

Imperial College London
Department of Life Sciences

**Modelling haematopoietic stem cells in their
niche**

Adam L MacLean

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy of Imperial College London

All of the work presented in this thesis is my own, except where clearly stated otherwise.

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it.

Abstract

Modelling haematopoietic stem cells (HSCs) mathematically allows us to probe their behaviour, test hypotheses and make predictions. HSCs are essential and elusive; despite many recent advances much remains unknown, including how the microenvironment (niche) influences HSC behaviour, and the nature of complex interactions between HSCs, the niche and disease.

This thesis comprises three sections. In the first we present new methods for the analysis of ODE systems that allow us to characterise steady state properties such as the probability of a state being stable (over some parameter range). We apply these methods to models of stem cell dynamics that differ in the form of regulation (feedback) imposed on the system and study the steady states of these systems.

Competition within the niche between HSCs and invading cancer cells disrupts the system. In the second section we model this using ODEs that describe the population dynamics of cellular species from an ecological perspective. From the analysis of two models differing in their treatment of the niche, using Bayesian inference, we find that maintaining a viable HSC population is necessary and almost sufficient in order to outcompete leukaemia and restore healthy haematopoiesis.

In the third section we extend the analysis of interactions between HSCs, leukaemia and the niche: performing a comparison of three models that describe the dynamics of chronic myeloid leukaemia. We study heterogeneous data from a clinical trial of the disease. All models fit these data, but do so in different ways. One model is discarded due to its unrealistic predictions, suggesting that direct competition between species is important. Each of the remaining two models makes testable predictions, but validating these is at current experimental limits. In the future, the results presented will aid experimental design and provide a framework from which to identify new therapeutic targets for haematopoietic diseases.

Isn't this enough? Just this world? Just this beautiful, complex,
wonderfully unfathomable world?

– Tim Minchin

Acknowledgements

I thank my supervisor, Michael Stumpf, for all the support, without which I would not have begun my PhD, let alone approach its completion. Your encouragement and guidance in pursuit of systems biological endeavours, and much beyond, have been invaluable to me every step of the way. I thank also my second supervisor Cristina Lo Celso, for welcoming me into her lab, helping to guide my projects all along the way, and patiently teaching me so much biology.

My PhD research was conducted far from isolation, and I am grateful to the following collaborators who have each made significant contributions: Paul Kirk and Delphine Rolando on Chapter 4, and Sarah Filippi on Chapter 6. I also thank Ingo Roeder who kindly supplied data that were used in Chapter 6. In addition to collaborative efforts, thank you Paul for guidance throughout, for many good ideas that have made an impact on my work, and for sharing in the delights of etymology.

Thank you to all of the Theoretical Systems Biology lab, present and past. I am most grateful for your contributions to my work both academic and social, each playing its own crucial role. From lunches and ping pong to your acceptance of the towel wearing episode, you have made it all a pleasure. Thanks to members of the Lo Celso lab for helping me understand more biology and one or two evenings of light pub relief. Thank you Narges Rashidi, for answering the many, many biological questions of an ex-physicist and for all the stories.

Thank you Myrsini Kaforou and Evan Bellos, with me from the very beginning of Imperial, for your companionship in exploring the science and, more-

over, exploring London. Thank you to Marta Carrara for coping with interruptions whenever my concentration waned and for the music. Thank you Kameswarie Nunna for teaching me control theory and patience. Thanks to Josh Patch and Joe Radford, for always being so close in times of need or beer. And thank you Aleksandra Skolimowska, for making my write up much more of a pleasure than it probably should have been and for rolling along with me.

Thank you to my parents Fran Lavendel and John MacLean, and my sister Laura, for your incredible support, passion and love; you give me the strongest foundation. Finally, thank you to everybody who makes the world a more beautiful place, which is all of you.

Contents

Abstract	ii
Acknowledgements	iv
Contents	vi
List of Tables	xi
List of Figures	xii
List of Abbreviations	1
1 Introduction	2
2 Background: haematopoiesis, leukaemia and the stem cell niche	5
2.1 Overview	5
2.2 Haematopoietic stem cells	6
2.2.1 Haematopoietic stem cell specification	8
2.2.2 Haematopoietic stem cell regulation	9
2.3 The haematopoietic stem cell niche	10

2.4	Leukaemia	13
2.5	Competition within the niche	17
2.6	Conclusion	19
3	Background: theoretical ecology and mathematical modelling	20
3.1	Overview	20
3.2	Lotka and Volterra: the beginnings of theoretical ecology . .	21
3.3	A century of ecological modelling	24
3.4	General strategies for modelling the haematopoietic system .	26
3.5	Modelling haematopoiesis: the state of the art	32
3.5.1	Modelling healthy haematopoiesis	33
3.5.2	Modelling malignant haematopoiesis and cancer . . .	36
3.6	Conclusion	40
4	A new framework for stability analysis and its application to stem cell niche models	41
4.1	Overview	41
4.2	Introduction	42
4.2.1	Motivation	42
4.2.2	Background on linear stability analysis	44
4.2.3	Extending linear stability analysis	47
4.3	Model development	48
4.3.1	Models $S1-S4$: three-species models	50
4.3.2	Models $S5-S10$: five-species models	52
4.4	Methods	57

4.4.1	A generalised stability analysis approach	57
4.4.2	Null distributions to describe the stability of a model	58
4.5	Results	60
4.5.1	Generalised stability analysis enables global characterisation of the stability of a model and its fixed points	60
4.5.2	Structure maintains stability across all S -models	63
4.5.3	Different parameter regimes control the proportion of progenitor and differentiated cells at steady state	68
4.6	Discussion and conclusions	71
5	Qualitative inference and desirable behaviour	77
5.1	Overview	77
5.2	Introduction	78
5.3	Model development	82
5.3.1	Model assumptions	82
5.3.2	Niche competition	84
5.3.3	Model A	86
5.3.4	Model B	87
5.4	Methods	90
5.4.1	Parameter sampling and qualitative inference	90
5.4.2	Robustness analysis via approximate Bayesian computation	94
5.5	Results	96
5.5.1	Simulation of models with fixed parameters gives biologically relevant trajectories	96

5.5.2	Desirable regimes reveal distinct dynamics for HSCs and LSCs	99
5.6	Discussion	104
5.7	Conclusion	107
6	The ecology of chronic myeloid leukaemia	109
6.1	Overview	110
6.2	Introduction	111
6.3	Model development	113
6.3.1	Model <i>M1</i> (MacLean et al., 2013) — niche competition	113
6.3.2	Model <i>M2</i> (Michor et al., 2005) — niche independence	115
6.3.3	Model <i>M3</i> (Foo et al., 2009) — partial niche dependence	117
6.4	Methods	118
6.4.1	Data processing	118
6.4.2	Bayesian parameter inference and model comparison	120
6.5	Results and discussion	121
6.5.1	Models cannot be distinguished based on fit to the data alone	121
6.5.2	The importance of competition (model checking) . .	123
6.5.3	<i>M1</i> and <i>M3</i> offer contrasting mechanisms of leukaemia suppression (model comparison)	125
6.5.4	CML patient outcomes are controlled by different fea- tures between models (model comparison)	129
6.5.5	Minimal descriptions of models <i>M1</i> & <i>M3</i> (model re- duction)	133

6.6	Conclusion	137
7	Conclusions and future work	140
A	Full posterior distributions	146
B	Spatial models of stem cell dynamics	153
B.1	Overview	153
B.2	Introduction	153
B.3	Applications of cellular automata	156
B.3.1	A model of loss of stem cell self renewal	156
B.3.2	A competition model for healthy versus cancer cells .	157
B.4	Discussion and Conclusions	161
	Bibliography	165

List of Tables

4.1	Description of parameters that characterise models <i>S1-S4</i>	52
4.2	Description of parameters that characterise models <i>S5-S10</i>	56
4.3	The stability probabilities for models <i>S1-S10</i>	65
5.1	Description of parameters that describe model A	87
5.2	Description of parameters that describe model B	90
5.3	Explanation of the terms of Equation 5.3 (model B)	91
5.4	Models A and B produce plausible dynamics for heuristically chosen parameter values.	97
6.1	Description of parameters used to characterise models <i>M1</i> and <i>M3116</i>	
6.2	Different treatment of species' niche reliance between models . .	118
6.3	Ranking based on model probability does not discriminate be- tween models	122

List of Figures

2.1	Haematopoiesis: all blood cells are produced from HSCs in a hierarchical process	7
3.1	The basic Lotka-Volterra equations model two species interacting through predation	22
4.1	Illustration of the circular law for a random matrix distribution	47
4.2	Graphical depiction of models $S1-S4$	53
4.3	True, independent and <i>i.i.d.</i> distributions over Jacobians of the Lorenz model	61
4.4	Eigenvalue spectra of the Lorenz model under three random matrix distributions	63
4.5	True, independent and <i>i.i.d.</i> distributions of eigenvalues derived from model $S5$	64
4.6	Graphical representation and eigenvalue spectra of models $S5-S9$	66
4.7	Production rates of progenitor and differentiated cells affect the steady state reached	69

5.1	Diagrammatic representation of stem cell competition within the bone marrow niche	85
5.2	Exploring model space through different parameter sampling regimes	93
5.3	Simulated trajectories for the species of models A and B under different conditions	99
5.4	Properties of the posterior distribution and comparison of parameter posterior values that allow leukaemia suppression for model A	101
5.5	Parameter posterior values that allow for leukaemia suppression under model B	102
6.1	Mechanistic models for haematopoiesis in CML	114
6.2	Experimental measurements of leukaemia progression in 69 patients	119
6.3	Models are not distinguishable based on fits to the data	121
6.4	Simulation of stem cell dynamics demonstrated the importance of niche competition	124
6.5	Dynamics of model species that lead to either remission or relapse of disease	126
6.6	Linear discriminant analysis predicts differences between models in how they control disease outcomes	130
6.7	<i>In silico</i> interventions predict that for $M1$ (but not $M3$) relapse can be avoided	132
6.8	Dynamics of model species for reduced-form versions of $M3$. . .	135

A.1	Posterior distribution for model A	147
A.2	Posterior distribution for model B	148
A.3	Posterior distribution for model M1 under remission condition .	149
A.4	Posterior distribution for model M1 under relapse condition . .	150
A.5	Posterior distribution for model M3 under remission condition .	151
A.6	Posterior distribution for model M3 under relapse condition . .	152
B.1	Update rule for a CA based on nearest neighbours on a 2D grid	155
B.2	Simulation of HSC loss of self renewal dynamics	158
B.3	Simulation of spatial competition model for healthy and cancer cells	160

Abbreviations

ABC	Approximate Bayesian computation
AML	Acute myeloid leukaemia
CA	Cellular automata
CML	Chronic myeloid leukaemia
FACS	Fluorescence-activated cell sorting
HSC	Haematopoietic stem cell
<i>i.i.d.</i>	Independent and identically distributed
LDA	Linear discriminant analysis
LSC	Leukaemia stem cell
ODE	Ordinary differential equation
PDE	Partial differential equation
PCA	Principal components analysis



Introduction

Understanding the biology of haematopoietic stem cells (HSCs) is of great importance. HSCs control the production of all the different types of blood cell through tightly orchestrated processes; by explaining these processes we will learn more about the haematopoietic system and about cell fate choice, cellular regulation and other basic mechanisms. This will have an impact on biomedical research: most directly for haematological disorders, including leukaemia. There is a further case for studying HSCs as a model for other (stem) cell systems, given that haematopoiesis is relatively well characterised and its outputs (blood cells) are abundant and easy to measure.

In Chapter 2 of this thesis, the biology of HSCs is presented with a description of how they are specified and regulated. The microenvironment re-

quired by HSCs to function correctly, known as the stem cell niche, is introduced and our knowledge about it summarised. We also discuss leukaemia and the idea of competition within the niche. This covers all of the biological topics that will be the focus of this work.

We seek to learn about HSCs using models and in Chapter 3 we describe the mathematical framework upon which we base this theoretical investigation. The central concepts used are taken from ecology; these are presented and placed within a historical context. In order to motivate the models that are subsequently developed, we give an overview of different strategies that could be suitable for modelling HSCs, with examples from current literature.

In Chapter 4 we develop methods for stability analysis in order to study models of stem cell niche dynamics. We develop a family of deterministic models for the purposes of this analysis and use them to demonstrate how we can study the linear stability of the steady states of such models in a general, probabilistic way.

Two new models are studied in Chapter 5, for the purpose of understanding how competition plays out in the niche. Specifically, we look at the interactions between HSCs and leukaemia stem cells (LSCs). The new models are based on similar assumptions to those of Chapter 4, but with the extension that they now include a description of cancer in the niche. They also treat interactions with the niche differently. We study these models using a qualitative inference framework that allows us to search globally for behaviour that we wish to observe, namely, the eradication of leukaemia from the niche. We describe conditions under which this is possible and

discuss them in an ecological context.

Chapter 6 continues our study of the population biology of HSCs and LSCs, applied here to the specific example of chronic myeloid leukaemia. We seek to describe data from patients undergoing treatment for the disease using three models, one from the previous chapter and two from literature. Adopting a model comparison approach, we look at where each model can or cannot explain the data and find that although all of the models can explain the data, significant differences in how they do so arise. This leads us to reject one of the models and propose different mechanisms for disease progression based on the other two.

In the final chapter we draw together the various themes of competition, stability and survival in the niche with regards to systems of stem cells and we make some general conclusions. We also discuss directions for future study.



Background: haematopoiesis, leukaemia and the stem cell niche

2.1 Overview

Haematopoietic stem cells (HSCs) are central to this work; in this chapter they are presented and the biology pertinent to their function and maintenance is described. We discuss the challenges of identifying HSCs and describe briefly how they are regulated. We go on to introduce the concept of a stem cell niche, and the knowledge that we have about the niche is described, along with its limitations. We discuss disease caused by mutations or environmental changes in the haematopoietic system, in particular

leukaemia. We give an outline of the ways in which leukaemia and leukaemia stem cells (LSCs) can interact with HSCs and the niche.

2.2 Haematopoietic stem cells

Driven by HSCs, haematopoiesis – or blood formation – occurs in all vertebrates and accounts for the production and renewal of all types of blood cell throughout the lifetime of an animal. In development, this process begins in the yolk sac, the aorta-gonad-mesonephros region and the placenta where haematopoietic progenitor cells are first found to be present (Lanza, 2009). They subsequently move through the fetal liver and, just before birth, migrate to the bone marrow (Mikkola & Orkin, 2006). The bone marrow remains thereafter the dominant site of haematopoiesis throughout the lifetime of the individual.

HSCs are responsible for the production of blood cells and can be pictured as the root of a tree-like structure where branching points mark lineage choices: early on the lymphoid and myeloid lineages split, followed by subsequent splits that further restrict each lineage until each branch produces one terminally differentiated (and specialised) cell type, such as an erythrocyte, B cell, T cell or neutrophil. A representation of the haematopoietic tree is shown in Figure 2.1.

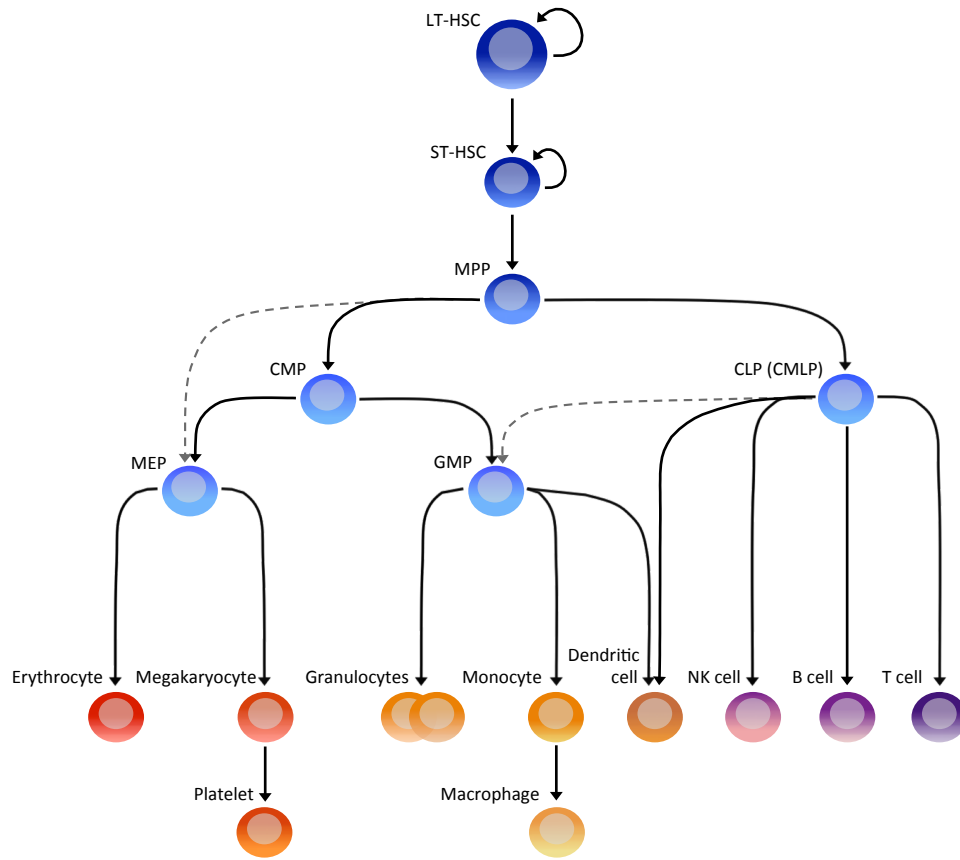


Figure 2.1: *Haematopoiesis: all blood cells are produced from HSCs in a hierarchical process.* Representation of haematopoiesis in the bone marrow showing some of the possible different cell populations (commonly identified using cell surface markers). Through a number of differentiation and branching steps (moving from top to bottom), HSCs lose their *stemness* and become specialised. Stem/progenitor cell species are shown in blue and the different specialised blood cells are shown in other colours. Self-renewal arrows for the HSC populations denote the long term (indefinite) repopulation potential of these cells. Dashed arrows denote putative transitions that could occur instead of (or as well as) the accepted transitions shown in bold. Abbreviations: LT, long term; ST, short term; MPP, multipotent progenitor; CMP, common myeloid progenitor; CLP (CMLP), common lymphoid progenitor (common myeloid-lymphoid progenitor); MEP, megakaryocyte-erythroid progenitor; GMP, granulocyte-macrophage progenitor; NK, natural killer. The granulocytes are a sub-population of specialised cells that include mast cells, basophils, eosinophils and neutrophils.

2.2.1 Haematopoietic stem cell specification

The property of *stemness* defines an HSC. The three main qualities that are encompassed by this term are the ability

- to generate multiple cell types through asymmetric division,
- to regenerate tissue when transplanted and
- to quiesce, enabling a long life span (Martinez-Agosto et al., 2007).

However, qualifying the stemness of a cell is not straightforward (Mikkers & Frisén, 2005; Potten & Loeffler, 1990). The long-term follow-up required to prove that the cell can regenerate tissue successfully, coupled with the heterogeneity of most (if not all) cell populations means that we cannot ever say with certainty that a particular cell is an HSC. Location can also influence and conflate HSC determination (Grassinger et al., 2010). As such, it is impossible to observe an HSC directly; the best we can do is to define as HSCs the population we obtain by sorting cells using the best available cell surface markers, and keep in mind that probably only a subpopulation of these are actually HSCs. In the work that follows, we adopt these standards: we refer to the haematopoietic cell population that is identified as being most primitive as HSCs. It is worth keeping in mind, however, that HSCs have never truly been isolated. The elusive nature of these stem cells means that analyses can be difficult and at times rely on inaccurate assumptions.

Haematopoietic cells that are Lineage⁻ lack expression of cell markers which define various lineage-committed cell types. The combination of Lineage⁻

plus positive expression of cell-surface markers Sca-1 and c-Kit defines a multipotent haematopoietic cell population (Okada et al., 1992; Spangrude et al., 1988). Within this population, to distinguish between multipotent progenitor cells, short-term HSCs and long-term HSCs, cluster of differentiation (CD) markers including CD34, CD48 and CD150 can be used (Wang & Wagers, 2011).

2.2.2 Haematopoietic stem cell regulation

There are a number of signalling pathways that are known to play important roles in the specification and maintenance of HSC function and location. Two key regulatory pathways – those of Wnt and Notch signalling – are briefly described here. These molecular mechanisms that underpin cellular function are introduced to help to explain the context of cells within the HSC niche.

The wingless type protein family (Wnt) pathways (canonical and non-canonical) exert an influence over haematopoiesis (Austin et al., 1997; Fleming et al., 2008). The canonical Wnt pathway is mediated by the Wnt receptor Frizzled on the cell membrane and ultimately affects the expression and location of β -catenin, which can translocate to the nucleus and lead to the transcription of target genes, including some that perform functions targeting self renewal. Although there is some controversy over the exact role of Wnt in HSC self renewal regulation, both the level and the duration of the signal seem to be important (Malhotra & Kincade, 2009; Staal & Luis, 2010). Non-canonical Wnt signalling has been shown to play a role in

maintaining quiescence in HSCs (Sugimura et al., 2012).

The Notch signalling pathway is crucially involved in maintaining HSC self renewal potential (Duncan et al., 2005; Karanu et al., 2000; Morrison et al., 1997). Notch also influences location: osteoblasts that express the Notch ligand Jagged-1 regulate HSCs since stimulated osteoblast production increases the number of HSCs found in the niche (Calvi et al., 2003; Zhang et al., 2003). There also remain important unanswered questions regarding this pathway including whether Notch signalling is required under homeostatic conditions and to what extent the crosstalk between Notch and other pathways influences haematopoiesis (Duncan et al., 2005; Renstrom et al., 2010).

These pathways appear to be amongst the most important with regard to HSCs, however the topic is far from exhausted. See Blank et al. (2008) for review and (for example) (Smith-Berdan et al., 2011; Yilmaz et al., 2006) for further discussion of pathways involved in the regulation of HSCs.

2.3 The haematopoietic stem cell niche

Understanding the niche in which HSCs reside has been shown to be crucial in order to understand HSCs. Here we describe what is known about HSC-niche interactions and discuss open questions and challenges.

The niche of an organism is defined as the space it occupies within its habitat and the resources that it requires to function. This definition is applicable to cellular species in addition to complete organisms. There

exists a mammalian niche – found within the bone marrow – that is occupied by HSCs. First introduced by Schofield (1978), our understanding of the HSC niche has grown dramatically in recent years.

In 2003, it was shown that osteoblasts regulate HSCs in the niche (Calvi et al., 2003; Zhang et al., 2003). Osteoblasts reside in the endosteum – the area between bone and bone marrow, thus suggesting that the endosteum could be an HSC niche location. Using cell surface markers (Kiel et al., 2005), and then using two-photon and confocal microscopy (Lo Celso et al., 2009a), it was demonstrated that HSCs are indeed found near to vessels and/or osteoblasts and the endosteal surface. The question of whether HSCs are located at or near the endosteum remains under debate, with evidence from Kiel et al. (2009) suggesting that HSCs reside near but not at the endosteum. An important role in defining HSC niches is also played by perivascular cells (residing at the periphery of blood vessels) (Kiel & Morrison, 2008; Kiel et al., 2005). Ding et al. (2012) suggest that HSCs occupy a perivascular niche with support from endothelial cells but without support from osteoblasts, given their work where HSC frequency and function were studied following conditional depletion of the stem cell factor ligand from different cell types.

At times, it seems that HSC scientists are in danger of becoming as entrenched in their disparate beliefs about the true location of the niche as the stem cells are themselves in their niches. Whether there are two distinct niches: endosteal and perivascular, or whether there is only one location that we are yet to isolate is an open question (Morrison & Spradling, 2008).

Either way, self renewal capacity and other HSC attributes will be affected by which niche (or part of the niche) an HSC is found in (Haylock et al., 2007). If it is the case that both endosteal and perivascular niches coexist, it has been proposed that dormant HSCs might reside in an endosteal niche, and active (cycling) HSCs in a perivascular niche (Wilson et al., 2007).

In addition to the endosteal vs. perivascular niche question, the influence of other cell types on HSC niche maintenance is also important. The progenitor cells that are progeny of HSCs can directly exert an influence on the niche (Hsu & Fuchs, 2012). Macrophages play a role in niche maintenance, as do mesenchymal stem cells (directly, in addition to the role they play by producing osteoblasts) (Ehninger & Trumpp, 2011; Méndez-Ferrer et al., 2010), as do CXCL12-abundant reticular cells (Sugiyama et al., 2006). The picture we have is thus complicated and far from complete. The complexity increases further when we consider functional aspects of the niche: how is HSC dormancy controlled? (Trumpp et al., 2010) and is the niche solely a place of quiescence or does it also allow for HSC self renewal? (Wilson & Trumpp, 2006)

In ecology, niches are commonly studied in order to better understand species' behaviour both in isolation and when interacting with other species. Here the definition of the niche changes slightly from its biological use, shifting towards a functional definition that is more broad than the biological one.

In the modelling work that follows, because of these unanswered questions regarding how the niche is defined (both competing biological descriptions

and biological versus ecological descriptions), choice of the niche definition will depend on the goal of the model and will need to be justified, in concordance with choosing its functional extent. It could be defined as the area in which an HSC is in direct contact with a niche-supporting cell, or the extent of influence of all extrinsic niche signals on HSCs, or (most widely) as the volume of the whole inner bone marrow cavity. In the case of each model introduced, the assumptions surrounded niche definition will be explained and justified.

2.4 Leukaemia

Leukaemias are haematological malignancies, that is, cancers of the blood system. Leukaemias can be characterised based on the haematopoietic branch they primarily affect (myeloid/lymphoid) and the speed of their progression (acute/chronic). In all cases, normal haematopoiesis is disrupted by an abnormal expansion of some haematopoietic cell type. Leukaemias present a serious health concern, particularly among older people: 1 in 72 people will be diagnosed with leukaemia during their lifetime and the age-adjusted death rate is 7.1 per 100,000 people per year [based on U.S. mortality rates from 2000 – 2010] (Howlader et al., 2013).

In Chapter 6 of this thesis, treatment of a particular leukaemia – chronic myeloid leukaemia (CML) – is studied. CML is characterised by a chromosomal translocation: $t(9;22)(q34;q11)$ in up to 95% of patients with the disease (Epstein et al., 1999). This creates a hybrid gene called BCR-ABL that is strongly associated with the disease, and molecular therapy via ty-

rosine kinase inhibitors that selectively target the BCR-ABL protein has proven to be very successful and indeed pioneered targeted cancer therapy (Muller et al., 2003). However even here, disease relapse can occur and we still do not understand at the fundamental level what balance of cellular processes drives disease progression. Other associations can be made between leukaemias and genetic abnormalities: for example, acute leukaemias with mixed-lineage leukaemia (MLL) fusion genes such as MLL-ENL, MLL-AF4 and MLL-AF9 (Cozzio et al., 2003; Domer et al., 1993; Krivtsov et al., 2006; Rubnitz et al., 1994), and acute lymphoblastic leukaemia with hypodiploidy (Harrison et al., 2004).

The “cell of origin” question remains a challenging one in the study of leukaemias: from what healthy haematopoietic cell population did the leukaemia arise? This question is not the focus of this work and it is raised here only so that two points can be made. For the purposes of our study, this question of cancer origin is tangential. We focus on the progression of disease from some (non-zero) initial malignant population. The moment at which the cell of origin question becomes relevant to us is when discussing which haematopoietic species leukaemia cells are competing against – stem, progenitor or differentiated. Also, note that the term *cancer stem cells* (see below) relates to the *stemness* properties of the cell population and not to its population of origin (Dick, 2008).

Regarding the driving mechanism of leukaemias, and of cancers more generally, there are two predominant theories. The first proposed mechanism is that cancer is driven by clonal evolution (Nowell, 1976), whereby all can-

cerous cells have tumorigenic potential and any successful cancer treatment must kill most – if not all – of the affected cells. The alternative proposal is that a tumour is driven by a subpopulation of cells that have the capacity to metastasise and form new tumours and that only this subpopulation needs to be targeted and eradicated in order to cure the disease. This subpopulation ought to exhibit stem-like properties and are referred to as *cancer stem cells*. Evidence for cancer stem cells was found about 15 years ago, by Bonnet & Dick (1997), and others.

According to the cancer stem cell theory, the tumour would have a hierarchical organisation similar in structure to the healthy haematopoietic system (as well as other adult-stem cell derived tissues). Some of the early evidence for such tumour organisation came from studying acute myeloid leukaemia (Bonnet & Dick, 1997; Lapidot et al., 1994). More recent work has studied whether stem cells can also be found in other types of leukaemia (Bernt & Armstrong, 2009; Jamieson et al., 2004; Krivtsov et al., 2006). Despite growing support for the cancer stem cell theory, evidence is not conclusive and many questions remain (Dick, 2008; Medema, 2013; Passegué et al., 2003; Wang & Dick, 2005). Much of the evidence used in support of the theory comes from transplantation experiments in mice, but it has not been shown that transplanted cancer ‘stem’ cells can indeed recapitulate in full the heterogeneity of the primary tumour (Shackleton et al., 2009). In the work that follows, the cancer stem cell theory is adopted. Even if the theory does not hold for all cancers, its application to leukaemias is well supported (Reya et al., 2001), and it provides a basis upon which to study the inter-

actions between healthy and cancerous haematopoietic cells. The reality of cancer progression is probably a mixture of clonal and stem cell-driven evolution.

By adopting the cancer stem cell theory, we consider a leukaemia as being composed of two subpopulations: leukaemia stem cells (LSCs) and non-stem (differentiated) leukaemia cells. It is assumed that LSCs share certain properties with their healthy counterparts – haematopoietic stem or progenitor cells. At least, they share the ability to self renew (indefinitely) and differentiate into a mature – non-proliferative or non-tumorigenic – cell population. LSCs and HSCs may also respond to many of the same signals, although given the often aggressive nature of leukaemia growth, it seems probable that at some stage LSCs no longer respond correctly to quiescence/cell death signals.

It is commonly assumed that LSCs will be rare within a tumour and make up only a small fraction of cancer cells (Lane et al. (2009) suggest that “as few as one in a million” acute myeloid leukaemia cells have leukaemia initiating potential). However, it has been shown both mathematically (Johnston et al., 2010) and biologically (Quintana et al., 2008) that LSCs could comprise any fraction of a tumour. In the subsequent modelling framework used, this is taken into account by not placing restrictions on the LSC population size relative to other leukaemia or healthy haematopoietic populations.

2.5 Competition within the niche

In Section 2.3 above, the concept of the HSC niche was introduced, followed by an introduction of leukaemia and cancer stem cell theory in Section 2.4. Here we bring these topics together and discuss competition with regards to haematopoiesis.

We study leukaemia as a stem cell disease, driven by LSCs, occurring within a stem cell niche inside the bone marrow. The assumption that LSCs occupy the same niche (or similar niches) to HSCs is central to this work. It has been shown *in vivo* that leukaemia progression disrupts different haematopoietic cell species in different ways within the niche (Hu et al., 2009; Miraki-Moud et al., 2013). There is also evidence to suggest that leukaemia might compete with healthy haematopoietic species by directly changing the microenvironment to create abnormal niches (Colmone et al., 2008). Schepers et al. (2013) demonstrate that LSCs can drastically alter the niche by affecting mesenchymal stem cell and, in turn, osteoblast function through a number of important signalling pathways. As such, bone marrow niches may need to be redefined once leukaemia is established.

What remains unknown are the similarities and differences in the ways that HSCs and LSCs respond to the niche, as well as the extent to which competition is indirect i.e. leukaemia exerts effects mainly on the niche. In addition, it might be that leukaemic competition exists between LSCs and haematopoietic progenitor cells (HPCs), leaving the parent HSC population not directly affected. This idea is supported by evidence that some

leukaemias result from HPC populations (Goardon et al., 2011; Passegué et al., 2003).

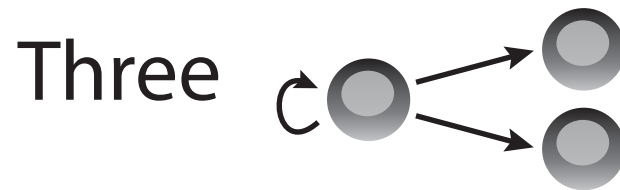
Questions regarding the nature of competition are relevant not just for leukaemias but also for solid tumours. In other tissues including the brain, colon and epidermis it has been shown that stem cells can be responsible for driving the disease (Fuchs et al., 1999; Polakis, 2000; Reya et al., 2001; Wechsler-Reya & Scott, 2001). In cases where cancers may be clonal (or non-hierchically organised), there is also competition between the malignant cells and the healthy cells surrounding them. Given the broad scope of the modelling undertaken in this work, the conclusions that we draw regarding competition and leukaemia can, in places, inform our understanding of competition and cancer more generally. There are a number of unanswered (and crucial) questions in cancer research that relate directly to competition, such as the poorly understood mechanisms of metastases and how the competition between healthy and malignant cell species within the body affects growth and dormancy (Uhr & Pantel, 2011).

In the work that follows, these questions of competition will be addressed through the development and analysis of models that each target competition in different ways. Many of the concepts relating to this competition between cellular species are taken from ecology; this will be described more fully in the next chapter. Here, it is sufficient to say that the processes governing both HSCs and LSCs are complex, and yet to be fully elucidated. Once we combine the two we have an even richer picture that will require investigation at multiple levels – from molecular to cellular – to successfully

identify all of their key components.

2.6 Conclusion

Here we have described the biological foundations of this thesis and some of the most important unanswered questions have been highlighted. At the heart of this work lies the HSC. An introduction to HSCs has been given, along with a description of the key regulatory processes governing them in the microenvironment in which they reside. Additionally, we have introduced leukaemias as haematological malignancies and outlined the ways that they can interact with the normal (healthy) haematopoietic system. Given this biological overview, we are now ready to delve into the mathematics and theoretical ecology that allow us to describe these biological processes using models. These models will in turn help us to shed light on the mechanisms underpinning these biological processes.



Background: theoretical ecology and mathematical modelling

3.1 Overview

Mathematical modelling already has a rich history of application to biology, despite the perceived dichotomy between the mathematical and biological sciences. Using examples from ecology, we describe how mathematics has advanced our biological knowledge over the last hundred years. Ecological examples are particularly relevant here because we base our modelling of the haematopoietic system on such foundations: we find that ecological concepts provide us with a strong framework upon which to model population dynamics between cellular species.

Subsequently, we present an overview of modelling frameworks appropriate for studying HSCs and leukaemia. The statistical and inferential tools that we employ are not presented here, but at the appropriate points in later chapters. Here, to provide background information about the literature regarding mathematical modelling of HSCs, we first present general modelling strategies applicable to problems of this type, and then discuss specific examples of recent models of HSCs and cancer. This allows us to summarise what has been done so far and what remains to be addressed.

3.2 Lotka and Volterra: the beginnings of theoretical ecology

One of the best-known models describing a dynamical system is given by the Lotka-Volterra equations, and owes its name to the work of physicist Alfred Lotka and mathematician Vito Volterra (Lotka, 1920; Volterra, 1926). These equations have been very widely used, and motivate the models that will be developed later.

The system describes two species, X and Y , interacting in the following way:

$$\begin{aligned}\dot{x} &= x(\alpha - \gamma y) \\ \dot{y} &= y(\gamma x - \beta)\end{aligned}\tag{3.1}$$

where x and y denote the abundances of species X and Y , respectively. In the absence of Y , X will grow exponentially at rate α ; in the absence of

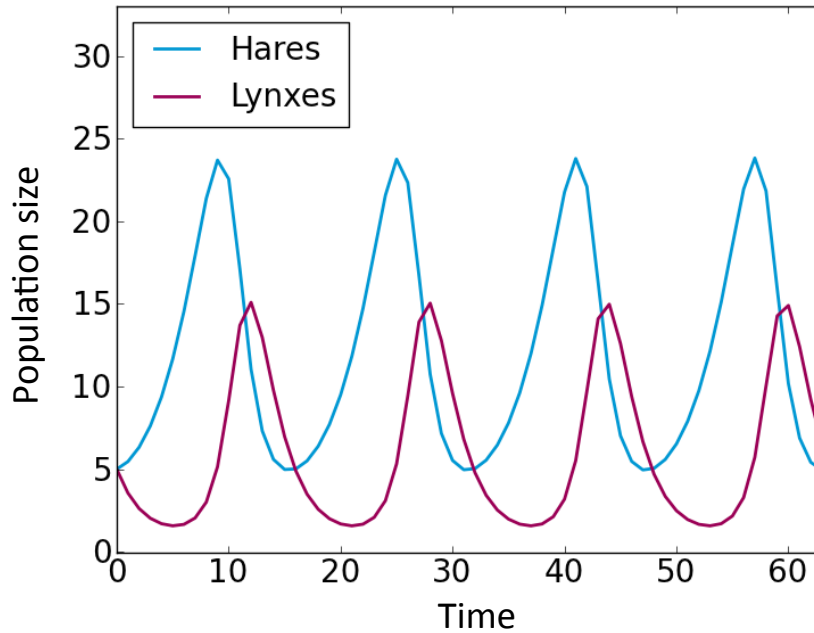


Figure 3.1: *The basic Lotka-Volterra equations model two species interacting through predation.* Simulation of the Lotka-Volterra equations as given by Equation 3.1 with parameters $\alpha = 0.3$, $\beta = 0.6$, $\gamma = 0.05$. Both species exhibit oscillations at steady state over some range of values. Time is given in arbitrary units.

X , Y dies (decays) at rate β . When both species are present, Y consumes X at rate γxy : proportional to parameter γ and the interaction between X and Y . This system is interesting dynamically because it permits stable oscillations between X and Y .

Lotka (1920) first proposed these equations and studied them in relation to the interaction between plants and herbivores. Volterra, unaware of Lotka's work, independently devised the same model whilst studying predatory fishes. Originally published in Italian in 1925, it then appeared in *Nature* the following year (Volterra, 1926).

Hares and lynxes are one of the commonly used examples of species interacting in a Lotka-Volterra type of system. Hares (the prey) would be species X and lynxes (the predator) species Y in the equations above. As a simple example, in Figure 3.1, we plot a simulation of Equation 3.1 for hares and lynxes at a given set of parameter values. It can be seen that both species exhibit stable oscillations; the stationary state for these parameters is a limit cycle. From Figure 3.1, we see that the behaviour of lynxes follows that of the hares: initially lynx growth follows hare growth until the population of hares reaches a maximum and begins to decline. Subsequently, the lynx population peaks and then decreases just behind the hare population until reaching a minimum value, and the cycle repeats.

There are three important advantages of such an approach that we ought to take away from the Lotka-Volterra example. The first is that by specifying our biological assumptions precisely using mathematics, we are able to test them quantitatively in a way that was not possible before. Rather than suggesting that “growth of a hare population is affected by the rate at which they are eaten by lynxes” we can quantify this statement by creating a mathematical model and then test it by simulating the model and studying its ability to describe biological data. Second, we see that this very simple model with only three parameters already describes quite well the interactions between two species. Additional terms are needed to provide a good fit to most real predator-prey datasets, but this most basic model is a good first step. That (vastly) complicated biological systems can be described successfully by (very) simple models is perhaps the most impor-

tant lesson here. Finally, by establishing the validity of this model, we are able to use it to make predictions and thus prompt new biological hypotheses. For example, we could change the parameter γ and study its effect on the dynamics of the system, mimicking a change in the behaviour of hares that allows them to better (or less well) evade capture by lynxes. Each of these points demonstrates the potential of studying biological systems with mathematical models.

3.3 A century of ecological modelling

Almost sixty years after the work of Lotka and Volterra began, *The Theory of Island Biogeography* was introduced in a book of the same title by MacArthur & Wilson (1967). Here the authors predicted how ecological dynamics would determine species' richness in island habitats; these can be islands of land surrounded by water, or less traditional 'islands' such as mountain peaks, trees or areas of a lake divided according to their temperatures (MacArthur & Wilson, 1967; Mason et al., 2008). Alternatively, as we consider in this work, 'islands' or *niches* could be areas within the bone marrow distinguished spatially or functionally. The theory of island biogeography took ecology in a new direction. MacArthur & Wilson (1967) predicted how many species an island could support using a mathematical formulation (they consider newly created theoretical islands), paving the way for future decades of niche modelling.

In 1972 and 1973 Robert May published a number of papers in which he made considerable advances in theoretical ecology (May, 1972, 1973a,b; May

& MacArthur, 1972). He proved a theorem that told us that as systems (of randomly interacting species) grow, their stability in general will decrease and tend towards zeros for large enough (or very highly connected) systems (May, 1972). This went against scientific opinion at the time, which stated that stability ought to increase with complexity (Elton, 1958; MacArthur, 1955). Together with Robert MacArthur, May developed the concept of niche overlap, showing how variability between species can impact their ability to occupy the same niche space (May & MacArthur, 1972). One year later, May published *Stability and Complexity in Model Ecosystems* in which these ideas are explained and consolidated (May, 1973b).

A related yet distinct branch of science that was developed alongside theoretical ecology is evolutionary game theory. Pioneered by John Maynard Smith and others, it provided new tools for studying dynamics within populations of species: based on a set of players and for each player a set of strategies, it can be found whether there are evolutionary stable strategies. These are ones for which, if adopted by the whole (large $\rightarrow$ infinite) population, no mutant can enter and displace the existing species (Maynard Smith, 1982).

There are interesting parallels between the concept of stability as applied to community interactions in May's work, and the evolutionary stable strategies of Maynard Smith. They both define conditions under which a system, if perturbed, will return to its initial state (state before perturbation). Evolutionary game theory provides a different framework to consider competition between species, and has been applied to the study of cancer growth

(Tomlinson & Bodmer, 1997). Such modelling approaches are not presented in this thesis, but do offer interesting avenues for future work.

Recently, there have been extensions to the stability theorem of May (1972) that move closer towards systems that we might expect to find in the real world. Allesina & Tang (2012) extended May’s theorem for structured random communities of species. That is, interactions between species are still chosen randomly, but structure is imposed by allowing only certain interaction types, for example predator-prey-type or mutualistic interactions. This is very interesting work, that moves us closer to ‘real’ biological systems, however there remains more to be done. The results presented by Allesina & Tang (2012) still rely on random matrices, thus as models are still a long way abstracted from the real world. This could be remedied by imposing further biological structure onto such interaction matrices.

This section has outlined very briefly the important steps in ecological modelling that have been made over the last hundred years. Despite the time that has passed, questions regarding the stability of ecosystems remain very topical today, and the debate rumbles on (Evans et al., 2013; Ives & Carpenter, 2007).

3.4 General strategies for modelling the haematopoietic system

Many different quantities can be modelled in biology: gene/protein abundances; gene/protein interactions in simple (e.g. Boolean) networks; abun-

dances of cellular species; the effects of cytokines, chemicals, drugs or other interventions on a system or tissue-level effects, to name a few. Here we are particularly interested in abundances of cellular species. We forsake the molecular details of the system at hand in order to progress with understanding what is happening at the cellular level.

Modelling strategies that are applicable for modelling cellular abundances can be classified as either deterministic or stochastic; we will discuss each approach in turn. In this section, strategies are discussed conceptually, together with a critical review. In the following section a number of concrete examples are given from the recent literature.

Deterministic approaches describe systems for which, given full details of the model (parameter values and initial conditions), the time evolution of the system can be known exactly. This means that if a system is restarted multiple times from the same initial state it will always return to the same future states. Ordinary differential equations (ODEs) and partial differential equations (PDEs) are two examples (Murray, 2008). PDEs are more flexible than ODEs, but the simulation and analysis of models based on PDEs quickly becomes computationally very expensive and often unfeasible. An ODE could be specified by Equation 3.2, where the dynamics of variable x are controlled by a function F :

$$\frac{dx(t)}{dt} = F(x(t)). \quad (3.2)$$

Equation 3.3 below gives an example of a PDE, where we now have a variable $u(x, t)$ that is a function of both x and t and whose dynamics are

controlled by a function G . For example, $G = c\frac{\partial^2 u}{\partial x^2}$ gives the diffusion equation, where c is constant.

$$\frac{\partial u(x, t)}{\partial t} = G(u(x, t)). \quad (3.3)$$

Deterministic methods provide good summaries of population-level behaviour: they are appropriate whenever population sizes are large enough. When populations become very small, the effects of random fluctuations are important and stochastic methods are required.

Stochastic approaches, in contrast to deterministic ones, describe systems for which temporal evolution has unpredictable elements due to the inclusion of randomness somewhere in the system. These are popular for modelling biological systems, since in biology examples of randomness and heterogeneity abound. Such approaches have yielded many important results, and ought to be used whenever population sizes are small enough that fluctuations cannot be ignored. In most cases, population averages will be recovered from a stochastic model when the abundances grow large enough. In some cases, where little is known about the system, it is best in the first instance to investigate with deterministic models to see whether they can already capture the behaviour that one is attempting to describe. If this is the case, then the additional complexity of a stochastic model is probably unnecessary.

A stochastic modelling formulation similar to use of ODEs is provided by stochastic differential equations (SDEs) (Gardiner, 2009). One way to create an SDE model is to take a set of ODEs and to each add a noise term,

often Gaussian, i.e. add a noise term η where η is a random variable distributed according to $\eta \sim N(0, \sigma^2)$. Such an approach can capture population dynamics that exhibit stable fluctuations around some mean value.

Alternatively, one can begin directly from the master equation. This describes the probabilistic evolution of a state (or set of states) over time. For example, if the states are discrete then the master equation can be written as

$$\frac{dP(t)}{dt} = P(t)Q \tag{3.4}$$

where $P(t)$ is a vector of states given at time t and Q defines a matrix of transition probabilities between the states of the system (Wilkinson, 2011). An algorithm that was introduced by Gillespie (1977) provides an exact method for simulating such a system, although it can be computationally very expensive for large numbers of molecules or over large timescales. In these circumstances, approximate methods exist to reduce simulation times such as the τ -leaping algorithm, an extension of the original Gillespie algorithm presented by Gillespie (2001). These methods account for the exact numbers of molecular/cellular species, rather than their concentrations that are given by, for example, ODE models.

Agent-based modelling provides a different strategy, and is usually attempted stochastically, although it can also be performed in a deterministic framework (Gilbert, 2008; Jost, 2005). Here, individual agents act according to a set of rules. For example, from within a population of HSCs, a cell chosen at random could either self-renew or differentiate, with a probability

associated with each process. Such a framework considers cells in a very different way than do differential equation methods, by treating each as an autonomous agent and thus studying the population dynamics in a “bottom up” manner. Different treatments of temporal evolution are possible: cells within the system can be updated synchronously or asynchronously (and if the update is asynchronous, it can also be with or without replacement). Agent-based approaches may appeal to our intuition – and the evidence – on HSC behaviour: each cell can act differently and follow distinct paths that can be seen as the result of a chain of decisions. The main limitation of this approach is that certain analyses are computationally expensive or infeasible due to the cost of treating each cell individually. Modelling feedback also presents a challenge because the state of the focal cell is dependent on the current states of other cells.

Cellular automata are a subset of agent-based models that impose spatial structure on the system by constraining the agents of the system on to a grid, in two or three dimensions (von Neumann, 1966; Wolfram, 1983). The agents are then updated according to a set of (deterministic or – more commonly – stochastic) rules. Each grid point can be occupied by at most one cell of a certain type (the model can contain multiple cell types). These rules are functionally similar to the rules a general agent-based model will follow, but with spatial criteria. For example, in a simple cellular automaton of one cell type where the eight nearest neighbours of a cell on a two dimensional grid influence its behaviour, a model could obey the following rule:

1. if surrounded by less than four cells, divide, creating a new cell at a free neighbouring grid point,
2. if surrounded by more than six cells, die
3. else, do nothing.

Cellular automata exhibit spatial relationships between different cell types and offer the advantage of being easy to interpret biologically. The main challenges associated with these models are that the update rules may not translate clearly into biological hypotheses, i.e. it is not clear how we should interpret the meaning of rules (1) and (2) above. Furthermore, simulation of cellular automata can be very expensive, as for other agent-based models. Fitting such models to data is at the limits of what is currently feasible since, despite great advances in cellular imaging technology, obtaining sufficient resolution of cell dynamics to fit to a model is rare. This is beginning to change however, and performing parameter inference on spatial models is an exciting future prospect.

Moving away from agent-based approaches, branching processes present yet another way to consider the dynamics of cell populations (Haccou et al., 2007). Here, a population evolves according to how many offspring (of one or more types) each individual produces at each time point (time may be discrete or continuous). In a simple case, the number of offspring produced by all individuals follows the same probability distribution, however this assumption can be relaxed allowing for different types of individual to produce offspring according to different distributions. Branching processes are

attractive for their easy interpretation and, for simple models, obtaining analytical results is possible. The main limitations of branching process models is that they quickly become intractable once different fates of offspring are considered, or if the progeny are drawn from different probability distributions. Additionally, the only two equilibria that they offer are extinction or exponential growth. Their lack of ability to reach non-zero steady states limits their biological plausibility, since although the ultimate fate of any individual (cell or person) is death, in modelling haematopoiesis we wish to capture homeostatic mechanisms, not just the eventual extinction of the system.

This overview of modelling approaches commonly used in systems biology, and that might be appropriate for problems relating to the haematopoietic system, is of course not exhaustive, but introduces commonly used methods that have led to recent advances in understanding HSCs. In the next section, specific examples of each approach are given, with critical appraisal.

3.5 Modelling haematopoiesis: the state of the art

Recent models of haematopoiesis and leukaemia are presented here, with brief discussion of their methods, findings, strengths and limitations. The models studied in subsequent chapters will build upon and extend these, so we must begin with a clear understanding of the current methods used to model HSCs: deterministic, stochastic, and those incorporating a descrip-

tion of the dynamics of cancer.

3.5.1 Modelling healthy haematopoiesis

ODE modelling accounts for the largest proportion of approaches taken here. Manesso et al. (2013) present a model containing details for many species in the haematopoietic hierarchy, and use it to study the system in cases of homeostasis and under perturbation. This model is appealing for its coverage of species within the hierarchy (including many stem and progenitor species explicitly), and it was able to recapitulate observed steady state levels of species within the hierarchy following a perturbation such as haemorrhage.

Huang et al. (2007) and later Chickarmane et al. (2009) study a particular mechanism within the hierarchy: the branching point where erythroid and myeloid lineages separate. A small circuit of two genes (GATA1 and PU.1) seems to control this process and studying its dynamics using ODEs yields interesting results about branching cell fates and lineage choice. The authors propose a gene circuit design that permits bistability and find that this leads to a loss of cooperativity between the two genes and as such the creation of a “primed” state: an interesting proposed step in cell fate determination.

Qu & Ortoleva (2008) offer another angle on cell fates: they model stem cell differentiation in an attempt to understand how the asymmetric/symmetric decision might be made. They model cell fate as being controlled by biochemical reactions occurring within the cell. The model is used to search for

gene-regulatory circuit motifs that could promote stem cell differentiation.

A mixture of stochastic kinetics and ODEs (as the mean approximation of the stochastic case) are used by Mangel & Bonsall (2008) to model HSC processes of self renewal and differentiation. Mangel & Bonsall (2008) use results from life history theory to describe conditions for which HSCs can survive for longest in the niche. In later work, using similar strategies, the same authors argue that stem cell biology ought to be considered in the light of population biology and in terms of the demands an organism makes on the HSC system (Mangel & Bonsall, 2013).

Schittler et al. (2012) study HSC proliferation and differentiation using a structured population model and a mixture of ODEs and PDEs. This model is interesting because it can capture different proliferation properties in different cell subpopulations.

Delay differential equations have also been used to study haematopoiesis. Such equations exhibit interesting dynamics with periodic or chaotic solutions, depending on parameter values. Mackey & Glass (1977) present delay differential equations with which they suggest that they can recapitulate the behaviour of periodic fluctuations in different blood cell populations.

Xu & Xu (2010) use a difference equation approach to address questions of HSC differentiation. They study the stability of their model and from their analysis derive conditions that allow for oscillatory behaviour of HSCs.

In addition to the (mostly) deterministic methods used above to study the healthy haematopoietic system, a number of recent stochastic models de-

scribe different properties of healthy haematopoiesis. As introduced above, branching processes describe the dynamics of a population by the number of offspring each member has at each time point, and are thus naturally suited to the study of cell division. Alexandersson (2000) has studied branching process models of cell fate, aimed specifically at the cell cycle, but which could also be used to model HSC fate. Also using branching processes, Nordon et al. (2011) describe the inheritance of generation times from mother to daughter cells and show that it is possible to describe both single cell and multi-cell systems with such methods.

Marr et al. (2012) use branching processes and – in addition – present a stochastic toggle switch model to describe the granulocyte-monocyte branching point in the haematopoietic hierarchy. They find that the differentiation probability of a granulocyte-monocyte progenitor cell can be inferred using a branching process model but the branching process model cannot capture time dependence of this probability that the data suggest. They do manage to describe this time dependence with their stochastic toggle switch model, and in fitting their model they see a separation of timescales in the rates governing branching cell fates; one happens on a faster timescale than the other.

The GATA1 – PU.1 gene switch that was considered above from an ODE perspective has also been analysed stochastically, via construction of a Boolean network in order to elucidate different attractors of the system (Krumsiek et al., 2011). Using this method, the authors claim to describe successfully the four different cell fates considered according to gene expres-

sion profiles.

Villani et al. (2011), in an approach similar to Krumsiek et al. (2011), present an interesting model that describes cell differentiation as an emergent property of the underlying gene transcriptional network. They do this using random Boolean networks and link network attractors to different cell fates. They use this model to stress the role that noise has in driving differentiation.

The above work represents current approaches to modelling the haematopoietic system unperturbed by disease. Many insights have recently emerged from such approaches, for example regarding cell fate determination and mechanisms for homeostasis. We now proceed to examine models of haematopoiesis in the presence of disease.

3.5.2 Modelling malignant haematopoiesis and cancer

The work presented here describes recent attempts to use models to predict the behaviour of HSCs when leukaemia is present in the system, alongside a few models of non-haematological cancers, introduced here for the purposes of studying cancer stem cell dynamics.

Roeder & Loeffler (2002) develop an agent-based model to describe the stem cell dynamics of self renewal and differentiation and from this work and its extensions (Horn et al., 2013; Roeder et al., 2006, 2005) the model demonstrates the ability to reproduce expected stem cell behaviours both

in health and when under competition from malignant HSCs. In application to predicting system behaviour in the presence of chronic myeloid leukaemia (CML), the authors suggest that the efficacy of drug therapy can be increased by stimulating proliferation during therapy (Roeder et al., 2006). They also provide a predictor for the chance that an individual patient will relapse, thus offering the potential for patients who are predicted not to relapse to be taken off treatment after some time (Horn et al., 2013).

In 2005, Michor presented a different model for CML, using ODEs. This model was used to suggest that leukemia progresses by a biphasic decline (under treatment), with the first phase representing death of differentiated leukaemia cells and the second representing death of leukaemia progenitor cells. This model was one of the first descriptions of CML dynamics and provides interesting results, but it does not take into account interactions between healthy and CML lineages (Michor et al., 2005).

Extensions to this model added interactions between HSCs and LSCs, and these new versions were used to model CML dynamics and attempt to provide better treatment practice. The authors predict that adding a stem cell stimulating factor does not enhance treatment of CML (Foo et al., 2009) and suggest that it might not be safe to discontinue treatment even when disease reaches very low levels (Tang et al., 2012).

Colijn & Mackey (2005) present a model of periodic chronic myeloid leukaemia that incorporates the dynamics of both differentiated haematopoietic populations (erythrocytes, leukocytes and platelets) and HSCs to describe the dynamics of this periodic type of CML. They use dif-

ferential delay equations to model the system and find that certain parameters are particularly important to simulate disease dynamics that accurately reflect patient data.

Moore & Li (2004) present another dynamical model of CML that studies the disease from a different angle, examining the interactions between CML cells and T cells. They use ODEs to model the system and latin Hypercube sampling to cover the parameter space of interest, and find that parameters controlling CML growth and death rates are the most important in determining outcome of the system (level of CML).

Traulsen et al. (2012) look at how disease in the haematopoietic system can arise from a neutral mutation (rather than one that confers a selective advantage to its progeny) using a Moran process (a type of branching process used to describe genetic drift). They suggest that leukaemias that arise due to a neutral rather than selective process may be harder to treat.

La Porta et al. (2012) use another branching process model, here to study senescence in tumours. Their results suggest that inducing senescence in cancer cells is a poor treatment strategy for most tumours, and instead that effort should be focused on eradicating cancer stem cells – cells that are likely resistant to senescence.

Enderling et al. (2009) present a cellular automaton model that studies tumour growth. This approach is a good example of what can be achieved using cellular automata, it manages to link update rules to biological processes quite well and produces some fascinating results. Most important is

the finding that tumour suppression – via increasing the tumour cell death rate – can increase the long term growth rate of the tumour by offering free space to cancer stem cells in which they can proliferate. If this paradoxical result does indeed match the biology of tumours, it has strong implications for cancer therapy.

Alarcón et al. (2003) also use cellular automata to study tumours. They are particularly interested in angiogenesis and its impact on cancer growth. They demonstrate that their cellular automaton model can simulate the effects of blood flow and red blood cell heterogeneity. This model is interesting and somewhat more sophisticated than that of (Enderling et al., 2009), although (perhaps to compensate) it is simulated on a much smaller grid. The authors suggest that blood flow heterogeneity has a strong influence on tumour growth and can limit it considerably.

Daukste et al. (2009) model tumour dynamics from another perspective: placing cell cycle times at the centre of their analysis. They assume that two subpopulations exist within a tumour (stem and non stem cells) with two different characteristic cycling times. They model these populations with ODEs and compare the model to data that describe cell cycle dynamics of tumour cell lines in culture.

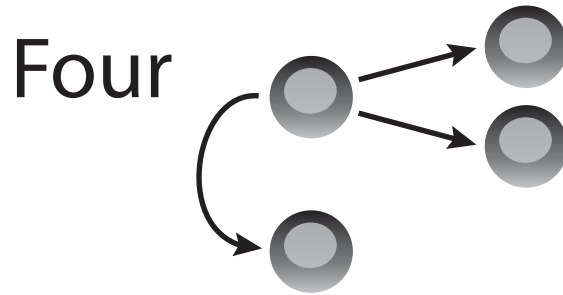
Zhu et al. (2011) also study tumour growth using ODEs under the assumption of the cancer stem cell model. They include a description of the stem cell niche in this model, and additionally model the impact of the epidermal growth factor signalling pathway on tumour dynamics. Their main findings propose important roles for niche size and over-expression of

epidermal growth factor receptor in tumour progression.

These examples from literature are chosen for interest and relevance and the list is of course far from being exhaustive. The above models have advanced our knowledge of the dynamics and effects of leukaemia, helping to elucidate the role of interactions between different types of healthy and cancer cells, proposing new hypotheses and suggesting new routes for experimentation. Despite such advances, many fundamental questions remain regarding disease incidence and – of greatest interest here – how interactions between haematopoietic cells, leukaemia and the niche affect disease progression.

3.6 Conclusion

In this chapter we discussed how the haematopoietic system has been treated mathematically, both in a historical context and through a number of current examples. There is a rich history of interplay between ecology and mathematical modelling, each benefitting the other. This history places current approaches in context, and introduces many of the concepts that remain pertinent today, especially that of competition between cellular species. We presented an overview of modelling approaches applicable to the haematopoietic system in health and disease, and went on to give a number of examples from the literature, covering both deterministic and stochastic approaches. This provides a platform from which we will develop new methods in an attempt to understand in greater depth the haematopoietic system.



A new framework for stability analysis and its application to stem cell niche models

4.1 Overview

Linear stability analysis is a useful tool for the analysis of nonlinear dynamical systems, and important results have arisen from its application to studying the behaviour of ecosystems. In this chapter we extend these methods and apply them to stem cell (eco)systems. We consider the stability properties of the fixed points of a system, not for a single parameter vector, but for many values, sampled over the parameter space. This gives us the probability that a fixed point is stable, over some range of param-

eters. We can use this framework also to investigate other properties of a model's fixed points, such as their 'basins of attraction' in parameter space. We apply these methods to a model from the literature to illustrate their use and we show from this example that dependencies and heterogeneity in biological systems dramatically change the stability properties of their fixed points. Subsequently, we focus our attention on models of stem cell niche dynamics and study the properties of their fixed points. We find that in most cases the fixed points are always stable when they are reachable from a certain point in parameter space. When models have multiple fixed points with biologically interesting properties (all species should be positive finite), we study what parameter values permit each and from this we can learn about the behaviour of the model. This probabilistic framework provides new methods for the analysis of the fixed points of a model and gives insight into how stability can be achieved, especially in the case of stem cell niche models.

4.2 Introduction

4.2.1 Motivation

Many mathematical models describing biological (and other) phenomena are constructed using ordinary differential equations (ODEs). The behaviour of an ODE system can be categorised into two classes: transient and steady state behaviour. The latter describes the system over large timescales ($\text{time} \rightarrow \infty$). For those systems that reach single valued (i.e.

not oscillating) steady states, we refer to values that system variables take at steady state as the fixed point of the system (Glendinning, 1994).

Given an ODE system that has one or more fixed points, it is interesting to ask: what happens to the system following a small perturbation in some direction away from the fixed point? There exists a well-established framework for such investigation known as linear stability analysis (Strogatz, 1994). According to this definition, a model's fixed point is called stable if, subsequent to perturbation from that fixed point, the model returns to the same place.

We are particularly interested in the application that this type of stability analysis has found in studying ecological systems. Here, it has been combined with concepts from random matrix theory to study the stability properties of ecosystems, as pioneered by Gardner & Ashby (1970) and May (1972). However, in order to be able to derive analytical results, these methods have only considered species whose interactions are drawn from random matrix distributions that dictate that entries be independent and identically distributed (*i.i.d.*). This situation is far away from what we expect to see in nature, where correlations and dependencies abound (Kokkoris et al., 2002). To remedy this we need to look at models of biological systems that account for such dependencies. We look at ODE examples of biological systems and use Monte Carlo methods to analyse their fixed point behaviour, since in these cases we cannot derive results analytically.

Our goal is to gain insight into the probabilistic stability properties of real-world systems, where previous analyses have only tackled highly idealised

i.i.d. versions of these. In the next section an overview of linear stability analysis is presented, with a description of its applications in ecological modelling.

4.2.2 Background on linear stability analysis

We begin with some definitions. A *system* is a set of interacting dynamic variables. Since here we are considering systems such as those found in an ecological setting, we refer to the variables as *species*. We consider (eco)systems whose dynamics can be described by ODEs, so we can write down the system dynamics in general form as

$$\dot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{x}(t)) \quad (4.1)$$

where $\mathbf{x}(t)$ is a vector of size n (n is the total number of species) that gives the abundance of each species at time t (we will use $\mathbf{x}(t) = \mathbf{x}$ i.e. time dependence implicit). The abundance could be represented by whole numbers of species or by their concentration, given by the fraction of a species occupying a certain space. $\mathbf{F}(\mathbf{x}(t)) = [F_1(\mathbf{x}(t)), F_2(\mathbf{x}(t)), \dots, F_n(\mathbf{x}(t))]$ are non-linear functions that define the rates of change of the species and include the model's parameter dependencies.

Since we are interested in the behaviour of the model close to steady state, we consider a Taylor expansion of $F(\mathbf{x}(t))$ about the point $\mathbf{x}_0$, where $\mathbf{x}_0$ defines a fixed point of the system. In doing so we aim to learn about how the model responds to perturbations from this steady state. Discarding

terms of order x^2 or higher from a Taylor expansion we have

$$\dot{\mathbf{x}} = F(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0) \frac{\partial F(\mathbf{x}_0)}{\partial \mathbf{x}}. \quad (4.2)$$

Since $\mathbf{x}_0$ is a fixed point of the system, $F(\mathbf{x}_0) = 0$. Substituting $\mathbf{y} = \mathbf{x} - \mathbf{x}_0$ we have the linearised system

$$\dot{\mathbf{y}} = \frac{\partial F(\mathbf{x}_0)}{\partial \mathbf{x}} \mathbf{y} \equiv A \mathbf{y} \quad (4.3)$$

where A is called the Jacobian matrix in a mathematical context or the community matrix in an ecological one (Levins, 1968). We can study A to determine properties of the system at hand. Specifically, we can find the eigenvalues of A (λ_i where $i \in (1, 2, \dots, n)$) by solving

$$\det(A - \lambda I) = 0 \quad (4.4)$$

for λ , where I is the $n \times n$ identity matrix. We can then use the λ to define stability criteria: if $\text{Re}\{\lambda_i\} < 0 \ \forall i$ then the system is *stable* about the point $\mathbf{x}_0$: if the system is perturbed slightly from state $\mathbf{x}_0$, it will return to the same state following some time delay. If this condition does not hold, the fixed point is unstable. See, for example, Strogatz (1994) for further discussion of linear stability analysis. This definition of stability is used in a broad variety of applications in dynamical systems; we are interested particularly in its ecological applications, by May (1973b) and others.

Forty two years ago, Robert May proved a theorem that demonstrated that as the complexity of a system increases, its stability will in general decrease and for large enough systems tends towards zero (May, 1972). Complexity here is quantified either by the total number of species in the system or by

the number of links between them. For a randomly connected system of n species, with connectance — the number of actual links over the number of all possible links — C and average link strength (or variance) α , then, for large n , a system will almost certainly be stable when

$$\alpha < \frac{1}{\sqrt{nC}} \quad (4.5)$$

(May, 1972). This result contradicted current opinion of the time which stated that in general as systems grow in complexity their stability should increase (Elton, 1958; MacArthur, 1955). The implications and uses of this theorem have been wide and varied; outside of ecology it has found application in fields including molecular biology and economics.

Given this theorem, a question arises, namely, how do large systems persist in nature? As their size and complexity increases, Equation 4.5 states that their probability of stability will eventually approach zero. To begin to answer this, note that Equation 4.5 gives a statement about the behaviour of systems where links between species are placed at random. Of course in the real world interactions are not completely random but guided by evolution. This imposition of structure upon a system can increase its stability: the relationship between structure and stability will be explored further in this chapter.

Results regarding the stability of biological systems by May, and recent extensions to his work by Allesina & Tang (2012), rely on interactions between species being *i.i.d.* because they appeal to a theorem known as the circular law, and its extensions (Girko, 1984; Tao et al., 2010). This

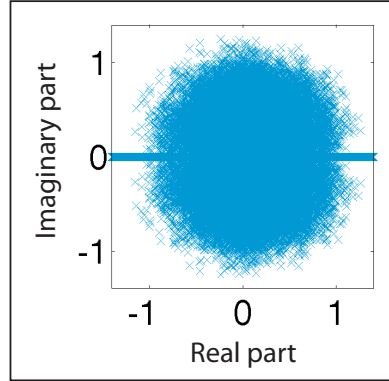


Figure 4.1: *Illustration of the circular law for a random matrix distribution.* Jacobian entries for 5×5 matrices are drawn from an empirical *i.i.d.* distribution. The eigenvalues for each Jacobian are calculated and can be plotted on the complex plane, as shown here.

states that, given a system of n species with interactions between them drawn according to an *i.i.d.* distribution, then in the limit $n \rightarrow \infty$, the eigenvalues of the system lie on a circle in the complex plane. In the case of finite n , the spectral distribution of eigenvalues still approaches a circular distribution, with some “bleeding” at the edges, especially along the real axis. As an example, shown in Figure 4.1 are the eigenvalues of 100,000 5×5 matrices with *i.i.d.* entries. In this chapter we use the circular law as a point of comparison with our numerical results.

4.2.3 Extending linear stability analysis

In the remainder of this chapter we will apply these concepts of stability and extend preexisting methods for the analysis of complex systems.

These extensions allow us to study the stability properties of ‘real’ biological systems where previously only *i.i.d.* systems have been studied. In the case of Allesina & Tang (2012), structure was imposed, for example in the form of predator-prey-type interactions or mutualistic interactions, but the sampling distributions used were still *i.i.d.*

We start with the assumption that stem cell species and their progeny act similarly to animal species in ecology: we are interested in learning about the behaviour of stem cell systems near to their fixed points. We seek to discover how the systems respond following perturbations.

The stem cell models used for analysis are introduced and described in the next section. Subsequently in section 4.4 we will present the methods that we have developed for generalised stability analysis. In section 4.5.1 we present results using this method for a general system — a version of the Lorenz model, and then, in sections 4.5.2 - 4.5.3 we analyse stem cell niche models via this framework. This analysis reveals how system structures, more than dependencies, maintain overall stability. By studying those systems that permit multiple biologically relevant steady states, we make predictions about cell lineage choices in a stem cell hierarchy. We end with discussion and some conclusions.

4.3 Model development

We develop 10 different ODE models that describe stem cell differentiation dynamics. They all share common features and differ in those parts that

we wish to compare and contrast. In each model we have a three-layer hierarchy, consisting of stem, progenitor and differentiated cells. Each model is built from a set of shared rules that describe cellular growth, death and differentiation. These rules also define how one species can regulate the growth of another through feedback.

The models differ in the number of species that exert and experience the effects of feedback. They also differ in the total number of species: models $S1 - S4$ each have three species and models $S5 - S10$ each have five. In $S5 - S10$, the extra two species result from a branching point that represents branching cell fates: when the stem cell divides it is able to differentiate into one of two progenitor species, each of which produces differentiated cell progeny. Such branching points occur frequently in the haematopoietic hierarchy, such as the myeloid/lymphoid branching (from a multipotent progenitor cell) or the monocyte/granulocyte branching from a granulocyte monocyte progenitor (see Figure 2.1).

All of the S -models share a common set of assumptions. Namely, that

- stem and progenitor cells can self renew, but differentiated cells cannot
- differentiation is irreversible (dedifferentiation is not allowed)
- stem and progenitor species cannot die or migrate out of the niche
- feedback enters the model through linear growth inhibition of one species on either itself or its parents/grandparents.

The models are parameterised by terms representing growth, differentiation and death rates. We are going to investigate the region of parameter space $p \in [0, 1]$ for each model parameter p . Parameters are assigned biological meaning, thus should never be negative, and we set the upper limit to be 1 so that each p is interpreted as a rate parameter. This limit can be set without loss of generality: different parameter ranges could be studied with a rescaling parameter that would not alter the results beyond a rescaling of time (here given in arbitrary units).

4.3.1 Models *S1-S4*: three-species models

Model *S1*: This model describes the dynamics of the following species in the haematopoietic hierarchy: HSCs (x_0), progenitor cells (x_1) and differentiated cells (x_2). Their dynamics are given by

$$\begin{aligned} \dot{x}_0 &= ax_0(1 - x_0 - x_1 - x_2) - bx_0 \\ \dot{x}_1 &= bx_0 + cx_1(1 - x_0 - x_1 - x_2) - ex_1 \\ \dot{x}_2 &= ex_1 - gx_2 \end{aligned} \tag{4.6}$$

where the parameters (a, b, c, e, g) characterise the phenotypes of the species and are described in Table 4.1. See Figure 4.2 for graphical representation of model *S1*. Here, populations x_0 and x_1 self-renew at rates a and c , respectively. They differentiate into daughter species at rates b and e , respectively, and x_2 dies/migrates out of the niche at rate g . There is also feedback onto the self renewal terms of x_0 and x_1 : these are affected by all species in the model.

Model S2: The only change made here compared with *S1* above is in the feedback: now x_2 does not inhibit stem and progenitor cell populations. The equations describing *S2* are

$$\begin{aligned} \dot{x}_0 &= ax_0(1 - x_0 - x_1) - bx_0 \\ \dot{x}_1 &= bx_0 + cx_1(1 - x_0 - x_1) - ex_1 \\ \dot{x}_2 &= ex_1 - gx_2 \end{aligned} \tag{4.7}$$

where the same parameter set characterises the model; the model is shown in 4.2.

Model S3: This model simplifies the three-species lineage further, removing another layer of feedback and leaving only saturating feedback from x_0 onto itself; $\dot{x}_1$ and $\dot{x}_2$ are now given by linear expressions. Model *S3* is governed by

$$\begin{aligned} \dot{x}_0 &= ax_0(1 - x_0) - bx_0 \\ \dot{x}_1 &= bx_0 - ex_1 \\ \dot{x}_2 &= ex_1 - gx_2 \end{aligned} \tag{4.8}$$

where the parameters (a, b, e, g) have the same meaning as above, described in Table 4.1.

Model S4: In the final three-species model we investigate the effects of keeping regulation of the stem cell pool but removing it from the progenitor

Symbol	Definition
a	Rate of self renewal of x_0
b	Rate of differentiation of $x_0 \rightarrow x_1$
c	Rate of self renewal of x_1
e	Rate of differentiation of $x_1 \rightarrow x_2$
g	Rate of death of x_2

Table 4.1: *Description of parameters that characterise models S1-S4.*

cell population. The equations governing $S4$ are

$$\dot{x}_0 = ax_0(1 - x_0 - x_1) - bx_0 \quad (4.9)$$

$$\dot{x}_1 = bx_0 + cx_1 - ex_1$$

$$\dot{x}_2 = ex_1 - gx_2$$

where the same parameter set characterises the model as for $S1$ and $S2$ and the model is shown in Figure 4.2.

By comparison of the fixed points of models $S1-S4$ we aim to discover how more or less feedback impacts upon their stability.

4.3.2 Models $S5-S10$: five-species models

In the second set of models of this family, we include two more species. HSCs are still denoted x_0 and we now have two progenitor cell populations, x_1 and x_2 , that differentiate into two differentiated cell populations, x_3 and x_4 , respectively.

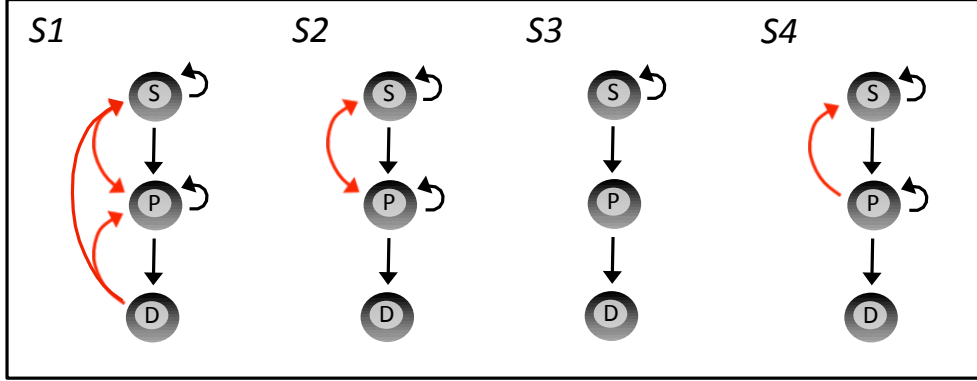


Figure 4.2: *Graphical depiction of models S1-S4.* The three cell species in the hierarchy are stem, progenitor and differentiated cells. Black arrows denote self renewal and differentiation processes and red arrows denote inhibitory feedback processes. *S2* and *S4* are subtly different: note the double arrowhead.

Model S5: This model contains feedback only onto x_0 from progenitor cells, and its equations are

$$\dot{x}_0 = ax_0(1 - x_0 - x_1 - x_2) - (c_1 + c_2)x_0 \quad (4.10)$$

$$\dot{x}_1 = c_1x_0 - e_1x_1$$

$$\dot{x}_2 = c_2x_0 - e_2x_2$$

$$\dot{x}_3 = e_1x_1 - g_1x_3$$

$$\dot{x}_4 = e_2x_2 - g_2x_4$$

where $(a, c_1, c_2, e_1, e_2, g_1, g_2)$ characterise the model. A graphical depiction of *S5* is given in Figure 4.5.2, alongside depictions of the models *S6* - *S10* that are introduced below. The parameters that specify this model are given in Table 4.2.

Model S6: This model extends *S5* by including a separate level of feedback: from differentiated cells onto progenitor cells, in addition to the feedback exhibited by *S5*. The equations for *S6* are given by

$$\dot{x}_0 = ax_0(1 - x_0 - x_1 - x_2) - (c_1 + c_2)x_0 \quad (4.11)$$

$$\dot{x}_1 = c_1x_0(1 - x_1 - x_3) - e_1x_1$$

$$\dot{x}_2 = c_2x_0(1 - x_2 - x_4) - e_2x_2$$

$$\dot{x}_3 = e_1x_1 - g_1x_3$$

$$\dot{x}_4 = e_2x_2 - g_2x_4$$

where the same parameter set as for *S5* describes the model.

Model S7: Here we remove the feedback onto stem cells and only consider feedback from differentiated cells onto their parent population. The equations that specify this model are

$$\dot{x}_0 = ax_0(1 - x_0) - (c_1 + c_2)x_0 \quad (4.12)$$

$$\dot{x}_1 = c_1x_0(1 - x_1 - x_3) - e_1x_1$$

$$\dot{x}_2 = c_2x_0(1 - x_2 - x_4) - e_2x_2$$

$$\dot{x}_3 = e_1x_1 - g_1x_3$$

$$\dot{x}_4 = e_2x_2 - g_2x_4$$

where the parameters are the same as before.

Model S8: This model combines all of the feedback effects we have considered in *S5-S7*. That is, we consider the joint effects of (differentiated $\rightarrow$ progenitor cell) feedback and (progenitor $\rightarrow$ stem cell) feedback. The

equations that specify $S8$ are

$$\dot{x}_0 = ax_0(1 - x_0 - x_1 - x_2 - x_3 - x_4) - (c_1 + c_2)x_0 \quad (4.13)$$

$$\dot{x}_1 = c_1x_0(1 - x_1 - x_3) - e_1x_1$$

$$\dot{x}_2 = c_2x_0(1 - x_2 - x_4) - e_2x_2$$

$$\dot{x}_3 = e_1x_1 - g_1x_3$$

$$\dot{x}_4 = e_2x_2 - g_2x_4$$

where the parameters are the same as before. In the final two models that we consider, additional parameters are added to model the effects of progenitor cell expansion.

Model $S9$: This model shares features with $S6$, but with the addition of progenitor cell self renewal. The equations for $S9$ are

$$\dot{x}_0 = ax_0(1 - x_0 - x_1 - x_2) - (c_1 + c_2)x_0 \quad (4.14)$$

$$\dot{x}_1 = c_1x_0 + p_1x_1(1 - x_1 - x_3) - e_1x_1$$

$$\dot{x}_2 = c_2x_0 + p_2x_2(1 - x_2 - x_4) - e_2x_2$$

$$\dot{x}_3 = e_1x_1 - g_1x_3$$

$$\dot{x}_4 = e_2x_2 - g_2x_4$$

where $(a, b, c_1, c_2, p_1, p_2, e_1, e_2, g_1, g_2)$ characterise the model: the p_i control self renewal of the progenitor cells.

Model $S10$: This model allows for progenitor cell expansion that is unregulated either by self or differentiated cells. It provides a bridge between

Symbol	Definition
a	Rate of self renewal of x_0
$c_{1/2}$	Rate of differentiation of $x_0 \rightarrow x_{1/2}$
e_1	Rate of differentiation of $x_1 \rightarrow x_3$
e_2	Rate of differentiation of $x_2 \rightarrow x_4$
g_1	Rate of death of x_3
g_2	Rate of death of x_4
$p_{1/2}$	Rate of self renewal of $x_{1/2}$

Table 4.2: *Description of parameters that characterise models S5-S10.*

models $S5$ and $S9$. The dynamics of $S10$ are given by

$$\dot{x}_0 = ax_0(1 - x_0 - x_1 - x_2) - (c_1 + c_2)x_0 \quad (4.15)$$

$$\dot{x}_1 = c_1x_0 + p_1x_1 - e_1x_1$$

$$\dot{x}_2 = c_2x_0 + p_2x_2 - e_2x_2$$

$$\dot{x}_3 = e_1x_1 - g_1x_3$$

$$\dot{x}_4 = e_2x_2 - g_2x_4$$

where the parameters are the same as for $S9$.

This completes the S -model family that will be investigated in this chapter. Before we study the properties of these models in detail, we introduce our method for stability analysis and give some general results using this method.

4.4 Methods

4.4.1 A generalised stability analysis approach

In order to conduct stability analyses for each of the models we evaluate the Jacobian matrix at each fixed point for a given parameter vector. We sample the parameter space randomly, keeping only those points that give rise to fixed points in the positive quadrant, i.e. we discard parameters giving rise to fixed points that are not feasible in the sense of being biologically relevant or ‘reachable’, from some initial conditions.

We follow these steps to determine the stability probability for a given model around a fixed point:

1. Solve $\dot{\mathbf{x}} = 0$ to find symbolic expressions for the values of each species (total number n) at each fixed point in terms of the parameters.
2. Sample a $1 \times d$ parameter vector $\mathbf{p}$ (model contains d free parameters) and evaluate the fixed point for these values. Check *reachability*: if the fixed point lies within the positive orthant, i.e. all species ≥ 0 , proceed. If not, resample.
3. For $\mathbf{p}$, find the eigenvalues λ_i , $i \in (1, 2, \dots, n)$ of the Jacobian, J . Recall that $J = \frac{\partial F(\mathbf{x}_0)}{\partial \mathbf{x}}$ where F is given by the model equations, $\dot{\mathbf{x}} = F(\mathbf{x})$ and evaluated at fixed point $\mathbf{x}_0$.
4. Repeat steps (2) and (3) until a total of N reachable points are found.

5. Construct $\mathbf{s}$, where $s_j = 1$ if $\max(\operatorname{Re}\{\lambda_i\}) < 0$ for λ_i of $p_j \in \mathbf{p}$, and $s_j = 0$ otherwise. i.e. s_j indicates whether or not the model is stable around the fixed point for p_j .
6. Calculate the stability probability: $P(\text{stable}) = \frac{1}{N} \sum_{k=1}^N s_k$.

Note that the stability probability $P(\text{stable}) = P(\text{stable}|\text{reachable})$. That is, we calculate the probability that a given fixed point is stable given that it is reachable from the starting point (parameter values and initial conditions).

Aside from calculating $P(\text{stable})$, given the distribution of Jacobian matrices that this analysis gives us, we can also investigate the fixed points of a model in other ways, such as by quantifying the basins of attraction of a model and looking at their intersection. We can also look at basins of attraction in parameter space: the effects that varying parameters (rather than initial conditions) have on the reachability of the fixed points.

4.4.2 Null distributions to describe the stability of a model

In this chapter, we set out to study the stability properties of stem cell models and in particular to determine which properties of the models confer to them more or less stability around a certain fixed point. So, in addition to comparison between models, for a single model we would like to compare its stability properties with those of a permuted version of the same model. In

doing so we essentially create a null distribution for the stability properties of a model around a given fixed point.

In fact, we consider two null distributions for each model. First, we sample with replacement from the empirical distributions that we have for each entry in the model Jacobian, maintaining the entry place in the matrix – denote this distribution $\phi_{i,j}$. For example, to sample from $\phi_{1,1}$ is to sample from the N values we have for the first entry in the Jacobian: $J_{1,1}$. This we call the independent null distribution: it removes dependencies between entries in the Jacobian. Second, we sample again from the distribution over all Jacobian entries but now we combine the values for all entries so that, for each entry in our new Jacobian, we sample with replacement from the empirical distribution containing the $n \times n \times N$ values – we denote this distribution θ . This is an empirical *i.i.d.* distribution since we are now sampling independently from the same distribution for each entry in the Jacobian.

To construct these two null distributions we follow steps (1)-(6) as above but replace (3) by (for the independent null distribution – ‘independent’ case in Figure 4.3):

- Construct a new Jacobian by sampling with replacement from $\phi(i, j)$ for each entry $J(i, j)$ and find its eigenvalues

and (for the *i.i.d.* null distribution – ‘*i.i.d.*’ case in Figure 4.3):

- Construct a new Jacobian by sampling with replacement from θ for each entry in J and find its eigenvalues.

Shown in Figure 4.3 are histograms and pairwise density plots for Jacobian entries corresponding to the model introduced by Lorenz (1963), which is described below. We see that the true distributions (those drawn from the unperturbed Jacobians of the ODE model) are different for different entries – not identical – and that they are not independently distributed, since a clear correlation is seen for the plotted entry pair, $(3, 2)$ and $(3, 3)$. In the independent case (introduced above), dependencies are lost although the distributions are still non-identical. Finally in the *i.i.d.* case, the distributions are now identical in addition to being independent.

4.5 Results

4.5.1 Generalised stability analysis enables global characterisation of the stability of a model and its fixed points

In order to demonstrate the method of probabilistic stability analysis, we first consider the model of Lorenz (1963), as an example of a system of

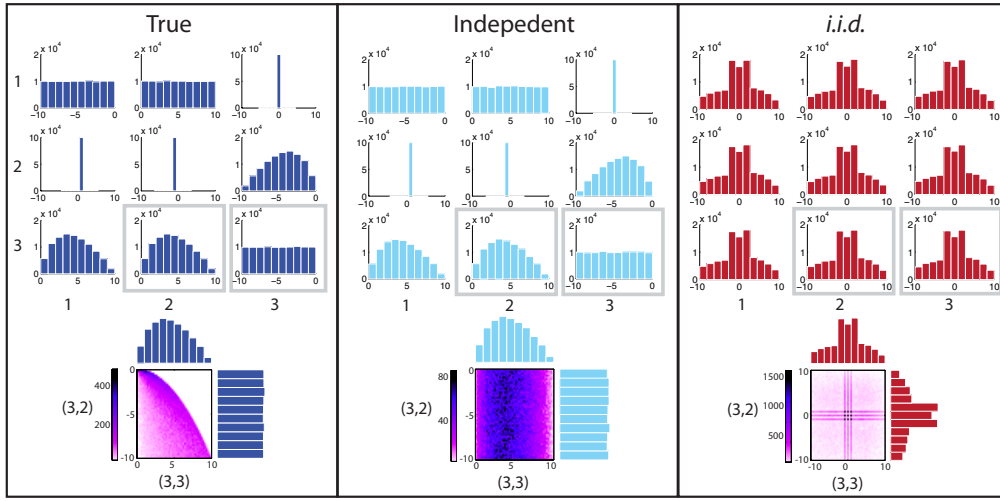


Figure 4.3: *True, independent and i.i.d. distributions over Jacobians of the Lorenz model.* Histograms for each entry in the 3×3 Jacobian for the Lorenz model are shown, for 100,000 randomly sampled parameters. Entries are marked and counted from top left. Histograms are shown for each of the true case, independent case and *i.i.d.* case (see text for how these were constructed). For one pair of entries – (3,2) and (3,3) – we plot the pairwise density using a heatmap, showing that only in the true case are there dependencies between entries.

interacting variables. This model is given by the equations

$$\dot{x}_0 = -\sigma x_0 + \sigma x_1 \quad (4.16)$$

$$\dot{x}_1 = \rho x_0 - x_1 - x_0 x_2$$

$$\dot{x}_2 = x_0 x_1 - b x_2$$

where the parameter set (σ, ρ, b) specifies the model. The model has two fixed points, one of which is the origin. We perform our analysis on the other non-zero fixed point. As described in the Methods section, we perform stability analysis by sampling 100,000 parameter sets and calculating the eigenvalues of the Jacobian and thus the stability for each. We also construct the two null distributions that we use for comparison: the independent and the *i.i.d.* distributions. In Figure 4.4 are the eigenvalue spectra plotted for these three distributions, both overlaid and as heat maps, where the density can be seen.

For the true distribution we find the fixed point has a probability of stability of 0.999: the fixed point is almost always stable from anywhere in the sampled region of parameter space. The corresponding probability of stability for the independent case is 0.841 and for the *i.i.d.* case is 0.105. So the probability drops in the independent case in comparison with the true case, but not by a large amount; under this distribution the fixed point is still stable for the majority of parameters. In the *i.i.d.* case however, the fixed point is stable only around 10% of the time.

Looking at the spectra in Figure 4.4 we see that the true and independent eigenvalue distributions are neither circular nor uniformly distributed. This

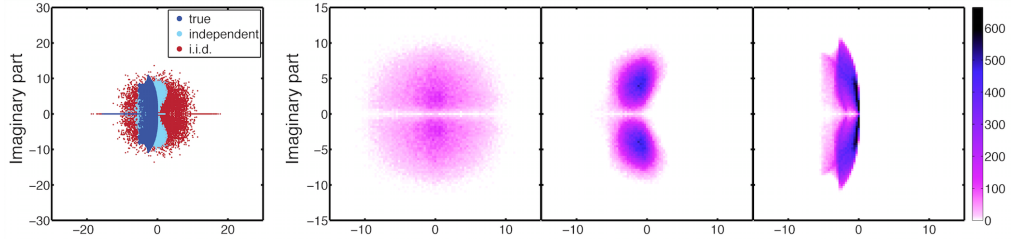


Figure 4.4: *Eigenvalue spectra of the Lorenz model under three random matrix distributions.* On the left is a scatter plot showing (left to right) the *i.i.d.*, independent and true distributions for the Lorenz model. on the right the same eigenvalue distributions are plotted using a heatmap, demonstrating that not only are the true and independent distributions non-circular, but that they are also non-uniform. Eigenvalues with real part = 0 are omitted from this plot.

highlights the fact that investigating the stability of Jacobian matrices derived from an underlying model rather than a purely random distribution requires methods other than the circular law.

4.5.2 Structure maintains stability across all

S-models

We now move on to study the fixed points of the stem cell models introduced above, using the framework we have developed here. In Figure 4.5 we show an example of the eigenvalue spectra of a fixed point for model *S5*, for the three distributions – true, independent and *i.i.d.* This illustrates how the true and independent distributions deviate from the circular *i.i.d.* distribution, as we also saw above for the Lorenz model.

In order to compare the fixed points for each stem cell model, we calculate the probabilities of stability for each; these are shown in Table 4.3. We

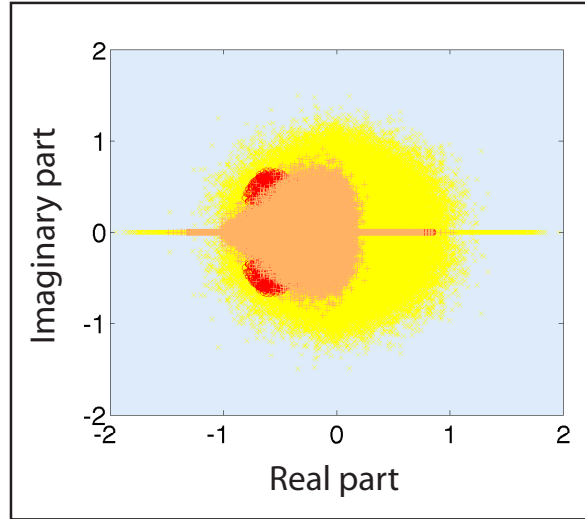


Figure 4.5: *True, independent and i.i.d. distributions of eigenvalues derived from model S5.* In yellow is the *i.i.d.* distribution and in orange and red are the independent and true distributions, respectively. While the *i.i.d.* distribution is close to circular, the other two distributions deviate considerably from a circle on the complex plane, illustrating that the circular law does not hold here.

also give in Figure 4.5.2 the eigenvalue spectra for certain fixed points of models $S5 - S10$, alongside the structure of these models. The $S5$ spectra shown here differ from that shown in Figure 4.5 because Figure 4.5 did not take into account whether a given fixed point was feasible or reachable, i.e. the system can reach it from some starting parameter values. Since the reachability of a fixed point is important biologically, we proceed to analyse parameter distributions that give rise to reachable fixed points, such as those shown in Figure 4.5.2.

Model	Fixed point	True	Independent	<i>i.i.d.</i>
S1	1	0.25	0.25	0.034
	2	0.75	0.75	0.13
	3	1.0	0.94	0.19
S2	1	0.25	0.25	0.034
	2	0.75	0.75	0.084
	3	1.0	0.97	0.14
S3	1	0.50	0.50	0.056
	2	1.0	1.0	0.080
S4	1	0.25	0.25	0.034
	2	1.0	1.0	0.080
S5	1	0.83	0.83	0.0039
	2	1.00	1.0	0.010
S6	1	0.83	0.83	0.0038
	2	—	—	—
	3	1.0	1.0	0.018
	4	1.0	1.0	0.018
S7	1	0.83	0.83	0.0038
	2	1.0	1.0	0.011
S8	1	0.83	0.83	0.0036
	2	—	—	—
	3	1.0	0.96	0.023
	4	1.0	0.96	0.023
S9	1	0.96	0.96	0.011
	2	0.46	0.46	0.0056
	3	0.0	0.0	0.0040
	4	0.96	0.96	0.011
S10	1	0.27	0.21	0.0021
	2	1.0	1.0	0.0

Table 4.3: *The stability probabilities for models S1-S10.* For each model, the true stability probability of each fixed point is given, alongside the associated independent and *i.i.d.* probabilities. Where probabilities are not given, this fixed point was not reachable for any parameter values.

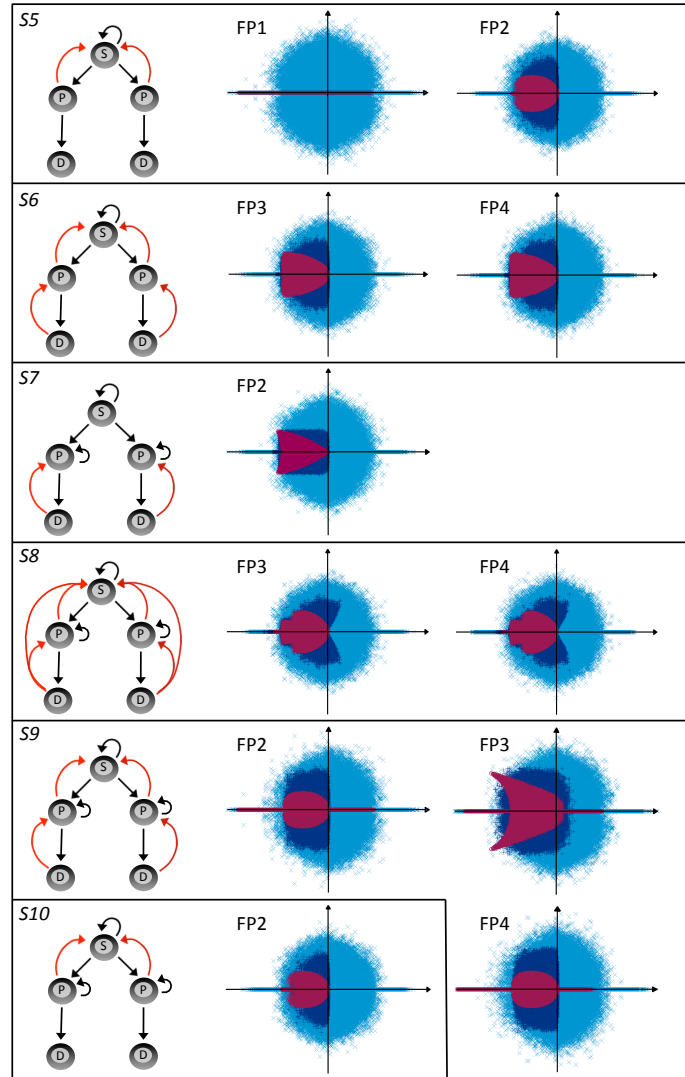


Figure 4.6: *Graphical representation and eigenvalue spectra of models S5-S9.* The left hand columns shows graphical representations of the models: S, stem cell; P, progenitor cell; D, differentiated cell. Black arrows show self renewal/production processes and red arrows show inhibition processes (self inhibition is not shown). Shown right are the eigenvalue spectra of certain fixed points of the models plotted on the complex plane. In purple is the real spectra, in dark blue the independent spectra and in light blue the *i.i.d.* spectra. The fraction of points with negative real part (to the left of the ordinate) represent the probability of stability for this fixed point. Not all fixed points of each model are shown. Note that from these representations one cannot necessarily extract the probability of stability since all eigenvalues for each fixed point are shown, not just the largest one.

In Table 4.3 we have the probabilities of stability for the fixed points of each model. Fixed point 1 is the origin in each case, the probabilities of which we are not particularly interested in biologically, so will not analyse them further here. Fixed point 2 of models *S6* and *S8* is not reachable for any parameter values in the sampling range which is why no probability is given for it.

From this table, we can see the striking impact that structure has on stability. In most cases the true probability of stability of each of the fixed points 2 - 4 is 1.0, so the fixed point is guaranteed to be stable whenever it can be reached from the starting point. Comparing this with the *i.i.d.* distribution we have constructed by permuting all Jacobian entries, we see a dramatic decrease in the probability of stability, in most cases the *i.i.d.* probability of stability is very close to 0.

In the case of the independent Jacobian distributions, we see that the stability probabilities are the same as, or slightly less than, the true stability probabilities. This tells us that it is the variance between entry distributions, rather than the dependencies between them, that crucially controls the stability of these models. These differences between variance in the Jacobian's entry distributions and dependencies are highlighted in Figure 4.3.

So by imposing structure we can bestow stability onto a system. Given a set of relationships between species (in this case described by an ODE model), we see that the species can be stable to perturbations about their fixed point values. If we take this same set of species and link them at random rather

than according to some interaction scheme, this stability is lost. In his original paper, May (1972) hinted at how structure can improve stability: discussing the improved stability of those multi-species communities that were ‘arranged in “blocks”’. Here we have demonstrated with examples how stability can be imposed by structure. The limitation of these results is that we are not able to distinguish between fixed points 2 - 4 of these models in terms of their stability probability. In the discussion we will consider how we arrived at these results and look at ways to distinguish between the fixed points.

4.5.3 Different parameter regimes control the proportion of progenitor and differentiated cells at steady state

For two of the models under investigation here, $S6$ and $S8$, there are two steady states – fixed points 3 and 4 – that yield positive population sizes for all of the species. In this section we consider only these biologically interesting fixed points; we are not interested in fixed point 1 (the origin) or fixed point 2 (not reachable for any parameter values sampled). We then investigate what parameter values can give rise to each of these two fixed points; in doing so we can characterise the behaviours that lead to them.

First we studied whether, for any parameters sampled, we are able to reach both fixed points with positive population sizes for all species. This bistability would be interesting to characterise by studying those parameters that permit it and linking them to behaviours of the system. In this case

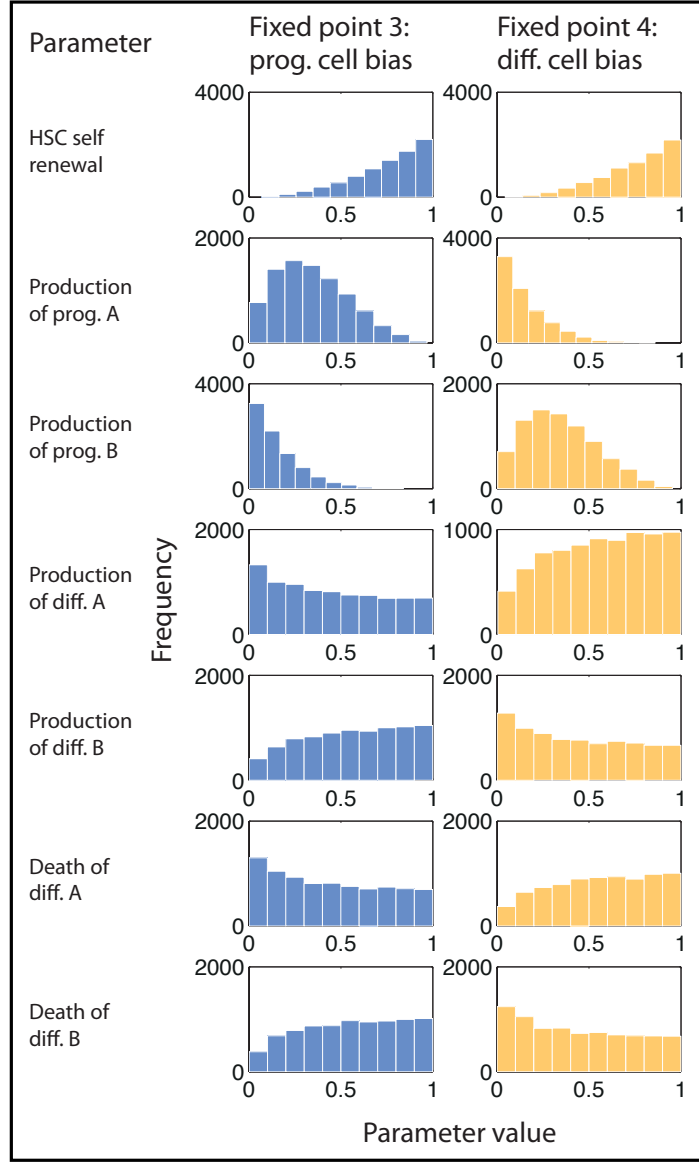


Figure 4.7: *Production rates of progenitor and differentiated cells affect the steady state reached.* Histograms of parameter values that lead to one of two fixed points for model *S6*. The first fixed point is characterised by a higher proportion of progenitor cells (prog. cell bias) and the second fixed point is characterised by a higher proportion of differentiated cells (diff. cell bias). ‘A’ and ‘B’ denote the two possible lineages that stem cells give rise to. To reach the progenitor cell-biased state, lineage A is favoured, whereas to reach the differentiated cell-biased state, lineage B is favoured.

however, there are no such parameters: those that give positive population sizes for all species at one fixed point do not at the other.

Given that certain parameters lead to one or the other fixed point, but never both, we look at what differences there are in the distributions of parameters leading to each. From the total sample of 100,000 parameter sets, we find that approximately 8000 lead to fixed point 3 and another approximately 8000 lead to fixed point 4. In order to ascribe significance to the results we obtain, we need to understand what state each of the fixed points corresponds to. Recall Figure 4.5.2 for a graphical depiction of model *S6*: a stem cell gives rise to two progenitor cell populations, which here we call A and B to distinguish them, although both lineages have the same qualitative dynamics. Each progenitor cell population then gives rise to a corresponding differentiated cell population.

Fixed point 3 corresponds to higher population sizes for progenitor cells by a factor of 10 – for both lineages – compared with the differentiated cell populations. Fixed point 4, in contrast, is characterised by higher levels of differentiated cell populations, again by approximately a factor of 10, relative to the progenitor cell populations. In Figure 4.7 the distributions of parameters that give rise to these two different states are plotted, along with a description of their meaning.

By studying Figure 4.7 we can describe the differences that lead to one fixed point or the other. To reach the state dominated by progenitor cells requires higher production rates of lineage A progenitors than lineage B. It also requires lower production and death rates of differentiated cells of lineage

A compared to lineage B. We observe symmetries between the distributions that lead to each state: to reach the state dominated by differentiated cells requires the converse rate relations than those just described. That is, a higher production rate for progenitor cells of lineage B than lineage A, and lower production and death rates for the differentiated cells of lineage B than lineage A.

Describing how lineage bias influences the proportions of different cell species at steady state is especially interesting given the importance of branching fates in the haematopoietic hierarchy. The analysis performed here on fixed points 3 and 4 extends the concept of mapping a model's basin of attraction in parameter space. We can find what regions in parameter space lead to one, or another, fixed point and begin to delineate a boundary between them. Characterising the behaviour of a model with respect to a broad area in parameter space, rather than only at some specific values, enhances our understanding of a model and its potential use. This idea is developed further in the next chapter where we study qualitative inference and model behaviour.

4.6 Discussion and conclusions

The relationship between complexity and stability in ecosystems, despite having a substantial body of research devoted to it over the last 50 years, is yet to be clearly elucidated (Allesina & Tang, 2012; Elton, 1958; Gardner & Ashby, 1970; Ives & Carpenter, 2007; May, 1972; Roberts, 1974; Saavedra et al., 2011). Here, we have shed some light on the matter by dropping the

i.i.d. assumption that much previous work has been based on, and investigating the stability properties of systems that we expect to be closer to the real world in that they exhibit heterogeneity and dependencies between species.

Performing analyses of these systems, where ‘real world’ properties are bestowed by sampling from an underlying ODE model rather than a random matrix distribution, we find that the probability that their fixed points are stable is much higher than it would be were these systems rewired at random. We also find that the heterogeneity between species’ interactions (given by non-identical distributions) – more so than the dependencies between them (given by non-independent distributions) – is important in determining the system’s stability. We demonstrate the ways in which structured systems are more stable and thus are able to persist in nature.

This probabilistic stability analysis was applied to a family of 10 simple model motifs describing the dynamics of interacting stem cell species and their progeny. The models differed in the total number of species and in the ways feedback entered the system. We were first interested in characterising the stability properties of the biologically relevant fixed points of each model and comparing them. This comparison did not however yield great insight, because all of the relevant fixed points had a stability probability of (or very close to) 1. It is possible that the different feedback motifs that we considered did not differ enough to provide many distinguishable fixed points. It is also possible that the probability of stability of the biologically relevant fixed points of a system is robust across a wide class of population

dynamical models. It would be an interesting next step to address larger systems of ODEs, or the stability properties of other types of system, such as delay differential equations.

One of the factors that led to this indistinguishability of the fixed points of these models was the condition that each fixed point be reachable. When we do not impose this constraint, the stability probability is lower. This tells us that there are unstable regions of the model, but that these do not correspond to biological behaviour regions. These results correspond well to early work done on the reachability of fixed points by Roberts (1974). In a similar way to our analysis, Roberts showed using a simple example how the stability of a system increased when only reachable (feasible) points were considered.

For two of the models studied here, *S6* and *S8*, two biologically relevant fixed points were found to be reachable from certain points in parameter space, and we analysed these two models further. The results are only presented for *S6* however, because the behaviour of *S8* was qualitatively the same, indicating that feedback from differentiated cells onto stem cells – the distinction between the two models – does not alter the fixed point behaviour. Our study of *S6* found that the balance of progenitor and differentiated cells in the system is controlled by the propensity of stem cells to favour production of progenitor and differentiated cells in one or the other lineages. Given the multiplicity of haematopoietic progenitor cell species and the possible interactions between them (Wang & Wagers, 2011; Wilson & Trumpp, 2006) (see Figure 2.1), such asymmetries are interesting to find

and could have important implications for haematopoiesis. To analyse this branching process further, we should compare this model to cell species data, allowing us to distinguish between the model's two lineages.

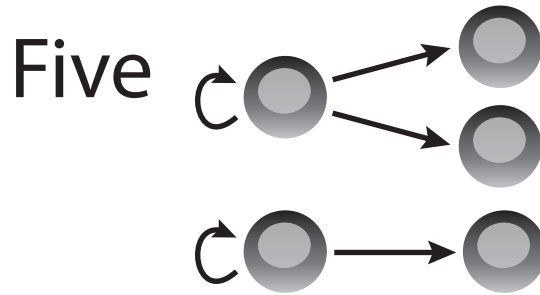
The definition of asymptotic stability used throughout this chapter – that a system at a fixed point will return to the same fixed point following a perturbation – at times may not match the biologically ‘stable’ properties that we aim to describe. One example of such a mismatch is that of oscillating systems, which can be stable in the sense of persistence. Another example is more subtle: if we compare two bistable systems, one where both fixed points have positive values for all species, and the other where for one fixed point at least one (or more species) goes to 0, we might wish to distinguish between them. That is, we might wish to call the perturbation from one state to another that causes at least one species to vanish greater (in the sense of being destabilising) than the perturbation that causes a state change that is not associated with the extinction of any species. This is an interesting avenue for future work where perhaps different criteria for stability could be used that reflect other aspects of biological homeostasis, such as species’ extinction (Saavedra et al., 2011) or the effects of neutral mutations (Traulsen et al., 2012).

Currently, our stability analysis relies on deriving analytical expressions for the fixed points of each model in order to keep computational cost down and so that we are able to keep track of each fixed point individually. Extending these methods so that analytical expressions for the system are not required represents another direction for future work. However, challenges remain,

such as how to characterise and distinguish each fixed point. If we proceed numerically, finding the value of each of (say) four fixed points, for a given parameter set, and then repeat with a new parameter set, we should keep track of which fixed point is which. This is straightforward when we have symbolic expressions, but otherwise it is not. Another challenge to address is that of defining satisfactory distance requirements between numerical simulation and the fixed point.

Overall, in this chapter we have characterised the behaviour of biologically motivated systems in a new way, based upon probabilistic analysis of their fixed points and their stability. We have seen that structure – in the sense of either an underlying interaction network (ODE model) or heterogeneity between species – bestows stability on such systems. We argue that both of these are biological ‘structures’: the former is structured in a natural sense; the latter is structured in the sense that the interaction terms of the Jacobian are structured. Some are always 0; some share distributions; some may have symmetrical or unique distributions. We have shown how the stability dramatically decreases when structure is removed. We found this to be the case for all of the stem cell models that we studied here, and were not able to distinguish their interesting fixed points based on stability properties alone. For models with multiple biological steady states, we identified how each could be reached and in doing so mapped out the ‘basins of attraction’ in parameter space. This provided insight into how branching decisions in the haematopoietic hierarchy can be made. In the next chapter we will develop new models based on similar principles to those introduced

here, and we shift our focus onto competition within the stem cell niche.



Qualitative inference: desirable behaviour regimes where leukaemia can be controlled

Parts of this chapter have been published in (MacLean et al., 2013).

5.1 Overview

Haematopoietic stem cells (HSCs) are responsible for maintaining the blood system throughout life and must be located in their ecological niche (the bone marrow) to function correctly. In this chapter we propose that cells with oncogenic potential — cancer/leukaemia stem cells (LSCs) — and their progeny will also occupy this niche. We study how the dynamical

relationships between HSCs, LSCs and the progeny of each evolve using an ecological framework that presents species' survival within the niche as a question of competition. Two mathematical models that describe the dynamics of these species allow investigation into the conditions necessary for defeating a malignant invasion of the niche. To characterise the models' behaviour we employ an inferential framework which allows us to study regions in parameter space that give rise to desired behaviour together with an assessment of the robustness of the dynamics. Using this approach we map out conditions under which HSCs can outcompete LSCs. We do this in a qualitative manner: we search for behavioural conditions rather than single trajectories. In therapeutic applications we clearly want to drive haematopoiesis into such desirable regimes and the current analysis provides some guidance as to how we can identify new therapeutic targets. Our results maintain that sustaining a viable population of HSCs and their progeny in the niche may often already be nearly sufficient to eradicate LSCs from the system.

5.2 Introduction

In this chapter, models are developed and studied that describe the behaviour of HSCs and their progeny under competition from malignant species – leukaemia stem cells (LSCs) and their progeny. As has been described previously, the key characteristic of an HSC is its *stemness*. Here, HSCs and LSCs share this characteristic and are modelled accordingly.

HSCs and – we assume here – also LSCs are characterised by their de-

pendency on the *niche*. In stem cell biology, the HSC niche comprises the complete set of environmental signals originating from HSC progeny, bone marrow stroma cells and other factors that regulate HSC function (Wang & Wagers, 2011; Wilson & Trumpp, 2006). There has been much recent work on specifying the location of the HSC niche within the bone marrow (Forsberg & Smith-Berdan, 2009; Renstrom et al., 2010; Trumpp et al., 2010) and identifying the cells and features that characterise the HSC niche, see Chapter 2 for full discussion.

Owing to the uncertainty about the anatomical constituents of the niche, an ecological approach is taken here. In ecological terms, a niche refers to the position of an organism within its environment defined by its habitat and the finite set of resources available to it. It is this broader definition that we will adopt throughout this analysis. Our modelling approach is thus flexible enough to encompass unknown components because these are (implicitly) contained within the definition of an ecological niche as considered here. We will argue below that this ecological perspective allows us to capture essential aspects of HSC biology and, in particular, the interplay between normal HSC lineages and leukaemic stem cells and their progeny.

Leukaemia, a cancer of the blood, which can occur in all types of blood cell, originates in the bone marrow (Franks & Teich, 1997). As introduced in Chapter 2, it is proposed that leukaemia is driven by LSCs that occupy a niche close to (or the same as) the HSC niche. Given this situation, it is natural to suggest that there exists competition between HSCs and LSCs for niche space and resources (such as interactions with niche-maintaining

cells). Studies have found evidence for such competition using an induced leukaemia mouse model, (Lane et al., 2011) and in cases when HSCs are impaired either as a result of ageing or irradiation (Henry et al., 2010; Marusyk et al., 2009). Although much attention has been paid to the molecular signalling and regulatory pathways controlling HSC behaviour and, to a lesser extent, how LSCs could interact with and disrupt this behaviour, to our knowledge there have been few studies that have framed this problem in terms of ecological competition (despite the fact that an increasing number of mathematical and computational approaches are gaining a foothold in stem cell biology (Aïnseba & Benosman, 2011; Daukste et al., 2009; Duff et al., 2012; Enderling et al., 2009; Johnston et al., 2010; Sieburg et al., 2011; Solé et al., 2008; Wolkenhauer et al., 2011)).

The key aims of this chapter are to elucidate the cell intrinsic factors that exert the greatest influence over the species they control, under the effects of competition from other species within the niche. The basic mathematical analysis of ecological competition between multiple species is now well developed, and a host of analytical results have been obtained for the canonical models (Hofbauer & Sigmund, 1998; May, 1973b). There is a large body of work on the gene regulatory pathways that are involved in haematopoiesis and the related cancers (see e.g. (Aïnseba & Benosman, 2011; Chaidos et al., 2012; Dalerba et al., 2007; Roy et al., 2012)); we take a slightly higher population-level view. In doing so we hope to address another aim which is to identify types of intervention that will allow us to better understand and control the differentiation and proliferation

of competing HSCs and LSCs. The current perspective complements the molecular and mechanistic studies which are aimed at elucidating cellular decision making processes. The population dynamics of cell lineages set constraints on the range of the dynamical behaviour of HSCs and their differentiation/proliferation behaviour and vice versa.

Whereas for some biological systems there is a wealth of data against which models can be fitted or compared to, in other cases we have to base our analyses on less detailed information. Here, rather than estimating parameter values from what might be insufficient or irrelevant information, it is prudent to take a more general approach. With a few notable exceptions (Boisset et al., 2010; Lo Celso et al., 2009b; Paik & Zon, 2010; Wang et al., 2003), we are not yet able to identify native HSCs *in vivo* with confidence and gather extensive amounts of data. It is possible to study leukaemia in mice, but the experimental models available do not fully reflect the dynamics of leukaemia development taking place in patients. Xeno-transplantation of human leukaemia cells into immunocompromised mice is widely used, but it leads to inconsistent engraftment levels and does not take into account the role of the immune system in recognising leukaemia development. Host conditioning prior to injection of human and mouse leukaemia cells is also widely used and provides a far-from-physiological starting point for the development of the disease. In experimental models that do not use conditioning, such as transgenic mice, leukaemia is initiated through expression of a driving oncogene in a specific cell population and not in a single leukaemia initiating clone through the use of artificial promoters.

These initial conditions are very different to what we would expect to see in leukaemia patients where disease will typically arise from mutations in a single cell or lineage. Furthermore, recent results (both experimental and theoretical) suggest that within a population of cancerous cells it may be hard to determine which proportion of cells exhibits stemness (Johnston et al., 2010; Quintana et al., 2008).

For these reasons we proceed to explore large regions of parameter space and characterise the solutions they produce, rather than investigate models based upon any particular parameter set. By analysing models according to distinct outcomes which we select, we can test their robustness with regards to changes in input. We are thus able to home in on dynamical regimes that exhibit certain types of behaviour in a computational affordable manner (even when other tools such as bifurcation analysis would be inappropriate because we are interested in quick responses). This, we propose, allows us to suggest new ways for therapies to coax the behaviour in the HSC niche into certain desirable directions.

5.3 Model development

5.3.1 Model assumptions

HSCs (species x_0 in the model) are ancestors of all types of blood cells, a diverse set performing many different functions (Wang & Wagers, 2011). Here we are not investigating all the possible blood cell fates, but the general mechanism of blood cell production; therefore we do not consider each

type of blood cell individually, but group them all under the category of terminally differentiated blood cells (x_2). In between these and the stem cells, we group together all haematopoietic progenitor cells that are lineage committed and progressively lose self-renewal capacity (x_1). For the questions under investigation in this work, this coarse-grained version of the haematopoietic tree is sufficient (Mangel & Bonsall, 2008). All species are produced within the bone marrow ecological niche, but the population of differentiated cells, x_2 , will leave the niche at a high rate as these cells enter the blood stream; in healthy adult humans, for example, some 2 million new erythrocytes enter the bloodstream from the bone marrow every second (Sompayrac, 2008).

Complementary to this model of the healthy haematopoietic system, we model leukaemia as two sub-populations of cancer cells, y_0 and y_1 . Here y_1 are terminally differentiated leukaemia cells and y_0 are proliferating leukaemia cells. We will mainly refer to the y_0 population as LSCs but in fact in our coarse-grained model y_0 represents any cell population driving the growth of leukaemia, including pre-LSCs, LSCs or leukaemia progenitor cells. This is important because both pre-LSCs and progenitor cells transforming to LSCs have been observed to play a role in the progression of leukaemias (Krivtsov et al., 2006; Lane et al., 2011).

We do not model the intermediate stage in leukaemia differentiation explicitly (leukaemic species analogous to the healthy progenitor population x_1). Progenitor cell species x_1 is included because we are interested in how its dynamics change during disease progression; the effects of healthy species

of a leukaemic progenitor population are of less interest. Furthermore, including an intermediate leukaemia species does not qualitatively change the results that are obtained.

5.3.2 Niche competition

Competition models are based upon the ideas introduced by Lotka and Volterra and later ecologists, who consider two or more species that rely on the same limited (environmental) resource for their survival and proliferation (Hofbauer & Sigmund, 1998; May, 1973b; Roughgarden et al., 1989). In the case of haematopoiesis, this resource is the niche space available to stem cells, defined through the molecular signals necessary to maintain correct function (Wilson & Trumpp, 2006).

In general terms, the competition, which exists between two species, X and Y , where x and y define the population size of each type, can be expressed in mathematical form as

$$\begin{aligned}\frac{dx}{dt} &= x(\alpha - \gamma(x + y)) \\ \frac{dy}{dt} &= y(\beta - \gamma(x + y)),\end{aligned}\tag{5.1}$$

where α and β represent the growth rates of species X and Y , respectively, γ is the strength of the feedback signal and $\alpha, \beta, \gamma \geq 0$; further examples of competition are given in Hofbauer & Sigmund (1998). This system assumes that the effect from feedback is linear and that both X and Y experience feedback proportional to the combined population sizes of the two species. In a more general case, the feedback effect from $(x + y)$ could vary for each

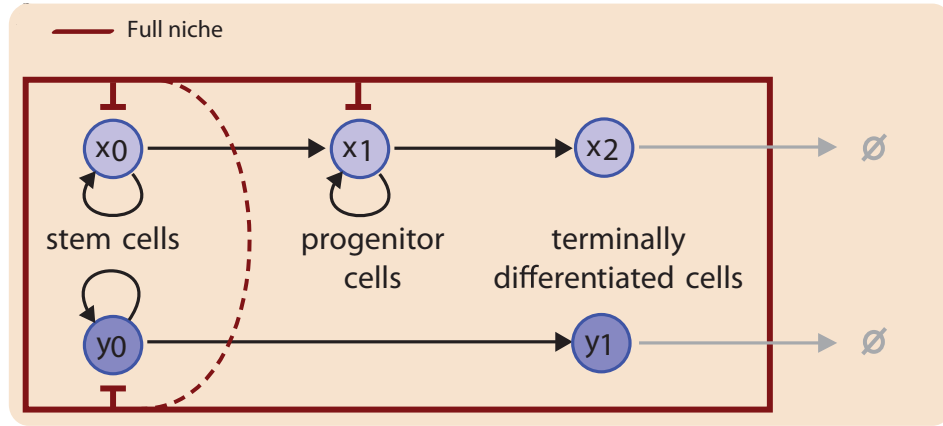


Figure 5.1: *Diagrammatic representation of stem cell competition within the bone marrow niche.* Normal haematopoiesis occurs through HSC (x_0) self renewal and differentiation to produce progenitor cells (x_1) that in turn differentiate into mature blood cells (x_2). Alongside this healthy lineage, there exists a pool of mature leukaemia cells (y_1) that is maintained by self renewing LSCs (y_0). Mature cells are able to leave the niche by migration into the blood stream. Bold/dashed lines represent different possible niche boundaries. Note that cell-cell feedbacks are not shown here.

species, or the individual contributions could be altered. In this (simple) example the equilibrium solutions can be trivially characterised. For all solutions with $\alpha \neq \beta$ the species with the greater growth rate will dominate and the other will go to zero. If Equation 5.1 is generalised so that the feedback parameters for each term vary, coexistence or bistability (in which case the initial conditions affect the steady state taken up by the system) are also possible.

5.3.3 Model A

In Model A we describe five species with competition modelled by linear feedback. The first three species are haematopoietic stem cells (x_0), progenitor blood cells (x_1) and terminally differentiated blood cells (x_2). Leukaemic stem cells (y_0) and fully differentiated leukaemia cells (y_1) complete the set; see Figure 5.1. The equations which fully specify the dynamics of Model A are given below.

$$\dot{x}_0 = a_x x_0 (k - z) - b_x x_0 \quad (5.2)$$

$$\dot{x}_1 = b_x x_0 + c x_1 (k - z) - e x_1$$

$$\dot{x}_2 = e x_1 - f x_2$$

$$\dot{y}_0 = a_y y_0 (k - z) - b_y y_0$$

$$\dot{y}_1 = b_y y_0 - g y_1$$

Here $z = x_0 + x_1 + x_2 + y_0 + y_1$, the parameter set $(a_x, b_x, c, e, f, a_y, b_y, g)$ characterises the phenotypes of the five species involved (see Table 5.1) and k represents the so called *carrying capacity*, i.e. the total number of cells that can exist within the bone marrow. In this case k is ‘hard’ in the sense that it provides an exact upper limit for the total population size of the niche (In model B below, the carrying capacity is ‘soft’ and is not equal to the maximal total population size). In simulations $k = 1$ is used such that the population sizes are given as concentrations.

Symbol	Definition
a_x	Rate of self renewal of x_0
b_x	Rate of transition $x_0 \rightarrow x_1$
c	Rate of self renewal of x_1
e	Rate of transition $x_1 \rightarrow x_2$
f	Rate of disappearance of x_2
a_y	Rate of self renewal of y_0
b_y	Rate of transition $y_0 \rightarrow y_1$
g	Rate of disappearance of y_1

Table 5.1: *Description of parameters that describe model A.*

5.3.4 Model B

Model A appeals due to its simple form which ought to facilitate explanation of its behaviour. There are, however, certain aspects of the haematopoietic system which have not been included but which will be important biologically (and mathematically due to the effects they would have on the dynamics of the system). These are addressed with our second model. Here stem cell fates are modelled explicitly; that is, the terms controlling the rates of symmetric renewal ($x_0 \rightarrow 2x_0$), asymmetric renewal ($x_0 \rightarrow x_0 + x_1$), symmetric differentiation ($x_0 \rightarrow 2x_1$) and loss of x_0 ($x_0 \rightarrow \emptyset$) are incorporated explicitly (similarly for y_0). This is in contrast to, for example, a_x in model A which accounts for production of HSCs by any means. This allows for more detailed investigation of the relationships between fate choices, at the cost of increasing the size of the parameter space. Mangel & Bonsall (2008) model HSCs and their descendants in a similar fashion. There, feedback terms associated with stem cell fates were given an exponential form; in model B this form is adopted. It is arguably more realistic that as the

niche fills (or becomes crowded) the feedback effect increases faster than linearly, due, for example, to complex patterns of feedback from other cells in the niche.

The third important distinction between the models regards the form of the niche. Within the bone marrow niche the location of HSCs (and LSCs) may be limited to specific sub-regions governed by particular signals (Aguila & Rowe, 2005; Calvi et al., 2003). In order to capture this behaviour we separate the feedback terms into one containing the stem cell species, and one containing all other species.

The assumption that there exists no overlap of species between these sub niches may be oversimplifying matters in principle, but ought not to matter in practice as we will argue later. Thus, for the present at least, this assumption is upheld, also in an attempt to keep model complexity under control and maintain parsimony. Equation 5.3 describes the dynamics of model B, and Equation 5.4 specifies the form of the feedback.

$$\dot{x}_0 = x_0\Phi_1(\lambda_1 - \lambda_3\Phi_3) - \mu_4x_0 \quad (5.3)$$

$$\dot{x}_1 = x_0\Phi_1(\lambda_2\Phi_2 + 2\lambda_3\Phi_3) + (\lambda_5 - \mu_6)x_1$$

$$\dot{x}_2 = \chi_1\mu_6x_1 - \mu_8x_2$$

$$\dot{y}_0 = y_0\Phi_1(\kappa_1 - \kappa_3\Phi_3) - \nu_4y_0$$

$$\dot{y}_1 = y_0\Phi_1(\kappa_2\Phi_2 + 2\kappa_3\Phi_3) - \nu_8y_1,$$

where

$$\begin{aligned}\Phi_1(\gamma_1, x_0, y_0) &= e^{-\gamma_1(x_0+y_0)} \\ \Phi_2(\gamma_2, x_1, x_2, y_1) &= e^{-\gamma_2(x_1+x_2+y_1)} \\ \Phi_3(\gamma_3, x_1, x_2, y_1) &= e^{-\gamma_3(x_1+x_2+y_1)}\end{aligned}\tag{5.4}$$

The parameter set required to fully specify the phenotypes of these species is now

$(\lambda_1, \lambda_2, \lambda_3, \mu_4, \lambda_5, \mu_6, \chi_1, \mu_8, \kappa_1, \kappa_2, \kappa_3, \nu_4, \nu_8, \gamma_1, \gamma_2, \gamma_3)$. These quantities are described in Table 5.2. In addition, in Table 5.3, an explanation is provided for each term included in Equation 5.3, with a description of the effects of feedback on this term, as controlled by different cell populations in the model.

In our model analysis, the first 13 of these parameters are varied; for simplicity we assume that γ_1, γ_2 and γ_3 which control feedback are identical and held fixed at the values $\gamma_1 = \gamma_2 = \gamma_3 = 0.01$ (resulting in an arbitrary rescaling of other parameters). Equation 5.4 that defines the feedback terms, accounts for the effects of the occupants of each sub niche on all differentiating species (x_0, x_1 and y_0). Φ_1 describes the effect of x_0 and y_0 on themselves; Φ_2 and Φ_3 describe the effects of x_1, x_2 and y_1 on the differentiating species.

The models presented here rely on certain other key assumptions which ought to be made explicit. The forms of y_0 and y_1 are similar to their healthy counterpart species: x_0 for y_0 and x_1 or x_2 for y_1 . Model A assumes that only x_2 or y_1 can leave the niche by migration. This is not the case for

Symbol	Definition
λ_1	Rate of symmetric renewal of x_0 ($x_0 \rightarrow 2x_0$)
λ_2	Rate of asymmetric renewal of x_0 ($x_0 \rightarrow x_0 + x_1$)
λ_3	Rate of symmetric differentiation of x_0 ($x_0 \rightarrow 2x_1$)
μ_4	Rate of migration/apoptosis of x_0
λ_5	Rate of amplification of x_1
μ_6	Rate of disappearance of x_1
χ_1	Rate of transition of $x_1 \rightarrow x_2$
μ_8	Rate of migration/apoptosis of x_2
κ_1	Rate of symmetric renewal of y_0 ($y_0 \rightarrow 2y_0$)
κ_2	Rate of asymmetric renewal of y_0 ($y_0 \rightarrow y_0 + y_1$)
κ_3	Rate of symmetric differentiation of y_0 ($y_0 \rightarrow 2y_1$)
ν_4	Rate of migration/apoptosis of y_0
ν_8	Rate of migration/apoptosis of y_1
$\gamma_1, \gamma_2, \gamma_3$	Terms controlling feedback strength

Table 5.2: *Description of parameters that describe model B.*

model B, where all species can migrate or die. In model A, y_1 has the same form as x_2 ; in model B, y_1 is expressed as a combination of the structure of x_1 and x_2 . That is, y_1 is produced directly from y_0 but cannot self-replicate.

5.4 Methods

5.4.1 Parameter sampling and qualitative inference

The problem of classifying the steady state solutions of models of complex systems is often substantial. However the rewards of such efforts — a qualitative analysis of the results a model may provide — are great. Analytical expressions for the steady states can be obtained by solving the system of ODEs after setting each equation to 0.

Term	Sign	Explanation	Effects of feedback
$x_0 \cdot \lambda_1 \cdot \Phi_1$	+	x_0 growth	$x_0, y_0 \uparrow \Rightarrow \Phi_1 \downarrow$ $\Rightarrow$ decrease in growth ^a
$x_0 \cdot \lambda_3 \cdot \Phi_1 \Phi_3$	−	x_0 diff. (sym.)	$x_0, y_0 \uparrow \Rightarrow \Phi_1 \downarrow$ $\Rightarrow$ decrease in diff. ^a $x_1, x_2, y_1 \uparrow \Rightarrow \Phi_3 \downarrow$ $\Rightarrow$ decrease in diff. ^b
$x_0 \cdot \mu_4$	−	x_0 loss	
$x_0 \cdot \lambda_2 \cdot \Phi_1 \Phi_2$	+	x_1 production (asym.)	$x_0, y_0 \uparrow \Rightarrow \Phi_1 \downarrow$ $\Rightarrow$ decrease in production ^a $x_1, x_2, y_1 \uparrow \Rightarrow \Phi_2 \downarrow$ $\Rightarrow$ decrease in production ^b
$x_0 \cdot 2\lambda_3 \cdot \Phi_1 \Phi_3$	+	x_1 production (sym.)	$x_0, y_0 \uparrow \Rightarrow \Phi_1 \downarrow$ $\Rightarrow$ decrease in diff. ^a $x_1, x_2, y_1 \uparrow \Rightarrow \Phi_3 \downarrow$ $\Rightarrow$ decrease in diff. ^b
$x_1 \cdot \lambda_5$	+	x_1 growth	
$x_1 \cdot \mu_6$	−	x_1 loss	
$x_1 \cdot \chi_1 \mu_6$	+	x_2 production	
$x_2 \cdot \mu_8$	−	x_2 loss	
$y_0 \cdot \kappa_1 \cdot \Phi_1$	+	y_0 growth	
$y_0 \cdot \kappa_3 \cdot \Phi_1 \Phi_3$	−	y_0 diff. (sym.)	
$y_0 \cdot \nu_4$	−	y_0 loss	
$y_0 \cdot \kappa_2 \cdot \Phi_1 \Phi_2$	+	y_1 production (asym.)	
$y_0 \cdot 2\kappa_3 \cdot \Phi_1 \Phi_3$	+	y_1 production (sym.)	
$y_1 \cdot \nu_8$	−	y_1 loss	

Table 5.3: *Explanation of the terms of Equation 5.3 (model B).* All terms are presented in the form (species $\times$ rate $\times$ feedback control). Feedback control of the terms controlling leukaemic species (y_0, y_1) are qualitatively similar to those for (x) and so are not repeated. Notation: diff. is differentiation; sym. is symmetric; asym. is asymmetric.

^adue to saturation of the stem cell compartment.

^bdue to saturation of the progenitor/differentiated cell compartment.

By sampling parameter sets and simulating a model from these, we can obtain unbiased dynamical trajectories as examples of a model's behaviour that we can classify, according to some suitable criteria. Here we classify trajectories according to whether HSC lineages win, LSC lineages win, or neither lineage wins (either due to coexistence, bistability or because the result falls into neither category). Obtaining enough samples for such outcome-clustering to be meaningful, however, may be difficult in practice whenever the number of parameters affecting the system's dynamics exceeds five or six. We can use Latin hypercube (LH) or Sobol sampling approaches in order to sample more evenly over the entire parameter space, but even so coverage is often too sparse (Santner et al., 2003). For a model with 10 parameters, for example, we require in excess of 1,000,000 samples to obtain just 4 samples per dimension. Four different possible sampling regimes are shown in Figure 5.2.

Furthermore, even given successful classification of a model's solutions, dissecting or further analysing such clustering within a high dimensional space is challenging. This is particularly true when we are investigating parameters that vary over orders of magnitude. For these reasons, we proceed using a targeted approach to determine the biological relevance of the solutions each model provides.

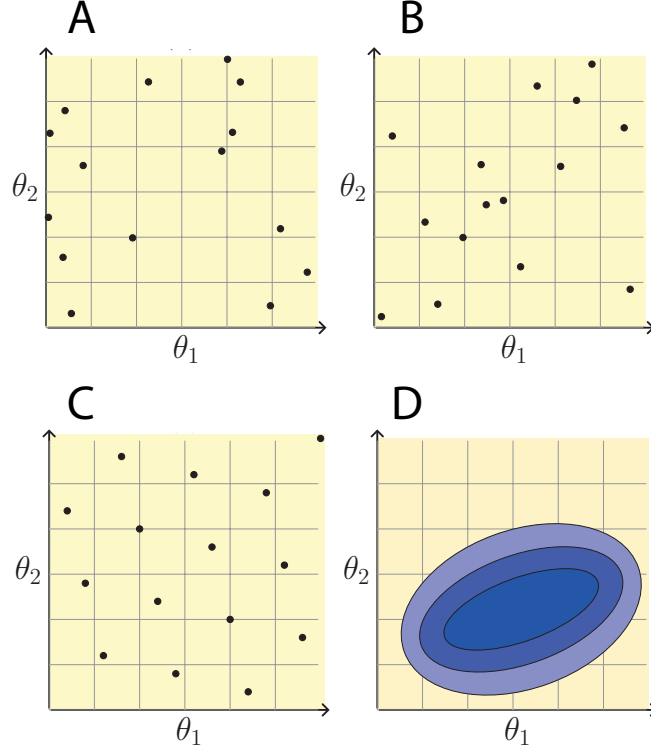


Figure 5.2: *Exploring model space through different parameter sampling regimes.* **(A)** Random sampling draws from a multivariate uniform distribution to cover the two dimensional parameter space of θ_1 and θ_2 . This is the most general (undirected) approach. **(B)** Subdividing the space using a Latin hypercube allows for more even sampling over the parameter range. **(C)** Quasi-Monte Carlo sampling based on the Korobov sequence defining a low discrepancy point set. In certain circumstances such methods outperform simple MC approaches such as (A) and (B). **(D)** In contrast to the global approaches of (A) - (C), ABC sampling is targeted towards regions of acceptable model behaviour, as represented by the blue circles growing brighter with each iteration. The advantage of such an approach is that time is not wasted searching through those areas of space that we are not interested in.

5.4.2 Robustness analysis via approximate Bayesian computation

Rather than following such a global (but non-adaptive) sampling strategy, we can target our investigation by systematically identifying regions in parameter space where a model produces some desired behaviour. That is, rather than fitting a model to experimental data, we specify (surrogate) data which correspond to a system state which we are interested in. The conditions under which this type of behaviour is realised are initially unknown but can be identified with approximate Bayesian computation (ABC). Originally developed for cases where computing the likelihood is not feasible, which is often the case for large, complex systems, ABC methods provide a convenient solution (Toni et al., 2009) to identify those parameters that have a high probability of having generated the data. ABC methods forego evaluation of the likelihood in favour of comparing real with simulated data (Beaumont et al., 2002; Marjoram et al., 2003). If simulated data x' for some parameter value θ' is found to be in good numerical agreement with the observed data, x_0 , then θ' is accepted as a valid draw from the (unknown) posterior distribution,

$$P_{\text{ABC}}(\theta|x_0) \propto \mathbf{1}(\Delta(x', x_0) \leq \epsilon) f(x'|\theta) \pi(\theta), \quad (5.5)$$

where $\pi(\theta)$ summarises our prior knowledge/expectations on θ and $f(x_0|\theta)$ is the data-generating model, $\Delta(a, b)$ is a suitable distance function and ϵ our tolerance level, which determines how closely real and simulated data have to agree. ABC methods can also be employed in the context of Bayesian model selection in order to choose the best model (given the

data and compared to the alternative models considered) (Toni & Stumpf, 2010). Bayesian model selection approaches strike an implicit balance between the predictive power of a model, its complexity (by favouring more parsimonious models over larger models unless these describe the data significantly better), and robustness. However, the same approach can also be used to scan models for the ability to generate certain outcomes. Here we are interested in situations where there is competition between HSCs and LSCs, and in particular we want to understand under what conditions the ‘ecological’ interactions between the two cell types (and, of course, their progeny) favour the HSCs and ideally result in suppression of LSCs.

To this end we use qualitative inference to seek to determine $P(\theta|y_0 = 0)$: the probability distribution over model parameters θ , that — reliably and efficiently — yields loss of LSCs (and stable HSC pools). Our method is qualitative in the sense that we do not fit our model to surrogate data, instead we fit it to criteria that specify the behaviour we require: stable, positive populations of healthy cell species and vanishing populations of leukaemia species.

The ABC Sequential Monte Carlo (SMC) sampler of Toni *et al.* (Toni et al., 2009) allows us to do this and is already implemented in the **ABC Sys-Bio** package (Liepe et al., 2010). ABC SMC proceeds by approximating the *posterior* distribution, $P(\theta|x)$ (where x is some data or specified system behaviour), in an iterative (sequential) fashion. Compared to other sampling methods of large-dimensional parameter spaces, e.g. random, LH and Sobol sampling (Santner et al., 2003), these Bayesian methods have

the advantage of homing in on regions in parameter space that have a high probability of resulting in the desired or observed outcome and do not waste much time on exploring the generally vast regions of parameter space that do not lead to the specified outcome, see Figure 5.2. The sequential nature of the SMC framework also incorporates a natural adaptive/learning component, which here allows us to assay potential model behaviour efficiently and comprehensively (Barnes et al., 2011).

5.5 Results

5.5.1 Simulation of models with fixed parameters gives biologically relevant trajectories

As a first step, we sought to determine whether, given certain specified parameter values, models A and B will predict species' dynamics that are realistic and plausible. Parameter values were set using estimates obtained from literature where possible, and where reliable estimates could not be made, values were assigned heuristically. We proceed with this analysis to demonstrate that model simulations yield results that are realistic and meaningful. We do not however assign significance to the values chosen here; in the subsequent section parameter values are not set but are inferred based upon the behaviour we want to observe.

For the self renewal rate of HSCs, we are able to obtain a value based upon the work of Wilson et al. (2008) who identify distinct populations of active and dormant HSCs and find that the populations have predicted doubling

Par	HSC win	LSC win	Par	HSC win	LSC win
a_x	0.27	0.09	λ_1	0.18	0.15
b_x	0.15	0.2	λ_2	0.1	0.1
c	0.3	0.3	λ_3	0.04	0.04
e	0.2	0.4	μ_4	0.05	0.15
f	0.15	0.4	λ_5	0.1	0.1
a_y	0.27	0.27	μ_6	0.15	0.15
b_y	0.2	0.2	χ_1	0.1	0.1
g	0.3	0.3	μ_8	0.01	0.01
			κ_1	0.18	0.18
			κ_2	0.1	0.2
			κ_3	0.04	0.04
			ν_4	0.2	0.1
			ν_8	0.15	0.15

Table 5.4: *Models A and B produce plausible dynamics for heuristically chosen parameter values.* Parameter values given are used for simulation of model A (left) and model B (right) for the two scenarios of the healthy lineage triumphing (HSC win) and the disease lineage triumphing (LSC win).

times of 36 days and 145 days, respectively (from the fit of a model to their data). This study also estimates that the fraction of active HSCs is 0.85 and the fraction of dormant HSCs is 0.15. From these figures we find an estimate for the proliferation rate of HSCs by taking the mean of the active and dormant fractions: HSC proliferation rate = 0.012 day^{-1} or 0.09 week^{-1} . It was not possible to find reliable estimates for progenitor/differentiated cell production rates from literature: these processes have yet to be successfully quantified. The corresponding parameter values were set to reflect the need for larger haematopoietic populations as we move down the hierarchy. In most cases, values were kept to be the same in the healthy and leukaemic lineages.

In Table 5.4 are given the full parameter sets used for the simulations visualised in Figure 5.3. For both models, rates between the lineages were kept equal for some processes, and changed in those cases where biologically there is motivation to do so. The timescale used here is weeks. In model A, we assume that the proliferation rate of LSCs will be considerably higher than that of HSCs, so it is set to 3×0.09 . This led to the leukaemic lineage taking over the niche at the expense of the healthy lineage. In order to simulate an HSC-enhancing treatment, we take $a_x \rightarrow 3a_x$ (so $a_x = a_y = 0.27$) and decrease the rate of production of x_2 (thus keeping a larger proportion of healthy stem/progenitors), this leads to successful eradication of the leukaemia lineage.

A similar analysis is performed for model B. Since here we explicitly model stem cell self renewal, asymmetric differentiation and symmetric differentiation, the estimated proliferation rate of 0.09 week^{-1} is split between these processes, with lower weighting given to the differentiation processes. Similarly to model A, by decreasing healthy differentiated cell production we enable eradication of leukaemia from the niche. If the leukaemia pool is boosted by an increase in asymmetric differentiation (κ_2), this allows the leukaemia lineage to beat the healthy species. The trajectories for each simulation are shown in Figure 5.3.

From these simple simulations we see that, given parameter values taken either from literature or chosen heuristically, we can recover niche dynamics of the sort that we expect to observe for different conditions in vivo. Given this proof of principle, we proceed to look at model behaviour globally,

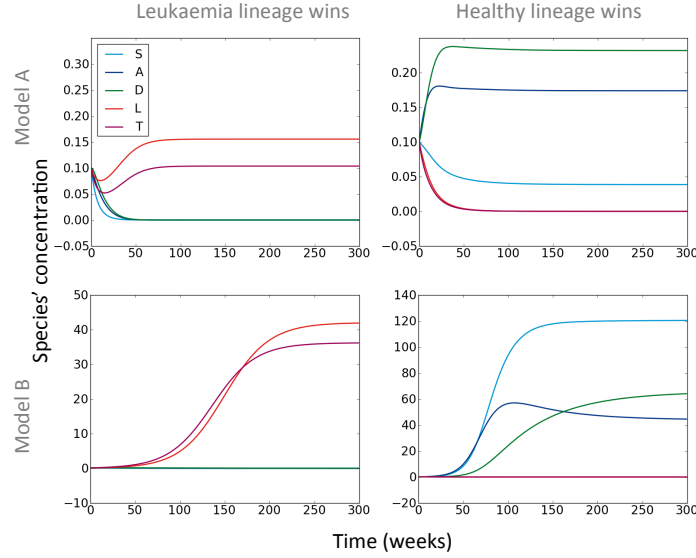


Figure 5.3: *Simulated trajectories for the species of models A and B under different conditions.* On the left are shown simulations for models A and B under conditions where the leukaemia lineage wins and drives the healthy lineage to extinction. On the right are shown simulations for another scenario where the healthy lineage outcompetes the leukaemic one. Notation: S, HSCs; A, progenitor cells; D, differentiated cells; L, LSCs, T, leukaemia differentiated cells. For parameter values used to simulate see Table 5.4.

guided not by certain parameters but certain desirable behaviours.

5.5.2 Desirable regimes reveal distinct dynamics for HSCs and LSCs

Here we investigate the circumstances under which the two models favour HSCs whilst suppressing LSCs, that is x_0, x_1 and $x_2 > 0$ and $y_0 = y_1 = 0$ on reaching steady state. Each parameter was sampled from a uniform prior distribution in the range $[0, 1]$. The posterior distribution that we obtain shows the regions of parameter space that are most likely to suppress

leukaemia. Each parameter is described by a histogram; pairs of parameters can be visualised on 2 dimensional density plots. This is illustrated in Figure 5.4A.

Overall we see that those parameters controlling LSC dynamics are less well inferred (less peaked) than those controlling HSC dynamics. This suggests that requiring a species to die out is a less stringent condition than requiring a species to reach some finite population size, as is the case for haematopoietic species x_0, x_1 and x_2 . A comparison between the parameters controlling healthy and leukaemia dynamics is shown in Figure 5.4B. The full set of results for model A is given in Appendix A.

The posteriors that we obtained appear to be unimodal. Therefore it is possible to perform principal component analysis (PCA) on the posterior distribution for the parameters in order to evaluate the robustness of these systems to (joint) changes to parameters (Erguler & Stumpf, 2011; Secrier et al., 2009). (It is not advisable to perform PCA on multimodal datasets as the technique is not designed to capture the multimodality.) The principal components are constructed by finding the eigenvalues and eigenvectors of the covariance matrix of the parameters. The first principal component (corresponding to the largest eigenvalue) corresponds to the direction where the posterior is widest; the last principal component (corresponding to the smallest eigenvalue) points into the direction of least variance. This provides us with an indication of those parameters that are tightly constrained by the imposed/desired system behaviour; i.e. varying these parameters by even moderate or small amounts can have appreciable effects on system

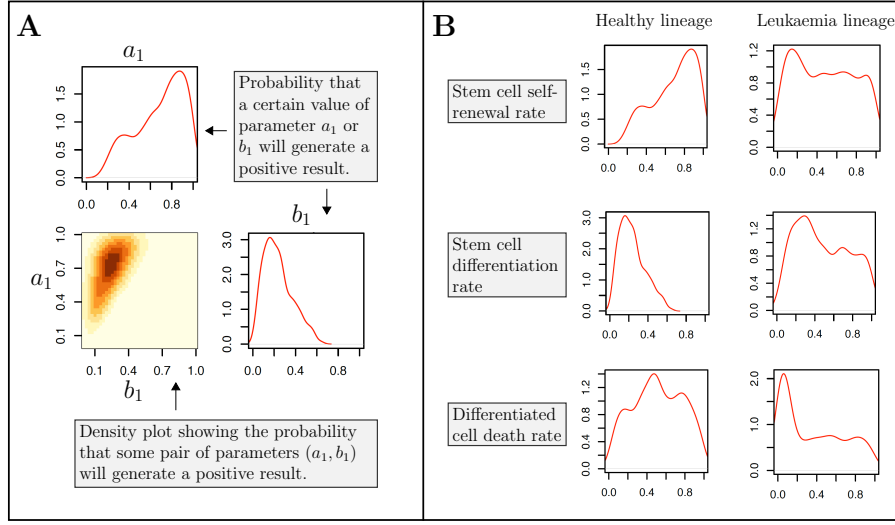


Figure 5.4: *Properties of the posterior distribution and comparison of parameter posterior values that allow leukaemia suppression for model A. (A)* Marginal posterior distributions (red histograms) and density plot. *(B)* For model A: comparison of marginal posterior distributions for different rates controlling healthy and leukaemia cell species. Overall, rates controlling leukaemic species are less well inferred than those controlling healthy species.

output (in the vernacular of Sethna *et al.* these are the “stiff” parameters) (Gutenkunst *et al.*, 2007). Here these smallest principal components are of primary interest unlike in most cases where PCA is used to try to identify parameters that control the largest variance in the data (i.e. distinguish between outcomes). Such “sloppy” parameter (combinations) are less critical for determining system dynamics and may sometimes even not be identifiable (i.e. extend over the whole prior range).

For model A, the majority of the last principal component (56%) was composed of contributions from the parameters a_x, b_x and e . The second last

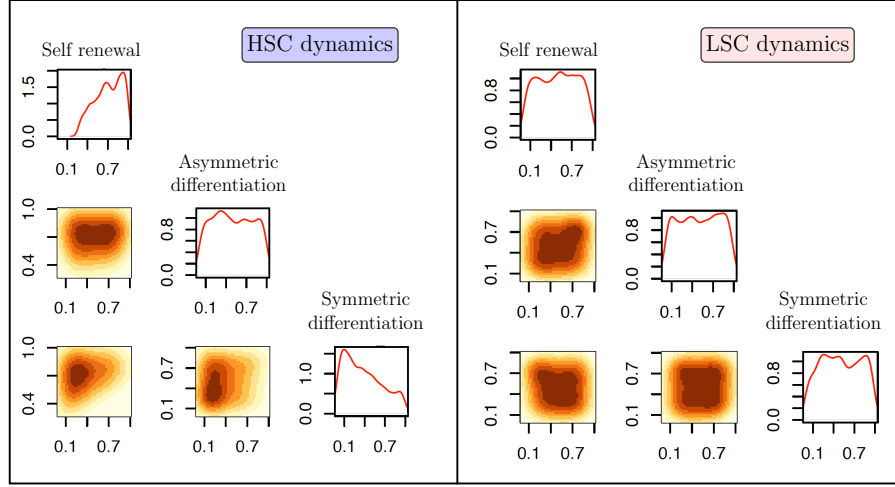


Figure 5.5: *Parameter posterior values that allow for leukaemia suppression under model B.* Shown are parameter ranges for healthy and leukaemia stem cell rates that allow suppression of leukaemia to occur. For HSCs, self renewal and symmetric differentiation should be controlled. For LSCs, little to no control is required. See Figure 5.4A for explanation of the plots.

principal component was mainly composed (63%) of e and f . So, given our analysis of the posterior via PCA, in order to suppress leukaemia by populating the niche with healthy blood cells, it is most important to observe specific rates of the production of progenitor blood cells, self renewal of the stem cells that are producing them, production of terminally differentiated cells from progenitor blood cells and the migration/death of these differentiated cells.

For model B, the 13 free parameters were also sampled from uniform priors on $[0, 1]$. Parameters controlling healthy cell dynamics are better inferred than parameters controlling leukaemia. As in the case of model A, the constraint $y_0, y_1 \rightarrow 0$ at equilibrium is more easily met than the constraint

for HSC species to reach a finite positive value above some threshold. In Figure 5.5 the posterior distribution for the stem cell dynamics is shown. The full output for model B is shown in Appendix A.

We see in Figure 5.5 that the posterior for HSC asymmetric differentiation is uninformative. The result that more information is given about self renewal and symmetric differentiation rates suggests that controlling the rate of HSC asymmetric differentiation does not assist in achieving LSC suppression. Controlling self renewal and symmetric differentiation of HSCs is sufficient in this case. In contrast to the parameters controlling HSC fate, none of the parameters controlling LSC fate need to be specified in order to achieve leukaemia suppression.

PCA was also performed on the results for model B. The last principal component is dominated (66 %) by λ_1 , λ_3 and μ_4 , which collectively control the fate of HSCs. This suggests that leukaemia can be beaten if a healthy population of HSCs is present. The second last principal component is mainly composed (64 %) of λ_5 and μ_6 . These parameters control healthy progenitor cell (x_1) dynamics and are analogous to e_1 and f in model A. There is also a significant contribution (8 %) from ν_4 to this principal component, suggesting that of all the parameters controlling leukaemia, the death rate of LSCs could influence model dynamics the most.

5.6 Discussion

Here we studied circumstances under which leukaemia stem cells and their progeny succumb to normal HSCs and their progeny. We have taken a population-level perspective which models a highly simplified representation of the ‘ecology’ in the haematopoietic niche. In order to interpret these results we have to address two questions: first, according to our simplified model, which parameters control the ecological balance, and which interventions would allow us to shift this balance in favour of normal non-leukaemic HSCs maintaining themselves and their progeny in the niche? And second, how can we relate these highly idealised model results to the processes occurring in the real world? The answer to the second question must depend on the answer to the first question and we address them in turn.

The results from both models suggest, perhaps somewhat counterintuitively, that it is more important to maintain the dynamics of the HSC lineage within certain bounds than to control parameters related to the LSC lineage. One reason for this is that there are many scenarios where suppression of the LSC lineage is possible, for example by assigning leukaemic species very low growth rates, but where nevertheless the HSC lineage cannot be maintained, recovers very slowly, or where both lineages co-exist. This is perhaps the most important lesson emerging from both models.

The models presented here were developed based on the biological feature set we aim to describe, the crucial components of which are stem cell differ-

entiation, competition for an ecological resource and feedback mediated by the niche. By characterising solutions according to the lineage outcomes, we show that we can compare the likelihood of niche dominance by one lineage or another given a certain model. There are certain regions of parameter space that produce results that are not permissible biologically, such as infinite population sizes. This is the case for model B in the region $\lambda_5 > \mu_6$, where results are unbounded. This can be deduced both from ABC parameter inference, which excludes this region even for large tolerances, and from simulation. So if species x_1 is self-replicating more quickly than it is disappearing, the system grows in an unstable manner.

The niche (of size K) can be divided into sub-niches, $K = K_1 + K_2$ as there is some evidence that HSCs and by extension LSCs reside in specific sub-compartments of the (ecological) niche. We must consider whether this coarse-grained description is realistic, or whether in reality we have a situation whereby some species move between sub-niches. Our central findings, however, appear to be robust features of niche-mediated feedback: in particular we believe that maintenance of the HSC lineage is important to drive down LSCs and their progeny. In turn, an assault on LSCs and their progeny on its own will likely not suffice to restore the haematopoietic system.

There are many different ways to study the behaviour of a biological system. In this work we look at a model globally, by performing targeted investigation of its whole behaviour space. This is facilitated by ABC. Thus we can understand the stability and robustness properties of our system. In

Kitano (2007), robustness is defined as the ability of a system to maintain its functions under perturbations, and contrasted with stability which has the narrower definition of maintaining certain system states. Using ABC we can characterise not only a system's stability but also its robustness by looking at the maintenance of system functions, in this case the survival of the haematopoietic lineage.

Throughout the analysis of the models presented here, we have been considering the case where the presence of LSCs is already established. This should, we feel, include pre-clinical dynamics as well. Small leukaemia populations could occur often in the ecological niche and be outcompeted by healthy species thus restoring normal haematopoiesis. These instances are interesting because they provide more clues about the conditions favourable for suppressing an invasion of the niche by leukaemia cells.

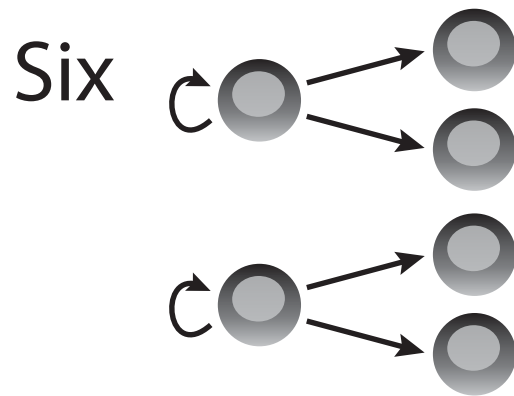
Models must strike a balance between simplicity and reality. In the endeavour to retain or, more correctly, establish simplicity, many known aspects of the biology of haematopoiesis have been left out of the current analysis. For example, more complicated hierarchies and feedback from the periphery into the niche could be included. However here too, details of the processes involved are largely unknown, and the explorative use of techniques such as ABC, which combine the ability to condition on desired or expected (quantitative and qualitative) behaviour with an assessment of the robustness of this behaviour are useful. In the future, such methods will aid experimental design and enable the study of how external (therapeutic) targets can be designed to control patterns of HSC differentiation.

We have investigated the ability of two similar but distinct models to produce certain behaviours and we ask if they do so in different ways. From the results, we find that the results for models A and B are complementary and that model B, with its added complexity, does not yield significantly greater insight into the system than (more simple) model A. It seems that the additional processes of model B – sub-niches, explicit stem cell fate processes and exponential (rather than linear) feedback – do not play crucial roles in determining model outcome: healthy vs. leukaemia lineage winning in the niche. Thus, although model B provides interesting behaviour, for reasons of parsimony we will proceed in future analyses with investigation only of model A.

5.7 Conclusion

In this chapter our main goals were twofold. Firstly, we sought to develop models of HSC/LSC competition within the niche and methods for their analysis, and secondly we sought to determine how leukaemia could be eradicated from the niche. The models developed built upon the basic structure of stem cell models illustrated in the previous chapter and extended them to include competitive effects. Upon completion of our first goal, we were able to identify favourable niche conditions and we found that controlling leukaemia lineage dynamics was the most important factor to achieve its successful eradication from the niche. In the next chapter we will continue to build upon these ideas of cellular competition, next applied in the context of biological (clinical) data. We will retain model A (model *M1* in Chapter

6) for future analysis but will not proceed to analyse model B further. Two new models are additionally presented, along with an investigation into how different patients exhibit different disease outcomes, specifically remission from – or relapse of – leukaemia.



Ecology in the haematopoietic stem cell niche determines the clinical outcome in chronic myeloid leukemia

Parts of this chapter have been published in (MacLean et al., 2014).

6.1 Overview

Chronic myeloid leukaemia (CML) is a blood disease that disrupts normal function of the haematopoietic system. Despite the great progress made in terms of molecular therapies for CML, there remain large gaps in our understanding. In this chapter light is shed on certain aspects of this disease by comparison of models that describe CML in different ways. This builds upon the work presented in Chapter 5 by introducing additional methods used for analysis and applying these methods to study the behaviour of models, given clinical data describing patients undergoing treatment for CML.

By comparison of mathematical models that describe CML progression and aetiology we sought to identify those models that provide the best description of disease dynamics and their underlying mechanisms. The essential element shared by the models used here to study CML is a description of the dynamics of healthy and leukaemia stem cell populations and their progeny within the haematopoietic stem cell niche. Data for two clinical outcomes – disease remission or relapse – are considered and we investigate these using Bayesian inference techniques throughout. We find that it is not possible to choose between the models based on fits to the data alone, however by studying model predictions we can discard models that fail to take niche effects into account. More detailed analysis of the remaining models reveals mechanistic differences: for one model leukaemia stem cell dynamics determine the disease outcome, whilst for the other model disease progression is determined at the stage of progenitor cells, in particular

by differences in progenitor death rates. This analysis also reveals distinct transient dynamics that will be experimentally accessible, but are currently at the limits of what is possible to measure. To resolve these differences we need to be able to probe the haematopoietic stem cell niche directly. Our analysis highlights the importance of further mapping of the bone marrow haematopoietic niche microenvironment as the ecological interactions between cells in this niche appear to be intricately linked to disease outcome.

6.2 Introduction

Chronic myeloid leukaemia (CML) is a blood cancer primarily affecting the myeloid lineage species of the haematopoietic system. As introduced in Chapter 2, in the majority of patients it is characterised by a specific chromosomal translocation: $t(9;22)(q34;q11)$ (Epstein et al., 1999). This results in a fusion gene made by combining the breakpoint cluster region gene (BCR) on chromosome 22 with the Abelson murine leukaemia viral oncogene (ABL) on chromosome 9. The protein product of this fusion – BCR-ABL – is a tyrosine kinase with ABL comprising a transmembrane region. Given the form of the protein, a successful therapy for (BCR-ABL+) CML has been devised using tyrosine kinase inhibitors (TKIs). These bind to the ABL domain of BCR-ABL and successfully block growth of BCR-ABL+ cells (Muller et al., 2003).

The cellular and molecular processes involved in healthy haematopoiesis must be disrupted in disease (Hu et al., 2009). But there is solid evidence that they operate in parallel to the normal haematopoietic system:

thus a branch of the haematopoietic system propagates deficient cells and generates the ensuing pathology, while other cells of the haematopoietic system continue to operate alongside (Hope et al., 2004; Jamieson et al., 2004; Lapidot et al., 1994; Wang & Dick, 2005). This is, of course, akin to problems frequently encountered in ecology. Based on this analogy we here attempt to understand the interplay between the healthy and CML systems. This is a particularly helpful perspective if interactions between healthy and diseased cell types can affect the disease progression, as is now widely believed.

Interactions will include shared reliance on nutrients, co-factors, and physiological conditions required to maintain stem cell properties. These factors can be subsumed under the umbrella of the stem cell niche. The dynamics in the niche determine the abundance of stem cells, and if there are different ways in which healthy and cancerous stem cells can exploit or adjust to the niche environment, then ecological dynamics can influence the progression of the disease in a way akin to classical ecological dynamics (Mangel & Bonsall, 2008).

Here we study CML dynamics in light of patients who are undergoing treatment for CML via TKIs. The level of leukaemia in patients is quantified by measuring the proportion of differentiated leukaemia cells in blood, samples of which are taken every few months. Our data fall into two different categories: data for patients who go into remission, and data for patients exhibiting relapse. We consider these cases separately in order to identify possible functional differences that give rise to – or at least are correlated

with – the different clinical outcomes.

Disease outcomes are investigated with respect to models that describe both healthy and leukaemia-related haematopoietic systems. It is assumed that both lineages are derived from a stem cell population – HSCs and LSCs, respectively. The models differ quite profoundly in the manner in which they consider interactions (or competition) between different cell types. We use an approximate Bayesian computation (ABC) framework throughout to calibrate models against data and compare their relative performance.

Given detailed analysis of the models, we will see that we can reject those models that fail to account for niche dynamics explicitly and from the outset. Furthermore we can show that the available time course data strongly suggest an important role of the HSC niche dynamics for CML disease progression and outcome. We find evidence for the niche effects extending beyond the HSC/LSC level and also affecting the dynamics of progenitor and quite possible other cell types in the haematopoietic hierarchy.

6.3 Model development

6.3.1 Model *M1* (MacLean et al., 2013) — niche competition

The first model we consider has already been introduced as model A in Chapter 5. Here we are going to refer to this model as *M1* for consistency with the others considered in this chapter. This model describes the dy-

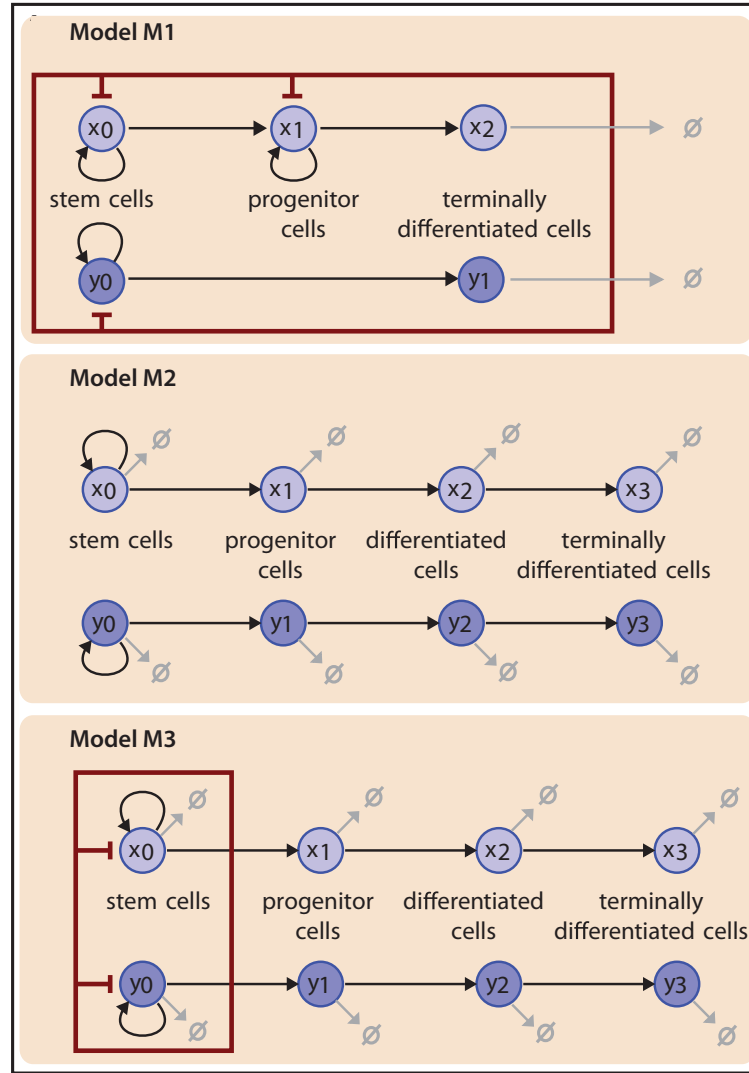


Figure 6.1: *Mechanistic models for haematopoiesis in CML*. Graphical description of models *M1*, *M2*, and *M3*. The haematopoietic niche is represented by a red rectangle: all species inside the rectangle influence the niche; only species marked by red inhibitory arrows are affected by it. For model *M2* there are no niche effects present. Cell death is illustrated by light grey arrows.

namics of five species of interacting cell populations; we will not specify its details again as these can be found in the previous chapter and in MacLean et al. (2013). This model is depicted once again in Figure 6.1 for comparison with models $M2$ and $M3$.

6.3.2 Model $M2$ (Michor et al., 2005) — niche independence

This model was developed to describe the dynamics of HSCs and LSCs and their temporal evolution under TKI therapy and contains eight species: HSCs with three progenitor populations, and LSCs and their corresponding three leukaemia progenitor populations (Michor et al., 2005). The model is depicted in Figure 6.1 and the equations that specify it are,

$$\dot{x}_0 = (\lambda(x_0) - d_0)x_0 \equiv 0 \tag{6.1}$$

$$\dot{x}_1 = a_x x_0 - d_1 x_1$$

$$\dot{x}_2 = b_x x_1 - d_2 x_2$$

$$\dot{x}_3 = c_x x_2 - d_3 x_3$$

$$\dot{y}_0 = (r_y(1 - u) - d_0)y_0$$

$$\dot{y}_1 = a_y y_0 - d_1 y_1$$

$$\dot{y}_2 = b_y y_1 - d_2 y_2$$

$$\dot{y}_3 = c_y y_2 - d_3 y_3$$

where x_0 are HSCs, y_0 are LSCs and the x_i, y_i for $i \in (1, 2, 3)$ are progeny of the two stem cell populations. $\lambda(x_0)$ describes a homeostatic function defined such that $\dot{x}_0$ is constant; u is the number of resistance mutations

	Symbol	Definition
<i>M1</i>	a_i	Self renewal rate of stem cells
	b_i	Production rate of progenitor cells
	c	Self renewal rate of healthy progenitor cells
	e	Production rate of healthy differentiated cells
	f	Death rate of healthy differentiated cells
	g	Death rate of leukaemia differentiated cells
<i>M3</i>	$\bar{x}_0$	Carrying capacity of the niche
	a_i	Production rate of progenitor cells
	b_i	Production rate of differentiated cells
	c_i	Production rate of fully differentiated cells
	r_i	Growth rate of HSC/LSC (rescaled)
	d_0	Death rate of stem cells [†]
	d_1	Death rate of progenitor cells [†]
	d_2	Death rate of differentiated cells [†]
	d_3	Death rate of fully differentiated cells [†]

Table 6.1: *Description of parameters used to characterise models M1 and M3.* Parameter symbols used by each model and their meanings are given. $i \in \{x, y\}$ denotes lineage where x is healthy and y is leukaemia. (†) death rates for model *M3* are shared between healthy and leukaemia lineages.

which here is assumed to be 0. (A version of the model that did include resistance mutations was briefly considered, but mechanistically it offers no more than the form presented here, since there are no additional interactions between species.) All other parameters are shared by model *M3* and described in Table 6.1. The main assumptions of this model are thus that each species is produced from its parent species at some rate proportional to the size of the parent population, and can die at a rate constant to its own population size. Also, the death rates (d_1, d_2, d_3) are shared between healthy and leukaemia lineages and ‘death’ here is used as a general term that can account for cell death, differentiation or migration away from the niche. In this model there are no regulatory feedback effects.

6.3.3 Model $M3$ (Foo et al., 2009) — partial niche dependence

Model $M2$ does not allow for competition between HSCs and LSCs: the two lineages progress independently. In (Foo et al., 2009), a model is introduced that describes the dynamics of the same eight species as in $M2$, but here the growth of stem cells is affected by the population size of HSCs and LSCs. In this way competitive effects are introduced and dynamics within the niche are taken into account. The form of model $M3$ used here is a rescaled version of the original proposed in (Foo et al., 2009) in order to allow comparison of the three models considered here, and we have,

$$\dot{x}_0 = (\Phi_x - 1)d_0x_0 \quad (6.2)$$

$$\dot{x}_1 = a_x x_0 - d_1 x_1$$

$$\dot{x}_2 = b_x x_1 - d_2 x_2$$

$$\dot{x}_3 = c_x x_2 - d_3 x_3$$

$$\dot{y}_0 = (\Phi_y - 1)d_0 y_0$$

$$\dot{y}_1 = a_y y_0 - d_1 y_1$$

$$\dot{y}_2 = b_y y_1 - d_2 y_2$$

$$\dot{y}_3 = c_y y_2 - d_3 y_3$$

with

$$\Phi_i = \frac{r_i + 1}{1 + \frac{r_i}{x_0}(x_0 + y_0)} \quad (6.3)$$

where $i \in \{x, y\}$ and the species (x, y) are the same as for $M2$. The parameters that specify this model are described in Table 6.1. Note that, as for

Model	Effectors	Niche reliant	Unconstrained
<i>M1</i>	all cell types	stem and progenitor cells	differentiated cells
<i>M2</i>	–	–	all cell types
<i>M3</i>	stem cells	stem cells	progenitor and differentiated cells

Table 6.2: *Different treatment of species’ niche reliance between models.* The relationship between a species and the niche can be categorised into the following groups: *effectors* – can impact the niche; *niche reliant* – can be affected by the niche; and/or can be fully *unconstrained* by the niche.

M1, the d_i that are referred to as ‘death’ rates in the succeeding analysis subsume the processes of differentiation and migration out of the niche.

In addition to these three models, we have studied one further established model that describes interactions between HSC and LSC, including competition, presented in (Roeder et al., 2006). Here similar cellular processes are modelled but using an agent-based modelling approach. The large computational cost of simulating this model, however, precludes the Bayesian analysis attempted here.

6.4 Methods

6.4.1 Data processing

CML levels are assayed through measurements of transcript levels of the BCR-ABL protein in the blood (see Figure 1). Each of the 69 patients in the study is being treated with tyrosine kinase inhibitors (TKIs) that selectively bind to the BCR-ABL protein present on CML cells. Blood

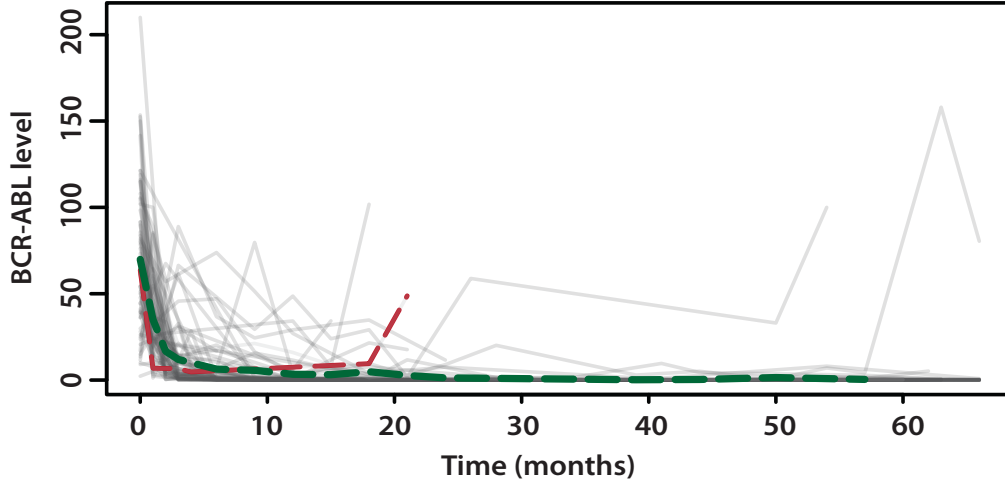


Figure 6.2: *Experimental measurements of leukaemia progression in 69 patients.* The change in the BCR-ABL level over time for each patient is represented by a grey line. The average change over all patients is illustrated by a dashed green line; the change in the BCR-ABL level for a patient who exhibits disease relapse is illustrated by a dashed red line. Model comparison is based on these two coloured trajectories.

samples are taken every few months although not always at the same time points across all patients.

The 69 patient trajectories over the time course studied are heterogeneous (Figure 6.2). We distinguish two distinct outcomes – remission and relapse. A patient is classified as in remission if the level of CML in the blood is (strictly) decreasing over time. The remission trajectory used is the mean over all 69 patients, averaged at each time point where there were at least 20 measurements taken. Behaviour is classified as relapse if an initial decrease is followed by a pronounced and continuing increase in CML levels. A single patient trajectory exhibiting such features was used for analysis.

6.4.2 Bayesian parameter inference and model comparison

We seek to perform model selection, that is, choose which model best describes the data that are available to us. There are several different ways to do this that are appropriate to systems biological problems (Kirk et al., 2013). We primarily use a likelihood-free parameter inference approach – approximate Bayesian computation (ABC) – to fit the models to data. Additionally, the probability that a given model could explain the data is computed with a likelihood-based sequential Monte-Carlo algorithm.

The ABC algorithm provided by Toni et al. (2009) allows us to infer the model parameters. ABC approaches use systematic comparison between simulated trajectories and the observed data in order to identify parameters associated with simulated trajectories that are closest to the observed data. The methods in the previous chapter give a more detailed explanation of ABC. Here, the time series for BCR-ABL is compared (using the Euclidian distance) to the ratio $T/(T + 2D)$ where T is the number of leukaemia differentiated cells and D is the number of healthy differentiated cells (Roeder et al., 2006). We use the *ABC – SysBio* package which provides a python implementation of this algorithm with GPU support (Liepe et al., 2010).

Use of a likelihood-based sequential Monte-Carlo sampler (Del Moral et al., 2006) enables us to compute the *model evidence* which is the probability to observe a dataset under a given model. We used our own implementation

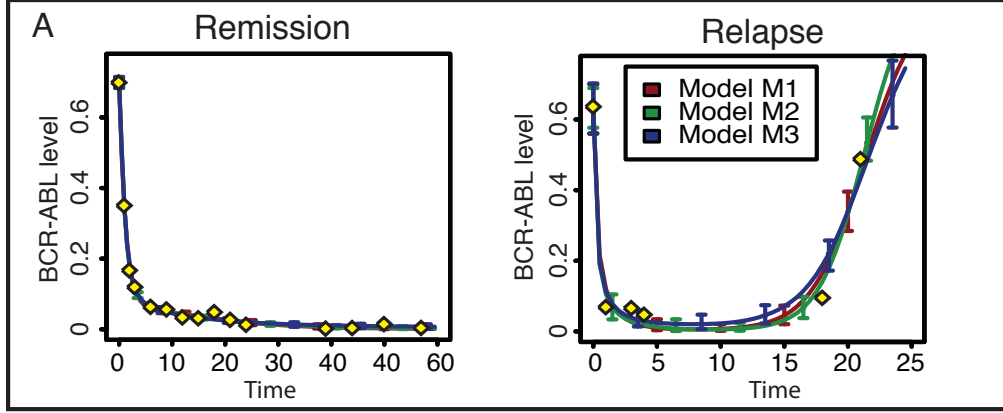


Figure 6.3: *Models are not distinguishable based on fits to the data.*

Fits to the data (yellow diamonds) for the three models and the two outcomes. All three models are able to explain the observed leukaemia evolution for the two outcomes.

of the SMC sampler algorithm in python as well as an interface to simulate the models in a computational efficient manner using a GPU accelerated ODE solver (Zhou et al., 2011).

6.5 Results and discussion

6.5.1 Models cannot be distinguished based on fit to the data alone

By the use of likelihood-free sequential Monte-Carlo Bayesian inference (Toni et al., 2009), we can perform parameter inference independently for the three models. As shown in Figure 6.3, the simulated behaviours obtained for each of the three models and two outcomes are very similar to the observed behaviours. The acceptance threshold between simulated tra-

Model	Remission	Relapse
<i>M1</i>	-40.2 (1.2)	-25.2 (0.04)
<i>M2</i>	-39.6 (1.6)	-25.4 (0.1)
<i>M3</i>	-39.9 (1.9)	-25.1 (0.03)

Table 6.3: *Ranking based on model probability does not discriminate between models.* Models are ranked by the probability of each to explain the observed datasets. For each model and each outcome, the logarithm of the model probability is computed using a sequential Monte-Carlo (SMC) sampler; the average over 10 independent runs is displayed and the number in brackets is the standard deviation. There are no significant differences between the model probabilities according to Jeffreys' scale.

jectories and observed data was chosen so that the final distance between the two was lower than $\epsilon = 4.0$ in the case of remission and $\epsilon = 15.0$ for relapse. From the similarity of the fits, we can say that all three models are able to explain these outcomes, over some range of parameters.

As a point of comparison with the ABC approach used to test the models, a likelihood-based Bayesian sequential Monte-Carlo sampler (Del Moral et al., 2006) was used to compute the model probabilities. To do this we define a likelihood function for the model giving rise to the observed data, and specify a noise term (we use an additive Gaussian noise term). We can then maximise this likelihood function to give us the probability that the model has given rise to the observed data. This again found no evidence in favour of any one of the three models. The model probability (log likelihood of the model given the data) for each case is shown in Table 6.3. According to Jeffreys' scale (Jeffreys, 1961), there is no statistically significant difference between the model probabilities.

Thus the available data alone does not allow us to decide between the three models considered here. Therefore we need to consider other properties of these models coupled with biological reasoning in order to elucidate the mechanisms underlying CML aetiology and progression.

6.5.2 The importance of competition (model checking)

The posterior distributions obtained from Bayesian parameter inference allow us to predict cell-type behaviour: trajectories of individual species are simulated using parameter values sampled from the posterior. In Figure 6.4 the trajectories for HSCs and LSCs are plotted for each model in the case of relapse. Models $M1$ and $M3$ both exhibit stem cell species' dynamics that reach finite and biological plausible steady states. This is not however the case for model $M2$. Here LSCs and all their progeny exhibit unbounded exponential growth.

Thus model $M2$ cannot represent the underlying biological mechanisms and cannot describe the dynamics of the haematopoietic system. Since our goal is to explain mechanisms of leukaemia growth and decline, we cannot hope to gain much insight using a model with such drastically implausible growth kinetics. In light of the behaviour found we can reject model $M2$ and focus our further attention on the models $M1$ and $M3$.

The distinguishing features of model $M2$ are related to competition and reliance on the niche (See Table 6.2). In model $M2$, species are independent

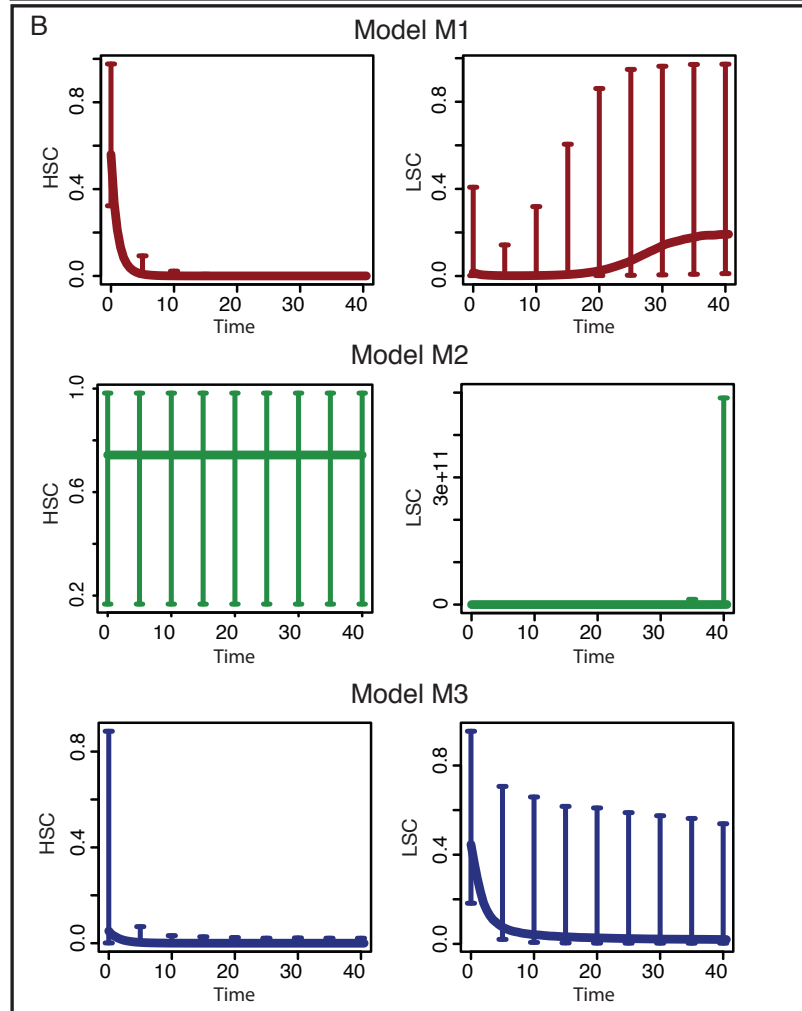


Figure 6.4: *Simulation of stem cell dynamics demonstrated the importance of niche competition.* Predicted trajectories for stem cell species for each model. The lines represent the median of the evolution and the error bars designate the 5% and 95% percentiles for a set of 1000 parameters sampled from the posterior distribution. For model *M2*, LSCs follow exponential growth for all sampled parameter values.

from reliance on the niche; there exists no explicit competition between any species. The rejection of $M2$ thus suggests that competition between haematopoietic species is a salient feature of a model that attempts to describe the dynamics of CML. Although direct competition is not the only manner by which the population could be controlled; one could also imagine other feedbacks or saturating functions that have not been modelled here. In the next section we study in more detail what form these interactions and competitive effects can take.

6.5.3 $M1$ and $M3$ offer contrasting mechanisms of leukaemia suppression (model comparison)

In order to study the differences between models, we exploit the posterior distributions obtained from the parameter inference. It is difficult to directly identify differences between models from these distributions because none of the model parameters are strongly inferred, and most of the parameters are highly correlated (full posterior distributions are shown in Appendix A); while this may be seen as failure of the inference, it is in fact, a reflection of the data (their nature, quality and quantity) and is itself informative. More crucially, perhaps, it highlights the importance of assessing the uncertainty of estimates instead of merely obtaining “optimal estimates” of parameter values. In the Bayesian framework we can still derive useful insights by making the posterior distribution the centre of the analysis and using it to predict the behaviour of the different cell-types over time. The corresponding predicted trajectories for each outcome are shown

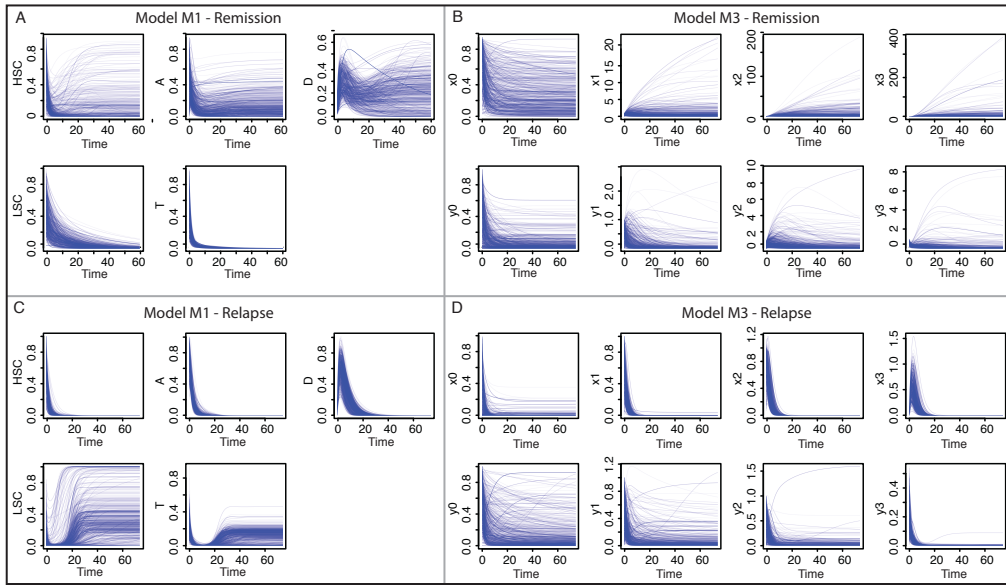


Figure 6.5: *Dynamics of model species that lead to either remission or relapse of disease.* The predicted change over time of each species for a timespan of six years (the total period of the trial), for each model (model *M1* in **A** & **C**, model *M3* in **B** & **D**) and outcome (remission in **A-B** and relapse in **C-D**). For each model and each outcome, 1000 parameter sets have been sampled from the posterior distribution and each line corresponds to the simulated trajectory for one of these parameter sets.

in Figure 6.5.

In the case of remission we note that the mechanisms of model $M3$ are harder to elucidate since here a large variety of behaviours is permitted (Figure 6.5B). Model $M3$ fitting remission also permits higher species' concentrations ($\gg 1$) than for any other outcome. Thus, while we can identify the main mechanisms leading to remission under model $M1$, direct comparison of the models' actions is difficult. For $M1$, the level of both leukaemia stem and differentiated cells declines over time, with the differentiated leukaemia cell level being more tightly constrained (Figure 6.5A). The HSC population initially decreases, but after a few time points, different behaviours — either HSC increase and decrease — are permitted. There is a sharp initial peak in the level of healthy progenitor cells, after which the behaviour is no longer tightly constrained by the parameters sampled from the posterior. This initial peak is also observed for the healthy differentiated cells, but with greater variance. We observe that disease remission for model $M1$ is a consequence of suppression of leukaemia species and expansion of the healthy differentiated cell populations through transient progenitor population growth. This prediction is directly testable, i.e. we need to study healthy blood progenitor abundances over the course of CML treatment to see whether a growth peak occurs after initial administration of TKIs.

For CML relapse, the species' behaviours under both models are more constrained (Figure 6.5C-D), even though we still have broad posterior distributions. Models can only fit the data for low levels of healthy differentiated

cells at steady state. We see that for both models the level of healthy progenitor and differentiated cells first increases to explain the temporary suppression of leukaemia, before quickly decreasing. This transient behaviour occurs more quickly for progenitor than for differentiated cells; there is a time lag as we move down the hierarchy. In model *M1*, both stem and differentiated leukaemia cells initially decrease before increasing to fill – at least a proportion of – the niche. This transient behaviour of leukaemia suppression before a return of the disease is in agreement with the literature on recurrence of CML (Mahon, 2012).

In contrast to model *M1*, in model *M3* leukaemia stem and progenitor cells can have very different behaviours. There are also some cases where HSC levels at steady state can be relatively high (occupying $\sim 40\%$ of niche space). This can be due to the number of layers in the hierarchy of *M3*, allowing effects originating at the stem cell level to be dampened in successive progeny levels; the fact that we measure only fully differentiated cells makes the details of the process hard to identify. The death/differentiation rates for progenitor and differentiated cells in *M3* are all high, meaning that the populations of each lineage are being exhausted through differentiation, migration and death of cells. The differentiation/death rates for the leukaemia lineage are slightly higher than for healthy cells. This increased rate of differentiation (or progression through the haematopoietic hierarchy) seems to be the most plausible explanation for the transient decrease of leukaemia, although from these results the mechanism by which leukaemia returns is unclear. The increased rate of differentiation has, to

our knowledge, not been reported in the literature although it should be measurable in principle if it does indeed exist.

Considering the prediction (made by both models) that the differentiated blood cell population should decline until close to zero in the case of relapse, we note that this does not imply a total loss of blood because the blood samples used consisted only of lymphocytes (Muller et al., 2003). The models predict that in the case of CML relapse, healthy lymphocyte numbers will be very low, but they do not make a prediction about the population size of myeloid lineage cells.

6.5.4 CML patient outcomes are controlled by different features between models (model comparison)

We next seek to determine which combinations of parameters best explain the differences between remission and relapse. For each model we look at the parameter combinations that give rise to either outcome; this way we propose the likely mechanisms which result in the different progressions. Linear discriminant analysis provides a convenient way to do this by finding a linear combination of parameters that best separates the two outcomes. This is similar to PCA (used in Chapter 5) but is appropriate when the variable of interest is discrete rather than continuous. The results of this analysis are summarised in Figure 6.6. For both models $M1$ and $M3$, a very good separation was found; it is interesting to note this separability since each individual posterior distribution is broad.

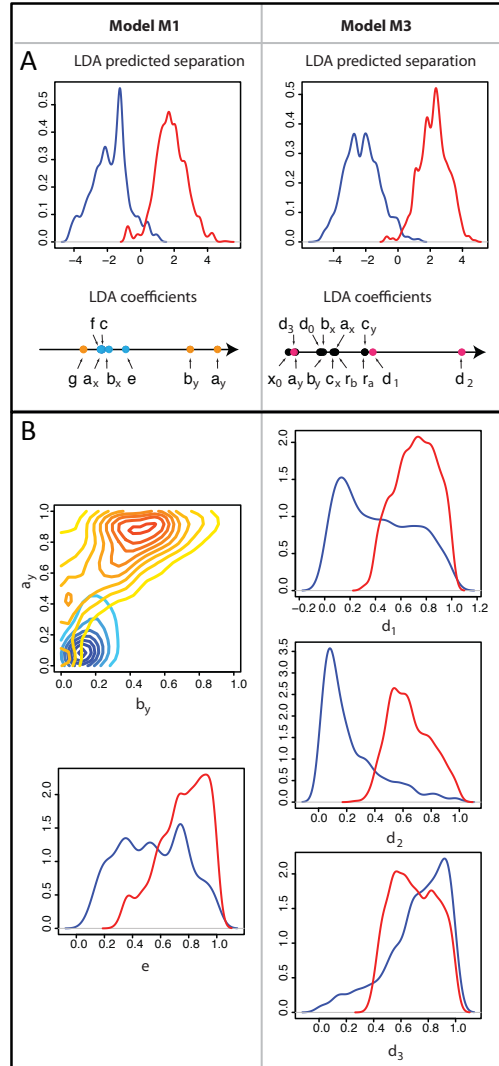


Figure 6.6: *Linear discriminant analysis (LDA) predicts differences between models in how they control disease outcomes. (A)* For each model, clear separation between disease outcomes can be made (histograms, top). The parameter contributions (coefficients) to the linear discriminant are shown on a horizontal axis. For model *M1*, the healthy parameters are represented by blue dots and the leukaemia parameters by orange dots; for model *M3*, the pink dots designate the death rates. **(B)** Shown for *M1* is a density plot of a_y against b_y with the location of remission outcomes marked in blue and the relapse outcomes in red: using only these two parameters a good separation is already achieved. We also show the posterior density distribution of parameter e alone in case of remission (blue) and relapse (red) – moderate separability is observed. For *M3* we show the posterior density distribution for each of the death rates and we see that for d_2 , remission (blue) and relapse (red) can be well distinguished.

For model $M1$ we find that parameters a_y , b_y and e explain most of the differences seen between disease outcomes. The pair of parameters contributing the largest amount to the linear discriminant, (a_y, b_y) , describe growth and differentiation of LSCs and production of differentiated leukaemia cells. In the case of remission, both these parameters have to be sufficiently low for leukaemia cell production to be kept in check, whereas for relapse $M1$ requires $a_y > b_y$, so that growth of LSCs can be maintained. Parameter e , whilst not constrained for remission, should be relatively high for relapse. This is required to produce the sharp peak in healthy differentiated cells. Overall for $M1$, we can say that in determining outcome, controlling leukaemia species is more important than controlling healthy species.

This leads us to predict that disease outcome could be controlled by variation only in parameters related to LSC function. To test this we simulate model $M1$ under relapse conditions until the onset of disease relapse and study the effect of varying a_y . We see that by decreasing a_y while keeping all other parameters constant, we can stop relapse from occurring and the patient goes back into remission, shown in Figure 6.7A. This intervention – selectively suppressing the growth of LSCs – thus seems promising, but is probably not yet possible to achieve clinically.

Performing a similar analysis for model $M3$, we see that a single parameter, d_2 , explains most of the variation between outcomes (see Figure 6.6). Relapse requires all of the differentiation/death rates to be high in order to produce the desired peak in healthy differentiated cells from peaks in the progenitor cell species. Contrastingly, for remission, d_2 should be moderate

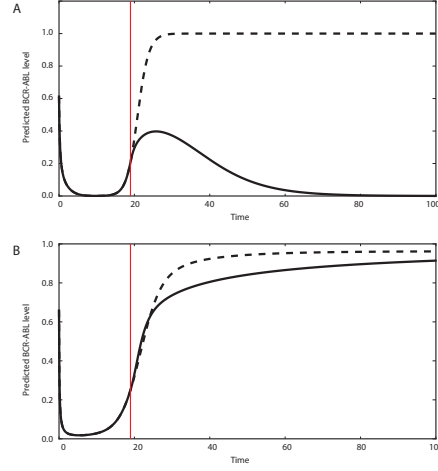


Figure 6.7: *In silico* interventions predict that for $M1$ (but not $M3$) relapse can be avoided. A. Response of model $M1$ to treatment intervention. Model is simulated using the maximum a posteriori parameter for relapse until 19 months, at which point the LSC growth parameter (a_y) is decreased from 0.92 to 0.07 and the model predicts that the patient returns to remission (bold line). The dashed line represents the predicted level of BCR-ABL without intervention. B. Response of model $M3$ to treatment intervention. Model is simulated using the maximum a posteriori parameter for relapse until 19 months, at which point the progenitor cell death parameter (δ_2) is decreased from 0.47 to 0.01. In this case the predicted time until relapse increases, but the outcome is not changed. As above, the bold line represents the intervention and the dashed line represents original relapse parameters.

or low. We observe that non-zero populations of partially differentiated cells (x_2, y_2) must be maintained (explaining the low value of d_2) so that fully differentiated cells (x_3, y_3) can be produced under the control of parameters c_x and c_y .

The parameters that are identified as important for discrimination here all relate to the more differentiated cell species. This is not surprising: the data describe only dynamics at the differentiated cell level and not

the stem cell level, so we would expect parameters controlling stem cell population dynamics to be less informative here. It is also worth considering the direct impact that the different model constraints will have on the inferred mechanisms of the models. Model *M1* does not allow stem or progenitor cell death, thus restricting the possible behaviours observed for these species. Model *M2*, on the other hand, allows cell death but couples the death rates for healthy-leukaemia species pairs that are at the same hierarchical level. This limits the possible dynamics of these pairs of species and constrains their behaviour.

6.5.5 Minimal descriptions of models *M1* & *M3* (model reduction)

In light of these differences in system behaviour, the comparable quantitative agreement between model and data (as revealed by Bayesian model selection) and the absence of discriminatory data, we need to consider models *M1* and *M3* from yet a different angle: robustness to the specification of the model. Model *M1* contains five species and model *M3* contains eight; in any case this is a vast oversimplification of the true extent of the haematopoietic system. Here we consider the effects of reducing the number of parameters and the number of species within each model.

In model *M1* there are three species in the healthy and two in the leukaemia lineage. Healthy progenitors are included because we know that the haematopoietic system contains a number of layers between stem and differentiated cells and we do not want to exclude these intermediate layers

altogether. In the case of leukaemia, there is evidence to suggest that at least a two-layered hierarchy exists, but not necessarily that intermediate layers exist, so only two layers have been included here. A model with three leukaemia species has been analysed previously and the qualitative changes introduced by the addition of this species were found to be negligible and do not impact on the overall model behaviour.

Model $M3$ has four layers of cell species for both the healthy and leukaemia lineages, which Foo et al. (2009) describe as stem cells, progenitor cells, differentiated cells and terminally differentiated cells. Their justification for this model hierarchy is based on the characterisation of the haematopoietic lineage as described by Weissman and colleagues (Morrison et al., 1995; Spangrude et al., 1988). Haematopoietic species do make up a multilevel hierarchy, but whether this number of species is required in a model used to analyse data for only the terminally differentiated cell population is questionable.

To test this, we ran a similar posterior comparison analysis to that performed above for reduced versions of $M3$ with only two or three layers in the hierarchy. Once again, rather than directly study the posterior distributions obtained, we investigate trajectories simulated for parameters sampled from the posterior. The results of this are shown in Figure 6.8.

Given the three variations we now have for model $M3$ (for both remission and relapse cases), we can compare them and we see that the trajectories obtained for the original and the reduced models behave similarly. For remission we see that trajectories for the three- and four-layer models are very

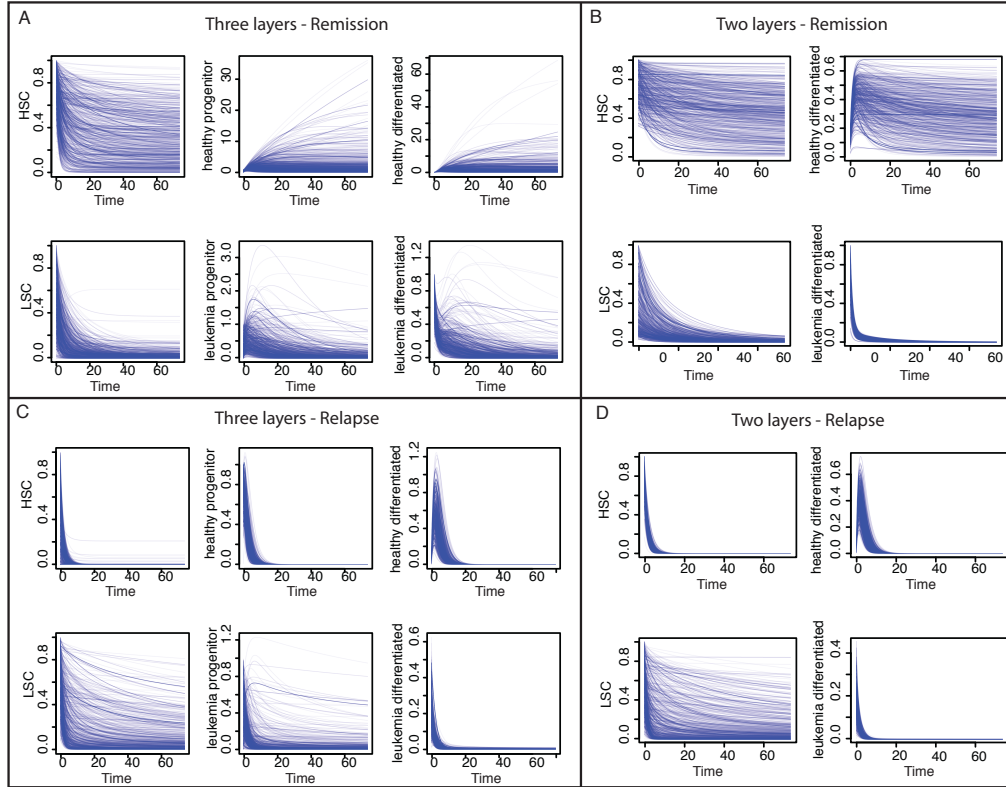


Figure 6.8: *Dynamics of model species for reduced-form M models.* Predicted evolution of each species for both outcomes (remission in **A-B**, relapse in **C-D**) under two reduced versions of model $M3$: three-layer (**A** & **C**) and two-layer (**B** & **D**) hierarchies, respectively. For each model and each outcome, 1000 parameter sets have been sampled from the posterior distribution and each line corresponds to the simulated trajectory for one of these parameter sets.

similar, so we could omit one layer without changing the model behaviour. In going from three to two layers we do observe more pronounced changes: now the dynamics of differentiated leukaemia cells are constrained, as are the dynamics of their parent (LSC) population (Figure 6.8B). These trajectories show similarities to those obtained for model *M1* after calibrating it to the remission data. The shared feature between models *M1* and two-layer *M3* is that the niche exerts its influence over all non-differentiated cells; this might explain the similarities seen here. The remaining difference between these models is the number of niche competitor cells: all species in *M1* vs. only stem cell species in two-layer *M3* (see Figure 6.1). We thus suggest that this could be a robust feature of niche competition models.

For relapse, the healthy and the leukaemia differentiated cells follow restricted trajectories in all versions of model *M3*. The populations of cells giving rise to fully differentiated healthy cells in each case (which can be progenitors or stem cells) are now more tightly constrained by the data. The parent population to differentiated leukaemia cells in each case can exhibit a range of behaviours without such restriction.

In summary, this analysis demonstrates that a hierarchy of four layers is not necessary to capture the observed CML dynamics within the context of model (or model family) *M3*. The main difference between models *M1* and *M3* is the level to which cells in the haematopoietic system experience the influence of the haematopoietic stem cell niche. The version of model *M3* with a two layer hierarchy thus naturally resembles the dynamics of model *M1*, where differentiation occurs inside the niche. To resolve the problem

of discriminating between models $M1$ and $M3$ we thus need to be able to probe the haematopoietic stem cell niche directly.

6.6 Conclusion

CML has been studied in great detail, and a number of successful models describing different aspects of disease progression have been developed in the past decade (Colijn & Mackey, 2005; Foo et al., 2009; Kim et al., 2008; MacLean et al., 2013; Michor et al., 2005; Moore & Li, 2004; Roeder et al., 2006; Werner et al., 2011). However, there remains uncertainty about the cellular processes that control disease aetiology and outcome. Here we have approached this question by comparing three models of CML dynamics and their ability to explain clinical data on CML progression.

We used a Bayesian inferential framework and found that all of the models can be fitted to data for both remission and relapse and receive the same level of statistical support. Because of the limitations of the data we thus have to adopt more detailed model analysis and model checking tools in order to discriminate between these different mechanistic hypotheses.

Model M2, for example, can be ruled out as it requires exponential and sustained growth of LSCs in order to explain the available data. The key difference between this and the other two models is that model M2 does not allow for interactions in the stem cell niche but assumed that all cell-types evolve independently. This suggests that cellular competition is an important and perhaps crucial aspect of a model seeking to describe the

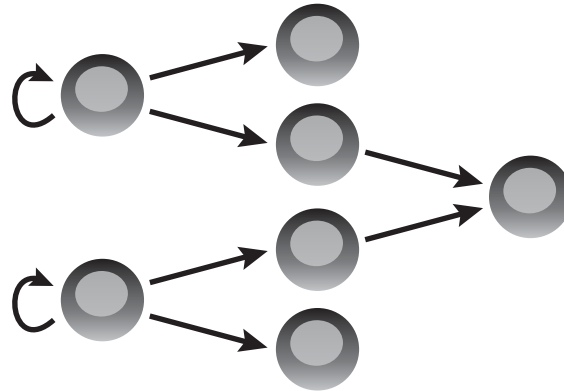
ecology of haematopoietic and leukaemia cells inside CML patients.

Despite the limited nature of the data our Bayesian framework allowed us to analyse the remaining models in more detail, and we used simulations from the posteriors to explore the behaviour of the different models in patients exhibiting relapse and remission. Posterior distributions, even though frequently broad, nevertheless can frequently give rise to constrained behaviour of the different species modelled here. Crucially, our analysis has allowed us to show the effects that the niche can exert on disease progression: as more cell types — HSCs and LSCs reside necessarily inside the niche micro-environment — are exposed or contribute to the cell population dynamics inside the niche, model M3 increasingly resembles model M1. If the niche does indeed extend its influence also over progenitor cells then our analysis suggests that clinical outcome is determined by how effectively therapy targets LSCs. In M3 by contrast we found that the differences between patients showing remission or relapse are less pronounced, with the death rate of progenitor cells being the parameter exerting most influence over a patient's prognosis.

Haematopoiesis is a model system and CML is a model disease. Despite the large body of knowledge gathered, and despite the successful treatments for CML, there remain large gaps in our understanding of the cellular processes that underpin this disease. Here we have identified mechanistic differences at the cellular level between different hypotheses of CML progression that we hope will become experimentally testable as methods for probing the HSC niche advance. The ability to map the haematopoietic system and

its dynamics *in vivo* will almost certainly also benefit our understanding of other haematological disorders.

Seven



Conclusions and future work

In this thesis we have used a variety of theoretical approaches to attempt to learn more about the interactions between haematopoietic stem cells (HSCs), the niche in which they reside, and cancer cells that can perturb this system. Here we tie together the conclusions that we reached from each chapter. We discuss briefly their short and longer term implications. We also describe what we believe to be some of the most important or exciting future research goals to pursue both theoretically and experimentally.

In Chapter 4 we studied stability properties of simple models of stem cell dynamics under a new framework. We identified the ways by which systems

that are more realistic – in the sense that their interactions are specified by an underlying model, rather than at random – gain stability. This occurs primarily through heterogeneity in the types of interaction between species, and, to a lesser extent, through dependencies between interactions. We studied a family of models that differed in their number of lineage branches and in their treatment of feedback. Our probabilistic framework for stability analysis demonstrated that the fixed points of each model were always or almost always stable for those fixed points with biologically interesting (i.e. positive) values. This shows us that in the case of these models, and also, we believe, more generally, stability depends less on the particular type and strength of interactions and more on whether or not these interactions are drawn from a distribution based upon a structured model.

From analysis of one model where two fixed points could be reached, we saw how different parameter regimes characterise the two fixed points. Prospects based on both mathematical and biological extensions of this analysis are exciting. Mathematically, we move towards performing a global multivariate bifurcation analysis, by characterising the different fixed points in parameter space and delineating the boundaries between them. Biologically, we aspire to understand (and even control) how cell lineage branching decisions are made. This is a desirable goal both for fundamental biology and for designing new strategies to overcome disease such as leukaemia, that results from misregulated cell growth within a certain branch of the haematopoietic hierarchy. Another extension to the stability framework that we use here will be to extend it so that numerical simulation yields

similar results for the probabilities of stability in cases where analytical expressions for the fixed points cannot be found.

In Chapter 4 we learned how stem cell models can be stable through general principles that are not specific to the model. We also began to understand how cell branching processes can occur. These models provide the inspiration for the models developed subsequently, that build upon these simple versions and consider additional process.

In Chapter 5 we introduce two new models that describe stem cell dynamics within the niche in the case where leukaemia is also present. These models treat the competition between healthy and diseased lineages in different ways and we seek to infer desirable conditions within the niche, that is, we study how leukaemia can be out-competed and healthy haematopoiesis restored.

These two models are extensions of the models of Chapter 4, describing the cellular processes that govern stem cell function. They also incorporate concepts from ecological modelling, specifically Lotka-Volterra-type competition between species and mathematical treatment of the stem cell niche. Due to different – and sometimes conflicting – definitions of the HSC niche, we define it functionally rather than spatially; this allows us to account for uncertain or as yet unknown niche effects in our framework.

We fit these models to a qualitative outcome that requires the system to maintain a viable healthy haematopoietic cell population that can drive out the leukaemia cell species from the niche. In doing so we find that

the most important parameters to achieve this correspond to growth and maintenance of healthy cells – not the direct killing of cancer cells. We thus propose that if it were possible to maintain a sufficient healthy HSC population this alone could already be almost sufficient to overcome leukaemia. We also suggest that focus on healthy HSC maintenance ought to become a central aspect of any therapeutic regime. Both models share this finding. We see that the second, more complicated model does not add extra insight with regards to the questions that we address here, so we do not proceed further with analysis of this model.

In Chapter 6 we continue to study questions of competition between healthy haematopoietic cell species and leukaemia. We take the simpler model from Chapter 5, introduce two new models from the literature and compare these three models against clinical data for patients undergoing treatment for chronic myeloid leukaemia. In doing so, we attempt to answer a distinct but closely related question compared to the one of the previous chapter. Where before we sought to understand how (theoretically) leukaemia could best be eradicated from the HSC niche, here we want to know whether different models can explain the data both of disease remission (moving towards complete leukaemia eradication) and of disease relapse (a return of disease following from initially successful therapy).

We discard one of the models because although it provides an equally good fit to the data as the others, it does so only by unrealistic leukaemia stem cell growth dynamics (unbounded exponential growth). Since this model is the only one not to include explicitly the competition between healthy

and leukaemia stem cells, we are led to suggest that this competition is an important and perhaps crucial feature of models that seek to describe the dynamics of chronic myeloid leukaemia.

Proceeding to analyse the two remaining models in light of the data and with attention to the ecological description of cell interactions that they provide, we find that each makes distinct predictions regarding disease progression. One predicts that once treatment has begun there will be a transient increase in the number of healthy progenitor cells. The same model predicts that controlling the growth of leukaemia stem cells is the most important factor in determining disease outcome (remission or relapse). The other model, in contrast, predicts that controlling the death rate of progenitor cells is most important in determining outcome. Testing these predictions should enable us to invalidate one of these models, and perhaps provide support for the other.

In addition to the experimental studies we hope to prompt from these results, we also highlight more and less important features of models of HSCs. One of the more important features is a description of the competition between cellular species. One of the less important features is the addition of extra progenitor cell populations in the hierarchy – especially given that we do not currently have data describing these populations. We thus suggest that wherever possible, models with fewer subpopulations should be used until we are able to gather *in vivo* human data for the healthy and leukaemia stem and progenitor cell species, not only in the blood but also in the bone marrow.

One exciting direction for future study is to extend the modelling of HSCs using ordinary differential equations (ODEs) to include detailed spatial information, such as using cellular automata. We include an outline of possible ways this could be accomplished in Appendix B.

Despite so much activity, there remains much to be understood about HSCs. Here, their dynamics have been studied from a variety of angles. We have made new predictions that can be tested experimentally and we have raised new questions that we hope will advance the discussion on HSCs, their niche and leukaemia.

Appendix A: Parameter posterior distributions

Here we present the marginal posterior distributions for the parameters of models that are inferred using approximate Bayesian computation. Parts of these distributions were presented in the results of chapters 5 and 6; they are given in full here for completeness.

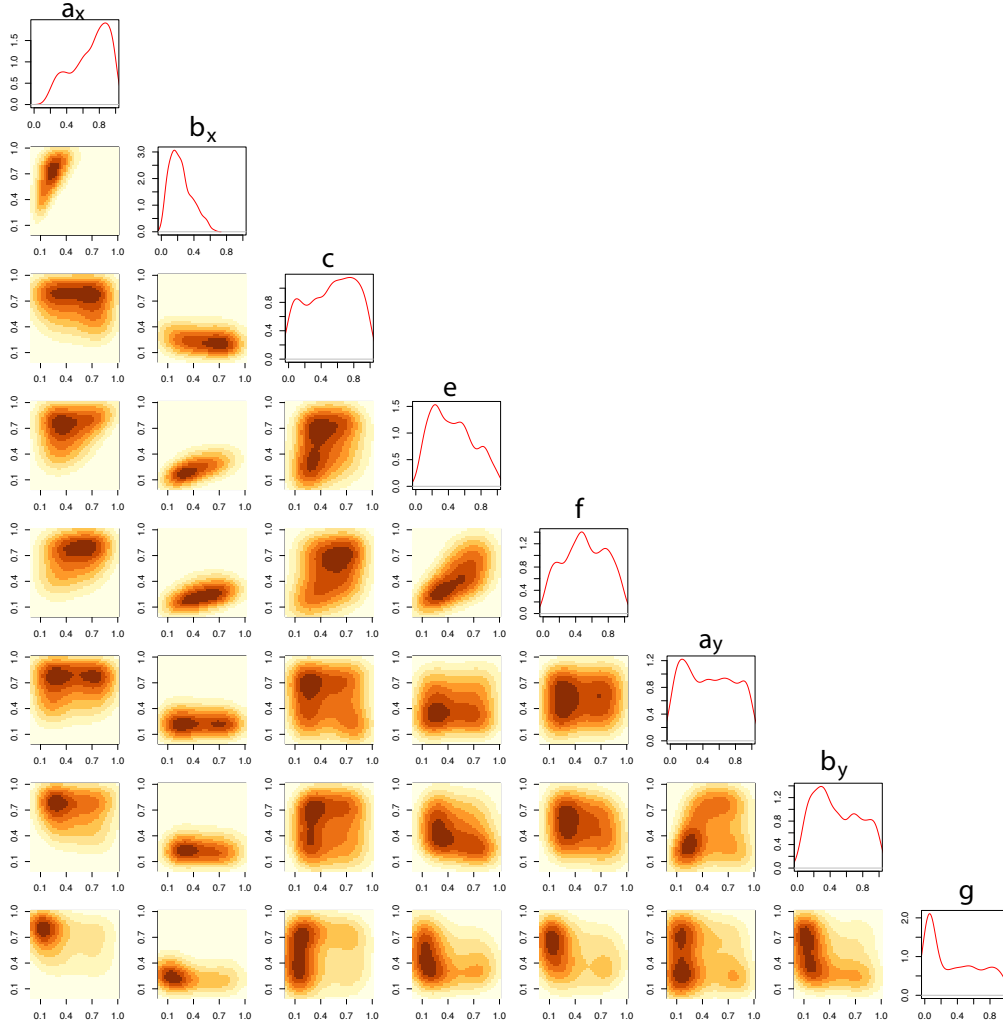


Figure A.1: *Posterior distribution for model A.* Based on fit to qualitative outcome: leukaemia eradication and healthy haematopoietic cell species survival. Marginal posterior distributions for each parameter are shown along the diagonal and surface density plots for each pair of parameters are given as heat maps.

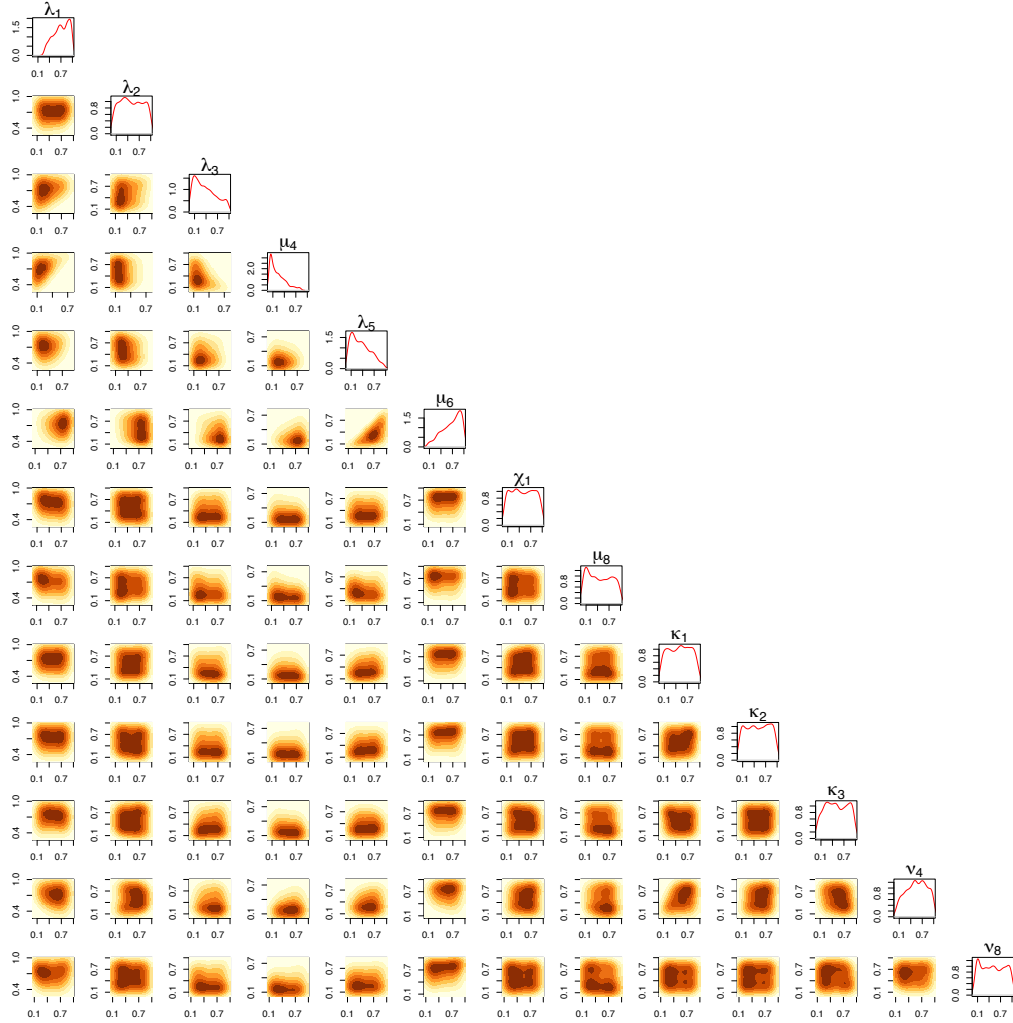


Figure A.2: *Posterior distribution for model B.* Based on fit to qualitative outcome: leukaemia eradication and healthy haematopoietic cell species survival. Marginal posterior distributions for each parameter are shown along the diagonal and surface density plots for each pair of parameters are given as heat maps.

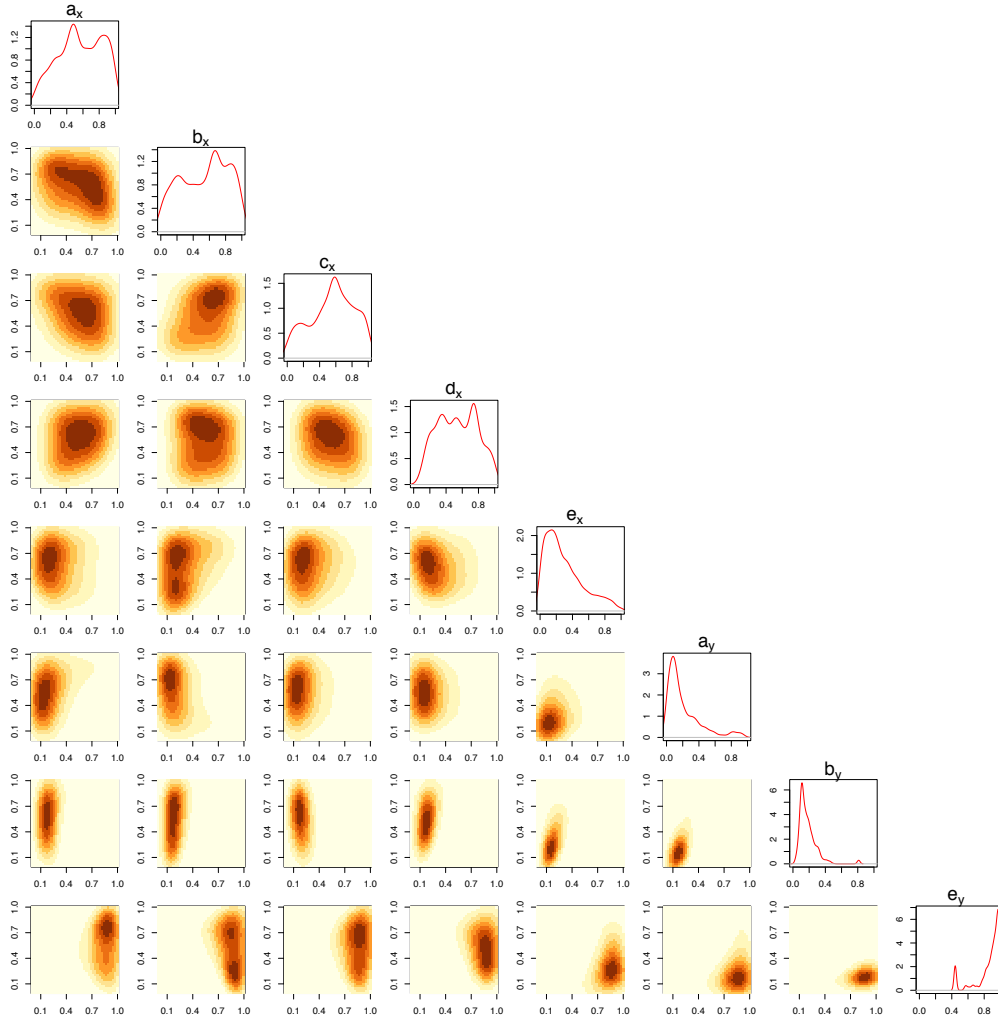


Figure A.3: *Posterior distribution for model M1 under remission condition.* Marginal posterior distributions for each parameter are shown along the diagonal and surface density plots for each pair of parameters are given as heat maps.

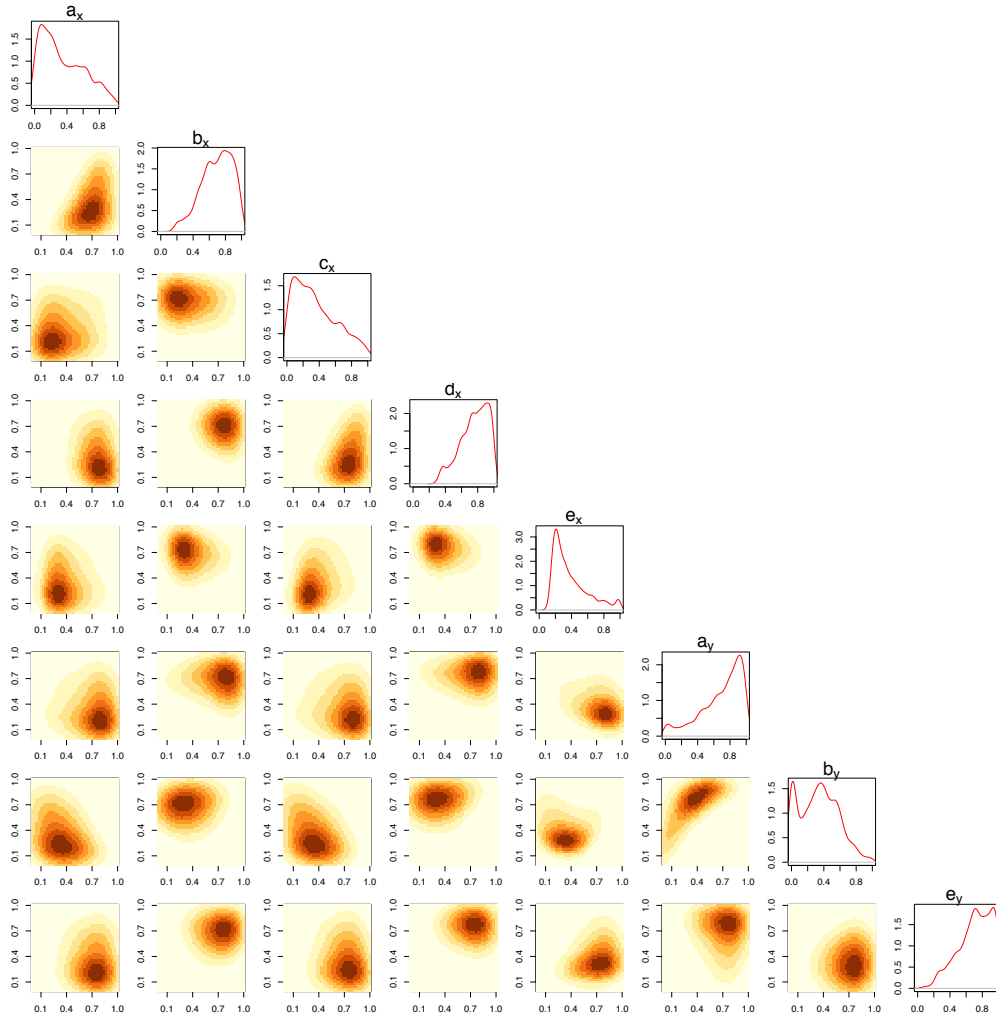


Figure A.4: *Posterior distribution for model M1 under relapse condition.* Marginal posterior distributions for each parameter are shown along the diagonal and surface density plots for each pair of parameters are given as heat maps.

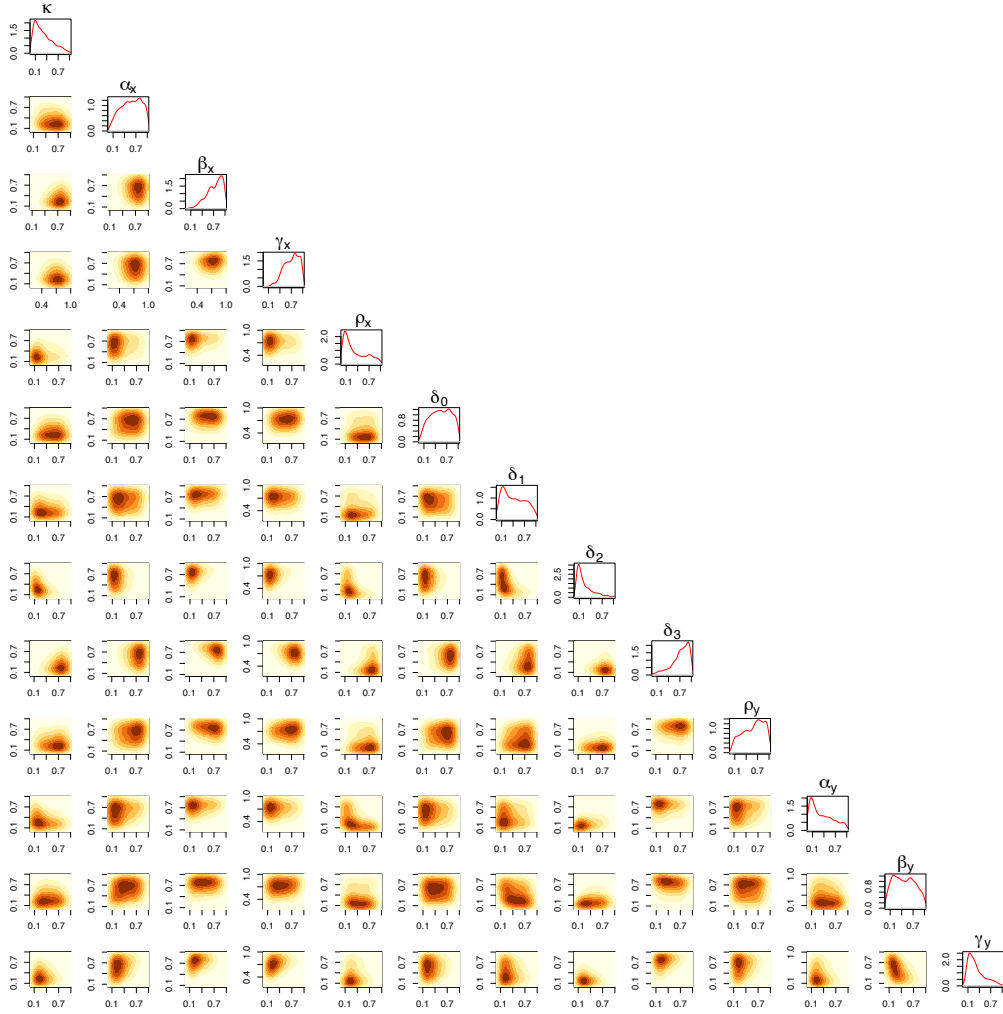


Figure A.5: *Posterior distribution for model M3 under remission condition.* Marginal posterior distributions for each parameter are shown along the diagonal and surface density plots for each pair of parameters are given as heat maps.

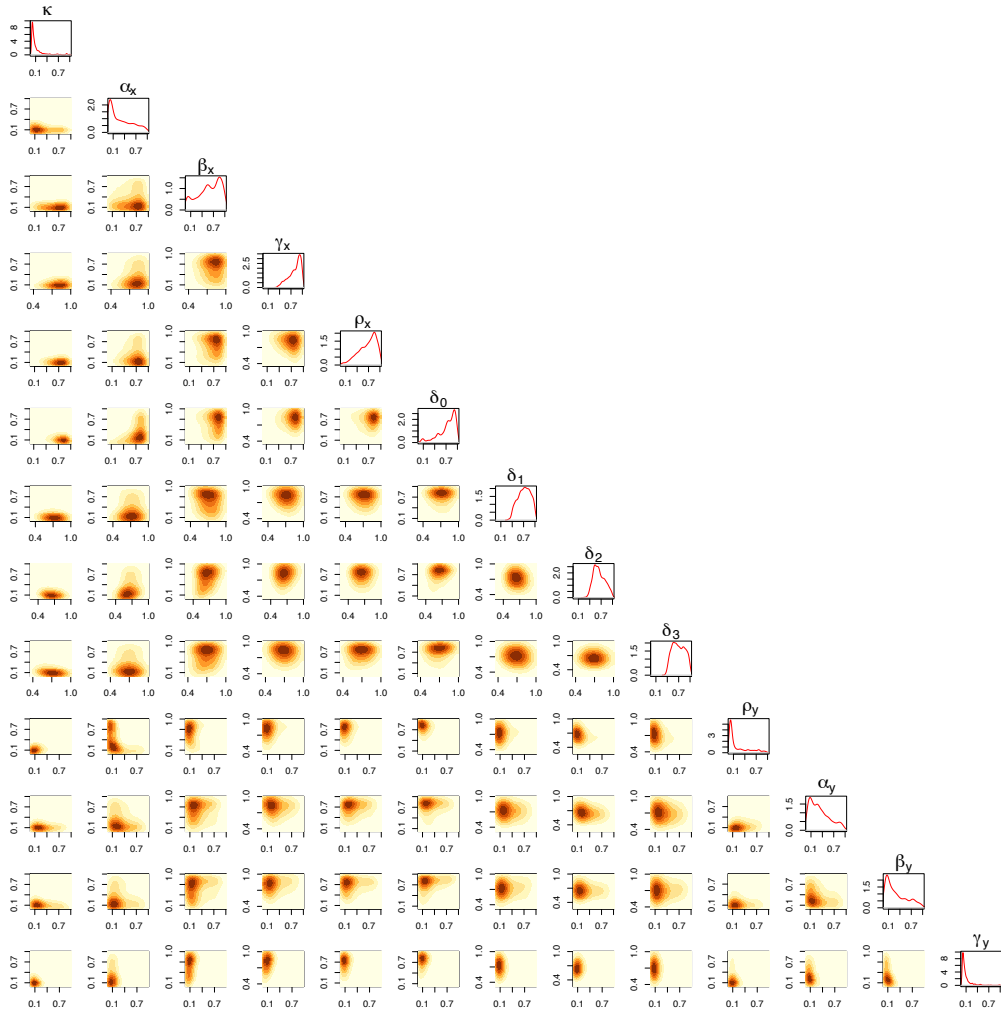


Figure A.6: *Posterior distribution for model M3 under relapse condition.* Marginal posterior distributions for each parameter are shown along the diagonal and surface density plots for each pair of parameters are given as heat maps.

Appendix B: Spatial models of stem cell dynamics with cellular automata

B.1 Overview

Cellular automata (CA) offer a stochastic, spatial, grid-based modelling framework and their application to modelling biological systems has many appealing characteristics. The development of CA to investigate stem cell niche dynamics and competition is beyond the scope of this thesis but does offer a very attractive avenue for future work. Here, we will briefly discuss two different CA pertinent to the biological problems that we have been investigating. We will demonstrate how these can be developed in future work and outline their potential advantages over deterministic (ODE) models and their limitations.

B.2 Introduction

In Chapter 3 we introduced CA in our overview of modelling techniques. We will not repeat that introduction here but give a brief reminder.

CA were initially developed by von Neumann (1966), Wolfram (1983) and others and describe the evolution of one or more types of agent on a grid of lattice points. (These agents are usually called cells but we avoid this name since they bear no relation to biological cells.) The grid defines the model's behavioural space; we consider 2D grids here but these methods are easily extendible to 3D, given appropriate boundary conditions and sufficient computational resources.

The dynamics of CA are controlled by an update rule. Update rules can be deterministic or stochastic as well as synchronous or asynchronous. here we consider asynchronous stochastic update rules, which we think resemble most closely the systems that we attempt to describe. Early work by John Conway and others did use deterministic update rules to study the evolution of CA that mimicked "life" in the sense that agents were governed by self replication and death (Sigmund, 1993).

Consider the grid shown in Figure B.1, as a small portion extracted from a larger grid. We construct an update rule based on the 8 nearest neighbours of the central grid point, thus only these are shown. In this example we have only one type of agent (red) and empty grid points are white. Based on the number of neighbours, the agent can either die, replicate, or remain in its place. These three options are shown in Figure B.1.

The rule used in Figure B.1 is the same as was presented as an example in Chapter 3, namely,

1. if surrounded by less than four cells, divide, creating a new cell at a

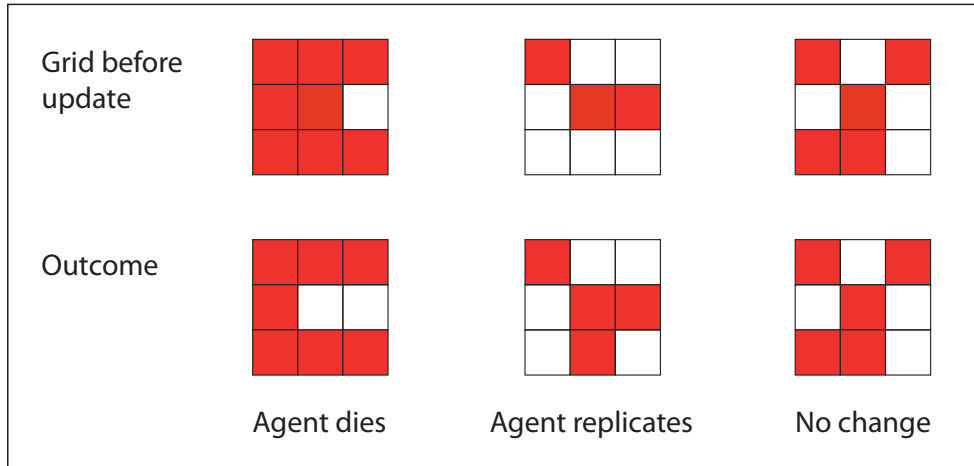


Figure B.1: *Update rule for a CA based on nearest neighbours on a 2D grid.* Agents are shown in red and white denotes empty spaces. The focal agent is in each case the one in the centre of the grid, surrounded by its 8 nearest neighbours. The top row shows three example grids that each lead to a different outcome for the focal agent. The bottom row shows the same grid after the update rule has been applied. See main text for a description of the update rule.

free neighbouring grid point,

2. if surrounded by more than six cells, die
3. else, do nothing.

By defining rules that as best as possible represent the underlying biology, and then simulating a model over some time period in order to observe how the population of agents evolves, we can make new predictions about the biological systems under investigation.

The agents in CA can correspond to many different phenomena. In biolog-

ical modelling, they often represent cells of parts of cells. Here we consider bijective CA where one agent = one cell; many other variations are possible.

In the next section we present two CA that address different questions regarding the spatial organisation of stem cell populations.

B.3 Applications of cellular automata

B.3.1 A model of loss of stem cell self renewal

Haematopoietic stem cells (HSCs) gradually lose their self renewal capacity through either activation from a quiescent state or differentiation towards specialised cells through multiple multipotent progenitor cell populations (Wang & Wagers, 2011). It has been shown that spatial aspects affect HSC self renewal (Kiel et al., 2005; Lo Celso et al., 2009a; Wilson et al., 2007), however, there remains much to be understood about the exact processes of how self renewal capacity is lost (Wilson & Trumpp, 2006). Here, we describe a CA that could be used to study loss of self renewal capacity along an axis.

Shown in Figure B.2 is an example simulation of this model. Along the abscissa is the cell position relative to the niche ‘wall’ – at 0 on this axis; along the ordinate is shown spatial HSC occupation of the niche. We make the assumption that as HSCs move farther away from the wall, they will gradually lose their self renewal capacity. We seek to model how this might occur using different functions to describe the strength of interaction between the niche and HSCs.

The model can be simulated by imposing an update rule that allows HSCs either to self renew or migrate away from the niche wall. In this example dedifferentiation (HSCs moving to the left on the grid) is not allowed. The update rule can be summarised as follows: if the niche strength (a decreasing function along the abscissa) is above a threshold, the cell will not migrate and divide with some probability, as long as there is free space to do so. If this niche strength is below the threshold then the cell will migrate to a free space one grid point further away from the niche wall, as long as there is space to do so.

In the snapshots of CA evolution shown in Figure B.2, we see an illustration of how the system might evolve. Depending on the definition of the niche strength function, we might find that all the HSCs eventually leave the niche wall begin to lose their self renewal capacity. Or instead, perhaps the strength of the niche will be enough to maintain the majority of HSCs at maximal self renewal capacity.

B.3.2 A competition model for healthy versus cancer cells

In this application, we consider a competition model that shares characteristics with the models studied in chapters 5 and 6, but with a description of how cell location influences competition. This model can be used to study how cancer arises in healthy tissue and to study requirements for cancer to invade successfully rather than being outcompeted by healthy cells. We can also look at the differences in outcome between competition initiated

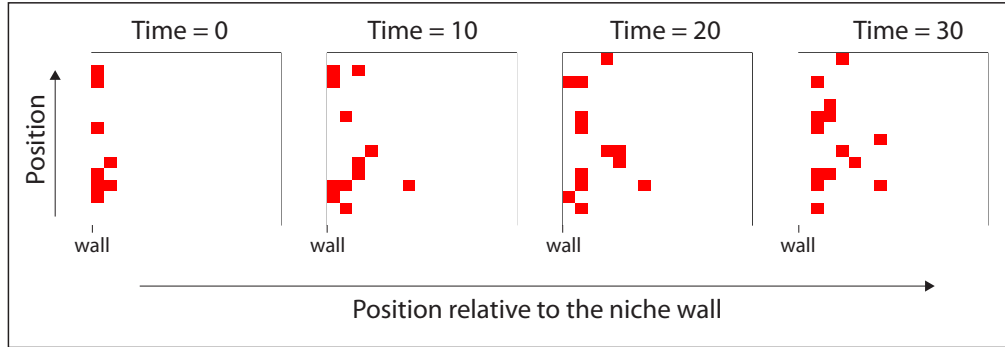


Figure B.2: *Simulation of HSC loss of self renewal dynamics.* In this CA the red agents are HSCs and the model describes how the HSCs lose their self renewal capacity as they move away from the niche wall (the ordinate). The update rule for HSCs encompassing cell replication and migration is described in the text.

by a single cancer cell, and that initiated by a small group of cancer cells.

In this simple illustration we do not consider to which particular tissue type the healthy and cancer cells belong: we refer to them generally. Each cell type has its own update rule that depends on its neighbours. These neighbours can now be one of two different cell types. We consider the case where both cell types are growing surrounded by free space (analogous to cells in a growth medium).

Figure B.3 shows an example simulation of the competition between cell types. The update rule that governs this simulation can be summarised by the general points:

- healthy and cancer cells surrounded by a small number of other cells can divide,
- healthy cells surrounded by roughly equal numbers of neighbouring cells and free spaces can migrate,
- cancer cells surrounded by equal or greater numbers of neighbouring cells than free spaces can migrate.

In this model no cells die. By giving cancer cells a migratory advantage over healthy cells, we can study the competition dynamics and discover (for example) the time taken until the cancer clone dominates the tissue.

Initial conditions were chosen such that a very small population of cancer cells arise amidst healthy cells, shown at time = 0 in Figure B.3. The system is then simulated with an asynchronous update rule for a specified time period. Snapshots of the populations of healthy and cancer cells are shown for three subsequent time points. From Figure B.3 we can see both populations growing alongside one another and that the cancer population quickly starts to overcome the healthy cells, due to its greater migratory capacity.

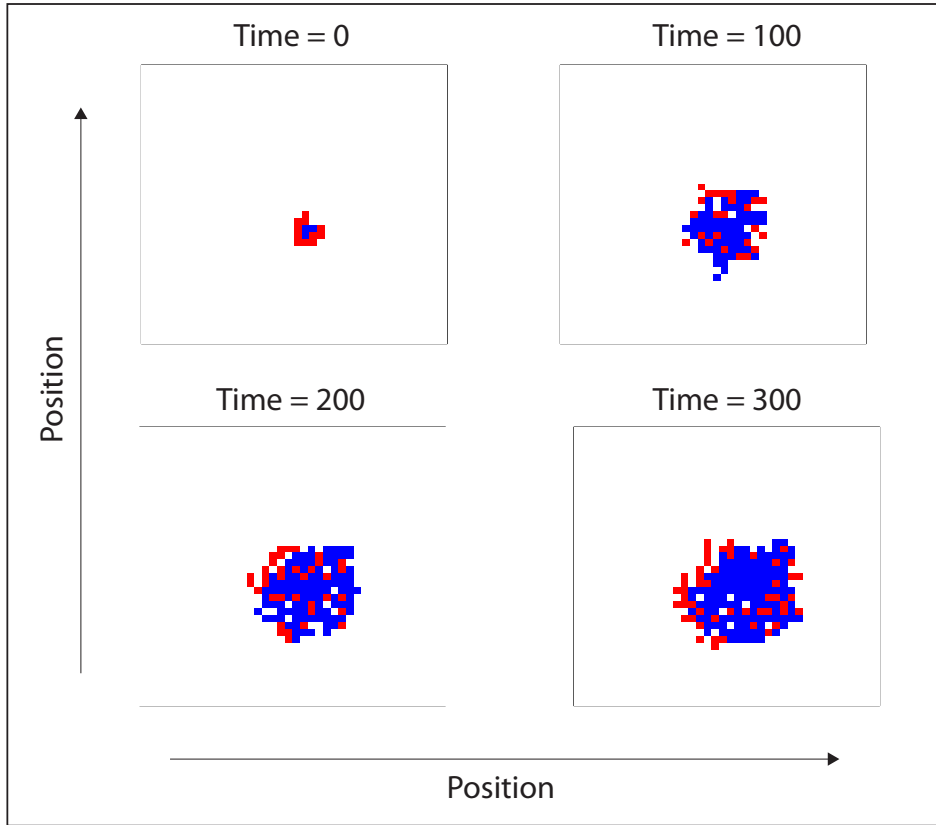


Figure B.3: *Simulation of spatial competition model for healthy and cancer cells.* Healthy cells are red, cancer cells are blue, and free space is white. Initial conditions were chosen such that a small population of cancer cells is surrounded by healthy cells. The number and location of healthy and cancer cells are then shown for three subsequent time points.

B.4 Discussion and Conclusions

The simple examples of CA given here serve as an illustration of the possibilities such models offer in attempting to understand biological systems. Their intuitive design and relatively simple implementation make them readily accessible to modellers who wish to perform grid-based simulations.

The myriad of subtle ways that CA can be modified represents both an advantage and a drawback of these methods. Their flexibility means that many different questions can be addressed by making changes to model design. On the other hand, this flexibility can be limiting because in many cases it might not be clear how the model should be constrained in order to best reflect the biology under investigation. This might bring in an arbitrary element to model design. For example, in the example rule given in the introduction, why should the threshold for cell death be 7 neighbours, not 6 or 8? However, as data resolution improves, these challenges may become surmountable.

Simulation of CA quickly becomes very expensive as the grid size or dimension increases, or the update rules become more sophisticated. This limits current applications to real biological systems, which are many orders of magnitude greater than the grid sizes most often used in CA modelling.

Despite the challenges, performing parameter inference with agent-based models and fitting them to spatial data, either *in vitro* or *in vivo*, is becoming possible, an example is given by Sottoriva & Tavare (2010). The challenges met in performing such inference are both methodological and

biological. Methodologically we need ways to reduce computational cost, probably through parallelisation, as well as measures of fit of model to data. Biologically, we need spatiotemporal data of sufficient resolution to perform the inference. More and more relevant data are becoming available (Flesken-Nikitin et al., 2013; Lane et al., 2011; Lo Celso et al., 2011; Méndez-Ferrer et al., 2010; Simons & Clevers, 2011; Zhu et al., 2013), although automating image analysis pipelines still represents a significant challenge.

CA are just one member of a family of agent-based models. We have focused on them here because we think they hold exciting potential for future biological modelling, but other types of model might also be appropriate for certain problems. For example, Cellular Potts models are a generalised version of CA that include an energy function within the update rule that is incorporated into the cellular dynamics (Scianna & Preziosi, 2013).

Hybrid modelling represents another attractive avenue for future research. Here, agent-based models such as CA are combined with ODEs or PDEs that describe continuous properties of the system. The additional flexibility that this combination offers can be advantageous to the modeller.

As a motivating example, consider the first CA application introduced above, where HSC agents lose self-renewal capacity as they move away from the niche wall. We might want to extend this example to include an evolving population of niche cells, such as osteoblasts, that control the niche strength. To do so we could use an ODE to model the size of the osteoblast population (that depends on the size of the HSC population), and calculate the niche strength (and thus update rule for HSCs) based on the osteoblast

population trajectory. This assumes that there are many more osteoblasts than HSCs, in most cases this will be a reasonable assumption.

In the above, we do not assume a spatial component to the osteoblast dynamics, this could be included if we were to use a PDE rather than an ODE to model osteoblasts and used, for example, a diffusion term to describe movement of osteoblasts.

The main advantage offered here is that simulation times will be much faster than if osteoblasts were to be treated as agents. Such an approach may be appropriate in many cases, where there are different molecular/cellular species numbers, or different timescales or length scales. Hybrid methods thus offer the chance to include effects on different scales, while maintaining reasonable efficiency. In order to successfully implement such a model, it must be ensured that the stochastic and deterministic components interact correctly – this requires a suitable choice of timestep between each update. It is also necessary to carefully consider the new interpretation of the system that this hybrid model will offer. That is, to consider under what conditions the division of model components into stochastic and deterministic is appropriate, and when these assumptions might break down.

In this chapter we have outlined two ways in which CA can be used to model biological systems, included because although they are outside the main scope of this thesis they represent exciting possibilities for future research. We hope that this outline may serve as a starting point for future work, and look forward to discovering where such ventures may lead.

Bibliography

- Aguila, H. L., & Rowe, D. W. (2005). Skeletal development, bone remodeling, and hematopoiesis. *Immunological reviews*, 208, 7–18.
- Aïnseba, B., & Benosman, C. (2011). Global dynamics of hematopoietic stem cells and differentiated cells in a chronic myeloid leukemia model. *Journal of Mathematical Biology*, 62(6), 975–997.
- Alarcón, T., Byrne, H. M., & Maini, P. K. (2003). A cellular automaton model for tumour growth in inhomogeneous environment. *Journal of Theoretical Biology*, 225(2), 257–274.
- Alexandersson, M. (2000). An application of general branching processes to a cell cycle model with two uncoupled subcycles and unequal cell division. *Int J Appl Math Comput Sci*, 10, 131–146.
- Allesina, S., & Tang, S. (2012). Stability criteria for complex ecosystems. *Nature*, 483(7388), 205–208.
- Austin, T. W., Solar, G. P., Ziegler, F. C., Liem, L., & Matthews, W. (1997). A role for the Wnt gene family in hematopoiesis: expansion of multilineage progenitor cells. *Blood*, 89(10), 3624–3635.

- Barnes, C., Silk, D., Sheng, X., & Stumpf, M. (2011). Bayesian design of synthetic biological systems. *Proc Natl Acad Sci U S A*, *108*, 15190–15195.
- Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate bayesian computation in population genetics. *Genetics*, *162*(4), 2025–2035.
- Bernt, K. M., & Armstrong, S. A. (2009). Leukemia stem cells and human acute lymphoblastic leukemia. *Seminars in hematology*, *46*(1), 33–38.
- Blank, U., Karlsson, G., & Karlsson, S. (2008). Signaling pathways governing stem-cell fate. *Blood*, *111*(2), 492–503.
- Boisset, J.-C., van Cappellen, W., Andrieu-Soler, C., Galjart, N., Dzierzak, E., & Robin, C. (2010). In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium. *Nature*, *464*(7285), 116–120.
- Bonnet, D., & Dick, J. E. (1997). Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nature Medicine*, *3*(7), 730–737.
- Calvi, L. M., Adams, G. B., Weibrecht, K. W., Weber, J. M., Olson, D. P., Knight, M. C., Martin, R. P., Schipani, E., Divieti, P., Bringhurst, F. R., Milner, L. A., Kronenberg, H. M., & Scadden, D. T. (2003). Osteoblastic cells regulate the haematopoietic stem cell niche. *Nature*, *425*(6960), 841–846.

- Chaidos, A., Barnes, C. P., Cowan, G., May, P. C., Melo, V., Hatjiharissi, E., Papaioannou, M., Harrington, H., Doolittle, H., Terpos, E., Dimopoulos, M., Abdalla, S., Yarranton, H., Naresh, K., Foroni, L., Reid, A., Rahemtulla, A., Stumpf, M. P. H., Roberts, I., & Karadimitris, A. (2012). Clinical drug resistance linked to inter-convertible phenotypic and functional states of tumor-propagating cells in multiple myeloma. *Blood*.
- Chickarmane, V., Enver, T., & Peterson, C. (2009). Computational modeling of the hematopoietic erythroid-myeloid switch reveals insights into cooperativity, priming, and irreversibility. *PLoS Computational Biology*, 5(1), e1000268.
- Colijn, C., & Mackey, M. C. (2005). A mathematical model of hematopoiesis—I. Periodic chronic myelogenous leukemia. *Journal of Theoretical Biology*, 237(2), 117–132.
- Colmone, A., Amorim, M., Pontier, A. L., Wang, S., Jablonski, E., & Sipkins, D. A. (2008). Leukemic cells create bone marrow niches that disrupt the behavior of normal hematopoietic progenitor cells. *Science*, 322(5909), 1861–1865.
- Cozzio, A., Passegué, E., Ayton, P. M., Karsunky, H., Cleary, M. L., & Weissman, I. L. (2003). Similar MLL-associated leukemias arising from self-renewing stem cells and short-lived myeloid progenitors. *Genes & Development*, 17(24), 3029–3035.
- Dalerba, P., Cho, R. W., & Clarke, M. F. (2007). Cancer Stem Cells: Models and Concepts. *Annual review of Medicine*, 58(1), 267–284.

- Daukste, L., Basse, B., Baguley, B. C., & Wall, D. J. N. (2009). Using a stem cell and progeny model to illustrate the relationship between cell cycle times of in vivo human tumour cell tissue populations, in vitro primary cultures and the cell lines derived from them. *Journal of Theoretical Biology*, 260(4), 563–571.
- Del Moral, P., Doucet, A., & Jasra, A. (2006). Sequential monte carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3), 411–436.
- Dick, J. E. (2008). Stem cell concepts renew cancer research. *Blood*, 112(13), 4793–4807.
- Ding, L., Saunders, T. L., Enikolopov, G., & Morrison, S. J. (2012). Endothelial and perivascular cells maintain haematopoietic stem cells. *Nature*, 481(7382), 457–462.
- Domer, P. H., Fakharzadeh, S. S., Chen, C. S., Jockel, J., Johansen, L., Silverman, G. A., Kersey, J. H., & Korsmeyer, S. J. (1993). Acute mixed-lineage leukemia t(4;11)(q21;q23) generates an MLL-AF4 fusion product. *Proceedings of the National Academy of Sciences of the United States of America*, 90(16), 7884–7888.
- Duff, C., Smith-Miles, K., Lopes, L., & Tian, T. (2012). Mathematical modelling of stem cell differentiation: the PU.1-GATA-1 interaction. *Journal of Mathematical Biology*, 64(3), 449–468.
- Duncan, A. W., Rattis, F. M., DiMascio, L. N., Congdon, K. L., Pazianos, G., Zhao, C., Yoon, K., Cook, J. M., Willert, K., Gaiano, N., & Reya, T.

- (2005). Integration of Notch and Wnt signaling in hematopoietic stem cell maintenance. *Nature Immunology*, 6(3), 314–322.
- Ehninger, A., & Trumpp, A. (2011). The bone marrow stem cell niche grows up: mesenchymal stem cells and macrophages move in. *Journal of Experimental Medicine*, 208(3), 421–428.
- Elton, C. S. (1958). *The Ecology of Invasions by Animals and Plants*. Chicago, IL: University of Chicago Press.
- Enderling, H., Anderson, A. R. A., Chaplain, M. A. J., Beheshti, A., Hlatky, L., & Hahnfeldt, P. (2009). Paradoxical Dependencies of Tumor Dormancy and Progression on Basic Cell Kinetics. *Cancer Research*, 69(22), 8814–8821.
- Epstein, F. H., Faderl, S., Talpaz, M., Estrov, Z., O'Brien, S., Kurzrock, R., & Kantarjian, H. M. (1999). The Biology of Chronic Myeloid Leukemia. *The New England journal of medicine*, 341(3), 164–172.
- Erguler, K., & Stumpf, M. P. H. (2011). Practical limits for reverse engineering of dynamical systems: a statistical analysis of sensitivity and parameter inferability in systems biology models. *Molecular BioSystems*, 7(5), 1593.
- Evans, M. R., Grimm, V., Johst, K., Knuuttila, T., de Langhe, R., Lessells, C. M., Merz, M., O'Malley, M. A., Orzack, S. H., Weisberg, M., Wilkinson, D. J., Wolkenhauer, O., & Benton, T. G. (2013). Do simple models lead to generality in ecology? *Trends in Ecology & Evolution*, 28(10), 578–583.

- Fleming, H. E., Janzen, V., Lo Celso, C., Guo, J., Leahy, K. M., Kronenberg, H. M., & Scadden, D. T. (2008). Wnt Signaling in the Niche Enforces Hematopoietic Stem Cell Quiescence and Is Necessary to Preserve Self-Renewal In Vivo. *Cell Stem Cell*, 2(3), 274–283.
- Flesken-Nikitin, A., Hwang, C.-I., Cheng, C.-Y., Michurina, T. V., Enikolopov, G., & Nikitin, A. Y. (2013). Ovarian surface epithelium at the junction area contains a cancer-prone stem cell niche. *Nature*, 495, 241–245.
- Foo, J., Drummond, M. W., Clarkson, B., Holyoake, T., & Michor, F. (2009). Eradication of chronic myeloid leukemia stem cells: a novel mathematical model predicts no therapeutic benefit of adding G-CSF to imatinib. *PLoS Computational Biology*, 5(9), e1000503.
- Forsberg, E. C., & Smith-Berdan, S. (2009). Parsing the niche code: the molecular mechanisms governing hematopoietic stem cell adhesion and differentiation. *Haematologica*, 94(11), 1477–1481.
- Franks, L. M., & Teich, N. (Eds.) (1997). *Introduction to the Cellular and Molecular Biology of Cancer*. Oxford University Press, third ed.
- Fuchs, E., Chan, E., Gat, U., & McNiff, J. M. (1999). A common human skin tumour is caused by activating mutations in beta-catenin. *Nature Genetics*, 21(4), 410–413.
- Gardiner, C. W. (Ed.) (2009). *Stochastic Methods: A Handbook for the Natural and Social Sciences*. Springer, fourth ed.

- Gardner, M. R., & Ashby, W. R. (1970). Connectance of Large Dynamic (Cybernetic) Systems: Critical Values for Stability. *Nature*, 228(5273), 784–784.
- Gilbert, N. (2008). *Agent-Based Models (Quantitative Applications in the Social Sciences)*. SAGE Publications.
- Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry*, 81(25), 2340–2361.
- Gillespie, D. T. (2001). Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, 115(4), 1716–1733.
- Girko, V. L. (1984). The Circular Law. *Theory of Probability and its Applications*, 29(4), 694–706.
- Glendinning, P. (1994). *Stability, Instability and Chaos: An Introduction to the Theory of Nonlinear Differential Equations*. Cambridge University Press.
- Goardon, N., Marchi, E., Atzberger, A., Quek, L., Schuh, A., Soneji, S., Woll, P., Mead, A., Alford, K. A., & Rout, R. (2011). Coexistence of LMPP-like and GMP-like Leukemia Stem Cells in Acute Myeloid Leukemia. *Cancer cell*, 19(1), 138–152.
- Grassinger, J., Haylock, D. N., Williams, B., Olsen, G. H., & Nilsson, S. K. (2010). Phenotypically identical hemopoietic stem cells isolated from

- different regions of bone marrow have different biologic potential. *Blood*, 116(17), 3185–3196.
- Gutenkunst, R. N., Waterfall, J. J., Casey, F. P., Brown, K. S., Myers, C. R., & Sethna, J. P. (2007). Universally sloppy parameter sensitivities in systems biology models. *PLoS Computational Biology*, 3(10), 1871–1878.
- Haccou, P., Jagers, P., & Vatutin, V. A. (2007). *Branching Processes: Variation, Growth, and Extinction of Populations*. Cambridge University Press.
- Harrison, C. J., Moorman, A. V., Broadfield, Z. J., Cheung, K. L., Harris, R. L., Reza Jalali, G., Robinson, H. M., Barber, K. E., Richards, S. M., Mitchell, C. D., Eden, T. O. B., Hann, I. M., Hill, F. G. H., Kinsey, S. E., Gibson, B. E. S., Lilleyman, J., Vora, A., Goldstone, A. H., Franklin, I. M., Durrant, J., Martineau, M., & Childhood and Adult Leukaemia Working Parties (2004). Three distinct subgroups of hypodiploidy in acute lymphoblastic leukaemia. *British Journal of Haematology*, 125(5), 552–559.
- Haylock, D. N., Williams, B., Johnston, H. M., Liu, M. C. P., Rutherford, K. E., Whitty, G. A., Simmons, P. J., Bertoncello, I., & Nilsson, S. K. (2007). Hemopoietic stem cells with higher hemopoietic potential reside at the bone marrow endosteum. *Stem Cells*, 25(4), 1062–1069.
- Henry, C. J., Marusyk, A., Zaberezhnyy, V., Adane, B., & DeGregori, J. (2010). Declining lymphoid progenitor fitness promotes aging-associated

- leukemogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 107(50), 21713–21718.
- Hofbauer, J., & Sigmund, K. (1998). *Evolutionary Games and Population Dynamics*. Cambridge, UK: Cambridge University Press.
- Hope, K. J., Jin, L., & Dick, J. E. (2004). Acute myeloid leukemia originates from a hierarchy of leukemic stem cell classes that differ in self-renewal capacity. *Nature Immunology*, 5(7), 738–743.
- Horn, M., Glauche, I., Muller, M. C., Hehlmann, R., Hochhaus, A., Loeffler, M., & Roeder, I. (2013). Model-based decision rules reduce the risk of molecular relapse after cessation of tyrosine kinase inhibitor therapy in chronic myeloid leukemia. *Blood*, 121(2), 378–384.
- Howlader, N., Noone, A., Krapcho, M., Garshell, J., Neyman, N., Altekruse, S., Kosary, C., Yu, M., Ruhl, J., Tatalovich, Z., Cho, H., Mariotto, A., Lewis, D., Chen, H., Feuer, E., & Cronin, K. (2013). Seer cancer statistics review, 1975-2010, national cancer institute. http://seer.cancer.gov/csr/1975_2010. Accessed: 2013-11-05.
- Hsu, Y.-C., & Fuchs, E. (2012). A family business: stem cell progeny join the niche to regulate homeostasis. *Nature Reviews Molecular Cell Biology*, 13(2), 103–114.
- Hu, X., Shen, H., Tian, C., Yu, H., Zheng, G., XuFeng, R., Ju, Z., Xu, J., Wang, J., & Cheng, T. (2009). Kinetics of normal hematopoietic stem and progenitor cells in a Notch1-induced leukemia model. *Blood*, 114(18), 3783–3792.

- Huang, S., Guo, Y.-P., May, G., & Enver, T. (2007). Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Developmental biology*, 305(2), 695–713.
- Ives, A. R., & Carpenter, S. R. (2007). Stability and Diversity of Ecosystems. *Science*, 317(5834), 58–62.
- Jamieson, C. H. M., Ailles, L. E., Dylla, S. J., Muijtjens, M., Jones, C., Zehnder, J. L., Gotlib, J., Li, K., Manz, M. G., Keating, A., Sawyers, C. L., & Weissman, I. L. (2004). Granulocyte-macrophage progenitors as candidate leukemic stem cells in blast-crisis CML. *The New England journal of medicine*, 351(7), 657–667.
- Jeffreys, H. (1961). *Theory of Probability*. Oxford University Press.
- Johnston, M. D., Maini, P. K., Chapman, S. J., Edwards, C. M., & Bodmer, W. F. (2010). On the proportion of cancer stem cells in a tumour. *Journal of Theoretical Biology*, 266(4), 708–711.
- Jost, J. (2005). *Dynamical Systems: Examples of Complex Behaviour*. Springer.
- Karanu, F. N., Murdoch, B., Gallacher, L., Wu, D. M., Koremoto, M., Sakano, S., & Bhatia, M. (2000). The notch ligand jagged-1 represents a novel growth factor of human hematopoietic stem cells. *The Journal of experimental medicine*, 192(9), 1365–1372.

- Kiel, M. J., Acar, M., Radice, G. L., & Morrison, S. J. (2009). Hematopoietic Stem Cells Do Not Depend on N-Cadherin to Regulate Their Maintenance. *Cell Stem Cell*, 4(2), 170–179.
- Kiel, M. J., & Morrison, S. J. (2008). Uncertainty in the niches that maintain haematopoietic stem cells. *Nature Reviews Immunology*, 8(4), 290–301.
- Kiel, M. J., Yilmaz, O. H., Iwashita, T., Yilmaz, O. H., Terhorst, C., & Morrison, S. J. (2005). SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell*, 121(7), 1109–1121.
- Kim, P. S., Lee, P. P., & Levy, D. (2008). A PDE model for imatinib-treated chronic myelogenous leukemia. *Bulletin of Mathematical Biology*, 70(7), 1994–2016.
- Kirk, P. D. W., Thorne, T., & Stumpf, M. P. H. (2013). Model selection in systems and synthetic biology. *Current opinion in biotechnology*, 24(4), 767–774.
- Kitano, H. (2007). Towards a theory of biological robustness. *Molecular Systems Biology*, 3(137).
- Kokkoris, G. D., Jansen, V. A. A., Loreau, M., & Troumbis, A. Y. (2002). Variability in interaction strength and implications for biodiversity. *Journal of Animal Ecology*, 71(2), 362–371.

- Krivtsov, A. V., Twomey, D., Feng, Z., Stubbs, M. C., Wang, Y., Faber, J., Levine, J. E., Wang, J., Hahn, W. C., Gilliland, D. G., Golub, T. R., & Armstrong, S. A. (2006). Transformation from committed progenitor to leukaemia stem cell initiated by MLL–AF9. *Nature*, 442(7104), 818–822.
- Krumsiek, J., Marr, C., Schroeder, T., & Theis, F. J. (2011). Hierarchical Differentiation of Myeloid Progenitors Is Encoded in the Transcription Factor Network. *PLoS ONE*, 6(8), e22649.
- La Porta, C. A. M., Zapperi, S., & Sethna, J. P. (2012). Senescent Cells in Growing Tumors: Population Dynamics and Cancer Stem Cells. *PLoS Computational Biology*, 8(1), e1002316.
- Lane, S. W., Scadden, D. T., & Gilliland, D. G. (2009). The leukemic stem cell niche: current concepts and therapeutic opportunities. *Blood*, 114(6), 1150–1157.
- Lane, S. W., Wang, Y. J., Lo Celso, C., Ragu, C., Bullinger, L., Sykes, S. M., Ferraro, F., Shterental, S., Lin, C. P., Gilliland, D. G., Scadden, D. T., Armstrong, S. A., & Williams, D. A. (2011). Differential niche and Wnt requirements during acute myeloid leukemia progression. *Blood*, 118(10), 2849–2856.
- Lanza, R. (Ed.) (2009). *Essentials of Stem cell Biology*. Academic Press.
- Lapidot, T., Sirard, C., Vormoor, J., Murdoch, B., Hoang, T., Caceres-Cortes, J., Minden, M., Paterson, B., Caligiuri, M. A., & Dick, J. E. (1994). A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature*, 367(6464), 645–648.

- Levins, R. (1968). *Evolution in Changing Environments: Some Theoretical Explorations*. Princeton, NJ: Princeton University Press.
- Liepe, J., Barnes, C. P., Cule, E., Erguler, K., Kirk, P. D. W., Toni, T., & Stumpf, M. P. H. (2010). ABC-SysBio—approximate Bayesian computation in Python with GPU support. *Bioinformatics*, *26*(14), 1797–1799.
- Lo Celso, C., Fleming, H. E., Wu, J. W., Zhao, C. X., Miake-Lye, S., Fujisaki, J., Côté, D., Rowe, D. W., Lin, C. P., & Scadden, D. T. (2009a). Live-animal tracking of individual haematopoietic stem/progenitor cells in their niche. *Nature*, *457*(7225), 92–96.
- Lo Celso, C., Lin, C. P., & Scadden, D. T. (2011). In vivo imaging of transplanted hematopoietic stem and progenitor cells in mouse calvarium bone marrow. *Nature protocols*, *6*(1), 1–14.
- Lo Celso, C., Wu, J. W., & Lin, C. P. (2009b). In vivo imaging of hematopoietic stem cells and their microenvironment. *Journal of Biophotonics*, *2*(11), 619–631.
- Lorenz, E. N. (1963). Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences*, *20*, 130–141.
- Lotka, A. J. (1920). Analytical Note on Certain Rhythmic Relations in Organic Systems. *Proceedings of the National Academy of Sciences of the United States of America*, *6*, 410–415.
- MacArthur, R. H. (1955). Fluctuations of animal populations and a measure of community stability. *Ecological Modelling*, *36*(3), 533–536.

- MacArthur, R. H., & Wilson, E. O. (1967). *The Theory of Island Biogeography*. Princeton University Press.
- Mackey, M. C., & Glass, L. (1977). Oscillation and chaos in physiological control systems. *Science (New York, N.Y.)*.
- MacLean, A. L., Filippi, S., & Stumpf, M. P. H. (2014). The ecology in the hematopoietic stem cell niche determines the clinical outcome in chronic myeloid leukemia. *Proceedings of the National Academy of Sciences of the United States of America*, 111(10), 3882–3888.
- MacLean, A. L., Lo Celso, C., & Stumpf, M. P. H. (2013). Population dynamics of normal and leukaemia stem cells in the haematopoietic stem cell niche show distinct regimes where leukaemia will be controlled. *Journal of the Royal Society Interface*, 10(81), 10.1098rsif.2012.0968.
- Mahon, F.-X. (2012). Is going for cure in chronic myeloid leukemia possible and justifiable? *Hematology / the Education Program of the American Society of Hematology. American Society of Hematology. Education Program*, 2012, 122–128.
- Malhotra, S., & Kincade, P. W. (2009). Wnt-related molecules and signaling pathway equilibrium in hematopoiesis. *Cell Stem Cell*, 4(1), 27–36.
- Manesso, E., Teles, J., Bryder, D., & Peterson, C. (2013). Dynamical modelling of haematopoiesis: an integrated view over the system in homeostasis and under perturbation. *Journal of the Royal Society Interface*, (p. 10.1098rsif.2012.0817).

- Mangel, M., & Bonsall, M. B. (2008). Phenotypic Evolutionary Models in Stem Cell Biology: Replacement, Quiescence, and Variability. *PLoS ONE*, 3(2), e1591.
- Mangel, M., & Bonsall, M. B. (2013). Stem cell biology is population biology: differentiation of hematopoietic multipotent progenitors to common lymphoid and myeloid progenitors. *Theoretical Biology and Medical Modelling*, 10(1), 5–34.
- Marjoram, P., Molitor, J., Plagnol, V., & Tavaré, S. (2003). Markov chain monte carlo without likelihoods. *Proc. Natl. Acad. Sci. USA*, 100, 15324–15328.
- Marr, C., Strasser, M., Schwarzfischer, M., Schroeder, T., & Theis, F. J. (2012). Multi-scale modeling of GMP differentiation based on single-cell genealogies. *FEBS Journal*, (pp. 10.1111/j.1742–4658.2012.08664.x).
- Martinez-Agosto, J. A., Mikkola, H. K. A., Hartenstein, V., & Banerjee, U. (2007). The hematopoietic stem cell and its niche: a comparative view. *Genes & Development*, 21(23), 3044–3060.
- Marusyk, A., Casás-Selves, M., Henry, C. J., Zaberezhnyy, V., Klawitter, J., Christians, U., & DeGregori, J. (2009). Irradiation alters selection for oncogenic mutations in hematopoietic progenitors. *Cancer Research*, 69(18), 7262–7269.
- Mason, N. W. H., Irz, P., Lanoiselée, C., Mouillot, D., & Argillier, C. (2008). Evidence that niche specialization explains species–energy relationships in lake fish communities. *Journal of Animal Ecology*, 77(2), 285–296.

- May, R. M. (1972). Will a Large Complex System be Stable? *Nature*, *238*, 413–414.
- May, R. M. (1973a). Qualitative Stability in Model Ecosystems. *Ecological Modelling*, *54*, 638–641.
- May, R. M. (1973b). *Stability and Complexity in Model Ecosystems*. Princeton, NJ: Princeton University Press.
- May, R. M., & MacArthur, R. H. (1972). Niche overlap as a function of environmental variability. *Proceedings of the National Academy of Sciences of the United States of America*, *69*(5), 1109–1113.
- Maynard Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge University Press.
- Medema, J. P. (2013). Cancer stem cells: The challenges ahead. *Nature Cell Biology*, *15*(4), 338–344.
- Méndez-Ferrer, S., Michurina, T. V., Ferraro, F., Mazloom, A. R., MacArthur, B. D., Lira, S. A., Scadden, D. T., Ma’ayan, A., Enikolopov, G. N., & Frenette, P. S. (2010). Mesenchymal and haematopoietic stem cells form a unique bone marrow niche. *Nature*, *466*(7308), 829–834.
- Michor, F., Hughes, T. P., Iwasa, Y., Branford, S., Shah, N. P., Sawyers, C. L., & Nowak, M. A. (2005). Dynamics of chronic myeloid leukaemia. *Nature*, *435*(7046), 1267–1270.
- Mikkers, H., & Frisé, J. (2005). Deconstructing stemness. *The EMBO journal*, *24*(15), 2715–2719.

- Mikkola, H. K., & Orkin, S. H. (2006). The journey of developing hematopoietic stem cells. *Development*, 133(19), 3733–3744.
- Miraki-Moud, F., Anjos-Afonso, F., Hodby, K. A., Griessinger, E., Rosignoli, G., Lillington, D., Jia, L., Davies, J. K., Cavenagh, J., Smith, M., Oakervee, H., Agrawal, S., Gribben, J. G., Bonnet, D., & Taussig, D. C. (2013). Acute myeloid leukemia does not deplete normal hematopoietic stem cells but induces cytopenias by impeding their differentiation. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 13576–13581.
- Moore, H., & Li, N. K. (2004). A mathematical model for chronic myelogenous leukemia (CML) and T cell interaction. *Journal of Theoretical Biology*, 227(4), 513–523.
- Morrison, S. J., Shah, N. M., & Anderson, D. J. (1997). Regulatory mechanisms in stem cell biology. *Cell*, 88(3), 287–298.
- Morrison, S. J., & Spradling, A. C. (2008). Stem Cells and Niches: Mechanisms That Promote Stem Cell Maintenance throughout Life. *Cell*, 132(4), 598–611.
- Morrison, S. J., Uchida, N., & Weissman, I. L. (1995). The Biology of Hematopoietic Stem Cells. *Annual Review of Cell and Developmental Biology*, 11(1), 35–71.
- Muller, M. C., Gattermann, N., Lahaye, T., Deininger, M. W. N., Berndt, A., Fruehauf, S., Neubauer, A., Fischer, T., Hossfeld, D. K., Schneller, F., Krause, S. W., Nerl, C., Sayer, H. G., Ottmann, O. G., Waller, C.,

- Aulitzky, W., Coutre, P. I., Freund, M., Merx, K., Paschka, P., König, H., Kreil, S., Berger, U., Gschaidmeier, H., Hehlmann, R., & Hochhaus, A. (2003). Dynamics of BCR-ABL mRNA expression in first-line therapy of chronic myelogenous leukemia patients with imatinib or interferon α /ara-C. *Leukemia*, 17(12), 2392–2400.
- Murray, J. D. (2008). *An Introduction to Mathematical Biology: Pt. 1*. Springer, third ed.
- Nordon, R. E., Ko, K.-H., Odell, R., & Schroeder, T. (2011). Multi-type branching models to describe cell differentiation programs. *Journal of Theoretical Biology*, 277(1), 7–18.
- Nowell, P. C. (1976). The clonal evolution of tumor cell populations. *Science*, 194(4260), 23–28.
- Okada, S., Nakauchi, H., Nagayoshi, K., Nishikawa, S., Miura, Y., & Suda, T. (1992). In vivo and in vitro stem cell function of c-kit- and Sca-1-positive murine hematopoietic cells. *Blood*, 80(12), 3044–3050.
- Paik, E. J., & Zon, L. I. (2010). Hematopoietic development in the zebrafish. *The International journal of developmental biology*, 54(6-7), 1127–1137.
- Passegué, E., Jamieson, C. H. M., Ailles, L. E., & Weissman, I. L. (2003). Normal and leukemic hematopoiesis: Are leukemias a stem cell disorder or a reacquisition of stem cell characteristics? *Proceedings of the National Academy of Sciences of the United States of America*, 100, 11842–11849.

- Polakis, P. (2000). Wnt signaling and cancer. *Genes & Development*, 14(15), 1837–1851.
- Potten, C. S., & Loeffler, M. (1990). Stem cells: attributes, cycles, spirals, pitfalls and uncertainties. Lessons for and from the crypt. *Development*, 110(4), 1001–1020.
- Qu, K., & Ortoleva, P. (2008). Understanding stem cell differentiation through self-organization theory. *Journal of Theoretical Biology*, 250(4), 606–620.
- Quintana, E., Shackleton, M., Sabel, M. S., Fullen, D. R., Johnson, T. M., & Morrison, S. J. (2008). Efficient tumour formation by single human melanoma cells. *Nature*, 456(7222), 593–598.
- Renstrom, J., Kroger, M., Peschel, C., & Oostendorp, R. A. J. (2010). How the niche regulates hematopoietic stem cells. *Chemico-Biological Interactions*, 184(1-2), 7–15.
- Reya, T., Morrison, S. J., Clarke, M. F., & Weissman, I. L. (2001). Stem cells, cancer, and cancer stem cells. *Nature*, 414(6859), 105–111.
- Roberts, A. (1974). The stability of a feasible random ecosystem. *Nature*, 251(5476), 607–608.
- Roeder, I., Horn, M., Glauche, I., Hochhaus, A., Mueller, M. C., & Loeffler, M. (2006). Dynamic modeling of imatinib-treated chronic myeloid leukemia: functional insights and clinical implications. *Nature Medicine*, 12(10), 1181–1184.

- Roeder, I., Kamminga, L. M., Braesel, K., Dontje, B., de Haan, G., & Loeffler, M. (2005). Competitive clonal hematopoiesis in mouse chimeras explained by a stochastic model of stem cell organization. *Blood*, *105*(2), 609–616.
- Roeder, I., & Loeffler, M. (2002). A novel dynamic model of hematopoietic stem cell organization based on the concept of within-tissue plasticity. *Experimental hematology*, *30*(8), 853–861.
- Roughgarden, J., May, R. M., & Levin, S. A. (Eds.) (1989). *Perspectives in Ecological Theory*. Princeton University Press.
- Roy, A., Cowan, G., Mead, A. J., Filippi, S., Bohn, G., Chaidos, A., Tunstall, O., Chan, J. K. Y., Choolani, M., Bennett, P., Kumar, S., Atkinson, D., Wyatt-Ashmead, J., Hu, M., Stumpf, M. P. H., Goudevenou, K., O'Connor, D., Chou, S. T., Weiss, M. J., Karadimitris, A., Jacobsen, S. E., Vyas, P., & Roberts, I. (2012). Perturbation of fetal liver hematopoietic stem and progenitor cell development by trisomy 21. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(43), 17579–17584.
- Rubnitz, J. E., Morrissey, J., Savage, P. A., & Cleary, M. L. (1994). ENL, the gene fused with HRX in t(11;19) leukemias, encodes a nuclear protein with transcriptional activation potential in lymphoid and myeloid cells. *Blood*, *84*(6), 1747–1752.
- Saavedra, S., Stouffer, D. B., Uzzi, B., & Bascompte, J. (2011). Strong

- contributors to network persistence are the most vulnerable to extinction. *Nature*, 478(7368), 233–235.
- Santner, T. J., Williams, B. J., & Notz, W. I. (2003). *The Design and Analysis of Computer Experiments*. Springer.
- Schepers, K., Pietras, E. M., Reynaud, D., Flach, J., Binnewies, M., Garg, T., Wagers, A. J., Hsiao, E. C., & Passegué, E. (2013). Myeloproliferative Neoplasia Remodels the Endosteal Bone Marrow Niche into a Self-Reinforcing Leukemic Niche. *Cell Stem Cell*, 13(3), 285–299.
- Schittler, D., Hasenauer, J., & Allgöwer, F. (2012). A model for proliferating cell populations that accounts for cell types. In *Proceedings of Ninth International Workshop on Computational Systems Biology*.
- Schofield, R. (1978). The relationship between the spleen colony-forming cell and the haemopoietic stem cell. *Blood Cells*, 4(1-2), 7–25.
- Scianna, M., & Preziosi, L. (2013). *Cellular Potts Models: Multiscale Extensions and Biological Applications*. Chapman and Hall/CRC.
- Secrier, M., Toni, T., & Stumpf, M. P. H. (2009). The ABC of reverse engineering biological signalling systems. *Molecular BioSystems*, 5(12), 1925–1935.
- Shackleton, M., Quintana, E., Fearon, E. R., & Morrison, S. J. (2009). Heterogeneity in Cancer: Cancer Stem Cells versus Clonal Evolution. *Cell*, 138(5), 822–829.

- Shipitsin, M., & Polyak, K. (2008). The cancer stem cell hypothesis: in search of definitions, markers, and relevance. *Laboratory Investigation*, 88(5), 459–463.
- Sieburg, H. B., Rezner, B. D., & Muller-Sieburg, C. E. (2011). Predicting clonal self-renewal and extinction of hematopoietic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, 108(11), 4370–4375.
- Sigmund, K. (1993). *Games of Life: Explorations in Ecology, Evolution, and Behaviour*. Oxford, UK: Oxford University Press.
- Simons, B. D., & Clevers, H. (2011). Strategies for Homeostatic Stem Cell Self-Renewal in Adult Tissues. *Cell*, 145(6), 851–862.
- Smith-Berdan, S., Nguyen, A., Hassanein, D., Zimmer, M., Ugarte, F., Ciriza, J., Li, D., García-Ojeda, M. E., Hinck, L., & Forsberg, E. C. (2011). Robo4 Cooperates with Cxcr4 to Specify Hematopoietic Stem Cell Localization to Bone Marrow Niches. *Cell Stem Cell*, 8(1), 72–83.
- Solé, R. V., Rodríguez-Caso, C., Deisboeck, T. S., & Saldaña, J. (2008). Cancer stem cells as the engine of unstable tumor progression. *Journal of Theoretical Biology*, 253(4), 629–637.
- Sompayrac, L. (2008). *How the Immune System Works*. Blackwell Publishing, third ed.
- Sottoriva, A., & Tavaré, S. (2010). Integrating Approximate Bayesian Computation with Complex Agent-Based Models for Cancer Research. In *Pro-*

- ceedings of COMPSTAT'2010*, (pp. 57–66). Heidelberg: Physica-Verlag HD.
- Spangrude, G. J., Heimfeld, S., & Weissman, I. L. (1988). Purification and characterization of mouse hematopoietic stem cells. *Science*, *241*(4861), 58–62.
- Staal, F. J. T., & Luis, T. C. (2010). Wnt signaling in hematopoiesis: crucial factors for self-renewal, proliferation, and cell fate decisions. *Journal of cellular biochemistry*, *109*(5), 844–849.
- Strogatz, S. H. (1994). *Nonlinear Dynamics And Chaos*. Perseus Books.
- Sugimura, R., He, X. C., Venkatraman, A., Arai, F., Box, A., Semerad, C., Haug, J. S., Peng, L., Zhong, X.-B., Suda, T., & Li, L. (2012). Non-canonical Wnt signaling maintains hematopoietic stem cells in the niche. *Cell*, *150*(2), 351–365.
- Sugiyama, T., Kohara, H., Noda, M., & Nagasawa, T. (2006). Maintenance of the hematopoietic stem cell pool by CXCL12-CXCR4 chemokine signaling in bone marrow stromal cell niches. *Immunity*, *25*(6), 977–988.
- Tang, M., Foo, J., Gonen, M., Guilhot, J., Mahon, F. X., & Michor, F. (2012). Selection pressure exerted by imatinib therapy leads to disparate outcomes of imatinib discontinuation trials. *Haematologica*, *97*(10), 1553–1561.
- Tao, T., Vu, V., & Krishnapur, M. (2010). Random matrices: Universality

- of ESDs and the circular law. *The Annals of Probability*, 38(5), 2023–2065.
- Tomlinson, I. P., & Bodmer, W. F. (1997). Modelling the consequences of interactions between tumour cells. *British journal of cancer*, 75(2), 157–160.
- Toni, T., & Stumpf, M. P. H. (2010). Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics*, 26(1), 7.
- Toni, T., Welch, D., Strelkowa, N., Ipsen, A., & Stumpf, M. P. H. (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, 6(31), 187–202.
- Traulsen, A., Lenaerts, T., Pacheco, J. M., & Dingli, D. (2012). On the dynamics of neutral mutations in a mathematical model for a homogeneous stem cell population. *Journal of the Royal Society Interface*, 10(79), 10.1098rsif.2012.0810.
- Trumpp, A., essers, M., & Wilson, A. (2010). Awakening dormant haematopoietic stem cells. *Nature Reviews Immunology*, 10(3), 201–209.
- Uhr, J. W., & Pantel, K. (2011). Controversies in clinical cancer dormancy. *Proceedings of the National Academy of Sciences of the United States of America*, 108(30), 12396–12400.

- Villani, M., Barbieri, A., & Serra, R. (2011). A dynamical model of genetic networks for cell differentiation. *PLoS ONE*, 6(3), e17703.
- Volterra, V. (1926). Fluctuations in the Abundance of a Species considered Mathematically. *Nature*, 118(2972), 558–560.
- von Neumann, J. (1966). *Theory of Self-Reproducing Automata*. University of Illinois Press.
- Wang, J. C. Y., & Dick, J. E. (2005). Cancer stem cells: lessons from leukemia. *Trends in Cell Biology*, 15(9), 494–501.
- Wang, L. D., & Wagers, A. J. (2011). Dynamic niches in the origination and differentiation of haematopoietic stem cells. *Nature Reviews Molecular Cell Biology*, 12(10), 643–655.
- Wang, X., Rosol, M., Ge, S., Peterson, D., McNamara, G., Pollack, H., Kohn, D. B., Nelson, M. D., & Crooks, G. M. (2003). Dynamic tracking of human hematopoietic stem cell engraftment using in vivo bioluminescence imaging. *Blood*, 102(10), 3478–3482.
- Wechsler-Reya, R., & Scott, M. P. (2001). The Developmental Biology of Brain Tumors. *Annual review of neuroscience*, 24, 385–428.
- Werner, B., Dingli, D., Lenaerts, T., Pacheco, J. M., & Traulsen, A. (2011). Dynamics of Mutant Cells in Hierarchical Organized Tissues. *PLoS Computational Biology*, 7(12), e1002290.
- Wilkinson, D. J. (2011). *Stochastic Modelling for Systems Biology*. Chapman and Hall/CRC, second ed.

- Wilson, A., Laurenti, E., Oser, G., van der Wath, R. C., Blanco-Bose, W., Jaworski, M., Offner, S., Dunant, C. F., Eshkind, L., Bockamp, E., Lió, P., Macdonald, H. R., & Trumpp, A. (2008). Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell*, 135(6), 1118–1129.
- Wilson, A., Oser, G. M., Jaworski, M., Blanco-Bose, W. E., Laurenti, E., Adolphe, C., Essers, M. A., Macdonald, H. R., & Trumpp, A. (2007). Dormant and self-renewing hematopoietic stem cells and their niches. *Annals of the New York Academy of Sciences*, 1106, 64–75.
- Wilson, A., & Trumpp, A. (2006). Bone-marrow haematopoietic-stem-cell niches. *Nature Reviews Immunology*, 6(2), 93–106.
- Wolfram, S. (1983). Statistical mechanics of cellular automata. *Reviews of Modern Physics*, 55(3), 601–644.
- Wolkenhauer, O., Shibata, D. K., & Mesarović, M. D. (2011). A stem cell niche dominance theorem. *BMC Systems Biology*, 5(1), 4.
- Xu, M., & Xu, S. (2010). Global behavior of a discrete haematopoiesis model with several delays. *Applied Mathematics and Computation*, 217(1), 441–449.
- Yilmaz, O. H., Valdez, R., Theisen, B. K., Guo, W., Ferguson, D. O., Wu, H., & Morrison, S. J. (2006). Pten dependence distinguishes haematopoietic stem cells from leukaemia-initiating cells. *Nature*, 441(7092), 475–482.

- Zhang, J., Niu, C., Ye, L., Huang, H., He, X., Tong, W.-G., Ross, J., Haug, J., Johnson, T., Feng, J. Q., Harris, S., Wiedemann, L. M., Mishina, Y., & Li, L. (2003). Identification of the haematopoietic stem cell niche and control of the niche size. *Nature*, 425(6960), 836–841.
- Zhou, Y., Liepe, J., Sheng, X., Stumpf, M. P. H., & Barnes, C. P. (2011). GPU accelerated biochemical network simulation. *Bioinformatics*, 27(6), 874–876.
- Zhu, X., Zhou, X., Lewis, M. T., Xia, L., & Wong, S. (2011). Cancer stem cell, niche and EGFR decide tumor development and treatment response A bio-computational simulation study. *Journal of Theoretical Biology*, 269(1), 138–149.
- Zhu, Y., Li, F., Vadakkan, T. J., & Zhang, M. (2013). Three-dimensional vasculature reconstruction of tumour microenvironment via local clustering and classification. *Interface Focus*, (p. 10.1098rsfs.2013.0015).