The spatiotemporal dynamics of inflammatory neutrophil populations

A thesis submitted to the University of Sheffield
for the degree of Doctor of Philosophy

Geoffrey Robert Holmes

Department of Automatic Control and Systems Engineering

March 2013

Abstract

Inflammation is a natural response of the innate immune system. It has evolved over the course of millennia to deal with the threat of injury and infection. The neutrophil is a powerful immune cell that plays a vital role in inflammation but when dysfunctional it can cause chronic diseases such as asthma, chronic obstructive pulmonary disease and arthritis. These illnesses have a devastating impact on the lives of sufferers. A complete understanding of the inflammation response is, therefore, a vital ongoing area of research where breakthroughs will potentially bring huge benefit to individuals and to society.

Neutrophils have been studied *in vitro* for over a hundred years. A new opportunity has recently arisen to observe their dynamics *in vivo* in transgenic zebrafish larvae which are transparent and have an immune system with similarities of form and function to our own. These new data present a modelling and system identification challenge: how to infer cell dynamics from limited amounts of data in a complex extra-cellular environment.

This thesis addresses the problem by modelling populations of neutrophils using a drift-diffusion model of cell dynamics. Firstly, a weighted regression framework is developed which uses observations of mean squared cell displacements to identify neutrophil migration coefficients during recruitment and resolution phases of inflammation. As a result the recruitment dynamics of inflammatory neutrophils are successfully quantified *in vivo*. Whilst this framework is more rigorous than existing approaches, it was not conclusive for model determination.

A second computational framework is therefore presented which reformulates the approximate Bayesian sequential Monte Carlo algorithm for use in the cell migration context. In particular, the Cha-Srihari distance is used to compare the distributions of cell populations. Drift-diffusion models are then extended to include chemoattractant receptor depletion dynamics and spatial variability in the extracellular environment. When this framework was applied to zebrafish neutrophils during inflammation resolution, a key result was that this migration is the unguided result of inherent stochastic cell movements. This contrasts with the externally guided dynamics during recruitment. An important conclusion is that the search for influences driving neutrophils away from a wound is futile and the focus should be on the mechanisms whereby neutrophils are desensitised to signals that retain them in the inflamed area.

Acknowledgements

I thank my supervisor Visakan Kadirkamanathan for being supportive of my return to academic study and for his insights and encouragement along the way; also to Debbie Proctor for arranging all the meetings.

I thank Stephen Renshaw whose collaboration has provided focus for the research in this thesis and whose help has been invaluable; also to Giles Dixon and Anne Robertson for all the data.

I thank Sean Anderson for technical discussions and for his perennial positivity and encouragement which have made a huge impact on my progress.

I thank Andrew Hills for introducing me to LaTeX, Python, Ubuntu, Bash, Iceberg and Git and then troubleshooting many of the problems I've encountered as a result. I also thank Tara Baldacchino, Eliza Condrea, Veronica Biga, Andrew Zammit Mangion, Michael Pelegrinis and Parham Aram, all of whom have made my time studying more enjoyable. And thank you to all my colleagues who bought eggs.

*

I am always grateful to my parents,
my mother Anne and late father Nigel
for all that they have given me.

*

I especially thank my wife, Rebecca, for her love and support
and for putting up with my crazy idea of studying for a PhD.

*

Finally,
I dedicate this to our children,
Imogen, Jemima, Reuben and Rosanna,
who brighten up every day.

Table of Contents

List of Figures

List of Tables

List of Algorithms

Nomenclature

A list of the variables and notation used in this thesis is defined below. The definitions and conventions set here will be observed throughout unless otherwise stated. For a list of acronyms, please consult page 131.

$\mathbf{a}$	vector
A	matrix
$\mathcal{Y}$	set
θ	parameter set
$\top$	transpose
∇	differential operator
$\mathcal{U}(a,\, b)$	uniform distribution taking values in the range $[a, b]$
$\mathcal{N}(\mu,\, \Sigma)$	normal distribution with mean μ and covariance Σ
P(.)	probability operator
P(. | .)	conditional probability operator
$\mathbb{E}\,(.)$	expectation operator

Chapter 1

Introduction

1.1 Background

The neutrophil is a type of white blood cell and one that plays a key role in the inflammatory response of the vertebrate innate immune system. During this process any damage to the host organism and invasion of infectious threats must be rapidly dealt with. On such occasions, it is vital that neutrophils be activated and deployed successfully. However, they must eventually also become de-activated and then disperse for normal equilibrium and health to return to the system. Activated neutrophils are powerful agents.

The ongoing study of these processes is important. Not least because the inflammatory response can be sometimes be inappropriately activated and sometimes may fail to resolve properly. This can result in severe damage to otherwise healthy tissue and organs. Such pathological states lie behind many chronic illnesses which are prevalent in modern society such as chronic obstructive pulmonary disease (COPD), bronchial asthma and arthritis. These are burdens which put huge demands upon the resources of our healthcare systems as well as upon the lives of individuals and their families.

To quote Carl Nathan [90]: *'Understanding the circuits that confer and control [neutrophil] behaviour is as challenging a problem as any other in cell biology. Meeting that challenge holds therapeutic promise in diverse settings of intense interest, ..., from metastatic tumours to ravaged joints.'*

Experimental studies are fundamentally important for the furthering of this understanding. More and more, however, mathematical modelling and system identification techniques are being seen as vitally important, to bring out the full implications of experimental results and help shape the priorities for the next generation of experiments. Modelling is the process by which a mathematical description of

a system is built up from known physical properties of the system. System identification determines the unknown parameters of candidate models of the system and weighs evidence to determine which is the best model to describe the system. This is a data-driven process in which prediction of the model systems are compared to measured outputs of the physical system being investigated. This thesis applies modelling and system identification principles to the study of neutrophil dynamics during the recruitment and resolution phases of an inflammation event.

The role and function of neutrophils in the immune system is an area that has been studied for over a hundred years and much has already been discovered and reported. However, with the advent of new experimental procedures and imaging techniques - especially those for experiments *in vivo*, new types of dataset are available. These new datasets bring new challenges to the task of data analysis which require the development of novel and innovative techniques for modelling, parameter estimation and model identification in order to explore and understand the cellular dynamics. Meeting these challenges means we will be better able to identify the key processes and hence to predict and control the outcomes.

Recently, a new opportunity to take this field of research forward has arisen with the development of the zebrafish (*danio rerio*) as a biological model* for the *in vivo* study of neutrophilic inflammation and its resolution. The zebrafish has the particular advantages of being genetically well known and amenable to alterations which allow fluorescent labelling of particular cell lines. The larvae are transparent which means the labelled cells can be observed with relative ease. And whilst being economical in terms of space and cost to maintain, it is, importantly, a species in which the immune system has significant functional similarities to our own.

The zebrafish model thus allows collection of data which records the whole of an inflammation and resolution event in which a particular cell line, the neutrophil, is singled out from others without confusion. Therefore, the new research challenge is to find novel frameworks within which to analyse this data, in particular to answer questions about the mechanisms by which neutrophil are recruited to and resolved from a site of inflammation.

In summary, the recent developments which allow specific visualisation of neutrophils through a complete inflammation cycle opens up the difficult challenge of modelling these cells. This thesis presents a framework for modelling them and thereby characterises important aspects of their dynamics.

*Model is used here in the biologist's sense of a genetically modified form of an organism which is amenable to experimental investigation.

1.2 Aims and objectives of this thesis

The central aim of this thesis is to characterise the nature of neutrophil dynamics during inflammation and inflammation resolution. Data is obtained from zebrafish larvae, but the conclusions have implications for understanding the human immune system due to similarities in form and function between that and the zebrafish immune system.

Whereas many recent neutrophil studies focus on the individual cell level, this thesis takes a previously unexplored cell population approach to this particular problem. Using the drift-diffusion partial differential equation (PDE) to draw analogies with molecular dynamics in fluids, it characterises the neutrophil dynamics in terms of drift and diffusion coefficients. These have a correspondence to biological realities of directed cell movement (drift-diffusion dynamics) and undirected, purely stochastic movement (which will be referred to here as pure-diffusion dynamics).

Within the central aim, a key question in current neutrophil research concerns neutrophil dynamics during inflammation resolution: are the resolving cells directed away from the inflamed area or are they moving freely according to inherent random patterns? This thesis answers this question on the population level and comes to a conclusion which contradicts a currently emerging consensus. The results presented here, therefore, have important implications for the direction of future research into the inflammation resolution process.

In order to achieve its aim, this thesis accomplishes the following objectives:

- Apply regression analysis to solutions of the drift-diffusion equation to develop a relatively straightforward procedure for estimating cell recruitment and resolution dynamics during inflammation. The aim of this approach is to be understandable and usable by experimentalists. It will be illustrated by application to the zebrafish data.

- Develop a Bayesian framework for analysis of cell dynamics. This framework will be applicable to any complex situation which can be simulated but not necessarily be expressed in an analytic probability model. The framework will have a built-in model selection strategy.

- Develop a range of candidate models, suggested by experimental studies, which are suitable for modelling the zebrafish neutrophil data within the Bayesian framework.

- Apply the Bayesian framework to zebrafish neutrophil data to select between the various candidate models and estimate the neutrophils' dynamic properties.

The data analysed is obtained from zebrafish larvae by research partners*. These datasets are novel in that, (a) they achieve neutrophil specific labelling of cells; and, (b) by labelling with photo-convertible Kaede protein, a subgroup of the neutrophils are differentially labelled. This twofold approach facilitates separate analysis of neutrophils that are being recruited to the inflammation region and neutrophils that are concurrently migrating away from that region.

1.3 Thesis overview

The structure of the remainder of this thesis can be summarised as follows.

Chapter 2 reviews the relevant literature relating to neutrophils, some of the models used to characterise them, and an overview of some common system identification techniques. It introduces the approximate Bayesian computation (ABC) approach which will be developed for the neutrophil migration problem and applied in Chapters 4 and 5.

Chapter 3 develops a regression based analysis framework which is applied to the zebrafish neutrophil data using models which are derived from the drift-diffusion PDE. The analysis is applied both to neutrophils that are being recruited to inflammation and neutrophils that are subsequently migrating away from the inflammation region. Whilst similar approaches have been used in a molecular context, to the best of the author's knowledge this represents a novel approach to cell dynamics and the framework can be applied in many other similar cell migration settings.

Chapter 4 develops further the key question posed above concerning inflammation resolution dynamics: whether there is directed or purely stochastic migration. It does this by developing a novel simulation based Bayesian cell modelling framework which allows parameter estimation and model selection between arbitrarily related models for which no likelihood function is easily computed. This framework is then applied to the zebrafish data using simulation models corresponding to the pure-diffusion and drift-diffusion models.

Chapter 5 seeks a more complete model description of the neutrophil migration dynamics in order to increase the robustness of the identification results. It does this by taking into account (a) temporal variation in the neutrophil dynamics via a novel model for attractant ligand receptors, and also (b) spatial variation via a novel way of characterising inhomogeneities in the extra-cellular matrix within which the neutrophils move. These new models are then applied to the neutrophil data using the framework developed in Chapter 4. Of particular interest here is whether the new results affect the conclusion concerning neutrophil migration dynamics during

*From the Department of Infection and Immunity, University of Sheffield

inflammation resolution.

Chapter 6 makes some conclusions and highlights areas of further work which follow on from this thesis.

1.4 Papers arising from this thesis

Material from this thesis has formed the basis for two published papers: from Chapter 3,

- G. Holmes, G. Dixon, S. Anderson, C. Reyes-Aldasoro, P. Elks, S. Billings, M. Whyte, V. Kadirkamanathan, and S. Renshaw. Drift-diffusion analysis of neutrophil migration during inflammation resolution in a zebrafsh model. *Advances in Hematology*, 2012. [53]

and from Chapters 4 and 5,

- G.R. Holmes, S.R. Anderson, G. Dixon, A.L. Robertson, C.C. Reyes-Aldasoro, S.A. Billings, S.A. Renshaw, and V. Kadirkamanathan. Repelled from the wound, or randomly dispersed? reverse migration behaviour of neutrophils characterised by dynamic modelling. *Journal of The Royal Society Interface*, September 2012. [52]

A third related paper to which the author contributed has been published during the period of research for this thesis,

- V. Kadirkamanathan, S.R. Anderson, S.A. Billings, X. Zhang, G.R. Holmes, C.C. Reyes-Aldasoro, P.M. Elks, S.A. Renshaw. The Neutrophil's Eye-View: Inference and Visualisation of the Chemoattractant Field Driving Cell Chemotaxis In Vivo. *PLoS One*, 2012. [65]

A further conference paper based on material from Chapters 4 and 5 is currently in submission,

- G.R. Holmes, S.R. Anderson, G. Dixon, S.A. Renshaw, V. Kadirkamanathan. A Bayesian framework for identifying cell migration dynamics.

Finally, a paper is in preparation which uses the Bayesian framework and models from Chapter 4 and 5, applying them to data collected from drug altered zebrafish neutrophil datasets,

- Anne L. Robertson, Aleksandra N. Bojarczuk, Geoffrey R. Holmes, Sean R. Anderson, Stuart N. Farrow, Roberto Solari, Visakan Kadirkamanathan, Moira K. B. Whyte and Stephen A. Renshaw. A novel anti-inflammatory mechanism defined by an in vivo zebrafish inflammation screen.

1.5 Summary of contributions

- A regression framework for identifying cell migration dynamics (Chapter 3).
- Quantification of zebrafish neutrophil recruitment dynamics during inflammation (Chapter 3).
- A Bayesian framework for identifying cell migration dynamics (Chapter 4).
- Novel models for neutrophil migration during inflammation and inflammation resolution (Chapter 5).
- Evidence that zebrafish neutrophil migration during inflammation resolution is purely stochastic (Chapters 3, 4, 5).

Chapter 2

Literature review

2.1 Neutrophils in the immune system

2.1.1 Background

Neutrophils are the most abundant type of mammalian leukocyte* in the circulation [3]. They play a crucial role in the innate immune system: their purpose is to find and neutralise microbial invaders in their host organism. In order to do this the are able to process complex signals and deploy a range of attacking strategies as part of the inflammation process [90]. Unfortunately, their power against infection carries a concomitant potential to cause serious collateral damage to the host [84].

Neutrophils are eukaryotic cells. This means that they have a cell nucleus which is in contrast to prokaryotic cells such as bacteria. Neutrophils have been known and studied since the late 19th century when Paul Erlich used a staining procedure to analyse subtypes of leukocyte [16]. The name 'neutrophil' arose from the tendency of these cells to retain neutral dyes. At the same time Elie Metchnikoff discovered the neutrophil's basic function as a microbial phagocyte [50], in other words they consume invading microbes. Observing their lobulated nuclei, Metchnikoff gave them the name polymorphonuclear leukocyte (PMN). Both neutrophil and PMN are still commonly used terms for these cells. Neutrophils are part of the granulocyte subtype of leukocytes [16]. The term granulocyte relates to the densely packed granules in their cytoplasm - structures that enables them to carry potent proteins safely. These proteins are encapsulated but ready to be released and deployed when the time and place is right.

Since their discovery over a century ago, neutrophils have proved difficult subjects of research. They are short lived cells with a 6-8 hour lifespan in the circulation and they are not amenable to growth in tissue culture [3]. Furthermore, the complex

*white blood cell

signalling processes they are part of in the complex environments in which they carry out their functions means that *ex vivo* studies are quite limited in what they can reveal. Modern techniques mean that much progress has been made but there are still many unanswered questions which remain to be addressed [118]. These questions include,

- What chemoattractants, and possibly what chemorepellants, drive neutrophil migration?
- What are the dynamics of neutrophil migration towards and away from inflammation sites?
- What are the functions and fate of these reverse migrating cells?
- What are the signalling mechanisms that underlie neutrophil polarisation and chemotaxis?

This thesis focusses on the second of these questions.

2.1.2 The neutrophil's role in inflammation

Neutrophils [3, 15] are generated within the bone marrow. From there they move into the bloodstream and are immediately receptive to any evidence of a threat to the host organism. Such a threat typically occurs in the epithelium*. An injury, together with any associated infectious incursion, immediately releases a cascade of chemical signals, some host derived and some microbial. When these reach the vicinity of the vasculature they interact with endothelial† cells which in turn produce markers on the walls of the blood vessels. Circulating neutrophils repeatedly come into contact with these walls and when they encounter such a marker it initiates them into a process of activation. First they start a rolling motion along the inner edge of the endothelium. Further contact with more of the pro-inflammation markers arrests them in a state of adhesion. They are then able to migrate through the endothelial wall and enter the interstitial space.

The neutrophils are now in a very different context to the vasculature and they are surrounded by the multi layered chemical field of pro-inflammatory agents and signals. Their heightened state of activation enables them to perform the multiple tasks of navigating towards the site of the threat, preparing to deploy their antimicrobial agents and releasing other chemicals that further facilitate the innate immune response and initiate the adaptive immune response. In doing this they are

*Epithelial tissues lines cavities and surfaces throughout the body including the skin.
†Endothelial cells form a thin surface on the interior of blood vessels.

interacting comprehensively with other immune cells and recruiting them into the inflammation area [117].

When they arrive in the target area, neutrophils reach the final stage of activation in which they have three strategies for dealing with infection. The first and most basic is phagocytosis - the engulfment of alien cells (and other cell debris) to nullify the threat these pose. Neutrophils will actively follow microbe derived signals and relentlessly pursue their target until they overtake and consume it. The second strategy is degranulation: the toxic proteins, up to now safely carried in the cytoplasmic granules, are released where they are needed. And thirdly, as part of programmed cell death (apoptosis) neutrophils may release extra-cellular traps [19].

Apoptosis of neutrophils, in turn, produces signals which lead to the dampening of the inflammatory response. Recruitment of further neutrophils slows down and eventually ceases. Phagocytosis of apoptic neutrophils by macrophages ensures that secondary death of apoptic neutrophils does not lead to a new inflammation event and also puts the macrophage cells into anti-inflammatory mode thus promoting tissue restoration. In addition to apoptosis, as will be described in Section 3.2 there is increasing evidence that inflammation resolution may be effected by neutrophils migrating away from inflammatory sites[20, 85, 131], and possibly even returning to the vasculature [21].

2.1.3 The importance of inflammation resolution and the means by which it may occur

The responses described above have evolved over hundreds of millions of years enabling a vigorous and effective response to invading micro-organisms. For millennia, human life was relatively short and such risks were ever present. Now, however, with the advent of modern medicine, life expectancy has increased. Over this longer span we are more aware of the potential pathological side effects arising from the potency of the neutrophil [84, 132].

Inappropriate activation of neutrophils or failure in the resolution process of an inflammation event often results in conditions of chronic inflammation [111]. This leads to harmful accumulations of toxic substances in the tissues which can lead, in turn, to one or more of several chronic illnesses. Examples include bronchial asthma, rheumatoid arthritis, cystic fibrosis and also COPD, which according to the World Health Organisation, is one of the leading causes of death in the developed world [132]. In these conditions the severity of inflammation often has a direct inverse correlation with positive outcome [3]. Adverse neutrophil action can also support disease progression in other illnesses such as cancer and autoimmune syndromes.

These negative consequences of inflammation clearly cause loss of life expectancy

and loss of life quality for millions of people worldwide. Associated with this is a huge drain on public healthcare resources. For these reasons, research which will lead to the understanding of the mechanisms of inflammation and its resolution as well as of the dysregulated versions of these process, is vital at this time. This thesis makes an important contribution to this area.

In the past it was thought that inflammation resolution was a passive process in the sense that the cues for inflammation decayed over time and neutrophil recruitment ceased. It is now believed that inflammation resolution is an active process, programmed in a similar manner to the way that inflammation itself is [111]. This has important therapeutic consequences: both pro-resolution and anti-inflammation interventions may be possible, and care needs to be taken that any anti-inflammatory interventions do not upset the inherent inflammation resolution programme and thereby cause unintended and unforeseen harm.

Resolution of inflammation, is clearly a multifaceted process [111]. The aspect of it that this thesis is particularly concerned with is how neutrophils leave or are removed from the affected tissues. It is recognised that this can happen in various ways:

- Apoptosis and subsequent clearance by other phagocytes.
- Migration away into the surrounding tissues.
- Migration back into bloodstream.

This thesis investigates how migration away into the surrounding tissue occurs. In particular, it investigates whether this is an actively directed process (in the sense described in the following section) or a purely passive (undirected) process which is the result of inherent random behaviours. It should be noted, however, that the belief that inflammation resolution as a whole is an active programmed process does not pre-empt or conflict with this particular issue to be addressed.

2.1.4 An inflammation model

Ordinary differential equation (ODE)s are frequently used to model biological systems. This approach has been applied to the inflammation response with a view to prediction and control. The following model of Parker and Clermont [99] illustrates

the approach.

$$\frac{\mathrm{d}P}{\mathrm{d}t} = k_P(1-P)P - k_{PM}MP \tag{2.1}$$

$$\frac{\mathrm{d}M}{\mathrm{d}t} = (k_{MP}P + L)M(1-M) - k_{PM}MP - k_M M \tag{2.2}$$

$$\frac{\mathrm{d}L}{\mathrm{d}t} = k_{LM}\left(1 + \tanh\left(\frac{M - M_0}{w}\right)\right) - k_L MP \tag{2.3}$$

where P is a pro-inflammation agent, M represents the immune system response, which could be measure, for instance, by the concentration of neutrophils, and L represents a quantification of dysfunction, such as host cell damage. Equations (2.1)-(2.3) are illustrated and explained in Figure 2.1. The figure also shows how a further term can be added to represent an anti-inflammation agent, C [31]. These models

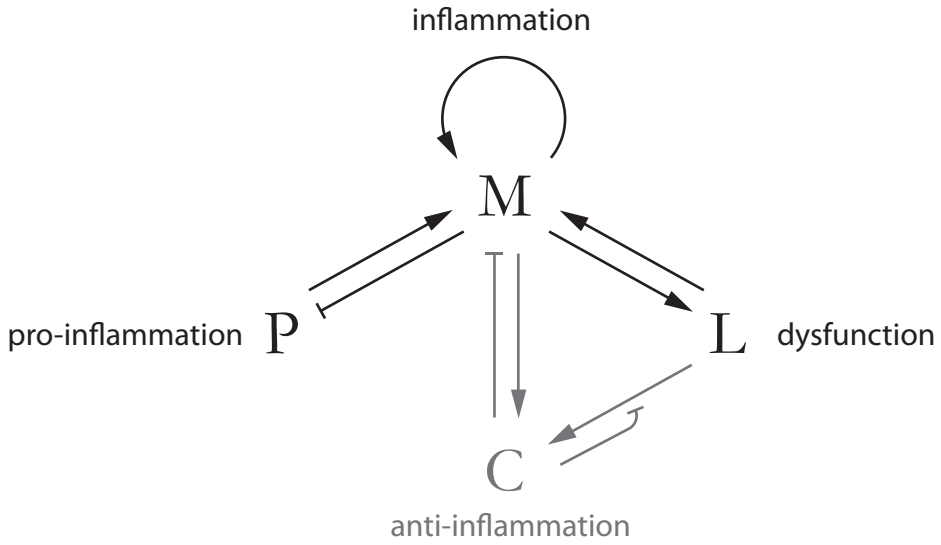

Figure 2.1: **Inflammation Model** [99]: Pro-inflammation agent, P, (e.g. bacterial concentration) promotes inflammatory response, M. M, in turn, is self promoting (e.g. neutrophil attract other neutrophils), suppressing P but also causes damage, L. The damage, L can also promote further inflammatory response. If an anti-inflammation agent, C is present, it introduces a negative feedback loop, reducing the possibility of unchecked inflammation which cannot resolve.

can be investigated by solving the equations with different parameter choices to see what the range of dynamic system responses can occur. They can also be calibrated or made to be biologically meaningful via combination of choosing parameters based on *a priori* knowledge or other studies in the literature and estimating parameters by minimising the difference between model predictions and experimental results.

ODE models such as this have no spatial dimension but can be given some spatial aspect by compartmental modelling. In compartmental modelling, systems of ODE

apply in a number of separate compartments and then transitions and interactions between compartments are also modelled on another level. Whilst such modelling reveals much about the inflammation process it is not suitable for modelling cell migration dynamics which requires a full spatio-temporal treatment.

2.1.5 Neutrophil migration

Cell migration [70] in its broadest sense is resolved and studied on three temporal scales: locomotion, translocation and migration [34]. Locomotion is the scale on which the mechanics of the individual cell are considered: cell polarisation, the formation of protrusions which reach out and adhere to the substrate* creating tractive force, and the release of rearward adhesions to allow forward movement (at or below this level one may also consider the binding of ligands and intracellular signalling which are effecting the locomotion). Translocation is the scale of several locomotion cycles of a cell. At this level, local directional correlations (persistence) are seen to occur and decay over time, and global directional bias may begin to be observed. On the migration scale, the detailed tracks of individual cells are of less interest. Migration relates to the position of cells at various times or the evolving distribution of a cell population as a whole. It is on this level of migration as populational distribution that neutrophil dynamics will be modelled in this thesis.

Migration is therefore the result of translocation, and it has been shown that translocation can be directed by signals external to the migrating cell in various different ways : haptotaxis is when a cell moves up an adhesion gradient [34] and chemotaxis is when a cell moves up a chemoattractant gradient [136]; to these can be added fugetaxis [127], the opposite of chemotaxis, with the cell migrating down the chemoattractant gradient, and galvanotaxis when a cell moves up a electrical field gradient [57].

If there are several forms of taxis, there are also various modes whereby these taxis events can be achieved [34]: Topotaxis is a mode of preferential turning toward the direction of maximum gradient; Orthotaxis is when the cell moves faster when moving in the gradient direction; Klinotaxis is the reduction of random turning when the cell is oriented in the gradient direction; orthokinesis is when cell speed increases with chemoattractant concentration; klinokinesis is when cell turning reduces with chemoattractant concentration.

The ability of neutrophils to direct their movements along a chemical gradient was first demonstrated by Zigmond in 1977 [135]. It has subsequently been shown that neutrophils can and perhaps must be guided sequentially towards their target by

*the substrate is the surface or medium on which a cell moves or is attached also known as the extra cellular matrix.

a sequence of competing chemoattractants [44]. The neutrophils are able to integrate the signals and use 'memory' to prioritise and respond to the newly distinguished signals each in turn [45]. Furthermore having migrated up an attractant gradient to a saturation level where orientation is no longer possible, they are able to navigate down the first gradient in response to a new more distant signal. It is found also that target (i.e. bacterium or dying cell) induced attractants are prioritised over regulatory cell induced attractants. This is accomplished by the former having an inhibitory effect on the ligand binding of the latter.

Arrieumerloua and Meyer [5] suggest and describe 3 basic chemotaxis modes for eukaryotic cells: Correlated Random Walking, Biased Random Walking and Persistent Migration. The first of these is purely stochastic but with short term directional persistence which decays over time. The difference between the latter two modes is that in a biased random walk, movement is possible in any direction but net movement is towards the target, whereas in persistent migration movement is never away from the target. A key point arising from the findings of this thesis is the importance of not misrepresenting short term directional correlation in purely random motion as though it is evidence of persistent migration.

2.1.6 The opportunity offered by transgenic zebrafish

Throughout most of the last century leukocyte research relied on *in vitro* methods to uncover new knowledge and understanding of cell phenotype and function [130]. More recently a number of animal models* have been developed which, together with modern video microscopy equipment, have allowed a revolution to occur in investigating how the cells behave in their native context. The preferred context for *in vivo* neutrophil study has been the mouse, and in the mouse a visualisation of neutrophil recruitment has been achieved with striking results [87, 131]. There are, however, limits to what can be achieved in this way. During recruitment the neutrophils are moving into the observed region of the animal but their positions and paths prior to arrival are not able to be observed. Furthermore, during inflammation resolution, by its nature, the cells are moving away and tend to disappear back into regions of tissue where observation is not possible. For this reason it is much more difficult to study the dynamics of these resolving cells as they are removed from inflammatory sites. In addition, genetic approaches allowing distinction *in vivo* of neutrophils from other immune cells are still uncommon.

The transgenic zebrafish model [97], in contrast, allows *in vivo* visualisation of individual immune cells during inflammation resolution [38, 107, 108]. Zebrafish are

*Model is used here in the biologist's sense of a genetically modified form of an organism which is amenable to experimental investigation.

genetically tractable and efficient in terms of both cost and space. Furthermore, the zebrafish has similar immune system characteristics to those in mammals. Not only do neutrophils have similar functionality in mammals as in bony fish generally [84], but more particularly, zebrafish neutrophils, as in humans, are the most numerous of the leukocytes [76] and are the first to be recruited to an inflammation event where they perform the usual vital functions of degranulation and phagocytosis [28, 40, 48, 118]. Using the zebrafish in its transparent larval form, neutrophils tagged with green fluorescent protein (GFP) markers can be directly observed and recorded *in vivo* and throughout the organism using video microscopy. In addition to this, by labelling the neutrophils with Kaede protein [4], which fluoresces green but changes to red on exposure to a certain frequency of light, sub-populations of the cells can be distinguished and observed separately. The behaviour of these distinct cell groups then be analysed to provide insight into the underlying patterns governing their movement. Data of this kind (though without the Kaede enhancement) has already been used to show that Hydrogen Peroxide (H_2O_2) is a likely candidate for rapid recruitment of neutrophils to wound sites in zebrafish [94].

2.2 Models for neutrophil migration

2.2.1 Individual cell models

On the individual cell level there is much work in the literature to try to uncover and understand the mechanics of cell migration using mathematical models. A particular goal has been to find a modelling framework which can unify the twin aspects of (a) random but directionally correlated motion in the absence of any guiding cue, with (b) the existence of directionally guided motion when such cues are present.

Stokes et al. [119] build on the earlier work of Uhlenbeck and Ornstein [124], Doob [35] and Dunn and Brown [37] to develop a stochastic migration model. This model is described in the following two equations for velocity, $\mathbf{v}$ and position, $\mathbf{x}$:

$$\mathrm{d}\mathbf{v}(t) = -\beta\mathbf{v}(t)\,\mathrm{d}t + \sqrt{\alpha}\,\mathrm{d}\mathbf{W}(t) + \mathbf{\Psi}(t)\,\mathrm{d}t \tag{2.4}$$

$$\mathbf{x}(t) = \int_0^t \mathbf{v}(s)ds \tag{2.5}$$

Equation (2.5) is self explanatory whilst (2.4) describes the change in velocity as the net result of two deterministic impulses and a random acceleration.

$-\beta\mathbf{v}(t)$ is the drag, directly proportional to and opposing the current velocity.

$\mathbf{\Psi}(t)$ is the force due to a biasing chemoattractant.

$\mathbf{W}(t)$ is the random velocity disturbances at time t.

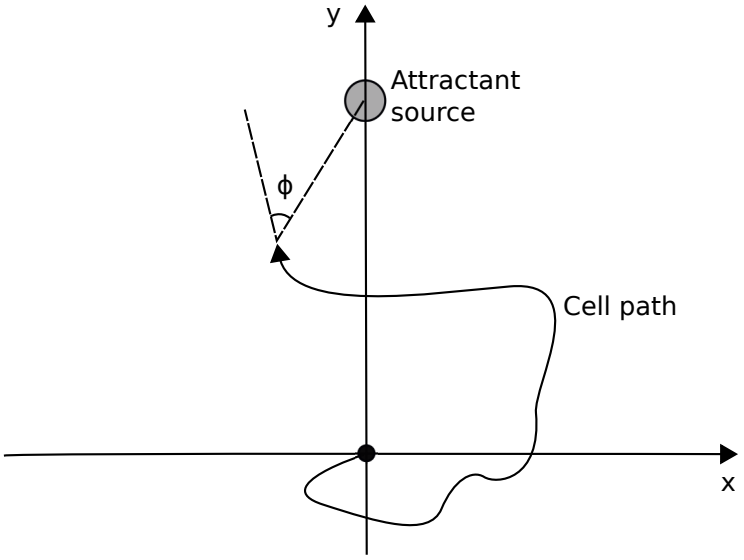

Figure 2.2: **Stochastic model for cell migration.** From Stokes et al. [119], adapted with permission*.

In this model it is assumed that the source of chemoattractant is localised and therefore the bias direction is spatially dependent. They use the results of Nossal and Zigmond [96] to derive a formula for the bias force:

$$\mathbf{\Psi} = \kappa \nabla a \sin \left| \frac{\phi}{2} \right| \tag{2.6}$$

where κ is the chemotactic responsiveness, ∇a is the gradient of attractant and ϕ is the current angle between $\mathbf{v}$ and ∇a, as shown in Figure 2.2.

Nossal and Zigmond [96] show, however, that the attraction increases with ϕ up to an angle of 90^o. There is no evidence given for angles over this. An alternative model of gradient sensing, proposed by Tranquillo et al. [122] and discussed below, implies instead that the sensed gradient would decrease as ϕ increases from 90^o to 180^o in which case a more appropriate formula would be:

$$\mathbf{\Psi} = \kappa \nabla a \sin |\phi| \,. \tag{2.7}$$

Also, whilst in the model proposed by Stokes et al. the bias is assumed to act as a force on the particle, in the light of the model of Tranquillo et al. [122], it might be more appropriate to treat it as a turning moment acting on the orientation of the particle.

Neutrophils and other leukocytes show short term directional persistence within

*With permission from Journal of Cell Science, The Company of Biologists.

globally random migration patterns when moving in the absence of any chemotactic signal. They can also move with directional accuracy under the influence of an attracting bias even when the gradient inducing the bias is very small (e.g. $1-2\%$). Tranquillo et al. [122] presented a model for leukocyte motility which reconciles these two facts.

The model, illustrated in Figure 2.3 assumes that the cells have constant polarisation and move forward at a constant speed independently of the turning mechanism described by the model. The leading edge of the polarised cell, or lamellipodium, is modelled as two interacting compartments. The surface of each compartment has a set of attractant receptors. When a receptor is occupied by the corresponding ligand, motile force effector, M, is produced within the compartment. The turning rate of the cell is proportional to the difference between the amount of effector in each compartment,

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} \propto M_1 - M_2 \tag{2.8}$$

The quantities M_1 and M_2 are calculated by solving differential equations which include terms for production, diffusion and decay of M. Production relies in turn on chemoattractant concentration, instantaneous and mean proportion of receptor occupancy and direction of travel through the chemoattractant field. The stochasticity which drives the model is introduced through a stochastic differential equation for instantaneous receptor occupancy.

Models such as Tranquillo et al. [122] and those based on it rely on internal signalling and integration of gradient sensing across the leading edge of the cell. In contrast to this some more recent models propose and find evidence for cell turning being due to the net result of independent lamellipod extensions on either side of the leading edge of the cell. For instance Arrieumerloua and Meyer [5] assumed that each receptor binding event within the leading edge triggers a local lamellipod extension, and that the average local lamellipod response to this creates a small turn in the direction of migration. On average the cell will tend to turn towards the source of the chemoattractant. From these assumptions they derive a distribution for cell orientation as a function of local chemoattractant gradient:

$$P(\alpha) = \gamma e^{\frac{g}{\delta}\cos(\alpha)} \tag{2.9}$$

where $-\pi < \alpha < \pi$, γ is a normaliser, δ is an incremental angle created by lamellipod extension in response to a single signalling event, and g is a relative measure of the chemoattracant gradient derived from the concentrations c1 and c2 at the left and

right side respectively of the leading edge,

$$g = 2\frac{c1 - c2}{c1 + c2} \tag{2.10}$$

Arrieumerloua and Meyer call the ratio $\frac{g}{\delta}$ the compass parameter. The higher the value of this ratio, the more efficiently the cell is able to find its way to the source of the chemoattracant.

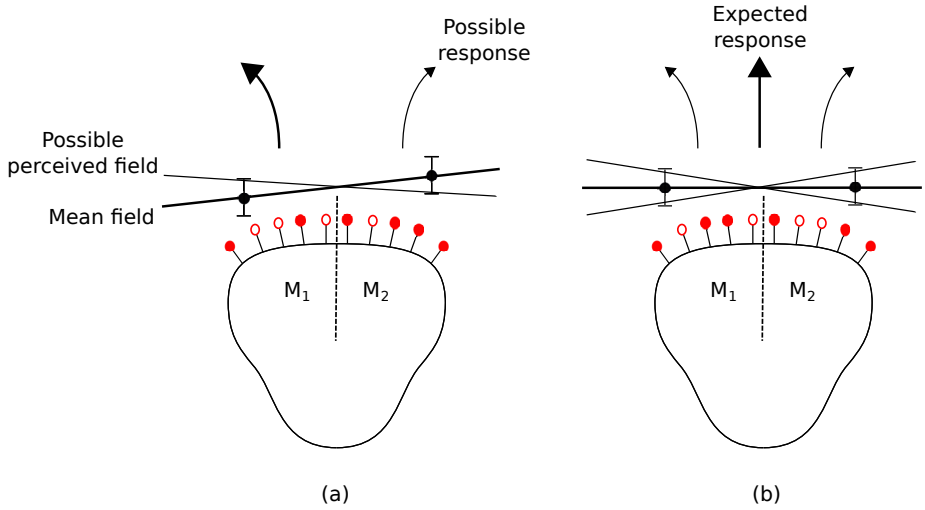

Figure 2.3: **Stochastic model for leukocyte random motility and chemotaxis.** Adapted from Tranquillo et al. [122] with permission*. (a) Chemotaxis: A cell is subject to a gradient of chemoattractant (mean field line). However, due to noise in the process of chemoattractant / receptor binding it is possible that the cell perceives the gradient to be operating in the contrary direction. Nevertheless, the mean direction perceived by the cell will be that of the true gradient. (b) Random motility: Similarly, in a uniform field the process noise means that the cell will perceive a gradient even though there is none present in fact. Globally the movements will be random. The model can be specified such that there is short term directional persistence which is in keeping with observations of neutrophil movements.

The models described above, and many more like them [60], aim to develop a mathematical description for individual neutrophil migration which is plausible and can give a good account of the types of behaviour that are observed experimentally. However, less work has been done on the reverse problem which is a thoroughgoing data-driven identification and estimation of migration models. Zigmond and Lauffenburger [134] discuss various ways in which cell migration data has been analysed. Their context is the evaluation of cells for migratory defects and the data is assumed to be collected *in vitro*. However, many current experimental studies still use

*©1988 Rockefeller University Press. Originally published in Journal of Cell Biology. 106:303-309. doi:10.1083/jcb.106.2.303

similar approaches. Typically, the analysis is based on comparing speed of movement, directional orientation or the proportion of cells responding to a directional cue. Common approaches include, counting the number of cells that migrate beyond a chosen threshold distance, measuring the distance travelled by a small group of leading cells, and determining the total number of cells migrating in a particular direction. Approaches that apply particularly to the evaluation of chemotaxis include the McCutcheon index (net direction travelled in the direction of chemotactic bias divided by total path distance), and the orientation bias (the proportion of cells moving in the direction of bias, at any point in time). These approaches are sensitive to the sampling time, since the full tortuosity of a cell path may be masked if the sampling time is too short. Whilst Zigmond and Lauffenburger were writing some time ago, the methods used in many experimental studies is still at this level.

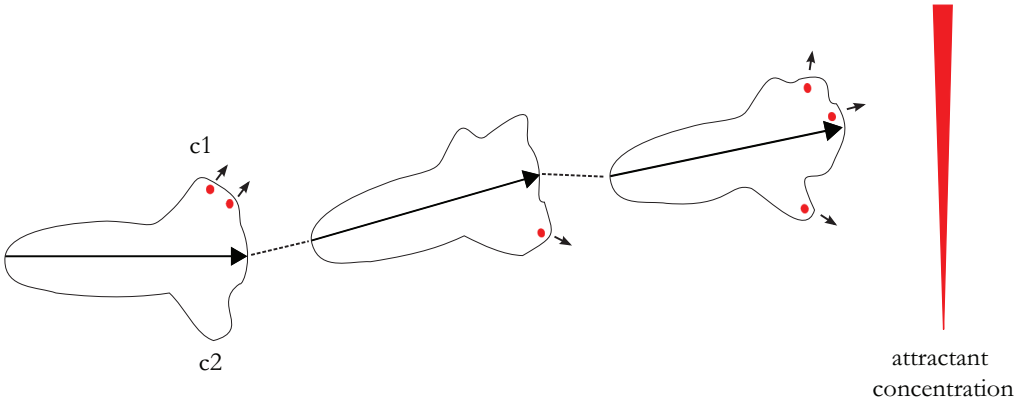

Figure 2.4: **Local coupling model.** Adapted from Arrieumerloua and Meyer [5]*.

2.2.2 Cell population models for chemotaxis

The idea of Chemotaxis has already been introduced in Section 2.1.5: it is the process whereby a cells movement is aligned to the gradient of a surrounding chemical field [136]. Two of the principle results of this thesis are the quantification of neutrophil chemotaxis during inflammation recruitment and the determination of whether neutrophils chemotaxis occurs during inflammation resolution. The analysis frameworks developed to do this will, however, have wider applicability because cell chemotaxis is a ubiquitous and fundamental process in living organisms [51]. Chemotaxis guides the sperm to the egg during fertilisation; it directs the positioning of cells in embryogenesis during gastrulation (an early folding process) and nervous system patterning; it encourages the formation of new blood vessels during angiogenesis; less happily it is implicated in the formation of cancerous cells into tumours. Chemotaxis is utilised

*With permission from Elsevier

for navigation by bacteria, and this has been extensively studied in the case of *Escherichia coli*, similarly in the slime mould *Dictyostelium discoideum*. It also plays a role in navigation by multi-cellular organisms such nematodes, fruit flies and moths.

Patlak introduced the idea of persistence, or directional correlation in Random Walks [26]. He was also the first to make a mathematical study of chemotaxis [100]. This was followed up by the well known work of Keller and Segel [68] yielding the model named after them. The Keller-Segel model has formed the foundation for much of the study of chemotaxis. It is a model that yields particular insights when the chemotaxing cells contribute to the directing field by producing chemoattractant chemical themselves. This auto-attractant feedback mechanism leads to pattern formation which is vital in many contexts mentioned above.

A simplified version of the Keller-Segel Model may be stated as follows [51]:

$$u_t = \nabla(D\nabla u - \chi u \nabla v) \tag{2.11}$$

$$v_t = \nabla^2 v + u - \gamma v \tag{2.12}$$

where u_t is the current cell concentration and v_t the concentration of attractant; D is the diffusion coefficient and ∇ is the gradient operator.

Equations (2.11) and (2.12) describe a cell population that is diffusing with constant rate D and with a drift proportional to the gradient of the attractant. The proportionality is determined by the chemotactic sensitivity, χ. The attractant diffuses (rapidly) and is produced by the cells themselves. It also decays at a constant rate, γ. In Random Walk Theory [26] this is referred to as a Reinforced Random Walk.

The Keller Segel equations are relatively simple in form as well as being analytically and numerically tractable — certainly in comparison to individual cell based approaches. They are able to capture the key characteristics of chemotaxis and have been applied with modifications to yield insight in several contexts, including those mentioned in the first paragraph of this section.

However, as described in the review by Hillen and Painter [51], there are a number of things about this model that are potentially unrealistic:

- There may be a saturating level of attractant above which a gradient can no longer be sensed.
- The cells may have limited velocity range which requires a non-linear relation between drift and gradient.
- Various aspects of the system (diffusion, chemotactic sensitivity) may be dependent on either or both of the density of cells and density of attractant.

The basic model can be adapted to address these issues. For instance, replacing (2.11) with,

$$u_t = \nabla(D\nabla u - \chi u \mathbf{F}(\nabla v)) \tag{2.13}$$

allows for a non-linear relationship between cell velocity and attractant gradient with a suitable choice of $\mathbf{F}$. For instance, with a sigmoidal $\mathbf{F}$ of the form,

$$F(s) = \frac{1}{1+e^{-s}} \tag{2.14}$$

the relationship is approximately linear in a central range but the velocity is effectively constant (and maximal) above a certain threshold attractant gradient. This prevents the physically unrealistic possibility of unbounded cell velocities.

Similarly, replacing (2.11) with,

$$u_t = \nabla\left(D\nabla u - \frac{\chi u}{(1+\alpha v)^2}\nabla v\right) \tag{2.15}$$

creates a model where the chemotactic sensitivity is reduced to zero when the attractant concentration is at a saturating level. The basic model is recovered when the constant $\alpha = 0$.

The Keller-Segel PDEs remain a cornerstone of current chemotaxis modelling and analysis. The are many studies in the literature developing variations of the basic model and investigating their resulting mathematical properties (e.g. [24, 41]) and the Keller-Segel approach also underpins models used in other methods such as agent based models and hybrid agent-PDE models [46].

In the analysis used in this thesis a simplifying assumption is made that the bias velocity induced by the chemoattractant is constant (and thus it is not necessary to model the dynamics of the attractant). In this case one can recover from (2.11) the drift-diffusion equation for species u,

$$u_t = \nabla\left(D\nabla u - bu\right) \tag{2.16}$$

where b is a constant encapsulating the gradient of the attractant which is assumed constant (see Appendix C).

It may be pointed out that the tissue within which neutrophils move *in vivo* is complex and porous rather than being a simple and completely homogeneous media. However, it has been shown by Nicholson [93] in relation to brain tissue that the dynamic coefficients are only affected by a scaling factor which takes into account the porosity and effectively averages the coefficients over a unit area. For instance,

the usual diffusion coefficient, D, is replaced by say $\hat{D}$ where,

$$D = \lambda^2 \hat{D}$$

with $\lambda^2 \geqslant 1$ being a measure of the tortuosity of the media.

2.3 System identification techniques

This thesis develops models, and associated system identification techniques, for cell migration. The models can describe a wide range of behaviours depending on the choice of parameters within the model. System identification is a multifaceted process. The key components, however, may be summarised following Ljung [77]:

- Experimental design for the collection of suitable data.
- Selection of candidate models.
- Data driven model choice.
- Validation of the chosen model.

These steps are generally used in an iterative loop, using the results and shortcomings of one cycle to inform the tuning of choices made for the next cycle.

Ideally, an experiment will be designed in such a way that in response to a sequence of known inputs the full dynamic range of the system being investigated is excited. If only a part of the dynamic range is explored, then only this aspect of the system can be identified since the data contains no information about other characteristics. Predictions of future behaviour may then fail if a different set of inputs is realised. Whilst such idealised experimental design may be achievable in many cases, especially in engineered systems, it will not be so in many biological systems which are an increasing area of research for engineers and systems analysts. The data analysed in this thesis is collected from cell observations in a living organism. The normal cell migration patterns are disrupted or excited by surgical wounding of a fish embryo. This is an 'input' to the system but clearly not one that can be turned on and off repeatedly or in varying degrees to elicit differing responses. Thus in this class of cases, system design, whilst still a vital consideration, has a limited extent.

As described at the end of Section 2.2, the initial candidate model used in this thesis is the drift-diffusion model which is a special case of the Keller-Segel equations. It will be considered in its general form and in the special case where the drift coefficient is zero. The latter will be called the pure-diffusion model. In Chapter

5 the shortcomings of these two models in relation to the experimental data will be used to propose and estimate enhanced models which better describe the full observed dynamics of the system.

Validation is a key part of the system identification process. The model identified may be the best available but the validity question still remains. It may take the form, 'Does the model give an accurate account of the data?', or, 'Is the model fit for the purpose in hand?'. Ljung provides a number of validation methods but admits that an amount of subjectivity must come into play. The standard methods [77] that will be employed in this thesis are,

- Comparison of estimated parameters to *a priori* knowledge.
- Evaluation of the physical plausibility of the parameters.
- Comparison of model simulations to experimental data.

When the choices of experimental design, candidate models and validation methods have been made, it remains to carry out the central task of system identification which is the estimation of model parameters and the identification of the preferred model. It is the comparison of model predictions to experimental data measurements via a suitable comparison criterion that make this a data-driven process. Two commonly used approaches to parameter estimation are represented by the following two questions [6]:

1. Which parameter set minimises the prediction error?
2. Which parameter set is most likely to have generated the observation set?

The following subsections will outline some methods that address these questions.

2.3.1 Least squares methods

A ubiquitous method for determining which parameter set minimises the prediction error, is to use the method of least squares (LS) [71, 92]. This minimisation requires an assumption about how the size of errors is to be measured and the name 'least squares' arises from the choice of using the standard Euclidean distance metric to measure the errors. Consider a system described by the following noisy linear model,

$$y_i = \theta^\top x_i + v_i \tag{2.17}$$

and suppose that, for realisations of the system $i = 1, \ldots, N$, the value of the (possibly vector) explanatory variable x_i and response variable y_i are known but the

random disturbance v_i is not known. The intention is to determine that value of the unknown parameter vector θ.

For any given choice of θ there is an associated prediction of the response,

$$\hat{y}_i = \theta^\top x_i \tag{2.18}$$

and thus an error between prediction and observation of

$$\epsilon_i = y_i - \hat{y}_i. \tag{2.19}$$

The ordinary least squares (OLS) estimate of θ is the value of θ which minimises the sum of the squares of all the errors, i.e.

$$\theta_{\text{LS}} = \underset{\theta}{\text{argmin}} \sum_{i=1}^{N} \epsilon_i^\top \epsilon_i. \tag{2.20}$$

Using (2.18) and (2.19), this can be written in the form of a cost function which must be minimised with respect to θ,

$$J(\theta) = \sum_{i=1}^{N} (y_i - \theta x_i)^\top (y_i - \theta x_i) \tag{2.21}$$

The right hand side of this cost function can be expanded, rearranged and differentiated to show that it is minimised when

$$\theta = \theta_{\text{LS}} = \left(\sum_{i=1}^{N} x_i x_i^\top \right)^{-1} \left(\sum_{i=1}^{N} y_i x_i^\top \right) \tag{2.22}$$

And this in turn can be written in matrix form as the well known normal equations [72],

$$\theta_{\text{LS}} = (X^\top X)^{-1} X^\top \mathbf{y} \tag{2.23}$$

where the element X_{ij} is the j^{th} element of x_i and $\mathbf{y}$ is the vector whose i^{th} component is y_i. Clearly a similar method could proceed with an alternative cost function. However, the terms within the summation in (2.21) must clearly all be positive and squaring has the advantage over taking the modulus, for instance, as the result is much more algebraically tractable and, in particular, amenable to differentiation.

The OLS method was published by the mathematician CF Gauss in 1809 and in 1822 he showed via the Gauss-Markov theorem that the estimate arrived at is the optimal unbiased estimate of the parameters when the model is linear and the errors, ϵ_i, have zero mean and are uncorrelated with identical variances. In the more

general situation where the variances of the error are unequal and have some degree of correlation OLS is no longer theoretically justified. However, generalised least squares (GLS) [66] deals with the heteroscedasticity and correlation via the error correlation matrix, V, to provide an unbiased estimate of the parameters. It finds θ which minimises the cost function,

$$\theta_{\mathrm{LS}} = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{N} \epsilon^{\top} V^{-1} \epsilon, \tag{2.24}$$

and the GLS estimate is reached via,

$$\theta_{\mathrm{GLS}} = (X^{\top} V^{-1} X)^{-1} X^{\top} V^{-1} Y \tag{2.25}$$

If the correlation matrix, V, is unknown it can be estimated along with the parameter vector, θ_{GLS}, in an iterative scheme.

A GLS method has been applied to identify cell movement indices from tracking data [33]. A special case of GLS is weighted least squares (WLS) [92], where all off-diagonal entries in the correlation matrix, V are zero. This essentially applies importance weights to the different observation instances. This approach will be discussed and used in Chapter 3.

In some applications of LS it may be unclear how many parameters are necessary to the explain the data being modelled. In this case the alternative method known as regularised least squares (RLS) [14] can be employed. Rather that simply minimising the squared error, which could lead to overfitting* the noisy data if the dimension of the parameter vector is unnecessarily high, a trade off is introduced between the size of the errors and the size of the elements of the parameter vector themselves. The latter has the effect of suppressing the number of effective dimensions of the parameter set, keeping those which have most explanatory power. One possible way of doing this is to replace the cost function in (2.21) with,

$$J(\theta) = \sum_{i=1}^{N} (y_i - \theta x_i)^{\top} (y_i - \theta x_i) + \lambda \theta^{\top} \theta \tag{2.26}$$

where λ is a user chosen parameter which controls the trade off described above. Increasing λ increases the preference for small parameter values. Because this cost function remains quadratic in θ it still has a closed form solution,

$$\theta_{\mathrm{RLS}} = (\lambda I + X^{\top} X)^{-1} X^{\top} y \tag{2.27}$$

*Overfitting of the data occurs when the size of observation noise is underestimated with the result that the identified function has much higher frequencies than the true underlying process.

analogous to (2.23). A more stringent method for ensuring sparsity of parameter vectors is to use the L1 norm of θ instead of the quadratic L2 norm. In this case we have

$$J(\theta) = \sum_{i=1}^{N}(y_i - \theta x_i)^\top (y_i - \theta x_i) + \lambda \sum_{j=1}^{d} |\theta_j| \tag{2.28}$$

where d is the size of the vector θ and the θ_j are its elements. Increasing λ in this case has the effect of driving increasing numbers of the elements of θ to zero. However, this more direct control of dimensionality comes at the cost of losing the analytic tractability of the quadratic form.

2.3.2 Maximum likelihood approaches

Maximum likelihood (ML) estimation [14] seeks to answer question 2 above (page 22): of all possible parameter sets, which one has the highest probability of producing the system observations, i.e.

$$\theta_{\mathrm{ML}} = \underset{\theta}{\mathrm{argmax}} \mathrm{P}(Y \mid \theta) \tag{2.29}$$

where $\mathrm{P}(Y \mid \theta)$, is the probability of Y given the parameter set θ. If equations (2.17) and (2.18) above are again considered, with the added assumption now that the v_i have identical normal distributions with variance Σ_v, then

$$\mathrm{P}(Y \mid X; \theta) = \prod_{i=1}^{N} \mathrm{P}(y_i \mid x_i; \theta) \tag{2.30}$$

where

$$\mathrm{P}(y_i \mid x_i; \theta) \sim \mathcal{N}\Big(\theta^\top x_i,\ \Sigma_v\Big) \tag{2.31}$$

Because the logarithm function is monotonically increasing, minimising a function is equivalent to minimising its logarithm. Logarithms can be taken, therefore, to simplify the manipulation of (2.31). In particular the logarithm of a product become the sum of logarithms, and the Gaussian form is reduced to a constant and a product of errors and their covariance matrix,

$$\log \mathrm{P}(Y \mid X; \theta) = \sum_{i=1}^{N} \{\log A + (y_i - \theta^\top x)^\top \Sigma_v^{-1} (y_i - \theta^\top x)\} \tag{2.32}$$

Algebraic manipulation shows that the minimising value, θ_{ML}, is achieved once again when,

$$\theta = \theta_{\mathrm{LS}} = \left(\sum_{i=1}^{N} x_i x_i^{\top}\right)^{-1} \left(\sum_{i=1}^{N} y_i x_i^{\top}\right) \tag{2.33}$$

The results presented here for OLS and ML illustrate that in the special case of independent and identical normal distribution of the errors, the two approaches yield the same estimate for the model parameters. This is not generally true, however.

The underlying difference between OLS and ML is illustrated in a simple way in Figure 2.5. The prediction (or expectation) of model A is 4, whilst that of model B is 5. Therefore, if the observed output is 4, a LS approach selects model A as the correct (error minimising) model as it produces an error of zero. ML, on the other hand looks at the probability of observing 4 under the assumption of each model. For model A, this is 20% for model B, 25%. Therefore ML chooses model B as the correct (most likely) model. A preference for the ML choice can be further justified by the following thought experiment. Both models are simulated 100 times. One would expect the outcome 4 to be observed 20 times from model A and 25 times from model B. A random selection from these 45 results is thus 25% more likely to have arisen from model B than model A.

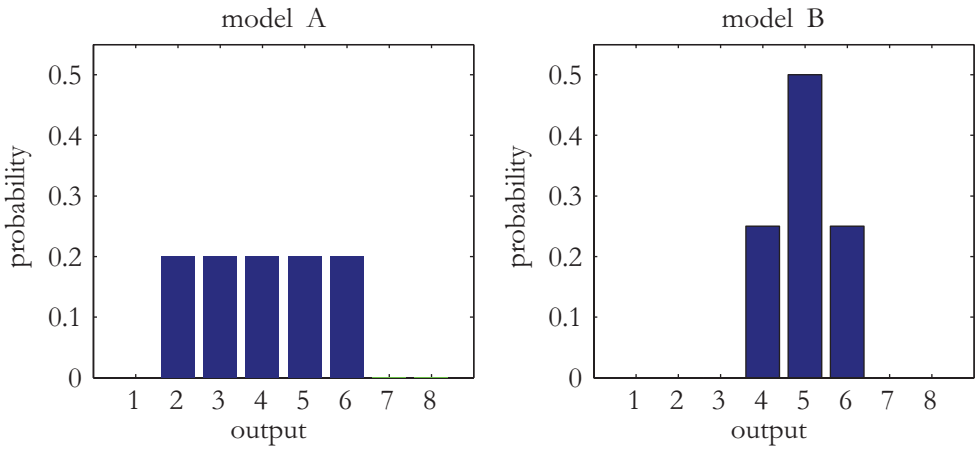

Figure 2.5: **Comparison of Least Squares and Maximum Likelihood**. Illustration of two discrete probability distributions where LS estimation and ML estimation differ in their conclusions.

A known problem with both LS and ML approaches, when errors are modelled using the normal distribution is their sensitivity to outliers, i.e. points which are exceptionally unlikely or points that are not rightly from the process under investigation. [47]. Outliers tend to make both LS and ML methods inefficient and bias the results away from the true underlying model. Also, the LS and ML methods arrive

at a point estimate of the parameter values. Providing such a point estimate can obscure the fact that there may be a range of parameter values that are also strong candidates for consideration. Both these issues can be addressed by the Bayesian methods which are discussed next.

2.3.3 Bayesian methods

The Bayesian method [47] is a general, thoroughgoing and flexible method for dealing with inferences and the associated uncertainties. It is a process whereby a generative probability model is chosen to describe the process behind an observed dataset. The estimation result produced by an application of the Bayesian method is expressed as a probability distribution on the parameters of the model and on any further unobserved quantities that may be of interest (e.g. predicted future observations). It is often contrasted to the frequentist approach which preceded it and now coexists with it. In brief, 'frequentists' consider the parameters underlying a process to be fixed and therefore amenable to identification via repeated sampling and null hypothesis testing or error minimisation. 'Bayesians', in contrast, model the parameters themselves as random variables. As such they are invested with a prior distribution (prior, that is, to receiving any evidence about them, or any new evidence from the current investigation). This prior distribution, often referred to simply as the prior, may be based on previous data-driven inferences about the parameters, or on expert knowledge about the parameters, or on a combination of these and other assumptions. The prior may be quite broad or 'uninformative', or it may be relatively tightly defined and therefore 'informative' about the likely range in which the parameters will be found to lie. An informative prior will exert more of an influence, or bias, on the subsequent inference results. At the heart of any Bayesian method is the theorem of Revd T Bayes [9]. For any two events A and B,

$$\mathrm{P}(A \mid B) = \frac{\mathrm{P}(B \mid A)\mathrm{P}(A)}{\mathrm{P}(B)} \tag{2.34}$$

where $\mathrm{P}(A \mid B)$ is the probability of event A occurring given that event B is known to have occurred.

Equation (2.34) can be re-written in notation more commonly used in Bayesian inference:

$$\mathrm{P}(\theta \mid y) = \frac{l(y|\theta)\mathrm{P}(\theta)}{\mathrm{P}(y)} \tag{2.35}$$

where θ is the parameter set for the model being used to describe the process which gives rise to observations y. So, $\mathrm{P}(\theta \mid y)$ is the desired posterior distribution of θ given the new evidence y. $l(y|\theta)$ is just another way of writing $\mathrm{P}(y \mid \theta)$ which is referred to

as the likelihood function – the likelihood of the observed evidence given a particular choice of θ. $\mathrm{P}(\theta)$ is the prior distribution of the parameters and $\mathrm{P}(y)$ is the probability of observing the evidence regardless of the underlying model for the process.

In practice $\mathrm{P}(y)$ may be unknown and impossible to calculate. However, since $\mathrm{P}(\theta \mid y)$ is a probability distribution it must be normalised. Hence it is sufficient to know that

$$\mathrm{P}(\theta \mid y) \propto l(y|\theta)\mathrm{P}(\theta) \tag{2.36}$$

and thus the evaluation of the likelihood function is the means by which the prior is transformed into the posterior.

Since the Bayesian method returns a distribution over the parameters of a model, and potentially also a distribution over candidate models as shall be seen in Chapter 4, it has distinct advantages over the LS and ML approaches which provide only a point estimate of the parameters. With a distribution over the parameters it is possible to integrate over the possibilities to get an average prediction for future of outcomes the system and a full sense of the uncertainties that pertain to it. However, if a point estimate is required this can be taken to be the parameter set corresponding to the maximum or modal value of the posterior distribution. That is why the Bayesian method is also referred to as the maximum *a posteriori* (MAP) approach.

The Bayesian and LS approaches are linked by a special case: If the Bayesian parameters are given a zero mean Gaussian prior then the solution reached is the same as that for regularised least squares using the quadratic L2 norm. This can be seen with reference to (2.26) by noting that

$$e^{-\frac{1}{2}(y_i-\theta x_i)^\top(y_i-\theta x_i)-\frac{1}{2}\lambda\theta^\top\theta} = e^{-\frac{1}{2}(y_i-\theta x_i)^\top(y_i-\theta x_i)}e^{-\frac{1}{2}\lambda\theta^\top\theta} \tag{2.37}$$

which is equivalent to (2.36). From this it can be seen that the Bayesian method has an implicit tendency towards dimensionality reduction.

2.4 Approximate Bayesian computation

There are many systems that can be studied in the realm of the life sciences and they tend to be very complex. As a result, for any given one of these systems there may be a multitude of models proposed to describe it. A model may be quite easy both to encode and to simulate, and it can be very tempting to compare model simulation output with observations of the real system and say that the agreement between the too is quite good, very good, or even astounding [91]. Of course judgements such those, though apparently confident, are subjective and vague at the best of times.

For instance, Figure 2.6 shows two plots reproduced from a well cited paper [33] where it is noted that the simulated tracks (a) and the experimentally observed tracks (b) "show similar track characteristics, that is the directional persistence over short times, and the meandering path characteristic of a random walk over longer time". In fact, the short term directional persistence is very different between the two cases. The simulated tracks are much smoother than the experimental tracks, having a significantly different distribution of turn angles. For this reason the persistence times for the two cases will in fact be very different. This highlights the fact that comparisons between simulation results and experimental results can easily be done in a vague way. It is possible however, to be principled and make explicit the way in which comparisons will objectively made.

The approximate Bayesian computation (ABC) [10] approach to estimation is a framework which allows and indeed relies on such a principled comparison of simulated and experimental results. In its most basic form it involves the choice of a model for which parameter values are simulated from a prior distribution. The model is then simulated and the output compared with the experimental observation using a chosen metric. If the distance between the two is zero (for discrete values) or is below some threshold (for continuously distributed values) the parameter set is accepted as being a sample from the posterior distribution. The ABC method thus carries with it all the advantages of the Bayesian paradigm whilst allowing estimation to be carried out without the evaluation of a likelihood function.

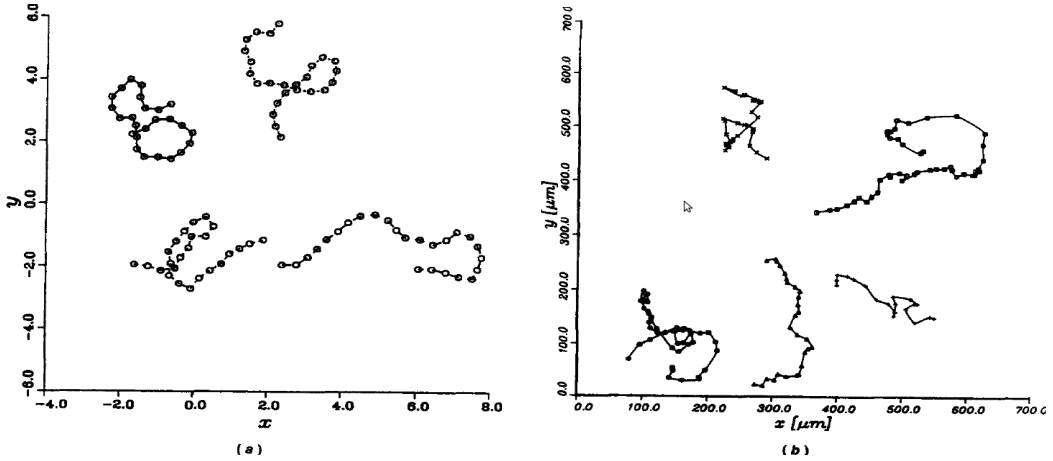

Figure 2.6: **Comparison of experimental mouse melanoma cells tracks to simulated tracks.** Reproduced from [33]*. (a) Simulated cell tracks. (b) Observed mouse melanoma cell tracks.

ABC methods have come to prominence over the last 10–15 years [10]. The recent

*With permission of John Wiley & Sons Inc.

proliferation of studies using the approach can be traced back to four formative papers in the domain of population genetics [11, 82, 103, 120]. In this research field, particular genetic markers are observed which may be linked in genealogical trees along a multitude of possible paths. This makes likehood evaluation approaches impractical.

2.4.1 The basic ABC rejection sampler

Though ABC methods started to gain popularity some 15 years ago, a similar approach was, in fact, present some 13 years earlier in an algorithm proposed by Rubin [110]. This is the basis of the most simple form of ABC rejection sampler. In this case the data are discrete and of low dimension such that the posterior can be sampled without approximation by repeated application of the following [10]:

1. Sample a process parameter vector, $\theta_i \sim \pi(\theta)$ from the prior.
2. Simulate observations of the process, $x_i \sim \mathrm{P}(x \,|\, \theta_i)$.
3. Reject θ_i if $x_i \neq y$, where y is the observation set.

The accepted set $\{\theta_i\}_{i=1}^{N}$ are a sampling representation of $\mathrm{P}(\theta \,|\, y)$ which is the probability distribution of the parameters that may have produced the observed data (assuming of course that the assumed model of the process is correct). The simplest way to portray this sampled distribution is as a histogram (for the cases of one or two dimensional parameter sets). Alternatively, kernel density methods [113] can be used to yield a smoothed version of the distribution which can be used to estimate the mode of the distribution, in situations where a point estimate of ML is required. In addition, either the discrete or smoothed version of the posterior can be used to compute a prediction of any quantity which may arise from future realisations of the process, i.e.

$$\mathrm{P}(f(x)) = \int \mathrm{P}(f(x) \,|\, \theta)\pi(\theta)\,\mathrm{d}\theta \tag{2.38}$$

where f is the quantity of interest and the integral is a summation if θ only takes discrete values.

In many applications, one or all of the observables may be a continuous quantity. In this case, $\mathrm{P}(x_i = y) = 0$. To enable the method to work in this situation the first approximation has to be introduced. This takes the form of an error threshold within which tolerance a parameter set will still be accepted. To be precise, an error threshold, ϵ, is chosen and step 3 above is replaced by:

3^a. Reject θ_i if $\rho(x_i(\theta_i) - y) > \epsilon$, where ρ is an appropriate distance metric.

Now the accepted set $\{\theta_i\}_{i=1}^N$ are a sampling representation of

$$\mathrm{P}(\theta \,|\, \rho(x(\theta) - y) < \epsilon). \tag{2.39}$$

The second main approximation in the ABC approach is the use of summary statistics. These are needed when the data are of high dimensionality and the resulting size of the parameter space effectively means that the target region in the parameter space, for which

$$\rho(x(\theta) - y) \leqslant \epsilon, \tag{2.40}$$

is impossible to find with any degree of efficiency. A summary statistic, $S(x)$, maps the observations into a space of lower dimensionality. Good summary statistics do this without significant loss of information. Ideally the summary statistics may be said to be sufficient, in which case no information is lost in mapping observation to observation summary and $\mathrm{P}(\theta \,|\, S(x)) = \mathrm{P}(\theta \,|\, x)$. With the inclusion of a summary statistic, step 3 above is replaced by:

3^b. Reject θ_i if $\rho(S(x_i(\theta_i)) - S(y)) > \epsilon$, where S is the chosen summary statistic.

Now the accepted set $\{\theta_i\}_{i=1}^N$ are a sampling representation of

$$\mathrm{P}(\theta \,|\, \rho(S(x(\theta)) - S(y)) < \epsilon)$$

with the distance metric, ρ, and error threshold, ϵ, being suitably chosen for the new context.

The use of summary statistics has been a point of criticism of ABC approaches, particularly when they have been adopted in an *ad hoc* manner. Over recent years there have been attempts to bring stronger principles to bear on their selection [64] and even automation [42]. However, an important part of any ABC implementation is the validation of the approach and any particular implementation decisions made, within the context of the specific application area. But as a general rule, the number of independent dimensions in the summary statistic needs to match the number of uncorrelated parameters in the model [10].

The foregoing text has presented three versions of the basic ABC rejection sampler. Unfortunately, this algorithm can be very inefficient in practice. Ideally, it is desirable that the tolerance, ϵ, be as small as possible so that the approximation made in 3^a or 3^b above will be correspondingly as good as possible. However, the smaller ϵ is, the greater is the proportion of sampled parameter vectors that will be rejected. This is exacerbated if the dimensionality of the problem is increased. In practice ϵ may be chosen to ensure that the acceptance rate is $\frac{N}{M}$, where N is the number of accepted samples deemed sufficient and M is that number of samples from

the prior that can realistically be sampled and simulated from in the computation time available. To combat the inefficiency of the basic ABC approach, a number of authors have developed the basic ABC paradigm to counter the inefficiency problem. There are three main ways in which this has been done:

- post processing of the ABC rejection results using weighting and a form of local linear regression which allows a greater proportion of the samples to be retained after appropriate adjustment to compensate for the size of the simulated observation error;
- setting the rejection step within a Markov chain Monte Carlo (MCMC) framework so that after a suitable burn-in time a higher proportion of samples will be accepted; and,
- sequential reduction of error tolerances to produce an iterative scheme known as approximate Bayesian computation sequential Monte Carlo (ABC-SMC) where at each iteration there is a proposal distribution which evolves from the prior towards the target posterior.

These broad ABC approaches will now be described in more detail.

2.4.2 ABC with local linear regression

ABC with local linear regression was introduced by Beaumont et al. [11] where two additions are made to the basic ABC rejection sampler: smooth weighting of the accepted samples and adjustment of the samples using local linear regression. It is clear that in the basic sampler all samples that pass through the error test are accepted with equal weight. Beaumont et al. propose that those with smallest error have greatest weight. This is achieved using an Epachnikov kernel (see Figure 4.4) which has the advantage of automatically assigning zero weight to any samples outside the error tolerance, ϵ,

$$W_\epsilon(s) = \begin{cases} c\epsilon^{-1}(1-(\frac{s}{\epsilon})^2) & t \leqslant \epsilon, \\ 0 & t > \epsilon, \end{cases} \tag{2.41}$$

where c is a normalising constant. The subsequent adjustment step is based on the assumption that (in the single parameter case), the accepted samples, ϕ_i, from the target distribution can be described by

$$\phi_i = \alpha + (S(x_i) - S(y))\beta + \gamma_i \tag{2.42}$$

where $S(.)$ is again a summarising function of the simulated and observed observations, x_i and y; α and β are regression constants and the γ_i are uncorrelated residuals with zero mean and a common variance. If $S(x_i) = S(y)$ then ϕ_i is a sample from the exact posterior which has mean α. Under these assumptions it is possible to obtain LS estimates for α and β, and thereby to make the adjustment,

$$\phi_i^* = \phi_i - (S(x_i) - S(y))\hat{\beta} \tag{2.43}$$

This will yield a true sample from the posterior distribution if the conditions are met. If the linear model is considered too restrictive an assumption, local linear regression can be performed instead where only points within the neighbourhood of $S(y)$ are included in the regression step. These two approaches, weighting and adjustment, can both be implemented on their own or together and in theory allow a much larger proportion of simulated parameter sets to be included in the estimation of the posterior distribution (thus increasing efficiency) as well as improving accuracy of the estimation for a given error tolerance. This weighting and adjustment can also be included as extra enhancements in the two methods that are discussed below. Whilst the regression adjustment version of ABC improves on the efficiency of the basic ABC algorithms, it nevertheless suffers from the drawback of sampling only from the prior distribution. If the prior distribution is uninformative relative to the data then many of the parameter samples will inevitably be of little use in the estimation. Two following enhancements of ABC seek to overcome this problem.

2.4.3 ABC Markov chain Monte Carlo methods

Generic Monte Carlo methods use repeated random samples from a distribution to build up a picture of that distribution. One subset of these methods is Markov chain Monte Carlo (MCMC) methods which greatly increase their efficiency by using a property of stationary Markov chains. A stationary Markov chain can reach an equilibrium whereby the distribution of the state of the chain remains unchanged from one iteration of the chain to the next. For a given posterior distribution it is usually possible to construct a Markov chain that has this distribution as its equilibrium state. Thus, propagating the Markov chain, and discarding samples from the 'burn in' period when the process is settling to its equilibrium, yields samples from the posterior distribution. Taken in the order in which they arise, these samples are correlated. However, taken as a whole over many iterations, they approximate an uncorrelated sample from the target distribution.

As an example of a mainstream MCMC method, consider a brief outline of the Metropolis Hastings (MH) algorithm [25]. This relies on a Markov process, f say,

which maps the a current state onto the subsequent one,

$$f(x \to \hat{x}) \tag{2.44}$$

For this process to be stationary with equilibrium distribution $\mathrm{P}(x)$, a detailed balance condition must be fulfilled,

$$\mathrm{P}(x) f(x \to \hat{x}) = \mathrm{P}(\hat{x}) f(\hat{x} \to x) \tag{2.45}$$

This condition is satisfied when $f(x \to \hat{x})$ is comprised of a proposal step, g, and an acceptance step, A, i.e.

$$f(x \to \hat{x}) = g(x \to \hat{x}) A(x \to \hat{x}), \tag{2.46}$$

with

$$A(x \to \hat{x}) = \min\left(1, \frac{\mathrm{P}(\hat{x})}{\mathrm{P}(x)} \frac{g(\hat{x} \to x)}{g(x \to \hat{x})}\right). \tag{2.47}$$

Beginning from an initial state x_0, the MH algorithm proceeds inductively as follows,

1. given state x_t propose a new state $\hat{x}$ according to the proposal $g(x_t \to \hat{x})$;
2. sample $u \sim \mathcal{N}(0,\ 1)$
3. set $x_{t+1} = \begin{cases} \hat{x} & \text{if } u \leqslant A(x \to \hat{x}) \\ x_t & \text{if } u > A(x \to \hat{x}) \end{cases}$

It is obvious from the expression for $A(x \to \hat{x})$ in (2.47), that the MH algorithm requires evaluation of a likelihood function to evaluate the update probability. Approximate Bayesian computation Markov chain Monte Carlo (ABC-MCMC), however, bypasses this problem by setting the MH step within an ABC rejection setting. Essentially there are now two decision steps: the ABC acceptance or rejection decision based on comparing simulation results and the MH decision based on relative probabilities.

An ABC-MCMC algorithm based on that of Marjoram et al. [82] can be outlined as follows, where $\mathrm{P}(x \,|\, \theta)$ is a probability model of the process for realising observations, x, conditional on parameters θ; $S(.)$ yields a summary statistic of the observations; $\rho(.,.)$ is a distance measure; ϵ is an error tolerance; and $g(.|.)$ is a suitably chosen Markov process proposal step.

1. Sample $\theta^{(0)}$ from the parameter prior, $\pi(0)$.
2. At iteration $t \geqslant 1$

(a) Sample $\theta' \sim g(\theta|\theta^{(t-1)})$

(b) Simulate $x \sim \mathrm{P}(x \,|\, \theta')$

(c) If $\rho(S(x), S(y)) > \epsilon$, set $\theta^{(t)} = \theta^{(t-1)}$, otherwise

i. Sample $u \sim \mathcal{U}(0, 1)$

ii. If $u \leqslant \frac{\pi(\theta')}{\pi(\theta^{(t-1)})} \frac{g(\theta^{(t-1)}|\theta')}{g(\theta'|\theta^{(t-1)})}$, set $\theta^{(t)} = \theta'$, otherwise set $\theta^{(t)} = \theta^{(t-1)}$

In [82] it is shown that after convergence, samples will come from the ABC posterior,

$$\mathrm{P}(\theta \,|\, \rho(S(x(\theta)) - S(y)) < \epsilon)$$

and at this time the proposal distribution in line 2(a) will be dominated by this target distribution and hence increased efficiency should be guaranteed.

The main practical drawback to ABC-MCMC is the unfortunate fact that if the chain is started in a low density region of the posterior, or if it happens to make a move into such a region, it can be very hard to leave again. The result of such an occurrence is that the chain gets stuck for a very long time on a particular parameter value. The proposal kernel may yield points in the vicinity that would be accepted by the MH step, 2(c)ii, but they are highly unlikely to first get past the ABC rejection step in the first line of 2(c). In theory, over sufficiently long realisations of the chain, these biases would eventually even out, but sufficiently long runs may not be practically possible. Steps have been taken to counter this problem by taking an adaptive approach to the tolerance parameter [17, 105]. However, another shortfall of ABC-MCMC is that like all MCMC methods it does not lend itself to parallelisation. This is in contrast to the next method considered.

2.4.4 ABC SMC

In order to address the problems faced by ABC-MCMC but still gain efficiency over basic ABC, Sisson et al. [114] proposed an integration of the ABC approach into a sequential importance sampling (SIS) framework.

In importance sampling [81], the aim is to approximate a target distribution, $\mathrm{P}(x)$, which cannot be sampled from but which can be evaluated at any point in its support up to an unknown normalising constant, i.e. $\mathrm{P}(x) \propto \mathrm{Q}(x)$, where $\mathrm{Q}(x)$ is the function that can be evaluated. This is achieved by sampling from a carefully chosen proposal distribution, $\eta(.)$, and then weighting the resulting sample according to its importance. The weights are given by

$$w(x_i) = \frac{\mathrm{Q}(x_i)}{\eta(x_i)} \tag{2.48}$$

Putting (2.48) into words, a sample is given increased weight for being from a region of high probability density in the target distribution and for being hard to sample from the proposal distribution. In some cases (such as a high dimensional target distribution) it can be hard to arrive at a suitable proposal distribution. In these circumstances SIS enables the target to be reached in stages where at each stage an intermediate distribution is approximated.

The initial ABC-SMC proposal of Sisson et al. was subject to a bias. This was subsequently corrected and the resulting closely related variants of ABC-SMC [12, 115, 121] provide one the most often implemented versions of ABC algorithm. It can be summarised as follows, where $\mathrm{P}(x \mid \theta)$ is a probability model of the process for realising observations x conditional on parameters θ; $S(.)$ yields a summary statistic of the observations; $\rho(.,.)$ is a distance measure; $\epsilon_1, \ldots, \epsilon_T$ is a decreasing sequence of error tolerances; and $K(.|.)$ is a parameter perturbation kernel.

1. On the first iteration, $t = 1$,
   (a) For $i = 1, \ldots, N$
   (b) Simulate $\theta_i^{(1)}$ from the prior, $\pi(\theta)$, and $x \sim \mathrm{P}(x \mid \theta_i^{(1)})$ until $\rho(S(x), S(y)) < \epsilon_1$
   (c) set $\omega_i^{(1)} = \frac{1}{N}$
2. Set τ_2^2 appropriately, e.g. twice the variance of $\{\theta_i^{(1)}\}_{i=1}^N$.
3. For subsequent iterations, $t = 2, \ldots, T$,
   (a) For $i = 1, \ldots, N$, repeat
      i. Choose $\hat{\theta}_i$ from $\{\theta_i^{(t-1)}\}_{i=1}^N$ with probability $\omega_j^{(t-1)}$
      ii. sample $\theta_i^{(t)} \sim K(\theta|\hat{\theta}_i; \tau_t^2), \quad x \sim \mathrm{P}(x \mid \theta_i^{(t)})$

      until $\rho(S(x), S(y)) < \epsilon_t$
   (b) Set $\omega_i^{(t)} \propto \frac{\pi(\theta_i^{(t)})}{\sum_{j=1}^N \omega_j^{(t-1)} K(\theta_i^{(t)}|\theta_j^{(t-1)}; \tau_t^2}$
   (c) Set τ_{t+1}^2 as twice the weighted variance of $\{\theta_i^{(t)}\}_{i=1}^N$.

In order to justify the weights used in the ABC-SMC algorithm, Fearnhead and Prangle [42] note that the ABC likelihood is approximated by,

$$\mathrm{P}(S(y) \mid \theta) = \int \mathrm{P}(x \mid \theta) \chi_{[-\epsilon,\epsilon]}(D_{xy}) \, \mathrm{d}x \tag{2.49}$$

where $D_{xy} = \rho(S(x) - S(y))$ and χ is the indicator function, i.e.

$$\chi_{[-\epsilon,\epsilon]}(s) = \begin{cases} 1 & -\epsilon \leqslant s \leqslant \epsilon \\ 0 & \text{otherise} \end{cases} \tag{2.50}$$

This in turn means that, by applying Bayes' theorem, the ABC posterior can be defined as,

$$\pi_{ABC}(\theta|S(y)) \propto \pi(\theta)\mathrm{P}(S(y)\,|\,\theta) \tag{2.51}$$

$$\propto \pi(\theta) \int \mathrm{P}(x\,|\,\theta)\chi_{[-\epsilon,\epsilon]}(D_{xy})\,\mathrm{d}x \tag{2.52}$$

where $\pi(\theta)$ is the prior distribution over the parameters. The standard importance sampling weights for the posterior are therefore,

$$\frac{\pi(\theta)}{\eta(\theta)} \int \mathrm{P}(x\,|\,\theta)\chi_{[-\epsilon,\epsilon]}(D_{xy})\,\mathrm{d}x \tag{2.53}$$

where $\eta(\theta)$ is the current proposal density. In the Monte Carlo step at the heart of ABC x is simulated from $\mathrm{P}(x\,|\,\theta)$ and accepted θ with probability $\chi_{[-\epsilon,\epsilon]}(D_{xy})$. If the accepted samples are assigned the weight,

$$\frac{\pi(\theta)}{\eta(\theta)}, \tag{2.54}$$

the expected weight over all samples will be given by (2.53). In the case of ABC-SMC the proposal density changes on each iteration so that it more closely resembles the posterior. It is a kernel approximation of the last estimate of the posterior:

$$\eta(x) = \sum_{j=1}^{N} \omega_j^{(t-1)} K(x|\theta_j; \tau^2) \tag{2.55}$$

and hence we have the weights that are assigned above in 3(b). In Chapter 4 this version of ABC and its extension to include model selection will be tailored to the identification of zebrafish neutrophil dynamics during inflammation and inflammation resolution.

2.4.5 Applications of ABC

Since its inception in the domain of population genetics as already described, ABC methods have been applied to filtering [58] and smoothing [83] problems and to diverse fields from chemoattractant field estimation [75] through psychology [123] to

cosmology [22]. With its increasing popularity a number of software tools have been developed to facilitate the application of the ABC methods in population genetics [29, 79, 129] and systems biology [74].

Chapter 3

Regression analysis of neutrophil migration during inflammation recruitment and resolution

3.1 Overview

During inflammation neutrophils are known to be recruited to sites of injury and infection. Neutrophils must also eventually be removed from these sites for inflammation to resolve satisfactorily and homeostasis to be restored. Neutrophil clearance from a site of inflammation can be effected by apoptosis, which is programmed death of the cell *in situ*. Recent work in zebrafish has suggested that neutrophils may also migrate away from inflammatory sites. Up to now rigorous mathematical treatment of these recruitment and inflammation resolution phenomena has been lacking. The dynamics of neutrophil recruitment are unquantified and a key open question in neutrophil research concerns what signals, if any, regulate the process of reverse migration. This chapter seeks to provide an answer to these questions by applying system identification techniques to study the behaviour of neutrophils during recruitment and reverse migration. The aim is twofold: firstly, is to confirm that neutrophils *in vivo* are actively recruited to inflammation sites and to quantify this recruitment dynamic; secondly, to investigate whether they move away from inflamed sites in a similar manner or whether the inherent random component of their migration was enough to account for this behaviour.

In order to achieve these aims, targeted data was obtained from experimental collaborators. They used neutrophil-driven photo-convertible Kaede protein in transgenic zebrafish larvae to label neutrophils at an inflammation site and they recorded the locations of these neutrophils over time. This chapter describes how regression

methods were applied to this data by fitting two separate but related models derived from the pure-diffusion equation and from the drift-diffusion equation. These models correspond to undirected or purely stochastic migration and directed migration respectively.

3.2 Introduction

The fate of neutrophils following completion of an inflammatory event is of critical importance for the outcome of episodes of acute inflammation and can determine whether there is prompt healing of a wound or the development of chronic inflammation and tissue injury. The uncertainty as to the *in vivo* fates of individual neutrophils relates in part to the difficulty in following individual cells during inflammation resolution *in vivo*. However, the transgenic zebrafish model is emerging as a key model for the study of vertebrate immunity [108]. It allows direct imaging of individual cells, and of populations of cells *in vivo* throughout the onset and resolution of inflammation.

The most widely accepted mechanism of neutrophil removal is the programmed cell death, or apoptosis, of the neutrophil. Apoptic neutrophils are subsequently cleared from the system by macrophage cells which consume the dead neutrophils [36]. This process has been observed in the zebrafish [40, 80]. However, there is increasing evidence that in addition to apoptosis, migration of neutrophils away from inflammatory sites is a significant process contributing to inflammation resolution. Neutrophils may move away from the inflamed site into the bloodstream ("reverse transmigration" [21]), by migration through other tissues ("retrograde chemotaxis" or "reverse migration" [20, 49, 54, 85, 133]), or may be exuded out of the wound [43, 125]. Because this process cannot be easily visualised in mammalian systems, definitive proof of its importance is lacking, but considerable evidence points to its potential significance in mammalian biology [21, 131]. Certainly, it appears to be a key regulatory step in inflammation resolution in the zebrafish [39]. In order to understand how this reverse migration is regulated, it is necessary to better understand what this process represents physiologically.

A number of papers have reported that retrograde chemotaxis plays a role in the reverse migration process [55, 85, 86, 88, 118] and although there is very limited analysis of neutrophil migratory patterns in the literature, this is becoming widely assumed to be the case. The implication of this belief is that directional guidance is exerted on the neutrophils to move them away from the inflammation site in an analogous way to that which they experience when being recruited. However, all the papers that report retrograde chemotaxis refer back to the one paper where hard

evidence for this process is reported to be present [85]. In that study zebrafish larvae at 3 days post fertilisation (dpf) were wounded in the ventral fin to induce inflammation. Neutrophils expressing green fluorescent protein (GFP) were then observed over periods of 20 mins (10s/frame), 3 hours (30s/frame) or 4 hours (60s/frame). Individual neutrophils were tracked for as long as could be distinguished from other cells and velocity, directional persistence and a directionality index (net distance travelled / path length) were calculated. The conclusion that retrograde chemotaxis occurred was based upon the following:

- The majority (80%) of cells entering the wound were also observed to track back towards the vasculature.
- Velocity and directionality estimates were comparable for cells in both directions of migrations (Figure 3.1).
- The directionality estimates were high and therefore presumed not to be due to random motility.

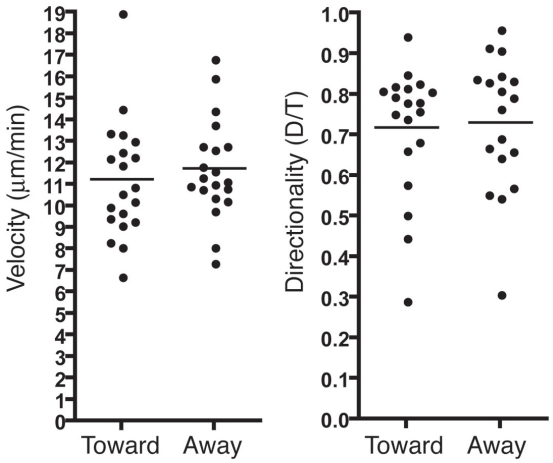

Figure 3.1: **Comparison of velocities and directional indicies for recruiting and resolving neutrophils** Reproduced from Mathias et al. [85]†. This represents the most rigorous analysis of zebrafish inflammatory neutrophils prior to the work contained in this thesis. The conclusion that the recruiting and resolving dynamics are of the same kind is drawn from the fact that the mean values are close even though the distribution of directionality indices are quite different. In any case the quantities evaluated are based on quite selective use of data.

Whilst [85] provides an important contribution to the field, it is concerning that a consensus should emerge based on a relative paucity of rigorous mathematical

†Republished with permission of The Society for Leukocyte Biology; permission conveyed through Copyright Clearance Center, Inc.

analysis. This thesis makes the novel contribution of analysing similar neutrophil data from a populational perspective. This bypasses any biases that would arise from the tracking or track selection process. Furthermore, because tracking is not required, data can be incorporated from all cell-position observations, thus gaining maximum information from the data. Observations can also be included over a long time period[‡] so that short time directional correlations can be ignored, which otherwise might be mistaken for direct motion.

3.3 Methods

3.3.1 Experimental methods

All experimental work was carried out by collaborators[§] from the Academic Unit of Respiratory Medicine, Department of Infection and Immunity, University of Sheffield, Sheffield, UK. All experiments were performed according to guidelines and legislation set out in UK law in the Animals (Scientific Procedures) Act 1986. Ethical approval was given by the University of Sheffield Local Ethical Review Panel. The *in vivo* zebrafish model for observing neutrophil inflammation and its resolution is derived from two transgenic lines: Tg(lyz:Gal4)i252 [39] and Tg(UAS:Kaede)s1999t [30]. These are described in their respective references. Briefly, the yeast transcription factor Gal4, fused to VP16 viral transcriptional activator sequence, recognises and drives transcription from the upstream activation sequence. The Gal4 sequence is inserted into a DNA vector 3 of a PCR generated promoter for Lysozyme C, which has almost complete overlap with the neutrophil-specific mpx promoter at early developmental stages. This construct is injected into fertilised eggs, allowing random incorporation into the genome and driving expression of Gal4 in neutrophils in subsequent generations. In parallel, the second transgenic line is generated in the same way, expressing the photo-convertible protein, Kaede, under an upstream activation sequence. Thus in the double transgenics, generated by crosses of the two single transgenics, Kaede is expressed in neutrophils. The zebrafish lines were maintained according to standard protocols [97].

Six zebrafish embryos from the double transgenic stock, at 3 days post fertilisation were subjected to tailfin transection under anaesthesia using a sterile scalpel. The embryos were allowed to recover. Each embryo was then mounted in 0.5% low melting point agarose and at four hours post injury the neutrophils then present within the tip of the tailfin were photoconverted so that they emitted red fluorescence. This

[‡]Tracking requires relatively fast data sampling and the ultraviolet exposure required for this has the effect of bleaching out the cells and damaging the health of the specimen. Both of these causes early experimental termination.

[§]Giles Dixon and Anne Robertson working under Stephen Renshaw.

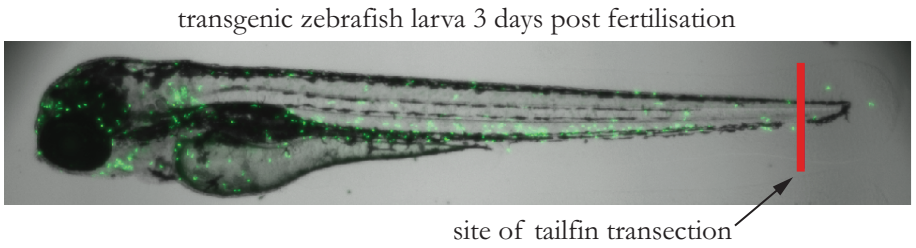

Figure 3.2: **Zebrafish embryo.** The neutrophils in the transparent double transgenic zebrafish embryo express a green fluorescent protein which facilitates visualisation of the neutrophils throughout inflammation and inflammation resolution. The approximate location for the tailfin transection is shown. The Kaede protein expressed by the neutrophils means that at the scheduled time after injury (4 hours in this case), neutrophils in the immediate vicinity of the tailfin wound can be photoconverted to show red fluorescence. Thus they are labelled as a distinct sub-population of cells and can be observed as they migrate around and away from the wound area.

was in contrast to the rest of the neutrophils present in the surrounding area which continued to emit green fluorescence. Thus it was possible to collect data for two distinct sub-populations of neutrophils. Photo-conversion of the Kaede protein was carried out using a Perkin Elmer UltraVIEW PhotoKinesis device, mounted on a Perkin Elmer UltraVIEW VoX ERS 6FR Laser Confocal Imaging System (Perkin Elmer INC, USA) with an inverted Olympus IX81 microscope, equipped with six diode laser lines and a Yokogawa CSU-X1 spinning disk. The device was calibrated using a glass microscope slide (Menzel-Gläzer) covered with fluorescent highlighter ink (Stabilo Boss) as a photo-bleachable substrate (according to manufacturers instructions). Photo-conversion was performed using 40% laser energy for 120 cycles of a 405 nm laser line.

The embryos were released from the agarose gel and transferred to fresh E3. The petri dishes containing the embryos were wrapped in tinfoil to prevent background photo-conversion. The embryos were then mounted on the microscope and the neutrophils were observed using a Nikon Eclipse TE2000-U Inverted Compound Fluorescence Microscope (Nikon UK Ltd) with a Hamamatsu 1394 ORCA-ERA camera (Hamamatsu Photonics Inc). During the time-lapse, separate images were taken using filter sets optimised for green fluorescence and for red fluorescence, in order to record the migration of both sub-populations of the neutrophils. Images were captured using Volocity (build 5.3.2, Perkin Elmer) every 300 seconds by taking widefield fluorescence Z-stacks as well as a background DIC[¶] image. Volocity was used for ini-

[¶]Differential interference contrast (DIC) is an enhanced imaging technique [89]. The DIC image is a confocal image of the specimen showing the tissues and cells as opposed to a fluorescent image of the neutrophils only. See Figures 3.3-3.5.

tial processing of the data to extract neutrophil centroid positions. These data were then exported to Matlab (MathWorks, MA) for the further analysis described below.

3.3.2 Dynamic Modelling of neutrophil behaviour

Random walk models are often used in biology to describe the movement dynamics of individuals and populations [26, 101] and particularly for cell migration patterns [2, 73, 102]. Over short time-scales neutrophils exhibit correlated random walk (CRW) behaviour. In a CRW the current migration direction usually changes only slightly from moment to moment. However, these short term correlations decay over time. The time between the observations in the experimental data analysed in this thesis is greater than typical neutrophil persistence times [122] and thus local correlations in direction can be ignored [101] and a simple random walk (SRW) or the variant biased random walk (BRW) model can be applied to the data. A random walk model for the individual cell has a corresponding Fokker-Plank equation which describes the resulting cell population dynamics. For the SRW and BRW this is the drift-diffusion equation:

$$\frac{\partial p(x,t)}{\partial t} = b\frac{\partial^2 p}{\partial x^2} + a\frac{\partial p}{\partial x} \tag{3.1}$$

where $p(x,t)$ is the distribution of cells (in this case) at the location described by coordinate x and at time t; b is the diffusivity coefficient and a the drift coefficient. The neutrophil locations were described by a single coordinate x which corresponds to their distance from the wound. The dynamics being considered relate to migrations and any biases to and from the wound, so the direction parallel to the wound is not relevant.

The diffusion equation, corresponding to a SRW, is a special case of the drift-diffusion equation where the drift coefficient, a is zero. The distinction will be made clear in this thesis by referring to the diffusion equation as the pure-diffusion equation. The dynamic models are derived from these related equations as follows.

Firstly, in order to model the photoconverted (red) cells which are initially in the vicinity of the wound and gradually migrate away, is was assumed that the wound in the fish can be mathematically characterised as a reflecting boundary and that the initial distribution of cells was a delta distribution at the origin. Equation (3.1) was solved using the method of reflections [27] to show that the system can be described by

$$p(x,t) = \frac{1}{\sqrt{4\pi bt}}\left(e^{\frac{-(x-at)^2}{4bt}} + e^{\frac{-(x+at)^2}{4bt}}\right) \tag{3.2}$$

Multiplying this by x^2 and then integrating over $x \in [0, \infty)$ yields:

$$\mathbb{E}\left(x(t)^2\right) = a^2t^2 + 2bt + k_1 \tag{3.3}$$

where k_1 is a constant determined by the initial positions of the cells. It should be noted that if the delta distribution assumption is relaxed, (3.2) will be a sum over the number of cells with two equivalent terms within this summation on the right hand side corresponding to a delta distribution on the initial position of each individual cell. However, the integration step will still yield result (3.3).

Equation (3.3) implies that $\mathbb{E}\left(x(t)\right)$ will have a quadratic form if $a > 0$ (when there is a bias present away from the wound) and a linear form if $a = 0$ (when there is no bias and the process is diffusion only).

Secondly, in order to model the non-photoconverted (green) cells which are initially at some distance away from the wound and still migrating inwards, it was assumed that the cells had an arbitrary distribution, which is far enough away from the boundary (i.e. the wound) that some time elapses before any cell reaches the boundary then, using the method described by Codling et al. [26], equation (3.1) is multiplied by x and integrated by parts to yield:

$$\mathbb{E}\left(x(t)\right) = at + k_2 \tag{3.4}$$

where k_2 is a constant determined by the initial positions of the cells. Equation (3.4) implies that the mean position of cells will be constant if there is no bias (diffusion only) but will move linearly with time in the direction of any bias.

3.3.3 Model identification: neutrophil recruitment

At each observation time the data provides a set values. Each value is the distance of an observed cell from the wound,

$$\mathcal{X}(t) = \{x_1(t), \ldots, x_{n_t}(t)\} \tag{3.5}$$

where n_t is the number of observed cells at time t. These values were averaged to produce an experimental value $\langle x(t) \rangle$ for each time t. This is the data derived version of the theoretical quantity $\mathbb{E}\left(x(t)\right)$ found in (3.4). $\langle x(t) \rangle$ is plotted against time in Figure 3.9. It can be seen that $\langle x(t) \rangle$ decreases linearly up to a certain switching time, say $t = t_S$. A suitable value of t_S was determined by visual inspection of the graph in each case.

In order to identify the bias parameter, a, closed form weighted least squares (WLS) estimation was applied [92]. The observation vector $\mathbf{z}$, the design matrix Φ

and the target parameter vector $\boldsymbol{\theta}$ were defined as follows,

$$\mathbf{z} = \begin{pmatrix} \langle x(t_1) \rangle & \dots & \langle x(t_S) \rangle \end{pmatrix}^\top \tag{3.6}$$

$$\Phi = \begin{pmatrix} t_1 & t_2 & \dots & t_S \\ 1 & 1 & \dots & 1 \end{pmatrix}^\top \tag{3.7}$$

$$\boldsymbol{\theta} = \begin{pmatrix} a & k_2 \end{pmatrix}^\top \tag{3.8}$$

It should be noted here that in keeping with the experimental observation rate one time unit $(t_k - t_{k-1})$ is 300 seconds. The model defined by (3.4) implies that:

$$\mathbf{z} = \Phi\boldsymbol{\theta} \tag{3.9}$$

In the case of the green cells, the variability in the number of visible cells was due to cells entering and leaving the domain as well as cells being indistinguishable, at times, from one another. The weighting vector was defined, therefore, in terms of the deviation from the mean number of visible cells, as follows:

$$W = \text{diag}([w_1 \dots w_S]) \tag{3.10}$$

$$w_k = 1 - \left(\frac{\langle n \rangle - n_k}{\langle n \rangle} \right)^2 \tag{3.11}$$

$$\langle n \rangle = \frac{1}{S} \sum_{k=1}^{S} n_k \tag{3.12}$$

Where diag(.) is the diagonal matrix constructed from its vector argument and n_k is the number of cells visible at time $t = t_k$. Then, according to the WLS scheme [92], the best estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is identified by:

$$\hat{\boldsymbol{\theta}} = (\Phi^\top W \Phi)^{-1} \Phi^\top W \mathbf{z} \tag{3.13}$$

and confidence in the identified value of u given the model, can be assessed via a standard deviation, $\boldsymbol{\sigma}$, which is calculated from the residuals as follows:

$$e(j) = \langle x(t_j) \rangle - (\hat{u}t_j + \hat{k}_2) \tag{3.14}$$

$$\mathbf{e} = (e(1) \dots e(S))^\top \tag{3.15}$$

$$\boldsymbol{\sigma} = \sqrt{\frac{\mathbf{e}^\top \mathbf{e}}{S - m} (\phi^\top \phi)^{-1} \phi^\top W^2 \phi (\phi^\top \phi)^{-1}} \tag{3.16}$$

where m is the number of parameters in the model. In this case $m = 2$.

3.3.4 Model identification: inflammation resolution

For the photo-converted (red) cells which are migrating away from the wound, the appropriate model is given by (3.3). This has two different forms, depending on whether or not the drift coefficient a is non-zero. In the $a = 0$ case, the design matrix is defined to be,

$$\Phi_0 = \begin{pmatrix} t_1 & t_2 & \dots & t_N \\ 1 & 1 & \dots & 1 \end{pmatrix}^\top \tag{3.17}$$

and if a can be non zero,

$$\Phi_1 = \begin{pmatrix} t_1^2 & t_2^2 & \dots & t_N^2 \\ t_1 & t_2 & \dots & t_N \\ 1 & 1 & \dots & 1 \end{pmatrix}^\top \tag{3.18}$$

In both cases, the observation vector z and parameter vector $\boldsymbol{\theta}$ are defined to be

$$\mathbf{z} = \begin{pmatrix} \langle x(1)^2 \rangle & \dots & \langle x(N)^2 \rangle \end{pmatrix}^\top$$

$$\boldsymbol{\theta} = \begin{pmatrix} 2b & a & k_1 \end{pmatrix}^\top$$

In this case the weights were defined as

$$W = \mathrm{diag}([n_1 \dots n_S])$$

This is because variability in the number of visible red cells is only caused by cells becoming obscured by proximity to each other. Very few if any cells leave the domain and no new cells can join this photo-converted sub-population. Proceeding analogously to the last section, allows calculation of

$$\hat{\boldsymbol{\theta}}_{\mathbf{2}} = \begin{pmatrix} 2\hat{b} & \hat{k}_1 \end{pmatrix}^\top$$

and

$$\hat{\boldsymbol{\theta}}_{\mathbf{3}} = \begin{pmatrix} \hat{a}^2 & 2\hat{b} & \hat{k}_1 \end{pmatrix}^\top$$

together with the associated confidence intervals.

An higher order model will always give a better fit to a data set. It will have a smaller associated error function

$$J(m) = \mathbf{e}^\top \mathbf{e}$$

(where m is the order of the model). The 'f' hypothesis test was applied, therefore, to evaluate the significance of the improvement. The higher order model is better with $1-\alpha$ certainty if $f > \chi^2_\alpha(1)$. Where $\chi^2_\cdot(.)$ is the Chi-Squared distribution and:

$$f = \frac{J(m-1) - J(m)}{J(m)}(N-m) \tag{3.19}$$

Put simply, this test says that if $f > 4$ the quadratic model should be chosen over the linear model [116].

3.4 Results

3.4.1 Experimental results

In six zebrafish embryos that had their tailfin transected, Kaede-expressing neutrophils were photoconverted in the vicinity of a wound at 4 hours post injury (Figure 3.3). These cells were initially in the range $0-100\,\mu m$ from the wound. The photo-conversion process effectively labelled two sub-populations of neutrophils. The evolving positions of both groups were observed at 5 minute intervals using time-lapse video-microscopy as described above. Cells were not individually tracked but their positions at each time point were observed in order to record the evolution of their populational distribution.

A representative sample of the raw images of the two resulting neutrophil sub-populations as they migrate over the duration of the experiment are shown in Figure 3.4 (photo-converted cells) and Figure 3.5 (non-converted cells). Plots of the extracted data for all neutrophils at photo-conversion and the evolving positions of the photo-converted cells are also shown in Figure 3.6. Throughout the course of the experiment the photoconverted red cells remained most densely clustered near the wound but their distribution widened over time as cells migrated away.

There were no green neutrophils observed at the site of injury at 4 hours post injury (hpi) because all the cells present had been photoconverted. However, there is a subsequent accumulation of green neutrophils at the site of injury from 6 hpi until 14 hpi. This shows that, in zebrafish, neutrophils continue to be recruited to the inflammation site even whilst neutrophils recruited earlier are beginning to migrate away. This is a contrast to mammalian inflammation, where neutrophil influx ceases early in the inflammatory response, at least in rabbit models of pneumonia [62, 63]. However, the difference could be due to the increased sensitivity for detecting continued influx in the transparent zebrafish model with the Kaede protein labelling.

Figure 3.7 shows the extracted neutrophil position data represented by plotting, for every observation time, the distance of each observed neutrophil from the wound.

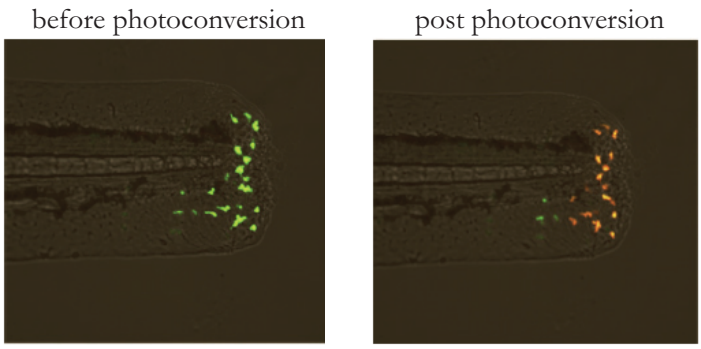

Figure 3.3: **Photoconversion of Kaede protein in neutrophils within the tailfin area.** Embryos at 3 days post fertilisation from transgenic zebrafish expressing Kaede in neutrophils were subjected to tailfin transection under anaesthesia using a sterile scalpel. The embryos were allowed to recover for 4 hours. At four hours after injury the embryo was mounted in 0.5% low melting point agarose for imaging on a Laser Confocal System (Perkin Elmer Inc). The PhotoKinesis device was first used to photoconvert all neutrophils present within the tip of the tailfin. Photoconversion was carried out according to the methods described (120 cycles of 40% 405 nm laser energy). The composite images of DIC (see footnote on page 43) overlaid with the red and green fluorescence channels show a representative zebrafish tail with the neutrophils in the vicinity of the wound before and after photoconversion. All experimental work was carried out by collaborators from the Academic Unit of Respiratory Medicine, Department of Infection and Immunity, University of Sheffield, Sheffield, UK.

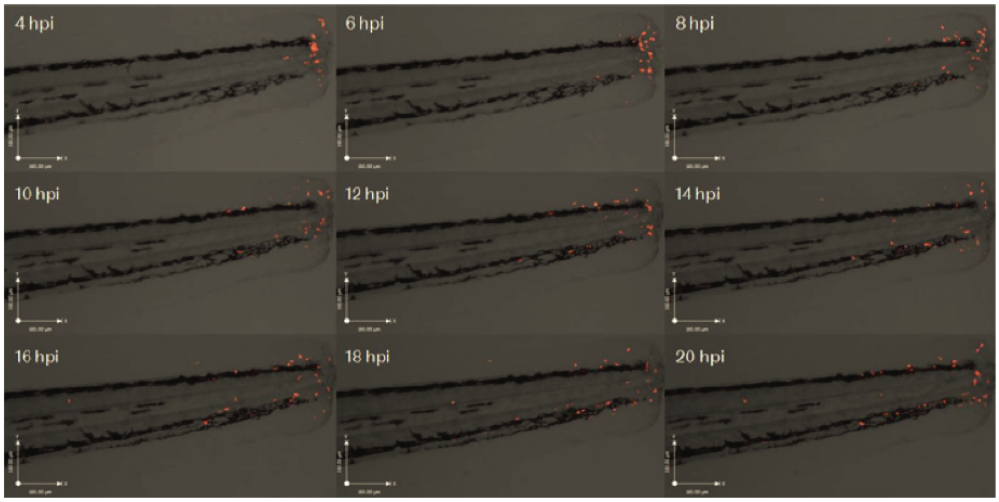

Figure 3.4: **Inflammatory neutrophils in zebrafish migrate away from the site of tissue injury.** A montage of DIC images overlaid with the red fluorescence channel at the timepoints indicated after tailfin injury (hpi is hours post injury). The redistribution of photoconverted cells can be clearly seen over time.

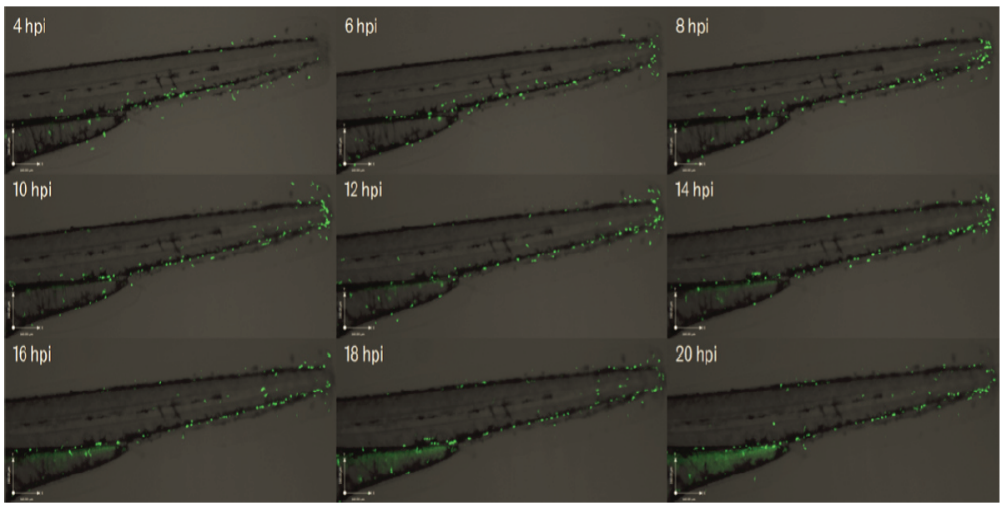

Figure 3.5: **At peak inflammation, further neutrophils are still being recruited to the site of injury.** A montage of DIC images overlaid with the green fluorescence channel at the timepoints indicated after tailfin injury (hpi is hours post injury). Green neutrophils can be seen to accumulate at the site of injury between 6 and 14 hours after injury. Thus recruitment continues whilst cells recruited previously have begun migrating away.

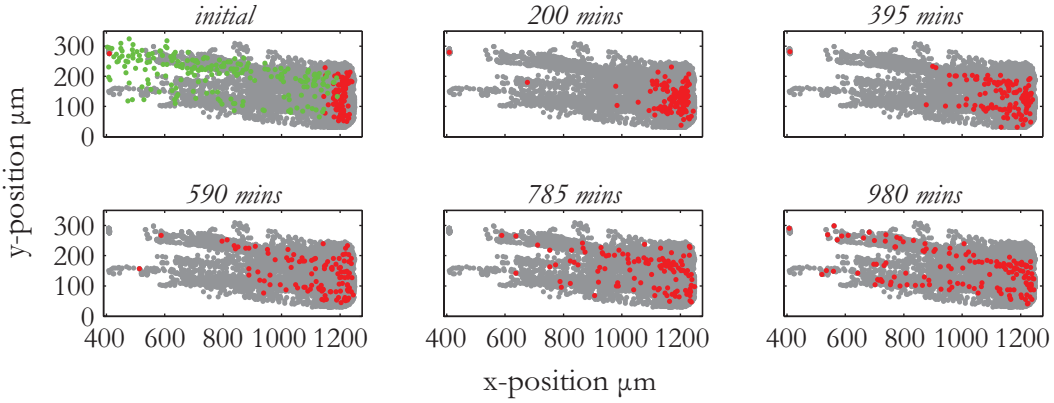

Figure 3.6: **Experimental data: reverse migration of photoconverted neutrophils away from a wound in zebrafish.** The top left subplot shows the initial observed position of all neutrophils. The photoconverted neutrophils are marked in red. All shown in grey are the positions of all photoconverted cells at all subsequent time points. The data is aligned and aggregated over 6 zebrafish and the wound was located in each case at approximately 1250 µm on the X position axis. The remaining subplots, likewise, show the evolving positions of photoconverted neutrophils at increasing time points relative to the start of observations.

These plots of the distance of each cell from the wound edge against time reveal a distinct pattern of neutrophil movement: neutrophils appear to be constrained in their behaviour, gradually increasing their mean distance from the wound, at a rate

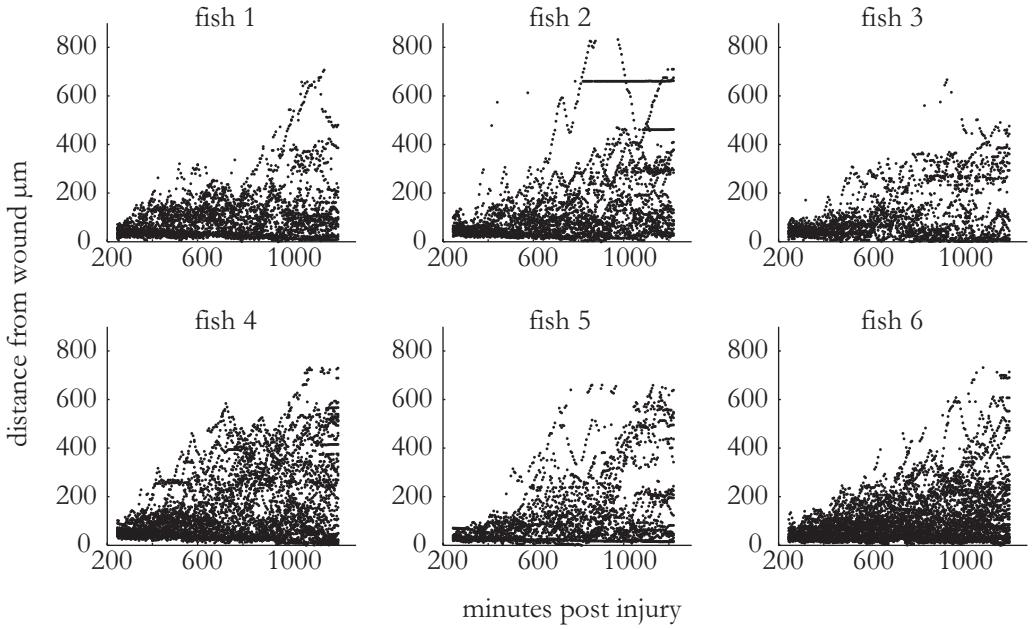

Figure 3.7: **Unprocessed neutrophil displacement data.** For each neutrophil in the six individual zebrafish embryos, the distance from the wound was calculated using algorithms within Volocity and plotted against time.

slower than their maximum speed would permit (Figure 3.7). Figure 3.8 shows the number of cells observed at each observation time for each of the two neutrophil sub-populations. For the red cells there is a generally a slight increase in numbers from the start of the experiment. This is due to the cells being densely packed near the wound initially, resulting in some indistinguishability or occlusion of cells behind other cells. Green cells generally show a decrease in the second half of the experiment. The reason for this is not clear. However, only the data from earlier times is used for calculating the green cell dynamics.

3.4.2 Modelling results: neutrophil recruitment

In order to identify and quantify and dynamics of neutrophils during inflammation recruitment, equation (3.4) was fitted to the data collected for the non-converted (green) neutrophils. These were initially all more than 100 µm from the wound. During the early part of the experiment they migrated towards the wound (Figure 3.5). This early portion of the data, prior to any of the green cells reaching the wound, was used to estimate the drift coefficient in (3.4). The mean number of observed cells was fairly constant over this time but with some noise, therefore a weighted least squares scheme was used as described in Section 3.3.3.

Fitting the drift-diffusion equation to the dataset treats the neutrophils as point

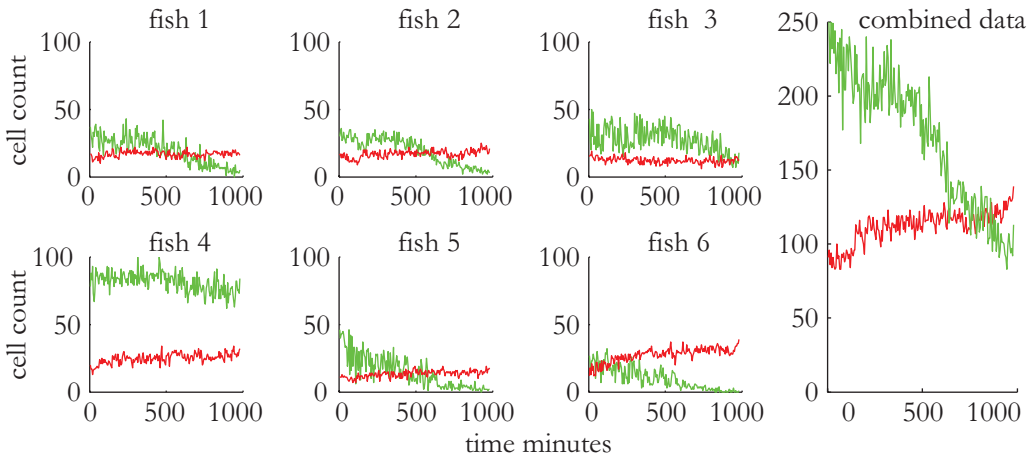

Figure 3.8: **Number of observed neutrophils** (a) The number of observed neutrophils at each observation time is shown for each individual zebrafish embryo. The red line shows the red, photoconverted neutrophils and the green line the green, non-photoconverted neutrophils. (b) The cell counts across all embryos in the combined dataset. For the red cells there is a generally a slight increase in numbers from the start of the experiment. This is due to the cells being densely packed near the wound initially, resulting in some indistinguishability or occlusion of cells behind other cells. These and similar effects also explain the noise on the cell counts. The green cells generally show a decrease in the second half of the experiment. However, only the data from earlier times is used for calculating the green cell dynamics.

embryo	drift coefficient	standard error
1	-0.85	(0.13)
2	-0.95	(0.06)
3	-0.11	(0.02)
4	-0.32	(0.02)
5	-0.48	(0.08)
6	-0.37	(0.06)
all	-0.35	(0.03)

Table 3.1: **Estimated drift coefficients for the model of drift-diffusion describing cell migration toward the wound.** Drifts are in $\mu m\,min^{-1}$, the standard deviation is given in brackets.

objects and asks whether they are behaving like simple particles redistributing stochastically (diffusion) or whether there is an element of active movement towards or away from a chemical gradient (chemotaxis or fugetaxis). The equation generates a value for the drift coefficient, for which non-zero values reflect an active rather than purely random migration. The drift was estimated from the linear relationship between time and mean cell distance from the wound (Figure 3.9). The results of

this analysis are shown in Table 3.1. For the data from the 6 individual embryos, the coefficient estimates ranged in value from $0.11-0.95\,\mu m\,min^{-1}$ and for the aggregated data the value was $0.35\,\mu m\,min^{-1}$. Graphical representation of the fitted data is shown in Figure 3.9. As expected, in all cases cell populations demonstrated active drift toward the wound. This is equivalent to saying that the neutrophils are moving according to a biased random walk (BRW) and is consistent with cell migration that is directed by a chemotactic process. This result, for the first time, quantifies the *in vivo* dynamics of neutrophils during inflammation recruitment in zebrafish.

3.4.3 Modelling results: inflammation resolution

In order to identify and quantify and dynamics of neutrophils during inflammation resolution, a similar analysis was performed for photoconverted (red) cells present at the site of the wound at the time of photo-conversion, 4 hours following the tail-fin transection. In this case, equation (3.3) was fitted to the data collected for the photo-converted (red) neutrophils which begin their migration from within $100\,\mu m$ from the wound (Figure 3.4). Two different versions of (3.3) were fitted to this data. In the first form, the drift coefficient, a was constrained to be zero, whilst in the second it was allowed to take any non-negative value. The first form describes pure-diffusion dynamics which corresponds to undirected random motion. The second form describes drift-diffusion dynamics which corresponds to some degree of chemotactic guidance which could be either attractive or repulsive [55]. Thus fitting the first form is to estimate a diffusion coefficient via a linear fit and fitting the second form both a diffusion and a drift coefficient via a quadratic fit (Figure 3.10). The estimated coefficients are displayed in Table 3.2. Visual inspection of Figure 3.10 suggests that the quadratic model is a better fit to the data than the linear model. However, this will always be the case for a higher order versus a lower order model. Therefore an F-test was applied to the two models as described in Section 3.3.4. The resulting F-test scores are also shown in Table 3.2 and again seem to confirm that the quadratic, drift-diffusion model should be the preferred model. However, further inspection of Figure 3.10 shows that the errors in the model predictions (equation 3.14) are correlated and thus the usual F-test cannot be relied on in this case.

Therefore, to differentiate between the identified models for the aggregated data, both were repeatedly simulated and the results averaged: the pure-diffusion model with $D = 41.8\,\mu m^2\,min^{-1}$ and the drift-diffusion model with $a = 0.26\,\mu m\,min^{-1}$, $D = 8\,\mu m^2\,min^{-1}$. This gave a telling result: the cell population mode of the drift-diffusion model moved away from the wound over time (Figure 3.11, red line), in contrast to the observed data, where the mode remained close to the wound (Figure 3.11, yellow bars). The pure-diffusion model accurately captured this qualitative behaviour, more

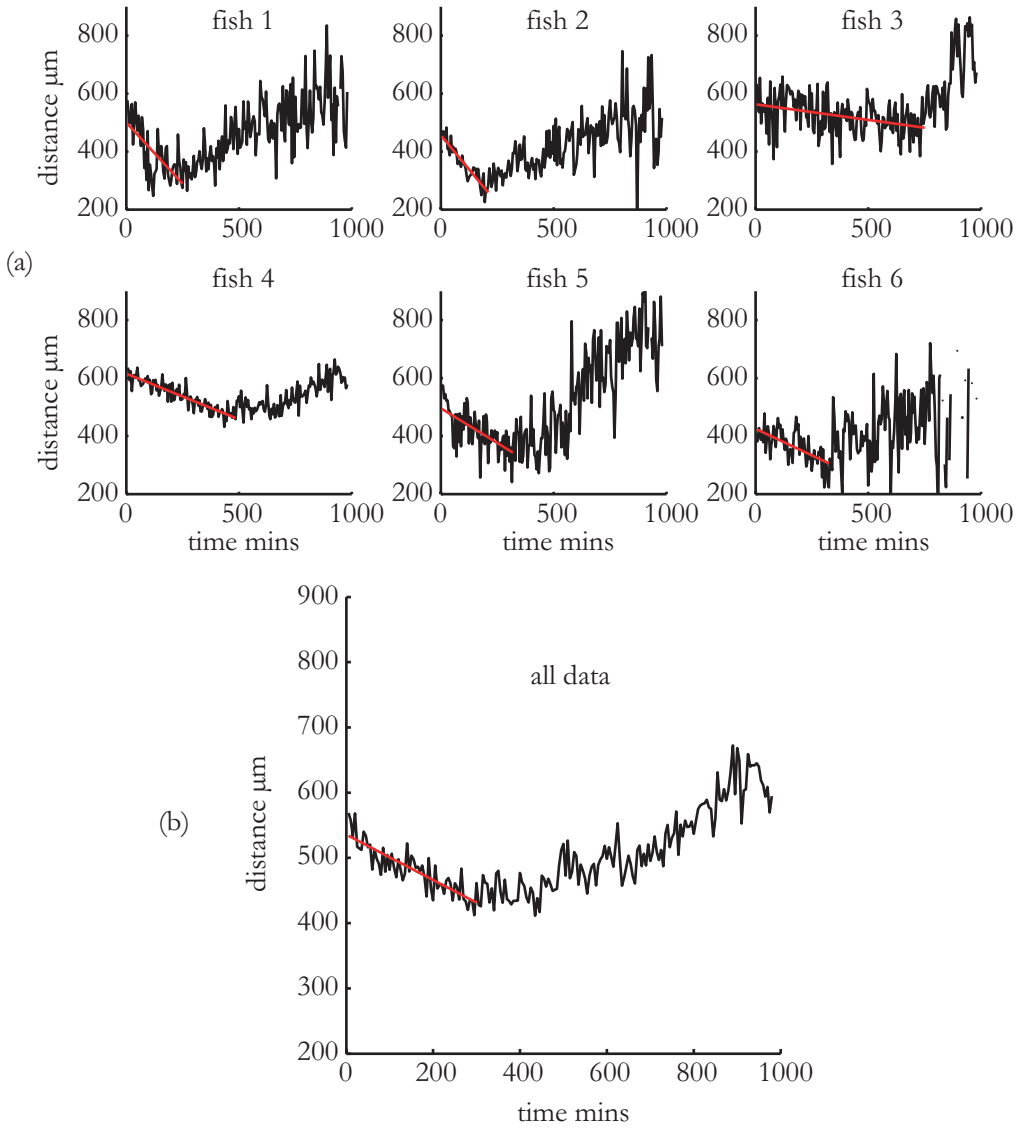

Figure 3.9: **Neutrophils actively migrate into the wound region.** (a) Mean cell distance from the wound against time for the non-photoconverted neutrophils in each of the six embryos (black line). Prediction of mean distance obtained from the linear model of initial drift (red line). Time is measured from the start of observations which commenced four hours after injury. The cell count in embryo 6 (bottom right) was low and sometimes zero near the end of the dataset, which explains the missing sections. (b) Data and model combined over all six embryos.

accurately reflecting the observed distribution of neutrophils over time (Figure 3.11, blue line).

The belief that neutrophil inflammation resolution migration is directed chemotaxis is becoming prevalent in the zebrafish community. The striking results presented here challenges that emerging consensus and suggests, to the contrary, that

stochastic redistribution best describes the phenomena of neutrophil behaviour during inflammation resolution.

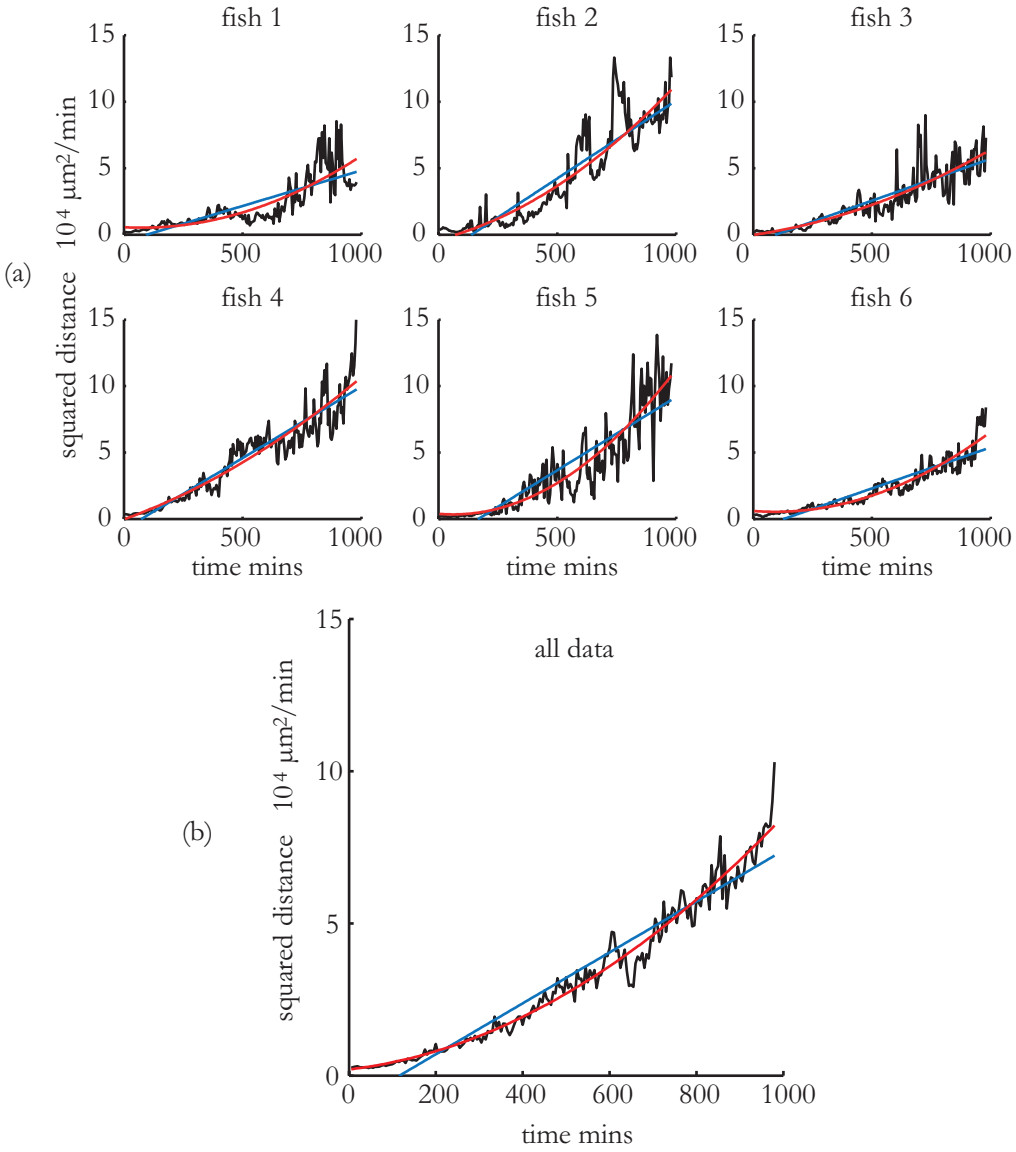

Figure 3.10: **Neutrophil inflammation resolution migration behaviour fitted to both pure-diffusion and drift-diffusion models.** (a) Plots of mean squared cell distance from the wound against time for the photoconverted (red) neutrophils for each of the six embryos. Also shown on each plot are the fits for the linear model corresponding to pure-diffusion with zero drift (blue line) and for the drift-diffusion model (red line). (b) Data and models combined over all six embryos.

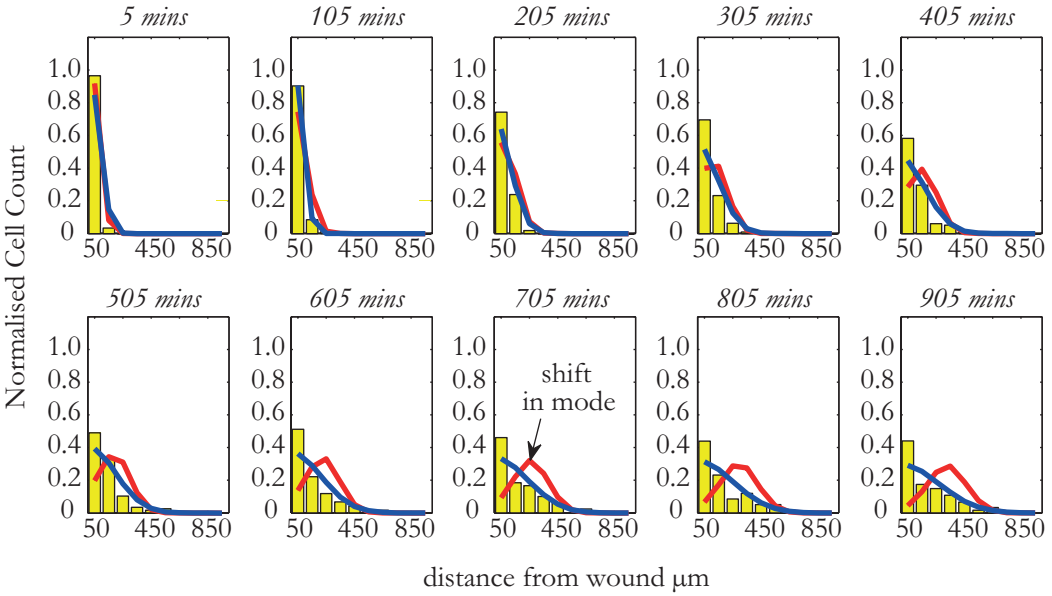

Figure 3.11: **Simulation reveals a pure-diffusion model to be a better fit to the real data.** Both the drift-diffusion model (red line) corresponding to drift ($0.26\,\mu m\,min^{-1}$) and diffusion ($8.0\,\mu m^2\,min^{-1}$) and the pure-diffusion model (blue line) corresponding to diffusion ($41.8\,\mu m^2\,min^{-1}$) were simulated 1000 times. The simulations were used to produce a distribution for the spatially binned data of each model. The mean values of cell distribution over space are shown by the red and blue lines, respectively (in terms of distance from the wound). Overlaid on these is a corresponding histogram representation (yellow) of the real data (combined over all fish). The histogram bins have width $100\,\mu m$ and are centered at 50 to $950\,\mu m$ from the wound. The pure-diffusion model shows a correct qualitative prediction of cell distribution whereas the drift-diffusion model predicts that the population mode moves away from the wound over time, in contrast to the observed data.

embryo	drift-diffusion model				pure-diffusion model		F-test
	drift coefficient		diffusion coefficient		diffusion coefficient		
1	0.25	(0.05)	-4	(10)	27	(2)	38
2	0.27	(0.07)	23	(15)	56	(4)	28
3	0.19	(0.05)	13	(10)	32	(3)	14
4	0.21	(0.05)	32	(11)	54	(3)	14
5	0.35	(0.07)	-8	(14)	55	(4)	82
6	0.27	(0.03)	-7	(6)	31	(2)	145
all	0.26	(0.02)	8	(3)	41.8	(0.10)	267

Table 3.2: **Estimated coefficients for the drift-diffusion model and pure-diffusion model of cell migration away from the wound.** Drifts are in $\mu m\,min^{-1}$ and diffusions in $\mu m^2\,min^{-1}$, standard deviation is given in brackets. Under normal circumstances, an F-test value greater than 4 indicates that the drift-diffusion model should be preferred to the pure-diffusion model.

3.5 Discussion

3.5.1 Neutrophil recruitment

The analysis of zebrafish neutrophil recruitment data in this chapter confirms that these cells are actively recruited towards the wound. This conclusion follows from the fact that the drift-diffusion description of their dynamics has a positive value of drift coefficient. The magnitude of this drift is small compared to usual migration speeds, which are of the order of $10\,\mu\mathrm{m\,min}^{-1}$. This in turn implies that the cells are performing a biased random walk rather than persistent migration [5] (see Section 2.1.5). In other words their natural stochastic search strategy is still dominating their migration but with a small bias towards the target direction. There is net movement towards the wound but the neutrophils are still able to track back and forth in the tissue.

However, unlike the photo-converted sub-population, the non-photoconverted cells are not a closed group: some of these cells may leave the observation area and other cells may arrive which cannot be distinguished from the original cells. Perhaps there will be a balance to these migrations such that they cancel each other out, but it is possible that these effects could be weakening the ability to detect the full amount of drift. It would be possible to investigate this further by repeating the experiment with earlier photo-conversion, this time of a sub-population of cells from a band in the middle of the observation area. This would label a distinct sub-population of cells that are being recruited and would overcome the issues mentioned above. However, it may be difficult to capture a large enough number of cells in this proposed photo-conversion process to make analysis feasible and reliable. This remains an avenue for future exploration.

A second alternative would be to analyse cell tracking data instead of the population approach taken here. Such an approach has been applied to proteins in living cells [104, 112] and to *in vivo* melanoma cell tracks [33]. However, care is needed when considering cell tracks as a naive approach could misrepresent short term correlations in track direction as biased migration, as will be discussed below. Furthermore it would be necessary to average the results obtained over a population and this in turn would require enough cells that can be tracked for a sufficient amount of time, at least greater than typical directional persistence times. In addition, to identify tracks requires faster sampling of observations. This would have to be balanced against total experiment runtime because there is a trade off between the two due to bleaching out of the cells and harm to the host organism from overexposure to UV light. These are some of the issues which is the analysis of this chapter has been designed to circumvent.

3.5.2 Inflammation resolution

The regression analysis of zebrafish neutrophil inflammation resolution data was not in itself conclusive. However, when combined with comparative simulations of the identified models, the pure-diffusion models was identified as the best model for the data. This is in sharp contrast to the results of [85] discussed in the introduction to this chapter. To summarise, Mathias et al. used tracking data for inflammatory neutrophils to calculate cell velocity and a directionality index. Finding these to be comparable for cells moving towards and away from the wound it was concluded that inflammation resolution migration, like recruitment migration, was directed motion as a result of chemotaxis. It was also reported that cells were observed to cyclically migrate towards and then away from the wound with the conclusion that there was directed movement alternating between the two directions. Whilst [85] is an important contribution to the field, a number of cautionary points need to be made:

- Neutrophils movements are usually modelled with a CRW process. Short term correlations in direction are therefore observed even when the motion is globally random. It is possible for these short term track portions to exhibit a high directionality index. This may be happening given the time scales over which the observations were made.

- Neutrophils may move at internally determined speeds that are roughly constant irrespective of whether they are migrating according to random or directionally biased patterns.

- Restricted channels in which neutrophils may preferentially move may further enhance the apparent directionality of random motion.

- In [85] neutrophils were 'observed to display simultaneous, bidirectional migration'. However, this could just describe a CRW observed over longer time scales.

- Some of the directionality in [85] is measured from track portions that are parallel to the wound. This suggests that the directionality is independent of attractive or repulsive forces which would be acting perpendicular to the wound and thus relate to inherent migration processes.

- It is stated that there is a degree of selectivity in the use of tracks (and the process of tracking is in itself implicitly selective). From a population of cells randomly migrating according to a CRW there will always be some that have

significantly high degree of directionality. This will be balanced globally by some which have very little directionality.

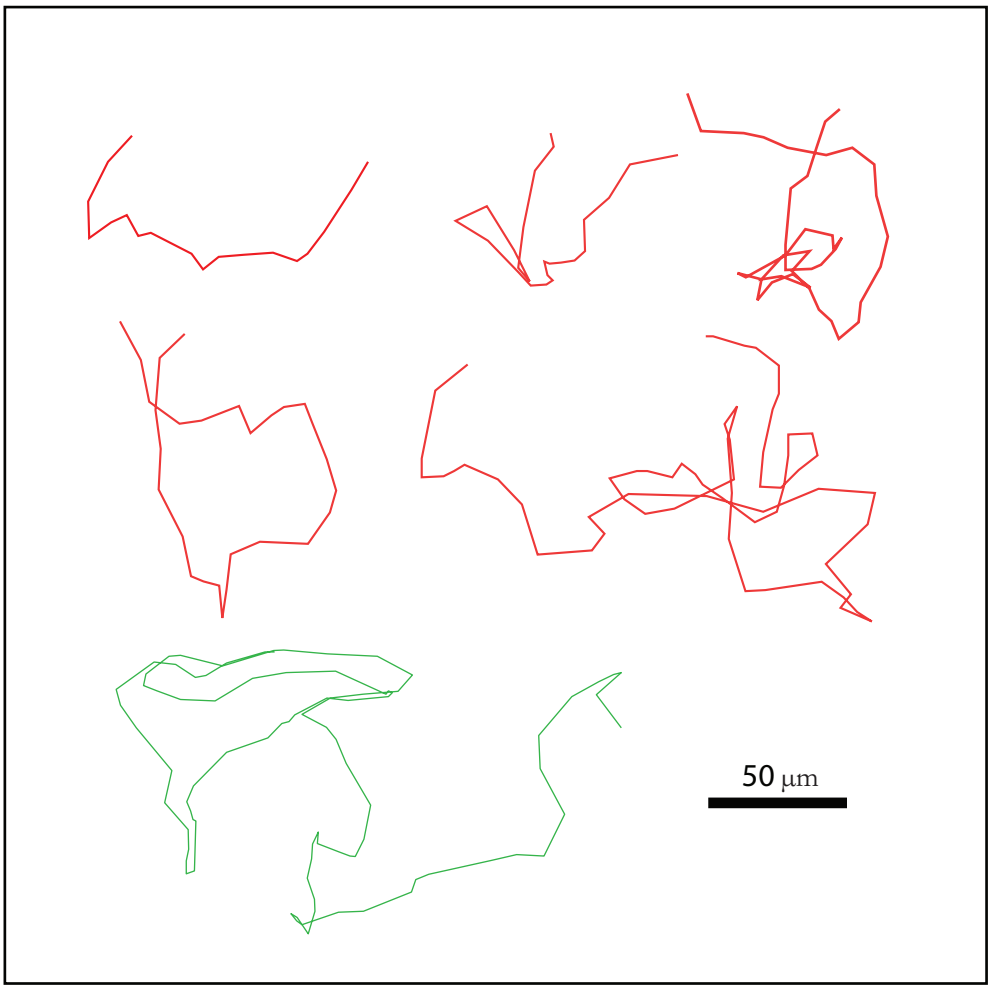

Figure 3.12: **Comparison of cell tracks assumed by [85] to be directed with a cell track which is not.** Red tracks are those presented in [85] as exhibiting bi-directional chemotaxis (directed migration). The green track was observed in the head of a zebrafish embryo in normal homeostasis such that it can be presumed no chemoattractant signals are presents. It could also be interpreted as showing directed or bidirectional chemotaxis. However, the migration is actually random search behaviour. This illustrates how it is easy to draw misleading conclusions.

In relation to the final point, attention is drawn to Figure 3.12. This figure shows neutrophil tracks presented in [85] as showing bidirectional chemotaxis (red tracks). The tracks are taken from two separate experiments from which one and four tracks are taken respectively. Where necessary the tracks have been separated to make them easier to see individually. In each case the neutrophil was migrating in the vicinity of an induced wound to the zebrafish embryo which was roughly below the

track as plotted. It is asserted in [85] that these tracks demonstrate directed motion, as opposed to random motion, in each direction (i.e. upwards and downwards, as shown here). However, the representative green track was observed in the head of zebrafish embryo in equilibrium conditions*. This embryo had not been subject to any inflammation inducing injury. It is presumed, therefore, that this neutrophil was not subject to any chemoattractant directional guidance. It was migrating according to inherent random search patterns. Nevertheless, the track show a large amount of correlation in short to medium term migration direction. In analogy to the findings in [85], this green tracks could be interpreted as exhibiting guided migration cyclically in the upward then downward directions. In this case it would be a spurious conclusion.

Most analyses of cell migration are subject to at least some of the issues highlighted in this section. One of the novel contributions of this thesis is to analyse neutrophil migration from a population perspective. This bypasses any biases that would arise in the tracking or track selection process. Because tracking is not required, data can be incorporated from all cell-position observations and furthermore, observations can be included over a long time period so that short time directional correlations can be ignored. Under the approach proposed here the opposite conclusion to that reached by Mathias et al. is established: migration of neutrophils away from an inflammation site is achieved by inherent random migration patterns and not by chemotaxis or fugetaxis.

The identified model for neutrophil inflammation resolution obtained in this chapter generally fits the data quite well. However, this is a relatively simplified model. The need to add further explanatory detail to the model is motivated by observing that the model consistently underestimates the number of photoconverted cells remaining adjacent to the wound. This suggests that some cells are actively retained at the wound site. Indeed, the experimental results showed that whilst early recruited neutrophils are migrating away from the wound, other cells are still being actively recruited. This implies a cell by cell switch off from the original chemoattractant field. To completely address this will require the development of a framework that can incorporate multiple models to reflect the dynamic mix of neutrophil behaviours present within a single population. This will be developed further in the following chapters.

3.5.3 Issues common to both analyses

It was found that the model errors are not un-correlated. This does not mean the model is wrong: any simulation of a pure-diffusion model will show that the errors

*Experimental data courtesy of Dr Katy Henry, Academic Unit of Respiratory Medicine, Department of Infection and Immunity, University of Sheffield, Sheffield, UK.

between the actual and theoretical mean squared distance are correlated. However, it does suggest that a more accurate result could be obtained by using a generalised least squares scheme (as implemented in [33]) in place of the ordinary least squares used here. This is another avenue for future work. Nevertheless, the ordinary least squares method is a well known one that is relatively easy to conceptualise. One of the aims of the work in this chapter has been to develop a method that is at a relatively simple theoretical level, in order that experimental collaborators will find the method accessible enough to consider utilising.

3.6 Conclusions

This chapter, which is based on already published work [53], provides a first mathematically rigorous attempt to analyse the dynamics of inflammatory neutrophils in zebrafish. The immune systems of fish and mammals have many similarities. Thus these dynamics of zebrafish neutrophils may shed important light on those in mammalian immune systems leading to the deepened understanding vital in developing therapies and preventions for many inflammation related chronic diseases that affect the human population.

Both the recruitment phase and resolution phase of inflammatory neutrophil migration have been investigated. The recruitment phase analysis showed that neutrophils undergo biased random walk migration towards an inflammation site. The presence of a bias confirms the widely held view that the cells are actively recruited to the wound area following chemotactic signals from the host organism. However, in addition, it was possible to quantify this bias by estimating a drift parameter for the cell population. This information will be useful in predicting neutrophil response to an inflammation trigger. If the recruitment analysis confirms existing views, the inflammation resolution analysis, in contrast, challenges a growing consensus. The inflammation resolution analysis suggests that an simple (unbiased) random walk best describes the neutrophils motion. On a population level this results in a pure-diffusion process and implies that the neutrophils are moving according to inherent stochastic processes and not in response to any attractive or repulsive signals.

The framework developed here for identifying reverse migration dynamics has stepped beyond the basic analyses currently available in the literature. However, it is apparent from our own evaluation of the framework in Section 3.5 that it has some limitations. Using standard methods alone, the framework was not able to correctly discriminate between the identified models in the inflammation resolution context and further simulation comparisons were needed to inform the decision. Furthermore, the framework is not readily extendible to more complex scenarios in which the

individual neutrophils may be in different migratory modes. In the next two chapters of this thesis a novel simulation based framework is developed which overcomes these shortcomings and which can give a more definitive answer on the question of inflammation resolution dynamics.

Chapter 4

Identifying neutrophil migration: a Bayesian modelling framework

In Chapter 3, a regression analysis framework was developed to identify neutrophil migration dynamics during the recruitment and resolution phases of inflammation episodes. When applied to inflammation resolution in zebrafish this gave an indication that the neutrophils were making purely diffusive movements and were not subject to any deterministic forces driving them away from the wound which induced the inflammation. The regression analysis framework provided a relatively quick and simple method for analysing the cell dynamics with a far greater mathematical rigour than has been applied before in this context. Two potential drawbacks of that framework, however, were that,

- not all information in the data was used as only the mean squared distance was extracted from the full observation set at each timepoint; and,
- an additional simulation step was required to distinguish between the competing identified models in order to rule out the existence of a small but nevertheless significant drift term.

In this chapter a more thoroughgoing analysis framework is developed – one which draws information from the full distribution of observed cell positions at each timepoint. Furthermore, the framework now to be proposed sits within a Bayesian paradigm which provides a number of advantages. Firstly, the parameters to be identified are physical motility coefficients and this gives plausibility ranges within which to search for them. These ranges can be incorporated into the Bayesian prior distribution over the parameters. Secondly, the key motivation in the identification process is a desire to test the fidelity of alternative models of neutrophil migration. These models, the pure-diffusion and the drift-diffusion models, correspond to alter-

native biological realities of the living system. In the former the cells are moving without external direction according to inherent random processes; in the latter, in addition to the random movements, there is also an attractive or repulsive bias influencing the migration. In a Bayesian framework, model selection can be included as a natural extension of the parameter estimation process. This provides a data-driven methodology for model discrimination. Thirdly, the Bayesian approach is a probabilistic one which yields distributions over identified parameters rather than point estimates. Thus the uncertainties associated with all identified parameters and models can be evaluated in a way that is both rigorous yet easy to interpret.

The development in this chapter of a Bayesian framework for cell migration is a central contribution of this Thesis. The framework is also applied to the zebrafish neutrophil data to confirm the answer to the key question – What kind of biological process is at work in the resolution dynamics of inflammatory neutrophils? A correct answer to this is vital as it will inform future research towards interventions and therapy for many chronic diseases.

4.1 Introduction

4.1.1 Bayesian methods

The benefits of a Bayesian approach can be summarised as follows:

- Prior knowledge about parameters can be included in the estimation process.
- Uncertainties in model choice and parameter values are made explicit in the result.
- These uncertainties are expressed in a form that makes intuitive sense to most readers (as compared to standard errors of classical estimation).
- Predictions can be based on the full posterior distribution over models and parameters rather than just point estimates, meaning they are more balanced.

In addition, sensitivity to the various parameters is built in to the results: a low variance posterior distribution for a parameter means that the fitted model is sensitive to that parameter, whereas a broad posterior means that the fitted model is insensitive to that parameter.

In some applications there may be sensitivity to the choice of prior distributions over parameters. Care must also be taken over the range of plausible models that are included or excluded in the set of candidate models. In this study the parameters to be identified are physical motion parameters for relatively large cells. As such

they are constrained to be non-negative and to be within a range that is physically plausible. This makes choosing the prior distributions less contentious than in some other contexts such as black box modelling where the output of a system is modelled but the functional forms and parameters used have no physically interpretable meaning. In the past a drawback of Bayesian estimation was the sheer amount of computation needed to evaluate the posterior distribution, especially in high dimensional systems. Certainly in this study the identification process is computationally intensive. However, since it is an offline process, it is easily handled on a desktop computer with typical run-times being in the region of 1-2 hours.

4.1.2 Why an approximate Bayesian approach is appropriate

Within the ensemble of Bayesian methods, even the sampling approaches such as Markov chain Monte Carlo (MCMC) require the evaluation of a likelihood function. Sometimes this is not possible: perhaps no analytic likelihood function is available or if it is available it may nevertheless be impossible to evaluate.

The data analysed in this thesis may be described as follows: at any timepoint, t, there are N_t observations of individual cell centroid positions. At time $t+1$ likewise there are N_{t+1} observations. In general $N_t \neq N_{t+1}$ and furthermore there is no correspondence between the observations at successive timepoints: cell are not tracked and so it is not known whether a particular cell moved from x_t to x_{t+1}. To compute a likelihood, therefore, one would have to take into account the $N_t N_{t+1}$ possible trajectories and thus the number of possible paths that would have to be considered would grow exponentially with the number of observation times. This correspondence problem could be avoided or at least made more manageable if cells that can be tracked were extracted from the data, but this would be to throw away a large amount of information and would distort the analysis towards cells which are migrating in a particular way. For instance, cells that have moved away from areas of higher cell density are more amenable to being tracked than those which stay closely packed. The decision not to track the cells also meant that the cells could be observed with a low inter-sampling time of 5 minutes. This, in turn meant that problems associated with over exposure of living specimens to ultra-violet light were avoided*

*Overexposure to UV light causes bleaching of the cells so that they are no longer observable and can cause death of the zebrafish larva specimen. In either case the experiment would terminate prematurely and the full inflammation resolution process would not be observed

4.1.3 ABC SMC with model selection

Chapter 2 reviewed the ABC approach and the different methods within it. To recap briefly, at the heart of the ABC method is the basic ABC rejection sampler. This can be finessed with parameter weighting and local linear regression [11] or built into a MCMC [82] or sequential Monte Carlo (SMC) framework [114]. Whilst the ABC-MCMC approach is efficient in theory, it is liable to get stuck in low density areas of the posterior distribution when there is noise in the simulation process. Having tested each of the ABC methods it was found that the ABC-SMC approach was efficient and without the problems of ABC-MCMC. It can also be extended to include model selection. Hence it is suitable to distinguish between the drift-diffusion and pure-diffusion models for neutrophil migration. The ABC-SMC approach was therefore adopted for the analysis reported in this thesis.

The ABC approach to model selection is based upon standard features of the Bayesian approach to model criticism, in particular the use of Bayes factors [47, 67]. If two models, m_1, m_2 are being compared for a system, then the Bayes factor for Model 1 with respect to Model 2 conditional on observed data, x, is defined as,

$$B_{12} = \frac{\mathrm{P}(m_1 \mid x)}{\mathrm{P}(m_2 \mid x)} \frac{\mathrm{P}(m_2)}{\mathrm{P}(m_1)} \tag{4.1}$$

A simple rearrangement of (4.1) shows that the Bayes factor, B_{12}, is the means for transforming the ratio of prior probabilities to the ratio of posterior probabilities. The interpretation of Bayes factors relies on the classification in the original edition of the work by Jeffreys [59]. An adapted summary of this classification is shown in Table 4.1. If the two models m_1, m_2, are the only candidates and their prior probabilities are equal then the Bayes factor, B_{12}, is equivalent to the betting odds for the model m_1. This provides an intuitive rationale for the classification by Jeffreys.

B_{12}	Evidence for m_1	$\log_{10} B_{12}$
< 1	not in favour	< 0
$1 - 3.2$	in favour but not significant	$0 - 0.5$
$3.2 - 10$	substantial	$0.5 - 1$
$10 - 32$	strong	$1 - 1.5$
$32 - 100$	very strong	$1.5 - 2$
> 100	conclusive	> 2

Table 4.1: **Interpretation of Bayes factor values.** B_{12} is the Bayes factor value for a model, m_1, with respect to an alternative model, m_2. This interpretation of the Bayes factor values is adapted from Jeffreys [59].

The version of ABC-SMC with model selection developed here is adapted from

[121]. It is described fully in Algorithm 4.2. The parameter set is augmented with a model index parameter. A parameter vector is therefore specified by,

$$\theta_i = (m_i, b_i, D_i)^\top \tag{4.2}$$

where m_i is the model index, b_i is the drift coefficient and D_i is the diffusivity coefficient. Coefficient b_i is set to zero if the model index is for the pure-diffusion model.

The model marginal posterior distribution is simply given by the proportionate representation of each model index value in the final population of samples. Thus if there are 1000 samples and of these 800 have model index 1 and the remaining 200 have model index 2, then the Bayes factor,

$$B_{12} = \frac{800}{200} = 4, \tag{4.3}$$

and this would be considered substantial evidence for Model 1 over Model 2.

Model selection can never be guaranteed to give the correct answer. Indeed it is questionable whether there ever is a correct model of a system*. There are always potential pitfalls associated with, for instance, an inappropriate choice of candidate models or an experimental design which fails to fully excite all the models [61]. More particularly, concerns have been raised over the validity of Bayesian model choice in the ABC setting [109]. Underlying these concerns is the question, whether the approximations involved allow the ABC version of the Bayes factor to approach the true value. In particular, even if a summary statistic is sufficient for all individual models (which itself is not always the case) it often is not sufficient to distinguish between models. This is an area of ongoing research [106] and for now any conclusions from Bayes factor analysis should be taken as indicative and backed up with further simulation studies to test the validity of the conclusions in the particular experimental context. This is addressed in Section 4.3.

4.2 Identification framework development

This thesis is addressing the open question of how cell population migration dynamics can be identified from time lapse video-microscopy data. It has been established that an ABC approach is suitable and necessary in this context and that ABC-SMC is a good choice of method. In order now to proceed and to apply ABC-SMC within the context of neutrophil migration dynamics it necessary to provide for the following,

*As the statistician George Box famously wrote, 'Essentially all models are wrong, but some are useful'[18]

- a **dynamic model**, $\mathrm{P}(x \mid \theta)$ and corresponding algorithm to simulate the forward process for a given parameter specification,
- a **summary statistic**, $S(x)$, which reduces the observations to a lower dimensional space,
- a **distance measure**, $\rho(S(x), S(y))$, to quantify the discrepancy between the summary of a simulation and the summary of the observed process.

This is done in the following sections, forming a key contribution of this thesis.

4.2.1 Dynamic model for cell migration

It has been reported that neutrophils move according to a correlated random walk (or biased correlated random walk if a directional cue if present) [122, 126]. However, over time-scales greater than the typical directional persistence times of the cells [122], this type of motion approaches the limit of a pure-diffusion (or drift-diffusion) process [26]. It is these time scales that are being considered here. These processes are described by the following stochastic differential equation, known as an Ito diffusion, [98],

$$\mathrm{d}x_t = b(t, x_t)\,\mathrm{d}t + \sigma(t, x_t)\,\mathrm{d}\omega_t \tag{4.4}$$

where x_t is the cell position at time t, $b(t, x_t)$ is the deterministic drift or migration bias, $\sigma^2(t, x_t)$ represents the magnitude of the diffusive dynamics and ω_t is a white noise process. If the simplifying assumption is made that both the drift and diffusivity are constant in space and time,

$$b(t, x_t) = b, \sigma(t, x_t) = \sigma, \tag{4.5}$$

then this motility can be approximated in one dimension with a boundary at $x = 0$ according to the following discrete time model,

$$x_{t+1}^{(i)} = \max(0, x_t^{(i)} + b_{\mathrm{out}}\Delta t + \omega_t^{(i)}\sqrt{2D\Delta t}), \tag{4.6}$$

where $x_t^{(i)}$ is the position of the i^{th} neutrophil at time t, for $i = 1, \ldots, N$; b_{out} is a bias velocity away from the boundary; D is the underlying diffusivity constant or magnitude of random movement of the neutrophils; $\omega_t^{(i)} \sim \mathcal{N}(0, 1)$ are a family of independent white noise processes; Δt is the time increment. The use of the $\max(0, .)$ function implies that cells do not migrate through the boundary and if they reach the boundary they remain there until the next time increment. The boundary is introduced to describe the wound inflicted on the zebrafish specimens

to induce inflammation. This is approximately straight and perpendicular to the net migration of cells. Cells are not observed to migrate across the line of the wound.

Equation (4.6) represents both the models considered: with $b_{\text{out}} = 0$ it is the pure-diffusion model, Model 1, which corresponds to an unbiased random walk and with b_{out} taking any non-negative value it is the drift-diffusion model, Model 2, which corresponds to a biased random walk. It both cases the random walk has a variable step size, where the steps sizes are drawn from a random normal distribution. It is, therefore, straightforward to simulate this system by sampling the $\omega_t^{(i)}$ from a Normally distributed random number generator. For a given set of experimental results, (4.6) was used to implement the process function, $\mathrm{P}(x \,|\, \theta)$, by applying it to all the initial x position coordinates of all observed cells and recording the evolving positions of the cells at times which correspond to the experimental observation times.

4.2.2 Summary statistic for the observations

The observations of cells in the experimental data at time t can be described by,

$$\mathcal{Y}_t = \{x_{t,i}\}_{i=1}^{M_t}, \tag{4.7}$$

where $x_{t,i}$ is the distance from the wound of i^{th} recorded cell observation at time t; $M_t \leqslant N_{\text{c}}$ is the number of observed cells at time t; and N_{c} is the total number of cells in the system. It should be noted that $x_{t,i}$ and $x_{t^*,i}$ observed at different times do not in general correspond to observations of the same cell as discussed in Section 4.1.2. Also, although the zebrafish is transparent, cells are not necessarily observed at every timepoint point because they sometimes appear to coalesce with or are occluded by other neutrophils. If T_{obs} is the total number of observation times then a complete observation set may now be defined as,

$$\mathcal{Y}^{\text{obs}} = \{\mathcal{Y}_t\}_{t=1}^{T_{\text{obs}}}. \tag{4.8}$$

It is clear that the event of reproducing an observed set of cell positions subsequent to the initial observation has probability zero. It is therefore necessary to construct a summary of the observations which is of low dimensional and yet preserves as much of the information in the data as possible. This was done by summarising the cell positions as a discrete distribution over space, as follows,

$$\mathbf{V}_t = \begin{pmatrix} \sum_{i=1}^{M_t} \chi_{B_1}(x_{t,i}) \\ \vdots \\ \sum_{i=1}^{M_t} \chi_{B_b}(x_{t,i}) \end{pmatrix} \tag{4.9a}$$

$$\mathbf{Y}_t = \frac{1}{\sum_i V_{t,i}} \mathbf{V}_t, \tag{4.9b}$$

where $B_j, j = 1, \ldots, b$ is a set of spatial intervals forming a partition of the range of the $x_{t,i}$; χ_{B_j} is the indicator function* of the interval B_j; and $\mathbf{Y}_t$ is thus the normalised form of $\mathbf{V}_t$.

4.2.3 Distance between simulated and observed cell distributions

Finally, in order to apply an ABC scheme for parameter estimation it was necessary to choose a metric for comparing the difference between the summarised observations for two complete observations sets.

A commonly used method for measuring the distance between two distributions, especially in the domain of information theory, is the Kullback–Leibler divergence (KLD) or relative entropy [8, 14]. In the discrete case, this quantity is defined by,

$$D_{\mathrm{KL}}(P||Q) = \sum_i P(i) \ln \frac{P(i)}{Q(i)} \tag{4.10}$$

where ln is the natural logarithm function. The KLD in this usual form is not a symmetric measure of difference. It measures the difference between the two distributions from the perspective of the first of them, in the sense that the differences at each point are weighted according the that distribution. When this divergence was originally introduced by Kullback and Leibler [69] it was in the symmetric form,

$$DS_{\mathrm{KL}}(P||Q) = \sum_i \left(P(i) \ln \frac{P(i)}{Q(i)} + Q(i) \ln \frac{Q(i)}{P(i)} \right) \tag{4.11}$$

A commonly used alternative, the Bhattacharyya distance (BD), is also widely used in statistical signal processing for measuring the distance between distributions [8, 13]. For two discrete distributions f and g over the same domain X, the Bhattacharyya distance (BD) D_{B} between them is symmetric and is relatively simple to compute,

$$D_{\mathrm{B}}(f, g) = -\ln \sum_{x \in X} \sqrt{f(x) g(x)}. \tag{4.12}$$

A problem is encountered with the BD and the KLD when the support of one distribution being measured is not identical to that of the second. In this case an infinite distance is returned even if the distributions are otherwise very similar, as

* $$\chi_B(s) = \begin{cases} 1 & \text{if } s \in B \\ 0 & \text{otherwise} \end{cases}$$

illustrated for the KLD in Figure 4.1. It has also been suggested that both the BD and

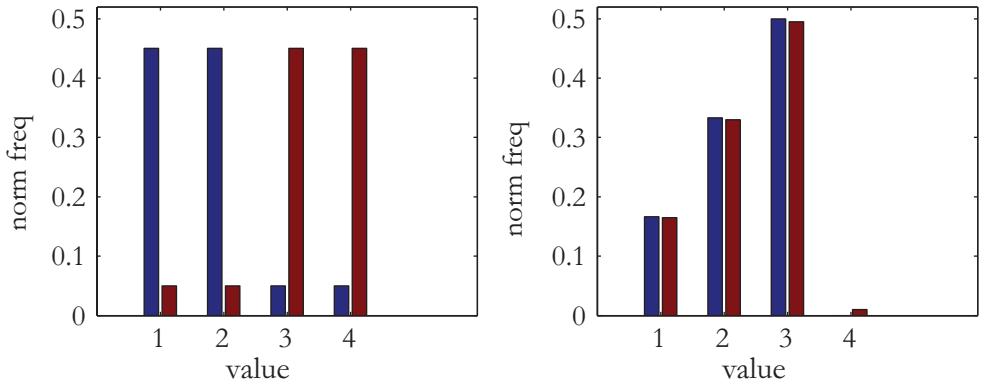

Figure 4.1: **A problem with using the symmetric Kulback-Liebler divergence.** (a) The KL divergence for the two discrete distributions coloured blue and red respectively is 3.5. (b) In this case, though the distributions are very similar, the have non-identical support and therefore the divergence is infinite.

KLD, whilst often used for comparing histogram data, are not suitable candidates for this task when the histograms are ordinal or modulo rather than nominal in nature*[23]. This is because BD and KLD only considers the difference between corresponding histogram bars and not the amount of 'work' needed to transform one histogram into the other. A naive way of taking this into account is to consider the minimal pairwise difference between all samples making up the histogram data. Computing this is exponential in time as there are $n!$ possible pair assignments if n is the number of samples.

Cha and Srihari [23] propose a quick-to implement algorithm to measure the distance between to histograms. They derive it by noting that the minimum difference of pairwise assignments is equivalent to the minimum cost of moving cells (the basic histogram bar size units) to transform from one histogram to the other. Their algorithm for computing it is shown in Algorithm 4.1.

The zebrafish data is summarised as a separate discrete distribution of cells for each of the T timepoints. Therefore, the Bhattacharyya and Cha-Srihari derived distances between two complete observation sets were defined as follows,

$$\rho_{\mathrm{B}}(\mathcal{Y}^{(p)}, \mathcal{Y}^{(q)}) = -\sum_{t=1}^{T_{\mathrm{obs}}} \log \sum_{i=1}^{b} \sqrt{Y_{t,i}^{(p)} Y_{t,i}^{(q)}}, \tag{4.13}$$

*In a nominal histogram (e.g. distribution over makes of car) the ordering of the bars does not matter whereas in an ordinal (e.g. heights) or modulo (e.g. angles) histogram there is an inherent order to the bars.

Algorithm 4.1 Cha-Srihari distance between ordinal histograms

Require: Histograms A, B with bar sizes $A_i, B_i, i = 1, \ldots, n$

Ensure: $D_{\mathrm{CS}}(A, B)$ the Cha-Srihari distance between the two histograms.

for $i = 1$ **to** n **do**

Compute the bar size differences, $d_i = A_i - B_i$.

Compute the cumulative sums of the differences, $c_i = \sum_{j=1}^{i} d_i$.

end for

Compute $D_{\mathrm{CS}}(A, B) = \sum_{i=1}^{n} |c_i|$

$$\rho_{\mathrm{C}}(\mathcal{Y}^{(p)}, \mathcal{Y}^{(q)}) = \sum_{t=1}^{T_{\mathrm{obs}}} D_{\mathrm{CS}}(Y_t^{(p)}, Y_t^{(q)}) \tag{4.14}$$

where $Y_{t,i}$ is the i^{th} component of the vector Y_t.

These two distance measures were considered and compared in order to test the robustness of the identification method with respect to choice of distance measure. It can be observed in Figures 4.2 and 4.3 that the two distance measures have similar trends when applied to simulated neutrophil migration data. However, due to the concerns noted above, the Cha-Srihari distance was used in what follows.

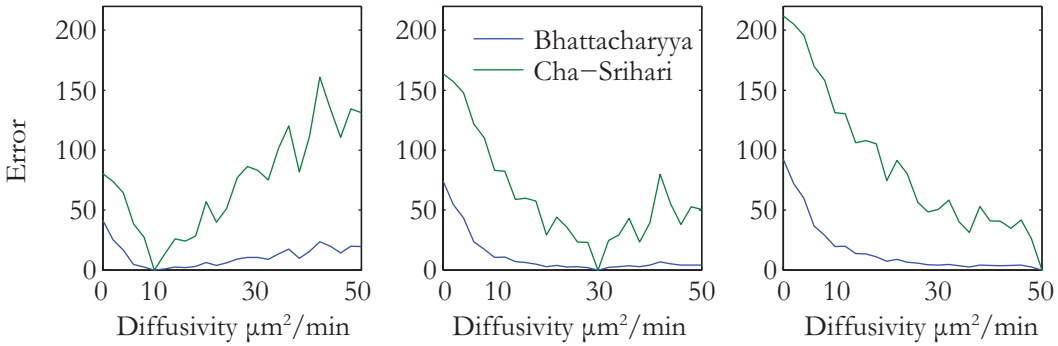

Figure 4.2: **Comparison of Bhattacharyya and Cha-Srihari distances for varying diffusivity.** The distances between a target observation set from a simulated pure-diffusion process with diffusivity, (a) 10, (b) 30, (c) $50\,\mu\mathrm{m}^2\,\mathrm{min}^{-1}$ and those from pure-diffusion processes in the range $0 - 50\,\mu\mathrm{m}^2\,\mathrm{min}^{-1}$, calculated using the two distance measures.

4.2.4 Implementation details

The resulting implementations of the ABC-SMC parameter estimation algorithm and ABC-SMC parameter estimation with model selection algorithm are set out in Algorithms A.1 and 4.2. In the algorithms $\rho(.,.)$ is defined as in (4.13) or (4.14) and

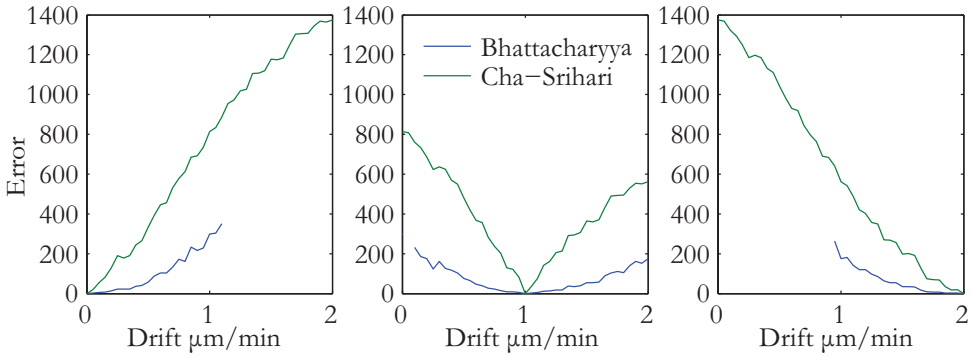

Figure 4.3: **Comparison of Bhattacharyya and Cha-Srihari distances for varying drift.** The distances between a target observation set from a simulated drift-diffusion process with diffusivity, $25\,\mu m^2\,min^{-1}$ and drift (a) 0, (b) 1, (c) $2\,\mu m\,min^{-1}$ and those with drift in the range 0-$2\,\mu m\,min^{-1}$, calculated using the two distance measures. The Bhattacharyya distance is infinite when the support of the two distributions is disjoint. This explains the missing data points for that measure.

thus implicitly includes the summarising of the data as a distribution via (4.9).

In all validation and application runs of the algorithm, unless otherwise stated, the user-set parameters for the algorithm were chosen as follows, $T = 4$ and $N = 4000$. A uniform prior was used over all parameters with the following ranges: diffusivity $0 - 200\,\mu m^2\,min^{-1}$, drift $0 - 2\,\mu m^2\,min^{-1}$.

Many ABC studies do not give much attention to how the choice of error tolerance schedule is made and possibly a process of trial and error is inevitable. It should be noted that it is the final tolerance value which directly affects the results of the estimation, the smaller the tolerance the closer the estimated posterior is to the true posterior. The intermediate tolerances affect the efficiency with which the algorithm reaches this approximate posterior. To give some examples, Beaumont et al. [12] use a linearly decreasing tolerance schedule, whereas Toni et al. [121] appear to have picked each value of tolerance to suit the application with by and large a linear decrease at first but then smaller reduction in error as the final iteration is approached.

The error tolerances in this study were chosen automatically as follows: in an initialisation run N parameter sets were chosen from the joint prior and the associated simulation errors, e_i, with respect to the experimental data were calculated. The initial error tolerance was chosen to be,

$$\epsilon_1 = \frac{1}{2}\max(e_i)$$

and the final tolerance, ϵ_T, was chosen as the first percentile of the e_i.* To in-

*If this final error tolerance was used with a basic ABC rejection sampler using the same prior

Algorithm 4.2 Model selection and parameter estimation using ABC-SMC

Require: data, $\mathcal{Y}^{\text{obs}}$; Monte Carlo population size, N; number of iterations, T; prior distributions on models $\pi(m)$ and on model parameters $\pi(\theta|m)$; simulation algorithm to sample replicated observations from the processes, $\mathcal{Y} \sim p(\mathcal{Y}|m,\theta)$; distance metric ρ, model perturbation kernel M and parameter perturbation kernel K; decreasing error tolerance schedule $\epsilon_1, \ldots, \epsilon_T$

Ensure: a set of parameter vectors θ_i augmented with model indicator m_i, with importance weights $\omega_i, i = 1, \ldots, N$ that together form a weighted sample from the joint posterior distribution, $p(\theta, m|\mathcal{Y}^{\text{obs}})$

for $i = 1$ **to** N **do**

simulate $m_i \sim \pi(m)$, $\theta_i \sim \pi(\theta|m_i)$ and $\mathcal{Y} \sim p(\mathcal{Y}|m_i, \theta_i)$

until $e_i = \rho(\mathcal{Y}, \mathcal{Y}^{\text{obs}}) \leqslant \epsilon_1$

end for

set each $\omega_i^{(1)} \propto \frac{1}{\epsilon_1}\left(1 - \left(\frac{e_i}{\epsilon_1}\right)^2\right)$, such that $\sum \omega_i^{(1)} = 1$

for $t = 2$ **to** T **do**

for each model, m, set $\tau(m)^2 = 2\text{Var}(\{\theta_i : m_i = m\})$

for $i = 1$ **to** N **do**

choose k from $\{1 \ldots \text{N}\}$ with probabilities $\{\omega_1 \ldots \omega_N\}$

set $m^* = m_k$ and $\theta^* = \theta_k$

simulate $\hat{m}_i \sim M(m|m^*)$

Re-choose θ^* from $\{\theta_j : m_j = \hat{m}_i\}$ with probabilities $\{\omega_j : m_j = \hat{m}_i\}$

simulate $\hat{\theta}_i \sim K(\theta|\theta^*; \tau(\hat{m}_i)^2)$ and $\mathcal{Y} \sim p(\mathcal{Y}|\hat{m}_i, \hat{\theta}_i)$,

until $e_i = \rho(\mathcal{Y}, \mathcal{Y}^{\text{obs}}) \leqslant \epsilon_t$

set $\tilde{\omega}_i = \frac{1}{\epsilon_t}\left(1 - \left(\frac{e_i}{\epsilon_t}\right)^2\right)$

end for

set each $\hat{\omega}_i \propto \frac{\tilde{\omega}_i \pi(\hat{\theta}_i)}{\sum_{j:m_j=\hat{m}_i} \omega_j K(\hat{\theta}_i|\theta_j; \tau(\hat{m}_i)^2)}$, such that $\sum \hat{\omega}_i^{(t)} = 1$

set each $m_i = \hat{m}_i$, $\theta_i = \hat{\theta}_i$, $\omega_i = \hat{\omega}_i$

end for

crease efficiency, parameter sets from the initialisation run were recycled into the first iteration if their associated error was less than ϵ_1.

After investigating various choices for intermediate tolerances such as linearly decreasing and a scheme where $\epsilon_{i+1} - \epsilon_T = \frac{1}{2}(\epsilon_i - \epsilon_T)$, it was found that a negative

then on average 1 in every 100 simulations would result in acceptance of a parameter set.

exponential scheme was the most efficient of those investigated. This scheme was achieved by setting,

$$\epsilon_i = \epsilon_1 e^{-\alpha i}, \quad i = 2, \ldots, N-1 \tag{4.15}$$

with

$$\alpha = \frac{\log \epsilon_1 - \log \epsilon_N}{N} \tag{4.16}$$

The efficiency of computation for this scheme with respect to various values of N are shown in Table 4.2.

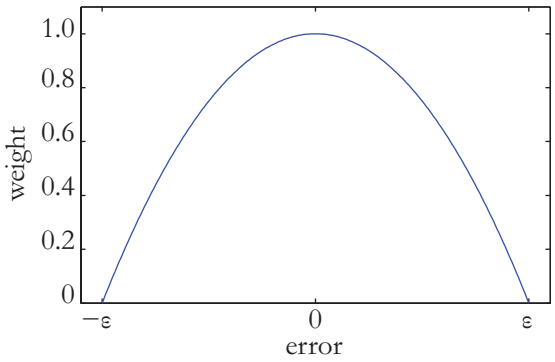

Figure 4.4: **Epanechnikov kernel** used for initial weighting of parameters.

Toni et al. [121] used a uniform acceptance kernel. This means that if the simulation error is just within the current error tolerance, the parameter set is accepted with as much importance as it would be if the error had been zero (prior to the subsequent importance weight allocation). Here, following Beaumont et al. [11] an Epanechnikov acceptance kernel (Figure 4.4) is used to pre-weight the accepted parameter sets with a weight w_k,

$$w_k = \frac{1}{\epsilon_i}\left(1 - \left(\frac{e_k}{\epsilon_i}\right)^2\right) \tag{4.17}$$

where ϵ_i is the current error tolerance and e_k is the calculated error.

The parameter perturbation kernel was chosen to be zero mean Gaussian with variance set to be twice the weighted empirical variance of the parameters in the previous population. Also, a model perturbation kernel was used in which the original model was kept with probability 0.6 and one of the r remaining* alternative candidate models with probability $\frac{0.4}{r}$.

*If a model has died out, i.e. has no representation in the previous generation, then this choice of kernel does not allow it to be re-introduced.

4.2.5 The validity of the posterior distribution

It is important to note that minimising the uncertainty of the estimated parameters is not a goal of the Bayesian approach. Nissen and Propach [95] investigated the performance of four optimisation schemes (a genetic algorithm, an evolution strategy, the classical pattern search and the threshold accepting method) when the cost function contains a significant degree of random noise in its evaluation. They found that averaging repeated evaluations of the function at each parameter value leads to a more efficient algorithm. Diffusion is essentially a noise process and the fewer diffusing cells are being modelled the noisier will be the resulting cell distribution. This poses the question, can a similar repeat-simulation-and-average step be beneficially applied in the ABC-SMC algorithm.

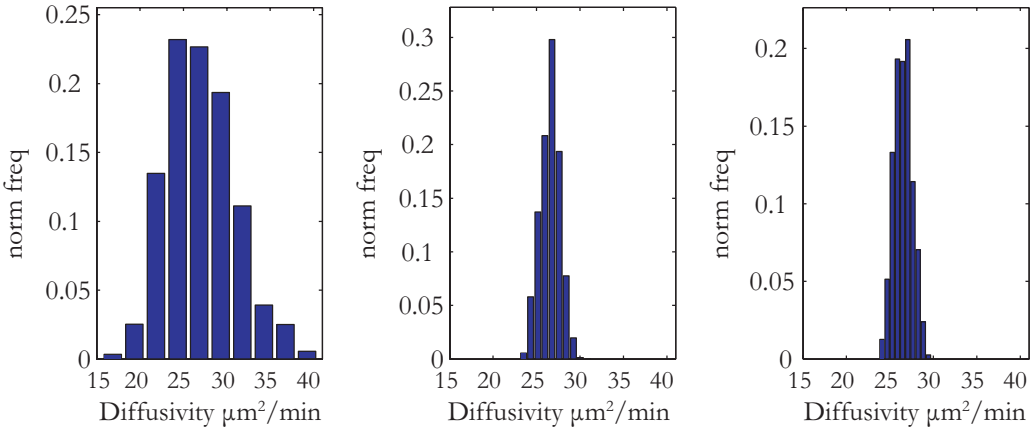

Figure 4.5: **An example of an over-fitted ABC posterior.** Averaging of repeat simulations results in a tighter distribution around the identified parameter value. The simulation was repeated (a) once, (b) 50, (c) 100 times before averaging. The true diffusivity parameter value is 25. In (b) and (c) the true parameter value can be seen to move into the tail of the posterior distribution.

Simulation studies showed that employing this step resulted in a tighter posterior distribution around the true parameter values as demonstrated in Figure 4.5. This suggests less uncertainty in the identified parameter values. However, it was not possible to justify this mathematically for the simple reason that, in general,

$$\mathbb{E}\left(\chi_{[-\epsilon,\epsilon]}(S(x|\theta) - S(y))\right) \neq \chi_{[-\epsilon,\epsilon]}(\mathbb{E}\left(S(x|\theta)\right) - S(y)) \tag{4.18}$$

where χ is the indicator function*. This is further demonstrated in Figure 4.6.

* $$\chi_{[-\epsilon,\epsilon]}(s) = \begin{cases} 1 & \text{if } -\epsilon \leqslant s \leqslant \epsilon \\ 0 & \text{otherwise} \end{cases}$$

Some implementations of ABC-SMC such as that of Toni et al. [121] do employ repeat simulation from each parameter set. However, in this case each simulation is considered in its own right without being averaged and an initial weight is given to the parameter set according to what proportion of simulated observations fall within the current error threshold.

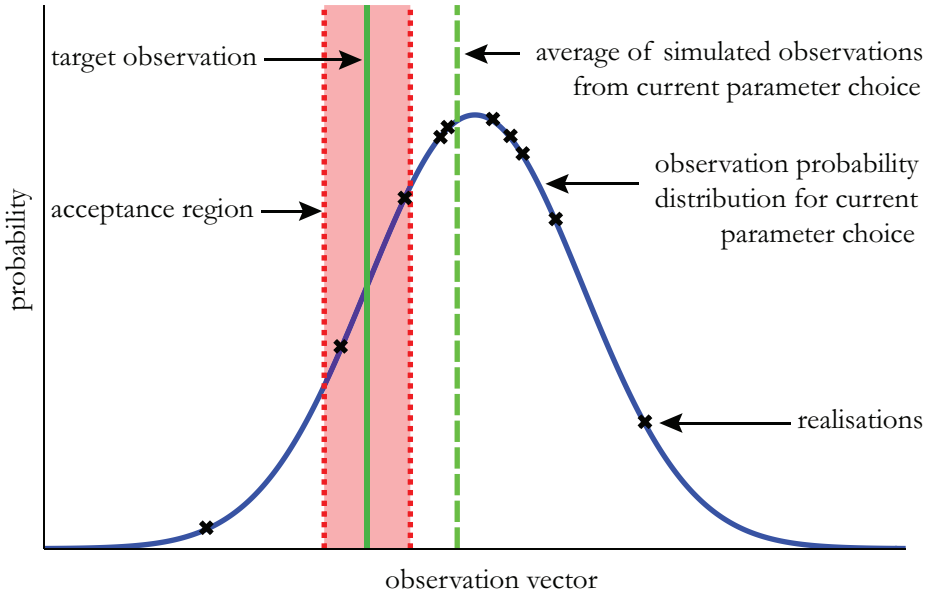

Figure 4.6: **Explanation of over-fitting the posterior distribution.** Illustration of how averaging repeated simulations from a single parameter set will lead to over-fitting. Parameter sets which are capable of creating the target observations are wrongly rejected. The target observation value is represented by the solid green line with dotted red shaded region indicating the current error margins for acceptance. The blue line is the probability density for simulated observations from the parameter set currently under consideration and the black crosses show ten realisations of this process. After taking the average of these realisations (dashed green line) this parameter set is rejected. Whereas, in fact, the crosses that are within the error margin show that this parameter set could have produced the target observations. Following this through, by increasing the number of realisation and decreasing the error margins the accepted parameter values will converge on a single value which would be equivalent to a maximum likelihood identification. This fatally compromises the identification of a posterior distribution which allows for the fact that the target observation may in fact be an outlier observation for the true system.

4.3 Validation of the new identification framework on simulated data

In any new modelling context, the proposed identification method must be validated to show that sensible and reliable results can be obtained. Populational modelling of neutrophils *in vivo* using ABC methods has not been implemented before, so a rigorous validation procedure must be carried out. Since the dynamics of neutrophils *in vivo* cannot be controlled or independently ascertained by others means, it is necessary to do this validation *in silico*.

The ABC-SMC algorithm will be used to estimate model parameters and select the optimal model for describing neutrophil migration. To give a brief résumé, ABC-SMC estimates model parameters by (a) randomly sampling parameter sets from a prior distribution; (b) simulating the process with those parameters; and (c) accepting the parameter set as a sample from the posterior distribution if the simulation results are within the current error tolerance from those of the experimental data. The error tolerance is sequentially decreased over a number of iterations. Moreover, the ABC-SMC algorithm is extensible to model selection by allocating each competing model an index value and treating this as part of the parameter vector to be estimated.

In order to test the ability of the developed framework to identify parameters and discriminate between pure-diffusion and drift-diffusion processes, it was applied to simulated cell data with a range of model and parameter values. All simulations began with cells in the configuration of experimental data, and observations were collected to correspond to those of the experimental data. Selected results are shown in Figure 4.7 where the diffusivity coefficient was set at $25\,\mu m\,min^{-1}$ in and the drift coefficient was varied from 0 to $0.4\,\mu m\,min^{-1}$. For higher values of drift, the correct model continued to be identified with 100% confidence and the parameter estimation had comparable accuracy and uncertainty to the results shown for $0.4\,\mu m\,min^{-1}$. The effect of increasing or decreasing the diffusivity was only to correspondingly increase or decrease the uncertainty in the parameter estimates. Likewise increasing or decreasing the number of cells had the inverse effect of decreasing or increasing the uncertainty respectively. However, for the data-driven estimates to be valid it was necessary to simulate with the same number of cells that are observed in the data.

To assess the quality of the model prediction, the Bayes Factor for Model 2, drift-diffusion, with respect to Model 1, pure-diffusion, was calculated for each of the simulations. The values are shown in Figure 4.11(b). For drift values below approximately $0.2\,\mu m\,min^{-1}$ model selection cannot be made with any confidence. Between

0.2 and 0.3 $\mu m\,min^{-1}$ there is substantial to very strong evidence for Model 2. Over 0.3 $\mu m\,min^{-1}$ the evidence for the Model 2 is overwhelming. Although there is insignificant evidence for Model 2 for drift values below approximately 0.2 $\mu m\,min^{-1}$, this does not, however, mean that there is evidence in favour of Model 1. It simply means that the models cannot be distinguished for that level of drift.

In order to further investigate the ability to detect the true model when the true simulated model is Model 1, the model selection algorithm was applied repeatedly with an increasing lower limit of the drift parameter prior. With this lower limit set to zero, Model 2 allows any drift value of zero or above, on the other hand if the lower limit is set to 0.5 $\mu m\,min^{-1}$, say, Model 2 requires a correspondingly significant level of drift. The results of this are shown in Figure 4.11(a).

The validation results shown in Figure 4.7 are for coefficients comparable to those that were estimated for the experimental data as will be reported in Section 4.4. The results suggest that a pure-diffusion process will be correctly identified with a confidence that grows rapidly as the minimum amount of drift considered significant is increased. But in any case the evidence for the correct pure-diffusion model will be substantial if any drift value is allowed. Also, a drift-diffusion process will be correctly identified if the drift coefficient is 0.2 $\mu m\,min^{-1}$ or above. The evidence for the correct drift-diffusion model rises rapidly as the true drift coefficient rises until the evidence is overwhelming for drift values above 0.3 $\mu m\,min^{-1}$.

	estimation time (mins)	
T	pure-diffusion	drift-diffusion
1	>240	>240
2	>240	>240
3	135	147
4	117	130
5	124	141
6	132	142
7	142	161
10	179	215

Table 4.2: **Performance of Algorithm 4.2 against number of iterations, T.** In each case an exponentially decaying error tolerance schedule was applied. Values greater than 240 mean that the algorithm was aborted after 4 hours without completing.

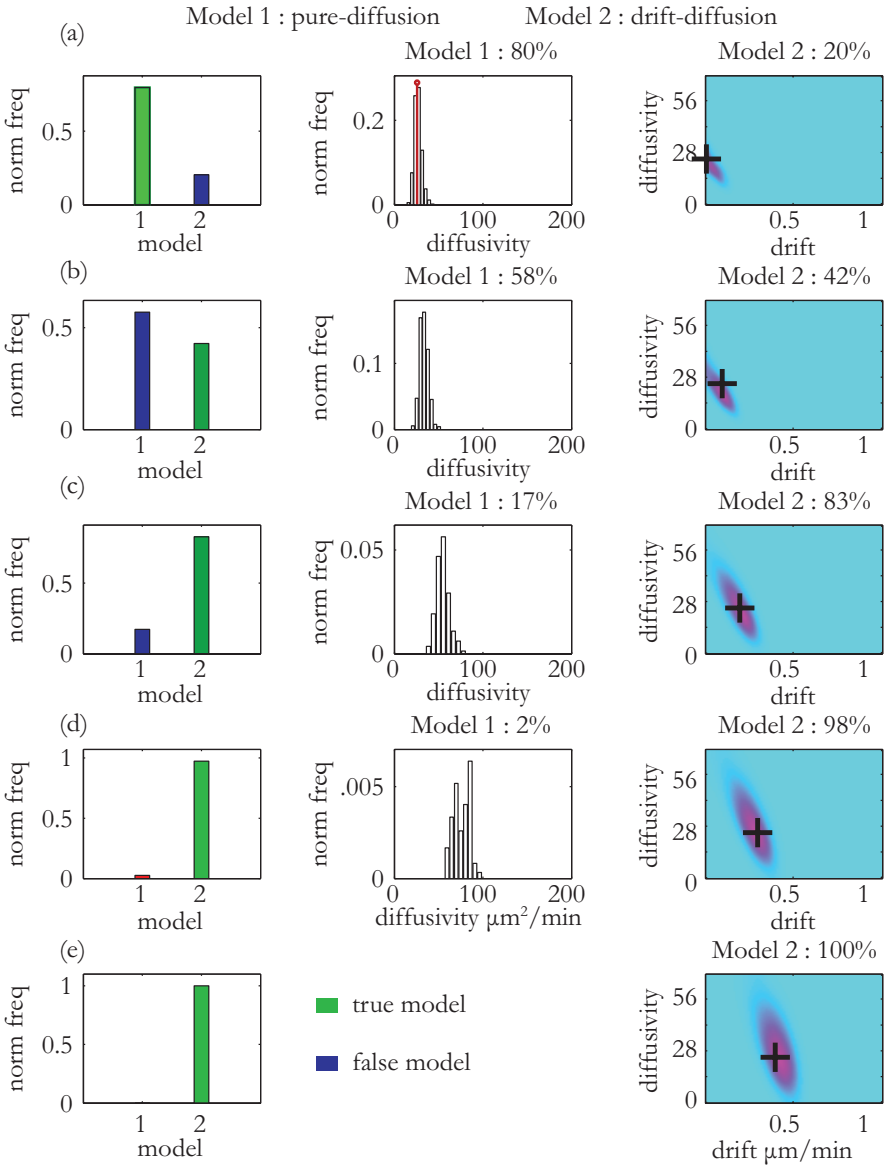

Figure 4.7: **Validation of the identification framework on simulated data.** In order to demonstrate the confidence with which the correct model is identified it was tested on simulated data for varying values of diffusivity and drift. The results shown are for a diffusivity, $D = 25\,\mu m^2\,min^{-1}$, and with drift values, b, of (a) 0, (b) 0.1, (c) 0.2, (d) 0.3, (e) $0.4\,\mu m\,min^{-1}$. Column 1 shows the posterior model marginal distribution, where Model 1 is the pure-diffusion and Model 2 the drift-diffusion model. Column 2 shows the identified posterior distribution of diffusivity ($\mu m^2\,min^{-1}$) for Model 1. Column 2 shows the identified joint posterior distribution for drift ($\mu m\,min^{-1}$) and diffusivity ($\mu m^2\,min^{-1}$) for Model 2. The (a) stem and (b)-(e) crosses show the true parameter values. Corresponding Bayes factor analysis is shown in Figure 4.11

4.4 Identification of zebrafish neutrophil dynamics during the resolution phase of inflammation

Understanding the processes by which inflammatory neutrophils resolve themselves away from a site of inflammation is vitally and urgently needed. Inflammation is a key process in the mammalian immune systems but one that may become dysfunctional. Compromised inflammation resolution is implicated in the causes of chronic inflammatory diseases such as COPD, asthma and rheumatoid arthritis. Understanding the drivers of the inflammation resolution process and why it may fail will help the search for more effective and hopefully cheaper interventions into the associated conditions. It has generally been assumed that this process is the mirror image of chemotaxis, where neutrophils are drawn into areas of infection or tissue damage by gradients of chemotactic cues. Indeed, efforts are under way to identify cues that drive neutrophils away by the reverse process, fugetaxis. Rigorous data driven justification of this assumption or its alternative, have however, up to now been lacking.

In order to address this open question, the preceding work of this chapter has developed an identification framework for cell migration dynamics. The framework uses ABC methods and cell position distribution data. The framework can now be applied to characterise the migration dynamics of zebrafish neutrophils *in vivo* with a view to confirming the result obtained in Chapter 3 on a more comprehensive basis.

4.4.1 Method

The ABC-SMC parameter estimation and model selection algorithms were applied to the same experimental neutrophil data from 6 zebrafish larvae analysed in Chapter 3 (Dataset A) and also a second dataset (Dataset B) obtained separately under the same experimental conditions. A full description of the data is given in Chapter 3, but briefly: A wound is induced to the tailfin of zebrafish larvae at 3 days post fertilisation (dpf). At 4 hours post injury, green fluorescent protein (GFP) expressing neutrophils in the vicinity of the wound are photoconverted to fluoresce red. These cells are then observed every 5 minutes over a total over 980 minutes which corresponds to the resolution phase of inflammation. The main results presented relate to Dataset A. Results for Dataset B are presented briefly to confirm the generality of the findings.

The candidate models are: Model 1, the pure-diffusion model and Model 2, the drift-diffusion model. These models were introduced in Section 4.2.1 and are recapped on page 82. Whilst Algorithm 4.2 estimates both model parameters and model index, the parameter estimates can be unreliable for a model that has a minority representation in the samples. For this reason, Algorithm A.1 was applied

first for the individual models. Algorithm 4.2 was then applied with both models as candidates in order to discriminate between them. From the resulting posterior marginal distribution over the two models a Bayes factor was calculated which was then interpreted according to Table 4.1. The default prior distribution for Model 2 allows drift values of greater than or equal to zero. In order to further explore the model preference the model identification algorithm, Algorithm 4.2 was applied several times with increasing minimum allowed drift values as was done for the validation in Section 4.3. The Bayes factor was calculated and plotted in each case. Finally, the preferred estimated model was simulated repeatedly and the distribution of resulting observations was compared to the original experimental data driving the estimation process.

Pure diffusion and drift-diffusion models

$$x_{t+1}^{(i)} = \max(0, x_t^{(i)} + b_{\text{out}}\Delta t + \omega_t^{(i)}\sqrt{2D\Delta t}), \tag{4.19}$$

where

- $x_t^{(i)}$ is the position of the i^{th} neutrophil at time t;
- b_{out} is a bias velocity away from the wound;
- $\omega_t^{(i)} \sim \mathcal{N}(0,1)$ are a family of independent white noise processes;
- D is the underlying diffusivity constant or magnitude of random movement of the neutrophils;
- Δt is the time increment;
- The use of the $\max(0,.)$ function implies that cells do not migrate through the boundary and if they reach the boundary they remain there until the next time increment.

With $b_{\text{out}} = 0$ this is designated as Model 1, and with $b_{\text{out}} \geqslant 0$, Model 2.

Model 1 & 2: pure-diffusion & drift-diffusion

4.4.2 Results

The result of applying Algorithm A.1 to Dataset A for Model 1, the pure-diffusion model, which has a single parameter, is shown in Figure 4.8. The MAP parameter estimate is

$$D = 25\,\mu\text{m}^2\,\text{min}^{-1}$$

with 90% confidence interval of $19 - 34\,\mu\text{m}^2\,\text{min}^{-1}$. The diffusivity is related to observed displacement via (4.6). Repeated simulation of this equation for a single

cell with $D = 25\,\mu\text{m}^2\,\text{min}^{-1}$ yielded a mean observed displacement of $5.5\,\mu\text{m}\,\text{min}^{-1}$ and a 90% confidence interval of between 0 and $14\,\mu\text{m}\,\text{min}^{-1}$. This is exactly in keeping with other reported neutrophil speeds in zebrafish (Mathias et al. [85] and Walters et al. [128] report mean speeds of approx $10\,\mu\text{m}\,\text{min}^{-1}$), and also comparable with measured speeds of human neutrophils [78].

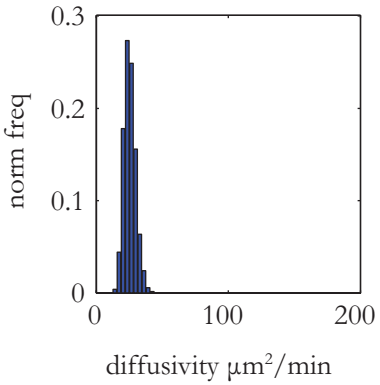

Figure 4.8: **Zebrafish neutrophil data estimation results for Model 1, the pure-diffusion model.** Algorithm A.1 for parameter estimation was applied with Model 1, the pure-diffusion model as the dynamic model. The MAP parameter estimate for the single diffusivity parameter is $D = 25\,\mu\text{m}^2\,\text{min}^{-1}$.

The results of Algorithm A.1 for Model 2, the drift-diffusion model, which has two parameters, are shown in Figure 4.9. The MAP parameter set estimate is

$$b_{\text{out}} = 0.05\,\mu\text{m}\,\text{min}^{-1}, D = 19\,\mu\text{m}^2\,\text{min}^{-1}$$

with 90% confidence intervals $0.005 - 0.15\,\mu\text{m}\,\text{min}^{-1}$ and $7 - 36\,\mu\text{m}^2\,\text{min}^{-1}$. These values imply mean observed displacement of $4.8\,\mu\text{m}\,\text{min}^{-1}$ and a 90% confidence interval of between 0 and $10\,\mu\text{m}\,\text{min}^{-1}$.

For the purpose of model selection, Algorithm 4.2 was applied with both Model 1 and Model 2 as candidates. The result is shown in Figure 4.10. The marginal distribution over the models was 88% for Model 1, the pure-diffusion model, and 12% for Model 2, the drift-diffusion model. This yielded a Bayes factor for Model 1 with respect to Model 2,

$$B_{12} = 7.3$$

Thus there was substantial evidence for Model 1 over Model 2. It should be noted that whilst Model 2 is the drift-diffusion model, nevertheless the model allows for drift values of zero or arbitrarily close to zero. To test the model selection result further Algorithm 4.2 was repeatedly applied, with increasing lower limits on the prior range for the drift parameter in Model 2. In other words, Model 2 becomes

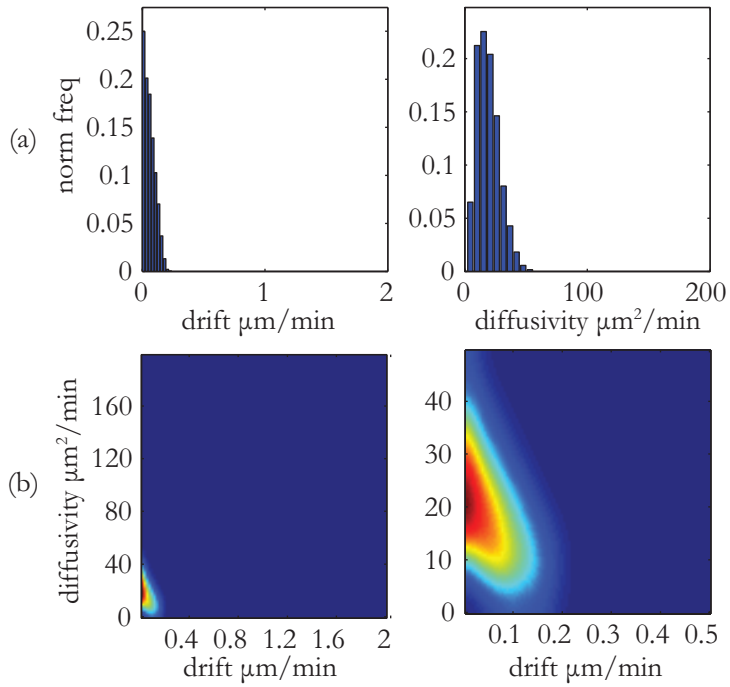

Figure 4.9: **Zebrafish neutrophil data estimation results for Model 2, the drift-diffusion model.** Algorithm A.1 for parameter estimation was applied with Model 2, the drift-diffusion model as dynamic model. (a) Posterior distribution over the individual parameters. (b) Joint distribution for the two parameters. The MAP parameter set is drift, $b_{\text{out}} = 0.05\,\mu\text{m}\,\text{min}^{-1}$, diffusivity, $D = 19\,\mu\text{m}^2\,\text{min}^{-1}$.

the drift-diffusion model with minimum drift great than b_0 for various values of b_0. The results of this are shown in Figure 4.11(c). Comparison to Table 4.1 shows that there is substantial evidence for Model 1, the pure-diffusion model, for all values of b_0 and very strong evidence when $b_0 > 0.08\,\mu\text{m}\,\text{min}^{-1}$.

The analysis was also performed on the second dataset, Dataset B, to check consistency of results collected in an identical manner to the main dataset. The main results were confirmed and even strengthened by this dataset, as shown in Figures 4.13 and 4.11(d).

4.4.3 Discussion

In order to validate the estimation framework, the ABC-SMC algorithm was first applied *in silico* in order to verify that discrimination could be made between different types of drift-diffusion process and correctly identify parameters within a reasonable confidence interval. This is an important step when applying ABC-SMC in a new context, especially in relation to the validity of model selection [109]. When the ABC-SMC identification algorithm was applied to a pure-diffusion process and the

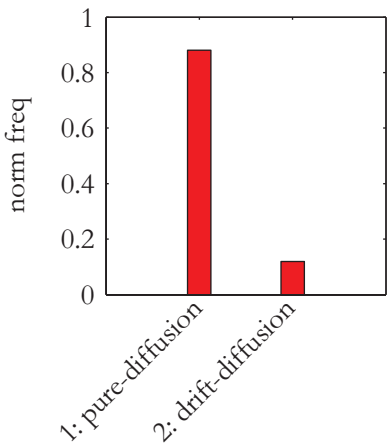

Figure 4.10: **Zebrafish neutrophil data model selection result.** Algorithm 4.2 was applied with both models as candidates to identify the preferred model. Model 1 is the pure-diffusion model, Model 2 the drift-diffusion model.

Bayes factor calculated, there was substantial evidence for the correct model. The alternative model had a low representation in the marginal posterior distribution and in addition the MAP drift coefficient was zero. When the alternative model was forced to look for drift values greater than a small tolerance value, the model selection confidence became stronger still. Moreover, when the ABC-SMC algorithm was applied to a simulated drift-diffusion process there was substantial evidence for the correct model for values over $0.2\,\mu m\,min^{-1}$ and overwhelming evidence for values over $0.3\,\mu m\,min^{-1}$. Hence, the validation procedure demonstrated that the ABC-SMC algorithm can accurately detect drift if a significant drift is present.

To investigate the reverse migration behaviour of the neutrophils, the estimation framework was applied to experimental data from zebrafish larvae during the resolution phase of an inflammation episode, using the same data that was studied in Chapter 3 (Dataset A). Substantial evidence was found for the pure-diffusion model which was represented by 88% of the posterior samples. The MAP drift coefficient of the alternative drift-diffusion model was estimated to be zero. When the drift-diffusion model was restricted to non-zero drift values, it was found that the evidence for the pure-diffusion model increased and was overwhelming if the drift was constrained to be not less than $0.1\,\mu m\,min^{-1}$. The estimation framework was also applied to data from a second independent but identical experiment (Dataset B). In this case there is strong evidence for the pure-diffusion model which was represented by 92% of samples from the posterior distribution. And if the drift-diffusion was constrained to take values not less than $0.06\,\mu m\,min^{-1}$ then the Bayes factor analysis provides conclusive evidence for the pure-diffusion model.

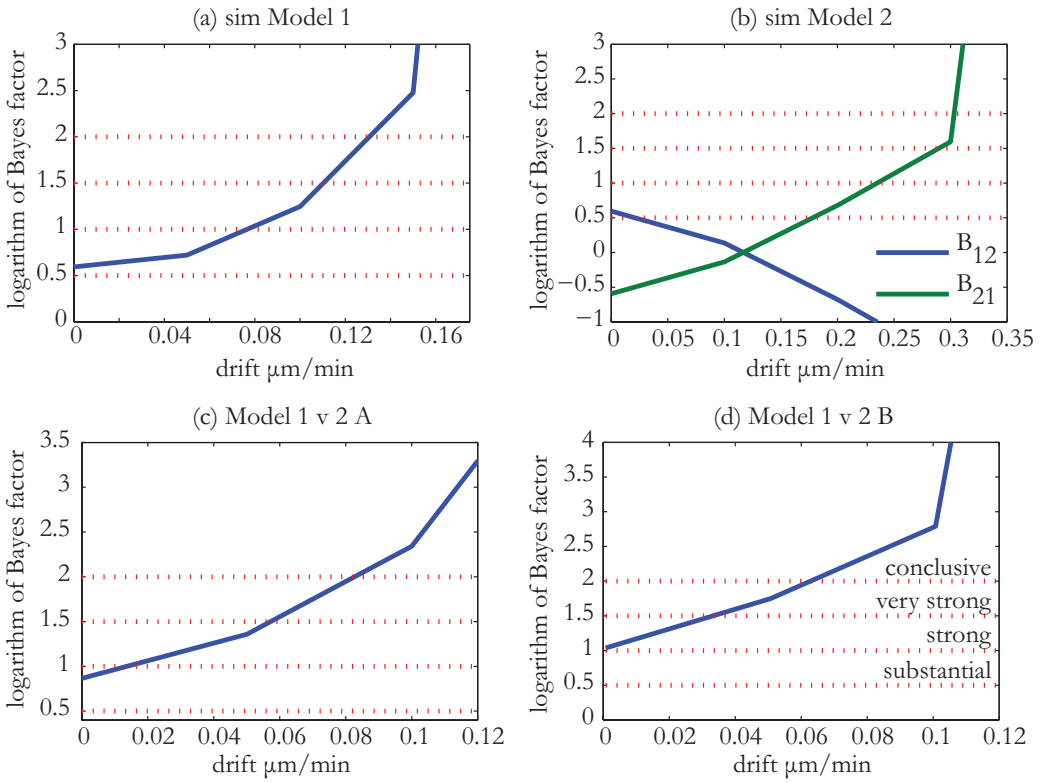

Figure 4.11: **Bayes factor analysis.** Model 1 is the pure-diffusion model. Model 2 is the drift-diffusion model. The Bayes factor, B_{12} for Model 1 with respect to Model 2 is plotted as a blue line on a logarithmic scale against the lower limit on the prior for the drift coefficient in Model 2. (b) also shows B_{21} in green.
(a) Simulated pure-diffusion data: Even when Model 2 is allowed a drift coefficient of zero, there is substantial evidence for the correct Model 1. As the lower allowed value is increased the evidence increases rapidly until, for drift greater than approximately $0.13\,\mu m\,min^{-1}$, the evidence is overwhelming.
(b) Simulated drift-diffusion data: When the drift coefficient is zero, there is substantial evidence for the correct Model 1. For drift values between zero and approximately $0.2\,\mu m\,min^{-1}$ model selection cannot be made with any confidence. Between 0.2 and $0.3\,\mu m\,min^{-1}$ there is substantial to very strong evidence for Model 2. For values over $0.3\,\mu m\,min^{-1}$ the evidence for the drift diffusion model is overwhelming.
(c) Dataset A: Similarly to the simulate data in (a), when Model 2 is allowed a drift coefficient of zero, there is substantial evidence for Model 1. As the lower allowed value is increased the evidence increases until, for drift greater than $0.1\,\mu m\,min^{-1}$, the evidence is overwhelming.
(d) Dataset B: For this dataset, there is already strong evidence for Model 1 when any drift value greater than zero is considered. If drift values less than approximately $0.06\,\mu m\,min^{-1}$ are considered insignificant, then there is conclusive evidence for Model 1, the pure-diffusion model.

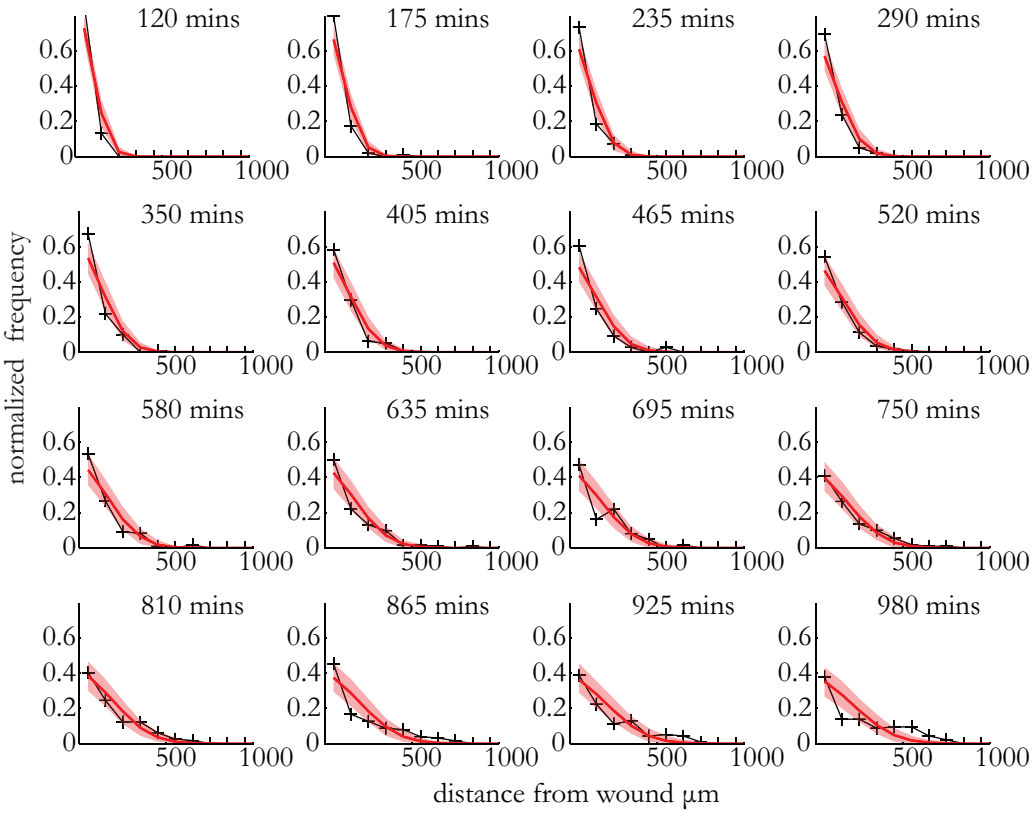

Figure 4.12: **Simulation comparisons of the estimated pure-diffusion model, Model 1, to the experimental data.** One thousand simulations were made of the estimated model using the MAP parameter estimate. The output is shown for representative times throughout the experiment. The red line shows the mean predicted distribution of cell positions and the shaded red region is the 90% confidence interval. The experimental observations are plotted in black for comparison.

These results strongly suggest that even in the event that the model was incorrectly identified and a small non-zero drift coefficient was present at the limiting value of $0.1\,\mu\text{m}\,\text{min}^{-1}$, this would only make a very small contribution to migration (an approx. $100\,\mu\text{m}$ mean shift over the 980 minute span of the experiment). This on its own would still leave the cells in the vicinity of the wound. Therefore, it is the identified diffusivity arising from the inherent migratory patterns of the neutrophils that is contributing the major part of the motility. This is also why cells are often seen to change direction during migration.

It is apparent from inspecting and comparing repeated simulations of the pure-diffusion model and the *in vivo* observations (Figure 4.12) that, whilst the fitted pure diffusion model gives a broadly accurate representation of the data, nevertheless it does not fully explain the response. The identified pure-diffusion model fits the experimental data well at earlier timepoints, but at later times (655-980 minutes) the

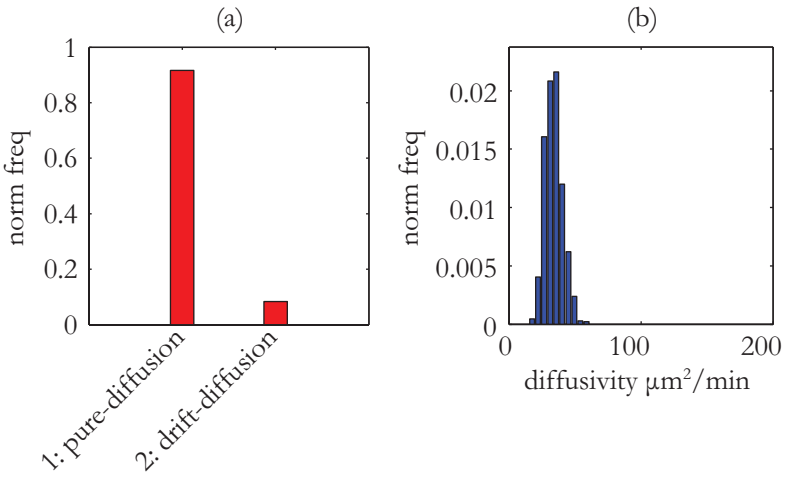

Figure 4.13: **Dataset B: model identificaion and parameter estimation.** Algorithm 4.2 was applied to an additional experimental dataset with Models 1 and 2 as candidates. Model 1 had a 92% representation in the posterior model marginal distribution. Algorithm A.1 was applied for Model 1 to estimate the Diffusivity parameter as $36\,\mu m^2\,min^{-1}$ and a 90% confidence interval of $27 - 50\,\mu m^2\,min^{-1}$.

predicted distribution of cells does not precisely describe the observed cell behaviour: the cell count at the wound is greater than predicted and less than predicted at $\sim 250\,\mu m$ from the wound. This suggests that the model is incompletely capturing the nature of neutrophil movements; specifically it suggests that neutrophils do not move away from the wound as easily as the pure diffusion model would suggest. The next chapter will consider the possible causes of this discrepancy and ways in which the model can be enhanced to obtain a more faithful representation of the data. It is also important to do this to ensure that any significant drift present in the dynamics has not been masked by shortcomings of the modelling process.

4.5 Conclusion

In this chapter a novel framework for parameter estimation and model discrimination for cell migration dynamics has been developed. The immediate aim has been the identification of neutrophil migration processes in inflammation resolution. However, the framework can also be applied in other cell migration settings. The candidate models can be quite general in nature as long as they are capable of being simulated. The framework makes use of approximate Bayesian computation (ABC) methods. ABC is being used increasingly in a wide range of research areas. It has been applied previously in a related context to estimate chemoattractant fields from distribution of straightness indices of zebrafish neutrophil tracks. However, it has not been applied

as here, to analyse and identify dynamics of cell migration and in particular of neutrophil migration in the inflammation resolution phase.

The identification framework was validated on simulated data comparable to the experimental data and it was shown that accurate and reliable results are achieved. The framework was then applied to data from two separate experiments yielding the conclusion that the process by which neutrophils migrate away from the wound in the zebrafish is best characterised as a stochastic redistribution without directional bias. This finding has important implications for further research informing future therapeutic interventions for chronic inflammation related illnesses. In particular, it is already guiding researchers to look for the mechanisms whereby neutrophils 'switch off' from the signals recruiting them to an inflammation event rather than to look for signals that are repelling them from the inflammation region.

Chapter 5

Spatiotemporal variability in the neutrophil migration model

Chapter 4 of this thesis developed a Bayesian framework for model identification and parameter estimation of cell population migration dynamics. The framework has a general applicability and was utilized to solve the particular problem of estimating zebrafish neutrophil migration dynamics in the resolution phase of inflammation, which is the focus application for this thesis. Using pure-diffusion and drift-diffusion models as candidates, it was shown that the pure-diffusion model was the preferred model to describe the zebrafish neutrophil data. In the modelling of any complex biological system, simplifying assumption must inevitably be made to allow analysis to proceed. Intrinsic to both the models developed in Chapter 4 were the following assumptions:

1. The sub-population of photoconverted neutrophils were all in inflammation resolution mode throughout the experiment.

2. The neutrophils were moving in a homogeneous environment.

The first assumption may be justified by the fact that the photoconverted cells were already in the vicinity of the wound - they had completed their recruitment migration prior to the start of the observations. Also, the photoconverted cells are subsequently seen to spread away from the wound in a persistent way. In other words, this is not an artefact of the random redistribution of cells that still have an attractive bias towards the wound region. The neutrophils are located in zebrafish tissue, not in the bloodstream. Clearly this is a complex environment (especially in contrast to the cell environment of *in vitro* studies) but the second assumption is that this complexity is uniform throughout the region of interest.

In this chapter, a limited relaxation of each of these assumptions is considered in turn, followed by a relaxation of both simultaneously. The objective of this is to achieve a more complete description of the data. In particular, it is intended to safeguard against the situation where simplifying assumptions bias the answer to the key question: whether the migration of neutrophils during inflammation resolution is a passive or a directed process. To pose this as a new question: Could a drift-diffusion process be mistakenly identified as a pure-diffusion process if certain complexities of the system are ignored? This question will be answered via the development of novel models of the system which include additional complexity. At that same time the development of these models illustrates the broader applicability of the estimation framework for varied types of cell migration model.

5.1 Extending the model to receptor depletion

It is clear that at the start of an inflammation cycle, neutrophils migrate towards the wound which is inducing the inflammation. At some later time they are observed to migrate away from the wound (Figure 3.4).

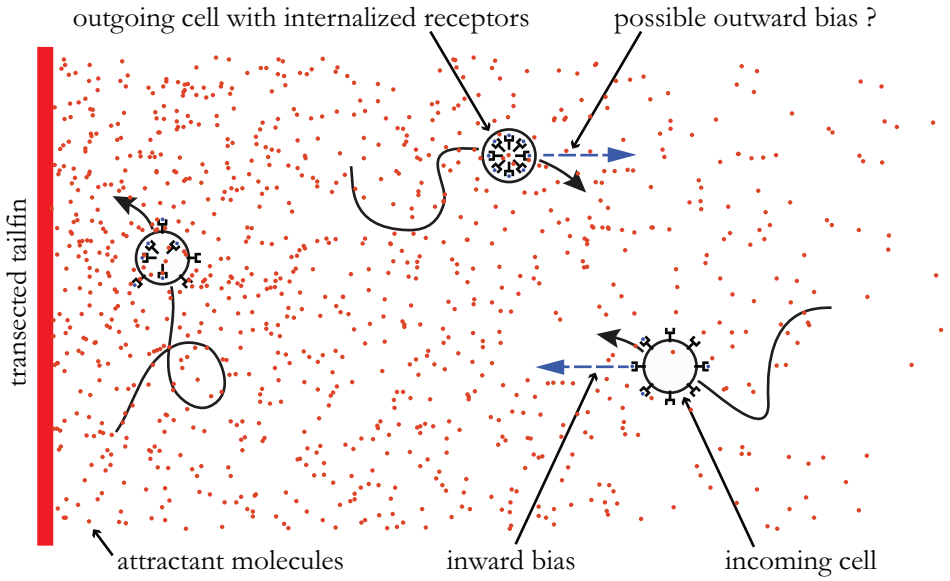

Figure 5.1: **The receptor depletion model.** A neutrophil that has newly entered the field of attractant ligand molecules has its full complement of receptors which are available on the cell surface for binding ligand. The resulting downstream internal signalling affects the cells migration behaviour, creating a bias towards the wound. Bound receptors also become internalised. This results in a gradual weakening of the response. Eventually, when the majority of receptors are internalised, the cell loses its response to the attractant field and it migrates randomly. Possibly it now recognises other guidance cues which direct it away from the wound region.

However, the experimental results of Chapter 3 showed that when the first cells begin their outward migration, other cells are still being actively recruited towards the wound region (Figure 3.5). This suggests that there is not a global switching off of the recruiting chemoattractant influence but rather a desensitization of individual cells to the effect of this chemotactic field. This observation in turn prompts the postulation of a change of dynamic mode of the neutrophils: from a recruitment mode, when they are responding to attractants guiding them to the wound region, to a inflammation resolution mode, when they no longer respond to these attractants and they move either randomly or under the effect of new attractive or repulsive biases away from the wound.

It is known that neutrophils are able to orientate themselves in a gradient of chemoattractant [135] and then to migrate via chemotaxis in the direction of orientation [136]. Furthermore, they are able to prioritize distinct attractant sources in a sequential manner [44], switching from a initial field to a subsequent one. In eukaryotic cells, such as neutrophils, the signalling pathway from chemoattractant gradient to molecular cell motors is a complex one which begins with the binding of chemoattractant ligand molecules to protein receptors on the cell surface [1]. It has been suggested that the number of different proteins involved in this signalling process is likely to be of the order of 100 [56]. Various models have been proposed to link receptors dynamics to cell migration dynamics [32, 56, 122] further details of some of these models has already been given in Chapter 2. In this section, in order to describe the change in neutrophil mode from recruitment to inflammation resolution, a simplified description of receptor dynamics is incorporated into the drift-diffusion models.

5.1.1 Model description

The basic models considered in Chapter 4 were extended to capture the change of dynamic mode from inflammation recruitment to its resolution. First, a number assumptions need to be stated.

- The attractant forms a steady state linear field, $A(x)$. The gradient is the negative of the rate, w, at which attractant is being produced at the origin. The field reduces to zero at the boundary of its domain, $x = L$, i.e.

$$A(x) = \max\left(0, w\left(\frac{L-x}{L}\right)\right) \tag{5.1}$$

  This assumption can be justified by assuming that the attractant molecules are fast diffusing (relatively to the migration rates of the neutrophils themselves) and by then solving the suitably posed diffusion problem (see Appendix C).

- All cells are assumed to have a fixed total number of receptors R_M. Whilst in the attractant field, the number of available receptors decays at a rate $\hat{\lambda}$ in proportion to the local field concentration. This is due to bound receptor-ligand complexes being internalised by the cell. It is assumed that R_M is large enough so that the number of receptors can be considered to be continuous without detriment to the model. It is also assumed that there is no recycling of receptors taking place, that is, once they are internalised they remain so.

- A cell's attraction or bias towards the wound is the product of a fixed bias from the field b_{in} (because of its constant gradient) and the current proportion of receptors that are available, R_t. An outward bias b_{out} is also allowed for. This takes over from the inwards bias as the receptor depletion reaches its conclusion. This will take the value zero if the motion away is purely diffusive.

Thus the model can be written down as follows:

$$x_{t+1}^{(i)} = \max\left(0, x_t^{(i)} + \left(b_{\text{out}} - R_t^{(i)} b_{\text{in}}\right)\Delta t + \omega_t^{(i)}\sqrt{2D\Delta t}\right) \tag{5.2}$$

$$R_{t+1}^{(i)} = R_t^{(i)} - \hat{\lambda} A(x_t^{(i)}) R_t^{(i)} \Delta t \tag{5.3}$$

Where, as before, D is the diffusivity of the cells and $\omega_t \sim N(0,1)$ is a white noise process representing the random nature of this underlying motion. $\hat{\lambda}A(x_t^{(i)})$ is thus the instantaneous rate of exponential decay of available ligand receptors for cell i at time t. The assumptions of this model are illustrated in Figure 5.1. It can be seen that when (5.1) is substituted into (5.3), w and $\hat{\lambda}$ form a composite parameter. So $\lambda = w\hat{\lambda}$ will be treated as a free single parameter.

When $b_{\text{out}} = 0$ this is denoted as the pure-diffusion-depletion model (Model 3) and with $b_{\text{out}} \geqslant 0$ the drift-diffusion-depletion model (Model 4). They key question remains as before: is the experimental data best described by Model 3 or by Model 4? In other words, are the cells moving away from the wound in a directed way, as a results of a new chemotactic influence, or are they migrating in a passive way, simply as a results of their inherent random search patterns?

5.1.2 Estimation results

Algorithm A.1 was applied to Model 3, the pure-diffusion-depletion model. The results of the estimation are shown in Figure 5.2. It can be seen that the support of the posterior distribution is almost as large as that for the prior. There is also some correlation between parameters, in particular between diffusivity and drift and between diffusivity and depletion rate. However, the data has clearly informed the posterior distribution and there is a definite MAP region centred on the parameter

Receptor depletion models

$$x_{t+1}^{(i)} = \max\left(0, x_t^{(i)} + \left(b_{\text{out}} - R_t^{(i)} b_{\text{in}}\right)\Delta t + \omega_t^{(i)}\sqrt{2D\Delta t}\right) \tag{5.4}$$

$$R_{t+1}^{(i)} = R_t^{(i)} - \lambda \max\left(0, \frac{L - x_t^{(i)}}{L}\right) R_t^{(i)} \Delta t \tag{5.5}$$

where

- $x_t^{(i)}$ is the position of the i^{th} neutrophil at time t;
- b_{out} is a bias velocity away from the wound;
- $R_t^{(i)}$ is the proportion of receptors available for the i^{th} cell at time t;
- b_{in} is a bias velocity towards the wound;
- $\omega_t^{(i)} \sim \mathcal{N}(0,1)$ are a family of independent white noise processes;
- D is the underlying diffusivity constant or magnitude of random movement of the neutrophils;
- Δt is the time increment;
- λ is the composite depletion constant described above;
- L is the nominal range of the chemoattractant field.

With $b_{\text{out}} = 0$ this is designated as Model 3, and with $b_{\text{out}} \geqslant 0$, Model 4.

Model 3 & 4: pure-diffusion-depletion & drift-diffusion-depletion

vector given by

$$b_{\text{in}} = 2.1\,\mu\text{m}\,\text{min}^{-1}, \quad D = 83\,\mu\text{m}^2\,\text{min}^{-1}, \quad \lambda = 0.0046\,\text{min}^{-1}.$$

The 90% confidence intervals calculated from the parameter marginals are 0.9-4.3 μm min^{-1}, 39-172 μm^2 min^{-1}, $2.4 - 9.2 \times 10^{-3}$min^{-1}, respectively. The drift and diffusivity values are physically plausible and of the same orders of magnitude as those identified by the regression methods in Chapter 3.

Algorithm A.1 was also applied to Model 4, the drift-diffusion-depletion model. The results of the estimation are shown in Figure 5.3. The MAP estimate is,

$$b_{\text{in}} = 2.8\,\mu\text{m}\,\text{min}^{-1},\ b_{\text{out}} = 0.3\,\mu\text{m}\,\text{min}^{-1},\ D = 120\,\mu\text{m}^2\,\text{min}^{-1},\ \lambda = 0.0025\,\text{min}^{-1}$$

and the 90% confidence intervals calculated from the parameter marginals are 1.0-4.7 μm min^{-1}, 0.01-1.6 μm min^{-1}, 42-190 μm^2 min^{-1}, $0.8 - 7.2 \times 10^{-3}$min^{-1}, re-

spectively. It can be seen that, for this model, the diffusivity parameter and to some extent the inward bias parameter have convergence issues. Also there is a correlation between diffusivity and inward bias which suggest a possible identifiability problem with the model for this data. Having estimated parameters for the

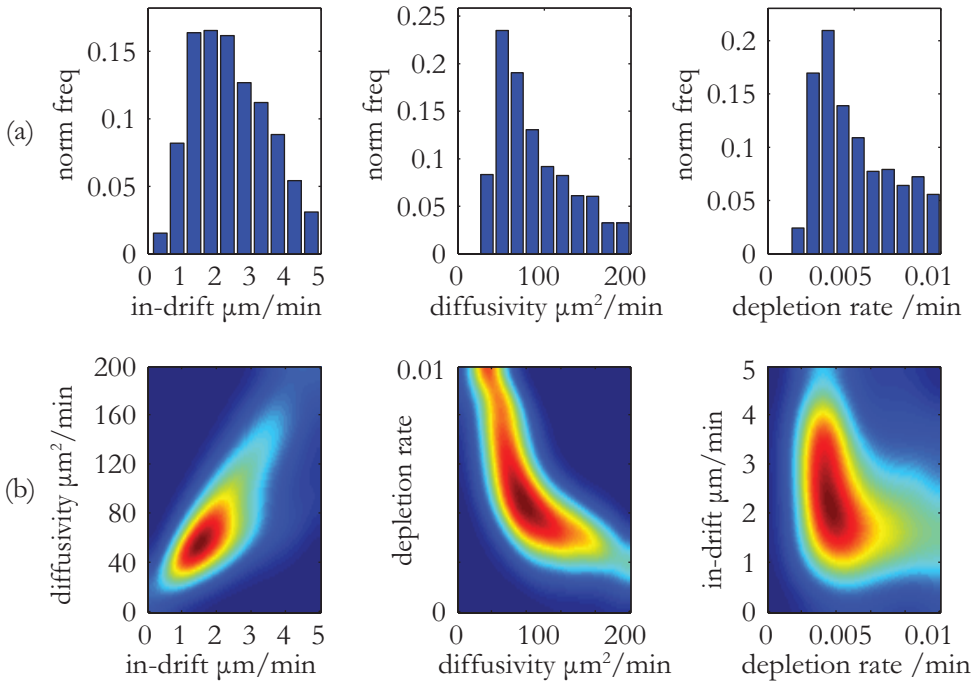

Figure 5.2: **Estimation results for Model 3, the pure-diffusion-depletion model.** (a) Posterior distribution over the individual parameters. (b) Joint distributions over each pair of parameters. The MAP parameter set is $b_{\text{in}} = 2.1\,\mu\text{m}\,\text{min}^{-1}, D = 83\,\mu\text{m}^2\,\text{min}^{-1}, \lambda = 0.0046\,\text{min}^{-1}$.

new models in isolation, the model identification algorithm, Algorithm 4.2, was applied to identify the preferred model. The two new models, pure-diffusion-restriction and drift-diffusion-restriction, were also compared to the previously preferred pure-diffusion model on a pairwise basis. The results of this are shown in Figure 5.4. The Bayes factors calculated from these results are

$$B_{31} \approx \frac{96}{4} = 24, \qquad B_{41} \approx \frac{71}{29} \approx 2.4, \qquad B_{34} \approx \frac{77}{23} \approx 3.3 \tag{5.6}$$

Referring to Table 4.1, the conclusion can be made that there is substantial evidence for preferring Model 3 over Model 4. Also, there is strong evidence for Model 3 over the previously preferred Model 1. In other words the introduction of the receptor depletion term in the model has significantly improved the performance of the identified model.

The model identification algorithm with Models 3 and 4 as candidates was also

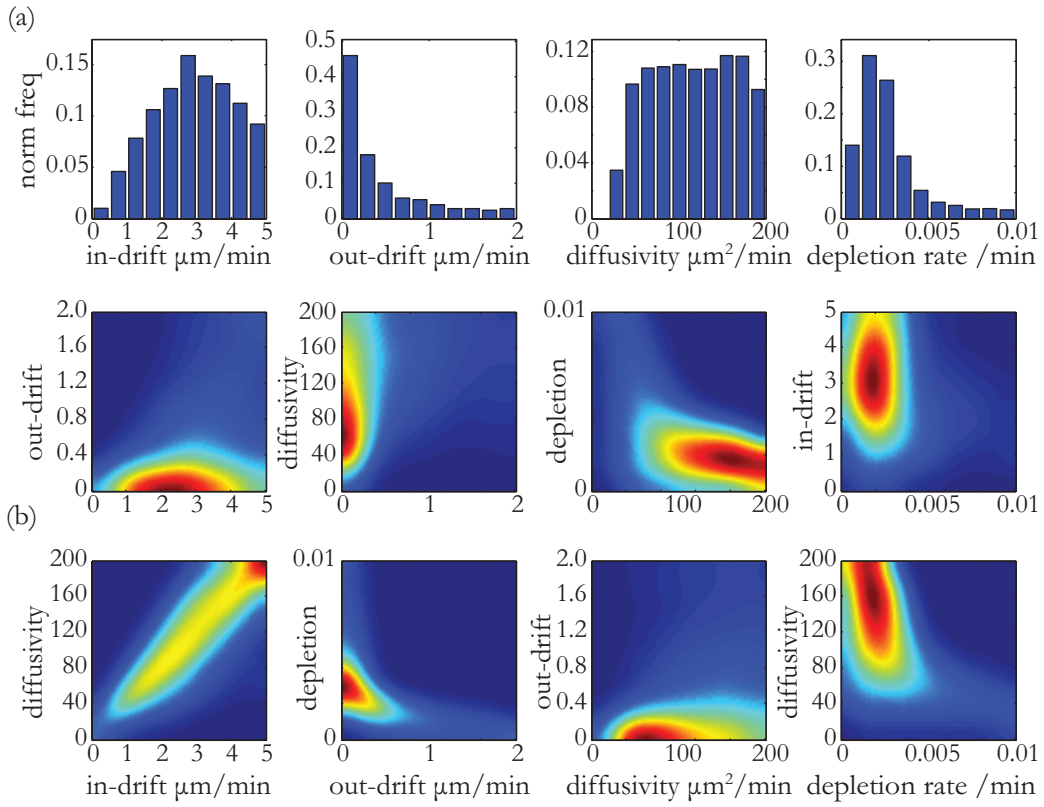

Figure 5.3: **Estimation results for Model 4, the drift-diffusion-depletion model.** (a) Posterior distribution over the individual parameters. (b) Joint distributions over each pair of parameters. The MAP parameter set is $b_{\text{in}} = 2.8\,\mu\text{m}\,\text{min}^{-1}$, $b_{\text{out}} = 0.3\,\mu\text{m}\,\text{min}^{-1}$, $D = 120\,\mu\text{m}^2\,\text{min}^{-1}$, $\lambda = 0.0025\,\text{min}^{-1}$.

repeated several times with incrementally increasing minimum allowed values for the drift parameter for Model 4. In each case the Bayes factor was calculated and the results are shown in Figure 5.7(a). For all minimum allowed values of drift there is substantial evidence for Model 3 over Model 4 with the evidence clearly increasing towards the strong evidence threshold as the minimum allowed value of drift increases.

5.1.3 Discussion

The positive identification of a model which includes a receptor depletion term suggests that this is an important addition to the dynamic model. The data driven framework suggests that this model is strongly preferred over the previously best identified model, the pure-diffusion model from Chapter 4. The discrimination between the two new models was somewhat ambiguous. Substantial evidence was found for the pure-diffusion version over the drift-diffusion version. However, the Bayes fac-

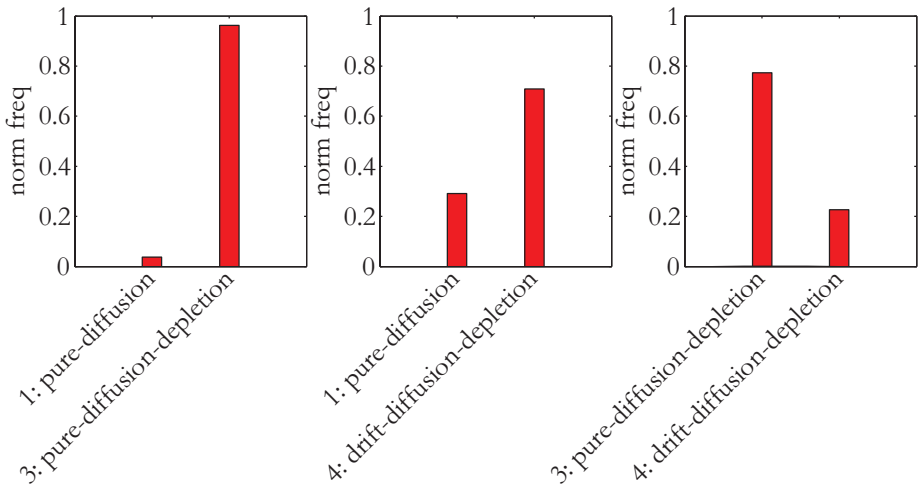

Figure 5.4: **Model selection for Models 1,3,4.** Algorithm 4.2 was applied to pairs of candidates models to identify the preferred model. The preferred pure-diffusion model from Chapter 4 is compared to the two new depletions models.

tor analysis did not show as much increase in confidence for the pure-diffusion version as was found in the similar analysis in Chapter 4. The probable reason for this is that there was not enough information in the data to provide discriminatory power for Model 4, the drift-diffusion-depletion model, which has an additional parameter to those in Model 3, the pure-diffusion variant.

This observation motivates the design of additional future experiments. For the receptor depletion dynamics to be fully explored and identified, it is necessary for the observed cells to spend significant time in both the recruitment and resolution modes of inflammation migration. The experimental design which produced the data analysed in this thesis means that, by and large, neutrophils have completed their recruitment migration and are switched off or in process of switching off from the recruitment signals. For this reason there will not be much information present pertaining to the receptor depletion activity. A proposed future experiment is outlined in Figure 5.6. This involves earlier labelling of neutrophils which are still migrating towards the inflammation site, in order to observe and capture the full dynamic shift from inflammation recruitment to its resolution. The challenge will be to capture a sufficient number of neutrophils to make meaningful analysis possible.

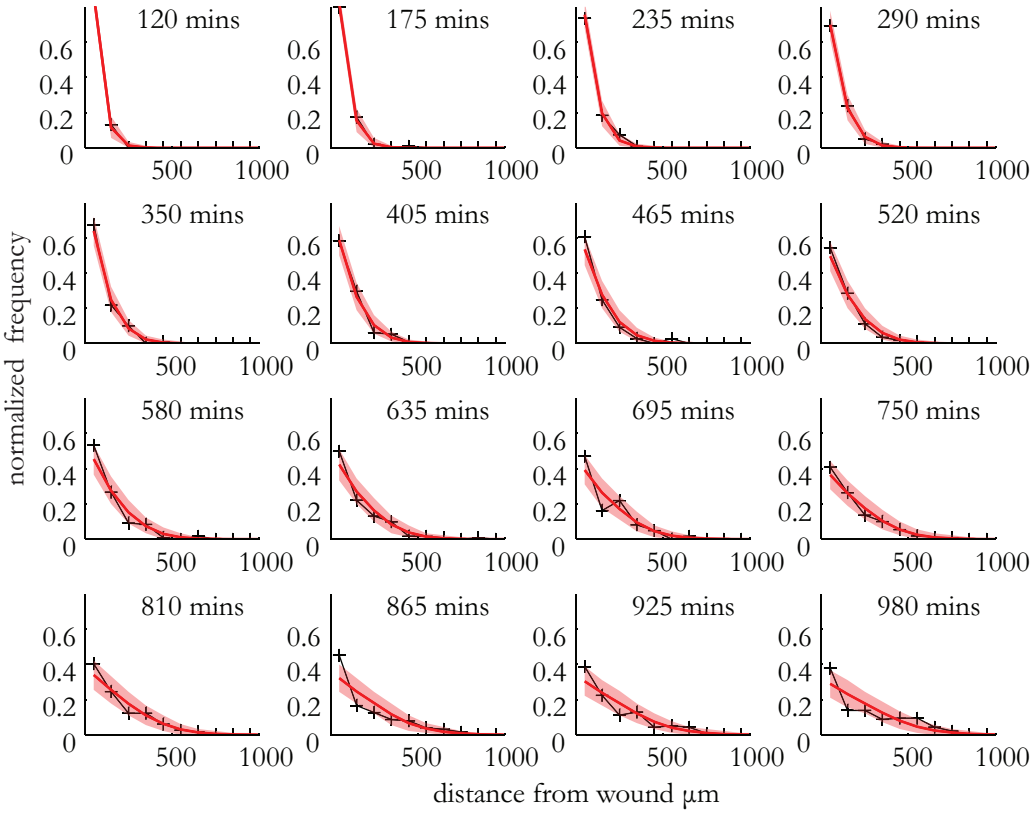

Figure 5.5: **Simulation comparisons of the estimated pure-diffusion depletion model, Model 3, to the experimental data.** One thousand simulations were made of the estimated model using the MAP parameter estimate. The output is shown for representative times throughout the experiment. The red line shows the mean predicted distribution of cell positions and the shaded red region is the 90% confidence interval. The experimental observations are plotted in black for comparison.

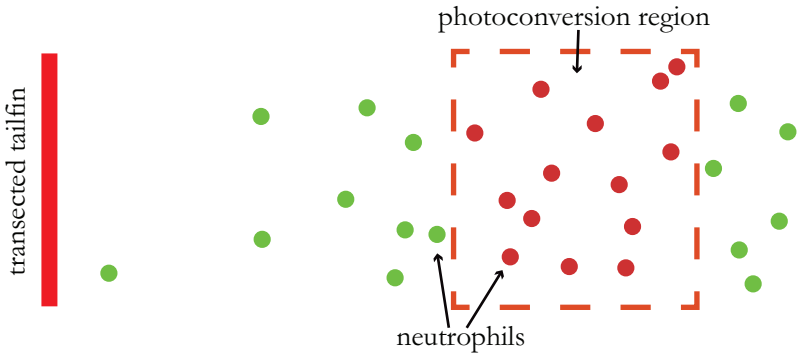

Figure 5.6: **Proposed experimental design for improved identification of receptor depletion models.** In the existing Datasets (A & B), the neutrophils are photoconverted in the region of the wound (within 100 µm) so that they can be observed as they subsequently move away from the wound (see Figure 3.6 on page 50). The receptor depletion models include terms for a recruitment bias towards the wound as well as any subsequent outward bias. In order to provide information for the identification of this initial bias it will be preferable to photoconvert cells that are still in the inward migration mode. In the proposed experimental design shown here, the photoconversion region is therefore displaced some distance away from the wound and photoconversion is applied earlier in the experiment to avoid labelling cells that are already resolving away from the wound. The challenge will be to capture sufficient cells in the labelling process to enable meaningful statistical analysis.

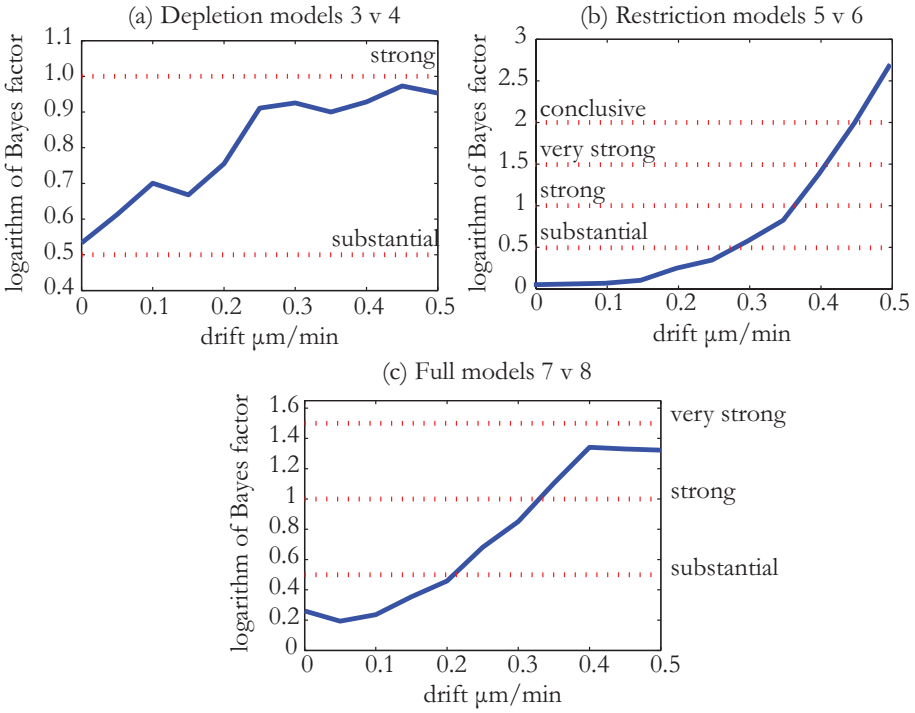

Figure 5.7: **Bayes factor analysis:** The Bayes factor is plotted on a logarithmic scale against the lower limit on the prior for the drift coefficient in the appropriate drift model.
(a) Bayes factor B_{34} for the pure-diffusion-depletion model, 3, with respect to the drift-diffusion-depletion model, 4. For all minimum allowed values of drift there is substantial evidence for Model 3.
(b) Bayes factor B_{56} for the pure-diffusion-restriction model, 5, with respect to the drift-diffusion-restriction model, 6. There is substantial evidence for Model 5 for drift values above approx. $0.27\,\mu\text{m}\,\text{min}^{-1}$.
(c) Bayes factor B_{78} for the pure-diffusion-depletion-restriction model, 7, with respect to the drift-diffusion-depletion-restriction model, 8. There is substantial evidence for Model 7 for drift values above approx. $0.21\,\mu\text{m}\,\text{min}^{-1}$

.

5.2 Physical restrictions included in the model

At the start of this chapter, two assumptions made in the modelling presented in Chapter 4 were reviewed. The last section investigated a relaxation of the first assumption. In this section a relaxation of the second assumption is now considered. This assumption stated that the neutrophils are migrating in a uniform environment. In relaxing this, one has to be realistic about what can be achieved given the experimental data at one's disposal. For instance, one cannot expect to identify drift and diffusivity parameters that have a high degree of spatial and temporal variation without having sufficient data to make identification of these parameters possible. The assumption is therefore relaxed in a limited way, one that is indicated by a global observation of the data as is now described.

Figure 5.8 shows a plot of all neutrophil position observations in a single zebrafish larva. The observations were recorded throughout the course of a 16 hour (approx.) experiment. Consideration of the data in this format suggested that, away from the vicinity of the tailfin area, the neutrophils tended to move preferentially along defined pathways. Similar evidence was seen in analogous plots for the other specimens in the experiment. From this evidence it was postulated that the cells were more likely to be able to leave the wound area if they were near to the entrance to one of these channels. Conversely, if they were not near the entrance of a channel their migration away from the wound area was likely to be restricted causing them to continue moving around in the vicinity of the wound. This restriction could result in significant disruption of the original pure-diffusion model. This, in turn, could explain the observation in Chapter 4: that the prediction of the identified model underestimated the concentration of cells remaining in the vicinity of the wound. Possibly, it could also mask the presence of a drift – making a drift-diffusion process appear to be pure-diffusion process and therefore compromising the conclusions reached so far. Therefore, in order to account for this restriction effect in a simplified way two additional models were proposed.

5.2.1 Model description

All of the models in this thesis are specified with one spatial dimension. This is the x-coordinate of the cells which is the perpendicular distance to the wound inducing the inflammation. The justification for this is that the wound goes right across the tailfin of the zebrafish, perpendicular to the spinal axis. This, together with the reasonable assumption that any chemotactic guidance will be either towards or away from the wound [85], implies chemoattractant fields that are constant in the direction parallel to the wound. In order to continue with one spatial dimension, the restriction effect was modelled in a simple probabilistic way. It was assumed that a cell moving away

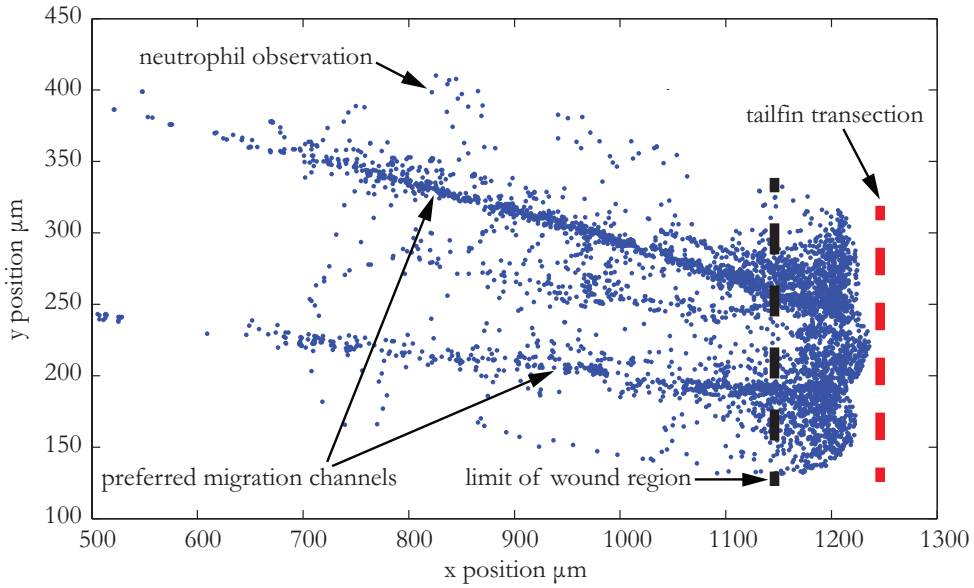

Figure 5.8: **Migration of neutrophils away from the wound indicates preferred channels of movement.** The positions of photoconverted neutrophils are shown at all timepoints for an individual zebrafish specimen (blue dots). Many of the cell positions appear to be gathered along specific pathways within the extracellular matrix, suggesting that the neutrophils migrate along preferred channels. All six zebrafish larva in Dataset A showed a similar pattern. See Appendix B for discussion of Dataset B. Also shown in this figure is a defined wound region which is used in the restriction modelling in this section.

from the wound and reaching the edge of the wound region (see Figure 5.8) has a probability r of being prevented from migrating away and a probability $(1-r)$ of migrating away successfully. Simplistically, the probability $(1-r)$ can be thought of as corresponding to the ratio of the cross-sectional area of the preferred migration channels to the cross-sectional area of the tailfin environment. However, the channels were observed experimentally not to coincide with any existing vascular or lymphatic structures in the fish. Rather, they are contingent channels through the tissue. It is possible that such channels could be formed as a result of earlier migration routes [7]. However, the positioning of the channels was similar across different experimental specimens, suggesting that they arise due to variations in the tissue rather than as a result of random migratory paths. In any case, the restriction parameter is intended to give an indication of the average ease with which migration away from the wound is navigated. This is achieved in the simulation model by using a test variable to decided whether cell move in its intended direction out of the wound region or whether it remains at the nominal boundary of the wound region. The full model is specified below, on page 104.

Restriction models

$$\hat{x}_{t+1}^{(i)} = \max(0, x_t^{(i)} + b_{\text{out}}\Delta t + \omega_t^{(i)}\sqrt{2D\Delta t}), \tag{5.7}$$

$$\pi_{t+1}^{(i)} \sim U(0,1) \tag{5.8}$$

$$x_{t+1}^{(i)} = \begin{cases} x_R & \text{if } x_t^{(i)} < x_R \text{ and } \hat{x}_{t+1}^{(i)} > x_R \text{ and } \pi_{t+1}^{(i)} < \beta \\ \hat{x}_{t+1}^{(i)} & \text{otherwise} \end{cases} \tag{5.9}$$

where

- $\hat{x}_t^{(i)}$ is a proposal for the position $x_t^{(i)}$ of the i^{th} neutrophil at time t;
- b_{out} is a bias velocity away from the wound;
- D is the underlying diffusivity constant or magnitude of random movement of the neutrophils;
- $\omega_t^{(i)} \sim \mathcal{N}(0,1)$ are a family of independent white noise processes;
- Δt is the time increment;
- $\beta \in [0,1)$ is a constant denoting the strength of the wound exit restriction ($\beta = 0$ means no restriction, $\beta = 1$ means 100% restriction);
- $\pi_t^{(i)} \sim \mathcal{U}(0,\,1)$ is a uniformly distributed test random variable for the i^{th} particle for simplified modelling of wound exit restriction – if the cell is moving outwards across the restriction region it has a probability, β of being unable to do so, i.e. the cell is restricted if $\pi_t^{(i)} \leqslant \beta$.
- x_R is the distance from the wound at which exit restriction is deemed to be experienced. In practice this was chosen as the minimum distance which contains all the cell positions at the time of photo-conversion which coincided with where the preferential migration channels were observed to begin relative to the wound. For Dataset A this was 100 µm.

With $b_{\text{out}} = 0$ this is designated as Model 5, and with $b_{\text{out}} \geqslant 0$, Model 6.

Model 5 & 6: pure-diffusion-restriction & drift-diffusion-restriction

5.2.2 Estimation results

In order to estimate the posterior distributions for the parameters of the new models conditional on the zebrafish data (Dataset A), Algorithm A.1 was again applied to each in turn. The results of the estimations are shown in Figure 5.9 and Figure 5.10 respectively. For Model 5, the pure-diffusion-restriction model, the MAP parameter estimates are

$$D = 38\,\mu\text{m}^2\,\text{min}^{-1}, \quad \text{restriction} = 60\%$$

and the 90% confidence interval in the marginal posteriors are $25 - 56\,\mu m^2\,min^{-1}$ and $25 - 75\%$ respectively. For Model 6, the drift-diffusion-restriction model, the MAP parameter estimates are

$$b_{out} = 0.1, \quad D = 28\,\mu m^2\,min^{-1}, \quad \text{restriction} = 60\%$$

and the 90% confidence interval in the marginal posteriors are $0 - 0.27\,\mu m\,min^{-1}, 12 - 53\,\mu m^2\,min^{-1}$ and $20 - 90\%$ respectively.

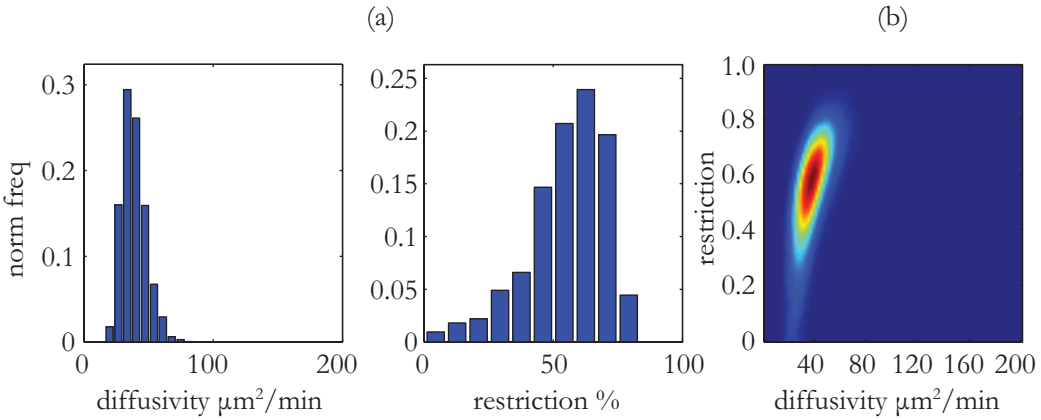

Figure 5.9: **Estimation results for, Model 5, the pure-diffusion restriction model.** (a) Posterior distribution over the individual parameters. (b) Joint distributions over each pair of parameters. The MAP parameter set is $D = 38\,\mu m^2\,min^{-1}$, restriction $= 60\%$.

To discriminate the data driven fidelity of these models with respect to each other and with respect to the preferred model from Chapter 4, Algorithm 4.2 was applied with pairs of these 3 models as candidates. The posterior model marginals for each of the 3 pairings is illustrated in Figure 5.11. The resulting Bayes factors were as follows,

$$B_{51} \approx 9.5, \quad B_{61} \approx 3.3, \quad B_{56} \approx 1.1 \tag{5.10}$$

These Bayes factors suggest (see Table 4.1) that there is substantial evidence for both the new restriction models over the original pure-diffusion model but no evidence to distinguish between the two new models. However, as before, Model 6, the drift-diffusion-restriction model, allows for drift values that are zero or arbitrarily close to zero. The Bayes factor analysis, where drift values are constrained above a increasing sequence of minimum thresholds, was repeated and the results are shown in Figure 5.7(b). Here it can be seen that if drift values less than approximately $0.28\,\mu m\,min^{-1}$ are considered insignificant, then there is substantial evidence for Model 5 over Model 6.

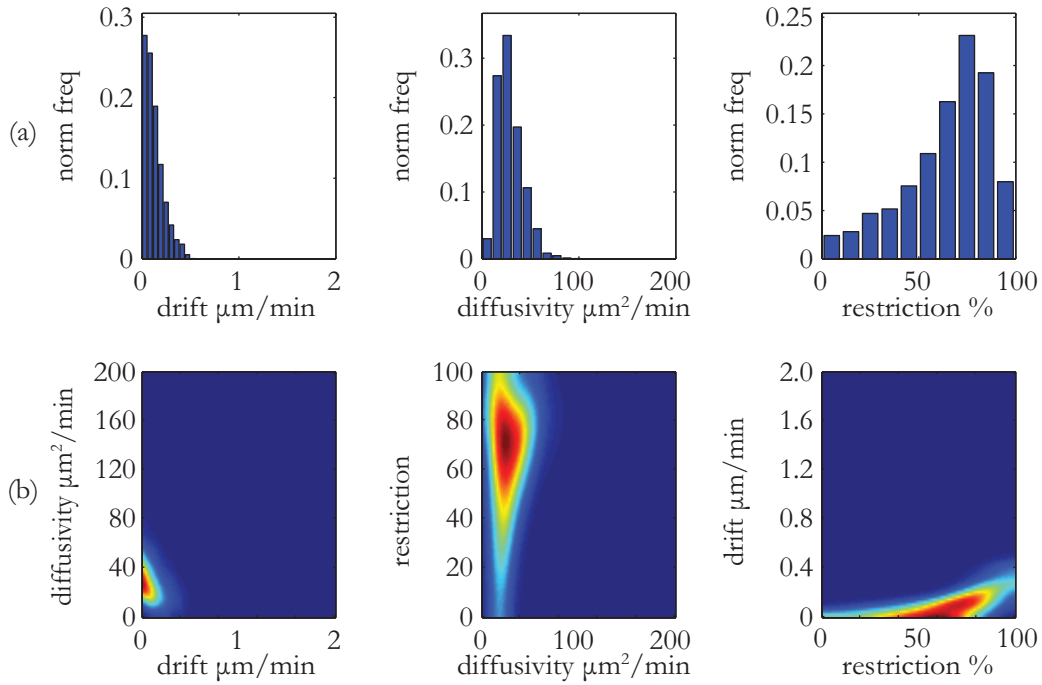

Figure 5.10: **Estimation results for, Model 6, the drift-diffusion-restriction model.** (a) Posterior distribution over the individual parameters. (b) Joint distributions over each pair of parameters. The MAP parameter set is $b_{\text{out}} = 0.1, D = 28\,\mu\text{m}^2\,\text{min}^{-1}$, restriction = 60%

5.2.3 Discussion

In this section, an additional component has been incorporated into the migration models developed in Chapter 4. The new component describes preferred pathways for cell movement through the tissues of the zebrafish and is characterised by a spatial restriction parameter. The addition of a spatial restriction parameter was motivated by the observation that cell positions appeared to be gathered in specific spatial channels and also that the pure-diffusion model from Chapter 4 tended to underestimate somewhat the concentration of cells remaining in the immediate vicinity of the wound (i.e. within 100 μm). In addition, there was a concern that omission of such a restriction parameter, if it was indeed necessary, might mask the presence of drift. By including this model term, the ability to detect drift, which corresponds to directed migration of the neutrophils, was enhanced. The parameter could have been rejected by the model selection algorithm or could have been assigned a zero value, indicating it was not necessary. Instead it was found to have a positive value of 60%. Such channels might arise due to physical characteristics of the local environment: it might be physically easier to displace tissue matrix around the spinal area than in the tailfin. However, they do not correspond to vascular or lymphatic structures,

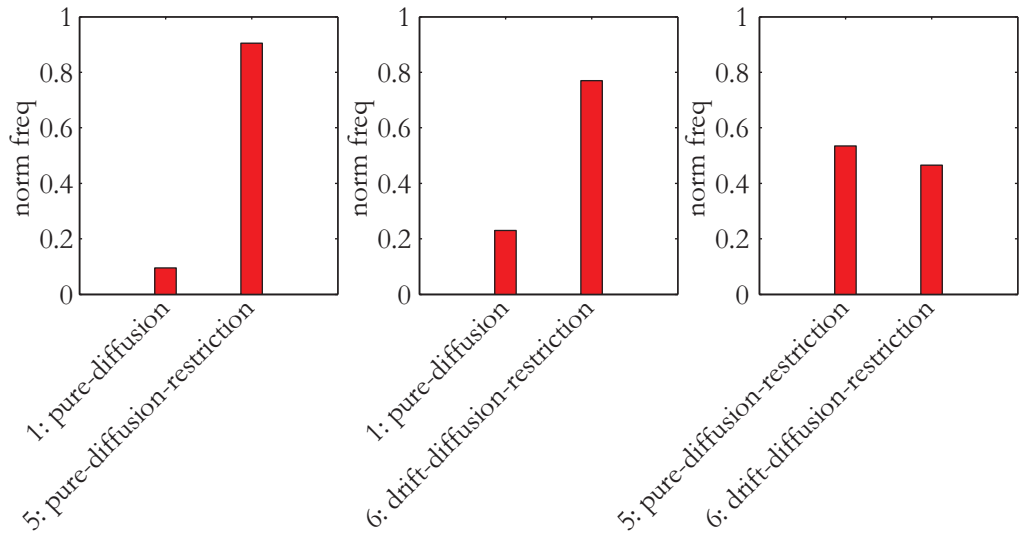

Figure 5.11: **Model selection for Models 1,5,6.** Algorithm 4.2 was applied to pairs of candidates models to identify the preferred model.

which would suggest they are a feature of extravascular tissues. In addition, the site of the channels is similar in different fish, suggesting they are not solely formed from random paths chosen by individual cells which define pathways of subsequent neutrophil migrations.

The ABC-SMC model selection algorithm (Algorithm 4.2) was used to select between pure-diffusion and drift-diffusion versions of the restriction model and the original pure-diffusion model of Chapter 4. The preferred model, in the light of the experimental data, was Model 5, the pure-diffusion-restriction model. Compared to Model 1, the pure-diffusion model, the drift-diffusion-restriction model found substantial supporting evidence (Bayes factor 9.5, with the substantial range being 3.2-10, see Table 4.1). However, the evidence for Model 5, the pure-diffusion-restriction model over Model 6, the drift-diffusion-restriction model was less emphatic. The representation in the model marginal posterior distribution was 53% versus 47%. The MAP value of drift in the identified drift-diffusion-restriction model was 0. However, not unless drift values under approximately $0.28\,\mu m\,min^{-1}$ were considered insignificant did the former have substantial evidence over the latter. When the parameters of the drift-diffusion-restriction model were estimated using Algorithm A.1 the 90% confidence interval for the drift parameter was 0-$0.27\,\mu m\,min^{-1}$. If the drift value

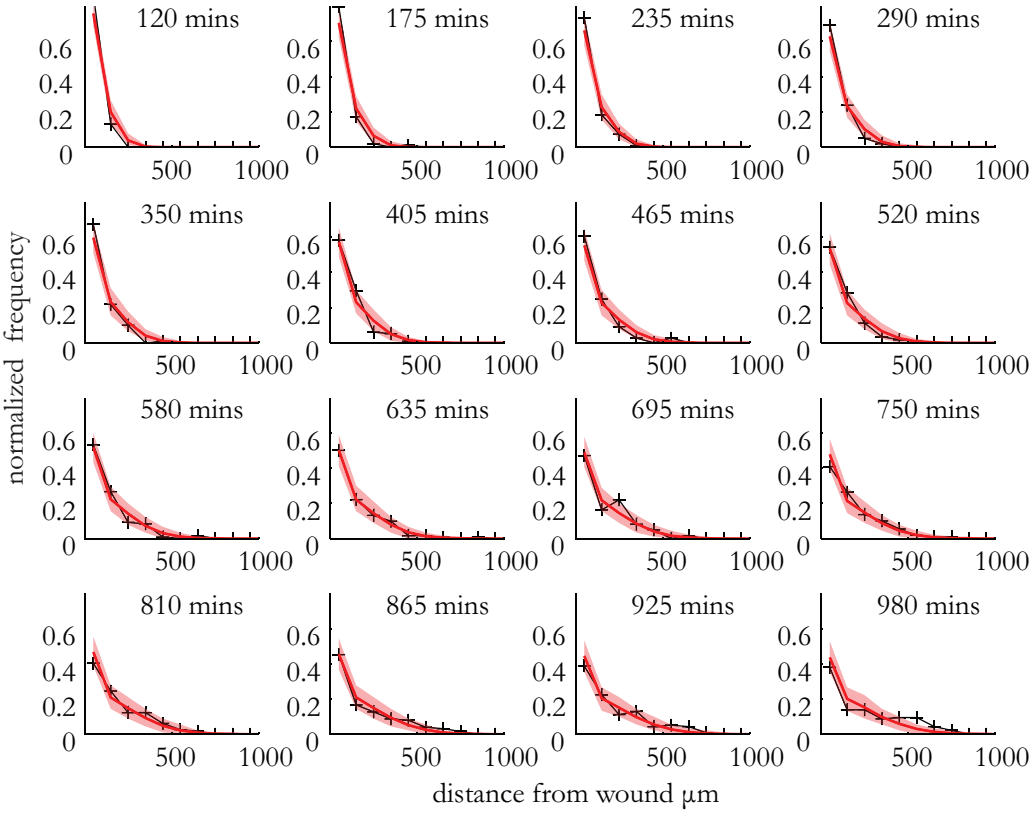

Figure 5.12: **Simulation comparisons between the estimated pure-diffusion restriction model, Model 5, and the experimental data.** One thousand simulations were made of the estimated model using the MAP parameter estimate. The output is shown for representative times throughout the experiment. The red line shows the mean predicted distribution of cell positions and the shaded red region is the 90% confidence interval. The experimental observations are plotted in black for comparison.

was at the limit of this range, it would correspond to a directed net migration of approximately 250 µm over the course of the experiment (980 mins). Nevertheless, the dynamics of such a model are still dominated by the random component (see chapter discussion in Section 5.5).

5.3 Combining the depletion and restriction models

In order to take a final approach to the problem, both the assumptions made in Chapter 4 were simultaneously relaxed by combining the two models proposed in the previous sections of this chapter. The resulting pair of models, therefore, allow for a change in the migration mode of the neutrophils through depletion of the chemoattractant ligand receptors on the surface of the cells, and also for a limited

heterogeneity in the extra-cellular environment through which the neutrophils move. Model 7, the pure-diffusion version, has 4 parameters and Model 8, the drift-diffusion version, has 5. It should be noted that the Bayesian estimation paradigm automatically includes a penalty for increased model complexity. This guards against the fact that a more complex model can always be made to fit the data better than a simpler model within which it is nested. The combined models are specified on page 110.

5.3.1 Estimation results

In order to estimate the posterior distributions for the parameters of the new models, Algorithm A.1 was again applied to each in turn, with the zebrafish neutrophil Dataset A. The results of the estimations are illustrated in Figure 5.13 and Figure 5.14 respectively. For Model 7, the pure-diffusion version, the MAP parameter estimates are

$$b_{\text{in}} = 3\,\mu\text{m}\,\text{min}^{-1}, \quad D = 46\,\mu\text{m}^2\,\text{min}^{-1}, \quad \lambda = 0.03\,\text{min}^{-1}, \quad \text{restriction} = 50\%$$

and the 90% confidence interval in the marginal posteriors are $0.8-4.8\,\mu\text{m}\,\text{min}^{-1}, 31-86\,\mu\text{m}^2\,\text{min}^{-1}, 0.005-0.09\,\text{min}^{-1}\text{and}10-70\%$ respectively. For Model 10 the MAP parameter estimates are

$$b_{\text{in}} = 1.8\,\mu\text{m}\,\text{min}^{-1}, \quad b_{\text{out}} = 0.2\,\mu\text{m}\,\text{min}^{-1}, \quad D = 80\,\mu\text{m}^2\,\text{min}^{-1},$$
$$\lambda = 0.004\,\text{min}^{-1}, \quad \text{restriction} = 40\%$$

and the 90% confidence interval in the marginal posteriors are $0.3-4.4\,\mu\text{m}\,\text{min}^{-1}, 0-0.86\,\mu\text{m}\,\text{min}^{-1}, 27-177\,\mu\text{m}^2\,\text{min}^{-1}, 0.001-0.009\,\text{min}^{-1}$ and $0-70\%$ respectively. These results can be compared with those for all the models of this and the previous chapter via Table 5.5.

In order to discriminate between the two new models, Algorithm 4.2 was applied. Once again, it was applied several times with incrementally increasing minimum allowed values of drift, in the drift version of the model. The Bayes factors were calculated and are shown in Figure 5.7(c). It can be seen that there is substantial evidence for Model 7 for drift values above approx. $2.1\,\mu\text{m}\,\text{min}^{-1}$ and this increases to strong evidence above values of approx. $3.3\,\mu\text{m}\,\text{min}^{-1}$.

Algorithm 4.2 was applied again on a pairwise basis to compare the three favoured models from this chapter: pure-diffusion-depletion, Model 3, pure-diffusion-restriction, Model 5, pure-diffusion-depletion-restriction, Model 7 (Figure 5.16). Only in the comparison of Model 7 and Model 3 is there any substantial evidence, with $B_{73} = 3.5$.

Combined receptor-depletion restriction models

$$\hat{x}_{t+1}^{(i)} = \max(0, x_t^{(i)} + \left(b_{\text{out}} - R_t^{(i)} b_{\text{in}}\right) \Delta t + \omega_t^{(i)} \sqrt{2D\Delta t} \tag{5.11}$$

$$R_{t+1}^{(i)} = R_t^{(i)} - \lambda \max\left(0, \frac{L - x_t^{(i)}}{L}\right) R_t^{(i)} \Delta t \tag{5.12}$$

$$\pi_{t+1}^{(i)} \sim U(0,1) \tag{5.13}$$

$$x_{t+1}^{(i)} = \begin{cases} x_R & \text{if } x_t^{(i)} < x_R \text{ and } \hat{x}_{t+1}^{(i)} > x_R \text{ and } \pi_{t+1}^{(i)} < \beta \\ \hat{x}_{t+1}^{(i)} & \text{otherwise} \end{cases} \tag{5.14}$$

where

- $\hat{x}_t^{(i)}$ is a proposal for the position $x_t^{(i)}$ of the i^{th} neutrophil at time t;
- b_{out} is a bias velocity away from the wound;
- $R_t^{(i)}$ is the proportion of receptors available for the i^{th} cell at time t.
- b_{in} is a bias velocity towards the wound;
- $\omega_t^{(i)} \sim \mathcal{N}(0,1)$ are a family of independent white noise processes;
- D is the underlying diffusivity constant or magnitude of random movement of the neutrophils;
- Δt is the time increment;
- λ is the composite depletion constant described on page 94.
- L is the nominal range of the chemoattractant field.
- $\hat{x}_{t+1}^{(i)}$ is a proposal for the position $x_t^{(i)}$ of the i^{th} neutrophil at time t;
- $\beta \in [0,1)$ is a constant denoting the strength of the wound exit restriction ($\beta = 0$ means no restriction, $\beta = 1$ means 100% restriction);
- $\pi_t^{(i)} \sim \mathcal{U}(0,\, 1)$ is a uniformly distributed test random variable for the i^{th} particle for simplified modelling of wound exit restriction – if the cell is moving outwards across the restriction region it has a probability, β of being unable to do so, i.e. the cell is restricted if $\pi_t^{(i)} \leqslant \beta$.
- x_R is the distance from the wound at which exit restriction is deemed to be experienced.

With $b_{\text{out}} = 0$ this is designated as Model 7, and with $b_{\text{out}} \geqslant 0$, Model 8.

Model 7 & 8: pure-diffusion version & drift-diffusion version

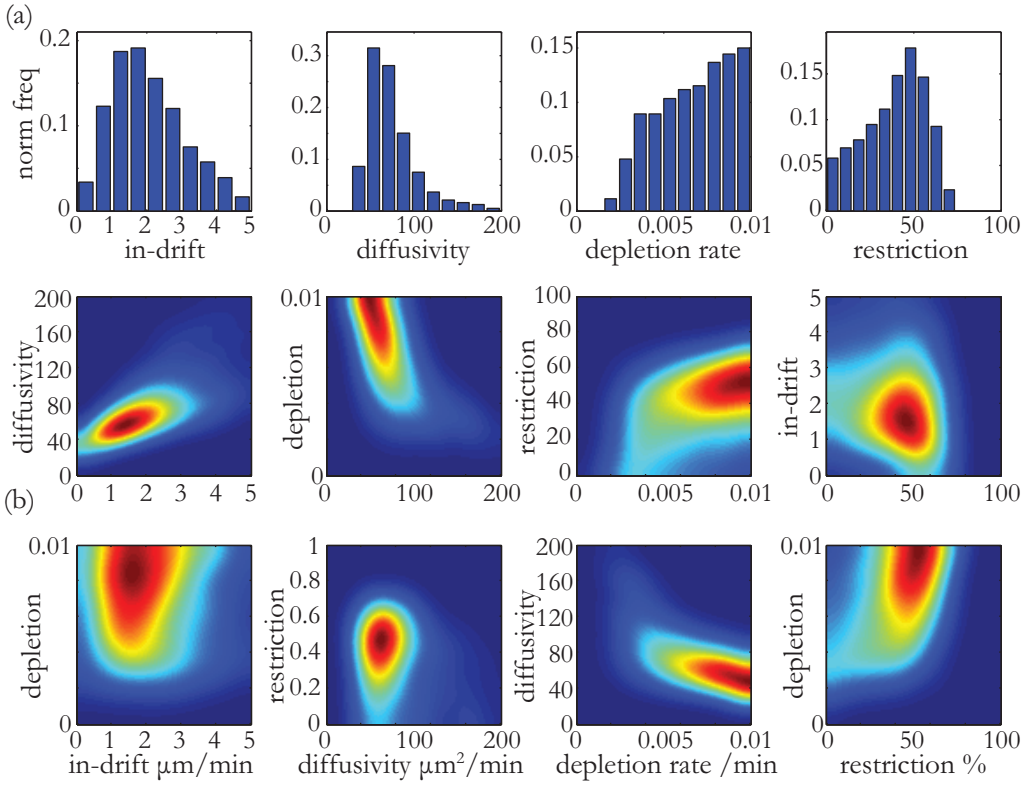

Figure 5.13: **Estimation results for, Model 7, the pure-diffusion-depletion-restriction model.** (a) Posterior distribution over the individual parameters. (b) Joint distributions over each pair of parameters. The MAP parameter set is $b_{\text{in}} = 3\,\mu\text{m}\,\text{min}^{-1}$, $D = 46\,\mu\text{m}^2\,\text{min}^{-1}$, $\lambda = 0.03\,\text{min}^{-1}$, restriction $= 50\%$

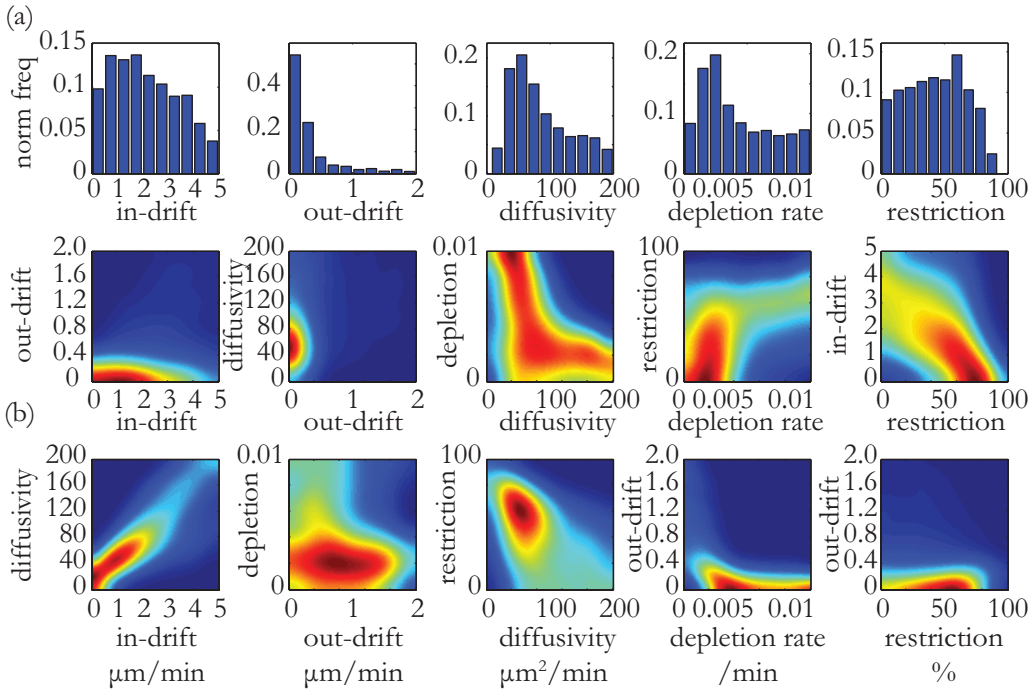

Figure 5.14: **Estimation results for, Model 8, the drift-diffusion-depletion-restriction model.** (a) Posterior distribution over the individual parameters. (b) Joint distributions over each pair of parameters. The MAP parameter set is $b_{\text{in}} = 1.8\,\mu\text{m}\,\text{min}^{-1}$, $b_{\text{out}} = 0.2\,\mu\text{m}\,\text{min}^{-1}$, $D = 80\,\mu\text{m}^2\,\text{min}^{-1}$, $\lambda = 0.004\,\text{min}^{-1}$, restriction $= 40\%$

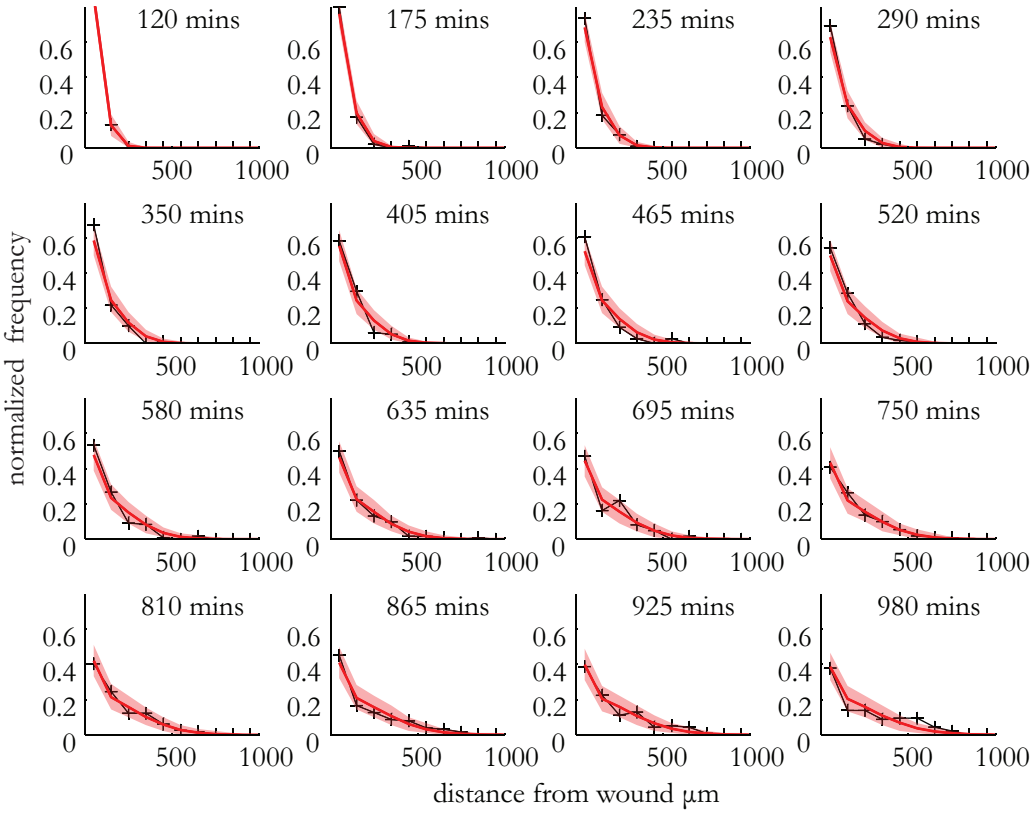

Figure 5.15: **Simulation comparisons of the estimated pure-diffusion-restriction-depletion model, Model 7, to the experimental data.** One thousand simulations were made of the estimated model using the MAP parameter estimate. The output is shown for representative times throughout the experiment. The red line shows the mean predicted distribution of cell positions and the shaded red region is the 90% confidence interval. The experimental observations are plotted in black for comparison.

5.4 Discussion

The analysis in this chapter has demonstrated the effectiveness of the ABC framework developed in Chapter 4, for estimating the parameters of complex models of cell migration and for discriminating between models not only when the models are nested but also when they have arbitrary relationships to each other. This confirms the framework as a useful tool in the exploration of cell migration and in particular of inflammatory neutrophil dynamics and the biological mechanisms that underlie these dynamics.

The motivation for developing the new models in this chapter was the shortcoming in the predictions of the basic pure-diffusion model. Figure 4.12 shows that the pure-diffusion model, which was the preferred model from Chapter 4, has a systematic underestimation of the density of cells in the immediate vicinity of the wound, especially throughout the first half of the experiment. It can be observed from Figures 5.5, 5.12 and 5.15 that this shortcoming has been significantly addressed and overcome by the expanded models. There was also a concern that the adoption of

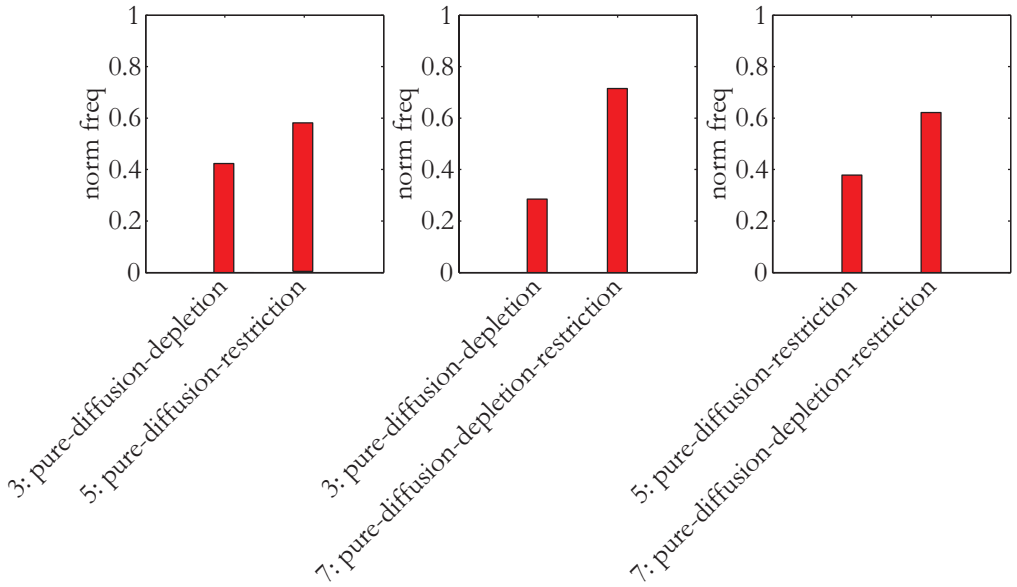

Figure 5.16: **Comparing the 3 favoured models, Models 3,5,7.** The evidence is not conclusive for deciding between the 3 different favoured models of this chapter – apart from the comparison between Models 3 and 7 where there is substantial evidence for Model 7 with a Bayes factor $B_{73} = 3.5$.

too simple a model in Chapter 4 might have hindered the ability to detect drift. For example a drift-diffusion process in which receptor depletion or physical restriction were active might look like a pure-diffusion process if the latter effects were not accounted for in the estimation process. When investigating the enhanced mod-

els it was found that there was scope for slightly higher values of drift coefficient when receptor depletion, physical restriction or both were included in the candidate models.

In Chapter 4 the MAP value of drift was $0.05\,\mu m\,min^{-1}$, whereas in this chapter the highest corresponding MAP value was estimated in the drift-diffusion-depletion model where the drift value was $0.3\,\mu m\,min^{-1}$. In both cases there was substantial evidence for the pure-diffusion version over the drift-diffusion version. However, the Bayes factor evidence reduced from 7.3 in the simple models to 3.3 in the depletion models. To summarise these observations, the introduction of increased model complexity does allow for an increase in possible drift-coefficient which corresponds to some element of bias away from the wound during inflammation resolution. However, the balance of probabilities is still in favour of a diffusion only process. Also, the magnitude of the largest possible drift coefficient for the alternative model is still of such a size to mean that the diffusive element of the dynamics dominates over any bias (see further discussion at the end of Section 5.5).

The results that have been presented in this chapter are derived only from Dataset A. The reason for this is that Dataset B was described much better by its identified pure-diffusion model. This is discussed in Appendix B where a summary of all the results for Dataset B are presented in Table B.2 and selected results are shown in the figures. In particular Figure B.2, in contrast to Figure 4.12 for Dataset A, shows that the cell density close the wound for Dataset B is much more accurately captured by the original pure-diffusion model, Model 1. It is not clear why there should be this difference in the results between the two datasets. There may be differences due to slight variation in environmental conditions or due to variation between different batches of fish embryos. However, it is important to emphasize that this thesis is performing data driven modelling. It should be stressed that it is the data from Dataset A which prompts the search for a better explanatory model and the data from Dataset B that does not require a more complex model. As and when further similar datasets become available, the framework made available here can be applied to obtain an overview of how often these extra factors are required to describe the data adequately. This is noted in Chapter 6 as potential future work.

Whilst both components introduced in this chapter, the receptor depletion and the physical restriction, may be important additions for accurate modelling of the neutrophil data, it is the receptor depletion that is more biologically interesting and which may have more potential for impacting on future therapies and treatments for inflammatory diseases. It is unlikely that the physical nature of the tissue of an organism can be readily controlled, whereas receptor dynamics may be amenable to alteration.

5.5 Conclusion

The ABC framework for cell migration developed in Chapter 4 was applied to two zebrafish neutrophil datasets. One of the datasets, Dataset B, was well described by the identified pure-diffusion model for inflammation resolution. Dataset A, however, was not fully described by the relatively simple pure-diffusion or drift diffusion models. This chapter has pursued two approaches for adding complexity to the models via chemoattractant receptor dynamics and spatial inhomogeneity. These two approaches have also been combined to form a composite model. Whilst a higher order model will always give a better fit to a dataset than a lower order model, the embedding of the identification process in a Bayesian framework means that there is an automatic penalty for model complexity. This means that the higher order model will not be chosen simply because it reduces prediction error but only if it explains the data more accurately. For Dataset A, each of the new models was preferred over the simple pure-diffusion model. There was no conclusive evidence to make a decision between the three new models. However, exploration of the receptor depletion model led to the proposal of a refined experimental design which would allow better determination of this model. It is hypothesized that if this experimental data is obtained, the receptor depletion model will present stronger evidence to be chosen as the preferred model of the migration system.

The new models in this chapter were also developed in order to investigate whether the present of a drift term (and hence migratory bias) might be overlooked if the full complexity of the system were oversimplified. In the drift versions of the new models, the drift coefficient is slightly higher than in the simple drift-diffusion model. But it is still of the same order or magnitude which is two orders of magnitude less than the dynamics associated with the diffusive component of the migration. Thus even in the unlikely event that drift is present, the migration is still dominated by stochastic search behaviour. In other words, persistent chemotaxis is not present.

model	in-drift $\mu m\,min^{-1}$	out-drift $\mu m\,min^{-1}$	diffusivity $\mu m^2\,min^{-1}$	depletion $10^{-3}min^{-1}$	restriction %
1	- -	- -	25 (19-34)	- -	- -
2	- -	0.05 (0.005-0.15)	19 (7-36)	- -	- -
3	2.1 (0.9-4.3)	- -	83 (39-172)	4.6 (2.4-9.2)	- -
4	2.8 (1.0-4.7)	0.3 (0.01-1.6)	120 (42-190)	2.5 (0.8-7.2)	- -
5	- -	- -	38 (25-56)	- -	60 (25-75)
6	- -	0.1 (0-0.27)	28 (12-53)	-	60 (20-90)
7	3.0 (0.8-4.8)	- -	46 (31-86)	30.0 (5.0-90)	50 (10-70)
8	1.8 (0.3-4.4)	0.2 (0-0.86)	80 (27-177)	4.0 (1.0-9.0)	40 (0-70)

Table 5.1: **Summary of the parameter estimates for Dataset A**. 1, pure-diffusion; 2, drift-diffusion; 3, pure-diffusion-depletion; 4, drift-diffusion-depletion; 5, pure-diffusion-restriction; 6, drift-diffusion-restriction; 7, pure-diffusion-depletion-restriction; 8, drift-diffusion-depletion-restriction. The 90% confidence intervals are shown in brackets.

Chapter 6

Conclusions and further work

This thesis has developed two novel frameworks to investigate and determine the dynamics of neutrophils during inflammation and inflammation resolution from cell population data. Recently, population modelling methods have tended to be overlooked in favour of cell tracking methods. However, a populational approach has a number of advantages:

- The sampling interval can be longer. Fast sampling is required for reliable tracking but the ultraviolet exposure required for these repeated observations causes harm to a living biological specimen such as the zebrafish larva. Less frequent sampling means that experimental run times can be longer and thus capture the full dynamics of an inflammation and resolution event.

- All cell observations can be included in the data for analysis. If tracks are analysed only those cell observations which are part of sufficiently long track portions can be included. This results in significant data loss and possible biasing of the results.

Both cell track analysis and cell population analysis will be important in pushing forward understanding of migration processes. This thesis makes an important contribution towards achieving this balance of approaches.

Chapter 3 developed a regression analysis approach for identifying migration dynamics and applied it to the zebrafish neutrophil population data. This allowed, for the first time, the identification of parameters of candidate models for both the recruitment and resolution phases of neutrophil migration during zebrafish inflammation. The method is fast in implementation (the analysis algorithm takes a fraction of a second once the data has been preprocessed into a suitable form), however, the framework does not allow confident discrimination between candidate models and a subsequent model simulation step was required in order to accomplish this. Related

to this is the fact that, by extracting one statistic from the data, namely the mean squared cell position at each timepoint, the method does not make maximal use of all the information available in the data. Nevertheless, this framework represent a significant step forward in rigorous mathematical analysis of zebrafish inflammatory neutrophil dynamics. Most notably, it raises evidence that goes against a currently emerging consensus. This consensus is that neutrophils are guided away from inflammation sites by chemical signals during inflammation resolution. The analysis of Chapter 3 suggests the opposite: that neutrophil inflammation resolution migration is solely the result of inherent stochastic migration patters.

Chapter 4 developed a more comprehensive framework, utilising the approximate Bayesian computation sequential Monte Carlo (ABC-SMC) approach, suitable for both parameter estimation and model identification for the neutrophil problem. This framework inherits the usual benefits of the Bayesian approach, specifically the ability to incorporate prior knowledge of appropriate parameter ranges and to estimate not only parameters and models but also associated uncertainties. The ABC approach requires tailoring to any new problem and this was done for the first time in the cell migration context by making use of an evolving cell distribution to summarise the data whilst maintaining all the relevant information it contains, and the Cha-Srihari distance [23] to measure the distance between summarised observations. This novel framework was applied to the zebrafish neutrophil data with candidate pure-diffusion and drift-diffusion models. The key finding of Chapter 3 was confirmed, i.e. that the pure-diffusion model, corresponding to undirected migration during inflammation resolution, was the one most likely to be a correct description of the process.

In Chapter 5 the Bayesian framework was applied again but now with three novel and more fully developed pairs of models, a pure-diffusion and drift-diffusion version in each case. The first model included ligand receptor depletion dynamics; the second included inhomogeneity in the extra-cellular matrix; and the third was a combined model including both these effects. The development of these novel models was motivated by:

- The fact that the predictions of the estimated pure-diffusion model had some shortcomings in its correspondence to the observed data. In particular, it underestimated the number of neutrophils that were retained near the wound throughout the course of the experiment.

- A desire to ensure that the detection of any bias would be robust.

The analysis of these models demonstrated the ability of the new Bayesian estimation framework to deal with complex models for which a likelihood function is unavailable or intractable, and also to inform choice between arbitrary models which are not

nested one within the other.

The drift-diffusion versions of these new models, which correspond to some degree of directed migration, like the drift-diffusion model of Chapter 4, were never represented by a majority of samples in the identification process. What is more, even if the neutrophils were moving according to the rejected drift version model, they would not be performing persistent chemotaxis, in which they purposefully move in the direction of the target. Instead they would be performing a very slightly biased random walk, where a cell's movement from moment to moment is very much dominated by inherent stochastic behaviour and the bias causes only a small net drift in the desired direction. In this scenario, portions of cell tracks would as usual have short term directionality but this would be due to short term directional correlations inherent in cell translocation rather than as a result of the chemotactic guidance. This suggests that it is a mistake to conclude that neutrophil tracks such as those shown in Figure 3.12 are moving under the influence of an external signal.

The mechanisms driving inflammation resolution are therapeutically important, and there is a fundamental mechanistic difference between the classes of molecular event that drive neutrophils away from inflammatory sites, and those that allow neutrophils to be blind to, or to ignore, chemotactic gradients that might retain neutrophils at inflammatory sites. The question of whether neutrophil behaviours are modelled best by chemotactic guidance or by stochastic redistribution is of fundamental importance in our understanding of inflammation resolution. This thesis has made an important contribution towards the search for a correct understanding of this process. It also provides a starting point for further work in the field.

1. The receptor depletion modelling in Chapter 5 has potential for enriching the biological understanding of how neutrophils change mode between recruitment and inflammation resolution. In Section 5.1.3 an enhanced experimental design was proposed to explore these dynamics more fully. When data from such an experiment have been collected the identification will be repeated and the models refined if appropriate. This will enable further testing of the hypothesis that the switch from recruitment to inflammation resolution is due to individual neutrophils being switched off from a recruitment signal rather than it being the signal itself that has ceased.

2. Two datasets were available for analysis in the preparation of this thesis. Both datasets led to a similar conclusion regarding the absence of directional bias on neutrophils during inflammation resolution. However, there was a difference between them in terms of the support found for the more complex receptor depletion and restriction models. If datasets from a significantly greater num-

ber of experiments can be obtained, these will be analysed in order to gain an overview of how often such additional model terms are found to be necessary.

3. The analysis presented in this thesis has been on neutrophils in a normal inflammation event. The results obtained are already affecting the direction of laboratory work in the University of Sheffield's Department of Infection and Immunity. There is potential for comparative analysis in modified contexts, where, for instance, the neutrophils have enhanced or reduced capacity for gradient sensing. Alternatively, various types of stress can be applied to the zebrafish embryo to investigate the effects of stress on the inflammation process. The methods of this thesis will therefore be used to further probe what are the important characteristics of healthy and pathological inflammation processes. This work had already been started with the analysis of neutrophils in zebrafish that have been treated with the drug tanshinone. Neutrophil dynamics appear to altered by tanshinone with inflammation resolution dispersion proceeding more rapidly. The new analysis in this context aims to uncover whether this is a result of more vigorous diffusive behaviour or earlier switching off from the recruiting chemotactic signal. A second treatment that is also being considered for investigation affects the hypoxia signalling pathways. This also modulates inflammation resolution but the way neutrophil behaviour is directly affected is as yet unknown.

4. There is scope for putting neutrophil track analysis alongside the population analysis used in this thesis. This will allow checking of the full consistency of the results obtained with the experimental data. Track data has been used in a related paper arising out of preparatory studies for this thesis [65] which used neutrophil observations to identify the unknown chemotactic field to which they were responding during inflammation recruitment. There is scope for developing enhanced tracking and mode detection algorithms based on dynamic models for neutrophil shape will complement and allow extension of the work in this thesis.

In conclusion, this thesis has achieved its aims and objectives. Building on this foundation, the program of future work outlined above will yield further important insights in the inflammation response of the innate immune system.

Appendix A

ABC-SMC parameter estimation

Algorithm A.1 Parameter estimation using ABC-SMC

Require: data, $\mathcal{Y}^{\text{obs}}$; Monte Carlo population size, N; number of iterations, T; prior distribution on parameter vector, $\pi(\theta)$; simulation algorithm to sample replicated observations from the process, $\mathcal{Y} \sim p(\mathcal{Y}|\theta)$ distance metric ρ and parameter perturbation kernel K; decreasing error tolerance schedule $\epsilon_1, \ldots, \epsilon_T$

Ensure: a set of parameter vectors θ_i with importance weights $\omega_i, i = 1 \ldots N$ that form a weighted sample from the posterior distribution, $p(\theta|\mathcal{Y}^{\text{obs}})$

for $i = 1$ **to** N **do**

simulate $\theta_i \sim \pi(\theta)$ and $\mathcal{Y} \sim p(\mathcal{Y}|\theta_i)$ until $e_i = \rho(\mathcal{Y}, \mathcal{Y}^{\text{obs}}) \leqslant \epsilon_1$

end for

set each $\omega_i^{(1)} \propto \frac{1}{\epsilon_1}\left(1 - \left(\frac{e_i}{\epsilon_1}\right)^2\right)$, such that $\sum \omega_i^{(1)} = 1$

for $t = 2$ **to** T **do**

Set $\tau^2 = 2\text{Var}(\{\theta_i : i = 1 \ldots N\})$

for $i = 1$ **to** N **do**

choose θ^* from the θ_j with probabilities ω_j

simulate $\hat{\theta}_i \sim K(\theta|\theta^*; \tau^2)$ and $\mathcal{Y} \sim p(\mathcal{Y}|\hat{\theta}_i)$ until $e_i = \rho(\mathcal{Y}, \mathcal{Y}^{\text{obs}}) \leqslant \epsilon_t$

set $\tilde{\omega}_i = \frac{1}{\epsilon_t}\left(1 - \left(\frac{e_i}{\epsilon_t}\right)^2\right)$

end for

set each $\hat{\omega}_i^{(t)} \propto \frac{\tilde{\omega}_i \pi(\hat{\theta}_i)}{\sum_{j=1}^{N} \omega_j^{(t-1)} K(\hat{\theta}_i|\theta_j; \tau^2)}$, such that $\sum \hat{\omega}_i^{(t)} = 1$

set each $\theta_i = \hat{\theta}_i$, $\omega_i = \hat{\omega}_i$

end for

Appendix B

Additional results for Dataset B

The results in the main body of this thesis are based on analysis of Dataset A, which was available throughout the research period. Towards the end of the period, a second dataset, Dataset B, of exactly similar data became available. This was used to confirm the main result at the end of Chapter 4. Selected further results from the analysis of Dataset B and presented in this appendix. The results from Dataset B confirm the major novel finding of this thesis, that regardless of the complexity of the model, the pure-diffusion version is preferred in model selection over the drift-diffusion version (see Figure B.1 and Table B.1). This in turn implies that the resolution migration of inflammatory neutrophils in zebrafish is governed by inherent stochastic search patterns and not by external chemotactic guidance.

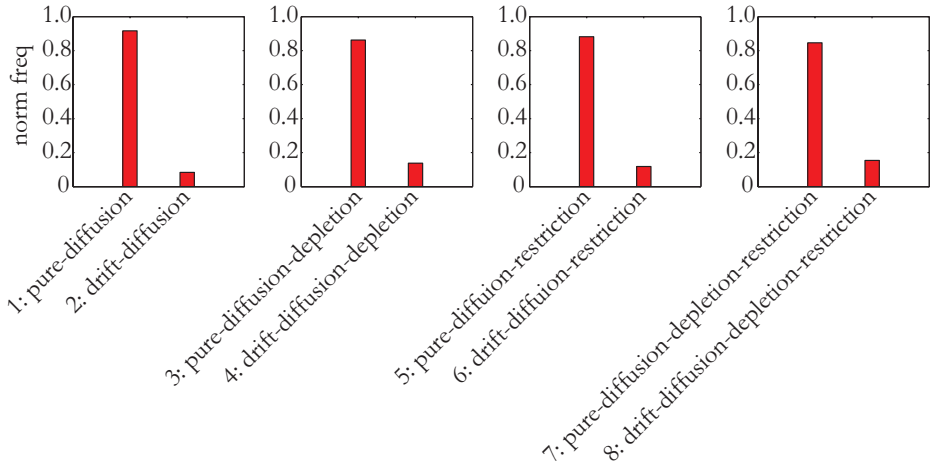

Figure B.1: **Model comparisons** between the pure-diffusion and drift-diffusion version of each model type. The Bayes Factor provides substantial evidence for the pure-diffusion version in each case and in the case of the original pair of models the evidence is strong.

	model			
	original	depletion	restriction	combined
drift $\geqslant 0\,\mu m\,min^{-1}$	1.02 strong	0.80 substantial	0.87 substantial	0.74 substantial
drift $\geqslant 0.1\,\mu m\,min^{-1}$	2.79 overwhelming	1.41 strong	1.81 very strong	1.23 strong

Table B.1: **Evidence in favour of the pure-diffusion model via logarithm of Bayes factor** The evidence is shown for the pure-diffusion version over the drift-diffusion version for each model pair. The second row shows values when the minimum allowed drift values in the drift version is not less than $0.1\,\mu m\,min^{-1}$. A logarithm of Bayes factor above 0.5 represents substantial evidence; above 1.0, strong evidence; above 1.5, very strong; and above 2.0, overwhelming evidence.

model	**in-drift** $\mu m\,min^{-1}$	**out-drift** $\mu m\,min^{-1}$	**diffusivity** $\mu m^2\,min^{-1}$	**depletion** $10^{-3} min^{-1}$	**restriction** %
1	- -	- -	36 (27-50)	- -	- -
2	- -	0.05 (0-0.1)	28 (11-52)	- -	- -
3	0.7 (0.05-2.90)	- -	41 (29-78)	60 (1-97)	- -
4	1.2 (0.1-4.2)	0.06 (0-0.21)	37 (18-184)	51 (0-92)	- -
5	- -	- -	40 (29-56)	- -	25 (3-51)
6	- -	0.06 (0-0.15)	33 (15-67)	-	37 (5-80)
7	0.74 (0.04-2.8)	- -	44 (32-67)	60 (8-97)	24 (2-56)
8	1.2 (0.08-4.2)	0.07 (0-0.17)	39 (18-83)	60 (3.6-96)	38 (4-77)

Table B.2: **Summary of parameter estimates for Dataset B** for all models considered in Chapters 4 and 5. 1, pure-diffusion; 2, drift-diffusion; 3, pure-diffusion-depletion; 4, drift-diffusion-depletion; 5, pure-diffusion-restriction; 6, drift-diffusion-restriction; 7, pure-diffusion-depletion-restriction; 8, drift-diffusion-depletion-restriction. The 90% confidence intervals are shown in brackets.

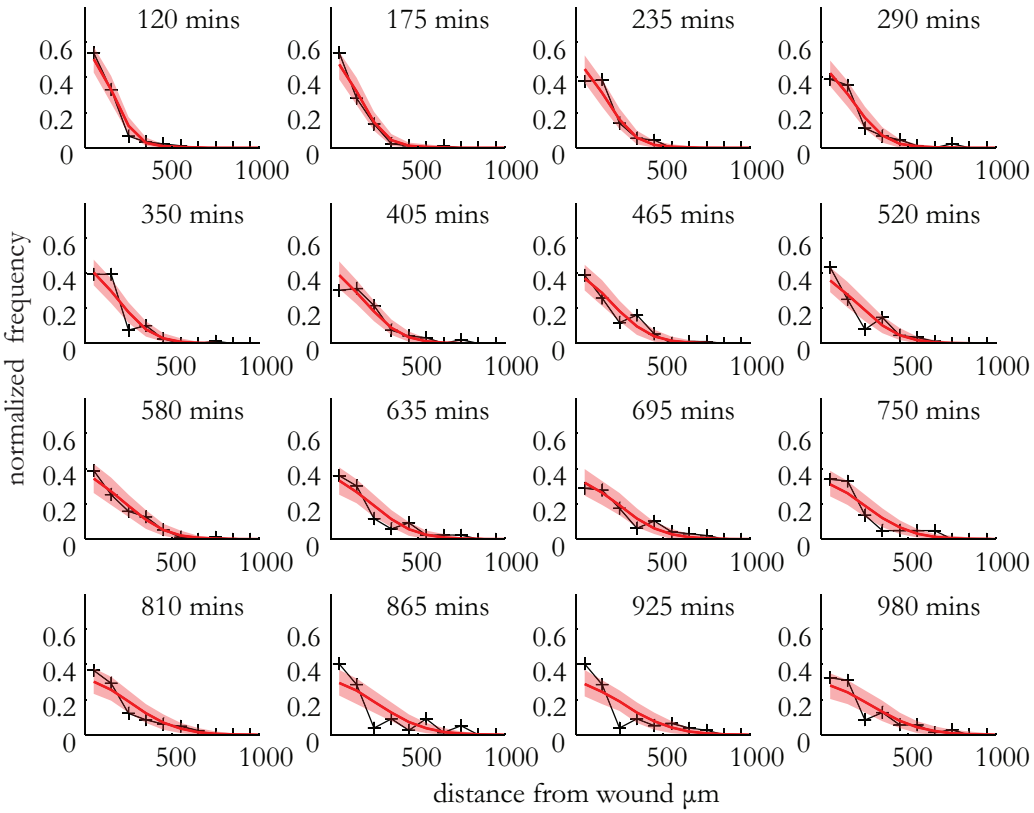

Figure B.2: **Simulation comparisons between the estimated pure-diffusion model, Model 1, and the experimental data.** One thousand simulations were made of the estimated model using the MAP parameter estimate. The output is shown for representative times throughout the experiment. The red line shows the mean predicted distribution of cell positions and the shaded red region is the 90% confidence interval. The experimental observations are also plotted in black for comparison.

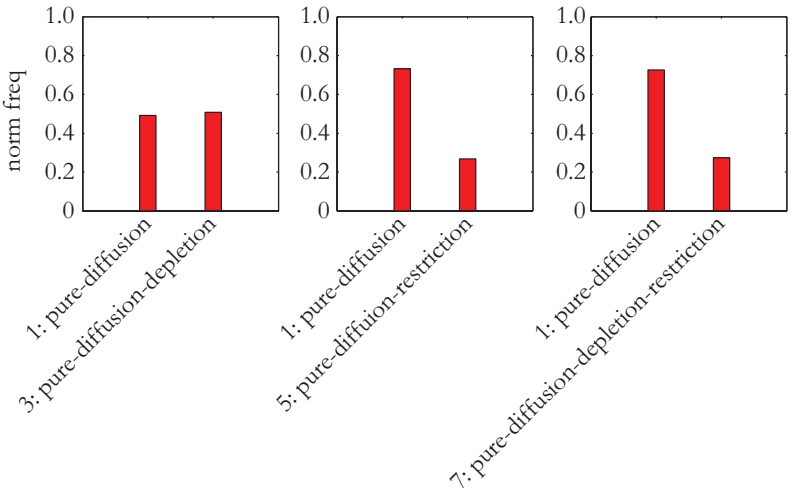

Figure B.3: **Model comparisons** between the original pure-diffusion model and the pure-diffusion version of the more complex models. In two cases the logarithm of Bayes factor evidence for the original pure-diffusion model is just below the 0.5 threshold for substantial evidence. In the other case there is no evidence to choose between the two models.

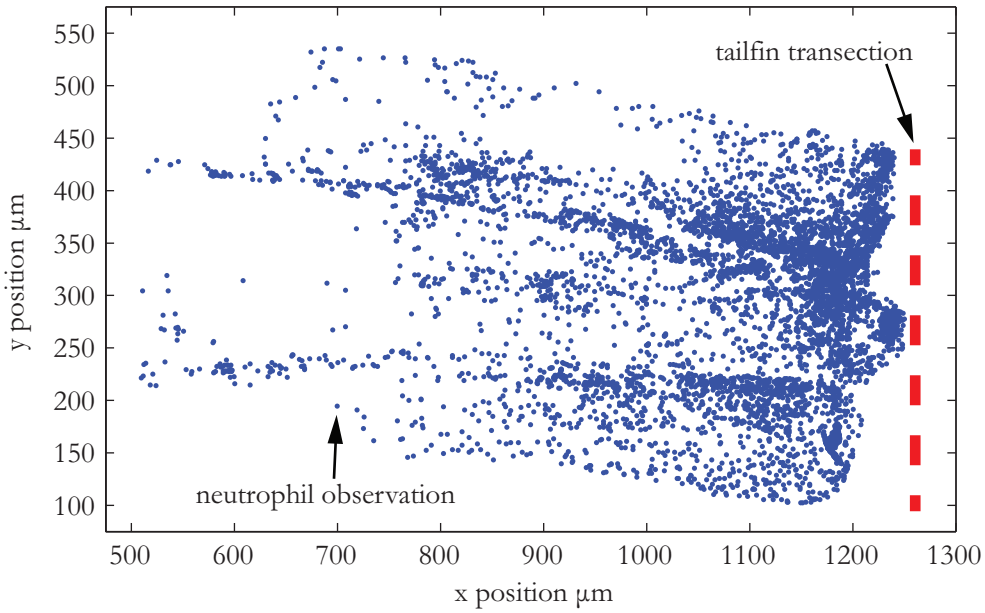

Figure B.4: **Neutrophil migration locations** For a representative individual zebrafish from those in Dataset B, the location of all photoconverted neutrophils are shown at all timepoint throughout the experiment. Comparing this to Figure 5.8, there is here less evidence of preferential migration channels than was the case in Dataset A.

Appendix C

Modelling the chemoattractant concentration

Neutrophils are known to align themselves with certain attractant chemicals and subsequently migrate in the direction of increasing concentration [136]. This appendix gives support for an assumption made in the thesis which takes the gradient of the chemoattractant field to be linear. If we assume the attractant diffuses away from a source (i.e. the wound in the case of the zebrafish experimental data) then it obeys the usual diffusion equation [27],

$$\frac{\partial A}{\partial t} = D_a \frac{\partial^2 A}{\partial x^2}, \tag{C.1}$$

where $A(x,t)$ is the concentration of attractant and D_a is its diffusivity. If we assume a constant input of attractant from a source at $x = 0$ this is represented by the boundary condition,

$$\frac{\partial A}{\partial x}(0,t) = -w. \tag{C.2}$$

If we assume further that the initial concentration is zero and the concentration decays to zero outside a certain distance we also have the initial and boundary conditions

$$A(x,0) = 0, \qquad A(L,t) = 0. \tag{C.3}$$

Assuming high diffusivity a steady state will quickly be reached but in any case steady state means no change with respect to time, i.e.

$$\frac{\partial A}{\partial t} = 0. \tag{C.4}$$

Thus (C.1) requires

$$\frac{\partial^2 A}{\partial x^2} = 0 \tag{C.5}$$

Following integration, this means the first derivative with respect to x is constant and so using (C.2) and (C.3),

$$\frac{\partial A}{\partial x} = -w \tag{C.6}$$

$$A(x,t) = w(L-x) \tag{C.7}$$

at steady state. Any change in input of attractant will disturb this equilibrium. But if the diffusivity is fast in comparison to these changes, equilibrium will be regained quickly and lowering or raising the input will respectively have the effect of lowering or raising the gradient throughout the field.

Acronyms

ABC approximate Bayesian computation. 4, 29–38, 67, 68, 71, 74, 75, 80, 84, 91, 116, 118, 121

ABC-MCMC approximate Bayesian computation Markov chain Monte Carlo. 34, 35, 67

ABC-SMC approximate Bayesian computation sequential Monte Carlo. 32, 36, 37, 67, 68, 74, 76, 80, 81, 84, 87, 88, 109, 121

BD Bhattacharyya distance. 71, 72

BRW biased random walk. 44, 52

COPD chronic obstructive pulmonary disease. 1, 9, 84

CRW correlated random walk. 44, 59

dpf days post fertilisation. 40, 84

GFP green fluorescent protein. 14, 41, 84

GLS generalised least squares. 24

hpi hours post injury. 49

KLD Kullback–Leibler divergence. 71, 72

LS least squares. 22, 24, 26, 28, 33

MAP maximum *a posteriori*. 28, 88, 96, 97, 106, 107, 109, 111, 117

MCMC Markov chain Monte Carlo. 32, 33, 35, 66, 67

MH Metropolis Hastings. 33–35

ML maximum likelihood. 25, 26, 28, 30

ODE ordinary differential equation. 10, 11

OLS ordinary least squares. 23, 24, 26

PDE partial differential equation. 3, 4, 20

PMN polymorphonuclear leukocyte. 7

RLS regularised least squares. 24, 28

SIS sequential importance sampling. 35, 36

SMC sequential Monte Carlo. 67

SRW simple random walk. 44

WLS weighted least squares. 24, 45, 46

Bibliography

[1] U. Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman & Hall, 2007.

[2] W. Alt. Biased random walk models for chemotaxis and related diffusion approximations. *Journal of Mathematical Biology*, 9(2):147–177, 1980.

[3] B. Amulic, C. Cazalet, G.L. Hayes, K.D. Metzler, and A. Zychlinsky. Neutrophil function: From mechanisms to disease. *Annual Review of Immunology*, 30(1):459–489, 2012.

[4] R. Ando, H. Hama, M. Yamamoto-Hino, H. Mizuno, and A. Miyawaki. An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein. *Proceedings of the National Academy of Sciences*, 99(20): 12651–12656, 2002.

[5] C. Arrieumerloua and T. Meyer. A local coupling model and compass parameter for eukaryotic chemotaxis. *Developmental Cell*, 8:215–227, 2005.

[6] Y. Bar-Shalom and X.R. Li. *Estimation and Tracking: Principles, Techniques and Software*. Artech House, 1993.

[7] V.H. Barocas and R.T. Tranquillo. An anisotropic biphasic theory of tissue-equivalent mechanics: The interplay among cell traction, fibrillar network deformation, fibril alignment, and cell contact guidance. *Journal of Biomechanical Engineering*, 119(2):137, 1997.

[8] M. Basseville. Distance measures for signal processing and pattern recognition. *Signal Processing*, 18(4):349–369, December 1989.

[9] T. Bayes. An essay toward solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53:370–418, 1763.

[10] M.A. Beaumont. Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics*, 41:379 – 406, 2010.

[11] M.A. Beaumont, W. Zhang, and D.J. Balding. Approximate Bayesian computation in popoulation genetics. *Genetics*, 162:2025–2035, 2002.

[12] M.A. Beaumont, J.M. Cornuet, J.M. Marin, and C.P. Robert. Adaptive approximate Bayesian computation. *Biometrika*, 96(4):983–990, January 2009.

[13] A. Bhattacharyya. On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society*, 35:99–109, 1943.

[14] C.M. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag New York Inc, 2007.

[15] N. Borregaard. Neutrophils, from marrow to microbes. *Immunity*, 33(5):657–670, November 2010.

[16] N. Borregaard and J.B. Cowland. Granules of the human neutrophilic polymorphonuclear leukocyte. *Blood*, 89(10):3503–3521, May 1997.

[17] P. Bortot, S.G. Coles, and S.A. Sisson. Inference for stereological extremes. *Journal of the American Statistical Association*, 102(477):84–92, March 2007.

[18] G.E.P. Box and N.R. Draper. *Empirical Model-Building and Response Surfaces*. Wiley, 1 edition, January 1987.

[19] V. Brinkmann, U. Reichard, C. Goosmann, B. Fauler, Y. Uhlemann, D.S. Weiss, Y. Weinrauch, and A. Zychlinsky. Neutrophil extracellular traps kill bacteria. *Science*, 303:1532–1535, 2004.

[20] S.B. Brown, C.S. Tucker, C. Ford, Y. Lee, D.R. Dunbar, and J.J. Mullins. Class III antiarrhythmic methanesulfonanilides inhibit leukocyte recruitment in zebrafish. *Journal of Leukocyte Biology*, 82(1):79–84, July 2007.

[21] C.D. Buckley, E.A. Ross, H.M. McGettrick, C.E. Osborne, O. Haworth, C. Schmutz, P.C.W. Stone, M. Salmon, N.M. Matharu, R.K. Vohra, G.B. Nash, and G.E. Rainger. Identification of a phenotypically and functionally distinct population of long-lived neutrophils in a model of reverse endothelial migration. *Journal of Leukocyte Biology*, 79(2):303–311, February 2006.

[22] E. Cameron and A.N. Pettitt. Approximate Bayesian computation for astronomical model analysis: A case study in galaxy demographics and morphological transformation at high redshift. *arXiv:1202.1426*, February 2012.

[23] S.H. Cha and S.N. Srihari. On measuring the distance between histograms. *Pattern Recognition*, 35(6):1355–1370, June 2002.

[24] F. Chalub, Y. Dolak-Struss, P. Markowich, D. Oelz, C. Schmeiser, and A. Soreff. Model hierarchies for cell aggregation by chemotaxis. *Mathematical Models and Methods in Applied Sciences*, 16(supp01):1173–1197, July 2006.

[25] S. Chib and E. Greenberg. Understanding the metropolis-hastings algorithm. *The American Statistician*, 49:327–335, 1995.

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

[27] M.P. Coleman. *An Introduction to Partial Differential Equations with MATLAB*. CRC Press, 2005.

[28] E. Colucci-Guyon, J.Y. Tinevez, S.A. Renshaw, and P. Herbomel. Strategies of professional phagocytes in vivo: unlike macrophages, neutrophils engulf only surface-associated microbes. *Journal of Cell Science*, 124(Pt 18):3053–3059, September 2011.

[29] J. Cornuet, F. Santos, M.A. Beaumont, C.P. Robert, J. Marin, D.J. Balding, T. Guillemaud, and A. Estoup. Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics*, 24(23):2713–2719, December 2008.

[30] J.M. Davison, C.M. Akitake, M.G. Goll, J.M. Rhee, N. Gosse, H. Baier, M.E. Halpern, S.D. Leach, and M.J. Parsons. Transactivation from Gal4-VP16 transgenic insertions for tissue-specific cell labeling and ablation in zebrafish. *Developmental Biology*, 304(2):811–824, 2007.

[31] J. Day, J. Rubin, Y. Vodovotz, C.C. Chow, A. Reynolds, and G. Clermont. A reduced mathematical model of the acute inflammatory response II. capturing scenarios of repeated endotoxin administration. *Journal of Theoretical Biology*, 242(1):237–256, September 2006.

[32] P. Devreotes and C. Janetopoulos. Eukaryotic chemotaxis: distinctions between directional sensing and polarization. *Journal of Biological Chemistry*, 278:20445–20448, 2003.

[33] R. Dickinson and R. Tranquillo. Optimal estimation of cell movement indices from the statistical analysis of cell tracking data. *AIChE journal*, 39(12):199–2010, 1993.

[34] R.B. Dickinson and R.T. Tranquillo. Transport equations and indices for random and biased cell migration based on single cell properties. *SIAM Journal on Applied Mathematics*, 55(5):1419–1454, October 1995.

[35] J.L. Doob. The brownian movement and stochastic equations. *The Annals of Mathematics*, 43:351–369, 1942.

[36] R. Duffin, A.E. Leitch, S. Fox, C. Haslett, and A.G. Rossi. Targeting granulocyte apoptosis: mechanisms, models, and therapies. *Immunological Reviews*, 236(1):28–40, 2010.

[37] G.A. Dunn and A.F. Brown. A unified approach to analyzing cell motility. *Journal of Cell Science Supplement*, 8:81–102, 1987.

[38] P.M. Elks, C.A. Loynes, and S.A. Renshaw. Measuring inflammatory cell migration in the zebrafish. *Methods in Molecular Biology (Clifton, N.J.)*, 769: 261–275, 2011.

[39] P.M. Elks, F.J. van Eeden, G. Dixon, X. Wang, C.C. Reyes-Aldasoro, P.W. Ingham, M.K.B. Whyte, S.R. Walmsley, and S.A. Renshaw. Activation of hypoxia-inducible factor-1 (Hif-1) delays inflammation resolution by reducing neutrophil apoptosis and reverse migration in a zebrafish inflammation model. *Blood*, 118(3):712–722, July 2011.

[40] F. Ellett, L. Pase, J.W. Hayman, A. Andrianopoulos, and G.J. Lieschke. mpeg1 promoter transgenes direct macrophage-lineage expression in zebrafish. *Blood*, 117(4):e49–e56, January 2011.

[41] Y. Epshteyn. Discontinuous galerkin methods for the chemotaxis and haptotaxis models. *Journal of Computational and Applied Mathematics*, 224(1): 168–181, February 2009.

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

[43] P. Follin. Skin chamber technique for study of *in vivo* exudated human neutrophils. *Journal of Immunological Methods*, 232:55–65, 1999.

[44] E.F. Foxman, J.J. Campbell, and E.C. Butcher. Multistep navigation and the combinatorial control of leukocyte chemotaxis. *Journal of Cell Biology,*, 139: 1349–1360, 1997.

[45] E.F. Foxman, E.J. Kunkel, and E.C. Butcher. Integrating conflicting chemotactic signals: The role of memory in leukocyte navigation. *The Journal of Cell Biology*, 147:577–587, 1999.

[46] B. Franz and R. Erban. Hybrid modelling of individual movement and collective behaviour. http://eprints.maths.ox.ac.uk/1393/, 2011.

[47] A. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin. *Bayesian Data Analysis*. Chapman & Hall, 2004.

[48] D.L. Guyader, M.J. Redd, E. Colucci-Guyon, E. Murayama, K. Kissa, V. Briolat, E. Mordelet, A. Zapata, H. Shinomiya, and P. Herbomel. Origins and unconventional behavior of neutrophils in developing zebrafish. *Blood*, 111(1): 132–141, January 2008.

[49] C. Hall, M.V. Flores, A. Chien, A. Davidson, K. Crosier, and P. Crosier. Transgenic zebrafish reporter lines reveal conserved toll-like receptor signaling potential in embryonic myeloid leukocytes and adult immune cell lineages. *Journal of Leukocyte Biology*, 85(5):751–765, 2009.

[50] M.B. Hampton, A.J. Kettle, and C.C. Winterbourn. Inside the neutrophil phagosome: Oxidants, myeloperoxidase, and bacterial killing. *Blood*, 92(9): 3007–3017, January 1998.

[51] T. Hillen and K.J. Painter. A users guide to PDE models for chemotaxis. *Mathematical Biology*, 58:183–217, 2009.

[52] G.R. Holmes, S.R. Anderson, G. Dixon, A.L. Robertson, C.C. Reyes-Aldasoro, S.A. Billings, S.A. Renshaw, and V. Kadirkamanathan. Repelled from the wound, or randomly dispersed? Reverse migration behaviour of neutrophils characterized by dynamic modelling. *Journal of The Royal Society Interface*, September 2012.

[53] G. Holmes, G. Dixon, S. Anderson, C. Reyes-Aldasoro, P. Elks, S. Billings, M. Whyte, V. Kadirkamanathan, and S. Renshaw. Drift-diffusion analysis of neutrophil migration during inflammation resolution in a zebrafish model. *Advances in Hematology*, 2012.

[54] J. Hughes, R.J. Johnson, A. Mooney, C. Hugo, K. Gordon, and J. Savill. Neutrophil fate in experimental glomerular capillary injury in the rat. *American Journal of Pathology*, 150:223–234, 1997.

[55] A. Huttenlocher and M.C. Poznansky. Reverse leukocyte migration can be attractive or repulsive. *Trends in Cell Biology*, 18(6):298–306, June 2008.

[56] P.A. Iglesias and P.N. Devreotes. Navigating through models of chemotaxis. *Current Opinion in Cell Biology*, 20:35–40, 2008.

[57] E.L. Ionides, K.S. Fang, R.R. Isseroff, and G.F. Oster. Stochastic models for cell motion and taxis. *Journal of Mathematical Biology*, 48:23–37, 2004.

[58] A. Jasra, S. Singh, J. Martin, and E. McCoy. Filtering via approximate Bayesian computation. *Statistics and Computing*, pages 1–15, 2010.

[59] H. Jeffreys. *Theory of Probability*. Oxford University Press, USA, 3rd edition, November 1998.

[60] A. Jilkine and L. Edelstein-Keshet. A comparison of mathematical models for polarization of single eukaryotic cells in response to guided cues. *PLoS Computational Biology*, 7(4):e1001121, April 2011.

[61] J.B. Johnson and K.S. Omland. Model selection in ecology and evolution. *Trends in Ecology & Evolution*, 19(2):101–108, February 2004.

[62] H.A. Jones, R.J. Clark, C.G. Rhodes, J.B. Schofield, T. Krausz, and C. Haslett. In vivo measurement of neutrophil activity in experimental lung inflammation. *American Journal of Respiratory and Critical Care Medicine*, 149(6):1635–1639, June 1994.

[63] H.A. Jones, S. Sriskandan, A.M. Peters, N.B. Pride, T. Krausz, A.R. Boobis, and C. Haslett. Dissociation of neutrophil emigration and metabolic activity in lobar pneumonia and bronchiectasis. *The European respiratory journal: official journal of the European Society for Clinical Respiratory Physiology*, 10 (4):795–803, April 1997.

[64] P. Joyce and P. Marjoram. Approximately sufficient statistics and Bayesian computation. *Statistical Applications in Genetics and Molecular Biology*, 7(1), January 2008.

[65] V. Kadirkamanathan, S.R. Anderson, S.A. Billings, X. Zhang, G.R. Holmes, C.C. Reyes-Aldasoro, P.M. Elks, and S.A. Renshaw. The neutrophil's eye-view: Inference and visualisation of the chemoattractant field driving cell chemotaxis in vivo. *PLoS ONE*, 7(4):e35182, April 2012.

[66] T. Kariya and H. Kurata. *Generalized Least Squares*. Wiley, 1st edition, July 2004.

[67] R.E. Kass and A.E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, June 1995.

[68] E.F. Keller and L.A. Segel. Model for chemotaxis. *Journal of Theoretical Biology*, 30:225–234, 1971.

[69] S. Kullback and R.A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, March 1951.

[70] D.A. Lauffenburger and A.F. Horwitz. Cell migration: A physically integrated molecular process. *Cell*, 84(3):359–369, February 1996.

[71] C.L. Lawson and R.J. Hanson. *Solving Least Squares Problems*. SIAM, 1995.

[72] D.C. Lay. *Linear Algebra and Its Applications*. Addison-Wesley, 2nd edition, July 1997.

[73] L. Li, S.F. Nørrelykke, and E.C. Cox. Persistent cell motion in the absence of external signals: A search strategy for eukaryotic cells. *PLoS ONE*, 3(5):e2093, May 2008.

[74] J. Liepe, C. Barnes, E. Cule, K. Erguler, P. Kirk, T. Toni, and M.P.H. Stumpf. ABC-SysBio approximate Bayesian computation in python with GPU support. *Bioinformatics*, 26(14):1797–1799, July 2010.

[75] J. Liepe, H. Taylor, C. Barnes, M. Huvet, L. Bugeon, T. Thorne, J. Lamb, M. Dallman, and M.P.H. Stumpf. Calibrating spatio-temporal models of leukocyte dynamics against in vivo live-imaging data using approximate Bayesian computation. *Integrative Biology*, 4(3):335–345, 2012.

[76] G.J. Lieschke, A.C. Oates, M.O. Crowhurst, A.C. Ward, and J.E. Layton. Morphologic and functional characterization of granulocytes and macrophages in embryonic and adult zebrafish. *Blood*, 98(10):3087–3096, November 2001.

[77] L. Ljung. *System Identification - Theory for the User*. Prentice Hall, Upper Saddle River, NJ, 2nd edition, 1999.

[78] M.A. Lokuta, P.A. Nuzzi, and A. Huttenlocher. Calpain regulates neutrophil chemotaxis. *Proceedings of the National Academy of Sciences of the United States of America*, 100(7):4006–4011, April 2003.

[79] J.S. Lopes, D. Balding, and M.A. Beaumont. PopABC: a program to infer historical demographic parameters. *Bioinformatics*, August 2009.

[80] C.A. Loynes, J.S. Martin, A. Robertson, D.M.I. Trushell, P.W. Ingham, M.K.B. Whyte, and S.A. Renshaw. Pivotal advance: Pharmacological manipulation of inflammation resolution during spontaneously resolving tissue

neutrophilia in the zebrafish. *Journal of Leukocyte Biology*, 87(2):203 –212, February 2010.

[81] D.J.C. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[82] P. Marjoram, J. Molitor, V. Plagnol, and S. Tavare. Markov chain monte carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States of America*, 100(26):15324–15328, December 2003.

[83] J.S. Martin, A. Jasra, S.S. Singh, N. Whiteley, and E. McCoy. Approximate Bayesian computation for smoothing. *arXiv:1206.5208*, June 2012.

[84] J.S. Martin and S.A. Renshaw. Using in vivo zebrafish models to understand the biochemical basis of neutrophilic respiratory disease. *Biochemical Society Transactions*, 37(Pt 4):830–837, August 2009.

[85] J.R. Mathias, B.J. Perrin, T.X. Liu, J. Kanki, A.T. Look, and A. Huttenlocher. Resolution of inflammation by retrograde chemotaxis of neutrophils in transgenic zebrafish. *Journal of Leukocyte Biology*, 80(6):1281–1288, 2006.

[86] J.R. Mathias, M.E. Dodd, K.B. Walters, J. Rhodes, J.P. Kanki, A.T. Look, and A. Huttenlocher. Live imaging of chronic inflammation caused by mutation of zebrafish hai1. *Journal of cell science*, 120(Pt 19):3372–3383, October 2007.

[87] B. McDonald, K. Pittman, G.B. Menezes, S.A. Hirota, I. Slaba, C.C.M. Waterhouse, P.L. Beck, D.A. Muruve, and P. Kubes. Intravascular danger signals guide neutrophils to sites of sterile inflammation. *Science*, 330(6002):362–366, October 2010.

[88] V. Mulero, M.P. Sepulcre, G.E. Rainger, and C.D. Buckley. Editorial: Neutrophils live on a Two-Way street. *Journal of Leukocyte Biology*, 89(5):645–647, January 2011.

[89] D. Murphy and M. Davidson. Differential interference contrast microscopy and modulation contrast microscopy. *Fundamentals of Light Microscopy and Electronic Imaging, 2nd Edition*, pages 173–197, 2012.

[90] C. Nathan. Neutrophils and immunity: challenges and opportunities. *Nature Reviews. Immunology*, 6(3):173–182, 2006.

[91] M.P. Neilson, D.M. Veltman, P.J.M. van Haastert, S.D. Webb, J.A. Mackenzie, and R.H. Insall. Chemotaxis: A feedback-based computational model robustly

predicts multiple aspects of real cell behaviour. *PLoS Biology*, 9(5):e1000618, May 2011.

[92] O. Nelles. *Nonlinear System Identification.* Springer, 2001.

[93] C. Nicholson. Diffusion and related transport mechanisms in brain tissue. *Reports On Progress In Physics*, 64:815–884, 2001.

[94] P. Niethammer, C. Grabher, A.T. Look, and T.J. Mitchison. A tissue-scale gradient of hydrogen peroxide mediates rapid wound detection in zebrafish (letter). *Nature*, 459:996–999, 2009.

[95] V. Nissen and J. Propach. On the robustness of population-based versus point-based optimization in the presence of noise. *IEEE Transactions on Evolutionary Computation*, 2(3):107–119, September 1998.

[96] R. Nossal and S.H. Zigmond. Chemotropism indices for polymorphonuclear leukocytes. *Biophysical Journal*, 16:1171–1182, 1976.

[97] C. Nüsslein-Volhard and R. Dahm. *Zebrafish: A Practical Approach.* Oxford University Press, 2002.

[98] B.K. Øksendal. *Stochastic Differential Equations: An Introduction with Applications.* Springer, 5th edition, November 2002.

[99] R.S. Parker and G. Clermont. Systems engineering medicine: engineering the inflammation response to infectious and traumatic challenges. *Journal of The Royal Society Interface*, 7(48):989–1013, June 2010.

[100] C.S. Patlak. Random walk with persistence and external bias. *Bulletin of Mathematical Biology*, 15(3):311–338, 1953.

[101] C.S. Patlak. The effect of the previous generation on the distribution of gene frequencies in populations. *Proceedings of the National Academy of Sciences of the United States of America*, 39(10):1063–1068, October 1953.

[102] A.A. Potdar, J. Jeon, A.M. Weaver, V. Quaranta, and P.T. Cummings. Human mammary epithelial cells exhibit a bimodal correlated random walk pattern. *PloS One*, 5(3):e9636, 2010.

[103] J.K. Pritchard, M.T. Seielstad, A. Perez-Lezaun, and M.W. Feldman. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution*, 16(12):1791–1798, January 1999.

[104] H. Qian, M. Sheetz, and E. Elson. Single particle tracking. analysis of diffusion and flow in two-dimensional systems. *Biophysical Journal*, 60:910–921, 1991.

[105] O. Ratmann, O. Jørgensen, T. Hinkley, M. Stumpf, S. Richardson, and C. Wiuf. Using likelihood-free inference to compare evolutionary dynamics of the protein networks of h. pylori and p. falciparum. *PLoS Computational Biology*, 3(11):e230, November 2007.

[106] O. Ratmann, P. Pudlo, S. Richardson, and C. Robert. Monte carlo algorithms for model assessment via conflicting summaries. *arXiv:1106.5919*, June 2011.

[107] S. Renshaw, C. Loynes, D. Trushell, S. Elworthy, P. Ingham, and M. Whyte. A transgenic zebrafish model of neutrophilic inflammation. *Blood*, 108(13): 3976–3978, December 2006.

[108] S.A. Renshaw and N.S. Trede. A model 450 million years in the making: zebrafish and vertebrate immunity. *Disease Models & Mechanisms*, 5(1):38–47, January 2012.

[109] C.P. Robert, J. Cornuet, J. Marin, and N.S. Pillai. Lack of confidence in approximate Bayesian computation model choice. *Proceedings of the National Academy of Sciences*, 108(37):15112–15117, September 2011.

[110] D.B. Rubin. Bayesianly justifiable and relevant frequency calculations for the applies statistician. *The Annals of Statistics*, 12(4):1151–1172, December 1984.

[111] C.N. Serhan, S.D. Brain, C.D. Buckley, D.W. Gilroy, C. Haslett, L.A.J. O'Neill, M. Perretti, A.G. Rossi, and J.L. Wallace. Resolution of inflammation: state of the art, definitions and terms. *FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology*, 21(2):325–332, February 2007.

[112] M.P. Sheetz, S. Turney, H. Qian, and E.L. Elson. Nanometre-level analysis demonstrates that lipid flow does not drive membrane glycoprotein movements. *Nature*, 340:284–288, 1989.

[113] B.W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall/CRC, 1st edition, April 1986.

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

[115] S.A. Sisson, Y. Fan, and M.M. Tanaka. Correction for Sisson et al., sequential monte carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 106(39):16889–16889, September 2009.

[116] T. Soderstrom and P. Stoica. *System Identification*. Prentice Hall International, 1989.

[117] O. Soehnlein and L. Lindbom. Phagocyte partnership during the onset and resolution of inflammation. *Nature Reviews Immunology*, 10(6):427–439, June 2010.

[118] T.W. Starnes and A. Huttenlocher. Neutrophil reverse migration becomes transparent with zebrafish. *Advances in Hematology*, 2012:1–11, 2012.

[119] C.L. Stokes, D. Lauffenburger, and S.K. Williams. Migration of individual microvessel endothelial cells: stochastic model and parameter measurement. *Journal of Cell Science 419-430*, 99:419–430, 1991.

[120] S. Tavaré, D.J. Balding, R.C. Griffiths, and P. Donnelly. Inferring coalescence times from DNA sequence data. *Genetics*, 145(2):505–518, January 1997.

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

[122] R.T. Tranquillo, D.A. Lauffenburger, and S.H. Zigmond. A stochastic model for leukocyte random motility and chemotaxis based on receptor binding fluctuations. *The Journal of Cell Biology*, 106(2):303–309, February 1988.

[123] B.M. Turner and T. Van Zandt. A tutorial on approximate Bayesian computation. *Journal of Mathematical Psychology*, 56(2):69–85, April 2012.

[124] G.E. Uhlenbeck and L.S. Ornstein. On the theory of the brownian motion. *Physical Review*, 36:823–841, 1930.

[125] L. Uller, C.G. Persson, and J.S. Erjeflt. Resolution of airway disease: removal of inflammatory cells through apoptosis, egression or both? *Trends in Pharmacological Sciences*, 27(9):461–466, 2006.

[126] P.J.M. Van Haastert. A model for a correlated random walk based on the ordered extension of pseudopodia. *PLoS Computational Biology*, 6(8):e1000874, 2010.

[127] F. Vianello, I.T. Olszak, and M.C. Poznansky. Fugetaxis: active movement of leukocytes away from a chemokinetic agent. *Journal of Molecular Medicine (Berlin, Germany)*, 83(10):752–763, October 2005.

[128] K.B. Walters, J.M. Green, J.C. Surfus, S.K. Yoo, and A. Huttenlocher. Live imaging of neutrophil motility in a zebrafish model of WHIM syndrome. *Blood*, 116(15):2803–2811, October 2010.

[129] D. Wegmann, C. Leuenberger, S. Neuenschwander, and L. Excoffier. ABC-toolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics*, 11:116, 2010.

[130] P.C. Wilkinson. Assays of leukocyte locomotion and chemotaxis. *Journal of Immunological Methods*, 216:139–153, 1998.

[131] A. Woodfin, M. Voisin, M. Beyrau, B. Colom, D. Caille, F. Diapouli, G.B. Nash, T. Chavakis, S.M. Albelda, G.E. Rainger, P. Meda, B.A. Imhof, and S. Nourshargh. The junctional adhesion molecule JAM-C regulates polarized transendothelial migration of neutrophils in vivo. *Nature Immunology*, 12(8): 761–769, 2011.

[132] World Health Organisation. Global surveillance, prevention and control of chronic respiratory diseases: a comprehensive approach, 2007.

[133] S.K. Yoo and A. Huttenlocher. Spatiotemporal photolabeling of neutrophil trafficking during inflammation in live zebrafish. *Journal of Leukocyte Biology*, 89(5):661–667, 2011.

[134] S.H. Zigmond and D.A. Lauffenburger. Assays of leukocyte chemotaxis. *Annual Review of Medicine*, 37(1):149–155, 1986.

[135] S. Zigmond. Ability of polymorphonuclear leukocytes to orient in gradients of chemotactic factors. *Journal of Cell Biology*, 75:606–616, 1977.

[136] S.H. Zigmond. Chemotaxis by polymorphonuclear leukocytes. *Journal of Cell Biology*, 77:269–287, 1978.