrin expressing) as the mother, and its initial position will

<sup>21</sup>In the case where the forces between same type cells are different between the two populations (i.e  $A^{eph} \neq A^{ephrin}$  or  $R^{eph} \neq R^{ephrin}$ ), we set two other parameters:  $A^{eph:ephrin}$  and  $R^{eph:ephrin}$  (meaning the attraction and repulsion amplitudes between not-like cells respectively) to avoid ambiguity.

always be the location of the mother at the time it was split. Indeed this is a specific case of the Galton-Watson branching process, where the number of offsprings is  $k = 2$  with probability  $= 1$  (see further details in section 2.2.4 and references therein).

**Assumption 5** (cell proliferation). *Assume each population size follows a pure birth process. At a given time  $t > t_0$ , each existing cell can give birth to a new cell of the same type with probability  $p = e^{-\lambda(t-t_0)}$ , where  $t_0$  is the last time at which that cell divided (or was first created). The initial position of the new cell is taken to be the location of the mother cell at the time it was split.*

At any given time  $t$ , the total number of cells  $N(t) = N^{ephin}(t) + N^{eph}(t)$  is a random variable. The expected value of the total number of cells at time  $t$  is:

$$\mathbb{E}N(t) = N(0)e^{\lambda t} \quad (6.2.3)$$

Since the cell population size doubles every 16 hours, we set:  $\lambda = \frac{\log 2}{16h}$ .

### 6.2.4 Model equations

This generalise version of our model gives rise to a form of a Branching Brownian Motion (BBM) with interactions. Indeed, as discussed in section 2.2.4, BBM is a well defined process which can be explored in terms of its weak limit or dual process. However, here we have an additional, fairly un-trivial, drift term which dictates interactions between the particles. Furthermore, the model accounts for two different types of particles which have complex interactions with like particles, and even more complex interactions with un-like particles.

With the biological question which motivated us to formulate this generalised model in mind, we choose not to go down the path of rigorous formulation and analysis. Instead, we find that the simulations of this model's predictions by generating numeric solutions of its equations (using a specially designed and executed computer code) is appropriate, and gives adequate and comprehensible answers.

Consolidate assumptions 1-5 above to formulate our model's equations:  
For each population type, at a given time  $t > t_0$ , each existing cell can give birth to a new cell of the same type with probability  $p = e^{-\lambda(t-t_0)}$ , where  $t_0$  is the last time at which that cell divided (or was first created). The location of the new cell is taken to be the location of the mother cell at the time it was split.

At any given time  $t$ , the total number of cells is

$$N(t) = N^{ephin}(t) + N^{eph}(t) \quad (6.2.4)$$

where  $N^{ephin}(t)$  and  $N^{eph}(t)$  both independently follow a pure birth process with rate  $\lambda$ .

$\mathbf{q}(t)$ , the vector holding all positions of all cell centres at time  $t$  is of length  $2N(t)$ , and is given by:

$$\begin{aligned} d\mathbf{q}(t) &= -\nabla U(\mathbf{q}(t))dt + \sigma(t)d\mathbf{B}(t) \\ \text{where } U &= \frac{1}{2} \sum_{i=1}^{N(t)} \sum_{j \neq i}^{N(t)} u_{ij} = \sum_{i=1}^{N(t)-1} \sum_{j=i+1}^{N(t)} u_{ij} \\ \text{with } u_{ij} &= rRe^{-\frac{d}{r}} - (1 - c_{ij})aAe^{-\frac{d}{a}} \\ \text{and } d_{ij} &= \sqrt{(q_i^x - q_j^x)^2 + (q_i^y - q_j^y)^2} \equiv d \end{aligned} \quad (6.2.5)$$

Note that  $\nabla_i^x u_{ij} = \left( -Re^{-\frac{d_{ij}}{r}} + (1 - c_{ij})Ae^{-\frac{d_{ij}}{a}} \right) \left( \frac{q_i^x - q_j^x}{d_{ij}} \right)$ , and the explicit representation of the SDE for the motion of the  $i$ 'th cell in the  $x$  direction gives:

$$\begin{aligned} dq_i^x &= -\nabla_i^x U dt + \sigma dB_x^i(t) \\ dq_i^x &= \sum_{j \neq i}^{N(t)} \left( Re^{-\frac{d_{ij}}{r}} - (1 - c_{ij})Ae^{-\frac{d_{ij}}{a}} \right) \left( \frac{q_i^x - q_j^x}{d_{ij}} \right) dt + \sigma(t) dB_x^i(t) \end{aligned} \quad (6.2.6)$$

and similarly for the  $y$  coordinate.

Our simulation follows a discretised version of these equations, where in particular, a cell that was born between times  $t_i$  and  $t_{i+1}$  will take the position of its ancestor at time  $t_i$ . More details regarding the simulation can be seen in appendix C.2.1.

## 6.3 Comparison of simulated and experimental cell-cell segregation

### 6.3.1 Dynamics of cluster formation

We commenced the evaluation of this extension to our mathematical model, initially by comparing simulated and experimental cell-cell segregation in co-cultures of two cell types: either expressing EphB2, together with a green fluorescent protein, GFP, for ease of visualisation (Poliakov et al., 2008; Janes et al., 2011), or expressing ephrin-B1, which is a EphB2 ligand. As previously reported (Nievergall et al., 2010; Janes et al., 2011), using time-lapse fluorescence microscopy to record the continuous changes in (cell) cluster size distribution, it was clear that within 49 hours, an initially mixed cell population consisting of a 1:3 ratio of GFP/EphB2 expressing (EphB2+) and ephrin-B1-expressing cells (ephrinB1+), gives rise to defined clusters of GFP-fluorescent EphB2-expressing cells that are surrounded by non-stained ephrin-B1 cells (Figure 6.1A). We simulated this experiment, using empirically-derived parameter values for the mean square cell displacement (MSD, see relation between MSD and  $\sigma$  in chapter 4, footnote 6), cell proliferation rate and an initial Eph:ephrin cell ratio of 1:3. Free parameter values were set to:  $A_{eph} = A_{ephrin} = 100$ ,  $R_{eph} = R_{ephrin} = 250$ ,  $a = 7.5$ , and  $r = 5.8$  so that both cell populations would display equal inter-cellular adhesion and surface tension forces (see chapter 3 for more details). Numerical solutions of our model equations, generating cell positions and graphic representations at defined time points, revealed that the relative distributions of (cell) cluster sizes were very similar in the simulated and experimental images through most of the time course of the experiment (Figure 6.1B).

A quantitative comparison between simulated and experimental cell-cell segregation is presented in figure 6.2, where images from 6 laboratory experiments and a corresponding set of 6 independent numeric realisations of the model were analysed. Direct comparison of experimental and simulated clusters suggested very similar size distributions. For ease of comparison though, (green) cell clusters were sorted by their areas into three groups: small ( $100 - 1500 \mu m^2$ ), medium ( $1500 - 15000 \mu m^2$ ) and large ( $> 15000 \mu m^2$ ) cell clusters. The group boundaries were chosen, as described previously (Nievergall et al., 2010), to reflect closely the actual cell patterning behaviour: Small clusters correspond to the average size of only 1-2 cells, and the medium bin captures transition to larger groups of cells representing cluster formation. The first two (small- and medium-size cluster) bins thus captured the majority of all observed clusters, and the less frequent, very large clusters that formed mainly at the end of the simulation were captured in the third bin.

Statistical analysis (multivariate-ANOVA) of cluster sizes suggested a good fit between experimental and simulated results considering frequencies of small and medium-sized clusters formed up to the 33 hour time point<sup>22</sup> ( $p = 0.5317$ ), see further discussion regarding statistical analysis methods we applied to these results in appendix C.2.4.

By the final time point (49h) however, significant differences ( $p = 0.0028$ ) between the simulated and experimental data emerged: The increased number of small ( $100 - 1500 \mu m^2$ ) clusters in the experiment as compared to the simulation is due to some single cells remaining separated from the larger clusters, as increasing cell density prevents their effective sorting so they therefore remain in small clusters. This apparent variability in the cell population (i.e. the occurrence of slow cells) may be caused by variability in Eph/ephrin surface expression levels between individual cells, which is not present in the simulated cell population in which all cells are identical.

Similarly, the significant number of large simulated clusters ( $> 15000 \mu m^2$ ) at the 49h time point was not reflected by the experiment. This is likely due to the very tight compaction of EphB2-positive cells which reduced the overall footprint of clusters and of individual cells at the end of the experiment, as was revealed in a detailed analysis of the corresponding microscopic images (Figure 6.1A, further analysis not included). By contrast, the simulation assumes equal adhesion and repulsion forces within both cell populations throughout the experiment, so that simulated clusters appear larger than the experimentally observed clusters. This suggests that during Eph/ephrin-regulated cell sorting, clustering of the EphB2 cell population may be modulated according to the expression level and kinase activity of the Eph receptors (Pasquale, 2010; Nievergall et al., 2011).

### 6.3.2 Modulation of the Eph-ephrin signalling strength effects cluster formation

We assessed the relationship between the Eph-ephrin signal strength and cluster size distribution on the basis of previously-published co-culture experiments, where the effects of modulating expression level and kinase signalling activity of Eph receptors on cell segregation and cluster formation had been analysed (Janes et al., 2011). That study revealed that changes in the composition of Eph signalling complexes, by additional exogenous expression and recruitment of kinase-active or kinase-inactive receptors, can either increase cell contraction and clustering or promote spreading and intermingling, respectively (Janes et al., 2011). To accommodate the complexity of Eph cell signalling in the model we modulated the attenuation constant  $C$  (see assumption 4 and equation 6.2.1), and to simulate the uneven repulsion and/or attraction forces within the two different populations we let  $R_{eph} \neq R_{ephrin}$  and  $A_{eph} \neq A_{ephrin}$  respectively.

In the experiments, over-expression of the inactive EphB2 mutant lacking the intracellular domain (Figure 6.3A, left column) completely blocked cell-cell segregation and cluster formation, compared to Wt EphB2 expressing cells (Figure 6.3A, third column), when co-cultured with cells expressing ephrinB1. In comparison, co-cultures of Wt EphB2-expressing cells with parental HEK293 cells, which express low but detectable levels of ephrin-Bs, resulted in an intermediate phenotype, revealing some segregation between the

<sup>22</sup>Note we do not use multivariate-ANOVA here in its conventional use, as we are seeking for similarities between the samples rather than a significant difference. We interpret the  $p$ -value obtained here merely as an indication of how close the two samples are.

The analysis was done considering only the first bin (very small clusters) at times 16 and 33 hours: The third bin assumes zero values for most of the data points, and therefore was excluded. The frequencies from the first and second bins always add up to one, and therefore data included in the first bin contains all the information. The first time frame ( $t=0$ ) was also left out, as it represents the initial conditions which were designed to have similar frequencies by definition. Considering data from times 16 and 33 hours gave a  $p$ -value of  $p=0.5317$ .

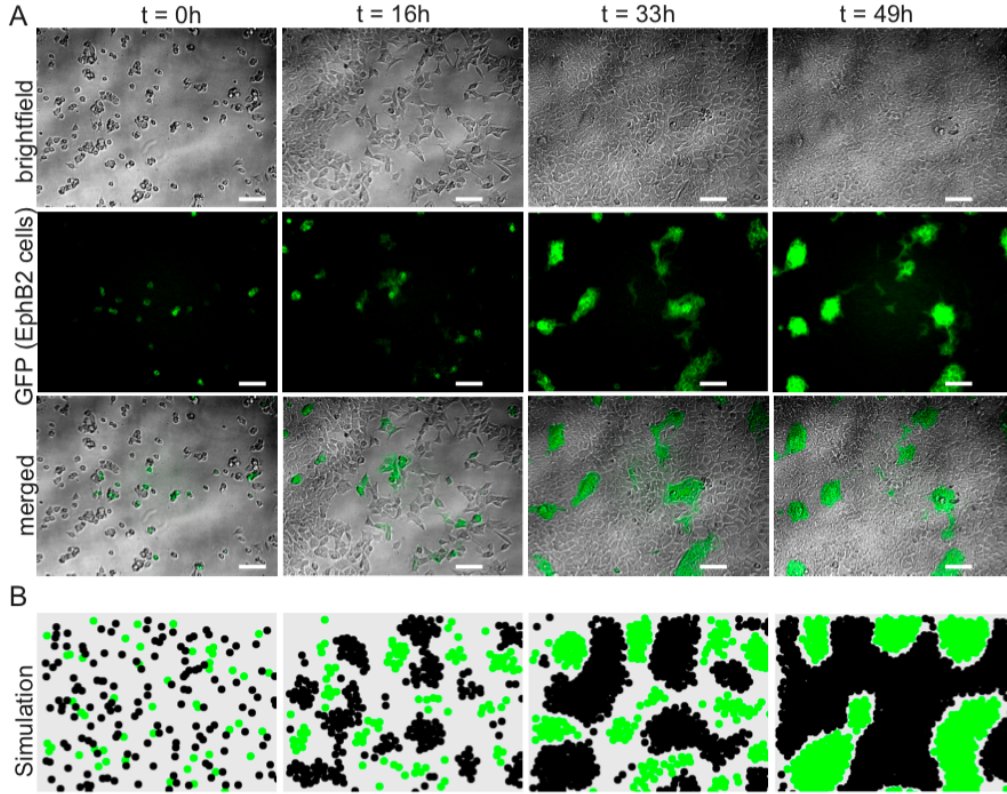


Figure 6.1: Time-lapse imaging and simulation of Eph/ephrin-driven cell-cell segregation (Aharon et al., 2014). A) Representative time-lapse microscopic images (taken every 20 minutes for indicated times) from co-cultured EphB2/GFP (green) and ephrin-B1 (unstained) expressing HEK293T cells. Bright-field micrographs (top panels), green-fluorescent images (middle panels) and merged images (bottom panels) are shown, scale bars,  $100\mu m$ ). B) Simulation of the same experiment; the corresponding time points are shown, ephrin-B1 and EphB2/GFP-expressing cells are represented as black and green circles, respectively.

two cells types, but lacking apparent clusters or clear cluster boundaries (Figure 6.3A, 2nd column). Modulation of the attenuation constant of the attraction force  $C$  allowed us to simulate these conditions, so that increasing the values of  $C$  from 0 to 1 effectively introduced increasing de-adhesion/segregation between the simulated Eph and ephrin cell populations (Figure 6.3B,C).

Further, an extreme case of clustering was examined, where over-expression of a second signalling-competent Eph receptor, EphA3, increased the cellular response in EphB2-expressing cells to the interaction with ephrin-B1 cells and resulted in enhanced cell segregation (Janes et al., 2011). In this experimental setting, the clusters of Eph-expressing cells appear brighter, as the cells condense and become more tightly packed than the corresponding ephrin cells, thereby effectively decreasing the footprint of individual clusters (Figure 6.3A, right-most panel). In this case, even a maximal attenuation constant of  $C = 1$  was inadequate for simulating the corresponding increase in the packing of individual EphB2 cells within clusters (Figure 6.3B). Therefore we considered that modelling of the increased cell compaction within the EphB2/EphA3 clusters might require lowering the repulsion term, or alternatively increasing the attraction force within the Eph-expressing cell population (see Figure 6.4 below), to effectively increase cell packing beyond that seen in the ephrin-expressing cells. By reducing the repulsion strength, which simulates cell

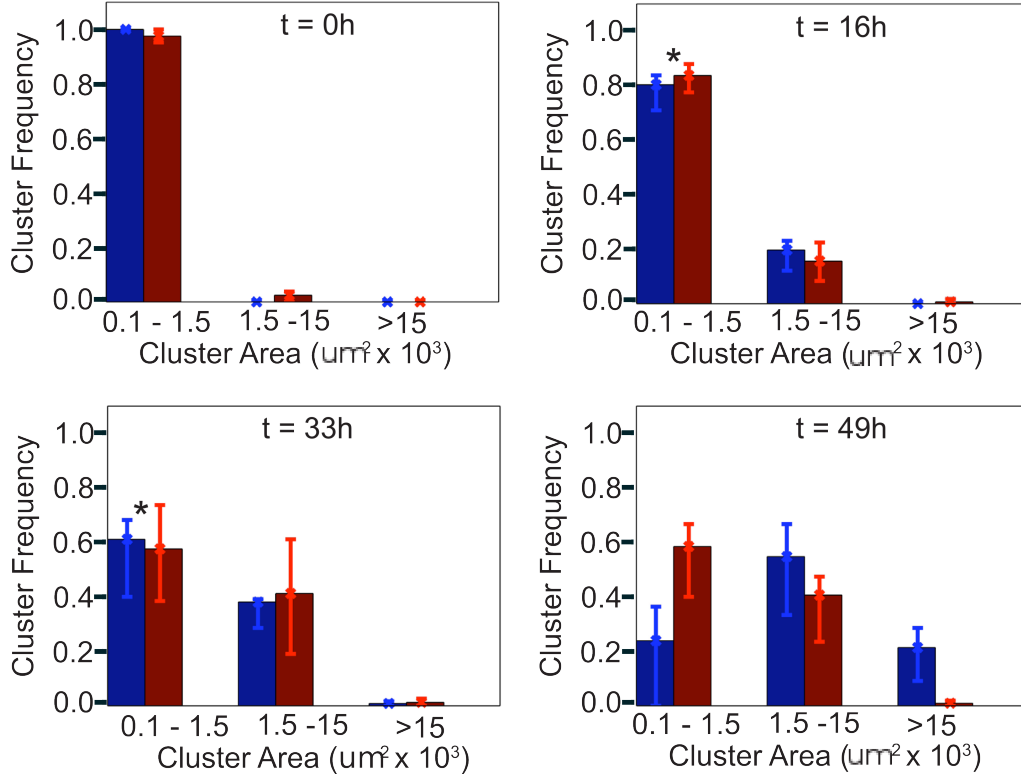


Figure 6.2: Distribution of cell cluster sizes during segregation (Aharon et al., 2014). Clusters of EphB2/GFP expressing cells from the time course experiment in Figure 6.1 were grouped according to their footprint into three groups as indicated; bar graphs illustrate frequency of clusters in each bin at indicated time points, (red bars), experimental microscopy data; (blue bars), simulated data. Error bars represent the upper and lower bounds, below which 75% and 25% of the data points are included. \*Multivariate-ANOVA,  $p > 0.5$ .

surface tension of EphB2 cells to  $R_{eph} = 220 \neq R_{ephrin} = 250$  (see equation 6.2.1), and thereby decreasing the effective radius of Eph expressing cells relative to that of ephrin expressing cells, the simulation now yielded a pattern of tightly-clustered Eph cells in a sea of ephrin-expressing cells that closely matched the experimentally-derived pattern (Figure 6.3B, right-most panel). Note, this decrease in the effective radius of Eph cells makes them pack more closely together, and therefore leaves more space for the ephrin cells, evident by the space left in the simulation even after 48 hours, whereas in the laboratory experiment, the ephrin cells expand to occupy the free space.

Quantitation and statistical (ANOVA) analysis of the experimental and simulated cell densities in the Eph cell clusters (Figure 6.3D), suggested a close fit between experimental and simulated data ( $p = 0.506$ ), whereby in particular, adjustment of the cell density by decreasing the cell surface tension of the EphB2/EphA3-expressing cells resulted in simulated cell densities which closely matched the experimental data. Our observations suggest that the cell morphology and cell-cell adhesion within these clustered Eph-expressing cell populations are changed compared to the surrounding ephrin-expressing cells, in agreement with a range of reports on mechanisms of Eph/ephrin-facilitated tissue patterning and boundary formation (Batlle and Wilkinson, 2012), possibly due to Eph receptor crosstalk with cell-cell adhesion molecules (Solanas et al., 2011).

### 6.3.3 The formation of tightly-packed islands-in-the-sea Eph cell clusters relies on enhanced cell-cell adhesion within clusters

We went on to examine crosstalk between Eph signalling and cell-cell adhesion proteins, by simulating the effect of altering the relative adhesion within one of the two cell populations. As a starting point we considered two equally sized populations of Eph and ephrin expressing cells with identical adhesion forces between the two distinct cell types ( $A_{eph} = A_{ephrin}$ ). While in this setting of the simulation, the two cell populations clearly segregated from each other, a preference for clustering of one cell type over the other was not apparent and the segregation did not yield the typical Islands-in-a-sea pattern observed in other experiments (Figure 6.4A, left panel). As expected, increasing the cell-cell adhesion term in one of the simulated cell populations over the other ( $A_{eph} \neq A_{ephrin}$ ), resulted in tighter clusters of this population and led to islands within the sea of cells of the other type (Fig 6.4A, right panel). We tested this prediction experimentally: cell-cell adhesion was increased within the EphB2-expressing cell population by overexpressing the cell-cell adhesion protein E-cadherin<sup>23</sup>. As illustrated in Figure 6.4B,C, overexpression of E-cadherin indeed resulted in the islands in a sea pattern predicted from the simulation, due to the formation of distinct EphB2 cell clusters with significantly enhanced cell density as compared to the parental cells which lack E-cadherin. Western blotting confirmed the expression of the m-cherry E-cadherin in the transduced cells, and that the transduced and control cells maintained similar levels of EphB2 (Figure 6.4D).

These results indicate that our mathematical model, based on the concept that the sum of various Eph and ephrin-derived signal inputs is translated into cellular responses which can be represented by adjustable terms for cell-cell adhesion and repulsion, provides the robustness and flexibility to allow realistic simulations of Eph/ephrin-regulated cell sorting under a range of experimental conditions.

---

<sup>23</sup>This was achieved by lentiviral transduction of mCherry-tagged E-cadherin (Solanas et al., 2011) into the EphB2-expressing HEK293 cells, which lack endogenous E-cadherin, but in agreement with their neural origin (Shaw et al., 2002) express N-cadherin.



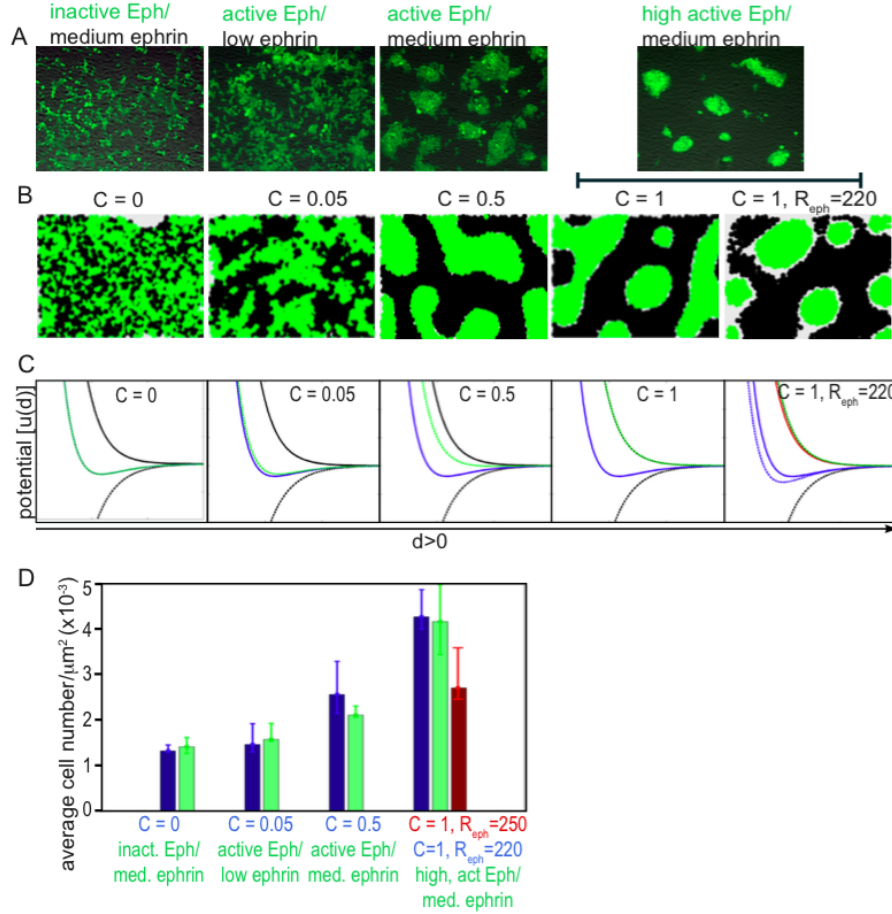


Figure 6.3: Eph expression level and signalling capacity regulate cell segregation: comparison of experimental versus modelling outcomes (Aharon et al., 2014). A) Increased Eph signalling capacity results in enhanced cell-cell segregation (Janes et al., 2011): HEK293 cells expressing cytoplasmic deleted (signalling-inactive), wild type EphB2 or co-expressing EphB2 and EphA3 were co-cultured with cells expressing low or high levels of ephrin-B1, as indicated. EphB2 cells were stained with Cell-Tracker green for ease of visualisation, images taken after 48 h co-culture when cell-cell segregation was regarded as complete. B) Simulation of the same experimental conditions, using parameter values of:  $A_{eph} = A_{ephrin} = 100$ ,  $R_{eph} = R_{ephrin} = 250$ ,  $a = 7.5$ ,  $r = 5.8$  and  $\sigma^2 = 10$ , apart from the right-most panel, where  $R_{eph} = 220$ . The initial ratio of Eph: ephrin cells in all cases is 1:1. C) Functions of the force potential,  $u(d)$  between two cells at distance  $d > 0$ , for the simulations illustrated in B. Unbroken black line,  $u_r$ , is the potential of the repulsion force at  $R_{eph} = 250$ ; Unbroken red line,  $u'_r$ , is the potential of the repulsion force at  $R_{eph} = 220$ ; Broken black line,  $u_a$ , is the potential of the attraction force; Unbroken blue line,  $u_r - u_a$ , is the potential of the total force between same type cells  $R_{eph} = 250$ ; Broken blue line,  $u'_r - u_a$ , is the potential of the total force between same type cells when  $R_{eph} = 220$ ; Broken green line,  $u_r - (1 - C)u_a$ , is the potential of the total force between cells of different types; D) Statistical analysis of cluster characteristics from a minimum of 5 independent data sets: Comparison of the cellular densities that were observed experimentally by microscopy (green bars), or by simulation under the conditions detailed in panels C (blue bars; except for  $C = 1, R_{eph} = 250$ , red bar). For microscopic images the number of cells in a cluster was estimated from the total fluorescence intensity of the cluster, divided by the average fluorescence intensity of a single cell as detailed previously (Janes et al., 2011). Error bars represent the upper and lower bounds, below which 75% and 25% of the data points are included.

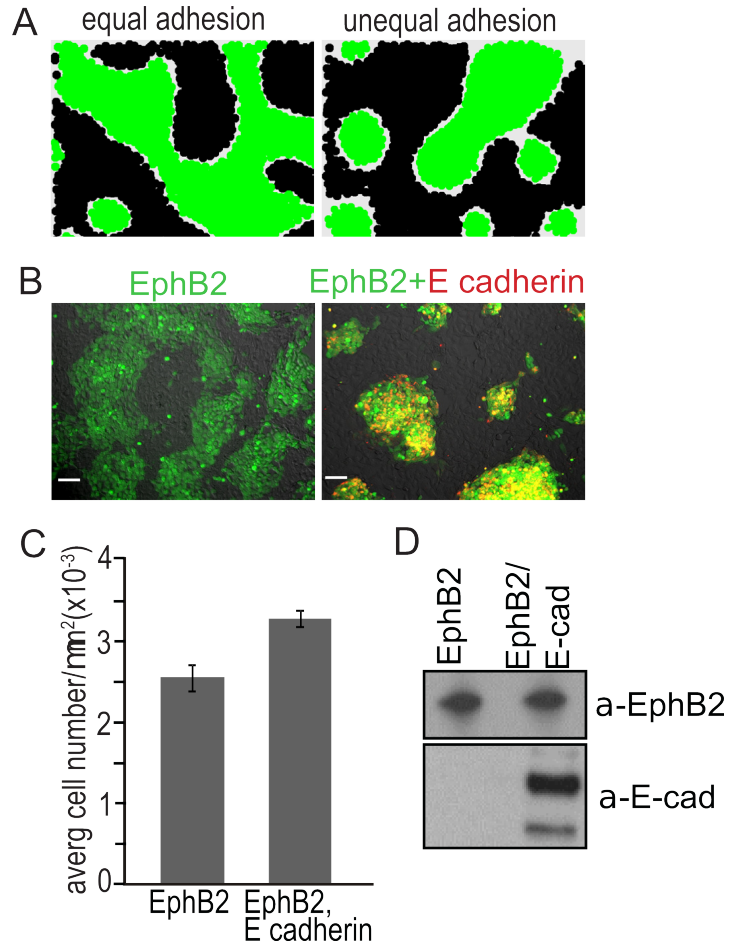


Figure 6.4: Increased cell-cell adhesion within one cell population is required for the formation of tightly packed cell clusters (Aharon et al., 2014). A) Simulation of cell-cell segregation using the same adhesion term in both cell types ( $A_{eph} = A_{ephrin} = 100$ , left panel) vs. increased adhesion only in the green cell population ( $A_{eph} = 110$ ,  $A_{ephrin} = 100$ , right panel). Both simulations started with the same number of (green) and ephrin (black) expressing cells. In the Equal adhesion case, an Islands-in-a-sea pattern is less apparent. B) Representative images from segregation assays of unlabelled ephrin-B1 cells co-cultured with Cell Tracker- green labelled EphB2 cells, without (left) or with E-cadherin-cherry expression (right). Scale bar 75 micron. C) Quantitation of EphB2 cell cluster density by image analysis of the experiment in B (n=10). D) Western blot analysis of lysates from parental and E-cadherin-cherry-transduced cells, using the indicated antibodies.

## 6.4 Discussion

Eph-ephrin interactions between neighbouring cells control the dynamic changes in cell morphology, adhesiveness and motility that underlie tissue patterning and boundary formation. As the number of cell-biochemical pathways implicated in these processes grows, the difficulty in predicting outcomes from knowledge of individual signalling pathways increases (Jorgensen et al., 2009). It is clear that prediction of *in vivo* and *in vitro* cellular behaviour during patterning can be assessed by mathematical modelling and several approaches have been used to aid our understanding of molecular and cellular concepts underlying cell positioning and tissue patterning (Salaita et al., 2010; Reber et al., 2004; Batlle and Wilkinson, 2012).

In this chapter we described a mathematical approach to simulate modes of Eph/ephrin-controlled 2-dimensional cell-cell segregation and aggregation. We developed a stochastic model based on the Lagrangian approach, which incorporates empirical parameters for cell movement and cell proliferation and theoretical knowledge regarding cell adhesion and surface tension. It allows both a realistic simulation of Eph/ephrin-directed segregation between two cell populations and the modelling of responses to modulation of adhesive and repulsive signalling mechanisms. Direct comparison of experimental and computer-simulated formation of cell patterns shows that our model recapitulates the spatial and temporal dynamics of the cell-cell segregation processes.

The cellular Potts model, a computational model based on the concept that differences in binding energies between two interacting cell populations direct the separation between non-like cells (the differential adhesion hypothesis), has been adopted to simulate EphB/ephrin-B controlled cell sorting along the intestinal crypt epithelium (Wong et al., 2010). An alternative approach, including a term for cell density-dependent cell motility (Taylor et al., 2011), was used to model segregation in 2-D tissue culture between two cell populations that differ in their expression of the cell-cell adhesion protein E-cadherin. However, in the case of Eph/ephrin regulated cell-cell interactions, the kinetics of the simulated sorting lagged behind the tissue culture experiment by an order-of-magnitude (Taylor et al., 2012), suggesting that additional terms may be required to model sorting between Eph- and ephrin-expressing cells.

Considering the characteristic dichotomous functions of Eph receptors, eliciting either cell-cell repulsion and segregation or cell-cell adhesion (Pasquale, 2010; Nievergall et al., 2012), we developed a model for the behaviour of two intermingled cell populations that display differing adhesive/repulsive forces within and between each population. We took a novel approach of complementing terms of surface tension (representing cell-cell repulsion) and attraction (representing cell-cell adhesion) that can be adjusted according to the Eph/ephrin signal strength. We are aware that on a molecular level Eph/ephrin-facilitated cellular patterning relies on a fine-tuned interplay of a large number of known and potentially an even larger number of unknown signalling pathways (Jorgensen et al., 2009), however, we propose that collapsing this complexity to a selected, small number of measurable mathematical terms, which can approximate the sum of all contributions, provides a model that realistically reflects the overall processes. Thus, we introduce a Brownian motion term to represent the movement of a single cell that is not influenced by any neighbours and bundle all possible interactions between two cells into two radial forces: a repulsion term that represents the sum of cell surface tension and spatial exclusion forces that prevent cells to occupy the same space, as well as an attraction term which represent all adhesion forces, including those governed by cadherin interactions. For any cell type, the size of parameters controlling the Brownian motion and the strength and length scale of both radial forces can be estimated directly from the paths of cells moving independently and at a density where interactions with neighbouring cells occur (chapter 4). In this model extension, the Eph-ephrin regulated interaction can act to cancel part or

all of the adhesion forces between Eph and ephrin expressing cells. The characteristics of these forces impose a stronger (spatial) separation between the two different types of cells. This heterologous interaction promotes the retraction of Eph- and ephrin- expressing cells from each other, as well as diminishing the adhesion between them.

Computer simulations, based on numerical solutions of our model, allowed the prediction of the dynamics of cell segregation and cluster formation. A comparison to experimental data demonstrated our models capacity to reproduce the spatial and temporal dynamics of cluster formation seen in the tissue culture experiments. In contrast to previous models of Eph/ephrin-regulated cell segregation, our model adjusts for the proliferative increase in number of cells in the two different cell populations throughout the experiment. As cell density increases, cells in our model are inherently slowed down due to interactions with increasing numbers of neighbouring cells. We find that cell proliferation is an important contributor to the dynamics of cluster formation.

We demonstrate the capacity of our model to accommodate variation in Eph expression and strength of signalling function, by modulating the amplitude of those model parameters that control the strength of the Eph-ephrin interaction (Figure 6.3). A comparison between our model-based predictions and experimental cell behaviour revealed the necessity to include asymmetry in the adhesive behaviour of the Eph and ephrin-expressing cell population in order to arrive at the typical cellular pattern, where one population of tightly-packed cells with increased adhesion (or accordingly-reduced repulsion) between like cells is surrounded by a sea of less-tightly packed cells of the other type (Figure 6.4). Specifically, our model highlighted that ephrin/Eph-driven cell retraction/repulsion at the interface between the two cell populations is insufficient to generate this typical Island-in-the-sea pattern and pointed to the critical role of additional cell-cell adhesion within the Eph-expressing population that is required to generate the tighter cell packing in this population. Interestingly, this same conclusion was previously drawn from experiments addressing EphA4/ephrin-B2-driven rhombomere formation in Zebrafish, where EphA4-expressing cells show increased adhesion to each other (Cooke et al., 2005). Since the cell-cell adhesion protein E-cadherin is known for its co-ordinated and co-localised expression with Ephs at sites of cell-cell contacts (Orsulic and Kemler, 2000) and is implicated in Eph/ephrin-facilitated boundary formation (Solanas et al., 2011), we over-expressed mCherry-E-cadherin in EphB2 cells. Our experiments confirmed a role for unequal adhesion in generating the asymmetry between the enclosed (i.e., the island) and the enclosing (i.e., the sea) cell populations. Eph/ephrin and E-cadherin interactions control the partitioning of two cell populations, by co-ordinately controlling the cell segregation between, and the adhesion within, each population.

In summary, we conclude that a simple model based on terms defining differential cell-cell adhesion, in which cell-cell adhesion forces are balanced to varying degrees by Eph-ephrin regulated cell-cell de-adhesion, is sufficient to describe the cell patterning observed between Eph and ephrin expressing cell populations migrating on a 2-dimensional surface over time.

# Chapter 7

## Summary

Mathematical models for cell motility range over a large scale of detail, from attempts to account for the biochemical reactions within a cell up to models for migration of a large cell population. When adding mutual interactions, complexity level increases even further, and agent based models employing random walk inclusion processes are commonly used. Computer simulations allow to easily incorporate a wide range of possible cell-cell interactions, and continuous approximations obtained by averaging were shown to capture simple interactions within a single population type (Deroulers et al., 2009; Fernando et al., 2010). More realistic interactions, however, can not be averaged to an analytic (continuous) model which still captures significant behaviours of the biological system (Simpson et al., 2010). Furthermore, models for interactions between two population types almost always resort to highly detailed computerised models, making them practically inaccessible to mathematical analysis (Glazier and Graner, 1993; Wong et al., 2010).

Here we constructed a mathematical model for cell migration and mutual interactions, aimed to describe the significant features of a highly complex biological system and thus help interpret experimental results (chapter 3). All internal processes that take place within an individual cell controlling its independent motion were reduced down to a single term made of an independent Brownian Motion function multiplied by a (possibly) time dependent function. All processes controlling any possible interaction between neighbouring cells were reduced down to two smooth (anywhere but at  $r = 0$ ), monotonic deterministic functions representing radial forces (which therefore have a gradient that represents the potential energy, and depend only on the distance between cell centres). The first bundles all attraction forces such as cell adhesion, and the second counts for all repulsion forces such as surface tension. Identifying these governing mechanisms resulted in a simple yet powerful mathematical model, which is flexible and versatile enough to accommodate a wide range of system behaviours. This allowed us to apply analytical methods to find its steady state solution and to develop a method for estimating model parameter values from experimental data. Computer simulations based on numeric solutions of our model equations further allowed direct comparison of model predictions to experimental results.

Girsanov Theorem was used to derive likelihood ratios for different parameter values given cell paths (which can be obtained either experimentally or by using computer simulation), and then estimated parameter values that maximise that likelihood. A computer program was written specifically to apply this algorithm and to generate error estimations. Studying the difference between estimated values given simulated cell paths and the known values that were used to generate those paths, for a few different choices of potential functions, revealed that decreasing sample rate and increasing integration time minimises the error. For the specific potential chosen to be used in our model (exponential attraction and repulsion forces), it was further found that the accuracy of the estimations obtained depend also on the distance between the cells observed: when looking at cells that

are too far apart so the potential between them diminishes, the method fails to estimate parameter sizes correctly.

The first application of our model was to study the effect of radiation treatment on cell migration and mutual interactions. Our method was applied to estimate model parameters from experimental data obtained by observing cells after radiation treatment. It unveiled a clear trend of an increase in individual cell motility and a decrease in mutual interactions as a result of an increase of radiation dosage from 0 to 10 Gy. This type of quantitation, which was not accessible to biologists before, points to the need of further experimental work in this field and provides a new tool for the analysis of experimental results.

The model was extended to account for cell proliferation and for interactions between two different cell types, and a second computer program was written based on its numeric solutions (chapter 6). In its application to a biological system concerning cell segregation and cluster formation regulated by Eph-ephrin interaction it was capable to replicate both the dynamics and the possible range of cell-cell segregation. It then generated further predictions highlighting the role of inner population (cadherin regulated) cell-cell adhesion in determining the outcomes of cell segregation. These were verified in the laboratory by our collaborators and backed up by further (biological) research.

We are aware that when attempting to answer biological questions using mathematical tools one should take extra care: On one hand, when translating the biological system into mathematical equations it is inevitable that some (if not a lot) of the intricate detail in the making of the biological system will be left out. It is important to conduct a careful assessment of the contributions of the different components in the system and identify the dominant mechanisms controlling its behaviour. On the other hand, in attempt to apply mathematical theories which are often designed to handle theoretical concepts (such as Brownian Motion or infinite domains) to real life questions (such as concerning particles which have a mass and a volume on a bounded domain), one will often have to ease some conditions or add some constraints. This is a common issue for all attempts in mathematical biology, (or any other mathematical application to that matter), and for this reason an ongoing examination of the mathematical model and its predictions against experimental results when possible adds valuable credibility.

We thus conclude that the model we develop and present here addresses a complex system that is usually hard to handle using analytical tools. Our minimalistic yet inclusive approach resulted in a flexible and powerful model which can address different experimental systems and accommodate for different experimental conditions. Its application to the investigation of radiation treatment highlighted trends which were not well noted before. In the Eph-ephrin regulation front, it added comprehensive understanding of the governing processes, along with being, to our knowledge, the first computational model which was found capable of replicating both the dynamics and range of possible segregation outcomes.

## Future work

One can deepen model analysis, e.g by exploring the existence of a steady state in the case of two cell populations. Then new mathematical predictions can be translated to biological terms and enhance the understanding of the biological system.

Work can be done on our JAVA program to improve efficiency and accuracy of our MLE algorithm. Error analysis can be extended further, e.g.: study what kind of cell path give the best estimations of model parameters (as was done by Bowden et al. (2013) using an agent based model), formalise the analysis of goodness of fit for the actual choice for the drift function, etc.

For the model applications presented here, adjustments to the model can be added (if required) as experimental work continues. More particularly, in the study of radiation treatment one can apply the methods developed in chapter 4 to data obtained from more experimental repetitions and a variety of cell types in order to formulate a stronger statement regarding the trends we saw in the response of cells to radiation treatment. When experimental data regarding cells' response to microbeam radiation treatment will become available, our model could be adjusted to help understand these new experimental results.

Further, our mathematical model and MLE algorithm can be applied to a variety of other 'real world' systems, where the response of a group of 'socially interacting', independently mobile agents to changes in their environment is studied, e.g molecular dynamics, wild life behaviour, etc.





# Appendix A

## Proofs

### A.1 Proof for lemma 4.2.2

*Proof.* Note this change of drift corresponds to taking  $H = \sigma^{-1}(\mu_\alpha(X(t), t) - \mu_{\alpha_0}(X(t), t))$  in Girsanov theorem above (theorem 4.2.1). Since  $\alpha_0$  is such that  $\mu_{\alpha_0} = 0$ , and due to changed notation  $\mu = \mu_\alpha$ , one has  $Z = -\int_0^t \sigma^{-1}\mu(X(s), s)dB(s)$ , and  $dB(t) = \sigma^{-1}(dX(t) - \mu(X(t), t)dt)$ . Substitute in Girsanov theorem and get:

$$\begin{aligned} \log \Lambda &= Z - \frac{1}{2} \langle Z, Z \rangle \\ &= -\int_0^t \sigma^{-1}\mu(X(s), s)\sigma^{-1}(dX(t) - \mu(X(t), t)dt) - \frac{1}{2} \langle Z, Z \rangle \\ &= -(\sigma^{-1})^2 \int_0^t \mu(X(s), s)dX(s) + \frac{(\sigma^{-1})^2}{2} \int_0^t \mu^2(X(s), s)ds \end{aligned}$$

because  $d\langle Z, Z \rangle = d\langle \sigma^{-1}\mu(X(s), s)B(s), \sigma^{-1}\mu(X(s), s)B(s) \rangle = (\sigma^{-1}\mu_\alpha(X(s), s))^2 ds$ .

Now one may use Itô formula to re-write the first (Itô) integral above. Recall Itô formula:

$$\int_0^t f'(X(s))dX(s) = f(X(t)) - \frac{1}{2} \int_0^t f''(X(s))d\langle X, X \rangle(s)$$

and since we chose our  $\mu$  so there is a  $U$  such that:  $\mu(x) = -\frac{dU}{dx}$ , it gives:

$$\begin{aligned} -\int_0^t \mu(X(s), s)dX(s) &= U(X(t)) - \frac{1}{2} \int_0^t U''(X(s))d\langle X, X \rangle(s) \\ &= U(X(t)) - \frac{\sigma^2}{2} \int_0^t U''(X(s))ds \end{aligned}$$

so:

$$\log \Lambda = (\sigma^{-1})^2 U(X(t)) - \frac{1}{2} \int_0^t U''(X(s))ds + \frac{(\sigma^{-1})^2}{2} \int_0^t \mu^2(X(s), s)ds \quad (\text{A.1.1})$$

□

## A.2 Proof for lemma 4.2.4

*Proof.*

$$\begin{aligned}
Z &= - \sum_{i=1}^d \int_0^t H_i(s) dB_i(s) \\
&= - \int_0^t \bar{H}^T \cdot d\bar{B}(s) \\
&\quad (\bar{a}^T \text{ implies the vector / matrix transposed, and then } \cdot \text{ is a vector / matrix product}) \\
&= - \int_0^t (\sigma^{-1} \bar{\mu}(\bar{X}(t)))^T \sigma^{-1} (d\bar{X}(s) - \bar{\mu}(\bar{X}(t)) dt) \\
&= - \int_0^t \bar{\mu}^T(\bar{X}(t)) \sigma^{-T} \sigma^{-1} (d\bar{X}(s) - \bar{\mu}(\bar{X}(t)) dt) \\
&= - \int_0^t \bar{\mu}^T(\bar{X}(t)) \Sigma^{-1} d\bar{X}(s) + \int_0^t \bar{\mu}^T(\bar{X}(t)) \Sigma^{-1} \bar{\mu}(\bar{X}(t)) dt
\end{aligned} \tag{A.2.1}$$

Due to substituting  $d\bar{B}(s) = \sigma^{-1} (d\bar{X}(s) - \bar{\mu}(\bar{X}(t)) dt)$ .

Substitute  $dZ(t) = (\sigma^{-1} \bar{\mu}(\bar{X}(t)))^T dB(t)$  in the quadratic variation gives:

$$\begin{aligned}
d\langle Z, Z \rangle(t) &= d\left\langle (\sigma^{-1} \bar{\mu}(\bar{X}(t)))^T B, (\sigma^{-1} \bar{\mu}(\bar{X}(t)))^T B \right\rangle(t) \\
&= (\sigma^{-1} \bar{\mu}(\bar{X}(t)))^T (\sigma^{-1} \bar{\mu}(\bar{X}(t))) dt \\
&= \bar{\mu}^T(\bar{X}(t)) \sigma^{-T} \sigma^{-1} \bar{\mu}(\bar{X}(t)) dt \\
&= \bar{\mu}^T(\bar{X}(t)) \Sigma^{-1} \bar{\mu}(\bar{X}(t)) dt
\end{aligned}$$

and:

$$\langle Z, Z \rangle = \int_0^t \bar{\mu}^T(\bar{X}(s)) \Sigma^{-1} \bar{\mu}(\bar{X}(s)) ds \tag{A.2.2}$$

Thus the log-likelihood is:

$$\begin{aligned}
\log \Lambda &= Z - \frac{1}{2} \langle Z, Z \rangle \\
&= - \int_0^t \bar{\mu}^T(\bar{X}(s)) \Sigma^{-1} d\bar{X}(s) + \frac{1}{2} \int_0^t \bar{\mu}^T(\bar{X}(s)) \Sigma^{-1} \bar{\mu}(\bar{X}(s)) ds
\end{aligned} \tag{A.2.3}$$

Where the first integral is an Itô integral.

One can use Itô formula to re-write the first integral above:

$$\int_0^t f'(X(t)) dX = f(X(t)) - \frac{1}{2} \int_0^t f''(X(s)) d[X, X](s)$$

And in its 2 dimensional form:

$$\begin{aligned}
df(X(t), Y(t)) &= \frac{\partial f}{\partial x}(X(t), Y(t)) dX(t) + \frac{\partial f}{\partial y}(X(t), Y(t)) dY(t) \\
&\quad + \frac{1}{2} \frac{\partial^2 f}{\partial x^2}(X(t), Y(t)) d[X, X](t) + \frac{1}{2} \frac{\partial^2 f}{\partial y^2}(X(t), Y(t)) d[Y, Y](t) \\
&\quad + \frac{\partial^2 f}{\partial x \partial y}(X(t), Y(t)) d[X, Y](t)
\end{aligned}$$

Generally for d-dimensions:

$$\begin{aligned}
df(\bar{X}(t)) &= \sum_{i=1}^d \frac{\partial f}{\partial x_i}(\bar{X}(t)) dX_i(t) + \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \frac{\partial^2 f}{\partial x_i \partial x_j}(\bar{X}(t)) d[X_i, X_j](t) \\
&= \sum_{i=1}^d \frac{\partial f}{\partial x_i}(\bar{X}(t)) dX_i(t) + \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d (\sigma \sigma^T)_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j}(\bar{X}(t)) dt \\
&= \nabla f(\bar{X}(t)) d\bar{X}(t) + \frac{1}{2} \nabla^2 f d[\bar{X}, \bar{X}] \\
\text{and: } \nabla f(\bar{X}(t)) d\bar{X}(t) &= df(\bar{X}(t)) - \frac{1}{2} \nabla^2 f d[\bar{X}, \bar{X}]
\end{aligned}$$

note the two last lines are true when all  $X_i$  and  $X_j$  are independent.

Now recall we chose  $\bar{\mu}$  so there is a  $U$  such that  $\bar{\mu} = -\nabla U$ , so the Itô integral in equation A.2.3 is:

$$- \int_0^t \bar{\mu}^T(\bar{X}(s)) \Sigma^{-1} d\bar{X}(s) = \int_0^t (\bar{\nabla} U(\bar{X}(s)))^T \Sigma^{-1} d\bar{X}(s)$$

When we take  $\Sigma = \sigma^2 I$ , where  $I$  is the identity matrix and  $\sigma$  is a scalar (all BM are independent, and have the same constant  $\sigma$ ), so  $d[X_i, X_j] = \Sigma_{ij} dt = \sigma^2 dt$  if  $i = j$  and 0 otherwise. The Itô integral is:

$$\begin{aligned}
&= (\sigma^{-1})^2 \int_0^t dU(\bar{X}(s)) - \int_0^t \frac{1}{2} \nabla^2 U ds \\
&= (\sigma^{-1})^2 (U(\bar{X}(t)) - U(\bar{X}(0))) - \int_0^t \frac{1}{2} \nabla^2 U ds
\end{aligned}$$

And the log-likelihood is:

$$\begin{aligned}
\log \Lambda &= Z - \frac{1}{2} \langle Z, Z \rangle \\
&= (\sigma^{-1})^2 \left( U(\bar{X}(t)) - U(\bar{X}(0)) - \frac{1}{2} \sigma^2 \int_0^t \nabla^2 U ds + \frac{1}{2} \int_0^t |\bar{\mu}(\bar{X}(s))|^2 ds \right)
\end{aligned} \tag{A.2.4}$$

□



# Appendix B

## Additional results

### B.1 Additional numerical demonstration

#### B.1.1 Linear drift

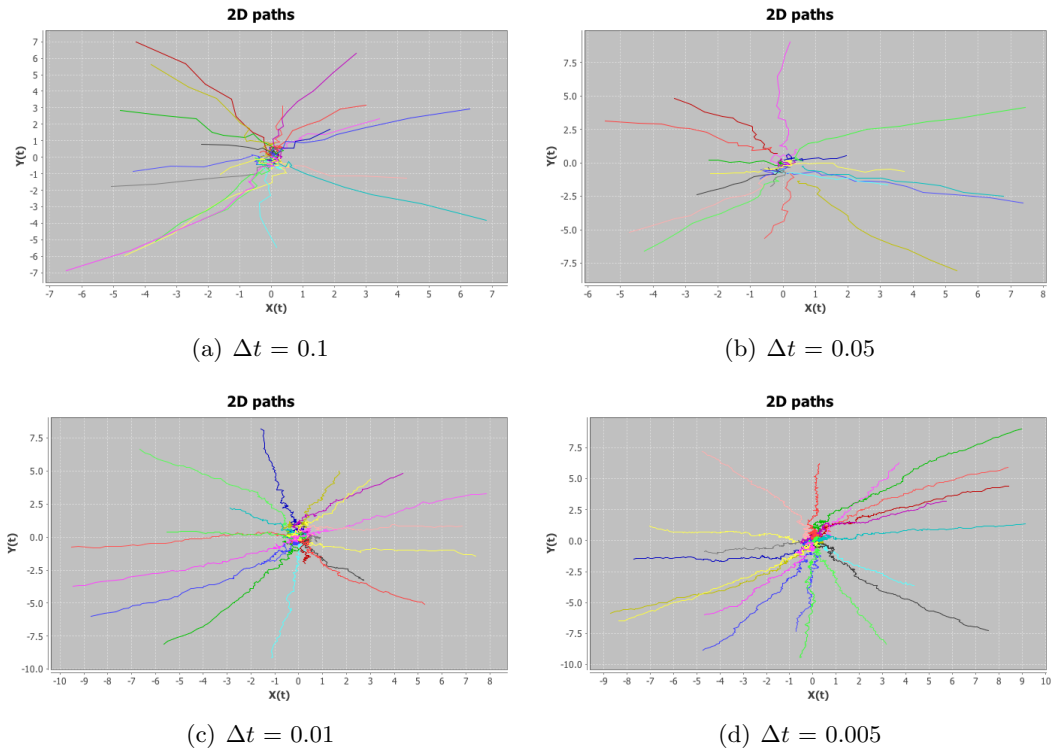


Figure B.1: Linear drift, computer generated sample paths,  $N = 20$ ,  $\alpha_x = 3$ ,  $\alpha_y = 2$ ,  $\sigma = 0.5$ . Random initial condition:  $\mathbf{q}(t = 0) \sim U[-10, 10]^{2N}$ .

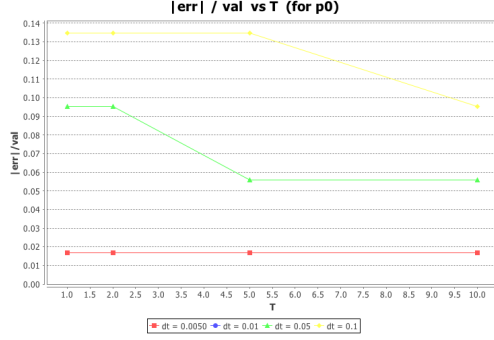
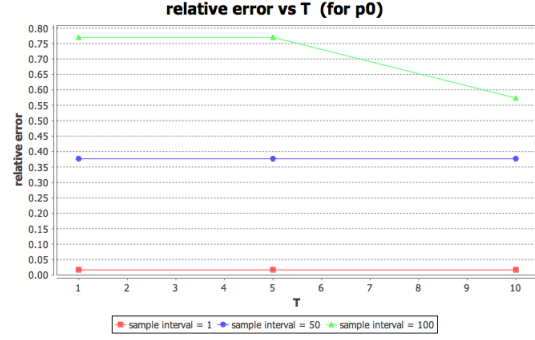
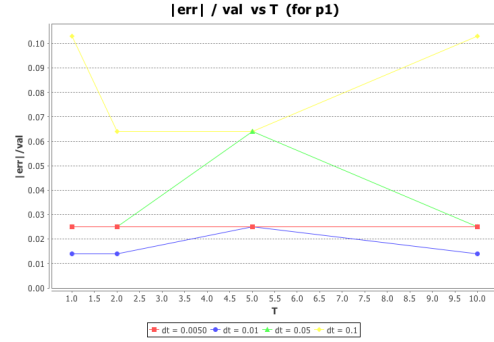
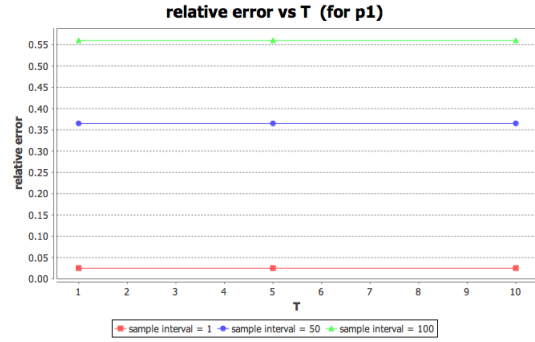
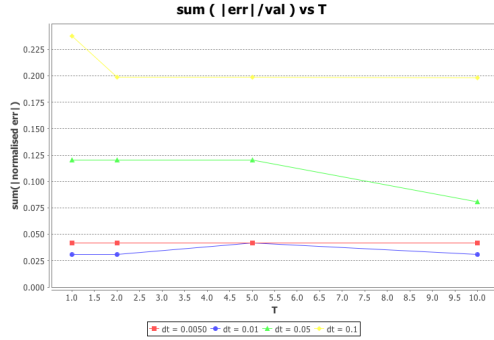
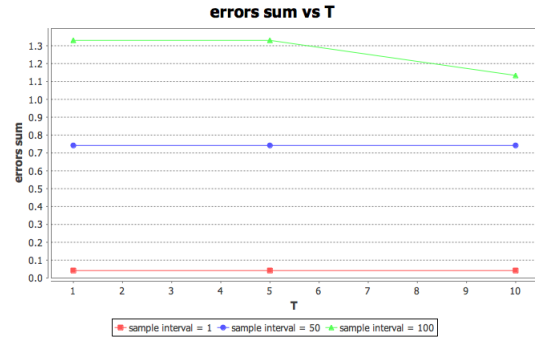
(a)  $err_x^{rel}$  for different  $\Delta t$  values(b)  $err_x^{rel}$  for different sample rates at  $\Delta t = 0.005$ (c)  $err_y^{rel}$  for different  $\Delta t$  values(d)  $err_y^{rel}$  for different sample rates at  $\Delta t = 0.005$ (e)  $err^{sum}$  for different  $\Delta t$  values(f)  $err^{sum}$  for different sample rates at  $\Delta t = 0.005$ 

Figure B.2: Linear drift, error estimations,  $N = 20$ . Here  $p_0 \equiv \alpha_x$  and  $p_1 \equiv \alpha_y$ . Left panels show error estimations obtained for different  $\Delta t$  values (integration intervals). Right panels show error estimations obtained for different sample rates at  $\Delta t = 0.005$

## B.1.2 Linear (attraction) interactions

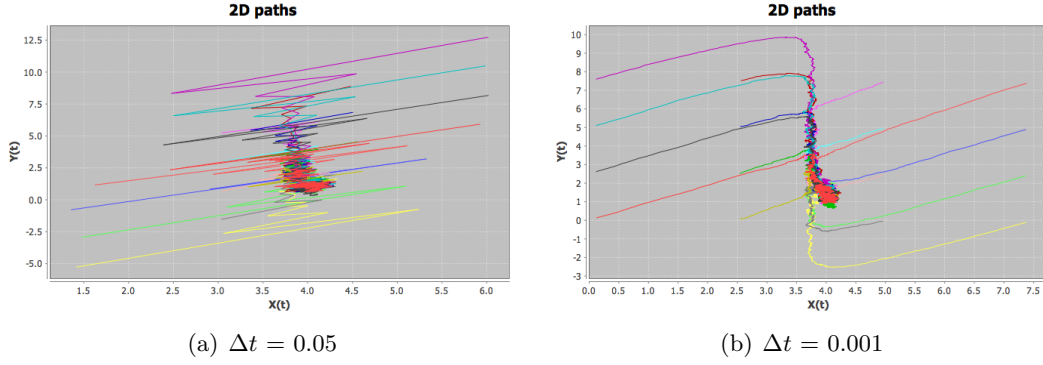


Figure B.3: Linear interactions (attraction only), computer generated sample paths.  $N = 16$ ,  $\alpha_0 = 1 (\equiv p_0)$ ,  $\sigma = 0.3$ . Initial condition was a grid of 4x4 nodes, over an 8x8 domain.

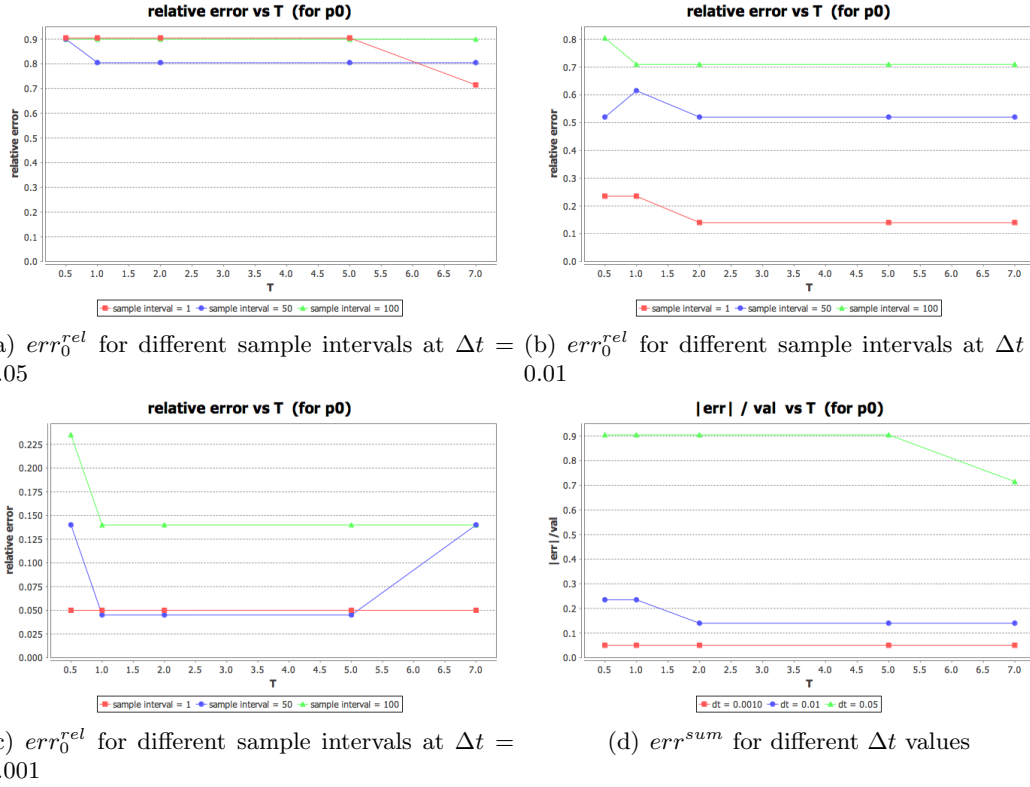


Figure B.4: Linear interactions (attraction only), error estimations,  $N = 16$ . (a)-(c) Error estimations obtained for different sample interval at  $\Delta t = 0.05, 0.01, 0.001$  respectively. (d) Error estimations obtained for different  $\Delta t$  values (with sample interval = 1).

### B.1.3 Exponential (repulsion and attraction) interactions

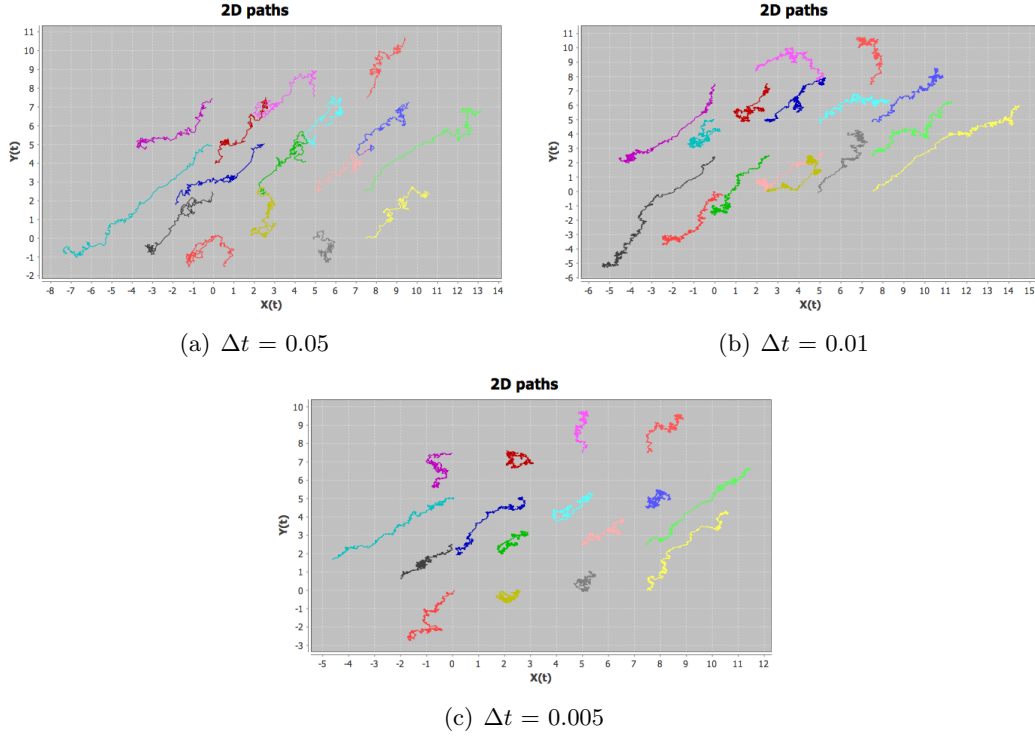
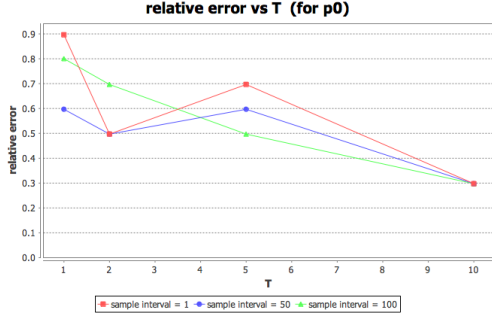
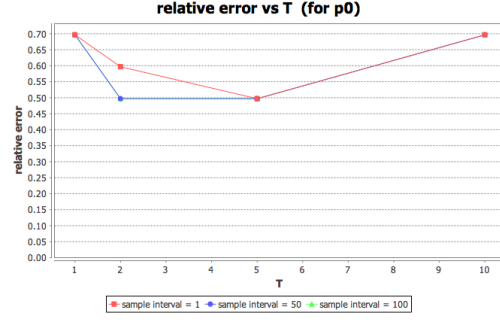
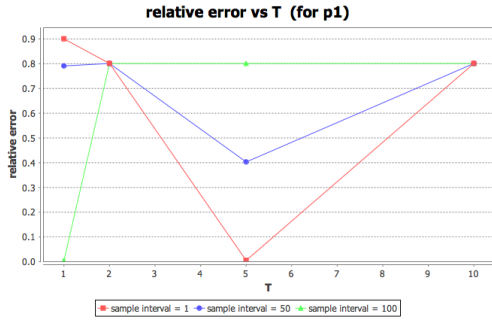
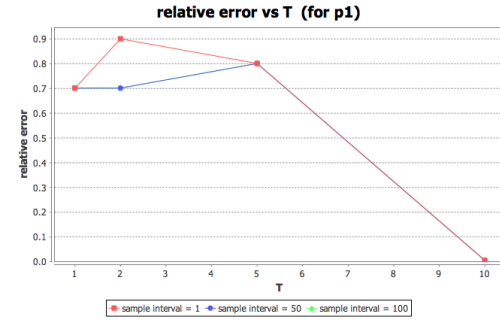
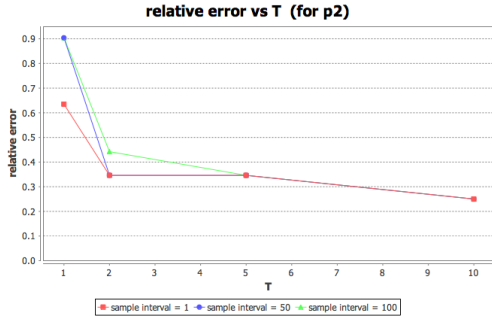
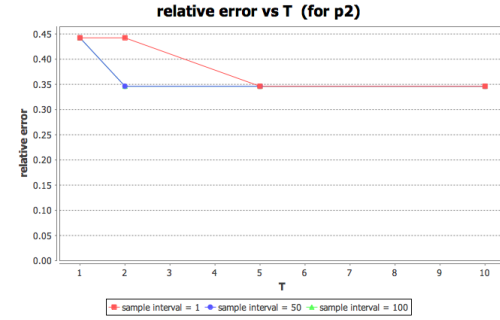
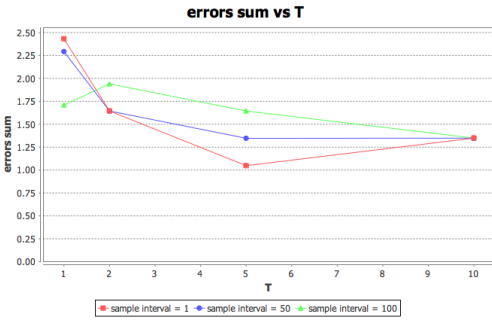
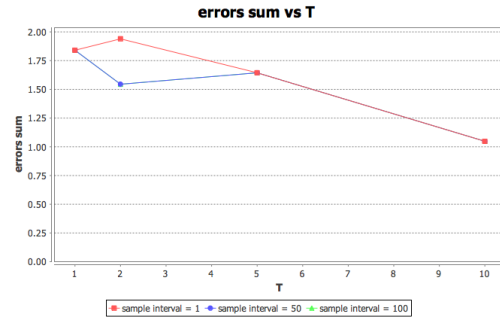


Figure B.5: Exponential interactions (repulsion and attraction forces included), computer generated sample paths.  $\bar{R} = 25 (\equiv p_0)$ ,  $\bar{A} = 10 (\equiv p_1)$ ,  $l = 1.3 (\equiv p_2)$  and  $\sigma = 0.3$  (number of particles  $N=16$ ). Initial condition was a grid of  $4 \times 4$  nodes, over an  $8 \times 8$  domain.



(a)  $err_0^{rel}$  for different sample intervals at  $\Delta t = 0.01$ (b)  $err_0^{rel}$  for different sample intervals at  $\Delta t = 0.005$ (c)  $err_1^{rel}$  different sample intervals at  $\Delta t = 0.01$ (d)  $err_1^{rel}$  for different sample intervals at  $\Delta t = 0.005$ (e)  $err_2^{rel}$  for sample intervals at  $\Delta t = 0.01$ (f)  $err_2^{rel}$  for different sample intervals at  $\Delta t = 0.005$ (g)  $err^{sum}$  for different sample intervals at  $\Delta t = 0.01$ (h)  $err^{sum}$  for different sample intervals at  $\Delta t = 0.005$ Figure B.6: Exponential interactions (repulsion and attraction), error estimations,  $N = 16$ . Here  $p_0 \equiv \bar{R}$ ,  $p_1 \equiv \bar{A}$  and  $p_2 \equiv l$ .

Left panels show error estimations obtained for different sample intervals at  $\Delta t = 0.01$ . Right panels show error estimations obtained for different sample intervals at  $\Delta t = 0.005$



# Appendix C

## Methods

### C.1 Methods for chapter 5

The experimental data analysed in this chapter was obtained by first irradiating cells using BBR with different levels of radiation dosage (0,2,5 and 10 Gy), and then observing them continually under the microscope for 48 hours, taking an image every 20 minutes. A tracking algorithm (a commercial product by LEICA microsystems) was used to track cells within those images and extract information about the position of each cell at each timeframe. This was done once with cells at a very low density, where cells were far from each other so effectively no interactions between cells could take place (low density). The second set up had cells at a high density (cells were grown till confluence) with a little gap that was made right before the radiation treatment took place (scratch assay), there cells were closely packed so interactions could take place, but also had a little room to move around so cell migration still occurred and cell paths could be observed.<sup>24</sup>

For each radiation level, the raw data for the analysis presented here came in 6 separate files: 3 files containing data from 3 different position within the low density set up, and 3 files containing data from 3 different position within the scratch assay. Each file contained lists of cell ID labels, positions and time stamps. First, an initial basic ‘rearranging’ and ‘cleaning’ of the data was performed, where the raw data was read and put into a suitable data structure from which cell paths can be readily drawn by either cell ID or a relevant time label. Zeros and gaps in the data were treated by extrapolation when possible, or by omitting data points which could not be resolved or identified with a particular cell path. The magnitude of cell motility, represented by  $\sigma$  in our model was estimated from cell displacements as explained in section 4.1.2. The MLE technique described in section 4.2 was applied to give an estimation of parameter values associated with cell-cell interactions given cell paths. At this stage, MLE was applied to data from a single repetition for each dosage level. Our program needs to be further automated in order to be able to identify groups of neighbouring cells within the data (this is done by the user at this stage), and pass their respective paths into the (already existing) part of the program which calculates the likelihoods and generates the estimations of parameter values.

---

<sup>24</sup>The experimental data analysed in this chapter was provided to us by Dr Jeffrey Crosbie from The Royal Women’s Hospital, and Dr Camden Lo from The Micro Imaging Centre at Monash University.

## C.2 Methods for chapter 6

### C.2.1 Numeric solutions

A JAVA script was written to solve a discretised version of the model's equations, and obtain  $N(t_i)$  and  $\mathbf{q}(t_i)$  (using a 1st order Euler method), for  $t_i \in [0, 1, \dots, 48 * 3600]_{sec}$ , given all other model parameters as inputs.

Model parameters set directly from observations of their physical sizes, are:  $\sigma^2 = 10$ ,  $r_{eff} = 15\mu m$  and  $\lambda = \frac{\log 2}{16} 1/h$ . Other parameters values are:  $A_{eph} = A_{eph rin} = 100$ ,  $R_{eph} = R_{eph rin} = 250$ ,  $a = 7.5$  and  $r = 5.8$  (unless mentioned otherwise).

Output was generated as images in which each cell is represented as a circle of radius  $r_{eff}$  around its calculated position, with a colour determined by the cell's type. A video file, made of all images, and a .csv file which includes the position of all cells at all calculated times were also generated.

### C.2.2 Image analysis

Analysis of both simulation and experimental results was done using ImageJ together with a designated MATLAB code written specifically for this purpose employing MATLAB's image analysis package.

For the analysis of microscopic images, thresholds were adjusted to guarantee optimal performance in capturing cell clusters with the least possible noise. Once clusters were identified, their areas and fluorescence intensities were recorded. The number of cells in a cluster was then estimated from the intensity by dividing by the average intensity of single cells determined from the images.

Images obtained from simulations have only three set colours (green, black and white) and therefore the threshold for identifying Eph (i.e. green) clusters was the same for all simulation setups. Once the clusters were identified, areas were recorded. The number of cells within each cluster was then obtained by crossing the data regarding all clusters positions and the .csv file containing all cell positions at the relevant time.

### C.2.3 Experimental setup

Segregation assays of EphB2- and ephrin-B1-expressing HEK293 cells were performed as described previously (Orsulic and Kemler, 2000), using approx. 50,000 cells seeded into a well ( $\sim 0.8 cm^2$  surface area) of a cell culture slide pre-coated with  $10 \frac{\mu g}{ml}$  fibronectin. To monitor cell movement over the course of the experiment (48-60h) we used time-lapse fluorescent microscopy (Leica AF6000LX microscope), taking images every 20 minutes. For experiments in Figure 3 and 4, images were taken at the end of the experiment when cells achieved confluence. In Figure 3, images are shown from previously reported experiments (Janes et al., 2011) using EphB2 cells stably transfected with the indicated constructs and labelled with Cell Tracker Green (Invitrogen). For experiments in Figure 4, EphB2 cells expressing E cadherin-cherry (kindly provided by E. Batlle) were produced by lentiviral transduction and sorted by flow cytometry. Western blots of cell protein extracts were with antibodies against EphB2 (R&D Systems) and E-cadherin (Abcam).<sup>25</sup>

### C.2.4 Statistical analysis of data regarding dynamics of cluster formation

In this experiment we look at the distribution of cluster sizes and its evolution over time. Ultimately, in order to show our mathematical model is capable of capturing the dynamics

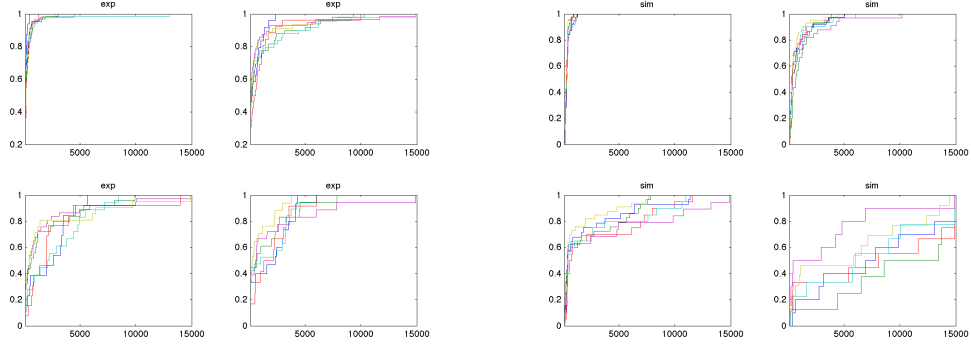
---

<sup>25</sup>These laboratory experiments were done by our collaborators Dr Peter Janes and Late Assoc. Prof. Martin Lackmann.

observed in laboratory experiments, we would like to perform a statistical test where the null hypothesis is that the two samples (the one obtained in the laboratory and the second obtained from simulations of numeric solutions to our model) are drawn from the same distribution. The choice of test is not as straight forwards as it first seems (the Kolmogorov Smirnov test for example), in particular due to the fact we have to also account for the time variable, which is a dependent variable.

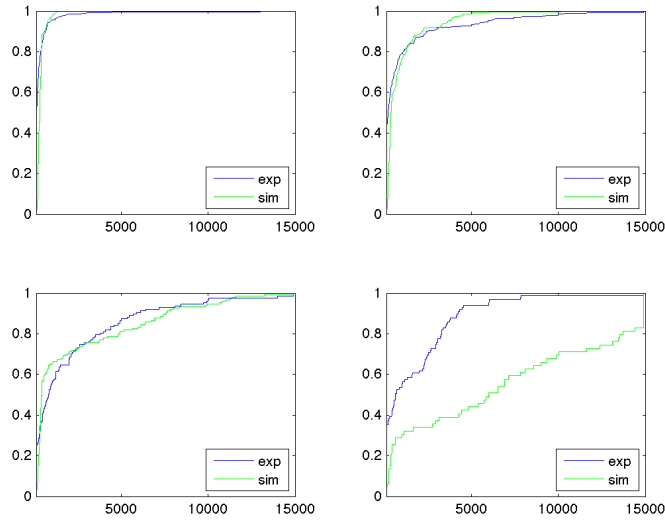
Figure C.1 presents semi-continuous snapshots of the cluster size distributions for both simulation and experimental results at four different times (0, 16, 33 and 49 hours). Cluster size distributions for the first three time frames appear close when comparing experimental to simulated results, but it is not possible to estimate the effect of time dependency in this manner. We therefore choose to group our measurements into three groups: small ( $100 - 1500 \mu m^2$ ), medium ( $1500 - 15000 \mu m^2$ ) and large ( $> 15000 \mu m^2$ ) cell clusters. This is often done in this specific field, in order to highlight the formation of clusters as the majority of clusters transition from small size to medium over time (Nievergall et al., 2010).

When the data is grouped in this manner, multivariate-ANOVA can be performed. Indeed this test is originally designed under the assumption the two data sets come from different distributions, and the  $p$ -value gives an indication of the significance level in which the hypothesis can be accepted or rejected by. We use multivariate-ANOVA here since we need to address the dependency of our sample on time, but unlike in its conventional use, as we are seeking for similarities between the sample rather than a significant difference. We therefore interpret the  $p$ -value obtained here merely as an indication of how close the two samples are.



(a) Experimental results

(b) Numeric results



(c) Averaged results

Figure C.1: Semi-continuous snapshots of cluster size distributions. Top left 0 hour, top right 16 hour, bottom left 33 hours, bottom right 49 hours. (a) Results obtained from 6 independent positions in a laboratory experiment. (b) Results obtained from 6 independent runs of our model simulation. (c) Averaged results appear close for all times but 49 hours. See explanation within main text (section 6.3.1).

# Vita

Publications arising from this thesis include:

- R. Aharon, P. Janes, A.W. Burgess, F. Klebaner, K. Hamza, M. Lackman. (2011).**  
A stochastic model for cell migration and segregation regulated by Eph-ephrin interactions. *International Conference on Applied Mathematics*. Vancouver, Canada.
- J.C. Crosbie, R. Aharon, C. Lo, C. Nowell, I. Mohammed-Johari, F. Klebaner, K. Hamza, and P.A.W. Rogers. (2012).** The effect of increasing radiation doses on normal and malignant cell migration. *European Radiation Research*. Vietri sul Mare, Italy.
- Aharon, R., Janes, P. W., Burgess, A. W., Hamza, K., Klebaner, F. and Lackmann, M. (2014).** A mathematical model for eph/ephrin-directed segregation of intermingled cells, *Plos One* 9(12).
- Aharon, R., Hamza, K. and Klebaner, F. (to be submitted).** A stochastic model for cell migration and interaction, *Mathematical Biology*.
- Aharon, R., Crosbie, J., Hamza, K. and Klebaner, F. (to be submitted).** Analysis of cells response to radiation treatment.

Permanent Address: School of Mathematical Sciences  
Monash University  
Australia

This thesis was typeset with  $\text{\LaTeX} 2_{\epsilon}$ <sup>26</sup> by the author.

---

<sup>26</sup> $\text{\LaTeX} 2_{\epsilon}$  is an extension of  $\text{\LaTeX}$ .  $\text{\LaTeX}$  is a collection of macros for  $\text{\TeX}$ .  $\text{\TeX}$  is a trademark of the American Mathematical Society. The macros used in formatting this thesis were written by Glenn Maughan and modified by Dean Thompson and David Squire of Monash University.





# References

- Aharon, R., Janes, P. W., Burgess, A. W., Hamza, K., Klebaner, F. and Lackmann, M. (2014). A mathematical model for eph/ephrin-directed segregation of intermingled cells, *PloS one* **9**(12).
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2008). *Molecular biology of the cell*, Garland Science.
- Alt, W. and Dembo, M. (1999). Cytoplasm dynamics and cell motion: two-phase flow models, *Mathematical Biosciences* **156**(1–2): 207 – 228.
- Baker, R. E., Yates, C. A. and Erban, R. (2010). From microscopic to macroscopic descriptions of cell migration on growing domains, *Bulletin of mathematical biology* **72**(3): 719–762.
- Batlle, E. and Wilkinson, D. G. (2012). Molecular mechanisms of cell segregation and boundary formation in development and tumorigenesis, *Cold Spring Harbor Perspectives in Biology* **4**(1).
- Bhattacharya, R. and Waymire, E. (1990). *Stochastic processes with applications*, John Wiley & sons, Inc.
- Bianconi, E., Piovesan, A., Facchin, F., Beraudi, A., Casadei, R., Frabetti, F., Vitale, L., Pelleri, M. C., Tassani, S., Piva, F., Perez-Amodio, S., Strippoli, P. and Canaider, S. (2013). An estimation of the number of cells in the human body, *Annals of Human Biology* **40**(6): 463–471.
- Bindschadler, M., Osborn, E., Jr., C. D. and McGrath, J. (2004). A mechanistic model of the actin cycle, *Biophysical Journal* **86**(5): 2720 – 2739.
- Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities, *The journal of political economy* pp. 637–654.
- Bovier, A. (2015). *Gaussian processes on trees: From spin glasses to branching Brownian motion*, Unpublished lecture notes, WS 2014/15, Bonn.
- Bowden, L. G., Simpson, M. J. and Baker, R. E. (2013). Design and interpretation of cell trajectory assays, *Journal of The Royal Society Interface* **10**(88).
- Brillinger, D. (2003). Simulating constrained animal motion using stochastic differential equations, *Lecture Notes-Monograph Series* pp. 35–48.
- Carter, N., Nakamoto, T., Hirai, H. and Hunter, T. (2002). Ephrina1-induced cytoskeletal re-organization requires fak and p130(cas), *Nature Cell Biology* **4**(8): 565–73.
- Chen, N., Glazier, J. A., Izaguirre, J. A. and Alber, M. S. (2007). A parallel implementation of the cellular potts model for simulation of cell-based morphogenesis, *Computer physics communications* **176**(11): 670–681.

- Cooke, J. E., Kemp, H. A. and Moens, C. B. (2005). EphA4 is required for cell adhesion and rhombomere-boundary formation in the zebrafish, *Current biology* **15**(6): 536–542.
- Cortina, C., Palomo-Ponce, S., Iglesias, M., Fernández-Masip, J., Vivancos, A., Whissell, G., Humà, M., Peiró, N., Gallego, L., Jonkheer, S. et al. (2007). EphB–ephrin-B interactions suppress colorectal cancer progression by compartmentalizing tumor cells, *Nature Genetics* **39**(11): 1376–1383.
- Crosbie, J., Anderson, R., Rothkamm, K., Restall, C., Cann, L., Ruwanpura, S., Meachem, S., Yagi, N., Svalbe, I., Lewis, R. et al. (2010). Tumor Cell Response to Synchrotron Microbeam Radiation Therapy Differs Markedly From Cells in Normal Tissues, *International Journal of Radiation Oncology\* Biology\* Physics* **77**(3): 886–894.
- Daar, I. O. (2012). Non-sh2/pdz reverse signaling by ephrins, *Seminars in cell and developmental biology*, Vol. 23, Elsevier.
- Davis, S., Gale, N. W., Aldrich, T. H., Maisonpierre, P. C., Lhotak, V., Pawson, T., Goldfarb, M. and Yancopoulos, G. D. (1994). Ligands for eph-related receptor tyrosine kinases that require membrane attachment or clustering for activity, *Science* **266**(5186): 816–9.
- Davy, A. and Soriano, P. (2005). Ephrin signaling in vivo: look both ways, *Developmental Dynamics* **232**(1): 1–10.
- Deroulers, C., Aubert, M., Badoual, M. and Grammaticos, B. (2009). Modeling tumor cell migration: From microscopic to macroscopic models, *Physical Review E* **79**(3): 031917.
- Dickinson, R. and Tranquillo, R. (1993). A stochastic model for adhesion-mediated cell random motility and haptotaxis, *Journal of Mathematical Biology* **31**(6): 563–600.
- Egea, J. and Klein, R. (2007). Bidirectional eph-ephrin signaling during axon guidance, *Trends in Cell Biology* **17**(5): 230–8.
- Espanol, P. and Warren, P. (1995). Statistical mechanics of dissipative particle dynamics, *EPL (Europhysics Letters)* **30**(4): 191.
- Etheridge, A. (2000). An introduction to superprocesses, volume 20 of univ, *Lecture Series. AMS, Rhode Island*.
- Fernando, A., Landman, K. and Simpson, M. (2010). Nonlinear diffusion and exclusion processes with contact interactions, *Physical Review E* **81**(1): 11903.
- Flanagan, J. G. (2006). Neural map specification by gradients, *Current Opinion in Neurobiology* **16**(1): 59–66.
- Foo, S. S., Turner, C. J., Adams, S., Compagni, A., Aubyn, D., Kogata, N., Lindblom, P., Shani, M., Zicha, D. and Adams, R. H. (2006). Ephrin-b2 controls cell motility and adhesion during blood-vessel-wall assembly, *Cell* **124**(1): 161–73.
- Gatenby, R. A. and Gawlinski, E. T. (1996). A reaction-diffusion model of cancer invasion, *Cancer Research* **56**(24): 5745–5753.
- Genander, M. and Frisn, J. (2010). Ephrins and eph receptors in stem cells and cancer, *Current Opinion in Cell Biology* **22**(5): 611–616.

- Genander, M., Halford, M. M., Xu, N. J., Eriksson, M., Yu, Z., Qiu, Z., Martling, A., Greicius, G., Thakar, S., Catchpole, T., Chumley, M. J., Zdunek, S., Wang, C., Holm, T., Goff, S. P., Pettersson, S., Pestell, R. G., Henkemeyer, M. and Frisen, J. (2009). Dissociation of ephb2 signaling pathways mediating progenitor cell proliferation and tumor suppression, *Cell* **139**(4): 679–92.
- Girsanov, I. V. (1960). On transforming a certain class of stochastic processes by absolutely continuous substitution of measures, *Theory of Probability & Its Applications* **5**(3): 285–301.
- Glazier, J. A. and Graner, F. (1993). Simulation of the differential adhesion driven rearrangement of biological cells, *Physical Review E* **47**(3): 2128.
- Gracheva, M. E. and Othmer, H. G. (2004). A continuum model of motility in ameboid cells, *Bulletin of Mathematical Biology* **66**(1): 167 – 193.
- Huynh-Do, U., Stein, E., Lane, A., Liu, H., Cerretti, D. and Daniel, T. (1999). Surface densities of ephrin-b1 determine ephb1-coupled activation of cell attachment through alphavbeta3 and alpha5beta1 integrins, *EMBO Journal* **18**(8): 2165–2173.
- Ionides, E., Fang, K., Rivkah Isseroff, R. and Oster, G. (2004). Stochastic models for cell motion and taxis, *Journal of Mathematical Biology* **48**(1): 23–37.
- Janes, P., Adikari, S. and Lackmann, M. (2008). Eph/ephrin signalling and function in oncogenesis: Lessons from embryonic development, *Current Cancer Drug Targets* **8**(6): 473–9.
- Janes, P. W., Griesshaber, B., Atapattu, L., Nievergall, E., Hii, L. L., Mensinga, A., Chheang, C., Day, B. W., Boyd, A. W., Bastiaens, P. I., Jørgensen, C., Pawson, T. and Lackmann, M. (2011). Eph receptor function is modulated by heterooligomerization of a and b type eph receptors, *The Journal of Cell Biology* **195**(6): 1033–1045.
- Janes, P. W., Saha, N., Barton, W. A., Kolev, M. V., Wimmer-Kleikamp, S. H., Nievergall, E., Blobel, C. P., Himanen, J. P., Lackmann, M. and Nikolov, D. B. (2005). Adam meets eph: An adam substrate recognition module acts as a molecular switch for ephrin cleavage in trans, *Cell* **123**(2): 291–304.
- Jorgensen, C., Sherman, A., Chen, G. I., Pasculescu, A., Poliakov, A., Hsiung, M., Larsen, B., Wilkinson, D. G., Linding, R. and Pawson, T. (2009). Cell-specific information processing in segregating populations of eph receptor ephrin-expressing cells, *Science* **326**(5959): 1502–1509.
- Karlin, S. and Taylor, H. (1981). *A second course in stochastic processes*, Academic press.
- Keller, E. F. and Segel, L. A. (1971). Model for chemotaxis, *Journal of Theoretical Biology* **30**(2): 225 – 234.
- Kil, W. J., Tofilon, P. J. and Camphausen, K. (2012). Post-radiation increase in vegf enhances glioma cell motility in vitro, *Radiation Oncology* **7**(25): 717X–7.
- Klebaner, F. C. (2011). *Introduction to stochastic calculus with applications*, Imperial College Press.
- Kutoyants, Y. A. (2013). *Statistical inference for ergodic diffusion processes*, Springer Science & Business Media.

- Lackmann, M. and Boyd, A. W. (2008). Eph, a protein family coming of age: more confusion, insight, or complexity?, *Science Signalling* **1**(15).
- Lawrenson, I., Wimmer-Kleikamp, S., Lock, P., Schoenwaelder, S., Down, M., Boyd, A., Alewood, P. and Lackmann, M. (2002). Ephrin-a5 induces rounding, blebbing and de-adhesion of epha3- expressing 293t and melanoma cells by crkii and rho-mediated signalling, *Journal of Cell Science* **115**(Pt 5): 1059–1072.
- Lin, C., Segel, L. and Handelman, G. (1988). *Mathematics applied to deterministic problems in the natural sciences*, Society for Industrial Mathematics.
- Lushnikov, P. M., Chen, N. and Alber, M. (2008). Macroscopic dynamics of biological cells interacting via chemotaxis and direct contact, *Physical Review E* **78**(6): 061904.
- Maini, P. K. (2000). *MATHEMATICAL MODELLING IN THE LIFE SCIENCES: APPLICATIONS IN PATTERN FORMATION AND WOUND HEALING*, WORLD SCIENTIFIC, chapter 7, pp. 201–237.
- Maini, P., Olsen, L. and Sherratt, J. (2002). Mathematical models for cell-matrix interactions during dermal wound healing, *International Journal of Bifurcation and Chaos* **12**(09): 2021–2029.
- McLaughlin, T. and O’Leary, D. D. (2005). Molecular gradients and development of retinotopic maps, *Annual Review of Neuroscience* **28**: 327–55.
- McLennan, R., Schumacher, L. J., Morrison, J. A., Teddy, J. M., Ridenour, D. A., Box, A. C., Semerad, C. L., Li, H., McDowell, W., Kay, D., Maini, P. K., Baker, R. E. and Kulesa, P. M. (2015). Neural crest migration is driven by a few trailblazer cells with a unique molecular signature narrowly confined to the invasive front, *Development* **142**(11): 2014–2025.
- Mellitzer, G., Xu, Q. and Wilkinson, D. (1999). Eph receptors and ephrins restrict cell intermingling and communication, *Nature* **400**(6739): 77–81.
- Métivier, M. (1982). *Semimartingales: a course on stochastic processes*, Vol. 2, Walter de Gruyter.
- Miao, H., Burnett, E., Kinch, M., Simon, E. and Wang, B. (1999). Activation of epha2 kinase suppresses integrin function and causes focal-adhesion-kinase dephosphorylation, *Nature Cell Biology* **2**(2): 62–69.
- Miao, H., Li, D. Q., Mukherjee, A., Guo, H., Petty, A., Cutter, J., Basilion, J. P., Sedor, J., Wu, J., Danielpour, D., Sloan, A. E., Cohen, M. L. and Wang, B. (2009). Epha2 mediates ligand-dependent inhibition and ligand-independent promotion of cell migration and invasion via a reciprocal regulatory loop with akt, *Cancer Cell* **16**(1): 9–20.
- Miao, H., Strebhardt, K., Pasquale, E. B., Shen, T. L., Guan, J. L. and Wang, B. (2005). Inhibition of integrin-mediated cell adhesion but not directional cell migration requires catalytic activity of ephb3 receptor tyrosine kinase: Role of rho family small gtpases, *Journal of Biological Chemistry* **280**(2): 923–32.
- Mogilner, A. (2009). Mathematics of cell motility: have we got its number?, *Journal of Mathematical Biology* **58**(1): 105–134.
- Mogilner, A., Edelstein-Keshet, L., Bent, L. and Spiros, A. (2003). Mutual interactions, potentials, and individual distance in a social aggregation, *Journal of mathematical biology* **47**(4): 353–389.

- Nielsen, J. N., Madsen, H. and Young, P. C. (2000). Parameter estimation in stochastic differential equations: an overview, *Annual Reviews in Control* **24**: 83–94.
- Niemierko, A. and Goitein, M. (1992). Modeling of normal tissue response to radiation: The critical volume model, *International Journal of Radiation Oncology* **25**(1): 135–145.
- Nievergall, E., Janes, P. W., Stegmayer, C., Vail, M. E., Haj, F. G., Teng, S. W., Neel, B. G., Bastiaens, P. I. and Lackmann, M. (2010). Ptp1b regulates eph receptor function and trafficking, *Journal of Cell Biology* **191**(6): 1189–203.
- Nievergall, E., Lackmann, M. and Janes, P. (2011). Eph-dependent cell-cell adhesion and segregation in development and cancer, *Cellular and Molecular Life Sciences* **69**(11): 1813–42.
- Nievergall, E., Lackmann, M. and Janes, P. W. (2012). Eph-dependent cell-cell adhesion and segregation in development and cancer, *Cellular and Molecular Life Sciences* **69**(11): 1813–1842.
- Noren, N. K., Foos, G., Hauser, C. A. and Pasquale, E. B. (2006). The ephb4 receptor suppresses breast cancer cell tumorigenicity through an abl-crk pathway, *Nature Cell Biology* **8**(8): 815–25.
- Orsulic, S. and Kemler, R. (2000). Expression of eph receptors and ephrins is differentially regulated by e-cadherin, *Journal of cell science* **113**(10): 1793–1802.
- Ouchi, N. B., Glazier, J. A., Rieu, J.-P., Upadhyaya, A. and Sawada, Y. (2003). Improving the realism of the cellular potts model in simulations of biological cells, *Physica A: Statistical Mechanics and its Applications* **329**(3): 451–458.
- Pasquale, E. B. (2008). Eph-ephrin bidirectional signaling in physiology and disease, *Cell* **133**(1): 38–52.
- Pasquale, E. B. (2010). Eph receptors and ephrins in cancer: bidirectional signalling and beyond, *Nature Reviews Cancer* **10**(3): 165–180.
- Pokern, Y., Stuart, A. M. and Vanden-Eijnden, E. (2009). Remarks on drift estimation for diffusion processes, *Multiscale Modeling & Simulation* **8**(1): 69–95.
- Poliakov, A., Cotrina, M. L., Pasini, A. and Wilkinson, D. G. (2008). Regulation of ephb2 activation and cell repulsion by feedback control of the mapk pathway, *The Journal of Cell Biology* **183**(5): 933–47.
- Rao, B. P. (1999). *Statistical inference for diffusion type processes*, Arnold.
- Reber, M., Burrola, P. and Lemke, G. (2004). A relative signalling model for the formation of a topographic neural map, *Nature* **431**(7010): 847–53.
- Rockne, R., Alvord, E.C., J., Rockhill, J. and Swanson, K. (2009). A mathematical model for brain tumor response to radiation therapy, *Journal of Mathematical Biology* **58**(4-5): 561–578.
- Ross, R. J., Yates, C. A. and Baker, R. E. (2015). Inference of cell–cell interactions from population density characteristics and cell trajectories on static and growing domains, *Mathematical Biosciences* **264**: 108–118.
- Rubenstein, B. M. and Kaufman, L. J. (2008). The role of extracellular matrix in glioma invasion: a cellular potts model approach, *Biophysical journal* **95**(12): 5661–5680.

- Sachs, R., Hlatky, L. and Hahnfeldt, P. (2001). Simple {ODE} models of tumor growth and anti-angiogenic or radiation treatment, *Mathematical and Computer Modelling* **33**: 1297 – 1305.
- Salaita, K., Nair, P. M., Petit, R. S., Neve, R. M., Das, D., Gray, J. W. and Groves, J. T. (2010). Restriction of receptor movement alters cellular response: Physical force sensing by epha2, *Science* **327**(5971): 1380–1385.
- Scianna, M. (2015). An extended cellular potts model analyzing a wound healing assay, *Computers in Biology and Medicine* **62**(2): 35–54.
- Scianna, M. and Preziosi, L. (2012). Multiscale developments of the cellular potts model, *Multiscale Modeling & Simulation* **10**(2): 342–382.
- Shao, D., Levine, H. and Rappel, W.-J. (2012). Coupling actin flow, adhesion, and morphology in a computational cell motility model, *Proceedings of the National Academy of Sciences* .
- Shaw, G., Morse, S., Ararat, M. and Graham, F. L. (2002). Preferential transformation of human neuronal cells by human adenoviruses and the origin of hek 293 cells, *The FASEB Journal* **16**(8): 869–871.
- Simpson, M. J., Landman, K. A., Hughes, B. D. and Fernando, A. E. (2010). A model for mesoscale patterns in motile populations, *Physica A: Statistical Mechanics and its Applications* **389**(7): 1412–1424.
- Smallbone, K., Gavaghan, D. J., Gatenby, R. A. and Maini, P. K. (2005). The role of acidity in solid tumour growth and invasion, *Journal of Theoretical Biology* **235**(4): 476 – 484.  
**URL:** <http://www.sciencedirect.com/science/article/pii/S0022519305000548>
- Soize, C. (1994). *The Fokker-Planck equation for stochastic dynamical systems and its explicit steady state solutions*, Vol. 17, World Scientific.
- Solanas, G., Cortina, C., Sevillano, M. and Batlle, E. (2011). Cleavage of e-cadherin by adam10 mediates epithelial cell sorting downstream of ephb signalling, *Nature Cell Biology* **13**(9): 1100–1107.
- Steinberg, M. S. (2007). Differential adhesion in morphogenesis: a modern view, *Current Opinion in Genetics & Development* **17**(4): 281 – 286. Pattern formation and developmental mechanisms.
- Stokes, C., Lauffenburger, D. and Williams, S. (1991). Migration of individual microvessel endothelial cells: stochastic model and parameter measurement, *Journal of Cell Science* **99**(2): 419–430.
- Taylor, W., Katsimitsoulia, Z. and Poliakov, A. (2011). Simulation of cell movement and interaction, *Journal of Bioinformatics and Computational Biology* **9**(01): 91–110.
- Taylor, W., Morley, R., Krasavin, A., Gregory, L., Wilkinson, D. and Poliakov, A. (2012). A mechanical model of cell segregation driven by differential adhesion, *PloS one* **7**(8).
- Timmer, J. (2000). Parameter estimation in nonlinear stochastic differential equations, *Chaos, Solitons & Fractals* **11**(15): 2571–2578.
- Watanabe, T. and Takahashi, Y. (2010). Tissue morphogenesis coupled with cell shape changes, *Current Opinion in Genetics & Development* **20**(4): 443–447.

- Wimmer-Kleikamp, S. H., Nievergall, E., Gegenbauer, K., Adikari, S., Mansour, M., Yeadon, T., Boyd, A. W., Patani, N. R. and Lackmann, M. (2008). Elevated protein tyrosine phosphatase activity provokes eph/ephrin-facilitated adhesion of pre-b leukemia cells, *Blood* **112**(3): 721–32.
- Wong, S. Y., Chiam, K.-H., Lim, C. T. and Matsudaira, P. (2010). Computational model of cell positioning: directed and collective migration in the intestinal crypt epithelium, *Journal of The Royal Society Interface* **7**(Suppl 3): S351–S363.
- Xu, Q., Mellitzer, G. and Wilkinson, D. (2000). Roles of eph receptors and ephrins in segmental patterning, *Philosophical Transactions of the Royal Society B: Biological Sciences* **355**(1399): 993–1002.
- Zheng, Q., Liu, Y., Zhou, H., Du, Y., Zhang, B., Zhang, J., Miao, G., Liu, B. and Zhang, H. (2015). X-ray radiation promotes the metastatic potential of tongue squamous cell carcinoma cells via modulation of biomechanical and cytoskeletal properties, *Human & Experimental Toxicology* .