# INAUGURAL-DISSERTATION

zur Erlangung der Doktorwürde der

NATURWISSENSCHAFTLICH-MATHEMATISCHEN
GESAMTFAKULTÄT

der

RUPRECHT-KARLS-UNIVERSITÄT
HEIDELBERG

vorgelegt von

**M. Sc. Ana Victoria Ponce Bobadilla**

aus Mexiko-Stadt, Mexiko

Tag der mündlichen Prüfung:

# Mathematical Models of Cell Migration and Proliferation in Scratch Assays

ANA VICTORIA PONCE BOBADILLA

Betreuer: Dr. Thomas Carraro
Prof. Dr. Tomás Alarcón

To my mom.
As Abraham Lincoln once said:
“All that I am
or hope to be,
I owe to my mother.”

*"Why use mathematics to study something as intrinsically complicated and ill understood as development, angiogenesis, wound healing, interacting population dynamics, regulatory networks, marital interaction and so on? We suggest that mathematics, rather theoretical modelling, must be used if we ever hope to genuinely and realistically convert an understanding of the underlying mechanisms into a predictive science."*

J.D Murray Mathematical Biology I. An Introduction

# ABSTRACT

Scratch assays are standard in-vitro experimental procedures for studying cell migration. In these experiments, a scratch is made on a cell monolayer. By imaging the recolonisation process of the scratched region, we are able to quantify cell migration rates. This experimental technique is commonly used in the pharmaceutical industry to identify new compounds that target cell migration, and to evaluate the efficacy of potential drugs that inhibit cancer invasion.

Given the key role this method plays in assessing the potential of new compounds for clinical use, it is important to develop robust quantification frameworks that accurately describe the movement of the front of migrating cells. We develop a migration quantification method that fits experimental data more closely than existing methods, provides a more accurate statistical classification of the migration rate between different assays and is able to cope with experimental data of lower quality than the classic quantification methods can handle. The robustness of our new method is validated using both in-vitro and in-silico data.

Developing robust quantification methods allows the validation of mathematical models that can be used to test hypotheses about the physical and biological mechanisms that govern cell migration. Typically scratch assays are modelled by continuum reaction-diffusion equations depicting cell migration by diffusion and carrying capacity-limited proliferation by a logistic source term. An age-structured population model is presented that aims to explain the two phases of proliferation in scratch assays previously observed experimentally: where an initial phase is observed where proliferation is not logistic, followed by a second phase where proliferation appears to be logistic. The cell population is modelled by a McKendrick-von Foerster partial differential equation. The conditions under which the model captures this two-phase behaviour are presented.

Finally, an important aspect of modelling biological systems is the development of efficient algorithms. The scratch assay is a classical example of a system in which there is low cell number in some regions of the spatial domain and high cell number in others. When the cell number is sufficiently high, mean-field models, like partial differential equations, can capture the relevant dynamics. However, when the cell number is low, such models are not appropriate and stochastic representations must be employed. Hybrid algorithms allow multiple modelling frameworks for the same species in different parts of the spatial domain. Typically hybrid algorithms consider heuristic methods based on the cell density for determining which compartments will be updated deterministically or stochastically. We introduce a hybrid algorithm that couples the mesoscopic description of a reaction-diffusion system with its mean-field analogue. We consider a natural indicator of when the mean-field approximation is valid: the system variance. We estimate the system variance using the intrinsic noise approximation and use this estimate to determine the regions in which the system is updated stochastically or deterministically over time. We apply the hybrid algorithm to the stochastic Fisher-Kolmogorov-Petrovsky-Piscounov model, a typical model of scratch assays. We analyse systematically how good is the approximation to the stochastic process and compare its performance to another hybrid algorithm.

# ZUSAMMENFASSUNG

Scratch Assays sind übliche experimentelle in-vitro Prozeduren, um Zellmigration zu erforschen. In diesen Experimenten wird ein Kratzer in eine einschichtige Zellebene gemacht. Durch Bildverarbeitung des Rekolonisierungsprozesses sind wir in der Lage, Zellmigrationsraten zu quantifizieren. Diese experimentelle Technik wird in der pharmazeutischen Industrie üblicherweise angewandt, um neue Komponenten zu identifizieren, welche auf Zellmigration zielen und um die Effektivität von potenziellen Medikamenten, welche Krebsinvasion verhindern können, zu evaluieren.

In Anbetracht der Schlüsselrolle, welche diese Methode beim Feststellen des Potentials neuer Komponenten in der klinischen Anwendung spielt, ist es wichtig, ein robustes Quantifizierungs-Framework zu entwickeln, welches die Bewegung der vordersten Zellfront der migrierenden Zellen akkurat beschreibt. Wir entwickeln eine Quantifizierungsmethode für Migration, welche experimentelle Daten besser als bislang existierende Methoden beschreibt, eine genauere statistische Klassifizierung der Migrationsrate von verschiedenen Assays liefert und es möglich macht, experimentelle Daten von niedrigerer Qualität zu verarbeiten, wozu klassische Methoden nicht in der Lage sind. Die Robustheit unserer neuen Methode wird durch das Anwenden auf sowohl in-vitro als auch in-silico Daten validiert.

Das Entwickeln robuster Quantifizierungsmethoden erlaubt es, mathematische Modelle zu validieren, welche benutzt werden können, um Hypothesen bezüglich der physikalischen und biologischen Mechanismen, welche Zellmigration steuern, zu testen. Normalerweise werden Scratch Assays durch kontinuierliche Reaktions-Diffusionsgleichungen modelliert, welche Zellmigration durch Diffusion und tragfähigkeitslimitierte Proliferation durch einen logistischen Quellenterm beschreiben. Ein altersstruktiertes Populationsmodell wird präsentiert, das darauf zielt, die zwei Phasen der Proliferation in Scratch Assays, welche bislang experimentell beobachtet wurden, zu erklären: Zunächst wird eine Initialphase beobachtet, in welcher Proliferation nicht logistisch ist, gefolgt von einer zweiten Phase, in welcher Proliferation logistisch zu sein scheint. Die Zellpopulation wird durch eine McKendrick-von Foerster partielle Differentialgleichung modelliert. Die Bedingungen, unter welchen das Modell dieses zweiphasige Verhalten wiedergibt, werden präsentiert.

Zu guter Letzt ist das Entwickeln von effizienten Algorithmen ein wichtiger Aspekt in der Modellierung biologischer Systeme. Der Scratch assay ist ein klassisches Beispiel eines Systems, in welchem in manchen Regionen des räumlichen Gebietes eine geringe Zellzahl und in anderen Regionen eine hohe Zellzahl vorliegt. Wenn die Zellzahl genügend groß ist, können Mittelfeld-Modelle, wie partielle Differentialgleichungen, die relevanten Dynamiken erfassen. Ist die Zellzahl allerdings niedrig, sind solche Modelle nicht mehr angebracht und stochastische Repräsentationen müssen genutzt werden. Hybride Algorithmen erlauben das Modellieren multipler Frameworks der gleichen Spezies in unterschiedlichen Teilen des räumlichen Gebietes. Typischerweise berücksichtigen hybride Algorithmen heuristische Methoden basierend auf der Zelldichte, um festzustellen, welche Kompartimente deterministisch oder stochastisch aktualisiert werden sollen. Wir führen einen hybriden Algorithmus ein, welcher die mesoskopische Beschreibung von Reaktions-Diffusions-Systemen an sein Mittelfeld-Analogon koppelt. Wir betrachten einen natürlichen Indikator für die Gültigkeit der Mittelfeldnäherung: Die Systemvarianz. Wir schätzen die Systemvarianz durch das Nutzen von intrinsischer

Störapproximation und nutzen diese Schätzung, um die Regionen zu bestimmen, in welchen das System stochastisch oder deterministisch über die Zeit aktualisiert wird. Wir wenden den hybriden Algorithmus auf das stochastische Fisher-Kolmogorov-Petrovsky-Piscounov-Modell an, ein typisches Modell für Scratch Assays. Wir analysieren systematisch, wie gut die Approximation an den stochastischen Prozess ist und vergleichen seine Leistung mit einem anderen hybriden Algorithmus.

# ACKNOWLEDGEMENTS

First and foremost I would like to thank my supervisors. Thank you for your support, guidance, patience and motivation throughout my PhD. Thank you for the constructive discussions despite the usual Skype connection problems. Thank you Thomas Carraro for the opportunity of pursuing a PhD in Mathematics at the Faculty of Mathematics and Computer Science at the Ruprecht Karl University of Heidelberg. Thank you for the freedom to pursue different research directions during my project. Thank you Tomás Alarcón for your creative ideas and endless help that guided my PhD. Thank you for introducing me to Simone and Ana, I always wanted to work in an interdisciplinary project and you made that wish come true. Thank you Philip Maini and Helen Byrne for all the support and constant feedback that you have given me since my master, I am honoured to have been working with you ever since. Thank you Tomás, Philip and Helen for allowing my multiple visits to the Centre de Recerca Matemàtica at the Autonomous University of Barcelona and the Wolfson Centre for Mathematical Biology (WCMB) at the University of Oxford.

I am very thankful to our experimental collaborators, Jazmine Arévalo, Eduard Sarró and Anna Meseguer, who provided us with the experimental data that enabled the work in Chapter 3. Thank you for your insightful explanations that helped me to demystify cancer cell biology. Special thanks go to Simone Balocco for hosting me in his lab during my stays in Barcelona and introducing me to the fascinating world of image processing.

Thank you Sven for your infinite patience and for all your help with programming and finite elements. Thank you Maria and Gegerly for helping me make sense of delay differential equations. My gratitude goes to Pablo for sharing your programming expertise and your passion for maths. Thank you Nils for helping me navigate german bureaucracy and cell biology.

My PhD would not have been possible without the funding from different sources. Primarily I would like to thank the Heidelberg Graduate School of Mathematical and Computational Methods for the Sciences (HGS MathComp) for funding this PhD project, which also enabled me to participate in several conferences, workshops and summer schools. Special thanks go to my principal supervisor for the additional funding for my last year of PhD. I would also like to thank the Heidelberg University Graduate Academy, the Mathematics for Industry Network and the Society of Mathematical Biology for the mobility grants that allowed me to participate in different research events.

This acknowledgements section would not be complete without my gratitute to the amazing people that I met in this PhD journey. I would like to thank my german family who have made Heidelberg feel like home: Asha, Diego, Andreas, Priscila, Danilo, Pablo, Lutz, Maggie, Julia, Lilo, Nils, Marcel, Michael and Pryata. Special thanks goes to the Botanik philosophers that made coffee time a heated debate, on politics, science and life. My gratitude also goes to several members of WCMB, specially to Rasa, Ana, Andrew and Masha, for making my stays in Oxford so enjoyable that I always want to come back. Thank you Andrew for your constant advice and help and for being a role model as a mathematician and as a human. Thank you

Adrián for your advice and humbleness that kept me sane and humble since our undergraduate time. Special thanks go to Bartosz for sharing your love for programming, for keeping me sane during the last stages of thesis writing and for always being there for me.

Last but not least, I am wholeheartedly grateful to my mom, for your continual love and support. This thesis is dedicated to you.

# Contents

# List of Abbreviations

| | |
|---|---|
| **DDE** | **D**elay **D**ifferential **E**quation |
| **FKPP** | **F**isher **K**olmogorov **P**etrovsky **P**iscounov |
| **KS test** | two sample **K**olmogorov **S**mirnov **test** |
| **MMS** | **M**ethod of **M**anufactured **S**olutions |
| **MEV** | **M**onolayer **E**dge **V**elocimetry |
| **ODE** | **O**rdinary **D**ifferential **E**quation |
| **PDE** | **P**artial **D**ifferential **E**quation |
| **RDME** | **R**eaction **D**iffusion **M**aster **E**quation |
| **t test** | unpaired two-sample **t test** |

# Chapter 1

# Introduction

Biological processes are intrinsically complex, as they involve multiple time and spatial scales, feedbacks and nonlinear interactions. The molecular mechanisms, at the nano-scale, and the cell behaviour, at micro-scale, can affect the tissue dynamics, at the centimeter-scale, and vice versa [52]. Studying a biological system through its structure and dynamics provides an insight into the mechanisms that control the system behaviour and help us understand the processes that trigger malfunctions [131]. From this system knowledge we can determine potential targets for disease treatments. Mathematical modelling, analysis, and computational simulations have great potential to help us understand multiscale interactions and to provide a quantitative description of the mechanisms of the biological process [52].

Mathematical models of biological systems are putative descriptions of the system and the underlying mechanisms [108]. These models focus on certain aspects of the object of study while other aspects are neglected or abstracted. Model analysis and computational simulations help us discover links between the system structure and its behaviour. They also allow us to explore the system in conditions that are experimentally unfeasible. By comparing the model dynamics to experimental data, one can analyse if the mathematical model is suitable and highlight gaps in our biological knowledge of the system [108].

Cell migration and proliferation play a key role in developing and maintaining the organisation of multicellular organisms while their malfunction can lead to pathological processes such as cancer invasion, chronic inflammatory diseases and vascular diseases [49, 77, 115, 146, 156]. To understand the biochemical and physical cues of these processes, in-vitro assays are used to quantify such attributes under controlled situations [3, 90, 135, 162]. A common in-vitro assay for examining collective cell migration is the scratch assay [142, 162].

In this thesis we introduce three mathematical models of cell migration and proliferation in scratch assays. Before introducing these models we first give a brief introduction to the biological and mathematical concepts needed to understand the material presented in the thesis. In Section 1.1 we describe briefly the main aspects of the biological objects of study. In Section 1.2 we introduce the mathematical models that have been previously developed. In Section 1.3 we introduce the main results and the contributions of the thesis and finally in Section 1.4 we present how the thesis is structured.

## 1.1 Biological background

In this section we introduce the biological and biomedical concepts necessary to understand the mathematical models presented in this thesis. First we give a brief introduction to cell proliferation and migration (Subsection 1.1.1). Then, we introduce the research techniques used to investigate the cell dynamics in-vitro (Subsection 1.1.2).

### 1.1.1 Cell proliferation and migration

We give a brief description of two main cell processes: cell proliferation and migration. For a more detailed description of cell components and processes we refer to [7, 51, 101].

Eukaryotic cell proliferation is regulated through a series of phases that lead to cell division called the *cell cycle*. The cell cycle consists of four phases: the first gap $G1$, synthesis $S$ phase, second gap $G2$ and mitosis $M$ phase (see Figure 1.1). The $G1$, $S$ and $G2$ are called the interphase and in this period the cell doubles its mass and prepares for cell division [7]. A cell spends most of its lifetime in the interphase. In the $S$ phase the DNA is replicated [228] and in the M phase two processes happen: *mitosis*, the nucleus divides and *cytokinesis*, the cytoplasm divides. $G1$ and $G2$ are the so-called gap phases. During these phases, cells prepare for the $S$ and the $M$ phases, respectively. This means that cells produce the necessary mRNAs and proteins for the subsequent phases and monitor its internal and external environment to ensure that the conditions are suitable for the next phase. If the tissue has reached its final size or if there are insufficient levels of nutrients, then the cell cycle stops and the cell enters a resting state called $G0$. Cells in $G0$ can re-enter the cycle if conditions change [101]. In Figure 1.1 we show a schematic of the cell cycle phases.

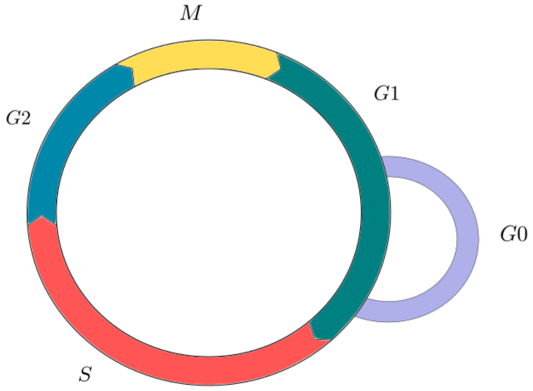


**Figure 1.1:** Schematic of the cell cycle phases. See text for details.

Progression through the cell cycle is regulated by environmental cues and intracellular checkpoints involving various proteins, in particular cyclins and cyclin dependent kinases [144, 228]. At these checkpoints, the cell checks, among other things: that DNA replication occurred, there are sufficent nutrients, and that the DNA is not damaged. There are three main checkpoints in the cell cycle [7]:

1. At the transition from G1 to S, the cell confirms that the cell environment is favorable: there are sufficient nutrients and no apoptotic signals from neighbouring cells.

2. At the transition from the G2 to M the cell checks that the DNA is undamaged and fully replicated.
3. During mitosis when the cell ensures that the duplicated chromosomes are attached to the mitotic spindle.

The duration of the cell cycle differs, depending on the cell type and the developmental state, from as short as 30 minutes to as long as 20 hours [7].

Cell migration comprises the movement of individual cells, cell sheets, or cell clusters from one location to another [90]. Two main types of migration can be distinguished: single cell migration and collective cell migration. Single cell movement has been observed for immune cells, germ cells and cancer metastasis, while collective cell migration is present in development, wound healing, and also in some forms of cancer metastasis [209].

There are two main modes of single cell migration: amoeboid and mesenchymal [76, 145]. Amoeboid migration is characterised by weak (or none) adhesion to the substrate, low (or none) degradation of the extracellular matrix (ECM) and a rounded cell morphology [105]. Hematopoietic stem cells, lymphocytes, leukocytes, dendritic and cancer cells exhibit this type of motility. Mesenchymal migration is characterised by adhesion to the surrounding ECM, ECM degradation and an elongated cell morphology [205]. Mesenchymal migration is a cyclic process in which four stages can be identified: (1) The cell polarises as a response to environmental cues or to internal biochemical changes [151, 227]. During this polarisation, the cell develops a leading and trailing edge [227]. (2) Protrusions extend from the leading edge of the cell. The extension is generally promoted by actin polymerisation. (3) The protrusions adhere to the substrate. (4) The trailing edge of the cell detaches from the substrate and retracts towards the cell body.

During collective cell migration multiple cells perform a coordinated movement regulated by cell-cell adhesion, collective cell polarisation, coordination of cytoskeletal activity and chemical and mechanical cues [77]. This type of migration uses similar mechanisms as in single cell migration (e.g. actin- and myosin-mediated protrusions, cell–extracellular matrix interactions [97]) however, since cells form a cohesive group, these mechanisms operate in a collective fashion. Collective migration can be: two-dimensional across a tissue surface, which is called sheet migration, or through three dimensional tissue scaffolds [77]. 2D cell sheets form a monolayer of cells that moves along tissues to form a single-layered or multilayered epithelium, e.g. the gut intestinal epithelium or the epidermal keratinocytes in wound healing [77]. 3D collective migration can vary widely: for example, vascular sprouting in angiogenesis, invasion of cancer masses to other tissues and bud sprouting during duct and gland formation. A more detailed description of this process can be found in [77, 97].

### 1.1.2 In-vitro research techniques

Proliferation assays focus on measuring the cell number or the proportion of cells that are dividing [182]. The simplest proliferation assay consists of growing a cell monolayer to low density on a two-dimensional substrate and measuring the cell number change with time [31, 35, 182]. See Figure 1.2 (a) for a schematic representation of the typical time progression in proliferation assay.

There are multiple cell migration assays available [13, 162]. Two-dimensional migration assays can be classified depending on how they set a cell-free area in which cells can migrate. In cell-removing assays, cells are first grown to confluence in an assay plate and then cells, from a specific region, are removed. In exclusions methods, cells are excluded from a specific area with a barrier, which is latter removed and cells are able to migrate [13]. Cell-removing assays differ mainly on the way cells are removed: scratch assays (mechanical), electric cell impedance sensing (electrical) and laminar flow (chemical). The biggest disadvantage of cell-removing assays is the damage on the cells and the underlying matrix caused by the removal [13]. An alternative to cell-removing methods are cell excluding methods. Some advantages of these methods are that the matrix in the cell-free area is not damaged and the assay can incorporate matrix complexity such as protein patterns [13]. A disadvantage of cell excluding assays is that the barrier can leave residues on the matrix which can alter the cell behaviour [13].

The most common in-vitro cell migration assay is the scratch or wound healing assay due to the low cost and simplicity of its experimental design [142, 162]. A scratch assay involves: growing a cell monolayer to confluence in a multiwell assay plate; creating a rectangular "wound", a cell-free zone in the monolayer, through a mechanical procedure, into which cells can migrate; and monitoring the recolonisation of the scratched region to quantify the cell migration [142] (see Figure 1.2 (b)).

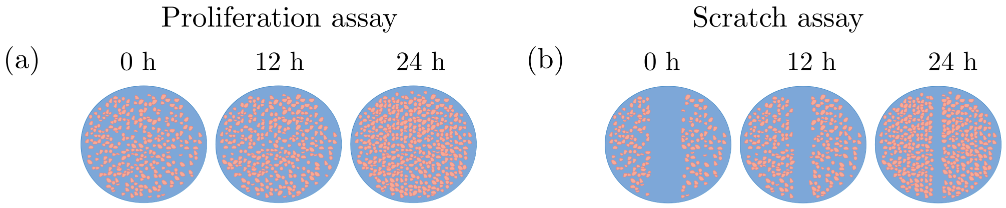


**Figure 1.2:** Schematic representation of how the (a) proliferation and (b) scratch assays evolve in time. Cells are represented by pink circles while the underlying assay plate is depicted in blue.

Typically cell migration assays consider either: cell-free areas (also called voids) shaped as rectangles and circles that are surrounded by a confluent (or near to confluence) assay; or nest assays which are dense cell populations in the shape of a rectangle or circle surrounded by cell-free area [13]. The evolution of these assays are recorded and processed to analyse the migration rates.

## 1.2 Mathematical models of scratch assays

Multiple mathematical models have been developed to describe in-vitro cell dynamics [27, 57, 83, 119, 162, 174, 185, 223, 224]. We first give a brief overview of the different modelling approaches available. We review the mathematical models of cell proliferation and migration, and finally of scratch assays.

Most mathematical models of biological systems fall into two categories: deterministic and stochastic [52, 204]. Deterministic models describe the evolution of the system through continuous concentrations whose evolution is given by differential equations. The key characteristic of determinisitic models is that the evolution of the process is completely determined by the parameter values, the initial condition, boundary conditions and in some cases it depends also on the previous states (e.g. delay differential equations). These models are typically formulated as ordinary, partial and delay differential equations. Deterministic models are accurate when the system consists of a large number of molecules so random fluctuations in the system can be ignored. They also have the advantage of being amenable to mathematical analysis. Furthermore, there exist fast numerical solvers to implement continuum models. In addition, mechanical factors (e.g. stresses and strains within the tissue) control several cellular and molecular processes and are best described by continuum techniques [208]. Also, it is difficult with this approach to describe processes where individual cell effects are important or dominant (e.g. epithelial-mesenchymal transition [46] and tumor angiogenetic sprout branching [114]).

Stochastic models are considered when the role of intrinsic and extrinsic noise in the system needs to be studied. When the number of molecules is low, it is more appropiate to model the evolution of the number of molecules and this is the approach of discrete stochastic models. A class of stochastic models are continuous-time Markov chain models [102]. The molecular population numbers are assumed to be random variables and the system state is modelled as a Markov process [229]. When the number of events in a given time interval is large, then the system can be modelled through Langevin equations: a type of stochastic differential equations [86]. If spatial effects are important, three frameworks are typically used: the Doi model, the Smoluchowski diffusion limited reaction model, and the reaction-diffusion master equation (RDME) [111]. We describe in more detail Markov chain models in Chapter 2, Section 2.4. Models that focus on a much smaller scale are Brownian, and molecular dynamics. Typically these approaches are considered for simulating very short time periods, nanoseconds to milliseconds.

Another important class of modelling approaches are agent-based models (ABMs). Agent-based models (also called individual-based models) are discrete-event and discrete-time modelling approaches [8, 158]. Individual components and agents of the system are given rules to interact with each other. ABMs can incorporate detailed descriptions of each agent dynamics and can easily incorporate experimentally inspired rules based on biomedical data or biomedical observations [158]. The biggest drawback of this technique is the large computational demand and the lack of mathematical techniques to determine the emergent behaviour [52]. There are different types of agent-based models, and we refer the reader to reviews on this topic [158, 167].

Hybrid models are more recent modelling frameworks. Hybrid modelling is characterised by the coupling of different modelling formalisms (deterministic and stochastic or different stochastic models) [196, 199]. A popular class of hybrid models are *multiscale models* in which different modelling frameworks are used to represent the dynamics at different spatial scales of the same biological system. The output of one modelling framework is used as the input of another and the coupling is typically bidirectional [199]. Another class of hybrid models are *spatially coupled hybrid models* in which multiple modelling frameworks are used in different regions of space for representing the same physical quantity. The different frameworks are commonly coupled together through an interface or overlap region [196]. In Chapter 2, Section 2.4.5 we describe in more detail spatially coupled hybrid models.

The choice of framework is typically determined by the aspects of the biological system we are interested in capturing [204]. Each modelling framework has its advantages and disadvantages and none can be said to be better than the rest but just more suitable for a certain problem. The biological question and the available experimental data are some of the factors that determine which modelling framework is most appropriate [82].

We describe now the different modelling approaches for describing cell proliferation. The simplest ones focus on the cell number changes as a result of cell proliferation. Typically these models describe the rate of change in the cell number by an ordinary differential equation (ODE). These models are also referred to as *continuous growth models* for which a conservation equation is assumed for the cell population. The simplest growth law is exponential growth, however in most cases it is unrealistic to assume exponential growth due to the constraints of nutrients and space in a given system. There are other growth laws that include more biological detail like Gompertz, logistic and Malthus laws [84]. Denoting $N(t)$ as cell population size at time $t$, $N(t)$ follows logistic growth if its dynamics is described by the following ODE:

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right), \tag{1.1}$$

where $r$ is referred to as the growth rate and $K$ is the carrying capacity. The carrying capacity represents the maximum population size that the system can support. It captures the limited space and resources in an environment. We refer the reader to [54, 84] for reviews that describe the different growth laws.

Another approach for modelling cell proliferation is integrating the molecular network that regulates cell cycle progression. Key works using this approach are by John J. Tyson and Béla Novák who performed extensive reasearch on the molecular networks that regulate the cell cycle of frog egg extracts, frog embryos and fission yeast cells [44, 45, 164, 213]. Logical dynamical modelling has also been consider to show the robustness and stability of the yeast cell cycle [32, 99, 140].

Researchers have also considered physiologically structured populations to capture cell proliferation [65, 78]. Typically mathematical models assume that all cells in a cell population are the same, however cells have individual physiological characteristics, like the time since division and size, that affect their proliferation. These types of models can describe the role of heterogeneity in these physiological traits and their effects on the population dynamics. We refer the reader to [157, 176] for more detail about size-structured models but we shall focus on age-structured models. We refer to a cell's *age* as the cell's temporal position within the cell cycle.

The McKendrick-Von Foerster equation is a linear first-order partial differential equation (PDE) typically used in population models to incorporate the age-structure of the population. It was first introduced by Heinz Von Foerster [74]. This type of PDE is derived using a conservation law accounting for the changes in the density of the population caused by aging, fertility and mortality. The mathematical properties of the McKendrick-Von Foerster equation have been extensively investigated over the past 60 years [11, 12, 47, 95, 160, 210, 211, 226] and extensions of the model have been developed to include, for example, nonlinear interactions, sex-structure, juvenile-adult dynamics, epidemic outbreaks, etc [133]. In Chapter 2, Section 2.1 we describe some of the mathematical properties of this equation.

By considering an age-structured model we can investigate how heterogeneity in the cell population age distribution affects the total cell population dynamics. In [117] it was shown that heterogeneity in cell age distribution may still generate logistic growth at the population scale. Other age-structured models have described how the cell age distribution affects the efficacy of anti-cancer drugs, particularly phase-specific drugs [26] and how growth factors regulate the duration of some phases in the cell cycle [27].

Similarly, most stochastic models of cell proliferation can also be classified in three types depending on what they focus on. Stochastic models that focus on cell number are mostly expressed as a continuous-time Markov birth-death process [19, 206]. Depending on the functional form of their transition rates, they can recreate the different growth laws macroscopically [206]. Other stochastic models of cell proliferation focus on progression through the cell cycle and the underlying subcellular regulatory networks [21, 124, 200]. These types of models allow the study of intrinsic noise in the signaling network (e.g. [21, 203]). Similarly as in the deterministic case there are also physiological structured models: age-structured birth-death processes and stochastic size dependent population models, we refer the reader to [19, 53, 91, 202] for more information about these approaches.

Regarding mathematical modelling of cell migration, both deterministic and stochastic models have been considered to model cell migration [17, 23, 48, 61, 82, 107, 169, 171]. Continuum models consider a macroscopic (tissue-level) phenomenological description of cell migration. Assuming that cells move randomly in space, one can show that the macroscopic description consists of a diffusion process [48, 64, 161]. Different extensions of the diffusion equation have been developed to capture different characteristics of cell migration: directed cell migration in response to chemical gradients (chemotaxis) [107, 125, 171], interaction with the ECM (haptotaxis, contact guidance) [43, 171] and interactions with other cells (cell-cell adhesion or contact guidance) [171, 212].

Stochastic models can be classified into lattice-based and off-lattice approaches [23, 48]. The most common ones are lattice-based random walk models in which cell motility events are represented by jumps to the nearest lattice neighbours and the lattice spacing is typically identified as the cell diameter [177, 194]. This framework is very popular since it is easy to implement, several motility aspects can be included, and there exist limit procedures to determine the corresponding mean-field approximation [172, 173]. Crowding effects can be added by allowing only one cell to occupy a lattice site and aborting a motility event if the lattice site is occupied [62, 109]. Chemotaxis can be modelled through a biased random walk assigning unequal probabilities to the jump rates to the nearest-neighbours [48]. We refer the reader to [104, 168] for more information about off-lattice approaches.

Another class of models of cell migration are mechanochemical models in which the tissue is represented as an active continuous media: viscoelastic solid or fluid [17]. These models typically incorporate contractility, cell polarisation, other forces driving collective cell motion [139, 231], and the interplay between mechanical stresses and the regulatory biochemistry [60]. Discrete stochastic models with mechanical effects are reviewed in [218].

We finally focus on the modelling of scratch assays and specifically on the models that include cell migration and proliferation. The most widely used deterministic model for scratch assays is the Fisher-Kolmogorov-Petrovsky-Piscounov (FKPP) equation [119, 150, 174, 185, 189, 191]. In this model cell migration is modelled as Fickian diffusion and cell proliferation is modelled by a logistic source term. This model has been widely used: to study the relationship between the diffusion coefficient, proliferation rate and the speed of an invading cell front [150], to quantify the effects of growth factors in cell migration and cell proliferation [119], to analyse the effect of experimental design in scratch assays [116, 174] to name only a small number of applications. Different extensions have been considered: non-constant coefficients [56, 98], non-linear diffusion to model contact inhibition [39] and the diffusivity factor dependent on external chemical factor [58], to name but a few.

An alternative approach has been to describe scratch assays by stochastic models [40, 48, 121, 169, 194]. Since experimental data from scratch assays involve individual cells, predictions of discrete models can be directly compared with data. The most common stochastic models are lattice-based random walk models [120, 121, 174, 194]. Different aspects of scratch assays have been analysed with these models: how the initial geometry of the wound affects the wound closure rates [116], how the leading edge evolution can help determine the migration and proliferation rates [120], and how contact inhibition affects migration [38]. We refer the reader to [10, 154, 201] for more information about other stochastic models of scratch assays.

## 1.3 Main results and contributions of the thesis

In this thesis we develop and analyse three mathematical models that focus on different aspects of scratch assays.

We first introduce a new migraton quantification method for scratch assays that can handle low-quality experimental data (irregular cell-free areas and crooked leading edges). We consider an agent-based model to produce in-silico data so we can perform a robust analysis and validation of the quantification method considering different proliferation and migration rates and different degrees of irregularity in the data. By analysing our quantification method through in-silico data, we are able to show that our method provides better statistical classification of migration rates than standard quantification methods. We additionally validate our quantification method using in-vitro data.

We then introduce an age-structured population model that accounts for the two phases of proliferation in scratch assays. In a recent work [118], Jin et al. observed experimentally that during the early stage of the scratch assay, there is a *disturbance phase* where proliferation is not logistic, and this is followed by a *growth phase* where proliferation appears to be logistic. The authors did not identify the precise mechanism that causes the disturbance phase but showed that ignoring it can lead to incorrect parameter estimates. Our aim is to show that a nonlinear age-structured population model can capture this dynamics. The model consists of an age-structured cell cycle model of a cell population, coupled with an ordinary differential equation describing the resource concentration dynamics in the substrate. The model assumes a resource-dependent cell cycle threshold age, above which cells are able to proliferate. By analysing our model and performing numerical simulations, we are able to show that the interplay between the resource concentration in the substrate and the cell subpopulations dynamics can explain the biphasic dynamics.

Finally we introduce a hybrid algorithm that couples a standard stochastic model for scratch assays, the stochastic description of the Fisher-Kolmogorov-Petrovsky-Piscounov (FKPP) model, with its mean-field analogue. Typically hybrid algorithms consider heuristic methods based on the cell density for determining which compartments will be updated deterministically or stochastically. We consider instead a natural indicator of when the mean-field approximation is valid: the size of fluctuations. We introduce a new hybrid algorithm that estimates the system variance a priori and uses this estimate to determine which compartments are updated deterministically or stochastically. We apply this new hybrid algorithm to the stochastic FKPP model, analyse how well it approximates the stochastic model and compare its performance to a heuristic hybrid algorithm presented in [149].

Some of the work presented in this thesis has been published in international peer reviewed journals:

- Ponce Bobadilla, A. V., Arévalo, J., Sarró, E., Byrne, H. M., Maini, P. K., Carraro, T., Balocco, S. & Alarcón, T. (2019). In-vitro cell migration quantification method for scratch assays. Journal of the Royal Society Interface, 16(151).
- Ponce Bobadilla, A. V., Carraro, T., Byrne, H. M., Maini, P. K., & Alarcón, T. (2019). Age Structure Can Account for Delayed Logistic Proliferation of Scratch Assays. Bulletin of Mathematical Biology, 1-19.

## 1.4 Thesis structure

The thesis is organised as follows.

In Chapter 2, we introduce the mathematical methods that are used in the rest of the thesis. We first describe standard results from age-structured models and delay differential equations. We then introduce the basic terminology and notations for finite element discretisation of partial differential equations. We describe the method of manufactured solutions. We end the chapter by introducing the basic principles of stochastic mathematical modelling of chemical reaction networks.

Chapter 3 focuses on introducing the new migration quantification method for scratch assays. We introduce the in-vitro and in-silico data that we consider to validate the quantification method. We introduce a new migration quantification method for scratch assays and describe the two quantification methods to which we compare it: the *percentage wound area method*, and the *closure rate method*. We then analyse our method and compare it against the other quantification methods using in-silico data. We then apply the three methods to in-vitro data.

Chapter 4 focuses on introducing an age-structured model that accounts for the two phases of proliferation in scratch assays. We first introduce the nonlinear age-structured model. We describe the dynamics of the full model in terms of the mature and immature subpopulation dynamics. We then derive conditions under which the per capita growth rate will exhibit biphasic behaviour and logistic-type proliferation. We investigate numerically these conditions and determine the parameter regimes in which the resource concentration regulates the biphasic dynamics.

Chapter 5 focuses on introducing a hybrid algorithm based on an apriori variance estimation. We first introduce the stochastic FKPP model and derive its mean-field and the coarse-grained factorial cumulant equations. We then analyse numerically the discretisation schemes to solve both equations. We describe a new hybrid algorithm that consider the system variance to determine which compartments will be updated determinstically or stochastically. We analyse the performance of the hybrid algorithm applied to the stochastic FKPP model and compare its performance to the one presented in [149].

Finally, in Chapter 6, we discuss the contributions made in this thesis, we give an overview of the future research directions and open questions that have arisen from the work in this thesis.

# Chapter 2

# Methods

In this chapter we introduce the theoretical background that is needed to understand the material presented in the rest of the thesis.

## 2.1 Age-structured models

We present some general results that we use in Chapter 4, where we will consider the dynamics of a cell population described by a McKendrick-Von Foerster type equation. Here, we introduce the mathematical properties of this type of equation that are relevant to the material presented in this thesis.

The McKendrick-Von Foerster equation considers the cell density $n(a,t)$ as a function of age $a$ and time $t$. Let us consider the time interval $[0,T]$ with $T > 0$. Assuming cells have an age- and time-specific fertility and death rate, the cell density function $n(a,t) : [0,\infty] \times [0,T] \to \mathbb{R}^+$ can be shown to satisfy the following equation:

$$\frac{\partial n(a,t)}{\partial t} + \frac{\partial n(a,t)}{\partial a} = -\beta(a,t)n(a,t), \tag{2.1a}$$

$$n(0,t) = \int_0^\infty \gamma(a,t)n(a,t)\mathrm{d}a, \tag{2.1b}$$

$$n(a,0) = v_0(a), \tag{2.1c}$$

where $\beta(a,t)$ is the age- and time-specific death rate and $\gamma(a,t)$ is the age- and time-specific fertility rate of the population. Both $\beta(a,t)$ and $\gamma(a,t)$ are assumed to be non- negative functions. The initial cell density (at time $t = 0$) is the non-negative function $v_0(a)$.

The standard method for finding a solution to Eq. (2.1) is to note the existence of special curves in the $(a,t)$-plane, called characteristics, along which Eq. (2.1) reduces to an ordinary differential equation (ODE) [127]. This method is known as *the method of characteristics* and we introduce it in the next subsection.

### 2.1.1 Method of characteristics

**Theorem 1.** *The solution of Eq.* (2.1) *is given by:*

$$n(a,t) = \begin{cases} v_0(a-t)\exp\left(-\int_0^t \beta(a-t+s,s)\mathrm{d}s\right) & \textit{for } a \geq t, \\ n(0,t-a)\exp\left(-\int_0^a \beta(s,t-a+s)\mathrm{d}s\right) & \textit{for } a < t. \end{cases} \tag{2.2}$$

*Proof.* Let $X(t)$ be a parametrisation of a curve in the $(a,t)$-plane and let us consider $U(t) = n(X(t),t)$. The derivative in time of $U(t)$ is given by:

$$\frac{d}{dt}U = \frac{\partial n}{\partial X}\frac{\partial X}{\partial t} + \frac{\partial n}{\partial t}. \tag{2.3}$$

If we consider a curve such that:

$$\frac{dX}{dt} = 1, \qquad X(t_0) = x_0, \tag{2.4}$$

then Eq. (2.1a) along $X(t)$ is reduced to:

$$\frac{d}{dt}U = -\beta(X(t),t)U. \tag{2.5}$$

$X(t)$ is called the *characteristic curve* of Eq. (2.1a). Solving Eq. (2.4) for $X(t)$, we obtain an explicit formula for the characteristic curve:

$$X(t) = t - t_0 + x_0. \tag{2.6}$$

We notice that the characteristics are lines with slope 1 and that the PDE (2.1a) along these lines reduces to the ODE (2.5). Therefore it is sufficient to know the value of the solution at one point of each characteristic curve to solve the ODE on each curve and therefore solve the problem in the whole domain. To determine the solution of the ODE along the characteristic curves, we consider the intercept, $\zeta$, of the characteristic curve with respect to the $a$-axis, $\zeta = -t_0 + x_0$. There are two cases: $\zeta \geq 0$ or $\zeta < 0$. These two cases are shown in Figure 2.1. Since we are interested in points where the solution is already defined, we notice that due to the boundary conditions, the solution is defined at the points of each axis. We plot the solution value for the points on each axis for points on the a-axis, i.e. $n(a,t) = v_0(a)$ for $(a,t) = (a,0)$ and for points on the t-axis, i.e. $n(a,t) = n(0,t)$ for $(a,t) = (0,t)$.

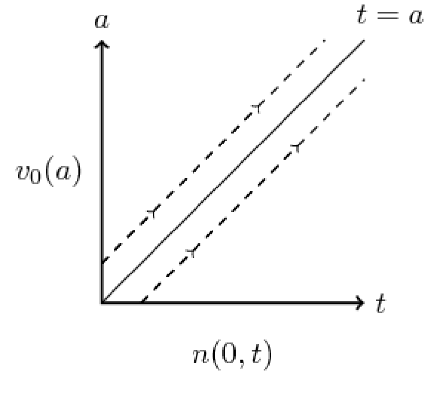


**Figure 2.1:** Plot of the characteristics of Eq. (2.1) in the plane $(a,t)$. We denoted the value of $n(a,t)$ for the points on the $a$-axis by $v_0(a)$ and those on the $t$-axis by $n(0,t)$.

For $\zeta \geq 0$, we consider as an initial position for the characteristic, the intercept is with the $a$-axis. $U(t) = n(X(t),t)$ then satisfies

$$\frac{d}{dt}U = -\beta(X(t),t)U \qquad U(0) = U(-t_0 + x_0, 0) = v_0(-t_0 + x_0). \tag{2.7}$$

Hence

$$U(t) = v_0(-t_0 + x_0) \exp\left(-\int_0^t \beta(X(t), t))\mathrm{d}s\right) \quad \text{for } t \geq 0. \tag{2.8}$$

As $U(t_0) = n(x_0, t_0)$ and $X(t) = t - t_0 + x_0$, we have:

$$n(x_0, t_0) = v_0(-t_0 + x_0) \exp\left(-\int_0^{t_0} [\beta(s - t_0 + x_0, s)]\,\mathrm{d}s\right) \quad x_0 \geq t_0. \tag{2.9}$$

Let $a = x_0$ and $t = t_0$, then we have that for $a \geq t$,

$$n(a, t) = v_0(a - t) \exp\left(-\int_0^t [\beta(s - t + a, s)]\,\mathrm{d}s\right). \tag{2.10}$$

If $\zeta < 0$, then we consider as the initial condition the intercept with the $t$-axis,

$$U(-x_0 + t_0) = n(X(-x_0 + t_0), -x_0 + t_0) = n(0, -x_0 + t_0). \tag{2.11}$$

Solving the ODE (2.5) with the initial condition (2.11), we obtain

$$U(t) = n(0, -x_0 + t_0) \exp\left(-\int_{-x_0+t_0}^t [\beta(X(s), s)]\,\mathrm{d}s\right) \quad \text{for } t \geq -x_0 + t_0. \tag{2.12}$$

Hence, using a similar argument to before, we have

$$n(x_0, t_0) = U(t_0) = n(0, -x_0 + t_0) \exp\left(-\int_{-x_0+t_0}^{t_0} [\beta(s - t_0 + x_0, s)]\,\mathrm{d}s\right) \quad \text{for } t_0 > x_0. \tag{2.13}$$

Setting $a = x_0$ and $t = t_0$, then for $t > a$,

$$\begin{aligned} n(a, t) &= n(0, t - a) \exp\left(-\int_{t-a}^t [\beta(s - t + a, s)]\,\mathrm{d}s\right) \\ &= n(0, t - a) \exp\left(-\int_0^a [\beta(s, s + t - a)]\,\mathrm{d}s\right). \end{aligned}$$

□

### 2.1.2 Long-time dynamics

In order to study the long-time dynamics of an age-structured model, it is necessary to introduce a special class of solutions of age-structured models: *stable age distributions* [226]. These solutions have the form:

$$n(a, t) = A(a)T(t) \tag{2.14}$$

where $A(a)$ and $T(t)$ are functions that depend only on $a$, $t$, respectively.

This class of solutions is called "stable" because while the total size of the population can change, the proportion in any age bracket $[a_1, a_2]$ remains constant, i.e,

$$\frac{\int_{a_1}^{a_2} n(a, t)\mathrm{d}a}{\int_0^\infty n(a, t)\mathrm{d}a} = \eta \tag{2.15}$$

for a constant $\eta \in \mathbb{R}$.

The following theorem gives the general form of this class of solutions.

**Theorem 2.** *Assuming the fertility and the death rate, $\beta$ and $\gamma$, are independent of time and that the solution of Eq.* (2.1) *is a stable age-distribution (Eq.* (2.14)*) then $n(a,t)$ is given by:*

$$n(a,t) = n(0,0)\exp\left[\lambda(t-a) - \int_0^a \beta(s)\mathrm{d}s\right], \tag{2.16}$$

*where $\lambda \in \mathbb{R}$.*
*In addition, $\beta(a)$, $\gamma(a)$ and $\lambda$ satisfy the following equation:*

$$1 = \int_0^\infty \gamma(a)\exp(-\lambda a - \int_0^a \beta(s)\mathrm{d}s)\mathrm{d}a, \tag{2.17}$$

*which is called the Lotka characteristic equation.*

*Proof.* By considering the solution of the PDE to be of the form (2.14) we find that $T(t)$ and $A(t)$ need to satisfy:

$$\frac{\frac{\mathrm{d}T}{\mathrm{d}t}}{T(t)} = -\left[\frac{\frac{\mathrm{d}A}{\mathrm{d}a} + \beta(a)A(a)}{A(a)}\right] \tag{2.18}$$

For this expression to hold, both sides have to be equal to the same constant $\lambda \in \mathbb{R}$. Solving for $T(t)$ we obtain:

$$T(t) = T(0)\exp(\lambda t). \tag{2.19}$$

Solving for $A(a)$ we obtain

$$A(a) = A(0)\exp\left[-\lambda a - \int_0^a \beta(s)\mathrm{d}s\right]. \tag{2.20}$$

Hence, from Eq. (2.14),

$$n(a,t) = n(0,0)\exp\left[\lambda(t-a) - \int_0^a \beta(s)\mathrm{d}s\right]. \tag{2.21}$$

Since the solution (2.21) also satisfies the boundary condition given by Eq. (2.1b) for all $t \in [0,T]$, we have that in particular for $t = 0$:

$$\begin{aligned} n(0,0) &= \int_0^\infty \gamma(a)n(a,0)\mathrm{d}a \\ n(0,0) &= \int_0^\infty \gamma(a)n(0,0)\exp\left[\lambda(0-a) - \int_0^a \beta(s)\mathrm{d}s\right]\mathrm{d}a \end{aligned}$$

and, given that $n(0,0) \neq 0$, the Lotka characteristic equation is satisfied. □

Theorem 2 describes an important feature of stable-age distributions. Since the solution in time is an exponential function, then the population density either grows, decreases exponentially or stays constant depending on the value of $\lambda$.

The next theorem states the importance of stable-age distributions for general age-structured models, specifically regarding the long-time dynamics of $n(a,t)$.

**Theorem 3** (Long time behaviour for age-structured models). *If there exists $\lambda \in \mathbb{R}$ that satisfies the Lotka characteristic equation:*

$$1 = \int_0^\infty \gamma(a) \exp(-\lambda a - \int_0^a \beta(s)\mathrm{d}s)\mathrm{d}a \tag{2.22}$$

*then the solution of the age-structured model (Eq.* (2.1)*) approaches a stable age-distribution:*

$$n(a,t) \sim e^{\lambda t} \exp\left[(-\lambda a - \int_0^a \beta(s)\mathrm{d}s\right] \qquad as \quad t \to \infty \tag{2.23}$$

*Proof.* The proof can be seen in [106]. □

**Remark 1.** *It is important to notice that Theorem 3 is valid for a general class of age-structured models, while Theorem 2 is valid for age-structured model whose solution is separable.*

## 2.2 Delay differential equations

In this section we introduce standard results for delay differential equations (DDEs) and we focus, in particularly, on results regarding the presence of oscillatory behaviour. We only introduce those concepts that are necessary to understand the material presented in Chapter 4.

A delay differential equation is a type of differential equation in which the derivative of the unknown function at a certain time, $\frac{\mathrm{d}x}{\mathrm{d}t}(t)$, does not only depend on the instantaneous value of the function, $x(t)$, but also on the function at previous times, $x(t-\tau)$ with $\tau > 0$ [4]. That is, DDEs are differential equations of the following form:

$$\frac{\mathrm{d}x}{\mathrm{d}t}(t) = f(t, x(t), x(t-\tau)), \quad \text{with} \quad \tau > 0 \quad \text{and} \quad t \geq t_0 \tag{2.24}$$

where $\tau$ is referred as a discrete delay.

The mathematical theory of DDEs has been extensively developed. Standard references of the topic are [70, 197]. We focus on linear delay differential equations and a standard stability result.

Let us consider a linear delay differential equation with constant coefficients and constant delay:

$$\frac{dx}{dt} = Ax(t) + Bx(t-r) \tag{2.25}$$

where $A, B$ and $r$ are scalars.

Assuming a solution of the form $e^{\lambda t}$, we obtain the characteristic equation for this equation:

$$h(\lambda, r) = \lambda - A - Be^{-\lambda t}. \tag{2.26}$$

We assume $\lambda \in \mathbb{C}$ and the values of $\lambda$ for which $h(\lambda, r) = 0$ are referred to as the characteristic roots of the DDE (2.25). The values of the real parts of the characteristic roots determine the stability of the solution.

The stability analysis of Eq. (2.25) is a standard result of DDE theory (eg. Theorem 4.7 p.53 in [197]), which we state here:

**Theorem 4** (Stability of linear delay differential equations). *Let us consider a linear delay differential equation as given by Eq* (2.25)*. Assume that* $A + B \neq 0$ *(otherwise* $\lambda = 0$ *is a root). The following results hold for Eq.* (2.25)*:*

- *If* $A + B > 0$*, then* $x = 0$ *is unstable.*
- *If* $A + B < 0$ *and* $B \geq A$*, then* $x = 0$ *is asymptotically stable.*
- *If* $A + B < 0$ *and* $B < A$*, then there exists* $r^* > 0$ *such that* $x = 0$ *is asymptotically stable for* $0 < r < r^*$ *and unstable for* $r > r^*$*.*

## 2.3 Discretisation of partial differential equations

In this section we introduce the basic terminology and notation that is used for finite element discretisation of partial differential equations (PDEs). We only introduce those concepts that are necessary to understand the material presented in Chapter 4 and 5 and refer the reader to standard reference books on finite elment methods for more in depth descriptions [33, 94].

Let $\Omega \subset \mathbb{R}^d$, for $d = 1, 2$, be an open bounded domain with boundary $\partial\Omega$. Throughout the thesis we consider second order PDEs of variables, which are defined in a space and time domain, i.e. functions $u(x,t) : \Omega \times [0,T] \to \mathbb{R}$ with $T > 0$. The *classical formulation* of a PDE problem consists of the PDE together with its initial and boundary conditions. A standard method to approximate the solution of the classical formulation of a PDE is to restate the problem in an integral form, called the *weak formulation*. We now introduce the functional spaces that are used to define the weak formulation.

We consider the following functional space

$$L^2(\Omega) := \{f : \Omega \to \mathbb{R} : f \text{ is measurable and } |f|^2 \text{ is Lebesgue integrable in } \Omega\} \tag{2.27}$$

where the associated inner product and norm are denoted by, respectively,

$$(f,g)_\Omega = \int_\Omega fg\mathrm{d}x, \quad \|f\|_{L^2(\Omega)} := \sqrt{(f,f)_\Omega}. \tag{2.28}$$

We sometimes omit the domain index if it is clear from the context.

Another functional space we consider is the space of locally integrable functions. The set of all such functions is denoted by $L^1_{loc}(\Omega)$:

$$\begin{aligned} L^1_{loc}(\Omega) := \{ f : \Omega \to \mathbb{R} : f &\text{ is measurable and } f|_K \text{ is Lebesgue integrable in } K \\ &\forall\, K \in \Omega,\ K \text{ compact}\}\,. \end{aligned} \tag{2.29}$$

A locally integrable function is weakly differentiable with respect to $x_i$ if there exists a function $g_i \in L^1_{loc}(\Omega)$ such that

$$\int_\Omega f \frac{\partial \varphi}{\partial x_i}\mathrm{d}x = -\int_\Omega g_i \varphi \mathrm{d}x \quad \forall \varphi \in C_c^\infty(\Omega), \tag{2.30}$$

The function $g_i$ is called the weak $i$-th partial derivative of $f$.

We mainly consider the Sobolev space $H^1(\Omega)$ which consists of locally integrable functions in $\Omega$ whose weak derivative belongs to $L^2(\Omega)$.

The weak formulation of a PDE is constructed by multiplying the PDE by functions from an appropriate functional space, which we refer to as the "test space", and integrating the resulting expression over the domain. The function in the appropriate functional space, which we refer to as the "solution space", that fulfils this new integral expression is known as the weak solution.

The finite element method involves approximating the weak solution $u$ by a solution $u_h$ of the weak formulation in a finite-dimensional subspace of the solution space $V_h$. If we assume that $V_h$ is an $n$-dimensional vector space of functions for which $\{\phi_i\}_{i=1}^n$ are the basis functions then, for $u_h \in V_h$, there exist coefficients $U_1, \dots, U_n$ such that $u_h = \sum_{i=1}^n U_i \phi_i$. By considering the basis functions of $V_h$, the weak formulation in $V_h$ is reduced to solving an algebraic finite-dimensional system with respect to the basis functions.

The basis functions of the space $V_h$ are typically constructed with respect to a mesh of the domain $\Omega$. A mesh of $\Omega$ can be obtained by partitioning the domain $\Omega$ into a series of elements $K_j$ with $1 \leq j \leq N$ such that $\Omega = \bigcup_{j=1}^N K_j$. We denote this partition by $\mathcal{T}_h$. We assume $K_j = T_{K_j}(E) \in \mathcal{T}_h$ in which $E = [0,1]^d$ is the reference element, and $T_{K_j}(E)$ is an invertible affine map. We assume the mesh is regular and, hence, must satisfy the following conditions:

(i) $\overline{\Omega} = \cup_{K \in \mathcal{T}_h} \overline{K}$

(ii) $K \cap K' = \emptyset$ for all elements $K, K' \in \mathcal{T}_h$ with $K \neq K'$

Given an element $K$ of a mesh $\mathcal{T}_h$, its *diameter* is defined as

$$h_K := \sup\left\{|x-y|, x, y \in K\right\}. \tag{2.31}$$

We denote the maximum diameter of the elements of the mesh as

$$h := \max_{K \in \mathcal{T}_h} h_K. \tag{2.32}$$

We consider as the solution space, the finite dimensional space $V_h \subset H^1(\Omega)$ where $V_h$ denotes the finite element space such that in each element of the mesh, $K \in \mathcal{T}_h$, the function reduces to a polynomial of *partial degree* at most $k \geq 1$:

$$V_h := \left\{v_h \in C^0(\overline{\Omega}) : \ v_h|_K \in \mathbb{Q}_k(K) \ \forall K \in \mathcal{T}_h\right\}, \tag{2.33}$$

where $\mathbb{Q}_k(K)$ denotes this set of polynomials:

$$\mathbb{Q}_k(K) := \{p \in C^0(\overline{\Omega}) : p(x) = \sum_{\substack{1 \leq i \leq d \\ \alpha_i \leq k}} a_i x_i^{\alpha_i}\}. \tag{2.34}$$

In this thesis we consider the solution space $\mathbb{Q}_k$ with $k = 1$. The standard basis functions for this space are known as the nodal functions: functions that take the value 1 at a specific node and are zero at all other nodes of the mesh.

We use the method of manufactured solutions (MMS) to verify that our numerical schemes solve the PDE accurately and that mesh and time step refinement reduces the numerical error at the expected rate [15, 181, 184].

Implementation of the MMS involves the following steps [15]:

1. Consider a PDE with boundary and initial conditions.
2. Fix a domain to solve the PDE and an initial mesh and time step.
3. Consider the domain in which the PDE is defined and consider a function in this domain.
4. Determine a source term, additional boundary terms and appropriate initial conditions of the PDE in step 1 so that the function defined in step 3 is the analytical solution.
5. Perform mesh and time convergence studies and quantify the observed order of convergence $r_{obs}$.
6. Compare the observed ($r_{obs}$) and the theoretical ($r_{theor}$) convergence rates of the discretisation scheme. If $r_{obs} \approx r_{theor}$ then the discretisation method is verified to be consistent.

If the discretisation method is appropriate, then the discretisation error will depend on $h$ where $h$ is defined in Eq. (2.32). For consistent discretisation methods, the discretisation error can be described as follows

$$e_h = \|u - u_h\|_{L^2(\Omega)} = ch^{r_{theor}} + \gamma(h^w, k). \tag{2.35}$$

where $r_{theor} > 0$ is referred to as the theoretical order of the method and $\gamma(h^w, k)$ contains either higher order terms of $h$ or terms that can be neglected [184].

To calculate the observed convergence rate, we rely on the assumption that as $h$ decreases to zero, $e_h \approx ch^r_{theor}$. We then make successive mesh refinements and quantify the observed convergence rate. Let us consider a mesh and its refinement: the original mesh has grid size $h$ and its refinement has grid size $h/2$. According to Eq. (2.35) and assuming $\gamma$ can be neglected, the discretisation errors corresponding to these two grids have the form:

$$e_h \approx ch^{r_{theor}}, \tag{2.36}$$

and

$$e_{\frac{h}{2}} \approx c\frac{h}{2}^{r_{theor}}, \tag{2.37}$$

so that

$$\frac{e_h}{e_{\frac{h}{2}}} \approx 2^{r_{theor}} \tag{2.38}$$

The observed convergence rate $r_{obs}$ is computed as

$$r_{obs} \approx \log\left(\frac{e_h}{e_{\frac{h}{2}}}\right) / \log(2). \tag{2.39}$$

## 2.4 Deterministic and stochastic modelling of chemical reaction networks

In this section we introduce the basic principles of stochastic mathematical modelling and explain briefly how to simulate chemical reactions. We focus on the concepts that are needed to understand the material presented in Chapter 5 and refer to the following references for details [9, 68, 80, 108, 132, 204, 216, 229].

Mathematical models typically abstract and simplify the complexity of the biological system. Instead of modelling all the mechanisms involved in the biological system explicitly, one may assume that the biological system consists of a series of discrete events [187]. These events are represented as *chemical reactions* and we use the standard abstract notation to represent them. The substances that participate in the reaction can represent proteins, DNA, RNA, chemical molecules, animals or cells. For example, the process in which a cell undergoes mitosis (i.e. cell duplication) can be represented by the following reaction:

$$A \xrightarrow{k} A + A.$$

Before continuing we introduce the standard nomenclature for chemical reaction networks, following [187].

A chemical reaction system consists of $N$ chemical species, $X_1, \ldots, X_N$ that interact via $R$ chemical reactions. We denote the $j$-th reaction as:

$$r_{1,j}X_1 + r_{2,j}X_2 + \ldots + r_{N,j}X_N \xrightarrow{k_j} p_{1,j}X_1 + \ldots + p_{N,j}X_N \quad \text{for } j = 1, \ldots, R, \tag{2.40}$$

where $r_{i,j}$ and $p_{i,j}$ are non-negative integers denoting the number of reactant and product molecules, respectively. The quantity $k_j$ in Eq. (2.40) is called the *rate constant* of the $j$-th reaction. The $j$-th reaction is of "order m" if it involves $m$ reactant molecules, i.e. $\sum_{i=1}^{N} r_{ij} = m$. If $m = 2$, then the reaction is called "bimolecular" and if $m = 0$ or 1 for all the reactions, then the system is called "linear". A chemical reaction system is generally assumed to be closed with volume $V$.

The ***stoichiometric matrix*** **S** is an $(N \times R)$ integer-valued matrix, $\mathbf{S} \in M_{N\times R}(\mathbb{Z})$, whose $S_{ij}$ entry is given by:

$$S_{ij} = p_{ij} - r_{ij}, \quad \text{for} \quad i = 1, \ldots, N, \quad j = 1, \ldots, R. \tag{2.41}$$

The entry $S_{ij}$ corresponds to the net molecule change of species $X_i$ when the $j$-th reaction occurs.

**Example 1.** *Consider the following reaction system which involves two reactions* $(R_1, R_2)$ *and three species* $(A, B, C)$*:*

$$R_1 : A + B \to 2C, \qquad R_2 : B + C \to 2B$$

*The corresponding stoichiometric matrix* $\mathbf{S} \in M_{3\times 2}(\mathbb{Z})$ *is:*

$$\begin{pmatrix} -1 & 0 \\ -1 & 1 \\ 2 & -1 \end{pmatrix}.$$

### 2.4.1 Deterministic rate equations and chemical master equation

Typically two assumptions are made for chemical reaction networks [108]:

1. The system is well-mixed which means that the reactants are uniformly distributed spatially in the system and that diffusion of the chemical species is so fast that it can be neglected.
2. Each species is abundant and can be described by a concentration that varies continuously in time.

Under these assumptions the classical model of a chemical reaction system is a set of reaction rate equations that follow the *law of mass action*. This theory was developed by the chemists Cato Maximilian Guldberg and Peter Waage between 1864 and 1879 [222]. The law of mass action states that "the rate of a chemical reaction is proportional to the product of the concentration of the reactants".

Let $\phi_i$ denote the concentration of molecules of species $X_i$ and let us consider the concentration vector $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_N)^T$. Following the law of mass action, the *rate function* of reaction (2.40), $f_j(\boldsymbol{\phi})$, is given by:

$$f_j(\boldsymbol{\phi}) = k_r \prod_{i=1}^{N} \phi_i^{r_{ij}}. \tag{2.42}$$

The rate at which the concentration $\phi_i$ changes due to the $j$-th reaction is given by $S_{ij} f_j(\boldsymbol{\phi})$. Summing contributions from all reactions, yields a system of coupled ODEs for the evolution of the concentrations:

$$\frac{\mathrm{d}\phi_i}{\mathrm{d}t} = \sum_{j=1}^{R} S_{ij} f_j(\boldsymbol{\phi}), \quad \text{for} \quad i = 1, \ldots, N. \tag{2.43}$$

This set of ordinary differential equations (ODEs) is called the system of deterministic *rate equations* of the reaction network [132]. The rate equations give a macroscopic, deterministic description of the dynamics of the chemical reaction network. The law of mass action and the associated rate equations assume continuous, well-mixed, concentrations and ignore the discreteness of molecule numbers. It has been observed experimentally that the relative fluctuations of the system scales like the inverse square root of the mean concentration and hence becomes small for systems with large number of molecules [20]. Therefore the rate equations are expected to give an accurate description when the system is well-mixed and has a large number of molecules. This description is also commonly called a *mean-field* description and the ODEs are called the mean-field equations.

Eqs. (2.43) can be rewritten in a matrix form as follows:

$$\frac{\mathrm{d}\boldsymbol{\phi}}{\mathrm{d}t} = \mathbf{S}\mathbf{f}(\boldsymbol{\phi}), \tag{2.44}$$

where $\mathbf{S}$ is the stoichiometric matrix and $\mathbf{f}(\boldsymbol{\phi}) = (f_1(\boldsymbol{\phi}), \dots, f_R(\boldsymbol{\phi}))^T$.

When the number of molecules is low, fluctuations become important. A chemical system is *dilute* if the combined volume of all the molecules is much smaller than the total volume. If the chemical system is dilute and well-mixed, then a suitable way to model it is via a continuous-time Markov jump process [88]. The state of the system is described via a probability distribution $P(\mathbf{n}, t|\mathbf{n}_0, t_0)$ in which $\mathbf{n} = (n_1, \dots, n_N)$ and $n_i$ is the number of molecules of species $X_i$. It is customary to abbreviate $P(\mathbf{n}, t) := P(\mathbf{n}, t|\mathbf{n}_0, t_0)$ and implicitly assume that we are conditioning on an initial state.

In this framework, reactions occur in a stochastic way. The probability for the $j$-th reaction to happen in an infinitesimal time step $dt$ is assumed to be given by $v_r(\mathbf{n})dt$ where $v_r(\mathbf{n})$ is the *propensity function* of the $j$-th reaction. The propensity function is proportional to the number of combinations of reactant molecules in $\mathbf{n}$ and also considers that the probability for two molecules to collide is proportional to $1/V$ where we recall that $V$ is the system volume [216]. The propensity function for reaction (2.40) is given by

$$v_j(\mathbf{n}, V) = k_j V \prod_{i=1}^{N} \frac{n_i!}{(n_i - r_{ij})! V^{r_{ij}}}. \tag{2.45}$$

In Table 2.1 we state the propensity functions for the most common reactions, which are reactions up to order two. Propensity functions of this form are called propensity of *mass-action kinetics type*. There are other propensity functions that consider time scale separation and other kinetics [216].

| Reaction order | Reaction | Propensity function |
|---|---|---|
| 0th | $\emptyset \xrightarrow{k}$ | $kV$ |
| 1st | $A \xrightarrow{k}$ | $kn_A$ |
| 2nd | $A + B \xrightarrow{k}$ | $\frac{k}{V} n_A n_B$ |
| 2nd | $A + A \xrightarrow{k}$ | $\frac{k}{V} n_A (n_A - 1)$ |

**Table 2.1:** Summary of simple reactions up to second order with their respective propensity functions.

The probability distribution follows the *chemical master equation* (CME):

$$\frac{\mathrm{d}}{\mathrm{d}t} P(\mathbf{n}, t) = \sum_{r=1}^{R} \left[ v_r(\mathbf{n} - \mathbf{S}_r, V) P(\mathbf{n} - \mathbf{S}_r, t) - v_r(\mathbf{n}, V) P(\mathbf{n}, t) \right]. \tag{2.46}$$

The CME can be rewritten in terms of the stoichiometric matrix $\mathbf{S}$ and the so-called step operator, $\mathbb{E}$. For an integer $k$ and a function $f : \mathbb{R}^N \to R$ and $1 \le i \le N$, is defined as follows:

$$\mathbb{E}_i^k f(\mathbf{n}) = \mathbb{E}_i^k f(n_1, \dots, n_N) = f(n_1, \dots, n_i + k, \dots, n_N). \tag{2.47}$$

The master equation can be rewritten as:

$$\frac{\mathrm{d}P(\mathbf{n},t)}{\mathrm{d}t} = \sum_{j=1}^{R} \left[ \prod_{i=1}^{N} \mathbb{E}_i^{-S_{ij}} - 1 \right] v_j\left(\mathbf{n}, V\right) P(\mathbf{n},t). \tag{2.48}$$

The connection between the CME framework and the deterministic rate equations is given by the evolution equation of the mean concentration. Let us denote the mean vector as $\langle \mathbf{n} \rangle = (\langle n_1 \rangle, \ldots, \langle n_N \rangle)$ where

$$\langle n_i \rangle = \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} n_i p(\mathbf{n},t) := \sum_{\mathbf{n}} n_i p(\mathbf{n},t), \tag{2.49}$$

is the mean number of molecules of the $X_i$ specie.

The evolution of the mean concentration $\langle \mathbf{n}/V \rangle = (\langle n_1/V \rangle, \ldots, \langle n_N/V \rangle)$ matches the one given by the deterministic rate equations if the system involves only zeroth or first order reactions [110]. If the system involves second or higher reactions then the evolution of the mean concentration is coupled to higher moments whose evolution may also depend on higher order moments, potentially leading to an infinite system of coupled equations. Two approximation methods allow describing the moments of order M as functions of lower order moments: the linear noise approximation and moment closure approximations. We describe the linear noise approximation in Section 2.4.4 and we refer the reader to [93, 186, 214] for more information about moment closure approximations.

### 2.4.2 Reaction-diffusion master equation

When the well-mixed assumption breaks down, spatial effects become important. Stochastic reaction-diffusion models are used to describe systems in which fluctuations and space play important roles. Three frameworks are typically used: the Doi model, the Smoluchowski diffusion limited reaction model, and the reaction-diffusion master equation (RDME) [111]. The RDME can be interpreted as an extension of the non-spatial chemical master equation (CME) model and is relevant when the well-mixed assumption breaks down in the whole system. In the RDME framework, space is partitioned into a distinct compartments and molecules are assumed to be well-mixed within each compartment. Diffusion of the molecules is modelled as a stochastic jump process between the compartments and the reactions are modelled locally.

Mathematically, the RDME can be described by a forward Kolmogorov equation for a continuous-time jump Markov process. Let us consider for example a chemical system consisting of $N$ chemical species in a spatial domain $V$ which is divided into $M$ compartments with volume $\frac{V}{M}$. The $j$-th chemical reaction in compartment $m$ has the form:

$$r_{1,j}X_1^m + r_{2,j}X_2^m + \ldots + r_{n,j}X_n^m \xrightarrow{k_j} p_{1,j}X_1^m + \ldots + p_{n,j}X_n^m \quad j = 1, \ldots, R, \tag{2.50}$$

where the local variable $X_i^m$ refers to species $X_i$ in the $m$-th compartment. The stoichiometric coefficients $r_{ij}$ and $p_{ij}$ are assumed to be compartment independent. The diffusive interchange of chemicals between neighbouring compartments is modelled by the reactions:

$$X_i^m \underset{d_i}{\overset{d_i}{\rightleftharpoons}} X_i^{m'}, \quad m' \in \mathrm{Ne}(m), \tag{2.51}$$

where $d_i$ is the diffusion rate of species $X_i$ which is assumed to be independent of the compartment index i.e. the diffusion coefficient is space independent, $\mathrm{Ne}(m)$ is the set of compartments neighbouring the $m$-th compartment. We denote the probability of the system being in state $\{\mathbf{n}^m\}_{m=1}^M$ at time $t$ as $P\left(\{\mathbf{n}^m\}, t\right)$.

The dynamics of the system is described by the RDME:

$$\begin{aligned}\frac{\mathrm{d}P\left(\{\mathbf{n}^m\},t\right)}{\mathrm{d}t} &= \sum_{m=1}^{M}\sum_{j=1}^{R}\left(\prod_{i=1}^{N}\mathbb{E}_{i,m}^{-S_{ij}}-1\right)v_j^m\left(\mathbf{n}^m,V\right)\times P\left(\{\mathbf{n}^m\},t\right)\\ &+\sum_{m=1}^{M}\sum_{m'\in\mathrm{Ne}(m)}\sum_{i=1}^{N}\left(\mathbb{E}_{i,m}^{1}\mathbb{E}_{i,m'}^{-1}-1\right)\frac{d_i}{h^2}n_i^m\times P\left(\{\mathbf{n}^m\},t\right)\end{aligned} \tag{2.52}$$

where $n_i^\lambda$ is the number of molecules of species $X_i^\lambda$. $\mathbb{E}_{i,m}^x$ is the operator which replaces $n_i^m$ with $n_i^m + x$. The local propensity function is defined as follows: if the state of the $m$-th compartment is $\mathbf{n}^m$, then $v_j(\mathbf{n}^m, \frac{V}{M})dt$ is the probability that one reaction of index $j$ occurs somewhere inside that compartment during the infinitesimal time interval $[t, t + dt)$.

For systems where only zeroth and first order reaction are involved, as the lattice spacing approaches zero, the time evolution of the mean concentration from the RDME converges the corresponding deterministic reaction-diffusion equation [111]. However for systems involving chemical reactions of second or higher order, in two or more dimensions, the RDME has been shown to inaccurate when modelling bimolecular reactions as the mesh spacing approaches zero [110–112]. Other frameworks, like the convergent RDME (CRDME) [111], have been developed that converge to a continuum model as the mesh spacing approaches zero.

An assumption of the RDME framework is that within each voxel, molecules are assumed well-mixed. A critical question is how to choose a suitable voxel (compartment) size for the RDME dynamics to be valid. Several criterion have been developed [25, 69, 113, 123] and most focus on ensuring that the diffusion time scale is smaller than the smallest reaction time scale. This means that the voxel size needs to be small enough that all species in a voxel can jump to other voxels on the time scale of the fastest reaction in the voxel, since only then can a compartment be considered well-mixed [123].

We describe the criterion for the voxel size introduced by [25]. Let us consider the voxel volume $V$ to be a $d$-dimensional parallelepiped such that its volume is given by $V = h^d$ where $h$ is the compartment size. Let us denote the time scale of reaction $r$ as $\tau_r$. As expected, the reaction time scale depends on the reaction order. In Table 2.2 we include the time scales $\tau_r$ for the zeroth, first and second order reactions. We assume that the jump rates $d_i$ are given by $d_i = \frac{D_i}{h^2}$ for all $i = 1, \ldots, N$. In this case the characteristic time scale of diffusion is given by $\tau_D = h^2/min_i D_i$ where the minimum is taken over all the species. In Table 2.2 for each reaction order considered, we include the constraint on the value $h$ that ensures that $\tau_D < \tau_r$. The compartment size should be chosen such that $\tau_D < \min_r \tau_r$ where the minimum is taken over all the reactions.

| Reaction order (r) | $\tau_r$ | Condition on $h$ |
|---|---|---|
| 0th | $\frac{c^*}{k_0 h^d}$ | $h^{2+d} < \frac{\min_i D_i}{k_0 (c^*)^{-1}}$ |
| 1st | $\frac{1}{k_1}$ | $h^2 < \frac{\min_i D_i}{k_1}$ |
| 2nd | $\frac{min_i D_i}{k_2 c^*}$ | $h^{2-d} < \frac{\min_i D_i}{k_1}$ |

**Table 2.2:** Reaction timescales and necessary conditions on the compartment size $h$ to ensure that $\tau_D < \tau_r$ for different reaction orders. The reaction rate for the $j$-th reaction is denoted by $k_j$. For the 0-th reaction order, we consider $c^*$ to be a reference concentration.

When the system contains reactions of order two or higher, there is also a lower bound for the compartment size. We notice that when the dimension is larger than two, the condition on the compartment size presented in Table 2.2 for bimolecular reactions becomes a lower bound for the compartment size. Other authors have proposed other lower bounds as well and we refer the reader to [69, 110] for more information.

### 2.4.3 Stochastic simulations

There are generally no analytic solutions known to the CME. However, it is possible to directly simulate the underlying process. The standard stochastic simulation algorithm (SSA) is a popular Monte Carlo method that allows the simulation of exact sample paths of the stochastic process described by the CME [85]. It was first proposed by Gillespie in the context of chemical kinetics, and several variants have been proposed in the literature, see [87, 170].

**Algorithm 2.1** Gillespie Algorithm

```
1: Generate a random number $\zeta_1$ uniformly distributed in $(0,1)$ and compute the
   time when the next chemical reaction takes place as $t + \tau$ where
       $\tau = \frac{1}{\alpha_0} \ln\left(\frac{1}{\zeta_1}\right).$    (2.53)
2: Generate a random number $\zeta_2$ uniformly distributed in $(0,1)$ and compute
   which reaction occurs at time $t + \tau$. Find $j$ such that
       $\zeta_2 \geq \frac{1}{\alpha_0}\sum_{i=1}^{j-1} \alpha_i(t)$ and $\zeta_2 < \frac{1}{\alpha_0}\sum_{i=1}^{j} \alpha_i(t).$    (2.54)
   Then, the $j$-th reaction takes place, so perform an update of the number of the
   reactants and products of the $j-$th reaction.
3: Continue with step 1 for time $t = t + \tau$.
```

### 2.4.4 System size expansion and intrinsic noise approximation

The system size expansion is a singular expansion of the CME originally developed by Van Kampen [217]. The system size ansatz, or Van Kampen ansatz, consists of expressing the number of molecules $n_i$ of species $X_i$ as

$$\frac{n_i}{\Omega} = \phi_i + \Omega^{-1/2}\epsilon_i, \quad i = 1, \dots, N, \tag{2.55}$$

where $\Omega$ is the system size, $\phi_i$ satisfies a deterministic equation and $\epsilon_i$ is a random variable that represents the fluctuations of the system. In most cases, $\Omega$ is proportional to the total volume $V$ or the total number of molecules, hence the name system-size expansion. The Van Kampen approximation at the lowest order results in the mean-field equations [216].

When the system size expansion is carried out only at first order, it is referred as the *linear noise approximation*. In this case $\epsilon_i$ solves a multivariate linear Fokker-Planck equation given by:

$$\frac{\partial \Pi(\boldsymbol{\epsilon}, t)}{\partial t} = \sum_{i=1}^{N}\sum_{k=1}^{N}\left[-A_{ik}\frac{\partial(\xi_k \Pi)}{\partial \xi_i} + \frac{B_{ik}}{2}\frac{\partial^2 \Pi}{\partial \xi_i \xi_j}\right]. \tag{2.56}$$

where $\boldsymbol{A}, \boldsymbol{B} \in M_{N\times N}(\mathbb{R})$.

The extension of the linear noise approximation for the RDME framework is referred to as the *intrinsic noise approximation* and was developed by Van Kampen in 1976 [215]. The aim of this procedure is to approximate the system dynamics by the first and second moment dynamics. To perform this approximation, the jump and the reaction processes are consider separately and evolution equations are derived for the first and the second moments of the stochastic process. Finally, the combined effect of the reactions and diffusion is obtained by adding the effects of the diffusion and the reactions on the moments.

Instead of determining the time evolution of the second moments directly, Van Kampen derived the evolution equation of the factorial cumulant and showed that, for several reaction systems, the factorial cumulants follow simpler evolution equations than the second moments [215]. In such cases the time evolution equation of the covariance can be obtained from the factorial cumulant evolution. The factorial cumulant for a reaction-diffusion system with local reactions (2.50) and the jump rates (2.51) for the number of molecules at the $m$-th and $m'$-th voxel is denoted by $[\boldsymbol{n}^m, \boldsymbol{n}^{m'}] \in M_{N\times N}(\mathbb{R})$ where:

$$[\boldsymbol{n}^m, \boldsymbol{n}^{m'}] := \langle \mathbf{n}^m (\mathbf{n}^{m'})^T \rangle - \langle \mathbf{n}^m \rangle \langle (\mathbf{n}^{m'})^T \rangle - \delta_{mm'}\mathrm{diag}[\langle \mathbf{n}^m \rangle]. \tag{2.57}$$

For a system with one species, this equation simplifies to $[n^m, n^{m'}] = \langle n^m n^{m'} \rangle - \langle n^m \rangle \langle n^{m'} \rangle - \delta_{mm'} \langle n^m \rangle$.

The methodology introduced by Van Kampen was formulated for general chemical systems in [188]. We now outline this procedure for reaction-diffusion systems with local reactions (2.50) and jump rates (2.51).

1. First one considers only the effects of the local reactions on the RDME (2.52).

2. In each voxel, the linear noise approximation is considered $\boldsymbol{x}^m = \Omega\boldsymbol{\phi}^m + \Omega^{1/2}\boldsymbol{\epsilon}^m$ by taking into account a system size $\Omega$.
3. The linear Fokker-Plank equation that the new random variable $\boldsymbol{\epsilon}^m$ follows is derived for each $m$-th voxel with $1 \leq m \leq M$.
4. The time evolution of the mean, the second moments and the factorial cumulant of the fluctuation term are derived.
5. The time evolution equation of the factorial cumulant of the number of molecules for each pair of voxels (Eq. (2.57)) is derived.
6. The factorial cumulant evolution equation is divided by the square of the voxel volume, $V^2$, and the limit of the equation as $V \to 0$ is considered.
7. The effects of the diffusion on the RDME (2.52) are now considered. The time evolution equation of the factorial cumulant due to the diffusion is derived.

### 2.4.5 Hybrid algorithms

In this section we have introduced deterministic and stochastic approaches for modelling reaction-diffusion systems. If the number of molecules is large, then using a stochastic framework such as the RDME would be inefficient, and a mean-field approximation would be a more suitable model. However, in some situations, the number of molecules is large in certain regions of the domain and close to zero in other parts of the domain. In such cases, a stochastic model may be inefficient in one region, and a PDE model inaccurate in another. To overcome this problem, *hybrid algorithms* have been developed that couple different modelling frameworks for the same species in a common domain. When the number of molecules is low, the stochastic framework is considered and when the number of molecules is large a PDE formalism is taken into account.

In this subsection we focus on hybrid algorithms that couple a RDME framework with its mean-field description given by a reaction-diffusion equation to which we refer as its mean-field PDE. Other hybrid algorithms have been proposed that couple RDME frameworks to molecular-based models, mean-field PDEs to molecular-based models and molecular models to more detailed molecular dynamics simulations. We further introduce some hybrid algorithms that couple the RDME framework to its mean-field PDE. We refer the reader to a review of hybrid algorithms for a description of alternative approaches [196].

Most hybrid algorthims that couple a RDME framework with its mean-field PDE split the domain into distinct regions in which the different modelling frameworks are used. We refer to these regions as deterministic or stochastic depending on how the number of molecules is updated. The regions may overlap to allow switching between the frameworks. Let us consider the algorithm developed by Spill et al. in [149]. The authors determine whether a voxel belongs to the stochastic or the mean-field domain by comparing the number of particles in that voxel with a threshold value. Voxels with a fewer number of particles than this threshold are modelled with the RDME framework; voxels with a higher number of particles are updated with a finite difference approximation to the mean-field PDE. The system update at each step consists of, first a stochastic update and then a deterministic update. The authors consider only one voxel between the two modelling domains. During the system update, the reactions and the flux from the stochastic domain to this voxel are updated stochastically during the stochastic update. The voxels contents are then transformed to concentration units and used to define a Dirichlet condition for the deterministic update. The authors showed that their algorithm can accomodate multi-species and multiple distinct stochastic and deterministic domains [149].

Another approach is to consider a large overlapping region (more than two voxels) between the deterministic and stochastic domains. For example, Harrison and Yates [100] proposed a hybrid algorithm in which a time-driven algorithm was used to simulate the stochastic domain and used the same time step for the deterministic update. Their algorithm couples the two frameworks via the overlapping region and relies on specification of a Dirichlet-type condition at the interface between the overlapping region and the deterministic domain and a flux-matching boundary condition at the other interface. The mesh used for the deterministic domain is much finer than that used for the stochastic domain (i.e. the voxels). The authors applied their algorithm to different reaction systems and showed that the resulting simulations approximate the stochastic model well.

Several alternative hybrid algorthims do not split the domain into regions where different modelling frameworks are implemented; we refer the reader to [196] for more details of algorithms of this type.

## Chapter 3

# Migration quantification method for scratch assays

The scratch assay is a standard technique in biomedical research that provides an initial assessment of the migration rates of cell lines (See Section 1.1.2 for more information about scratch assays). Given the key role this method plays in assessing the potential of new compounds for clinical use, it is important to develop robust methods that accurately measure and compare migration rates of different scratch assays. In this chapter we introduce a new migration quantification method for scratch assays. Parts of this chapter are based on the peer reviewed publication: [178].

## 3.1 Introduction

Multiple methods are used to assess collective cell migration in scratch assays [207]. The most common approaches focus on wound width or area change [90, 153]. They use a range of metrics to quantify migration, including the percentage difference in the wound width at different time points [180, 225], wound width at specific time points [37], and the slope of a linear approximation to the change in wound area [122] (see below). Other more computationally intensive methods rely on cell tracking or determining a velocity field across the full monolayer [152, 159]. A weakness of the previously mentioned methods is that they do not perform well when the two borders of the scratch are not perfectly straight, a feature which is common in experimental data [14, 122]. The lack of a reproducible wounding procedure results in non-uniform cell-free areas with irregular leading edges, as can be seen in Figure 3.1. Furthermore, migration rate measurements have been shown to be sensitive to the initial degree of confluence [174] and the initial geometry of the wound [116]. Current quantification methods require high quality experimental data which are difficult to obtain; therefore, frequently the experimental data are discarded and repeat experiments performed.

Another issue with many existing quantification methods is that they use time-specific measurements to determine differences in migration rates of cell samples. Typically, the wound area or width of each sample is measured at a series of time points. Then, statistical tests are performed on these measurements with the purpose of detecting significant differences [90]. The times at which data are collected are not standardised and vary across studies [24, 89, 225]. Differences in the time points at which data is collected may be due to differences in the proliferation rates of the cell types under consideration. However, there is no standard procedure for choosing the comparison time even though the choice can affect the comparison and may render the results from the analysis unreliable.

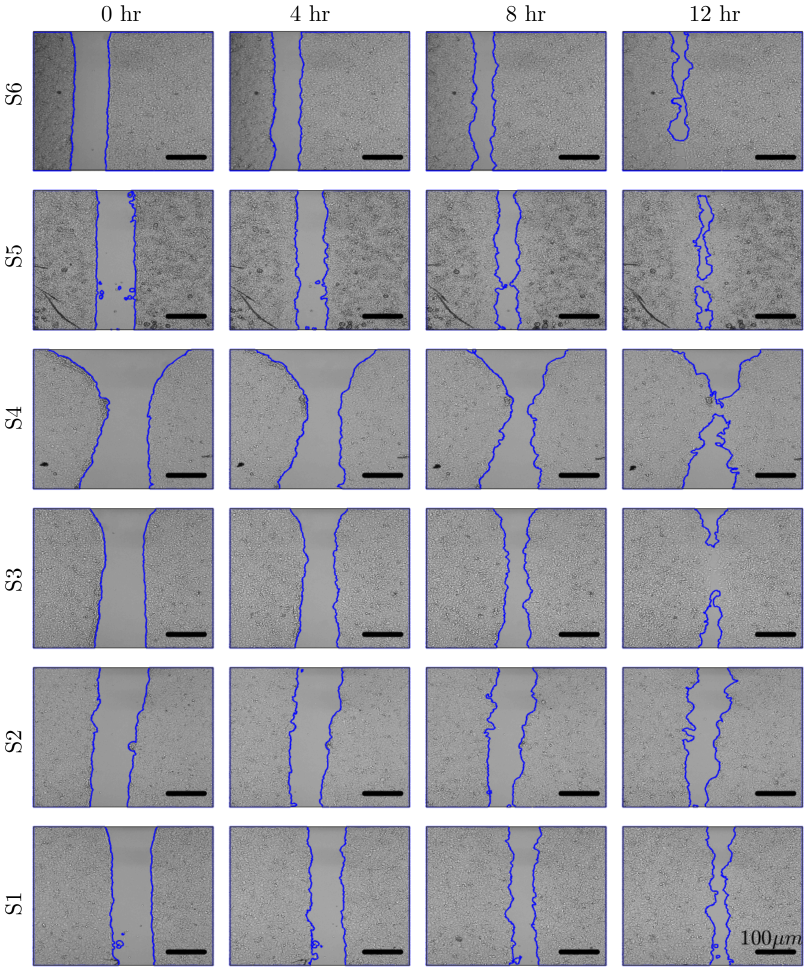


**Figure 3.1:** Time-lapse images from representative scratch assays from six experimental cell groups (S1-S6) are plotted at times $t = 0, 4, 8$ and 12 hr. In each image, the leading edges were detected by applying the segmentation algorithm presented in Section 3.3. The resulting interfaces/cell fronts are plotted in blue.

In this chapter, we introduce a new quantification method that can deal with non-smooth leading edges. Irregular leading edges are accounted for by approximating the front by a piecewise constant function, which is constant over windows with a fixed size, $w^*$. We assume that within each window, the contour moves with constant speed in the perpendicular direction until the left and right leading edges meet. Cell migration in the scratch assay is characterised by a series of linear approximations to the interface's position over time in these windows.

The remainder of the chapter is organised as follows. In Section 3.1.1 we introduce two common quantification methods for scratch assays: the *percentage wound area method* and the *closure rate method* [122]. In Section 3.2 we describe the scratch assay data that we use to validate our quantification method. We consider data from in-vitro experiments performed in the Vall d'Hebron Research Institute and simulations from an agent-based model. In Section 3.3 we describe the image-processing algorithm that we use to identify the leading edge in the scratch assay. In Section 3.4

we introduce a new migration quantification method for scratch assays: the monolayer edge velocimetry (MEV) method. In Section 3.5 we perform a sensitivity analysis to determine how the methods are affected by the cell migration and proliferation rates. In Section 3.6 we assess our quantification method by applying it to in-silico data. After showing that the method correctly classifies cells with different migration and proliferation parameters, we then compare it against the other methods. In Section 3.7 we present results obtained by applying the three methods to an experimental data set. Finally in Section 3.8, we discuss our results and present our conclusions.

### 3.1.1 Migration quantification methods for scratch assays

We first introduce two established quantification methods for scratch assays. Then, we introduce a new method that quantifies the x-component of the velocity of the leading edge of the cell monolayer.

#### The area method

One of the most widely used quantification methods, which we refer to as the *area method*, assesses migration indirectly [90]. During an experiment, the wound area percentage, $\hat{A}(t)$, is tracked:

$$\hat{A}(t) := \frac{A(t)}{A(0)} \times 100\%$$

where $A(t)$ is the wound area at time $t$ and $A(0)$ its initial area. The migration rate is then indirectly evaluated as the percentage wound area at a specific time point.

#### The closure rate method

In [122] cell migration is quantified by assuming that the wound area reduces linearly over time. We refer to this method as the *closure rate method*. The change in wound area $A(t)$ is first approximated by a linear function:

$$A(t) \approx m \times t + b \tag{3.1}$$

where $m$ and $b$ are scalars. The wound area is assumed to be the length of the field-of-view ($l$) times the width of the gap ($W(t)$). Since $l$ is constant during the course of the experiment, Eq. (3.1) becomes:

$$\frac{dA}{dt} = l \times \frac{dW}{dt} \approx m. \tag{3.2}$$

The migration rate, $C_r$, is defined to be half of the width closure rate

$$C_r := \frac{1}{2}\frac{dW}{dt}. \tag{3.3}$$

Combining Eqs. (3.2) and (3.3), we have

$$C_r = \frac{|m|}{2 \times l}. \tag{3.4}$$

## 3.2 Scratch assay data

In this section we introduce the scratch assay data that we consider to analyse the different quantification methods. We consider data from in-vitro experiments performed in the Vall d'Hebron Research Institute and in-silico data: simulations from an agent-based model. We describe in more detail each dataset.

### 3.2.1 In-vitro data

In-vitro experiments were performed by members of the Renal Physiopathology Group at Vall d'Hebron Research Institute lead by Prof. Anna Meseguer. Six site-specific mutations in a latent transcription factor that regulates downstream genes involved in essential biological processes, including migration, were generated. The corresponding mutants S1, S2, S3, S4, S5 and S6 were then transduced into a human renal carcinoma cell line, 769-P (ATCC CRL-1933), through lentiviral particles. The 769-P mutants were cultured in Dulbecco's Modified Eagle's Medium (DMEM) (#42430, Gibco) supplemented with 10% of fetal bovine serum (FBS) (#10270, Gibco), 1% of sodium pyruvate solution 100 mM (#03-042-1B, Biological Industries) and 1% of antibiotic-antimycotic solution 100X (#15240, Gibco). Cells were maintained at 37°C in 5% $CO_2$.

For the wound healing assay, the 769-P mutants (S1-S6) were seeded at $0.025 \times 10^6$ cells per well in a 2 well silicone insert with a defined cell-free gap (Ibidi #81176, Germany), incubated and allowed to grow for 48 hr. Once the cells reached 100% confluence in both wells, the culture insert was removed and the area that remained clear of cells was quantified over a period of 24 hr using the Live Cell-R Station (Olympus). Digital images were obtained every 30 minutes.

The data set consisted of 24 wound healing assays: four replicates for each of the six groups (S1-S6). Each assay consisted of 48 images. The imaged region size was $500 \times 500$ $[\mu m]^2$.

### 3.2.2 In-silico data: agent-based model simulations

We employ an agent-based model that has been previously used to simulate in-vitro cell cultures [120, 155, 194]. The simulation domain is a two-dimensional square lattice, with the same dimensions as the experimental images: $[0, D] \times [0, D]$ where $D = 500$ $\mu m$. The lattice spacing, $\Delta$, which is interpreted as the average cell diameter, is set to 10 $\mu m$ unless otherwise specified.

In this model on each time step, a cell can either proliferate, move or do nothing. We consider an end time of $T = 24$ hr and an update time of $\tau = 0.1$ hr. We include crowding effects by assuming that each lattice site is occupied by at most one cell. A cell with centre at $(x, y)$ is said to be at $(x, y)$. Reflective boundary conditions are imposed. At each update time, agents move and/or proliferate with migration and proliferation probabilities $p_m$ and $p_p$, respectively. Since typical estimates of the cell doubling time are approximately 15-30 h [150, 195], whereas the time required for a cell to move a distance equal to its diameter is of the order of 10 min [128], we consider migration and proliferation probabilities in the ranges $p_m \in [0, 1]$ and $p_p \in [0, 0.01]$, respectively. In Figure 3.2 we plot the evolution of a typical realisation of the agent-based model for which the migration and proliferation parameters are given by $p_m = 0.3$ and $p_p = 0.01$.

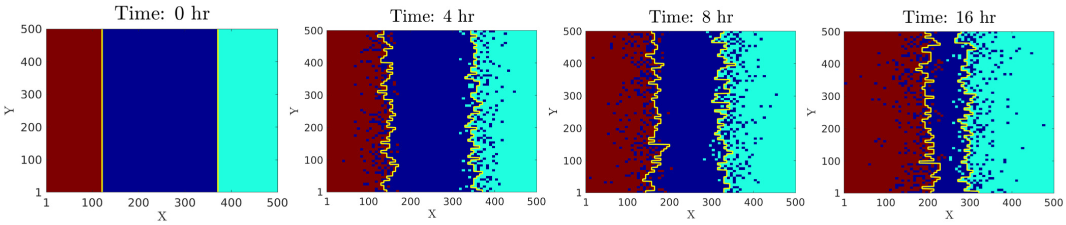


**Figure 3.2:** Evolution of an agent-based simulation. We considered an idealised initial condition and fixed the migration and proliferation parameters so that $p_m = 0.3$ and $p_p = 0.01$, respectively. The recolonisation of the wounded region is shown at times $t = 0, 4, 8$ and 16 hr. For ease of visualisation, the two cell monolayers (right and left) are plotted with different colours (red and turquoise), while the area devoid of cells is coloured blue. The leading edges detected by the segmentation algorithm (see details below) are plotted in yellow.

### Algorithm and parameter values for agent-based model

**Algorithm 3.1** Agent-based model algorithm

```
1: Initialisation. Consider an idealised initial condition unless otherwise specified: all lattice points $(x, y)$, for which $x < D/4$ or $x > 3D/4$, are occupied by an agent.
2: while $t < T$ do
3:     Update the system from $t + \tau$: Let $N(t)$ denote the number of agents at time $t$. To update the agent-based model at time $t$ to the next simulation time $t + \tau$, we proceed as follows:
4:     First, $N(t)$ agents are chosen sequentially at random and given the opportunity to move. An agent at $(x, y)$ attempts to move with probability $p_m$ to $(x \pm \Delta, y)$ or $(x, y \pm \Delta)$, with any possible target site being chosen with equal probability.
5:     Further $N(t)$ agents are then selected sequentially at random and given the opportunity to proliferate. An agent at $(x, y)$ attempts to proliferate with probability $p_p$ and places its daughter agent at $(x \pm \Delta, y)$ or $(x, y \pm \Delta)$, target sites being chosen with equal probability. Agents can move and/or proliferate only if the target site is vacant.
6: end while
```

A list of the parameters that are used in the agent-based model is presented in Table 3.1.

## 3.3 Image analysis

The spatial coordinates of the leading edges of the cell monolayers are extracted from the experimental images by applying a segmentation algorithm based on the Growcut method [219]. The method is already employed in several computer vision applications. It performs a binary image segmentation.

| Parameter | Description | Value | Units |
|---|---|---|---|
| $p_m$ | Motility probability | 0.3 | N/A |
| $p_p$ | Proliferation probability | 0.1 | N/A |
| $T$ | End time | 24 | hr |
| $\tau$ | Time step | 0.01 | hr |
| $D$ | Domain length | 500 | $\mu m$ |
| $\Delta$ | Lattice spacing | 10 | $\mu m$ |

**Table 3.1:** List of the parameters used in the agent-based model, together with their default values.

The Growcut algorithm requires the initial specification of a subset of pixels from each type of region: cell monolayer and unoccupied space; these pixels are referred to as *seeds*. The seeds should be located far from the leading edges, where all the pixels of such an area belong to one of the two classes. The algorithm evolves as follows: at each iteration, the pixels surrounding the initial seeds are assigned to one class or the other, adjusting the size of each region. The classification depends on the similarity of the pixel intensity with respect to the pixel intensity of the seeds.

In our implementation, the seeds are chosen as follows: for the cell region, the Canny and Roberts edge contour methods [41, 192] are used to select the pixels with the highest variability, corresponding to the cell contours. For the background region, the seeds are set in areas of low variability ( i.e. where the pixel intensity has a standard deviation less than 500).

After applying the detection algorithm to each image, we have a record of the positions of the left and the right interfaces at all times for which images were taken. At each vertical position, the interface is considered to be the closest pixel to the wound.

## 3.4 Proposed quantification method: monolayer edge velocimetry (MEV)

We propose a new method for quantifying front migration in a scratch assay using a set of representative velocities. We denote by $t_0, \dots, t_N$, the times at which data are collected. Let $X \times Y$ represent the square domain of the processed image, $X = Y = \{1, \dots, D\}$ where D is the number of pixels. For each $j \in Y = \{1, \dots, D\}$, we denote the interface position in the horizontal direction, at the $j-$th vertical position and at time point $t_n$, as $i_j(t_n)$ where $1 \leq i_j \leq D$. See Figure 3.3 (a) for a schematic representation.

To determine the velocities of the leading edge, we assume that the front position changes linearly over time within a window of size $w$. The linear approximation is determined in two steps:

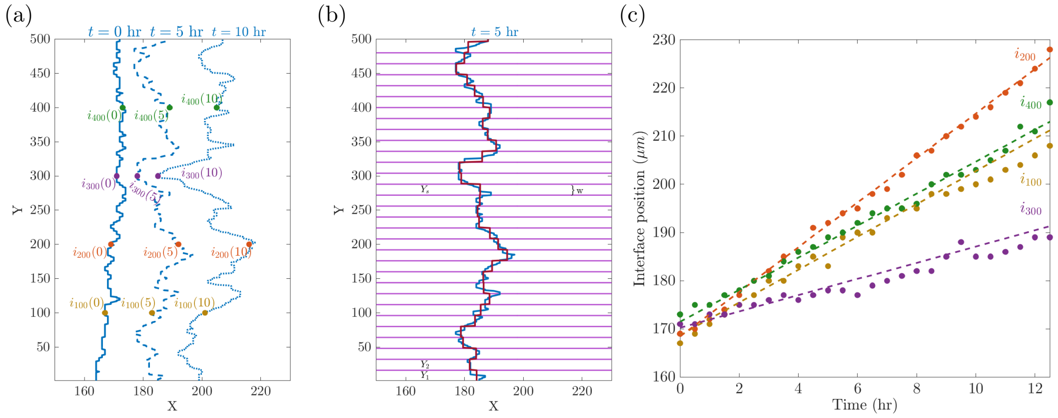
1. First, the front position is approximated for the window size, $w$. $Y$ is divided into $M = D/w$ segments denoted $Y_s$, $1 \leq s \leq M$, each of length $w$. The front position $\hat{i}_s$ in each segment $Y_s$ is approximated by its mean position,

$$\hat{i}_s(t_n) = \langle i_j(t_n) \rangle_{i_j \in Y_s}. \tag{3.5}$$

This procedure is illustrated in Figure 3.3 (b).
2. The dynamics of the front position in each window are approximated by linear regression, so that

$$\hat{i}_s(t_n) \approx m_s t_n + b_s. \tag{3.6}$$

In Figure 3.3 (c) the dotted lines indicate how the front position changes over time at selected y-coordinates; the dashed lines represent the corresponding linear approximations for a window of size $w = 16$.



**Front position at different times**

**Front approximation with respect to window size $w$**
$\hat{i}_s(t_n) = \langle i_j(t_n) \rangle_{p_j \in Y_s}$

**Linear approximation to the front position over time**
$\hat{i}_s(t_n) \approx m_s t_n + b_s$

**Figure 3.3:** Linear approximation of the front position over time with respect to a window size $w$ for one of the scratch assays from the experimental cell groups. (a) To introduce the notation, the positions of the left front at times $t = 0, 5$ and 10 hr are plotted in blue. The solid line corresponds to $t = 0$ hr; the dashed line to 5 hr and the dotted line to 10 hr. The front positions at the 100, 200, 300 and 400 y-coordinates for these times are plotted: yellow, orange, purple and green, respectively. (b) The left front at $t = 5$ hr is approximated by a window size $w$. $Y$ is partitioned into $M$ segments denoted $Y_s$, $1 \leq s \leq M$, each with length $w$. A magenta horizontal line delimits each segment. The front position is plotted in blue and the approximated front position, taken as an average over each $Y_s$, is plotted in red. (c) The time evolution of the interfaces at the 100, 200, 300 and 400 y-coordinates and the linear approximation with respect to the window size $w = 16$ are plotted using dotted and dashed lines, respectively. The window size $w = 16$ is the window size that maximises the objective function (3.7).

By performing this approximation for the left and right interfaces, we obtain a set of velocities $\{|m_s|\}_{s=1}^{2M}$ we refer to as the windowed velocities for window size $w$.

An important step in our method is to determine the window size at which to perform the linear approximation (3.6). We observe that as the window size decreases, the discrepancy between the linear approximation and the front evolution decreases (see Figure 3.17 a)). This observation prompts us to consider the smallest window size, however, for window sizes smaller than the average cell size, the left and right windowed velocity distributions are significantly different. This is a finite size sampling effect, since the scale on which the velocities are quantified is much smaller than the cell size scale. Therefore the individual velocity of each cell at the front is counted multiple times and its value is over represented, producing a bias in the overall windowed velocity distribution. To deal with the finite size sampling effect, we choose a window size for which the left and right windowed velocity distributions are similar. We note that if two different cell types are seeded on either side then the assumption of left and right similarity can no longer be made. However, in the standard experimental setup of the scratch assay, the left and right interfaces are from the same cell type.

We use two criterion to select the optimal window size, $w^*$: (i) fitness of the approximation, and (ii) similarity of the left and right windowed velocity distributions. In more detail, we introduce an objective function, $F(w)$, that has three terms and enables us to find the optimal window size based on the stated criterion:

$$F(w) = Fit_{resid}(w) + Fit_{Rsquared}(w) + Fit_{KS_{distance}}(w). \tag{3.7}$$

- $Fit_{resid}(w)$ measures the discrepancy between the interface position over time and the linear approximation (Eq. (3.6)).
- $Fit_{Rsquared}(w)$ considers the coefficient of determination, $R^2$, which describes how well the dynamic changes in the variance of the interface position can be explained by the linear approximation [103].
- $Fit_{KS_{distance}}(w)$ is the Kolmogorov-Smirnov distance [103]; it calculates the distance between the left and right front windowed velocity distributions.

The three terms that appear in Eq. (3.7) are scaled such that the window size that maximises the objective function takes into account the trade-off between giving the best fit and having the left and right velocity distributions closest to each other.

The procedure used to determine the optimal window size can be interpreted as optimizing the number of velocities needed to characterise cell migration. Detailed information about how to modify the objective function when multiple assays are considered can be found in Section 3.4.1.

The procedure used to determine the set of representative velocities is summarised in Algorithm 3.2.

**Algorithm 3.2** Monolayer edge velocimetry

```
1: Determine the optimal window size for the linear approximation by using the objective function (3.7)
       $w^* = \max_{1 \le w \le D} F(w)$        (3.8)
   where $F(w)$ is given by Eq. (3.7).
2: Perform a linear approximation with respect to the window size $w^*$; indicating how the positions of the left and right interfaces change over time,
       $\hat{i}_s(t_n) \approx m_s t_n + b_s$
   where $\hat{i}_s(t_n) = \langle i_j(t_n) \rangle_{i_j \in Y_s}$, $Y = \bigcup_{s=1}^{M} Y_s$ in which $|Y_s| = w^*$ and $M = D/w^*$.
Ensure: $\{|m_s|\}_{s=1}^{2M}$ is the representative set of velocities that quantify cell migration in the scratch assay.
```

We refer to the "Monolayer edge velocimetry method" as the MEV method for the rest of the chapter.

### 3.4.1 Objective function used to determine optimal window size

In this section we describe the objective function used to calculate the optimal window size in detail. For a given window size $w$, divide $Y$, which has size $D$, into $M = D/w$ segments of length $w$, denoted by $Y_s$ for $1 \leq s \leq M$ . $Y_s$ is the set of pixels that belong to a segment of length $w$. When $D$ is not divisible by $w$, $Y$ is divided into $M = \lfloor \frac{Y}{w} \rfloor + 1$ segments where $\lfloor \cdot \rfloor$ denotes the floor function. The first $\lfloor \frac{Y}{w} \rfloor$ segments have length $w$ and the last one has length $D - w \times \lfloor \frac{D}{w} \rfloor$. After applying the linear approximation with a window of size $w$, we obtain the following approximation to the position of the interface at time $t_n$ for each $j \in Y_s$,

$$i_j(t_n) \approx m_s t_n + b_s, \tag{3.9}$$

where $m_s$ and $b_s$ are determined by Eq. (3.6) and are fixed over time.

The following fitness functions are used to determine how well Eq. (3.9) approximates the position of the interface:

1. Residual sum of squares, $E(w)$:

$$E(w) = \frac{1}{2D} \sum_{j=1}^{D} \sum_{n=1}^{N^*} \left( {e_{l,j,n}}^2 + {e_{r,j,n}}^2 \right) \tag{3.10}$$

   where $e_{l,j,n} = i_j(t_n) - (m_s t_n + b_s)$ is the residual of the linear approximation (3.9) at time $t_n$ for $j \in Y_s$. The first subscript denotes the interface (left or right). $N^*$ denotes the time step at which the two interfaces first meet.

2. Linear fitness, $R(w)$ :

$$R(w) = \frac{1}{2D} \sum_{j=1}^{D} \left( {R_l}_j^2 + {R_r}_j^2 \right) \tag{3.11}$$

   where ${R_l}_j^2$ is the coefficient of determination $R^2$ of the linear approximation for $j \in Y_s$, i.e. let $\bar{i_j}(t_n) = \frac{1}{N^*} \sum_{n=1}^{N^*} i_j(t_n)$, $SS_{tol} = \sum_{n=1}^{N^*} (i_j(t_n) - \bar{i_j})^2$ and $SS_{res} = \sum_{n=1}^{N^*} (m_s t_n + b_s - i_j(t_n))^2$, then ${R_l}_j^2 := 1 - \frac{SS_{res}}{SS_{tol}}$. The subscript denotes the interface (left or right).

3. Fitness between left and right velocities
   We consider a distance metric derived from the Kolmogorov-Smirnov test statistic [30]. Given two data sets $\{a_s\}_{s=1}^{n}$ and $\{b_r\}_{r=1}^{m}$ and their empirical cumulative distributions $F_n$ and $F_m$, respectively, the Kolmogorov-Smirnov statistic is:

$$D_{n,m} = \sup_{x \in \mathbb{R}} |F_n(x) - F_m(x)|. \tag{3.12}$$

   Normalising the statistic by the effective number of data points, we obtain the KS distance that was introduced in [71]:

$$Dist(\{a_s\}_{s=1}^{n}, \{b_r\}_{r=1}^{m}) = \frac{nm}{n+m} D_{n,m}, \tag{3.13}$$

   where $\{a_s\}_{s=1}^{n}$, $\{b_r\}_{r=1}^{m}$ are two set of data points. We consider the KS distance between the left and right windowed velocity distributions

$$KS_{distance}(w) = Dist\left(\{m_l\}, \{m_r\}\right), \tag{3.14}$$

where $m_r$ and $m_l$ are the slopes of the linear approximations determined by Eq. (3.6) for the left and right interface, respectively. We consider this term to assure the KS distance between the left and right windowed velocity distributions is as small as possible.

The objective function used to determine the optimal window size combines the three fitness functions introduced above and is defined as follows: We consider the following global function:

$$F(w) = \frac{1}{3}\left(Fit_{resid}(w) + Fit_{Rsquared}(w) + Fit_{KS_{distance}}(w)\right), \tag{3.15}$$

where

$$Fit_{resid}(w) := \frac{\max\limits_{1\leq w\leq D}(E(w)) - E(w)}{\max\limits_{1\leq w\leq D}(E(w)) - \min\limits_{1\leq w\leq D}(E(w))}, \tag{3.16}$$

$$Fit_{Rsquared}(w) := \frac{R(w) - \min\limits_{1\leq w\leq D}(R(w))}{\max\limits_{1\leq w\leq D}(R(w)) - \min\limits_{1\leq w\leq D}(R(w))}, \tag{3.17}$$

and

$$Fit_{KS_{distance}}(w) := \frac{\max\limits_{1\leq w\leq D}(KS_{distance}(w)) - KS_{distance}(w)}{\max\limits_{1\leq w\leq D}(KS_{distance}(w)) - \min\limits_{1\leq w\leq D}(KS_{distance}(w))}. \tag{3.18}$$

The terms (3.16)-(3.18) are normalised to lie between 0 and 1 and the largest value is the one maximising each criterion so that $0 \leq F(w) \leq 1$. The width of the window with the largest $F(w)$ value is considered the optimal value with respect to the three fitness functions, weighted equally through the global function.

If additional scratches are considered to calculate the optimal window, we take a weighted sum of objective functions for each of the scratches considered. In such cases, the objective function can be written as

$$F(w) = \frac{1}{S}\sum_{s=1}^{n} F_s(w), \tag{3.19}$$

where $F_s$ is the individual objective function for scratch assay $s$ and $S$ is the number of scratch assays considered.

### 3.4.2 Similarity between left and right velocities

In this subsection we discuss why similarity between the left and right velocities is necessary for optimal window size calculation.

If the term $Fit_{KS_{distance}}$ is neglected when calculating the objective function (Eq. (3.15)), then the optimal window size will be a window of size $1\mu m$ since the front position is not averaged over a window and the actual interface position is considered for the linear fit, reducing the overall fitness error. We now outline the problems of considering a window of size significantly smaller than the cell diameter.

For windows of size significantly smaller than the cell diameter, velocity values are correlated (See Figure 3.4 (a) and (b)) and not amenable to standard statistical analysis. In order to apply standard statistical tests, the velocities must have the same statistics as an independent and identically distributed (iid) sequence. One way to test for iid sequences is to consider the sample autocorrelation function [190] which, for an iid sequence with $n$ elements and finite variance, gives iid values that approximately follow a normal distribution $\mathcal{N}(0, 1/n)$ [34]. Let $x_1, \ldots, x_n$ be observations and $h$ an integer such that $|h| < n$, then the sample autocorrelation function at $h$ is given by:

$$\rho(h) := \frac{\gamma(h)}{\gamma(0)},$$

where $\gamma(h)$ is the sample autocovariance function:

$$\gamma(h) := \frac{1}{n} \sum_{t=1}^{n-|h|} \left(x_{t+|h|} - \overline{x}\right) \left(x_t - \overline{x}\right),$$

where $\overline{x}$ is the sample mean.

A general observation from the in-silico and in-vitro data is that the autocorrelation function values from velocities corresponding to windows of size significantly smaller than the cell diameter differ greatly from an iid sequence. On the other hand, when considering windows whose sizes are similar to the size of a cell diameter, the autocorrelation function resembles one for an iid sequence. To illustrate this, we simulate our agent-based model with the parameters in Table 3.1. We plot the autocorrelation function of the left velocities for windows of sizes $1, 2, 10$ and $12\mu m$ in Figure 3.4. The horizontal red lines in each plot delimit the 95% confidence interval from a normal distribution $\mathcal{N}(0, 1/n)$ where $n$ is the number of velocities. We observe that for windows of size smaller than the cell diameter (1 and $2\mu m$), the autocorrelation values exceed the upper limit of the confidence interval and decrease slightly as the lag increases (see Figure 3.4 (a) and (b)). On the other hand, for windows of sizes similar to that of the cell diameter, most of the autocorrelation values of the velocities are within the 95% confidence interval and, as such, can be considered to belong to an iid sequence (See Figure 3.4 (c) and (d)).

Another problem with windows of size significantly smaller than the cell diameter is that the velocity values of individual cells at the leading edge are over-represented and affect the overall distribution. This effect can be analysed by looking at the similarity between the left and right velocities of the in-vitro and in-silico data. Since the cells on each side come from a common monolayer, they are subject to the same biological conditions. As such, we expect the velocities from the left and right sides to exhibit the same statistics. A general observation from the in-silico and in-vitro data is that for windows of size significantly smaller than the cell diameter, the velocity distributions of each side differ while for windows whose sizes are similar to that of the cell diameter, the distributions more closely resemble each other. To illustrate this, we simulate the agent-based model with the parameters in Table 3.1. Violin plots of the velocity distributions for the left and right velocities corresponding to windows of size $1, 2, 10$ and $12\mu m$ are presented in Figure 3.5 (b). For windows that are smaller than the cell diameter, the velocity distribution of the left and right interfaces differ but for windows whose sizes are similar to that of the cell diameter, the distributions are more similar. The similarity between the left and the right velocity distributions can be quantified using the Kolmogorov-Smirnov (KS) distance, Eq. (3.14). In Figure 3.5 (a) we plot the mean KS distance between the left and right velocities for 500 simulations of the agent based model using the parameter values listed in Table 3.1, and we observe that the mean KS distance decreases as the window size increases.

We account for the statistical problems associated with windows which are significantly smaller than the cell diameter, by including the KS distance in the objective function. We evaluated this objective function for agent-based model simulations with different motility and proliferation rates. In this case, the optimal window size is the cell diameter and this window gave the best fit and the best similarity between the left and right velocities (see Section 3.5). Deviations from this result indicates lateral movement of the cells.

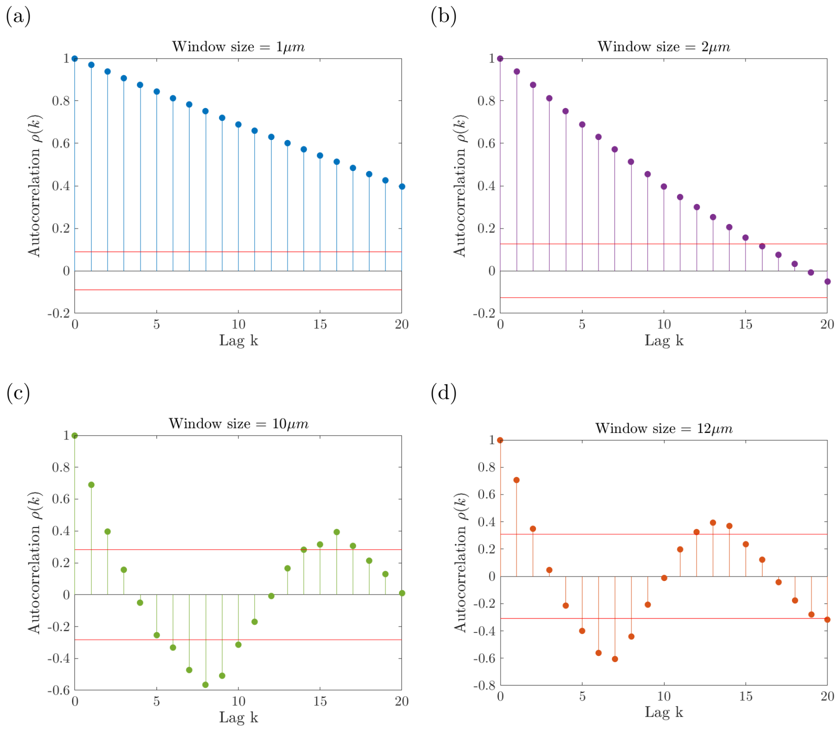


**Figure 3.4:** Plot of sample autocorrelation functions of left velocities corresponding to different window sizes for an agent-based simulation with the parameters in Table 3.1. We consider windows of size: (a) 1, (b) 2, (c) 10 and (d) $12\mu m$. For each integer $k$ between 0 and 20 we plot the sample autocorrelation with time lag $k$ with a horizontal line. The velocities show high correlation between neighbours for windows of size significantly smaller than the cell diameter, while for windows whose sizes are similar to that of the cell diameter, the autocorrelation values are within the expected confidence interval of an iid sequence that follow a normal distribution $\mathcal{N}(0, 1/n)$ where $n$ is the number of velocities.

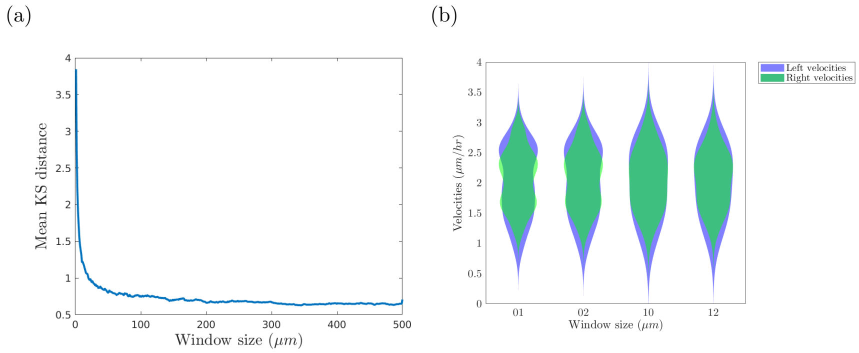


**Figure 3.5:** (a) Mean KS distance for the left and right velocity distributions while varying the window size. (b) Violin plots of the left and right velocities in blue and green, respectively, for different windows of size: 1, 2, 10 and 12 $\mu m$.

## 3.5 Sensitivity analysis of the quantification methods

We start analysing the MEV method performance against the area and the closure rate methods by investigating how each quantification method is affected by the cell migration and proliferation rates. Our aims are to assess how the proliferation rate changes the migration measurements and how does the variability of the migration measurements changes as the cell migration and proliferation rates change.

We use the agent-based model and vary the migration and proliferation probabilities for fixed initial conditions. We decompose the parameter space of migration and proliferation probabilities $[0, 1] \times [0, 0.01]$ into a regular $11 \times 11$ grid with 121 parameter pairs $(p_m, p_p)$. For each parameter combination, 500 simulations were performed and the migration rates analysed using the three quantification methods.

We investigate first how the windowed velocities are affected by the rates of cell migration and proliferation. To calculate the optimal window we consider in the objective function all simulations under the same parameter combination. In Figure 3.6 (a) we present a contour plot of the mean windowed velocity which shows how, as the relevant probabilities increase, the mean velocity increases. A similar trend is observed for the standard deviation (see Figure 3.6 (b)).

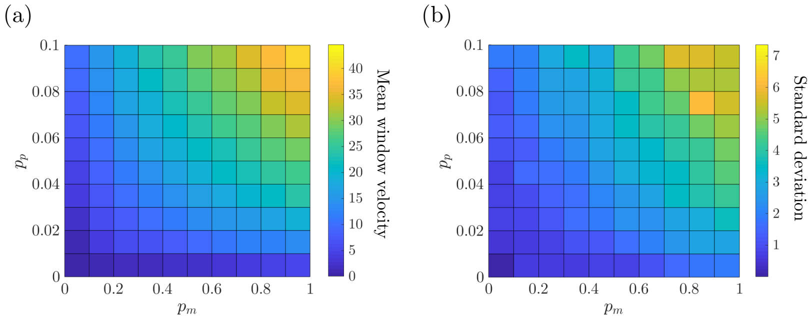


**Figure 3.6:** Sensitivity analysis of the window velocities with respect to the agent-based model parameters. We analyse the variability of the windowed velocities with respect to the proliferation and migration probabilities $(p_m, p_p) \in [0,1] \times [0,0.1]$. In (a) and (b), we plot the mean and the standard deviation of windowed velocities of 500 simulations under each of these 121 parameter pairs.

We analyse the variability of the closure rate and the area methods with respect to the proliferation and migration rates as well. In Figure 3.7 (a) and (b) we plot the mean and standard deviation of the closure rates for the 500 simulations, respectively. The standard deviation of the closure rate is significantly smaller than the one for the windowed velocity and the area method. Similarly as in the velocity distribution case, as the migration and proliferation probabilities increase, the closure rate measurement increases.

We measure the percentage wound area at 6 hours. For most parameter values considered the standard deviation is smaller than 4. Unlike the MEV and the closure rate methods, the standard deviation of the percentage wound area is highest when the motility probability is highest and the proliferation probability lowest. The standard deviation is the lowest, almost zero, when both the proliferation and motility probabilities are at their highest or lowest values. For the area method, as the probabilities increase, the mean percentage area decreases. In Figure 3.7 we observe a similar behaviour when we compare the percentage wound areas at other time points.

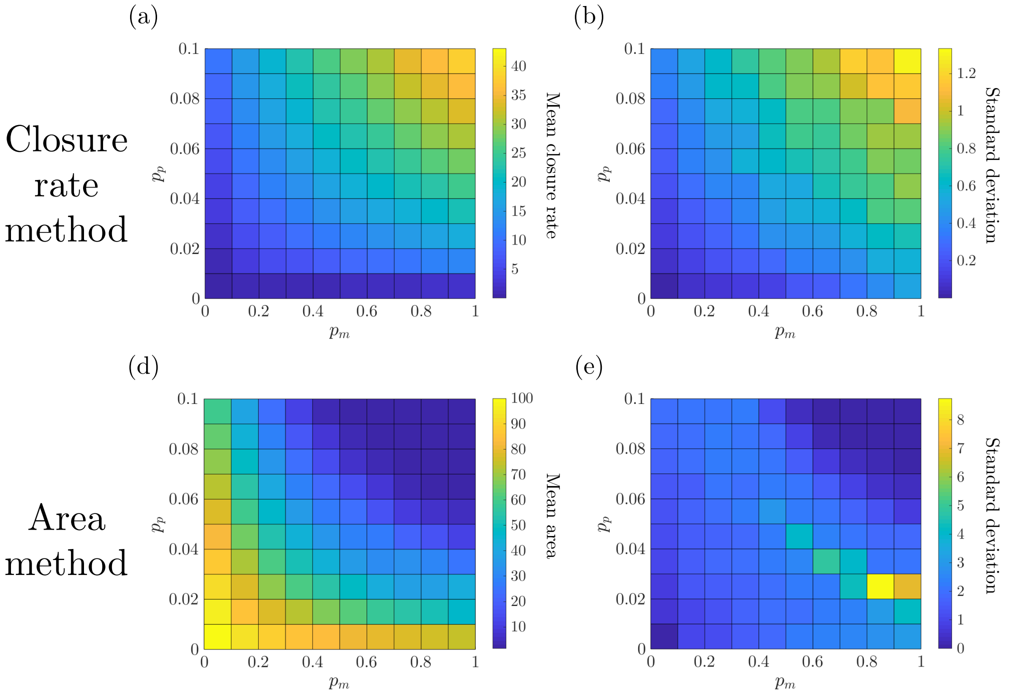


**Figure 3.7:** Sensitivity analysis of the closure rate measurements and the percentage wound area with respect to the agent-based model parameters. We analyse the variability of the area and the closure rate with respect to the proliferation and migration probabilities $(p_m, p_p) \in [0,1] \times [0, 0.1]$. In (a) and (b) we plot, respectively, the mean and the standard deviation of the percentage wound area of 500 simulations under each of these 121 parameter pairs. In (c) and (d) we plot the mean and the standard deviation of the closure rate measurements.

To clarify the difference between the measurement distributions, in Figure 3.8 we construct the violin plots of the measurements as the motility parameter varies and for a fixed proliferation probability ($p_p = 0.1$). While for the closure rate method and the windowed velocity, the measurements increase as the motility parameter increases, the percentage wound areas decrease as we increase the motility parameter.

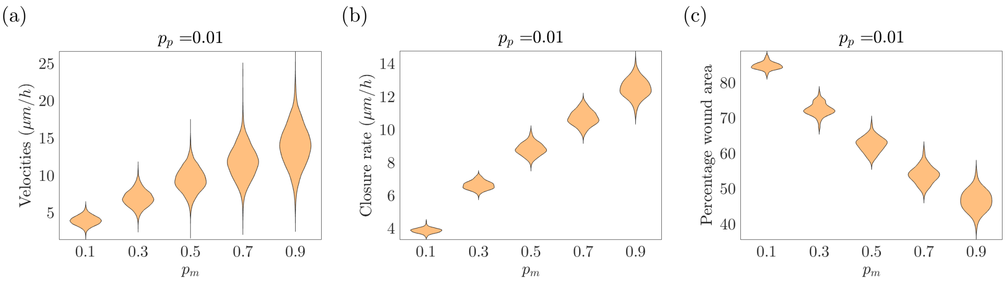


**Figure 3.8:** Violinplots for measurements associated with the three quantification methods while the motility probability and the proliferation probability is held fixed at $p_p = 0.01$. We consider the violinplot of the distribution of (a) window velocities (b) closure rates and (c) areas of 500 simulations for each parameter pair considered.

Finally we analyse the variability of the fitness function with respect to the parameters of the agent-based model. In Figure 3.9 we plot the fitness function as the motility parameter varies and the proliferation probability is held fixed at $p_p = 0.01$.

For all motility parameters considered ,the differences between the fitness functions are not significant and the maximum is attained at $w = 10$. We observe that the optimal window is the preset cell diameter of the agent-based model. This analysis shows that the optimal window size of the objective function in Eq. (3.15) is the cell diameter size for the cell migration and proliferation rates considered. This observation is true for all parameter pairs except when $p_m = p_p = 0$ which is when the cells do not present any dynamics, which is biologically unrealistic.

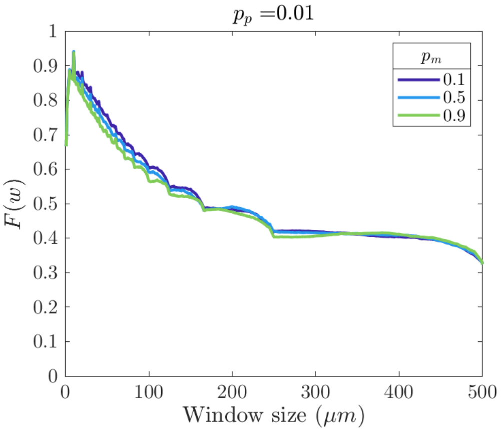


**Figure 3.9:** Sensitivity analysis of the fitness function when varying the agent-based model parameters. We consider different motility parameters and fixed the proliferation probability at $p_p = 0.01$.

These results showed us that the MEV and the closure rate method produce similar mean closure rates and velocity values. We observed also that the distribution of windowed velocities of simulations under the same parameter pair have a larger variance than the closure rates values. The percentage wound areas for simulations under the same parameter pair in general also have a smaller variance than the windowed velocities. The variance of the measurements will affect how well the the measurements characterise the simulations that belong to the same parameter pair. We study the performance of the quantification methods to characterise simulations that belong to the same parameter pair in the next section.

## 3.6 Validation of the quantification method through in-silico data

### 3.6.1 Classification performance test

In order to assess the performance of the three quantification methods in a controlled way, we use the agent-based model to generate in-silico scratch assays data. In particular, we compare the ability of the different methods to distinguish between cell populations with different proliferation and migration parameters. We consider the following classification test:

1. We fix a focal parameter combination $\hat{P} = (p_{\hat{m}}, p_{\hat{p}}) \in [0,1] \times [0, 0.01]$ and run $n$ simulations of the agent-based model using these parameter values.
2. We decompose the parameter space of migration and proliferation probabilities $[0,1] \times [0, 0.01]$ into a regular $11 \times 11$ grid with 121 parameter pairs $(p_m, p_p)$.

For each parameter combination, we run $n$ simulations of the agent-based model.

3. We calculate the cell migration rate in all simulations using the three quantification methods. The migration measurements are windowed velocities, closure rates or areas at specific time points, depending on the quantification method.

4. For each quantification method, we determine whether the migration measurements of each sampled parameter combination $(p_m, p_p)$ are statistically significantly different from those for the focal parameter pair $\hat{P}$. We perform two tests: the two-sample Kolmogorov-Smirnov test and the unpaired two-sample t-test, to which we refer as the K-S test and t-test, respectively. We fix a $p$-value $< 0.05$ to define statistical significance.

We consider a K-S test and a t-test to test for differences at the distribution level and in the mean. We test our data for normality and in case the migration measurements are not normally distributed, we consider a Wilcoxon rank-sum test. We account for stochasticity of the agent-based model by repeating this test for 20 times and analyse the mean and variance of the classification results.

When applying the classification test for the MEV method, we consider a global optimal window size for determining the windowed velocities of the simulations. In this way, we obtain the same number of windowed velocities for each simulation. To determine this global optimal window, we consider a weighted sum of the individual objective functions of each simulation (see Eq. (3.19)). When applying the classification test to the area method, we must specify the time point at which the wound areas are measured and compared. We fix the comparison time to be half the time it takes the leading edges to touch each other in the first simulation, which is a common choice in an experimental setting.

### 3.6.2 Performance analysis

We assume that the focal parameter combination, $\hat{P} = (\hat{p}_m, \hat{p}_p)$, takes values in $\{0.1, 0.5, 0.9\} \times \{0.01, 0.05, 0.09\}$ in order to test the classification for small, medium and high values of cell migration and proliferation in our parameter space. We consider $n = 4$ simulations as the sample size for our test, so as to coincide with experimental settings in which four samples are typically used. We repeat the classification test 20 times to produce results that account for the stochasticity of the system.

In Figure 3.10, we plot the results of the mean behaviour of the classification tests when considering the K-S test and the three focal parameter combinations: $\hat{P} = (0.1, 0.01)$, $(0.5, 0.01)$ and $(0.9, 0.01)$. On each plot, the focal parameter combination is indicated by a red circle. At each position $(p_m, p_p)$, we plot a circle whose colour corresponds to the percentage of times the migration measurements of that parameter pair are statistically significantly different to those for the focal parameters $\hat{P}$ with respect to the colourbar at the left of the plots. For parameter pairs different from the focal parameter, $(p_m, p_p) \neq \hat{P}$, the percentage of times the migration measurements of that parameter pair are statistically significantly different to those for the focal parameters $\hat{P}$ is smaller than 100%, is an indication of the presence of false negative results. For the focal parameter, $(p_m, p_p) = \hat{P}$, if this percentage is higher than 0% then it indicates the presence of false positive results. We note that for $\hat{P} = (0.1, 0.01)$, the method does not produce false negative results: the K-S test indicates that the windowed velocities from simulations of parameter pairs different from the focal parameters, $(p_m, p_p) \neq \hat{P}$, are statistically significantly different to the windowed velocities from simulations of the focal parameter pair 100% of the time (Figure 3.10 (a)). For $\hat{P} = (0.5, 0.01)$, there are four parameter pairs for which the velocities were 80%, 85%, 85% and 95% times statistically significantly different to those for the focal parameter (Figure 3.10 (b)). For $\hat{P} = (0.9, 0.01)$, the number of parameter pairs for which the percentage is not 100% increases (Figure 3.10 (c)). We observe that as the migration rate increases, the classification performance worsens. The intra-sample difference is accounted for when considering the classification tests for the parameter pair $(p_m, p_p) = \hat{P}$. We observe that the method gives false positive results fewer than 34% of the time. This percentage decreases as the proliferation probability increases (see Appendix A.1).

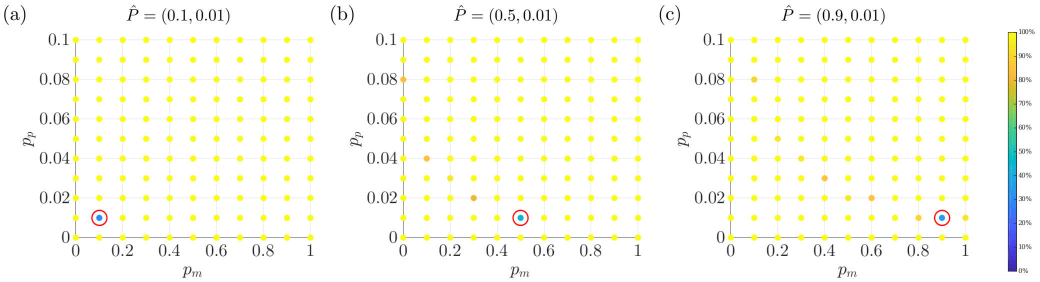


**Figure 3.10:** Plots of the mean behaviour of the classification tests for the monolayer edge velocimetry method. The classification tests are performed by considering a K-S test, a sample set of n=4 simulations and the focal parameters (a) $\hat{P} = (0.1, 0.01)$, (b) $\hat{P} = (0.5, 0.01)$, and (c) $\hat{P} = (0.9, 0.01)$. In each plot at each parameter pair $(p_m, p_p)$, the colour of the circle denotes the percentage of times the migration measurements of that parameter pair are statistically significantly different to those for the focal parameter $\hat{P}$. We indicate the focal parameter pair with a red circle. The plots illustrate how the classification performance of the method varies as the migration parameter varies. The method performs better when the migration parameter is small.

By applying the classification test to simulations generated using the same parameter values as the focal parameter set, we can establish the degree of similarity between velocity distributions from simulations generated using the same set of parameter values. Another more common way to assess this is to consider a set of simulations under the same parameter combination and analyse if there are statistical significant differences between the velocity distributions. For each parameter pair in our sampled parameter space, we consider a set of 4 simulations and apply a Kruskal-Wallis test between the velocity distributions of each simulation. We consider a Kruskal-Wallis test since the velocities do not follow a normal distribution. To account for the stochasticity of the agent-based model, we repeat this test 100 times. In Figure 3.11 we plot the percentage of times the Kruskal-Wallis test detected a statistically significant difference between the velocity distributions. We observe that as the proliferation and the motility probabilities decrease, the percentage of times that the Kruskal-Wallis test detected no statistically significant differences, increases. In general the percentages are between 30 and 40 percent which indicates a high variability between the velocity distributions under the same parameter combination.

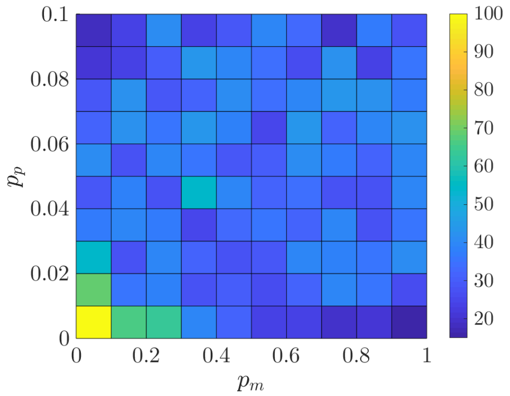


**Figure 3.11:** Contour plot of the percentage of times the Kruskal-Wallis test detected no differences between the velocity distributions from simulations under the same parameter combination.

### 3.6.3 Statistical comparison performance of MEV to standard migration quantification methods

We compare the classification performance of the MEV method with the closure rate and the area methods [90, 153]. As before, the focal parameter combination, $\hat{P}$, takes values in $\{0.1, 0.5, 0.9\} \times \{0.01, 0.05, 0.09\}$. We consider $n = 4$ simulations as the sample size and repeat the classification test 20 times.

In Figure 3.12, we plot the mean behaviour of the classification tests for the three quantification methods by applying the K-S test and the focal parameter combinations $\hat{P} = (0.1, 0.01)$, $(0.5, 0.01)$ and $(0.9, 0.01)$. We observe that for a focal parameter pair, the MEV method yields fewer incorrect classifications. We also observe that as the proliferation rate increases, the percentage number of incorrect classifications increases for the three methods.

The results of the classification tests for all other focal parameter combinations in $\{0.1, 0.5, 0.9\} \times \{0.01, 0.05, 0.09\}$ are presented in the Appendix A.1. Overall we observe that our method outperforms the closure rate and the area method. For all focal parameter combinations tested, the MEV method yields a greater percentage of correct classifications. The performance of the area method is the worst while the performance of the closure rate method is intermediate between our method and the area method. The performance of all three methods declines as the values of the migration and proliferation rates of the focal parameters $\hat{P}$ increase. The optimal window sizes for each classification test are of similar size as the cell diameter.

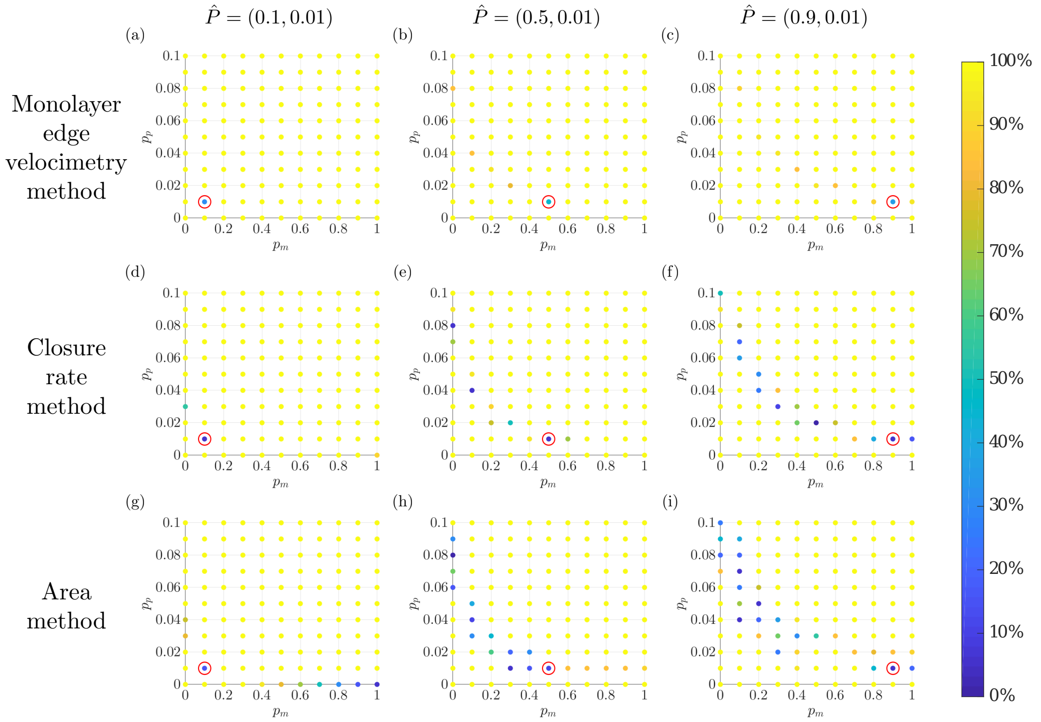


**Figure 3.12:** Series of plots showing how the performance of the three quantification methods changes as the migration rate of the focal parameters varies. In each plot, the colour of the circle at each parameter pair $(p_m, p_p)$ indicates the percentage of times the migration measurements associated with the parameter pair are statistically significantly different from those associated with the focal parameters $\hat{P}$. The focal parameters $\hat{P}$ are indicated by a red circle. The results reveal that the monolayer edge velocimetry method yields a better statistical classification than the other methods. We note also the performance of all three methods declines as the migration rate of the focal parameters $\hat{P}$ increases.

### 3.6.4 Performance analysis considering non-smooth leading edges

We have analysed the classification performance of the MEV method through agent-based model simulations. These simulations were initialised using ideal (i.e. smooth) initial conditions such as that depicted in Figure 3.2. We now analyse the performance of the quantification method for initial conditions with irregular leading edges. We obtain realistic initial conditions for the agent-based simulations, from the initial images of the in-vitro scratch assays. We use only 20 initial images from the 24 scratch assays since four assays have large holes within the monolayer. The 20 initial images of the in-vitro scratch assays can be seen in Figure 3.13.

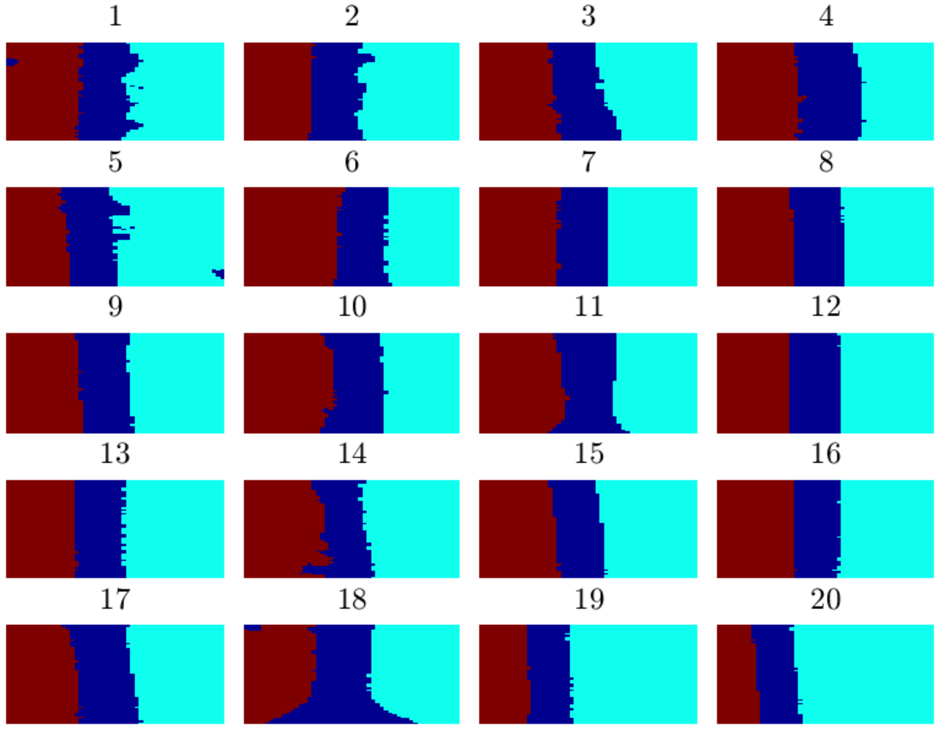


**Figure 3.13:** Initial images of the in-vitro assays that are used as the initial condition of the agent-based simulations.

We perform 500 simulations of the agent-based model randomly selecting as an initial condition one of the 20 initial images of the in-vitro scratch assays. We apply the three quantification methods to the simulations and perform the same classification performance test that was introduced in Section 3.6.1. As in the previous section, we assume the focal parameter combination, $\hat{P} = (\hat{p}_m, \hat{p}_p)$, takes values in $\{0.1, 0.5, 0.9\} \times \{0.01, 0.05, 0.09\}$, we consider $n = 4$ simulations as the sample size and repeat the classification test 20 times.

We include the results of all the classification tests for all the possible focal parameter combination values in Appendix A.1. To compare the performance of the quantification methods for smooth and irregular leading edges, in Figure 3.14 we present the results of the classification method for the focal parameter $\hat{P} = (0.5, 0.05)$ when considersing smooth leading edges (left) and non-smooth leading edges (right). The performance of all quantification methods decreases when analysing the non-smooth leading edges. The area method performs worst; it is unable to detect significant differences for most of the parameter pairs in the parameter space. The MEV method is superior to the other methods for smooth and non-smooth leading edges, and this trend persists for all other focal parameter pairs considered (see Appendix A.1).

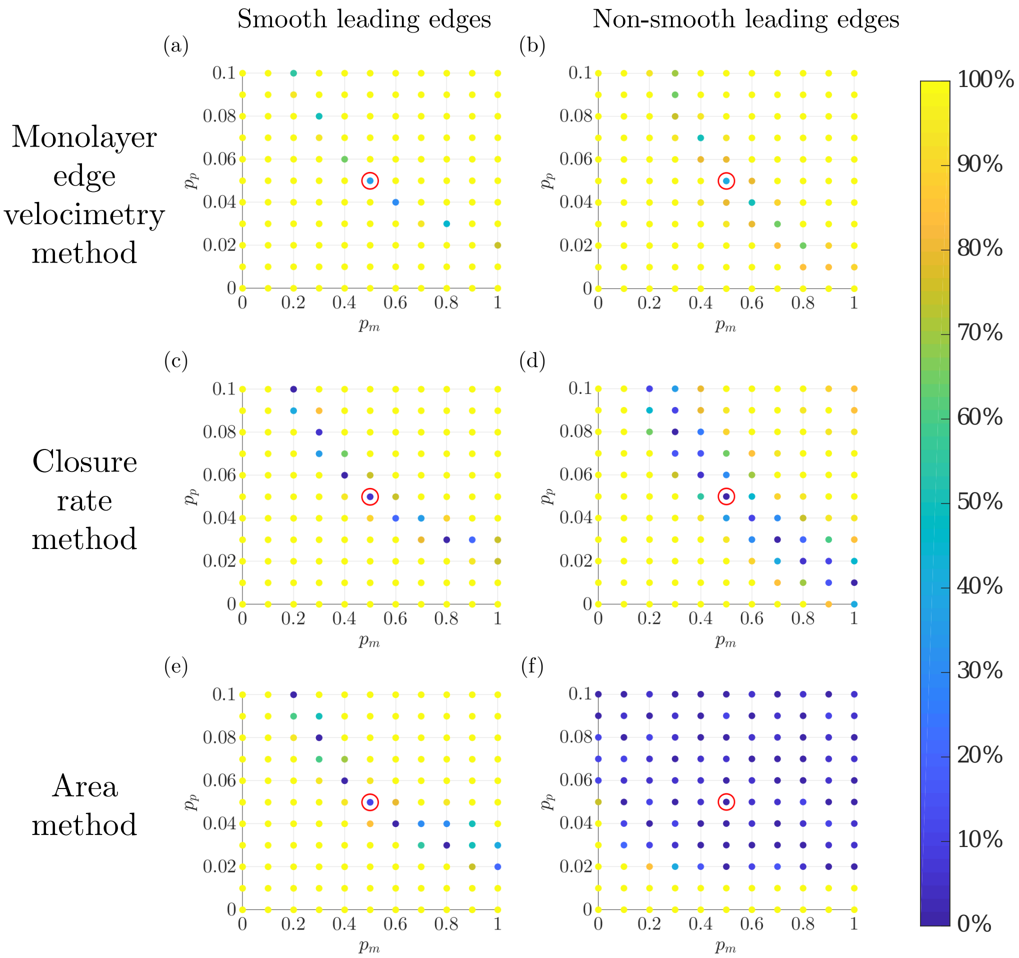


**Figure 3.14:** Plots of the mean behaviour of the classification tests of the three quantification methods when considering smooth (left) and non-smooth leading edges (right). The classification tests are performed by considering a K-S test, a sample set of n=4 simulations and the focal parameter $\hat{P} = (0.5, 0.05)$. In each plot, at each parameter pair $(p_m, p_p)$, the colour of the circle denotes the percentage of times the migration measurements of that parameter pair are statistically significantly different to those for the focal parameter $\hat{P}$. We indicate the focal parameter pair with a red circle. The results reveal that the monolayer edge velocimetry method yields a better statistical classification than the other methods when analysing ideal and non smooth leading edges. We note also that the performance of all three methods declines when the data are non-smooth.

## 3.7 Application of the quantification methods to in-vitro data

We quantify the migration velocity of scratch assays for the different cell types using the MEV method. We determine the global optimal window by calculating the objective function for the 24 scratch assays. We vary the window size $w$ from 1 to 500$\mu m$ with a step size of 1$\mu m$ and use Eq. (3.7) to calculate the objective function $F(w)$. The objective function and the three fitness functions that contribute to its calculation are shown in Figure 3.15. The maximum value is attained for a window size of 16$\mu$m. For a fixed window size ($w = 16\mu$m), we use a linear approximation to describe the position over time of the fronts and determine the 64 representative windowed velocities for each scratch assay and visualise their boxplots in Figure 3.16.

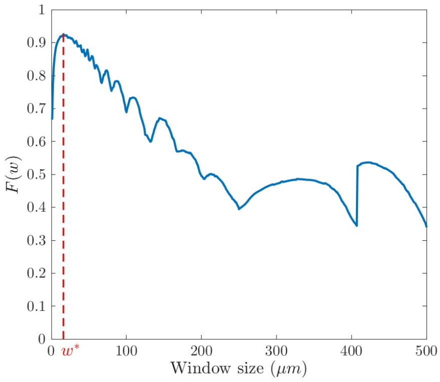


**Figure 3.15:** Plot of the objective function for the in-vitro data. The optimal window size, $w^* = 16\ \mu$m is indicated with a dashed line in red.

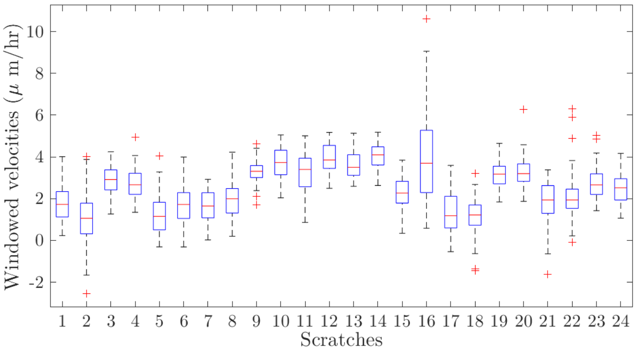


**Figure 3.16:** Boxplots of the windowed velocities with respect to the optimal window size 16 $\mu$m for each experimental scratch assay.

We present the plots of the three fitness functions that contribute to the objective calculation, applied to the in-vitro data in Figure 3.15 .

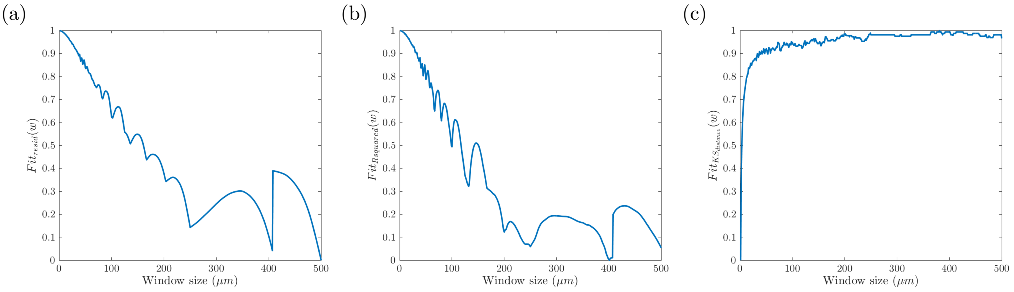


**Figure 3.17:** Plot of the fitness functions that constitute the objective function: (a) $Fit_{resid}$, (b) $Fit_{Rsquared}$ and (c) $Fit_{KS_{distance}}$. The functions have been rescaled so their values are between 0 and 1.

### 3.7.1 Statistical classification by the MEV method

After grouping the velocities of scratch assays from the same cell type, the migration rate of each cell group is represented by 256 velocities. The boxplots associated with the velocity distributions for the six groups are shown in Figure 3.18. To determine how different the migration rate of cell group S1 is from the others, we perform a

K-S test to test the null hypothesis that the velocities from group S! and one of S2, S3, . . ., S6 come from the same distribution. The null hypothesis was rejected for groups S2, S3 and S4 with statistical significance level of $p_{value} \leq 0.0001$. The null hypothesis was rejected for group S6 with statistical significance level of $p_{value} \leq 0.05$. For group S5, the null hypothesis was not rejected. We performed a t-test between S1 and each of the other groups to determine whether the mean difference is statistically significant. The mean difference between the velocities for cell groups S1 and S2, S3 and S4 is statistically significant at the 0.0001 level. There was statistical significance in the mean difference with respect to S6 at the 0.05 significance level. The statistical results for the K-S tests and t-tests are reported in Figure 3.18.

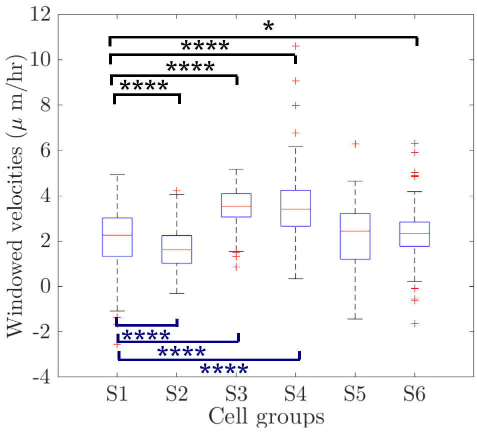


**Figure 3.18:** Statistical analysis of the experimental data using the MEV method. The migration measurements are grouped into the six different groups (S1-S6). Above the data, in black, we have reported the statistical significance results from performing a K-S test with respect to the S1 group. Below the data, we have done the same for the t-tests. Considering the windowed velocities, with respect to the K-S test and t-test, the null hypothesis was rejected when testing group S1 against group S2, S3 and S4 at the 0.001 significance level. The statistical significance level is encoded in the symbols: *:= $p_{value} \leq 0.05$, **:= $p_{value} \leq 0.01$, ***:= $p_{value} \leq 0.001$ and ****:= $p_{value} \leq 0.0001$.

We present the results of performing the unpaired two-sample t-test (t-test), the Wilcoxon rank-sum test and the two-sample Kolmogorov-Smirnov test (K-S test) between the migration measurements of the S1 group against the other groups' measurements. We fix a $p$-value $< 0.05$ to define statistical significance. We indicate in each row the statistical test performed and in each column we indicate if the hypothesis was rejected ($h = 1$) or not ($h = 0$), and the corresponding p-value for each hypothesis test. The exact value of the $p_{value}$ for each test is reported in Table 3.2.

| | S2 | S3 | S4 | S5 | S6 |
|---|---|---|---|---|---|
| t-test | h=1,<br>p=1.101e-07 | h=1,<br>p=3.489e-46 | h=1,<br>p=5.102e-27 | h=0,<br>p=3.822e-01 | h=0,<br>p=1.026e-01 |
| Wilcoxon rank sum test | h=1,<br>p=3.626e-09 | h=1,<br>p=7.146e-41 | h=1,<br>p=4.907e-25 | h=0,<br>p=3.401e-01 | h=0,<br>p=4.520e-01 |
| K-S test | h=1,<br>p=4.026e-09 | h=1,<br>p=1.668e-33 | h=1,<br>p=5.865e-18 | h=0,<br>p=1.041e-01 | h=1,<br>p=1.093e-02 |

**Table 3.2:** Hypothesis test results comparing S1 and the other groups' windowed velocities. In each row the statistical test performed is indicated.

Each scratch is represented by a group of 64 representative windowed velocities and there are four scratches per cell line. We analyse if there are statistically significant differences between the velocity distributions of scratches from the same cell line. Since the velocity distributions are not distributed normally, we assess statistically differences between the velocities from different scratches using the non-parametric Kruskal–Wallis test. For each cell line in Table 3.3 we include the p-value when assessing statistically significant differences. We notice that for all the groups the p-value is smaller than 0.05 so there are statistical significant differences between the distributions. This result is expected since for the in-vitro data we observed that even if the simulations are from the same parameter combination, the null hypothesis is rejected for the Kruskal–Wallis test, i.e. there are statistically significant differences between the distributions (see Section 3.6.2 ).

| | S1 | S2 | S3 | S4 | S5 | S6 |
|---|---|---|---|---|---|---|
| p-value | 7.389e-22 | 5.418e-05 | 7.407e-08 | 8.848e-19 | 2.945e-34 | 3.822e-08 |

**Table 3.3:** Results of the Kruskal-Wallis tests when analysing if the velocity distributions of scratches from the same cell line have statistical significant differences. The table shows the p-value of the Kruskal-Wallis test for each cell line. If the p-value is smaller than 0.05, the test concludes that there is a statistically significant difference between the velocity distributions.

For each cell line, we consider the largest subset for which there is no statistically significant difference between the velocity distributions. In Table 3.4 we show the p-values of applying the Kruskal-Wallis test to these subsets.

| | S1 | S2 | S3 | S4 | S5 | S6 |
|---|---|---|---|---|---|---|
| p-value | 1.041e-01 | 1.893e-01 | 9.127e-01 | 6.269e-01 | 5.356e-01 | 8.938e-01 |

**Table 3.4:** Kruskal-Wallis results for analysing the within-group and between-group variation of each cell line of the in-vitro data.

We group the velocities of each cell line such that the intra-group and inter-sample variance are not statistically significant and perform a t-test, Wilcoxon rank sum test and Kolmogorov Smirnov test between the windowed velocities of cell group S1 against the windowed velocities of each of the other groups. We include the results of the statistical tests in Table 3.5.

| | S2 | S3 | S4 | S5 | S6 |
|---|---|---|---|---|---|
| t-test | h=1,<br>p=1.370e-25 | h=1,<br>p=6.908e-13 | h=1,<br>p=7.338e-06 | h=1,<br>p=8.785e-09 | h=1,<br>p=1.724e-06 |
| Wilcoxon rank sum test | h=1,<br>p=2.285e-23 | h=1,<br>p=8.747e-15 | h=1,<br>p=1.496e-05 | h=1,<br>p=4.387e-09 | h=1,<br>p=4.534e-07 |
| ks-test | h=1,<br>p=3.755e-18 | h=1,<br>p=1.779e-13 | h=1,<br>p=3.159e-05 | h=1,<br>p=3.576e-08 | h=1,<br>p=1.724e-05 |

**Table 3.5:** Statistical analysis of the experimental data using the MEV method. We compare the windowed velocities of the cell group S1 against the windowed velocities of each of the other five groups. In each row the statistical test performed is indicated.

Table 3.5 reveals statistically significant differences between the windowed velocities of the cell group S1 and the windowed velocities of all the other groups. The MEV method detects more differences than before.

### 3.7.2 Comparison of the statistical classification from other standard migration quantification methods

We now compare the statistical results of our quantification method against those for the area and closure rate methods. In Figure 3.19 we plot the closure rates of each group and report the results from performing the K-S test and t-test between S1 and the other groups. We consider this comparison since it was the one our experimental collaborators were interested. S3 was the only group for which the null hypothesis of the K-S test and the t-test was rejected at the 0.05 significance level. When we performed the statistical tests for the percentage area measurements, no significant difference was found. In Table 3.6 and 3.8 we include the results of the K-S and t-tests for the percentage wound area measurements.

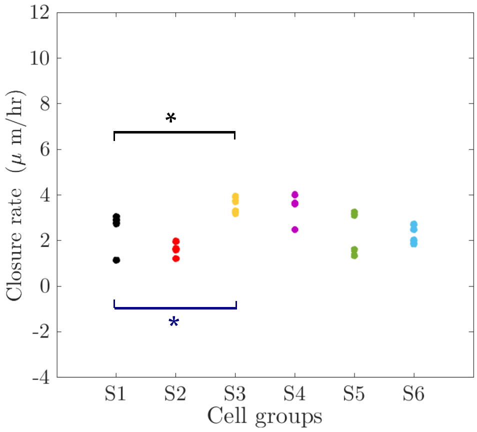


**Figure 3.19:** Statistical analysis of the experimental data using the closure rate method. The closure rates are grouped into the six different groups (S1-S6). Above the data, in black, we have reported the statistical significance results from performing a K-S test with respect to the S1 group. Below the data, we have done the same for the t-tests. The null hypothesis was rejected at the significance level of 0.05 between S1 and S3. The statistical significance level is decoded in the symbols: *:= $p_{value} \leq 0.05$, **:= $p_{value} \leq 0.01$, ***:= $p_{value} \leq 0.001$ and ****:= $p_{value} \leq 0.0001$.

We present the results of performing the unpaired two-sample t-test (t-test), the Wilcoxon rank-sum test and the two-sample Kolmogorov-Smirnov test (K-S test) between the migration measurements of the S1 group against the other groups' measurements. We fix a $p$-value $< 0.05$ to define statistical significance. We indicate in each row the statistical test performed and in each column we indicate if the hypothesis was rejected ($h = 1$) or not ($h = 0$), and the corresponding p-value for each hypothesis test.

| | S2 | S3 | S4 | S5 | S6 |
|---|---|---|---|---|---|
| t-test | h=0,<br>p=1.199e-01 | h=0,<br>p=6.562e-02 | h=0,<br>p=1.291e-01 | h=0,<br>p=8.490e-01 | h=0,<br>p=7.062e-01 |
| Wilcoxon rank sum test | h=0,<br>p=3.429e-01 | h=1,<br>p=2.857e-02 | h=0,<br>p=2.000e-01 | h=0,<br>p=6.857e-01 | h=0,<br>p=3.429e-01 |
| K-S test | h=0,<br>p=1.075e-01 | h=1,<br>p=1.107e-02 | h=0,<br>p=1.075e-01 | h=0,<br>p=5.344e-01 | h=0,<br>p=1.075e-01 |

**Table 3.6:** Hypothesis test results comparing S1 and the other groups' closure rates. In each row the statistical test performed is indicated.

The evolution of the percentage wound area of each scratch assay during the course of the experiment is shown in Figure 3.20.

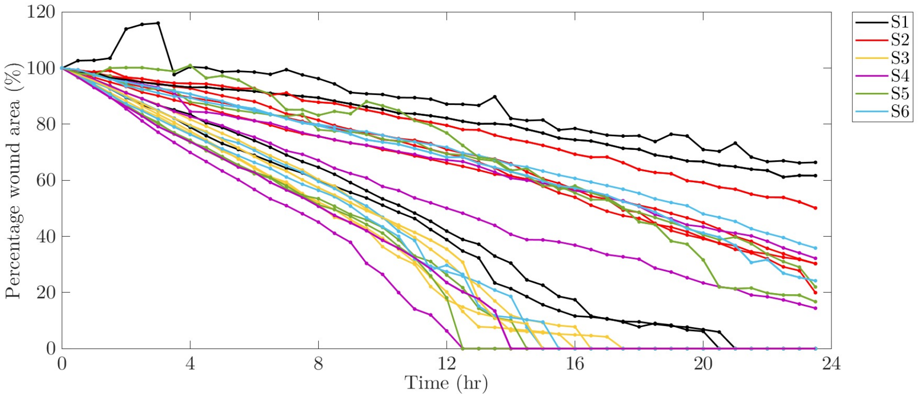


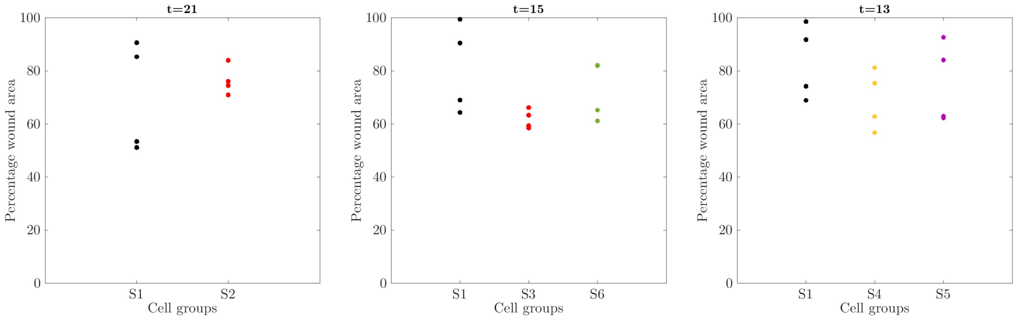
**Figure 3.20:** Evolution of the percentage wound area during the course of the experiment of each scratch assay. Experiments of the same cell group are plotted with the same colour.

We calculated the time of comparison for the different groups, and the results are presented in Table 3.20.

| | S2 | S3 | S4 | S5 | S6 |
|---|---|---|---|---|---|
| Time of comparison (hr) | 21 | 15 | 13 | 13 | 15 |

**Table 3.7:** Comparison times for the wound percentage areas of the different groups. The time of comparison was set to be half the time it takes for the first scratch being compared in which the leading edges touch.

The wound percentage area at the comparison times are shown in Figure 3.21.



**(a)** $t = 21$ hr **(b)** $t = 15$ hr **(c)** $t = 13$ hr

**Figure 3.21:** Percentage wound area between S1 and the other groups at the time of comparison.

The results of the statistical tests applied to the percentage wound area at the times of comparison are shown in Table 3.8.

| | S2 | S3 | S4 | S5 | S6 |
|---|---|---|---|---|---|
| t-test | h=0,<br>p=5.819e-01 | h=0,<br>p=6.936e-02 | h=0,<br>p=1.632e-01 | h=0,<br>p=4.790e-01 | h=0,<br>p=4.447e-01 |
| Wilcoxon rank sum test | h=0,<br>p=1.000e+00 | h=0,<br>p=5.714e-02 | h=0,<br>p=3.429e-01 | h=0,<br>p=4.857e-01 | h=0,<br>p=4.857e-01 |
| K-S test | h=0,<br>p=5.344e-01 | h=0,<br>p=1.075e-01 | h=0,<br>p=5.344e-01 | h=0,<br>p=5.344e-01 | h=0,<br>p=5.344e-01 |

**Table 3.8:** Hypothesis test results comparing S1 and the other groups' percentage wound areas. In each row the statistical test performed is indicated.

## 3.8 Discussion

In this chapter we have introduced a new migration quantification method for scratch assays that characterises the horizontal component of the front velocity of cell monolayers. The method involves three steps: (1) determination of an optimal window $w^*$ with which to approximate the cell front by a function which is piecewise constant in segments of length $w^*$; (2) approximation of the interface with respect to the window size $w^*$ at each time point; and (3) linear approximation of the position over time of the interface in each of these windows. In this way we characterise cell migration in the scratch assay by the slopes of a series of linear approximations to the interface position over time in these windows. The optimal window is chosen to be the one that best fits a constant velocity profile and for which the left and right front velocities can be considered to be samples from the same distribution.

We used an agent-based model that mimics the scratch assay, to tested the ability of our quantification method to distinguish between cell lines with known cell migration and proliferation rates. As the migration and proliferation rates increased, the mean and variance of the windowed velocities increased. This behaviour was expected since increases in cell migration and proliferation rates increase the interface velocity and the variance of the system increases as the migration and proliferation rates increase. We observed that the optimal window size matched the cell size given for the agent-based model. By comparing our quantification method with two existing methods, we observed that our method outperforms both yielding a greater percentage of correct classifications than the other methods across a range of parameter values. Our method also made significantly fewer statistical errors than the other two tested methods. Despite being widely used, the performance of the area method was the worst, while the performance of the closure rate method was intermediate between our method and the area method. The poor performance of the area method is due to the presence of irregular cell-free areas and to the indirect quantification of migration by a single time-point measurement as we observed when analysing the non-smooth leading edges. The poor performance of the closure rate method is also related to irregularities in the data since it effectively quantifies migration from the slope of the linear approximation to the mean position of the interface over time.

After showing that our quantification method performs well on in-silico data, we then used it to analyse our experimental data set. We calculated an optimal window size of 16 $\mu m$, which is of the same order as the mean cell diameter size, and then determined the corresponding windowed horizontal velocities. By performing two sample Kolmogorov-Smirnov (K-S) and unpaired two-sample t-tests, we identified a statistically significant difference between the S1 group and groups S2-S4. The K-S test also indicated statistically significant differences with respect to group S6. We used these two tests since we wanted to detect differences at the distribution level (through the K-S test) and at the mean level (through the t-test). The closure rate method only detected statistically significant differences between S1 and S3. The closure rate data are of poor quality: more samples or new samples with smoother leading edges are needed to analyse the migration rate with this method. The area method was unable to detect any statistically significant differences in the dataset. Even when we tried different time points, there was no significant difference. We observed that the S1 cell group also exhibited the highest levels of expression of target genes associated with malignancy and poor prognosis, when analysed by qRT-PCR techniques (data not shown) in agreement with the detected significant differences in migration.

This study has some limitations which could be addressed in future work. For the in-vitro experiments, the scratch was created by the removal of silicone inserts. It has been observed that the removal of silicone inserts can damage the culture surface coating and affect the cell migration rate [143]. We assume this effect is present in all cell samples so the comparative migration analysis performed by the quantification methods is not affected. Using our quantification method, we would be able to establish how the different procedures for creating the scratch affect migration rates. The statistical performance of our quantification method can be further validated on publicly available wound healing experiment data sets such as those in [230], which provide sets of assays and replicates under different experimental conditions. We observed that there is a high variability between velocity distributions under the same parameter, this behaviour can give rise to false positive classifications as we observed in the in-silico data. This weakness can be addressed by characterising the dynamics of the scratch assays with additional information such as the evolution of the pair-correlation function over the scratch domain [5]. In spite of that the classification performance of our quantification method is superior to the area and the closure rate method. In this study, the agent-based model accounted solely for proliferation and migration events. Additional interactions could be included to give a more realistic description of the scratch assay and the impact of how these factors affect the quantification method could be assessed. The quantification algorithm assumes a linear approximation of the horizontal cell monolayer displacement. Our method could be adjusted to account for an initial phase during which the cells react to the presence of the wound, so the cell monolayer front position over time can be fitted to a Richards function, a non-symmetrical sigmoid function, as in [207]. Currently the method does not address image boundary effects that could potentially affect the quantification. The method could be improved to allow it to deal with the image boundary in a more precise way. One of the drawbacks of our quantification method is the uncertainty in the optimal window size since it depends on the samples. However, in our study we found that the optimal window size is of the order of the mean cell diameter and the statistical power and measurements of the quantification method are the same for window sizes of the same order. If the mean cell diameter is known, then the optimisation procedure can be omitted and the linear approximation can be performed using the mean cell diameter as the window size. The framework can also be extended to consider a time-dependent velocity field across the full monolayer, such as Cell Image Velocimetry (CIV) [159]. The challenge in this case would be to determine to which measurements we should apply statistical tests to detect significant differences in migration rates.

In summary, we have introduced a new method for migration quantification of typical scratch assay data, which can be of low quality. Many of the challenges we have overcome with our method could be avoided through improved experimental design. However, the latter would require: repeating the experiment or using more sophisticated experimental tools to create the same degree of confluence and uniform “wounds”. These are costly solutions. We note that our quantification method can be applied to the existing data. Furthermore, through the classification test based on in-silico data, we show that even when the quality of the scratches is ideal, our quantification method is better at detecting differences in migration than the other standard methods.

## Chapter 4

# Age-structured model that captures delayed logistic proliferation behaviour in scratch assays

Mathematical models can help in the detection of the underlying mechanisms responsible for experimental observations. Analysing a mathematical model of a biological phenomenon can help us in the understanding of the main mechanisms underlying the observed phenomenon and highlight gaps in our understanding.

In this chapter we introduce an age-structured model that accounts for the different proliferation phases that have been previously observed experimentally in scratch assays [118]. Parts of this chapter are based on the peer reviewed publication: [179].

## 4.1 Motivation

As we mentioned in Sec.1.2 several mathematical models have been developed to describe scratch and proliferation assays and to test hypotheses concerning the mechanisms that govern cell migration and proliferation. Given the importance of cell migration and proliferation in biology, it is imperative to evaluate how well mathematical models describe these processes and analyse if they are capable of replicating experimental observations. If they are not, then it is necessary to investigate which of the mechanisms that were neglected need to be included, or if some biologically relevant mechanisms have yet not been discovered.

In [118] the ability of the logistic growth model to describe cell proliferation in scratch assays was studied. A series of scratch and proliferation assays using PC-3 prostate cancer cells were performed and the changes in cell density in two subregions located far from the scratch were quantified (see Section 1.1.2 for a description of scratch and proliferation assays). These two subregions were chosen so that the changes in cell density were not due to cell migration which occurs at the border of the scratch. To assess the suitability of the logistic growth model, the authors analysed the model fit to experimental data regarding the per capita growth rate of the experimental data, $\sigma(N) = \frac{dN}{dt}\frac{1}{N}$, as a function of the cell density, $N$. Calibrating solutions of the logistic model to the experimental data showed a good fit for both assays, however, analysing the per capita growth rate revealed different behaviours between the proliferation and the scratch assays. The authors observed that for proliferation assays, the per capita growth rate could be well described by a linearly decreasing function of the cell density (see Figure 4.1 (c)), a result consistent with the logistic model. However, for the scratch assay data, during the first 18 hr of the experiment, the per capita growth rate was found to increase with cell density, whereas for $8 \leq t \leq 48$ hr, the per capita growth rate was found to decrease approximately linearly with the cell density (see Figure 4.1 (d)).

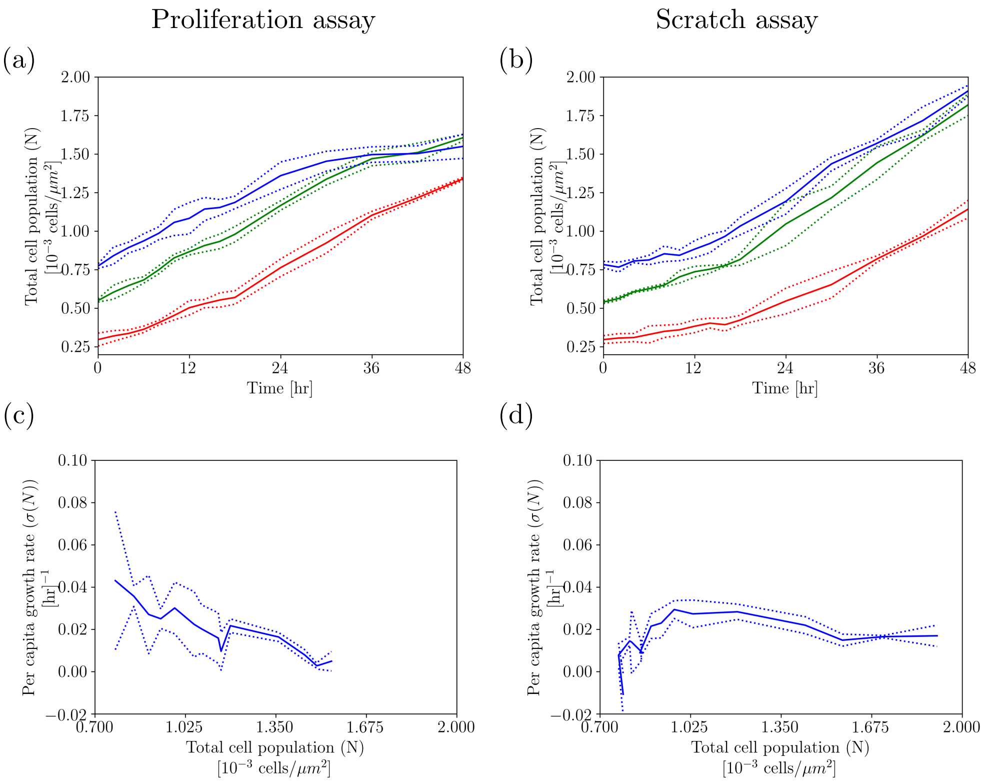


**Figure 4.1:** Experimental data from the proliferation and scratch assays performed in [118]. The authors consider three replicates for three initial degrees of confluence. In (a) and (b) the mean and the standard deviation of the cell population are plotted for each initial degree of confluence (denoted by the different colours). In (c) and (d) the mean and the standard deviation of the per capita growth rate $\sigma(N) = \frac{dN}{dt}\frac{1}{N}$ is plotted with respect to the mean cell population $N$. For plots (c) and (d) we consider only the assays with the highest degree of confluence, the plots for the other degrees of confluence can be found in Appendix B.2. We calculate the per capita growth rate in the same way as in [118]. A biphasic trend can be observed in (d) (for the scratch assay) but not in (c) (for the proliferation assay).

Guided by their experimental observations, the authors in [118] proposed that cell proliferation in scratch assays involves two phases: an initial *disturbance phase* during which proliferation is not logistic and a *growth phase* during which proliferation is approximately logistic, see Figure 4.2 for a schematic representation of their hypothesis. Since this behaviour was not observed in the proliferation assays, the authors concluded that it was caused by the scratching procedure. They hypothesised that scratching might create chemical or mechanical disturbances but did not test their hypothesis.

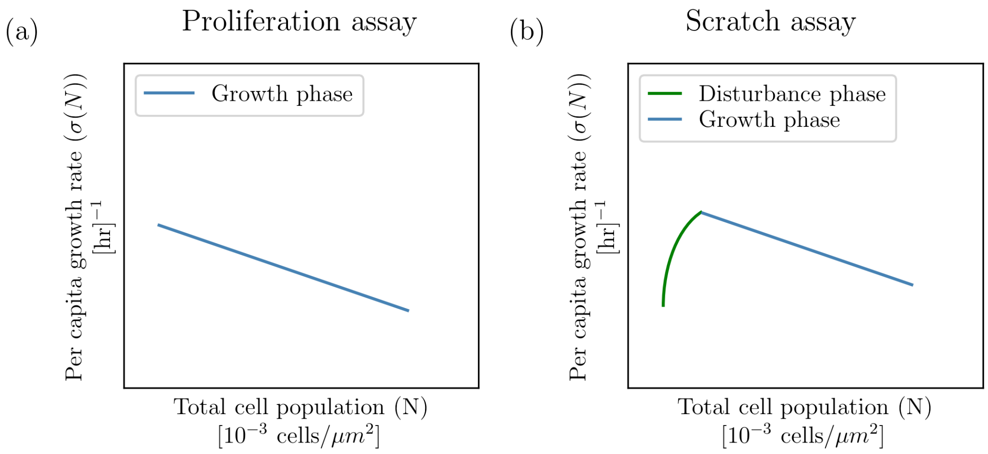


**Figure 4.2:** Schematic illustration of the proliferation phases present in proliferation and scratch assays postulated by [118]. The proliferation assays only exhibits the growth phase while the scratch assay exhibits first a disturbance phase and then the growth phase.

In this chapter, we show that a nonlinear age-structured model can account for the disturbance and growth phases observed in scratch assays. We refer to *age* of a cell as the time elapsed since cell division.

By considering an age-structured model we can investigate how heterogeneity in the cell population age distribution affects the total cell population dynamics. Previous work has shown that heterogeneity in cell age distribution may still generate logistic growth at the population scale [117]. Other age-structured models have described how the cell age distribution can affect the overall cell population dynamics in scratch assays: the speed with which the cells invade the vacant region [81, 220], the efficacy of anti-cancer drugs, particularly phase-specific drugs [26] and the influence of growth factors [27]. Agent-based models have also been considered to study the effect of heterogeneity in cell age distribution on the overall dynamics in in-vitro assays [223, 224].

We consider an age-structured model that was first introduced in [53]. The model considers the interplay between a single cell population and the resource concentration available in the substrate. The model assumes a resource-dependent G1/S transition age, above which, cells are able to proliferate. This critical age naturally divides the cell population into “immature” and “mature” cells. By studying the dynamics of the full system in terms of these subpopulations, we are able to find conditions under which the cell population evolution is of logistic-type and the per capita growth rate follows a biphasic behaviour. We verify our predictions via numerical simulations and show that the resource concentration regulates the disturbance phase.

## 4.2 Age-structured model with resource-regulated proliferation

To take into account the cell cycle heterogeneity of a cell population, we consider a model of the McKendrick-von Foerster type [126, 137]. The model considered in this work was first used to describe the mean-field dynamics of a stochastic multiscale model of a cell population with oxygen-regulated proliferation [53]. Here the model describes the dynamics of an in-vitro cell population assumed homogeneously distributed in space, like cell cultures in proliferation assays or cells in far away regions from the wound in scratch assays, as in [118]. We are interested in the population dynamics with respect to the resources available in the medium. We denote by $n(a,t)$, the number of cells of age $a$ at time $t$. We consider $a \in [0,\infty]$. We denote by $T > 0$ the duration of the experimental observation. The model focuses on the cell population proliferation dependence on the growth factors, oxygen and nutrients available in the medium. We refer to these components as a single, generic *resource* and denote it by $c(t)$.

We assume cells mature with constant speed 1, die with rate $\mu$ and proliferate at a rate $b(a,c(t))$ which we consider to be age and resource-dependent [166, 198]. Combining the above assumptions, it is straightforward to show that the evolution of the cell density function $u(a,t) : [0,\infty] \times [0,T] \to \mathbb{R}^+$ is therefore given by

$$\frac{\partial u(a,t)}{\partial t} + \frac{\partial u(a,t)}{\partial a} = -(\mu + b(a,c(t)))u(a,t). \tag{4.1}$$

We suppose further that when a cell divides it produces two daughter cells of age $a = 0$. By considering all possible division events, we deduce that

$$u(0,t) = 2\int_0^\infty b(a,c(t))u(a,t)\mathrm{d}a. \tag{4.2}$$

We consider a coarse-grained description of the cell cycle by lumping S, G2 and M into one phase, so we consider a two phase model G1 and S-G2-M (see Section 1.1.1 for a description of the cell cycle). We focus on the cells' G1/S transition and neglect the G0 phase. We assume cells proliferate at a constant rate, $\tau_p^{-1}$, provided they successfully enter the S-G2-M phase. Entry to the S-G2-M phase is regulated by the G1/S checkpoint which has been shown to depend on the presence of resources needed for cell growth [75]. Therefore, we assume that the proliferation rate is given by

$$b(a,c(t)) := \tau_p^{-1} H(a - a_{G1/S}(c(t))) \tag{4.3}$$

where $H$ is the Heaviside function and $a_{G1/S}(c(t))$ denotes *the G1/S transition age*, the age at which a cell passes from the G1 to the S phase. In Figure 4.3 we plot the proliferation rate $b(a,c(t))$ at time $t \in [0,T]$ .

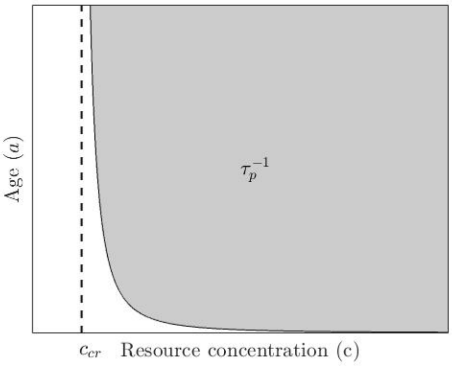


**Figure 4.3:** Contour plot of the proliferation rate $b(a, c(t))$ (Eq. (4.3)) at time $t \in [0, T]$. In the blank region, the proliferation rate is 0 while in the shaded region the proliferation rate is $\tau_p^{-1}$.

The G1/S transition age, $a_{G1/S}(c)$ is given as a function of $c$ to capture the dependence of the checkpoint on the available resource concentration. The authors in [53] describe this dependence by a simple scaling which was derived from analysing an oxygen-dependent cell cycle progression model,

$$a_{G1/S}(c) = \begin{cases} a_{-}\left(\frac{c}{c_{cr}} - 1\right)^{-\beta} & \text{for } c > c_{cr}, \\ \infty & \text{for } c \leq c_{cr}, \end{cases}$$

where $a_{-}$ and $\beta$ are positive constants. The positive quantity $c_{cr}$ is the critical resource concentration that allows cell proliferation. For high resource values, the transition age becomes smaller, so the cell requires less time to proliferate. On the other hand, for resource values slightly higher than $c_{cr}$, the transition age becomes bigger and cells take longer to transition to the G1/S phase and to proliferate. For $c \leq c_{cr}$, cells are assumed to enter the hypoxic-induced quiescent state.

If we assume that the resource, $c(t)$, is supplied at a constant rate $\overline{S}$ and consumed by all cells at a constant rate $k$, then

$$\frac{dc}{dt} = \overline{S} - kNc, \tag{4.4}$$

where $N(t) = \int_0^\infty u(a,t)\mathrm{d}a$ denotes the total number of cells at time $t$.

The full model consists of Eqs. (4.1) - (4.4) and the initial conditions

$$u(a, 0) = v_0(a) \quad \text{for } a > 0 \quad \text{and} \quad c(0) = c_0 \quad \text{where} \quad c_0 > 0. \tag{4.5}$$

To fully understand how the age-structured population dynamics is affected by the G1/S transition age we begin by considering an age-structured model with a fixed G1/S transition age.

## 4.3 Simplified model

Let us assume a fixed G1/S transition age, $a_{G1/S}(c) := a_{G1/S}$, and focus on the dynamics of the age-structured model. In this case, the model given by Eqs. (4.1) - (4.4) reduces to:

$$\left\{\begin{array}{l} \frac{\partial u(a,t)}{\partial t}+\frac{\partial u(a,t)}{\partial a}=-(\mu+b(a))u(a,t), \\ u(0,t)=2\int_0^\infty b(a)u(a,t)\mathrm{d}a, \\ u(a,0)=v_0(a), \end{array}\right. \tag{4.6}$$

where

$$b(a)=\left\{\begin{array}{ll} \tau_p^{-1} & \text{if } a\geq a_{G1/S}, \\ 0 & \text{if } a<a_{G1/S}. \end{array}\right.$$

in which $\tau_p, \mu$ and $a_{G1/S}$ are positive constants.

We analyse the long-time dynamics of this model using the general theory introduced in Sec. 2.1.

### 4.3.1 Transient and long-time dynamics

*Mature and immature subpopulations dynamics*

The G1/S transition age, $a_{G1/S}$, divides the cell population naturally into two subpopulations: the cell subpopulation whose age is above the G1/S threshold and can proliferate, and the cell subpopulation whose age is below the G1/S threshold. As mentioned above, we refer to the cells whose age is above the G1/S threshold as mature cells, and as immature cells otherwise. Let us denote by $X(t)$ and $Y(t)$, respectively, the subpopulations of mature and immature cells, so that

$$X(t)=\int_0^{a_{G1/S}} u(a,t)\mathrm{d}a \quad \text{and} \quad Y(t)=\int_{a_{G1/S}}^\infty u(a,t)\mathrm{d}a. \tag{4.7}$$

The full population is given by

$$N(t)=X(t)+Y(t). \tag{4.8}$$

**Theorem 5.** *Let X, Y be defined as* (4.7). *Then X and Y satisfy the following coupled system of ODEs:*

$$\left\{\begin{array}{ll} \frac{dX}{dt} & =-\mu X+2\tau_p^{-1}Y-u(a_{G1/S},t), \\ \frac{dY}{dt} & =-[\mu+\tau_p^{-1}]Y+u(a_{G1/S},t), \end{array}\right. \tag{4.9}$$

*where*

$$u(a_{G1/S},t)=\left\{\begin{array}{ll} v_0(a_{G1/S}-t)\exp(-\mu t) & \text{for } t\leq a_{G1/S}, \\ 2\tau_p^{-1}Y(t-a_{G1/S})\exp(-\mu a_{G1/S}) & \text{for } a_{G1/S}<t. \end{array}\right.$$

*Proof.* We integrate Eq. (4.6) over $[0,a_{G1/S}]$ and $[a_{G1/S},\infty]$ to obtain the evolution equations for $X$ and $Y$, respectively. The expression of $u(a_{G1/S},t)$ can be determined using the method of characteristics presented in Sec. 2.1.1. According to Theorem 1, considering $\beta(a,t)=\mu+b(a)$ and $\gamma(a)=2b(a)$, the solution of the simplified model (4.6), $u(a,t)$, is given by:

$$u(a,t)=\left\{\begin{array}{ll} v_0(a-t)\exp\left(-\int_0^t(\mu+b(a-t+s))\,\mathrm{d}s\right) & \text{for } a\geq t, \\ u(0,t-a)\exp\left(-\int_0^a(\mu+b(s))\,\mathrm{d}s\right) & \text{for } a<t. \end{array}\right. \tag{4.10}$$

Considering $a=a_{G1/S}$ and $t\leq a_{G1/S}$ in Eq. (4.10) we obtain:

$$\begin{aligned} u(a_{G1/S},t) &= v_0(a_{G1/S}-t)\exp\left(-\int_0^t (\mu + b(a_{G1/S}-t+s))\,\mathrm{d}s\right) \\ &= v_0(a_{G1/S}-t)\exp(-\mu t)\exp\left(-\int_{a_{G1/S}-t}^{a_{G1/S}} b(s)\mathrm{d}s\right) \\ &= v_0(a_{G1/S}-t)\exp(-\mu t). \end{aligned} \tag{4.11}$$

For $a_{G1/S} < t$

$$\begin{aligned} u(a_{G1/S},t) &= u(0,t-a_{G1/S})\exp\left(-\int_0^{a_{G1/S}} (\mu + b(s))\mathrm{d}s\right) \\ &= 2\tau_p^{-1} Y(t-a_{G1/S})\exp(-\mu a_{G1/S}). \end{aligned} \tag{4.12}$$

The expression for $u(a_{G1/S},t)$ in this case is derived using the definition of $Y(t)$ (Eq. (4.7)) and the boundary condition (Eq. (4.6)). ☐

Analysing these two possible values for $u(a_{G1/S},t)$ we can obtain a deeper insight into the dynamics of $X(t)$ and $Y(t)$. For $t \le a_{G1/S}$, the system (4.9) reduces to:

$$\begin{cases} \frac{dX}{dt} &= -\mu X + 2\tau_p^{-1} Y - v_0(a_{G1/S}-t)\exp(-\mu t), \\ \frac{dY}{dt} &= -[\mu + \tau_p^{-1}]Y + v_0(a_{G1/S}-t)\exp(-\mu t), \end{cases} \tag{4.13}$$

which is a non-homogeneous linear system of ODEs and an explicit solution can be derived using standard techniques [175].

For $t > a_{G1/S}$, the system (4.9) reduces to:

$$\begin{cases} \frac{dX}{dt} &= -\mu X + 2\tau_p^{-1} Y - 2\tau_p^{-1}\exp(-\mu a_{G1/S})Y(t-a_{G1/S}), \\ \frac{dY}{dt} &= -[\mu + \tau_p^{-1}]Y + 2\tau_p^{-1}\exp(-\mu a_{G1/S})Y(t-a_{G1/S}), \end{cases} \tag{4.14}$$

in which the equation of $Y(t)$ is independent of $X(t)$ and it is a linear delay differential equation with a constant delay $a_{G1/S}$ . By solving the DDE for $Y(t)$, the equation of $X(t)$ simplifies to a linear ODE and we can solve it with standard techniques [175].

Taking into account this observation, we now describe how to obtain the explicit solutions of $X(t)$ and $Y(t)$ for $t \le a_{G1/S}$ and $t > a_{G1/S}$.

**Corollary 5.1.** *For $t \le a_{G1/S}$, the explicit formulae for the mature and immature subpopulations $X(t)$ and $Y(t)$ of the simplified age-structured model (Eq. (4.6)) are given by:*

$$\begin{aligned} X(t) =& X(0)e^{-\mu t} + 2Y(0)e^{-\mu t} - 2Y(0)e^{-[\mu+\tau_p^{-1}]t}, \\ &+ e^{-\mu t}\int_0^t m(s)e^{\mu s}\mathrm{d}s - 2e^{-[\mu+\tau_p^{-1}]t}\int_0^t m(s)e^{[\mu+\tau_p^{-1}]s}\mathrm{d}s, \end{aligned}$$

*and*

$$Y(t) = Y(0)e^{-[\mu+\tau_p^{-1}]t} + e^{-[\mu+\tau_p^{-1}]t}\int_0^t m(s)e^{[\mu+\tau_p^{-1}]s}\mathrm{d}s \tag{4.15}$$

*where $m(s) = v_0(a_{G1/S}-s)\exp(-\mu s)$.*

*Proof.* The system (4.13) is equivalent to

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}(t) \tag{4.16}$$

where

$$\boldsymbol{A} = \begin{pmatrix} -\mu & \tau_p^{-1} \\ 0 & -[\mu + \tau_p^{-1}] \end{pmatrix} \qquad \boldsymbol{b}(t) = \begin{pmatrix} -m(t) \\ m(t) \end{pmatrix}$$

and $m(s) := v_0(a_{G1/S} - s)\exp(-\mu s)$.

Since (4.16) is a nonhomogeneous linear system, we first find a fundamental solution of

$$\frac{d\boldsymbol{\phi}}{dt} = \boldsymbol{A}\boldsymbol{\phi}, \quad \boldsymbol{\phi}(0) = \boldsymbol{I}. \tag{4.17}$$

where $\boldsymbol{\phi}$ is a nonsingular $2 \times 2$-matrix and $\boldsymbol{I}$ is the identity $2 \times 2$-matrix.

Using standard techniques of ODEs [175] the solution of (4.16) is then given by:

$$\boldsymbol{x}(t) = \boldsymbol{\phi}(t)\boldsymbol{\phi}^{-1}(0)\boldsymbol{x}_0 + \boldsymbol{\phi}(t)\int_0^t \boldsymbol{\phi}^{-1}(s)\boldsymbol{b}(s)\mathrm{d}s. \tag{4.18}$$

where the fundamental solution , $\boldsymbol{\phi}(t)$, is given by:

$$\boldsymbol{\phi}(t) = \begin{pmatrix} e^{-\mu t} & 2e^{-\mu t} - 2e^{-[\mu+\tau_p^{-1}]t} \\ 0 & e^{-[\mu+\tau_p^{-1}]t} \end{pmatrix}. \tag{4.19}$$

Combining (4.18) with the fundamental solution (4.19), we derive the desired expressions of $X(t)$ and $Y(t)$. □

To derive an explicit solution for $X(t)$ and $Y(t)$ when $t > a_{G1/S}$, we can solve the equation for $Y(t)$ using the technique for solving delay differential equations with constant delay known as the "method of steps" [197]. This method consists of solving the DDE sequentially in the time intervals $[na_{G1/S}, (n+1)a_{G1/S}]$ for $n \geq 1$. By doing this, the DDE reduces to a nonautonomous differential equation in each time interval. We consider the case $n = 1$ to illustrate the method. In the time interval $[a_{G1/S} < t < 2a_{G1/S}]$, the problem consists of solving the linear delay differential equation:

$$\frac{dY}{dt} = -(\mu + \tau_p^{-1})Y + 2\tau_p^{-1}\exp(-\mu a_{G1/S})Y(t - a_{G1/S}) \tag{4.20}$$

with

$$Y(t - a_{G1/S}) = \hat{Y}(t) \quad \text{for } t \leq a_{G1/S}, \tag{4.21}$$

where $\hat{Y}(t)$ is the value of $Y(t)$ given by Eq. (4.15). Eq. (4.20) is then reduced to a nonautonomous differential equation which can be solved by standard techniques.

### *Long-time dynamics*

As we discussed in Section 2.1.2, it is enough to determine the Euler-Lotka characteristic equation and the stable-age distribution of an age-structured model to determine its long time dynamics. The following corollary summarises the long-time dynamics of Eq. (4.6).

**Corollary 5.2** (Long-time dynamics of the simplified model)**.** *For the simplified age-structured model given by Eq.* (4.6)*, the Euler-Lotka characteristic equation is:*

$$1 = \frac{2\tau_p^{-1}}{\tau_p^{-1} + \mu} \exp[-(\mu + \lambda) a_{G1/S}], \tag{4.22}$$

*and the stable age-distribution has the following explicit form:*

$$u(a,t) = \begin{cases} \kappa \exp\left[\lambda(t-a) - \mu a\right] & a < a_{G1/S}, \\ \kappa \exp\left[\lambda(t-a) - \mu a - \tau_p^{-1}(a - a_{G1/S})\right] & a \geq a_{G1/S}. \end{cases} \tag{4.23}$$

*where* $\kappa$ *is a positive constant.*

*Proof.* First we notice that in the simplified model $\gamma(a,t) = 2b(a)$ and $\beta(a,t) = \mu + b(a)$. We now consider Theorem 3, which describes the long-time dynamics of the age-structured model. The Euler-Lotka characteristic equation is given by using the values of $\gamma(a)$ and $\beta(a)$ in Eq. (2.22). The stable age-distribution is derived similarly by using the functional forms of $\gamma$ and $\beta$ in Eq. (2.23). ☐

For $\lambda = 0$ to satisfy the Euler-Lotka characteristic equation (Eq. (4.22)), the G1/S transition age needs to be given by

$$a^*_{G1/S} = -\frac{1}{\mu} \log\left[\frac{\tau_p(\tau_p^{-1} + \mu)}{2}\right], \tag{4.24}$$

This equation is obtained by solving Eq. (4.22) for $a_{G1/S}$ while assuming that $\lambda = 0$. This critical value of $a_{G1/S}$ will be referred to as the *critical G1/S transition age*.

**Remark 2.** *For* $a_{G1/S}$ *to be positive, the argument of the logarithm needs to be less that 1 and this is possible for*

$$\mu < \tau_p^{-1}. \tag{4.25}$$

**Remark 3.** *If* $a_{G1/S} < a^*_{G1/S}$ *then* $\lambda > 0$ *and therefore* $u(a,t) \to \infty$ *as* $t \to \infty$*. Similarly, if* $a_{G1/S} > a^*_{G1/S}$ *then* $\lambda < 0$ *and therefore* $u(a,t) \to 0$ *as* $t \to \infty$*.*

From the previous corollary and the last remarks, we observe that the G1/S transition age determines the long time dynamics.

### 4.3.2 Analysis of oscillatory behaviour

We now investigate the presence of oscillatory dynamics for the long-time dynamics of $X$ and $Y$ given by Eq. (4.9). We consider the case in which $\lambda = 0$ satisfies the Euler-Lotka characteristic equation, i.e. when $\frac{dN}{dt} \to 0$ as $t \to \infty$ where $N(t) = X(t) + Y(t)$.

Since we are interested in the behaviour of the system (4.9) when $t \gg 1$, we focus our analysis on the DDE given by Eq. (4.20). From the long-time analysis we obtain that when $\lambda = 0$, the G1/S transition age, $a_{G1/S}$, is given by (4.24) and substituting this in Eq. (4.20) results in:

$$\frac{dY}{dt} = -[\mu + \tau_p^{-1}]Y + [\mu + \tau_p^{-1}]Y(t - a_{G1/S}). \tag{4.26}$$

IF we use the notation $M = \mu + \tau_p^{-1}$ and $r = a_{G1/S}$ then Eq. (4.26) can be written compactly as

$$\frac{dY}{dt} = -MY + MY(t - r) \tag{4.27}$$

where $M$ and $r$ are positive constants.

equation 4.27 is a linear delay differential equation and the standard theorem of stability for this type of equation is Theorem 4 in Section 2.2. However, we notice that we can not apply this Theorem since $\lambda = 0$ is a root of the characteristic equation. However we are able to arrive at the following results for Eq. (4.27).

**Lemma 6.** *The characteristic roots of the DDE* (4.27) *are* $\lambda = 0$ *or* $\lambda$ *non-zero with* $Re(\lambda) < 0$.

*Proof.* Let us consider the characteristic equation for the DDE (4.27):

$$h(\lambda, r) = \lambda + M - M\exp(-\lambda r) \tag{4.28}$$

A root of the DDE (4.27), $\lambda$, is such that $h(\lambda, r) = 0$. We notice that $\lambda = 0$ is a root of this function. For a non-zero root of the characteristic equation,

$$\lambda = -M + M\exp(-\lambda r). \tag{4.29}$$

Multiplying by $r$ and denoting $z = \lambda r \in \mathbb{C}$, we have

$$z = -Mr + Mr\exp(-z). \tag{4.30}$$

If we write $z = x + iy$, Eq. (4.30) simplifies to an equation for $x$ and $y$:

$$x = -Mr + Mre^{-x}cos(y), \tag{4.31}$$

$$y = -Mre^{-x}sin(y). \tag{4.32}$$

We focus on the real part of $z$:

$$x = -Mr + Mre^{-x}cos(y) = Mr\left(e^{-x}cos(y) - 1\right). \tag{4.33}$$

We prove that $x < 0$ by contradiction. Let us assume $x > 0$, then

$$0 < \exp(-x) < 1. \tag{4.34}$$

Since $cos(y) \leq 1$ for an arbitrary value of $y$ by multiplying Eq. (4.34) by $exp(-x)$ we have

$$\exp(-x)\cos(y) \leq \exp(-x) < 1. \tag{4.35}$$

From Eq. (4.35) we can conclude that

$$\exp(-x)\cos(y) - 1 < 0. \tag{4.36}$$

Considering Eq. (4.36) in Eq. (4.33) and that $Mr > 0$, we conclude that $x < 0$. But this is a contraction, since we assume $x > 0$. Therefore we have $x < 0$. ☐

From the previous theorem we can conclude that the solution can be asymptotically stable or have undamped oscillations that do not increase with time.

## 4.4 Full model behaviour analysis

We now analyse the behaviour of the full model (Eqs.(4.1) - (4.4)) taking into consideration what we have learned from the simplified model.

### 4.4.1 Long-time dynamics

The long-time behaviour of the full model can be described by Theorem 7 [53].

**Theorem 7** (Long-time dynamics of the full model)**.** *Assuming the average waiting time to division after the G1/S transition, $\tau_p$, is smaller than the average cell life span, $\mu^{-1}$, i.e. $\tau_p < \mu^{-1}$ and that the resource concentration converges to a a positive steady-state value $c_\infty > 0$ then as $t \to \infty$ the cell population and the resource concentration of the full model (Eqs.*(4.1) - (4.4)*) reach steady state values given by:*

$$N_\infty = \frac{\overline{S}}{kc_\infty}, \qquad c_\infty = a^{-1}_{G1/S}\left(a^*\right). \tag{4.37}$$

*where*

$$a^* = -\frac{1}{\mu}\log\left(\frac{\tau_p(\mu+\tau_p^{-1})}{2}\right). \tag{4.38}$$

*Proof.* Assuming $\frac{dc}{dt} = 0$ in Eq. (4.4) and that the steady value $c_\infty \in \mathbb{R} \neq 0$ we obtain (4.37). To analyse the long-time dynamics of the age-structured model, we consider the Euler-Lotka characteristic equation of the full model:

$$1 = 2\int_0^\infty b(a, c_\infty)\exp\left(-\lambda a - \int_0^a \left[\mu + b(y, c_\infty)\right] \mathrm{d}y\right) \mathrm{d}a \tag{4.39}$$

which we derive by following Theorem (3). Since $N_\infty$ is a constant real value then $\lambda = 0$ satisfies the Euler-Lotka equation and the critical G1/S transition age that is obtained by setting $\lambda = 0$ in Eq. (4.39). The value of $c_\infty$ is determined by using the inverse function of the G1/S transition age function. □

We notice that as $t \to \infty$ the dynamics of the full model reduces to the simplified model with a fixed G1/S transition age given by Eq. (4.38). Therefore the results we obtain for the simplified model extend to the full model in the long-time limit.

### 4.4.2 Mature and immature subpopulation dynamics

In Sec. 4.3.1 we defined the mature and immature subpopulations, and investigated their dynamics in the simplified model. We now analyse the full system with respect to these subpopulation dynamics.

Let us denote by $X(t)$ and $Y(t)$, respectively, the subpopulations of mature and immature cells so that

$$X(t) = \int_0^{a_{G1/S}(t)} n(a,t)\mathrm{d}a \quad \text{and} \quad Y(t) = \int_{a_{G1/S}(t)}^\infty n(a,t)\mathrm{d}a, \tag{4.40}$$

where $a_{G1/S}(t) := a_{G1/S}(c(t))$.

With these definitions, we obtain the following result:

**Theorem 8.** *Let X and Y be defined as in* (4.40) *and let $n(a,t)$ and $c(t)$ satisfy the model Eqs.* (4.1) - (4.5)*. Then the dynamics of X and Y are given by the following system:*

$$\begin{aligned} \frac{dX}{dt} &= 2\tau_p^{-1}Y - \mu X + \left[\frac{da_{G1/S}}{dt}(t) - 1\right] n(a_{G1/S}(t),t), \\ \frac{dY}{dt} &= -(\mu+\tau_p^{-1})Y + \left[1 - \frac{da_{G1/S}}{dt}(t)\right] n(a_{G1/S}(t),t), \end{aligned} \tag{4.41}$$

*in which $a_{G1/S}(t) := a_- \left(\frac{c(t)}{c_{cr}} - 1\right)^{-\beta}$ and $c(t)$ is formally given by*

$$c(t) = \left[c_0 + \overline{S}\int_0^t \exp\left(k\int_0^{t'} N(s)\mathrm{d}s\right)\mathrm{d}t'\right]\exp\left(-k\int_0^t N(s)\mathrm{d}s\right).$$

*Furthermore, if $c(t) \geq c_{cr}$ $\forall t \in [0,T]$, then the number of cells with G1/S transition age $a_{G1/S}(t)$ at time t, $n(a_{G1/S}(t),t)$, is formally given by*

$$n(a_{G1/S}(t),t) = v_0(a_{G1/S}(t)-t)\exp\left(-\int_0^t [\mu + b(s-t+a_{G1/S}(t),c(s))]\mathrm{d}s\right) \tag{4.42}$$

*for $t \leq a_{G1/S}(t)$ and*

$$n(a_{G1/S}(t),t) = 2\tau_p^{-1}Y(t-a_{G1/S}(t))\exp\left(-\int_0^{a_{G1/S}(t)} [\mu + b(s,c(s+t-a_{G1/S}(t)))]\mathrm{d}s\right) \tag{4.43}$$

*for $t > a_{G1/S}(t)$.*

*Proof.* Integrating Eq. (4.1) over the domains $[0, a_{G1/S}(t)]$ and $[a_{G1/S}(t), \infty]$, we obtain the dynamics of $X$ and $Y$, respectively. The evolution of $c(t)$ is given by solving Eq. (4.4) while assuming $N(t)$ is a known function. Finally, the formula for the cell concentration with G1/S transition age $a_{G1/S}(t)$ at time $t$, $n(a_{G1/S}(t),t)$, is obtained by solving the age-structured model by the method of characteristics. The characteristic curves of Eq. (4.1) are the lines $a = t + \mu$ where $\mu$ is a constant. By solving Eq. (4.1) along the characteristic curves and taking into account the boundary condition (4.2) and initial condition (4.5), we obtain that $n(a,t)$ is given by:

$$n(a,t) = \begin{cases} v_0(a-t)\exp\left(-\int_0^t [\mu + b(s-t+a,c(s))]\mathrm{d}s\right) & \text{for } t \leq a, \\ n(0,t-a)\exp\left(-\int_0^a [\mu + b(s,c(s+t-a))]\mathrm{d}s\right) & \text{for } a < t. \end{cases} \tag{4.44}$$

By considering $a = a_{G1/S}(t)$ and $n(0, t - a_{G1/S}(t))$ in terms of the mature subpopulation $Y$ in Eq. (4.44), we derive the expressions (4.42) and (4.43) for $n(a_{G1/S}(t),t)$. □

Theorem 8 reduces the analysis of the full model to the analysis of the system (4.41). Given Eqs. (4.42) and (4.43), we notice that for $t \leq a_{G1/S}(t)$, (4.41) consists of a non-homogeneous coupled linear system of ordinary differential equations that depends on the initial age-distribution and for $t > a_{G1/S}(t)$, the dynamics of $Y$ is described by a state-dependent delay differential equation,

$$\frac{dY}{dt} = -(\mu+\tau_p^{-1})Y + 2\tau_p^{-1}\overline{a}(t)Y(t-a_{G1/S}(t)) \tag{4.45}$$

in which $\overline{a}(t) = \left[1 - a'_{G1/S}(t)\right]\exp\left(-\int_0^{a_{G1/S}(t)} [\mu + b(s,c(s+t-a))]\mathrm{d}s\right)$.

Unlike the simplified model, we are not able to determine explicit solutions for $X(t)$ and $Y(t)$. However we are able to describe the overall proliferation and the per capita growth rate in terms of the dynamics of these two subpopulations. Let us denote by $\sigma(N(t)) := \frac{1}{N(t)}\frac{dN(t)}{dt}$, the per capita growth rate. The following theorem gives the evolution of the per capita growth rate and the total population evolution in terms of the mature and immature cell subpopulations.

**Theorem 9.** *The evolution of $N(t)$ and $\sigma(N(t))$ for $t \in [0,T]$ is given by:*

$$\frac{dN}{dt} = [\tau_p^{-1} - \mu]Y - \mu X, \tag{4.46}$$

$$\frac{d}{dt}\left(\sigma(N(t))\right) = \tau_p^{-1}\frac{d}{dt}\left[\frac{Y}{N}\right]. \tag{4.47}$$

*Proof.* By integrating Eq. (4.1) in the age domain $[0,\infty]$ we obtain Eq. (4.46). To derive Eq. (4.47), we consider the definition of the per capita growth rate and Eq. (4.46) as follows,

$$\begin{aligned}\frac{d}{dt}\left(\sigma(N(t))\right) &= \frac{d}{dt}\left(\frac{1}{N}\frac{dN}{dt}\right) = \frac{d}{dt}\left[\frac{(\tau_p^{-1} - \mu)Y - \mu X}{N}\right] \\ &= \frac{d}{dt}\left[\tau_p^{-1}\frac{Y}{N} - \mu\right] = \tau_p^{-1}\frac{d}{dt}\left[\frac{Y}{N}\right]\end{aligned}$$

from which we obtain Eq. (4.47). ☐

### 4.4.3 Conditions for delayed logistic proliferation

In this section we identify conditions under which the per capita growth rate of the age-structured model exhibits biphasic dynamics and logistic-type growth. First, we interpret the experimental behaviour in mathematical terms:

I) The cell population undergoes logistic-type behaviour, i.e,

- The population increases in size monotonically,

  $$\frac{dN}{dt} > 0 \qquad \forall t \in [0,T].$$

  where $T$ is the final observation time.
- The population saturates

  $$N(t) \to N_\infty. \quad \text{as} \quad t \to T,$$

  where $N_\infty > 0$.

II) The per capita growth rate exhibits biphasic behaviour, i.e. there exists $t_1 > 0$ such that the per capita growth rate, $\sigma(N(t))$, has the following behaviour:

$$\frac{d\sigma}{dN}\begin{cases} > 0 & \text{for } t \in [0,t_1], \\ \leq 0 & \text{for } t \in [t_1,T]. \end{cases}$$

**Theorem 10.** *The necessary and sufficient conditions for the age-structured model defined by Eqs.* (4.1) - (4.5)*, to exhibit the behaviour I-II are as follows*

$$\mu < \tau_p^{-1}, \tag{4.48}$$

$$(\tau_p^{-1} - \mu)Y(t) > \mu X(t) \quad \forall t \in [0, T], \tag{4.49}$$

*and there exists* $t_1 > 0$ *such that*

$$\frac{d}{dt}\left[\frac{Y}{N}\right] > 0 \textit{ for } t \in [0, t_1] \quad \textit{and} \quad \frac{d}{dt}\left[\frac{Y}{N}\right] < 0 \textit{ for } t \in [t_1, T]. \tag{4.50}$$

*Proof.* The condition (4.48) needs to be satisfied so that the age-structured model has a stable steady-state solution as described in [53]. Given Eq. (4.46), it follows that for $\frac{dN}{dt} > 0$, Eq. (4.49) must hold. To derive (4.50), we notice that from the chain rule that $\frac{d\sigma(N(t))}{dt} = \frac{d\sigma(N)}{dN}\frac{dN}{dt}$ and given that $\frac{dN}{dt} > 0$, the sign of $\frac{d}{dt}\left(\sigma(N(t))\right)$ is the same as that of $\frac{d\sigma(N)}{dN}$. □

Theorem 10 gives necessary and sufficient conditions for the model to exhibit the behaviour described in I-II. We note that the conditions for the biphasic behaviour, given by Eq. (4.50), can be interpreted biologically as an initial increase in the fraction of mature cells followed by a stabilisation phase in which the mature cell fraction decreases to its steady-state value. Since the fraction of mature cells is regulated by the transition age, $a_{G1/S}$, and this critical age is regulated by the resource concentration, we deduce that changes in the resource dynamics may influence the biphasic behaviour.

## 4.5 Numerical study

In this section we introduce the discretisation scheme that we use to simulate the simplified and full models. We check the accuracy of the discretisation scheme and calculate the observed convergence rate using the method of manufactured solutions that is introduced in Section 2.3.

Multiple numerical methods have been developed for discretising age-structured population models [1, 72, 129, 130]. The focus has been on finite difference methods [1] and few authors have consider finite elements [33]. Here we present a discretisation scheme based on finite elements and analyse its rate of convergence.

### 4.5.1 Discretisation scheme

We focus first on the discretisation of the age-structured population evolution equation. We consider a generalised model that can describe either the cell evolution of the simplified model or the cell evolution of the full model by specifying the G1/S transition age function. Therefore by presenting a discretisation scheme for this model we will be presenting a discretisation scheme for the evolution equation of the age-structured population of the simplified and full model. Let us consider the cell density $u(a,t) : [0,\infty] \times [0,T] \to \mathbb{R}$ and suppose that it satisfies the following system:

$$\frac{\partial u(a,t)}{\partial t} + \frac{\partial u(a,t)}{\partial a} = -(\mu + b(a,t))u(a,t), \tag{4.51a}$$

with boundary conditions

$$u(0,t) = 2\int_0^\infty b(a,t)u(a,t)\mathrm{d}a, \tag{4.51b}$$

and

$$u(a,0) = v_0(a), \tag{4.51c}$$

where

$$b(a,t) = \begin{cases} \tau_p^{-1} & \text{if } a \geq a_{G1/S}(t), \\ 0 & \text{if } a < a_{G1/S}(t), \end{cases}$$

for a function $a_{G1/S}(t)$. We describe the particular form that this function has later in the chapter.

If $a_{G1/S}(t)$ is a constant function, i.e. $a_{G1/S}(t) := \hat{a}$, then Eq. (4.51) reduces to the simplified model. In the case when $a_{G1/S}(t) := a_{G1/S}(c(t))$ then Eq. (4.51) reduces to the evolution of the cell population of the full model. Therefore presenting a discretisation scheme for this model will define a discretisation scheme for the PDE of the simplified and full models.

equation (4.51) is a hyperbolic partial differential equation, with a nonlocal boundary condition. We discretise it following the Rothe method: first in time and then in age [94]. For the time discretisation, we consider the backward Euler method and, for the age domain discretisation, we consider finite elements. Before proceeding, we introduce some notation and assumptions in order to present the scheme.

To discretise the age domain $[0,\infty]$, we notice that the cell age is naturally bounded given that biological cells do not have infinite age, therefore we assume that the initial condition is bounded. Under certain assumptions on the parameters that $u(a,t)$ is bounded for all $t \in [0,T]$ [141]. Hence there exists $a_{max} > 0$ such that $u(a,t) = 0$ for $a \in (a_{max},\infty)$ so we restrict our attention to $[0,a_{max}]$, which we denote by $\Omega$.

We consider the following discretisation of the time and the age domain. Let us partition the interval $[0,T]$ into $M$ subintervals of length $h_t = \frac{T}{M}$ and denote by $t_n = nh_t$ for $n = 0,1,\ldots,M$. We consider a mesh $\mathcal{T}_h$ of $\Omega$, namely, $\Omega = \bigcup_{j=1}^{N} K_j$ and refer to $K_j$ as an element of $\mathcal{T}_h$. We assume the mesh is regular (see Sec.2.3 for more details on regular meshes). We define $h_a$ as the maximum diameter of the elements of the mesh:

$$h_a := \max_{K\in\mathcal{T}_h} h_K. \tag{4.52}$$

We consider a non-uniform mesh for the age domain. We consider a specific point in the domain $\hat{a}$ so the elements, $K_i = [a_0^i, a_1^i]$, are divided into two groups: the elements located in the subdomain $[0,\hat{a}]$ whose length is given by $a_1^i - a_0^i = \hat{a}/2^m$ for an integer $m$; and the elements located in the subdomain $[\hat{a}, a_{max}]$ whose length is given by $a_1^i - a_0^i = (a_{max} - \hat{a})/2^m$ for the same integer $m$. By considering a non-uniform mesh, we can have a more efficient implementation. In Figure 4.4 we show the age-time mesh we consider the the finite element discretisation of Eq. (4.51).

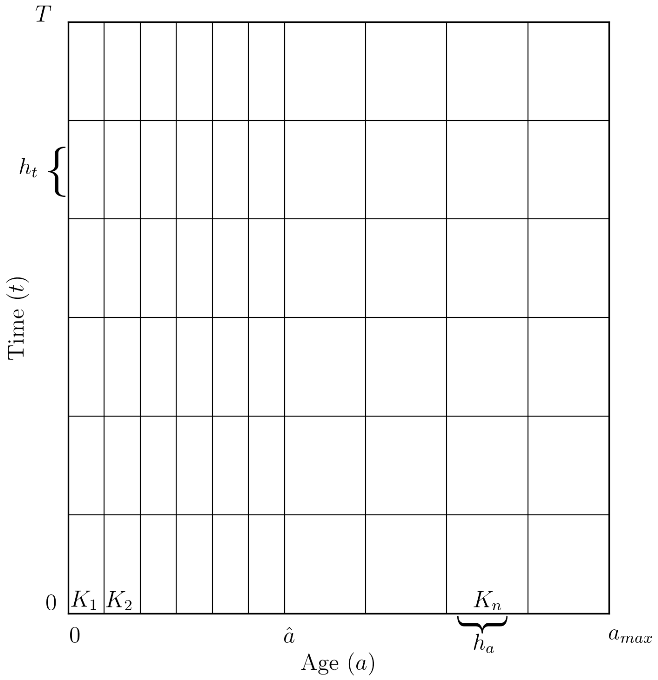


**Figure 4.4:** Schematic of the age and time mesh for the discretisation of Eq. (4.51). The time domain $[0, T]$ is partitioned uniformly with a step of size $h_t$. For discretising the age domain, $\Omega = [0, a_{max}]$, we consider a mesh $\mathcal{T}_h = \{K_i\}_{i=1}^{N_{\mathcal{T}_h}}$ where $K_i$ are intervals $K_i = [a_0^i, a_1^i]$. We can observe that the elements located in the left of the value $\hat{a}$ are smaller than the elements located in the right of $\hat{a}$.

We denote the solution of Eq. (4.51) in the age domain at $t_n$ as $u^n(a) := u(a, t_n)$. Similarly we denote the proliferation function in the age domain at $t_n$ as $b^n(a) := b(a, t_n)$.

We assume that the solution of Eq. (4.51) is defined in the following functional space:

$$V_0 \overset{\text{def}}{=} \left\{ v \in H^1(\Omega) : \; v|_{a=0} = 2 \int_0^{a_{max}} b(a) v(a) \mathrm{d}a \right\} \tag{4.53}$$

The definition of the functional space $H^1(\Omega)$ is included in Sec. 2.3.

To approximate the solution in $V_0$ we consider the finite dimensional space $V_h$ which we introduced in Sec.2.3.

### Variational formulation

We now present our discretisation scheme. Discretising the time derivative in Eq. (4.51) through the backward Euler method, we obtain the following update rule for $u^n(a)$ with $n \geq 1$:

$$\frac{u^n(a) - u^{n-1}(a)}{h_t} = \left[ -\nabla u^n(a) - (\mu + b^n(a)) u^n(a) \right]. \tag{4.54}$$

where we use the gradient notation for the derivate in the age domain: $\nabla u^n(a) = \frac{\partial u^n(a)}{\partial a}$.

Discretising the boundary condition (4.51b) with the backward Euler method we obtain the condition that needs to be satisfied at $a = 0$ for each $t_n$ with $n \geq 1$:

$$u^n|_{a=0} = 2\int_0^{a_{max}} b^n(a)u^n(a)\mathrm{d}a. \tag{4.55}$$

Discretisation of hyperbolic equations can produce unwanted oscillations if not done appropriately [29]. We stablise the age discretisation by considering the streamline upwind/Petrov Galerkin formulation [136]. Accordingly we consider the test functions in $W_h$ which is defined by:

$$W_h \stackrel{\text{def}}{=} \{w_h(a) : w_h(a) = v_h(a) + \delta\nabla v_h(a);\ v_h(a) \in V_h\} \tag{4.56}$$

where $\delta > 0$ denotes a stabilisation parameter and $\nabla$ denotes the gradient in the age domain which coincides with the partial derivative on age. $\nabla := \frac{\partial}{\partial_a}$.

The semi-discrete variational problem at each $t_n$ with $n \geq 1$ involves determining $u^n \in V_h$ such that

$$\left(\frac{u^n(a) - u^{n-1}(a)}{h_t}, w_h(a)\right) + T(u^n(a), w_h(a)) = F(u^n(a), w_h(a)) \quad \forall w_h(a) \in W_h \tag{4.57}$$

where

$$T(u^n(a), w_h(a)) = (\nabla u^n(a), \nabla w_h(a)), \tag{4.58}$$

$$F(u^n, w_h) = -((\mu + b^n(a))u^n, w_h), \tag{4.59}$$

and the function $u^n(a)$ also satisfies the boundary condition (4.55).

By considering the nodal basis of $V_h$, $\{\varphi_j\}_{j=1}^{N_h}$, the weak problem (4.57) reduces to a linear system. We define first the linear system for all nodes not located at $a = 0$ and then describe how we discretise the nonlocal boundary condition at $a = 0$.

### Linear system of equations for the basis functions $n = 1, \ldots, N_h - 1$.

We first approximate the solution of Eq. (4.57) by its Ritz projection in the space $V_h$.

$$u^n(a) \approx u_h^n(a). \tag{4.60}$$

Let $\{\varphi_0(x,y) \ldots, \varphi_{N_h-1}(x,y)\}$ denote the nodal basis functions of $V_h$, then $u_h^n(a)$ is given by:

$$u_h^n(a) = \sum_{j=0}^{N_h} U_j^n \varphi_j(a). \tag{4.61}$$

where $U_j^n$ are real constants for $0 \leq j \leq N_h - 1$.

The variational formulation (4.57) reduces to:

$$\left(\frac{u^n(a)-u^{n-1}(a)}{h_t},\varphi_i+\delta\nabla\varphi_i(a)\right)+T(u^n(a),\varphi_i(a)+\delta\nabla\varphi_i(a))$$
$$=F(u^n(a),\varphi_i(a)+\delta\nabla\varphi_i(a)) \tag{4.62}$$

for all $i=0,\dots,N_h-1$.

The matrix formulation of the linear system (4.62) is obtained substituting $u^n$ with $u_h^n$ :

$$\boldsymbol{M}^n\boldsymbol{U}^n=\boldsymbol{Y}^n \tag{4.63}$$

where $\boldsymbol{M}^n$ is an $N_h\times N_h$ matrix whose entries are given by:

$$\begin{aligned}\boldsymbol{M}^n_{i,j}&=\left(\varphi_i(a)+\delta\nabla\varphi_i(a),\varphi_j(a)\right)+h_t\left(\varphi_i(a)+\delta\nabla\varphi_i(a),\nabla\varphi_j(a)\right)\\&+h_t\left(\varphi_i(a)+\delta\nabla\varphi_i(a),\left(\mu+b^n(a)\right)\varphi_j(a)\right).\end{aligned}$$

$\boldsymbol{Y}^n$ is a $N_h$-dimensional vector whose entries given by:

$$\boldsymbol{Y}^n_i=\left(\varphi_i(a)+\delta\nabla\varphi_i(a),u^{n-1}(a)\right)$$

and $\boldsymbol{U}^n$ is the $N_h$-dimensional solution vector.

### Discretisation of the boundary condition

The boundary condition at each $t_n$, $n\geq 1$, Eq. (4.55) constraints in the solution value at $a=0$. Since we are dealing with Lagrange finite elements, without lost of generality, we can assume that the basis nodal function that corresponds to the point $a=0$ is $\varphi_0$ so the value of the solution at time $t_n$ and age $a=0$ (the boundary value at 0) is given by $U_0^n$.

We discretise the boundary condition (4.55) by approximating the integral in the age domain with a Gaussian-Legendre quadrature rule [134] and considering the approximation of the solution at time $n$, $u^n(a)$, by its linear representation with respect to the nodal basis functions of $V_h$ given in Eq. (4.61). Eq. (4.55) is approximated as follows:

$$U_0^n\approx 2\sum_{j=0}^{N_h-1}\sum_{q=1}^{Q}U_j^n\varphi_j(x_q)b^n(x_q)w_q, \tag{4.64}$$

where $x_q$ and $w_q$ are the corresponding quadrature points and the quadrature weights of the Gaussian quadrature and $U_j^n$ are the corresponding linear coefficients of the linear representation in terms of the nodal basis function (See Eq. (4.61)).

Solving Eq. (4.64) for $u_0^n$ yields the constraint that must be satisfied:

$$U_0^n=\sum_{j=1}^{N_h-1}\frac{p_j}{d_0}U_j^n \tag{4.65}$$

where

$$d_0 = 1 - \left( \sum_{q=1}^{Q} 2\varphi_0(x_q) b^n(x_q) w_q \right),$$

and

$$p_j = \sum_{q=1}^{Q} 2\varphi_j(x_q) b^n(x_q) w_q.$$

### Discretisation of the resource concentration

The full age-structured model given by Eqs. (4.1) - (4.4) is discretised via a time-splitting scheme. Since we have already introduced the discretisation of the age-structured model, it remains to introduce the discretisation of the evolution equation for the resource concentration, Eq. (4.4).

We denote the solution of the resource concentration evolution Eq. (4.4) at time $t_n$ as $c^n := c(t_n)$. By discretising in time Eq. (4.4) via a backward Euler method, we obtain that for $n \geq 1$:

$$\frac{c^n - c^{n-1}}{h_t} = (\bar{S} - kN^n c^n), \tag{4.66}$$

where $N^n = \int_0^{a_{max}} u^n(a) \mathrm{d}a$.

Solving Eq.(4.66) for $c^n$ we obtain the update rule for $c^n$ with $n \geq 1$:

$$c^n = \left( \frac{1}{1 + kh_t N^n} \right) \left( c^{n-1} + \bar{S} h_t \right). \tag{4.67}$$

### Update algorithms

We present below the update algorithms 4.1 and 4.2 used to implement the discretisation of the autonomous age-structured models (Eqs. (4.6) and (4.51)) and the coupled model for the age-structured population and the resource concentration (Eqs. (4.1) - (4.4)), respectively.

When we consider an age-structured population independent of the resource concentration (as for the simplified model) the update algorithm is the following:

**Algorithm 4.1** Solution update for an autonomous age-structured model

```
Initialise the solution with the initial conditions $v^0(a)$ and set the initial time index $n = 0$ and final time index $n = M$.
for $n = 1$ to $n = M - 1$ do
    Update rule $U^{n-1} \to U^n$: Solve the augmented linear system that consists of the linear system (4.63) and the constrained system (4.65) to obtain $U^n$.
end for
```

For the coupled system, the update algorithm is the following:

**Algorithm 4.2** Solution update for the coupled age-structured model and the resource concentration

```
Initialise the solutions for the age-structured distribution $U^0$ and the resource concentration $c^0$ with the initial conditions $v^0(a)$ and $c_0$, respectively. Set the initial time index $n = 0$ and final time index $n = M$.
for $n = 1$ to $n = M - 1$ do
    Update rule $U^{n-1} \to U^n$: Solve the augmented linear system that consists of the linear system (4.63) and the constrained system (4.65) to obtain $U^n$..
    Update rule $c^{n-1} \to c^n$: Calculate $c^n$ using the formula (4.67).
end for
```

The finite element discretisation was implemented in C++ using the software DEAL.II [18]. We consider the stabilisation parameter to be $\delta = 0.1h_K$ for $K \in \mathcal{T}_h$ since good numerical results have been observed when using this parameter value [36, 136].

### 4.5.2 Convergence analysis

In this section we analyse the convergence properties of our discretisation schemes. We aim to show that the schemes are second order in age and first order in time. Since an explicit solution of the age-structured model is not generally known, we consider the method of manufactured solutions to analyze the convergence of the discretisation scheme. The method of manufactured solutions (MMS) involves considering an exact solution of the governing equation with modified forcing terms and studying the convergence rate through numerical experiments [22]. We explained the implementation steps of the MMS in Section 2.3.

**Autonomous age-structured model**

We consider the following modified system: let $u : [0, \infty] \times [0, T] \to \mathbb{R}$ satisfy

$$\begin{cases} \frac{\partial u(a,t)}{\partial t} = -\left[\frac{\partial u(a,t)}{\partial a} + (\mu + b(a))u(a,t)\right] + G(a,t), \\ u(0,t) = 2\int_0^\infty b(a)u(a,t)\mathrm{d}a - 2\int_0^\infty b(a)F(a,t)\mathrm{d}a + F(0,t), \\ u(a,0) = F(a,0), \end{cases} \tag{4.68}$$

where $G(a,t) = \frac{\partial F(a,t)}{\partial a} + (\mu + b(a))F(a,t) + \frac{\partial F(a,t)}{\partial t}$,

$$b(a) = \begin{cases} \tau_p^{-1} & \text{if } a \geq a_{G1/S}(t), \\ 0 & \text{if } a < a_{G1/S}(t), \end{cases}$$

in which

$$a_{G1/S}(t) = a_0 + a_1 t. \tag{4.69}$$

with $a_0, a_1 \in \mathbb{R}$.

This system has an explicit solution given by

$$F(a,t) = A(a,t)T(t) \tag{4.70}$$

where

$$A(a,t) = \begin{cases} \exp(-\mu a) & \text{if } a < a_{G1/S}(t) \\ \exp(-\mu a - \tau_p^{-1}(a - a_{G1/S}(t))) & \text{if } a \geq a_{G1/S}(t) \end{cases}$$

and

$$T(t) = 50 \tanh\left(\frac{t}{\alpha}\right) + 1.$$

where $\alpha$ is a real scalar.

By construction we have considered the explicit solution $F(a,t)$ to have the same shape as the stable age-distribution (see Eq. (4.23)). We consider $a_{G1/S}(t)$ to be a linear function so we can analyse: a constant G1/S transition age, as in the simplified model, and a moving G1/S transition age, as in the full model. By varying the values of $a_0$ and $a_1$ we can study the effects of the G1/S transition age position and velocity on the performance of the discretisation scheme.

We analyse the performance of the discretisation scheme by solving Eq. (4.68) using our discretisation scheme and calculating the error between the numerical solution and the exact solution, $F(a,t)$, at each time step. Let $h_a$, $h_t$ denote the age and time mesh size, respectively. For a given simulation we consider the following error measure in the age domain and time:

$$e(h_a, h_t) = \|u_h(a,t) - F(a,t)\|_{L^2([0,a_{max}]\times(0,T))}\,. \tag{4.71}$$

We compute this error by approximating the time integral by a trapezoidal rule and the integration in the age domain $[0, a_{max}]$ by a Gaussian-Legendre quadrature.

It has been shown that for time-dependent transport equations that are solved using backward Euler method and Lagrange finite elements under certain assumptions [36] the error measure satisfy the following inequality:

$$0 \leq e(h_a, h, t) \leq C(h_a^2 + h_t). \tag{4.72}$$

for a real scalar $C > 0$.

We estimate the rate of convergence of the age (time) discretisation by refining the age (time) mesh and calculating the error measure $e(h_t, h_a)$ at each refinement cycle. To estimate the age convergence rate, after fixing an end time $T$ we consider a time step, $h_t$, small enough so that $e(h_t, h_a)$ is mainly the error in age and the time error can be neglected. We then consider an initial mesh and refine it sequentially: given $h_a$ the size of the initial mesh, for each $n$-th refinement cycle we calculate a mesh size of $h_a/2^n$. We consider the following quantity between each sequential refinement cycle:

$$r_a(h_a, h_t) = \frac{\log\left(e(h_a,h_t)/e(\frac{h_a}{2},h_t)\right)}{\log 2}. \tag{4.73}$$

The estimated convergence rate of the age discretisation is $r_a$ as $h_a \to 0$. It is possible to show that the age discretisation convergence rate is second order by showing that $r_a \approx 2$ (see Section 2.3).

We estimate the time convergence rate in a similar way. After fixing an end time $T$, we consider a refined mesh for the age domain so $e(h_t, h_a)$ is mainly the time error and the error of the age discretisation can be neglected. We consider an initial time step $h_t$ and refine sequentially: for each subsequent refinement cycle we consider a time step $h_t/2^n$. We consider the following quantity between each sequential refinement cycle:

$$r_t(h_a, h_t) = \frac{\log\left(e(h_a, h_t)/e(h_a, \frac{h_t}{2})\right)}{\log 2}. \tag{4.74}$$

The estimated convergence rate of the time discretisation is $r_t$ as $h_t \to 0$. It is possible to show that the time convergence rate is first order by showing that $r_t \approx 1$.

**Remark 4.** *When estimating the observed convergence rate of the time (age) discretisation, the number of refinement cycles needs to be limited since otherwise we can reach a time (age) mesh size so small that we can not neglect the age (time) error.*

In the numerical tests that we use to estimate the age and time convergence rates, we consider the proliferation rate, the death rate and the maximum age parameter values as in Table 4.5. A critical point in the discretisation of Eq. (4.68) is the location of the G1/S transition age with respect to the age mesh. By construction, the solution of this PDE has a derivative discontinuity at $a_{G1/S}$ (see Eq. (4.70)), since we are using linear finite elements, the functions of the solution space are linear between the nodes so the functions from this space are unable to properly approximate the derivative discontinuity at $a_{G1/S}$. The approximation deteriorates the further the $a_{G1/S}$ is from the mesh nodes. In order to analyse how the position of the G1/S transition age and velocity affects the convergence rate of the discretisation scheme, we consider the following cases and specify the situation that each case captures:

**Case 1:** $a_0 = 7.5$, $a_1 = 0$: constant $a_{G1/S}$ transition age located at a mesh point.

**Case 2:** $a_0 = a_{G1/S}$, $a_1 = 0$: constant $a_{G1/S}$ transition age which is not located at a mesh point.

**Case 3:** $a_0 = 7.5$, $a_1 = 0.1$: moving $a_{G1/S}$ transition age with slope 0.1.

**Case 4:** $a_0 = 7.5$, $a_1 = 1$: moving $a_{G1/S}$ transition age with slope 1.

The first two cases consider a constant $a_{G1/S}$; case 1 considers it at a mesh point while case 2 the is in the middle of two mesh points. Case 3 and 4 assume a moving $a_{G1/S}$ which is initially at a mesh point and has a small and high velocity, respectively. By considering a mesh like this, the point $a_{G1/S}$ will be at a mesh point for point 1,3,4 and for case 2 it will be between mesh nodes. However since the age refinement is much finer in the domain $[0, 15]$, the distance from $a_{G1/S}$ will be small so we expect the discretisation method will still have the theoretical convergence order.

We consider $\alpha = 2$ for the convergence analysis simulations. All other parameter values are as per Table 4.4.

*Age discretisation convergence analysis*

We consider an end time $T = 0.01$ and a time step of 0.001. The age mesh is set such that at the $n$-th refinement level two subdomains are discretised differently: the subdomain $[0, 15]$ is discretised with an step size of $h_a = 15/2^n$, and the domain $[15, a_{max}]$ is discretised with a step size of $h_a = (a_{max} - 15)/2^n$.

In Table 4.1, we summarise the results when estimating the age discretisation convergence order for the case in which $a_0 = 7.5$ and $a_1 = 0$. Each row represents a different refinement cycle where we include the time mesh size, $h_t$, the age mesh size, $h_a$, the error measure, $e(h_a, h_t)$, and the estimated convergence rate of the age discretisation, $r_a(h_a, h_t)$. Since the estimated convergence rates is calculated between the errors from subsequent refined meshes, we include the observed convergence rates from the second refinement onwards. We observe that as the age mesh size decreases the error measure decreases. Furthermore, we can observe that $r_a \approx 2$ which is the theoretical convergence rate.

| $h_t$ | $h_a$ | $e(h_a, h_t)$ | $r_a(h_a, h_t)$ |
|---|---|---|---|
| 1.000e-03 | 2.500e+01 | 2.223e-02 | - |
| 1.000e-03 | 1.250e+01 | 6.318e-03 | 1.81 |
| 1.000e-03 | 6.250e+00 | 1.649e-03 | 1.94 |
| 1.000e-03 | 3.125e+00 | 4.187e-04 | 1.98 |
| 1.000e-03 | 1.562e+00 | 1.062e-04 | 1.98 |
| 1.000e-03 | 7.812e-01 | 2.870e-05 | 1.89 |

**Table 4.1:** Estimates of the age convergence rates for discretisation scheme defined by Algorithm 4.1 when solving the system (4.51) for case 1: $a_0 = 7.5$ and $a_1 = 0$.

In Figure 4.5 we use Eq. (4.71) to plot the error measure when performing the convergence analysis for the other three cases. The summary tables are found in Appendix B.1. For all cases $r_a \approx 2$ and we can observe that the second case has the highest $L^2$ errors in all the age refinement cycles considered during the convergence analysis. This happens because the G1/S transition age is not initiated at a point in the mesh unlike in the other cases. Considering cases 3 and 4 we can investigate the effect of the velocity on the order of convergence. Both cases show a higher error than in case 1 and we can conclude that the slope of the moving G1/S transition age affects the convergence order of the discretisation scheme. While this effect is less pronounced in Figure 4.5, it can be observed by analysing the complete summary tables (see B.1). We notice that case 4 (the moving G1/S age with slope 1) has higher error measures so we can conclude that the performance of the discretisation scheme deteriorates as the velocity increases.

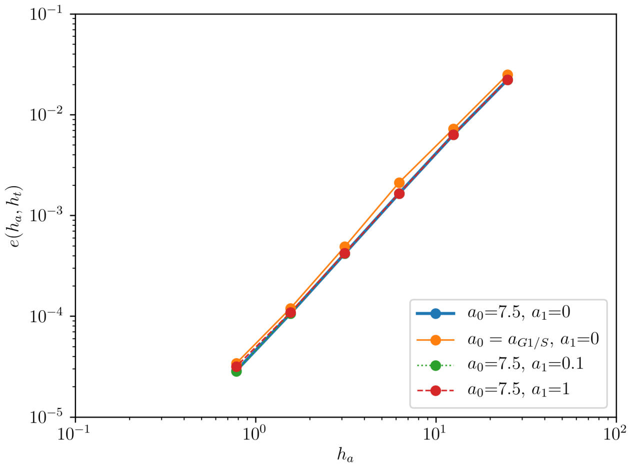


**Figure 4.5:** Plots of the error measures (Eq. (4.71)) while performing the convergence analysis of age discretisation when solving the modified system (4.51). The numerical solution is compared against the explicit solution given by Eq. (4.70). The error measures are plotted against $h_a$ on a loglog scale.

*Convergence analysis of the time discretisation*

We consider an end time $T = 2.5$ and age mesh refined as follows: two subdomains are discretised differently: the subdomain $[0, 7.5]$ is discretised with a step size of $h_a = 7.5/2^{12} \approx 3.66 \times 10^{-3}$, and the domain $[15, a_{max}]$ is discretised with a step size of $h_a = (a_{max} - 15)/2^{12}$. We consider an initial time step of 2.5 and in the n-th refinement cycle we consider a time step of $h_t = 2.5/2^n$.

In Table 4.2 we summarise the results of the convergence analysis of case 1. Each row represents a different refinement cycle. For each refinement cycle we include the time mesh size, $h_t$, the age mesh size, $h_a$, the error measure, $e(h_a, h_t)$, and the estimated convergence rate of the time discretisation, $r_t(h_a, h_t)$. We observe that as $h_t$ decreases then the error measure decreases. Furthermore we observe that $r_t \approx 1$ which is the expected convergence rate.

| $h_t$ | $h_a$ | $e(h_a, h_t)$ | $r_a(h_a, h_t)$ |
|---|---|---|---|
| 2.500e+00 | 2.441e-02 | 8.997e+01 | - |
| 1.250e+00 | 2.441e-02 | 3.764e+01 | 1.26 |
| 6.250e-01 | 2.441e-02 | 1.711e+01 | 1.14 |
| 3.125e-01 | 2.441e-02 | 8.200e+00 | 1.06 |
| 1.562e-01 | 2.441e-02 | 4.025e+00 | 1.03 |

**Table 4.2:** Time convergence rate estimates of the discretisation scheme defined by Algorithm 4.1 when solving the system (4.51) for case 1: $a_0 = 7.5$ and $a_1 = 0$.

In Figure 4.6 we plot the error measure (Eq. (4.71)) when performing the convergence analysis for the other three cases. The full summary of the convergence tests for the three cases are found in Appendix B.1. For all cases $r_t \approx 1$. We observe a similar behaviour as in the age convergence test: case 2 has the highest errors, case 3 and case 4 have higher errors than case 1, and case 4 has a higher error than case 3. We can then draw the same conclusions as in the age refinement case: the location of the G1/S transition age mesh with respect to the mesh affects the time convergence rate. We can also conclude that the performance of the time discretisation deteriorates as the velocity increases.

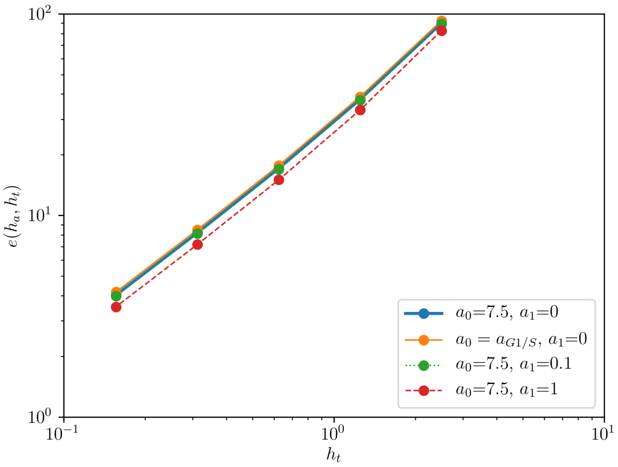


**Figure 4.6:** Plots of the error measures (Eq. (4.71)) when performing the convergence analysis of the time discretisation when solving the modified system (4.51). The numerical solution is compared against the explicit solution given by Eq. (4.70). The error measure is plotted against $h_t$ on a loglog scale.

The two previous convergence analyses revealed that considering an age mesh that is more refined in the region where the G1/S transition age is located, is sufficient for the estimated convergence rates to match the theoretical rates for similar PDEs. We also notice that the velocity of the G1/S transition age can affect the discretisation performance so we should take this into account when considering in the full system.

### Coupled age-structured model with resource concentration dynamics

We now consider the method of manufactured solutions to estimate the convergence rates of the discretisation method for the coupled age-structured model with resource concentration dynamics. As described in Section 2.3 we modify Eqs. (4.1) - (4.4), by adding the source terms, $G(a,t)$ and $H(t)$, additional boundary conditions and appropriate initial conditions, so that $F(a,t)$ and $\hat{c}(t)$ are the explicit solutions of Eq. (4.75). The modified full system that we consider is the following: we consider $u(x,t) : [0,\infty] \times [0,T] \to \mathbb{R}$ that satisfies:

$$\begin{cases} \frac{\partial u(a,t)}{\partial t} = -\left[\frac{\partial u(a,t)}{\partial a} + (\mu + b(a,c))u(a,t)\right] + G(a,t), \\ u(0,t) = 2\int_0^\infty b(a)u(a,t)\mathrm{d}a - 2\int_0^\infty b(a)F(a,t)\mathrm{d}a + F(0,t), \\ \frac{dc}{dt} = \overline{S} - kNc + H(t), \\ u(a,0) = F(a,0) \quad c(0) = c_0, \end{cases} \tag{4.75}$$

in which

$$b(a,c) = \begin{cases} \tau_p^{-1} & \text{if } a \geq a_{G1/S}(c) \text{ and } c \geq c_{cr}, \\ 0 & \text{if } a < a_{G1/S}(c) \text{ and } c \geq c_{cr}, \\ 0 & \text{if } c < c_{cr}, \end{cases}$$

$$a_{G1/S}(c(t)) = a_-(\frac{c(t)}{c_{cr}} - 1)^{-\beta},$$

$$N(t) = \int_0^\infty u(a,t)\mathrm{d}a,$$

$$G(a,t) = \frac{\partial F(a,t)}{\partial a} + (\mu + b(a,\hat{c}))F(a,t) + \frac{\partial F(a,t)}{\partial t},$$

and

$$H(t) = -[\overline{S} - k\hat{N}\hat{c}] + \overline{S} - k\hat{c}.$$

We have modified the right hand side of the full coupled system (Eqs. (4.1) - (4.4)) so that the explicit solutions for the above system are $F(a,t) = A(a,t)T(t)$ and $\hat{c}(t)$ where

$$A(a,t) = \begin{cases} \exp(-\mu a) & \text{if } a < a_{G1/S}(t) \\ \exp(-\mu a - \tau_p^{-1}(a - a_{G1/S}(t))) & \text{if } a \geq a_{G1/S}(t) \end{cases}$$

$$T(t) = 50\tanh\left(\frac{t}{\alpha}\right) + 1. \tag{4.76}$$

and

$$\hat{c}(t) = \left(C_0 + \frac{\overline{S}}{k}\left[exp(kt) - 1\right]\right) exp(-kt) \tag{4.77}$$

in which $\hat{N}(t) = \int_0^\infty F(a,t)\mathrm{d}t$.

$F(a,t)$ has the same shape as in the modified simplified model and we have chosen $\hat{c}(t)$ to be the solution of the ODE (4.4) taking $N(t) = 1$ in the ODE, in this way both functions have a similar shape as the expected solution of the full coupled system (Eqs. (4.1) - (4.4)).

In a similar fashion as in the previous system, we estimate the age and time convergence rates. We consider the parameter values as per Table 4.5 (see Section 4.6.1 below). As in the previous system, we consider an age mesh with two subdomains $[0, a_{G1/S}]$ and $[a_{G1/S}, a_{max}]$. We consider the following error measures for the age-structured population and the resource concentration solution:

$$e^u(h_a, h_t) = \|u_h(a,t) - F(a,t)\|_{L^2([0,a_{max}]\times(0,T))} \tag{4.78}$$

and

$$e^c(h_a, h_t) = |c_h - \hat{c}|_{L^2([0,T])}, \tag{4.79}$$

where we denote as $c_h$ the numerical solution of the resource concentration.

We denote the estimates of the age and time discretisation convergence rates as $r_a^x(h_a, t)$ and $r_t^x(h_a, t)$, respectively, where the superscript can be $c$ or $u$, denoting the error of the age-structured population solution or the error of the resource concentration.

For the age discretisation convergence analysis we consider an end time $T = 0.01$ and a time step of 0.001. The age mesh is set such that at the $n$-th refinement cycle two subdomains are discretised differently: the subdomain $[0, a_{G1/S}]$ is discretised with a step size of $h_a = a_{G1/S}/2^n$, and the domain $[a_{G1/S}, a_{max}]$ is discretised with a step size of $h_a = (a_{max} - a_{G1/S})/2^n$. We use a mesh like this so the mesh is more refined where the solution is non-zero so then the mesh is refined only where is needed. In

Table 4.3 we summarise the results of the age discretisation convergence analysis. Each row consists of a different refinement cycle, and includes the time and age step sizes, the error of the age-structured population and the resource concentration solution, and the estimated convergence rate for the age discretisation when discretising the cell population or the resource concentration. We can observe that that $r_a^u \approx 2$, which is the expected convergence rate. We observe also that by decreasing the age mesh size, the error measure of the resource concentration, $r_a^c(h_a, h_t)$ also decreases. This is expected since the system is coupled and a better estimates of $u(a,t)$ gives a better approximation of $\hat{c}(t)$.

| $h_t$ | $h_a$ | $e^u(h_a,h_t)$ | $r_a^u(h_a,h_t)$ | $e^c(h_a,h_t)$ | $r_a^c(h_a,h_t)$ |
|---|---|---|---|---|---|
| 1.000e-04 | 2.500e+01 | 2.540e-02 | - | 1.465e-08 | - |
| 1.000e-04 | 1.250e+01 | 1.027e-02 | 1.31 | 6.346e-09 | 1.21 |
| 1.000e-04 | 6.250e+00 | 3.172e-03 | 1.70 | 2.485e-09 | 1.35 |
| 1.000e-04 | 3.125e+00 | 8.516e-04 | 1.90 | 9.913e-10 | 1.33 |
| 1.000e-04 | 1.562e+00 | 2.171e-04 | 1.97 | 6.241e-10 | 0.67 |
| 1.000e-04 | 7.812e-01 | 5.502e-05 | 1.98 | 5.041e-10 | 0.31 |

**Table 4.3:** Estimates of the age convergence rates for discretisation scheme defined by Algorithm 4.2 when solving the system (4.75).

For the convergence analysis of the time discretisation we consider an end time of 0.15 and an age mesh discretised as follows: the subdomain $[0, a_{G1/S}]$ is discretised with a step size of $h_a = a_{G1/S}/2^{12}$, and the domain $[a_{G1/S}, a_{max}]$ is discretised with a step size of $h_a = (a_{max} - a_{G1/S})/2^{12}$. We consider an initial time step of 0.15 and at the n-th refinement cycle we consider a time step of $h_t = 0.15/2^n$. In Table 4.4 we summarise the results of the convergence analysis. We observe that in all the refinement cycles considered, the error measure of the resource concentration, $e^c(h_a, h_t)$ is below $10^{-6}$ while the time error of the age-structured population solution exceeds this value. Therefore the estimated time convergence rate is determined by the convergence rate from the age-structured population, $e^u(h_a, h_t)$, since the error of the resource concentration is sufficiently small that it can be neglected.

| $h_t$ | $h_a$ | $e^u(h_a,h_t)$ | $r_a^u(h_a,h_t)$ | $e^c(h_a,h_t)$ | $r_a^c(h_a,h_t)$ |
|---|---|---|---|---|---|
| 1.562e-02 | 9.766e-02 | 4.380e-04 | - | 2.877e-08 | - |
| 7.812e-03 | 9.766e-02 | 1.900e-04 | 1.20 | 2.919e-07 | -3.34 |
| 3.906e-03 | 9.766e-02 | 8.992e-05 | 1.08 | 3.781e-07 | -0.37 |
| 1.953e-03 | 9.766e-02 | 4.294e-05 | 1.07 | 4.346e-07 | -0.20 |
| 9.766e-04 | 9.766e-02 | 2.008e-05 | 1.10 | 4.675e-07 | -0.11 |
| 4.883e-04 | 9.766e-02 | 8.921e-06 | 1.17 | 4.852e-07 | -0.05 |
| 2.441e-04 | 9.766e-02 | 4.279e-06 | 1.06 | 4.943e-07 | -0.03 |

**Table 4.4:** Time convergence estimates of the discretisation scheme defined by Algorithm 4.2 when solving the system (4.75).

## 4.6 Numerical simulations

### 4.6.1 Model parametrisation

We consider as a case study, PC-3 prostate cancer cells. In order to parametrise our model, we use as a guideline the parameters that were estimated in [117] from the experimental time-course data of scratch assays. The authors estimated a proliferation rate of $\hat{\lambda} = 0.053 \pm 0.005 \quad [\text{hr}]^{-1}$. We assume their estimate corresponds in our model to $\tau_p^{-1}$, i.e.

$$\tau_p^{-1} = 0.053 \quad [\text{hr}]^{-1}. \tag{4.80}$$

Estimates of doubling times of PC-3 prostate cancer cells are in the range of 25-33 hr [16, 55]. In our model, the doubling time, which we denote as DT, is given by

$$DT := a_{G1/S}(t) + \tau_p,$$

therefore,

$$25 \leq DT \leq 33. \tag{4.81}$$

When the cell population reaches the steady-state age distribution, DT is given by:

$$DT := a^* + \tau_p = -\frac{1}{\mu} \log \left( \frac{\tau_p(\mu + \tau_p^{-1})}{2} \right) + \tau_p. \tag{4.82}$$

Using (4.82), estimates for the proliferation rate $\tau_p^{-1}$ (Eq. (4.80)) and the range of values for the doubling time (Eq. (4.81)), we obtain a range of values for the death term, $\mu \in [0.0233 - 0.0333]$. We assume that $\mu = 0.0283\ [\text{hr}]^{-1}$ (the midpoint value).

In [117] the carrying capacity was also estimated with $\bar{K} = 2.3 \times 10^{-3} \pm 2 \times 10^{-4}\ [\text{cells}]/[\mu m]^2$. Since the size of the subregion where they estimated $\bar{K}$ is $1430 \times 200\ [\mu m]^2$, the carrying capacity in cell number is

$$N_\infty = 657.8 \pm 57.2\ [\text{cells}]. \tag{4.83}$$

We therefore assume that

$$N_\infty = 657.8\ [\text{cells}]. \tag{4.84}$$

In [50] the oxygen consumption rate of PC-3 prostate cancer cells was estimated to be approximately $\frac{2[nM]}{[\text{min}]\,[10^6\text{cells}]} = 1.2 \times 10^{-4} \frac{[nM]}{[\text{hr}]\,[\text{cell}]}$. Using this value for the consumption rate, $k$, and the estimate of the carrying capacity (Eq. (4.84)) in the formula for the steady-state value of the total population in our model (Eq. (4.37)), we determine the value of the resource flux, $\bar{S} = 26.63 \times 10^{-4} \frac{[nM]}{[\text{hr}]}$.

Regarding the parameters that describe the resource dependence of the G1/S transition age, we take the parameter values that were derived from an intracellular model for an oxygen-regulated proliferation rate in [53], namely, the dimensionless parameter values $\beta = 0.2$ and $c_{cr} = 0.23$. For determining the parameter value of $a_-$, we notice that the admissible range of values for $a_{G1/S}(t)$ is $6 \leq a_{G1/S}(t) \leq 14$ given the range values of DT (Eq. (4.81)) and the estimate of $\tau_p$ (Eq. (4.80)). We consider $a_- = 8.25$ [hr], so $a_{G1/S}(t)$ is in this range for all our simulations.

We consider the following initial age-distribution that is a multiple of the steady-state age distribution:

$$v_0(a) = \begin{cases} \kappa \exp(-\mu a) & \text{if } a < a^* \\ \kappa \exp(-\mu a - \tau_p^{-1}(a - a^*)) & \text{if } a \geq a^* \end{cases}$$

where $a^*$ is the steady-state value of the transition age (Eq. (4.38)) and $\kappa$ is estimated so that the initial number of cells corresponds to the average number of cells considered as initial condition in [117] for the most confluent initial condition, $N(0) = 223$. We set $c_0 = c_\infty$ as the steady-state value of the resource concentration given by Eq. (4.37). The default parameters are listed in Table 4.5.

**Table 4.5:** Summary of model parameters

| Parameter | Description | Value | Units | Reference |
|---|---|---|---|---|
| $\tau_p^{-1}$ | Proliferation rate | 0.053 | $[\text{hr}]^{-1}$ | [118] |
| $\mu$ | Death rate | 0.028 | $[\text{hr}]^{-1}$ | Estimated |
| $k$ | Consumption rate | $1.2 \times 10^{-4}$ | $[\text{hr}]^{-1}$ | [50] |
| $\overline{S}$ | Resource flux | $26.635 \times 10^{-4}$ | $\frac{[nM]}{[\text{hr}]}$ | Estimated |
| $c_{cr}$ | Critical resource value | 0.023 | [nM] | [53] |
| $\beta$ | Scaling constant | 0.2 | dimensionless | [53] |
| $a_-$ | Scaling constant | 8.25 | [hr] | See text. |
| $c_0$ | Initial resource value | 0.034 | dimensionless | Estimated |
| $\kappa$ | Scaling constant | 12.488 | dimensionless | Estimated |
| $a_{max}$ | Maximum age | 100 | [hr] | Estimated |

### 4.6.2 Simplified model

The discretisation method was described and analysed in Section 4.5. Here we present the numerical results of the simplified model and compare with our analytical results.

We use Table 4.5 to fix the parameter values of $\tau_p^{-1}$ and $\mu$ and set $a_{G1/S} = 5$ hr. The simulations are run for $0 \leq t \leq 800[hr]$ while for the age domain $[0, a_{max}]$ we fix $a_{max} = 100$.

We consider as an initial condition for the simplified model an exponential function of the following form:

$$v_0(a) = m \exp(-na). \tag{4.85}$$

The value of $n$ is fixed so that the compatibility condition between the nonlocal boundary condition and the initial condition is fulfilled, i.e.

$$v_0(0) = 2 \int_0^\infty b(a) v_0(a) \mathrm{d}a. \tag{4.86}$$

Substituting the initial condition (4.85) and the proliferation rate function in Eq. (4.86), we obtain that $n$ needs to satisfy:

$$\begin{aligned} v_0(0) =& 2\int_0^\infty b(a)v_0(a)\mathrm{d}a \\ m =& 2\int_0^\infty b(a)m\exp(-na)\mathrm{d}a \\ 1 =& 2\int_{a_{G1/S}}^\infty \tau_p^{-1}\exp(-na)\mathrm{d}a \\ 1 =& \frac{2\tau_p^{-1}}{n}\exp(-na_{G1/S}) \\ n =& 2\tau_p^{-1}\exp(-na_{G1/S}) \end{aligned} \tag{4.87}$$

Using the parameter values of $\tau_p$ and $a_{max}$ as per Table 4.5 in Eq. (4.87), we determine the value of $n = 0.06$. The value of $m$ is fixed so the initial number of cells is 223, therefore $m = 13.32$. This initial number of cells matches our initial number of cells in the full model.

Considering the parameter values as per Table 4.5 and the initial condition (4.85), in Figure 4.7, we show the plots of interest when running a simulation of the simplified age-structured model. Figure 4.7 (a) shows the evolution over time of the number of cells in X, Y and N while Figure 4.7 (b) shows the evolution of the age-distribution over time). We can observe that as time progresses, the age-structured converges to the stable age-distribution given by Eq. (4.23). We now corroborate the analytical results that were presented in Sec. 4.3 with numerical simulations.

From Corollary 5.2 we know that the long time dynamics of the age-structured model is determined by the value of $a_{G1/S}$ with respect to the critical G1/S transition age $a^*_{G1/S}$. Considering $\mu$ and $\tau_p^{-1}$ we use Eq. 4.24 to estimate the critical age for the $G1/S$ transition to be

$$a_{G1/S}* = -\frac{1}{\mu}\log\left[\frac{\tau_p(\tau_p^{-1}+\mu)}{2}\right] = 9.60\ [hr] \tag{4.88}$$

As expected from Remark 3, since $a_{G1/S} < a^*_{G1/S}$ in Figure 4.7 (a) we see that the cell population grows exponentially. 4.7 (b) confirms Corollary 5.2 showing that as $t \to \infty$, the age-distribution converges to a stable-age distribution function.

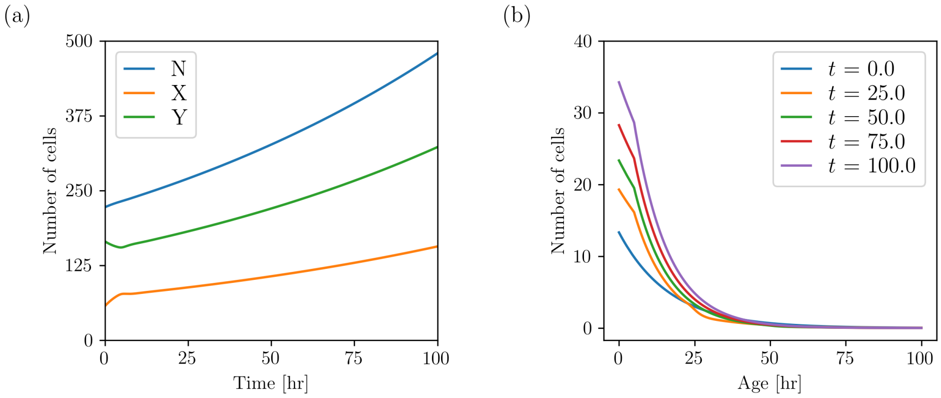


**Figure 4.7:** Series of plots showing the dynamics of the simplified age-structured model (Eq. (4.6)). In (a) the evolution of the number of cells of X, Y and N is plotted over time. $X, Y$ and $N$ are given in Eqs. (4.7) and (4.8). We can observe that after an initial transient phase, the number of cells of each subpopulation, $X$ and $Y$, grows exponentially. In (b) the age-distribution, $u(a, t)$, is plotted for distinct time points. As time increases the age-distribution settles to the stable age-distribution while still the cell number increases. Parameter values are as per Table 4.5.

To further corroborate the conclusions of Corollary 5.2 that the long time dynamics depends only on the value of the G1/S transition age, we vary $a_{G1/S}$ between 2.61 and 16.61 hr. The results presented in Figure 4.8 show how the total number of cells changes over time for different values of $a_{G1/S}$. As expected for $a_{G1/S} < a^*_{G1/S}$ the cell population grows exponentially, where for $a_{G1/S} > a^*_{G1/S}$ it decays exponentially and when $a_{G1/S} = a^*_{G1/S}$, the population remains at a constant level.

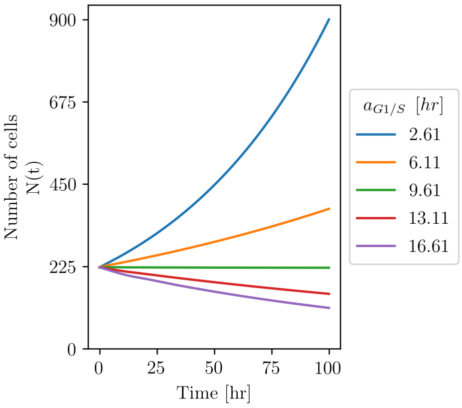


**Figure 4.8:** Series of simulations from the simplified model (Eq. (4.6)) showing how the total number of cells evolves over time for different values of $a_{G1/S}$. Other parameter values are as per Table 4.5.

We now validate through a numerical simulation the relationship between the dynamics of subpopulations X and Y given by the age-structured model (Eq. (4.6)) and the ordinary differential equation (4.13) and the delay differential equation (4.14). We compare the numerical solutions of the age-structured model with the explicit solutions of the ordinary differential equation and the numerical solutions of the delay differential equation for $[a_{G1/S}, 2a_{G1/S}]$, the latter obtained from the method of steps. The results presented in Figure 4.9 that there is an excellent agreement between the two methods of solutions.

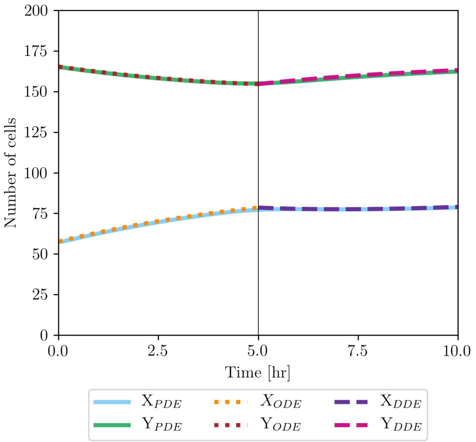


**Figure 4.9:** Comparison of the evolution of the mature and immature subpopulation dynamics given by the simplified model (4.6) and the ODE (4.13) and the DDE (4.14). The subpopulations evolution is obtained from the simplified model by integrating the numerical solution $n(a,t)$ over $[0, a_{G1/S}]$ and $[a_{G1/S}, \infty]$ to obtain the evolution equations for $X$ and $Y$, respectively. Parameter values are as per Table 4.5.

We analyse the presence of damped oscillations in the solution of the simplified model (4.6). We fix $a_{G1/S} = a^*_{G1/S}$ so the total number of cells converges to a steady-state value $N^*$ as $t \to \infty$. In Theorem 5 we show that the long-time dynamics of the subpopulations of the simplified model can be described by the DDE (4.14). We analyse how does the oscillatory dynamics depends on the delay. We show in Eq. (4.27) that when $a_{G1/S} = a^*_{G1/S}$, the delay is given by $a^*_{G1/S}$. We vary the proliferation rate $\tau_p^{-1}$ from 0.08 to 0.12 and since $a^*_{G1/S}$ depends on the proliferation rate, $a^*_{G1/S}$ is varied from is the delay in Eq. 4.27, we are increasing the size of the delay as well. In Figure 4.10 we plot the evolution of the total number of cells, $N$, over time for different values $\tau_p^{-1}$. We can observe that the larger the delay size, the oscillations have a smaller amplitude.

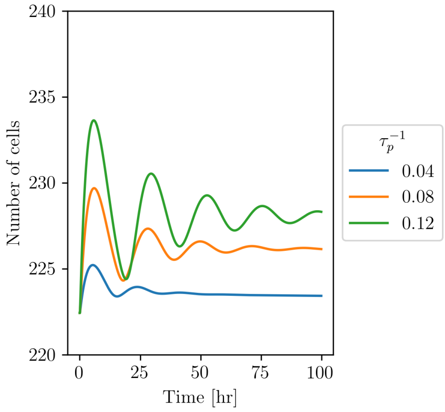


**Figure 4.10:** Evolution of the total cell number of the simplified model (Eq. (4.6)) while varying the proliferation rate $\tau_p^{-1}$. We find different behaviours depending on the value of $a_{G1/S}$ with respect to the critical G1/S transition age, $a^*_{G1/S}$.

### 4.6.3 Full model

#### Reference dynamics

In Figure 4.11 (a) and (b) we plot the time evolution of the total cell population $N(t)$ and of the resource concentration $c(t)$, respectively of the full age-structured model (Eqs. (4.1) and (4.4)) with the parameter values as per Table 4.5. Since the model satisfies the conditions of Theorem 10, the total cell population follows a logistic-type behaviour: it grows exponentially and then saturates, as expected. The resource concentration increases initially towards a maximum value and then decreases until it reaches its steady-state value. In Figure 4.11 (c) we plot the evolution of the transition age $a_{G1/S}$ and observe its dependence on the resource concentration: it decreases to a minimum value and then increases towards the steady-state value given by Eq. (4.38). In Figure 4.11 (d) we plot the per capita growth rate $\sigma(N)$ as a function of the total cell population $N$; as expected, we observe biphasic behaviour, with an initial increase in $\sigma$ with respect to cell density followed by a linear decrease.

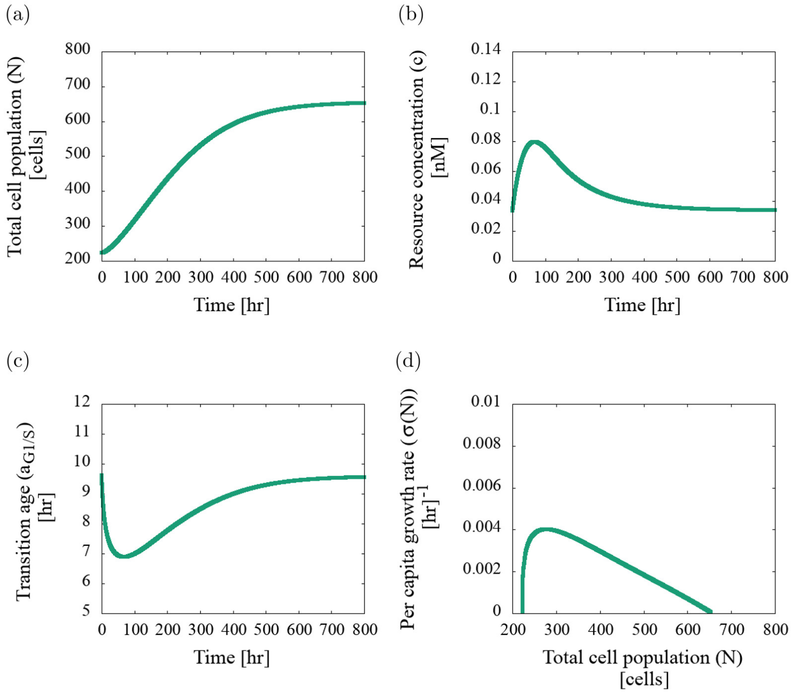


**Figure 4.11:** Evolution of the age-structured model with resource-regulated proliferation, Eqs. (4.1) and (4.4). In (a) we plot the total cell population evolution over time and observe that it follows a logistic-type growth. In (b) we plot the resource evolution and observe that it rapidly increases initially and then monotonically decreases to the steady-state value. In (c) we plot the evolution of the transition age $a_{G1/S}(t)$ over time and the inverse dependence of the resource concentration on the transition age can be observed. In (d) we plot the per capita growth rate against the total cell population evolution for which two proliferation phases can be observed (a rapid increase, then slower decrease). Parameter values as per Table 4.5.

### Sensitivity analysis

Here we perform a sensitivity analysis, focusing on those parameters that modulate the resource dynamics in order to understand their role in the biphasic behaviour of the per capita growth rate.

We first focus on two parameters which we expect that they can be experimentally manipulated: the resource flux, $\overline{S}$, and the initial concentration of resource, $c_0$. We consider $\overline{S}$ in the domain $[25.63 \times 10^{-4} - 28.63 \times 10^{-4}]$ so that the steady-state value of the cell population number, $N_\infty$, is within the confidence interval given by Eq. (4.83). Consistent with our analysis (see Eq. (4.37)), $N_\infty$ increases as $\overline{S}$ increases (Figure 4.12). We observe also that the maximum value of $c(t)$ increases and the minimum value of $a_{G1/S}(t)$ decreases as $\overline{S}$ increases (see Figure 4.12 (c) and (d)). We observe in Figure 4.12 (d) that the biphasic behaviour persists for all values of the resource flux considered.

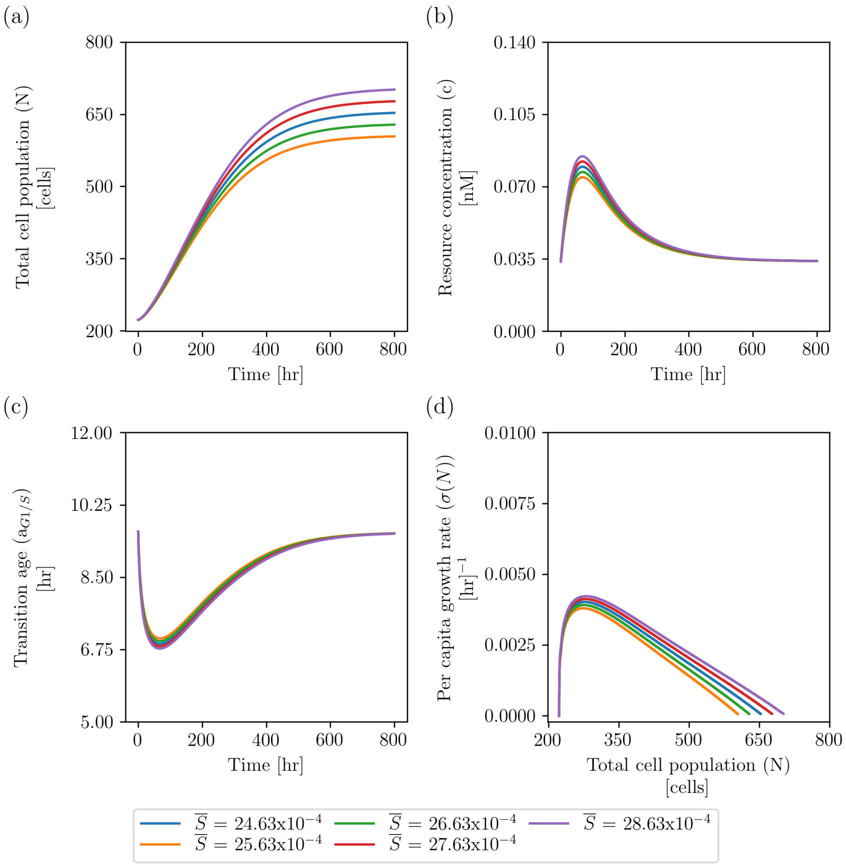


**Figure 4.12:** Series of plots showing the dynamics of the coupled age-structured model with resource-regulated proliferation given by Eqs. (4.1) and (4.4), as we vary the resource flux, $\overline{S}$. Increasing the value of $\overline{S}$: (a) increases the steady-state value of the total cell population, $N_\infty$, (b) increases the maximum resource concentration and the time it takes to reach the steady-state value, and (c) reduces the minimum value of the transition age $a_{G1/S}$. (d) The plots of the per capita growth rate against the total cell population exhibit biphasic dynamics for the selected values of $\overline{S}$. Parameter values as per Table 4.5.

We now vary the initial resource concentration $c_0$. We restrict our analysis to values of $c_0$ with $c_0 > c_\infty$, since when $c_0 < c_\infty$ the population decreases which is incompatible with the experimental observations. In Figure 4.13 we show that as $c_0$ increases from 0.034 to 0.114, the duration of the disturbance phase is shortened. The model predicts that by increasing $c_0$, the total cell population evolution does not change significantly, however the resource evolution changes from biphasic to monotonic, i.e. the duration of the disturbance phase decreases and eventually disappears as $c_0$ increases (see Figure 4.13 (b)). Increasing the value of the initial resource concentration shortens the duration of the initial phase and it is only possible to observe the latter phase in which the resource concentration declines to the steady-state value. This effect can also be observed in Figure 4.13 (c) where the initial decrease of the transition age $a_{G1/S}$ is shortened as $c_0$ increases. We observe that increasing $c_0$ decreases the duration of the disturbance phase where the dependence of the per capita growth rate with respect to the total cell population is not linear (see Figure 4.13 (d)). We notice that during the disturbance phase, the per capita growth rate does not increase monotonically with respect to the total cell population as we saw previously for the reference dynamics and when varying the resource flux, $\overline{S}$. For the simulations with $c_0 > c_\infty = 0.034$, the per capita growth rate decreases initially, then increases and finally decreases linearly with respect to the total cell population.

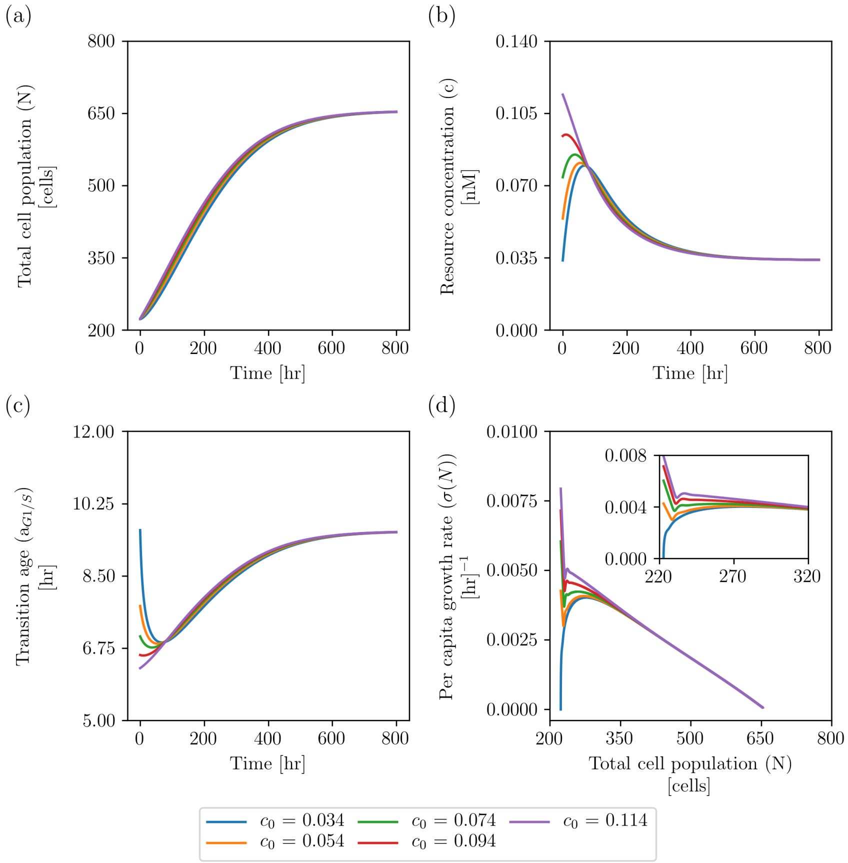


**Figure 4.13:** Series of plots showing the dynamics of the coupled age-structured model with resource-regulated proliferation, given by Eqs. (4.1) and (4.4), as we vary the initial resource concentration value, $c_0$. Increasing $c_0$: (a) does not affect the total cell population, (b) increases the maximum resource concentration but reduces the time it takes to reach it, and (c) makes the initial decrease of the transition age $a_{G1/S}$ disappear. (d) Increasing the initial resource concentration, $c_0$, affects the plots of the per capita growth rate against the total cell population by shortening the duration of the disturbance phase where there is no monotonic decreasing dependence. Parameter values as per Table 4.5.

We now perform a sensitivity analysis on other model parameters to show that our model predictions are robust. We consider as reference the initial age distribution and the default parameter values given in Section 4.6.1. Under these conditions, the total cell population undergoes logistic behaviour and the per capita growth rate exhibits two proliferation phases, as can be seen in Figure 4.11. We investigate whether small variations in the parameter values affect the predicted behaviour. When we vary a specific parameter, all other parameters are held fixed at the default values reported in Table 4.5.

To fix the range in which we vary a parameter, we calculate the minimum and maximum value of that parameter so that $550 \leq N_\infty \leq 750$ while all the other parameter values are fixed at their reference values. In this way we ensure that the carrying capacity is consistent with the estimates for PC-3 prostate cancer cells, between 600 and 700 cells (see Eq. (4.83)) [118].

For each simulation, we extract the following quantities (see Figure 4.14):

- $a_{min} := \min\limits_{0 \leq t \leq T} a_{G1/S}(t)$
- $c_{max} := \max\limits_{0 \leq t \leq T} c(t)$
- $\sigma_{max} := \max\limits_{0 \leq t \leq T} \sigma(N(t))$
- $t_{amin}$ such that $a_{G1/S}(t_{amin}) = a_{min}$
- $t_{cmax}$ such that $c(t_{cmax}) = c_{max}$
- $t_{\sigma max}$ such that $c(t_{\sigma max}) = \sigma_{max}$

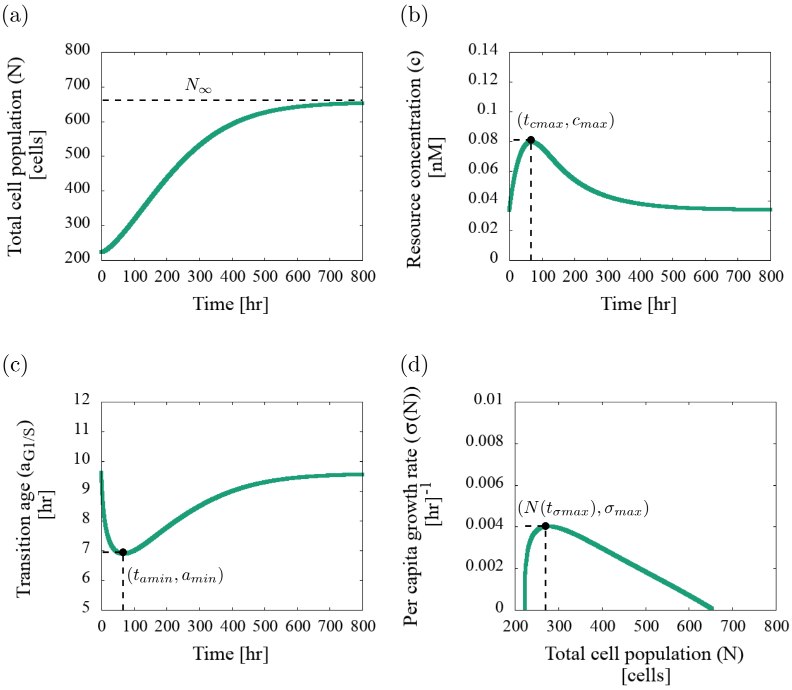


**Figure 4.14:** Results from a typical simulation indicating how the quantities of interest are defined. The quantities of interest are: the maximum value of the resource concentration, $c_{max}$, the minimum value of the transition age, $a_{min}$, the maximum value of the per capita growth rate, $\sigma_{max}$, and the times, $t_{cmax}$, $t_{amin}$ and $t_{\sigma max}$ at which these extrema are attained.

For all choices of the model parameters considered, the qualitative behaviour of the dependent variables was the same as the one presented in Figure B.6 in Appendix B.1. In Table B.7 we include the mean and standard deviation of the quantities of interest when performing the sensitivity analysis for each of the model parameters. For each parameter, we include the range of values where the parameter was varied, and the mean and the standard deviation of each quantity of interest. We notice that $t_{amin} = t_{cmax}$ and this is expected given the definition of $a_{G1/S}$ (Eq. (4.2)) so we only include $t_{cmax}$. We observe that on average there is less than ten percent variation in the quantities of interest in all the parameter sensitivity analyses, so we can conclude that the parameter uncertainty has little effect on the overall dynamics.

### Sensitivity analysis of the initial age-distribution

In this section we investigate how varying the initial age-distribution affects the model predictions. We consider the initial age-distributions to have the general form:

$$v_0(a) = m \exp(-na), \tag{4.89}$$

where $m$ and $n$ are positive real numbers.

In order for the population to initially increase, i.e. $\frac{dN}{dt} > 0$, $n$ must satisfy the following inequality,

$$n < \frac{-\log(\mu\tau_p)}{a_{G1/S}(0)}. \tag{4.90}$$

Since the initial value of the total cell concentration is given by $N(0) = \frac{m}{n}$, we set $m$ so $N(0)$ has the same value as the one used for the reference dynamics ($N(0) = 223$; see Section 4.3). Thus we fix

$$m = 223n. \tag{4.91}$$

Guided by Table 4.5 and Eq. (4.90) we require $n < 0.06642$. Hence, we choose $n \in [0.053, 0.065]$ and calculate $m$ via Eq. (4.91). For each pair $(m, n)$ we consider the corresponding initial age-distribution and perform a numerical simulation. In Figure 4.15 we plot the evolution of different simulations for different values of $(m, n)$. We observe that the total cell population, the resource concentration and the transition age evolution do not change when $(m, n)$ vary. Only the dependence of the per capita growth rate on the total cell population exhibits small changes as we vary $(m, n)$ (see Figure 4.15 (d)). In all cases, we observe that for all initial age-distributions considered, the total cell population follows a logistic behaviour and the plot of the per capita growth rate against the total cell population possesses the two proliferation phases.

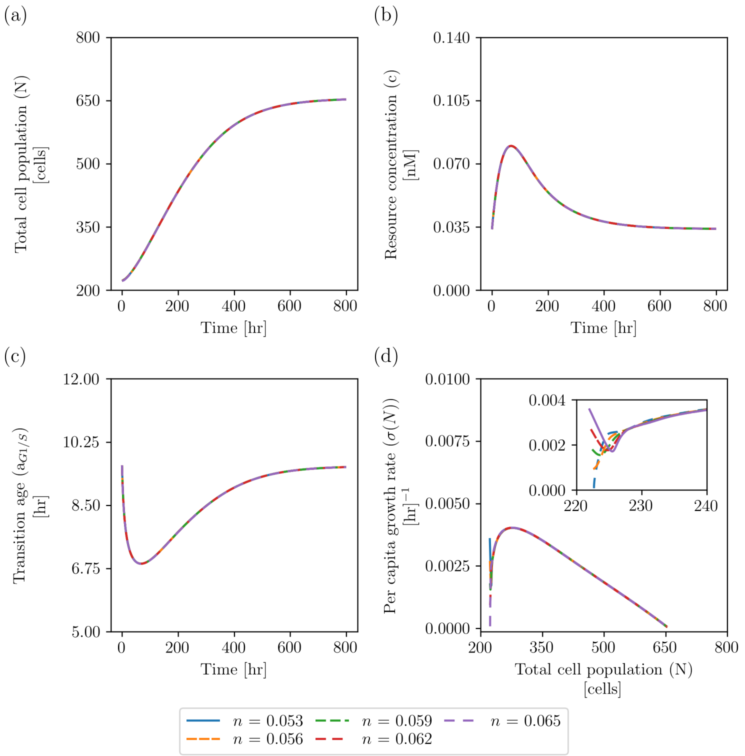


**Figure 4.15:** Series of plots showing the dynamics of the coupled age-structured model with resource-regulated proliferation, Eqs. (4.1) and (4.4), as we vary the initial age-distribution. The initial age-distributions have the general form given by Eq. (4.89). The value of $m$ is given by Eq. (4.91). Varying the initial age-distribution does not affect the: (a) total cell population, (b) resource concentration, (c) transition age evolution. (d) The plots of the per capita growth rate against the total cell population exhibit the two phases of proliferation and the initial dynamics changes slightly as we vary the initial age-distributions. Model parameters are as per Table 4.5.

### Disturbance and growth phase analysis

We have observed that varying the initial resource concentration controls the duration of the disturbance phase. We now analyse other aspects of the model behaviour when the dynamics are biphasic and when only the growth phase is visible.

The differences between the dynamics for the default parameter values and those for which $c_0 = 0.114$ nM are more visible when we analyse the evolution of the subpopulation fraction. In Figure 4.16 we observe that for the reference dynamics (where the dynamics are biphasic), the mature subpopulation increases initially and then relaxes to its stationary value, while when $c_0 = 0.114$ nM the mature population only increases. The immature fraction exhibits the opposite behaviour. In practice, we could validate our model by comparing it to experimental data plotted in this way. This type of data can now be recorded using the FUCCI technology [183] and recently mathematical models of in-vitro assays have used such data for parameter estimation [193, 220, 221].

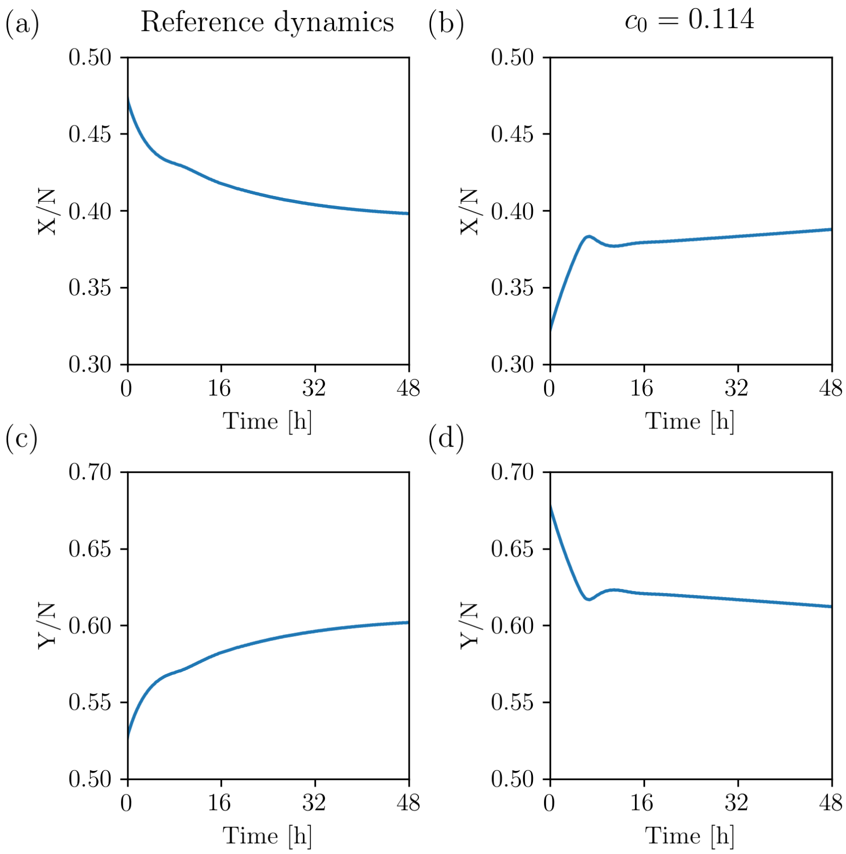


**Figure 4.16:** Series of plots showing how the dynamics of the subpopulations $X$ and $Y$ change over time. In (a) and (b) we plot the immature subpopulation dynamics while in (c) and (d) we plot the mature subpopulation dynamics for the full age-structured model (Eqs. (4.1) and (4.4)) for the default parameter values and when $c_0 = 0.114$ [nM] respectively.

### Model analysis when the resource concentration is below the critical value

If $c(t) < c_{cr}$ then, from the definition of the G1/S transition age, it follows that $a_{G1/S}(t) = \infty$ and therefore the number of cells with age $a_{G1/S}(t)$ at time $t$ is zero, $n(a_{G1/S}(t), t) = 0$. In this case, the mature compartment is empty ($Y(t) = 0$). As a result, cell proliferation does not occur so the cell population decreases. Furthermore, since $Y = 0$, from Eq. (4.47) we determine that the time derivative of the per capita growth rate is zero so the per capita growth rate is constant when $c(t) < c_{cr}$. Thereafter depending on the parameter values, three cases can arise:

I) The resource concentration increases until it reaches its critical value, $c_{cr}$, and then the cell population follows the dynamics given by Eq. (4.41) where $n(a_{G1/S}(t), t)$ is determined by Eqs. (4.42) and (4.43).

II) The resource concentration increases but never reaches $c_{cr}$ so the cell population becomes extinct.

III) The cell population oscillates between the two regimes described above since the resource concentration oscillates around $c_{cr}$.

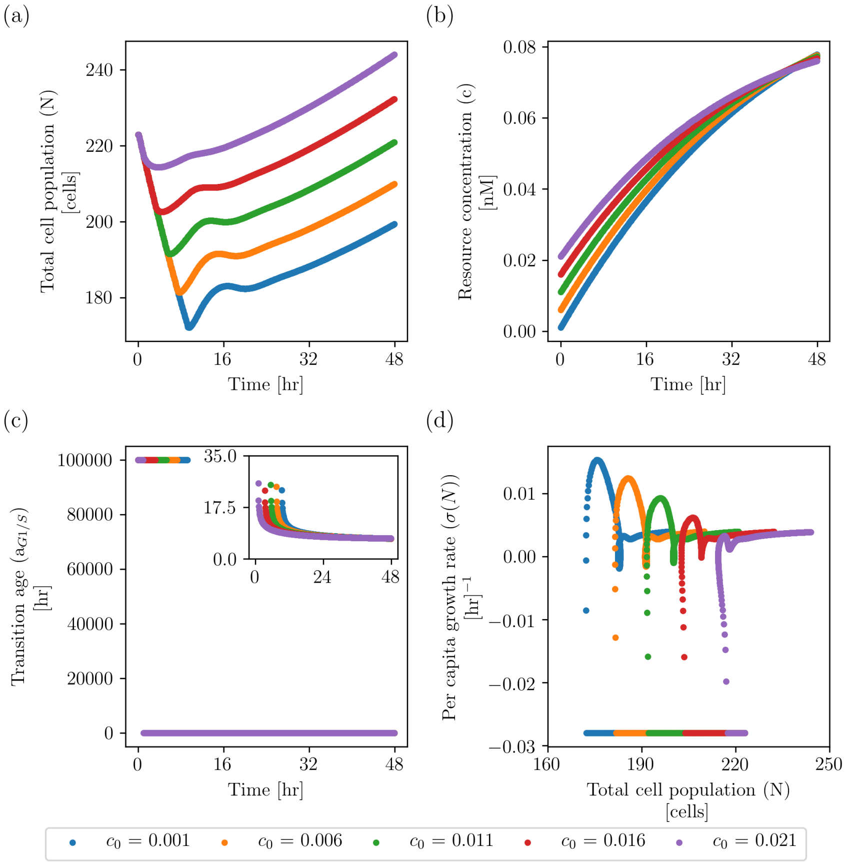


**Figure 4.17:** Series of plots showing the dynamics of the coupled age-structured model with resource-regulated proliferation as we vary the initial resource concentration value, $c_0$, below the critical value. We show the dynamics of (a) the total cell population, (b) the resource concentration, (c) the transition age and (d) the per capita growth rate against the total cell population.

To validate our claims, we perform numerical simulations in which we vary the initial resource concentration below the critical value $c_{cr}$, holding all other parameters fixed at the values stated in Table 4.5. For the simulations when $a_{G1/S}(t) = \infty$ we consider $a_{G1/S}(t) = 100,000$ hr. The results are presented in Figure 4.17 and they reveal that case (I) occurs for all values of $c_0$ considered. Since the cell population decreases, the model can not capture logistic-type behaviour. We also observe that for all values of $c_0$ considered, the per capita growth rate is constant while the cell population number decreases (see Figure 4.17 (d)).

## 4.7 Discussion

In this chapter we have presented an age-structured model for a cell population with a resource-dependent proliferation rate. Our model captures, for the first time, the biphasic behaviour in the per capita growth observed experimentally in [118]. We considered a simplified version of the model that helped us in understanding how the age-structured model is affected by the G1/S transition age, the age at which a cell can proliferate. For this simplified model we were able to derive explicit formulae for the evolution of the two subpopulations: cells which can proliferate and those which can not (mature and immature cells). We analysed the the long-time dynamics and determined the stability of the solutions. We then analysed the full model in terms of mature and immature cells. We derived necessary and sufficient conditions under which the model exhibits a logistic type behaviour and two phases of proliferation: an initial phase, which in [118] was named *disturbance phase,* in which proliferation does not follow a logistic growth, and a *growth phase,* where proliferation is approximately logistic. The biphasic behaviour was found to be the result of an initial increase in the fraction of mature cells followed by a decline to their steady-state concentration. Afterwards we presented our numerical scheme to discretise the simplified and the full model and showed they reached numerically the expected convergence rates. We then parametrised the model using PC-3 prostate cancer cells as a case study. Finally, through numerical simulations, we corroborated the results for the simplified and the full model. For the simplified model we showed the different long-time dynamics behaviour and the presence of damped oscillations as we vary the proliferation rate. For the full model, we showed that varying the resource initial value results in a change in the dependence of the per capita growth rate on the total cell population. The model predicts that the duration of the disturbance phase decreases as the initial resource concentration increases.

Analysis of the simplified model yielded insight into the dynamics of the full model. While the simplified model has been thoroughly investigated by other authors [28, 96], in this chapter we did not perform the typical analysis but rather investigated the role of the mature and immature dynamics in the per capita growth rate and the total cell population evolution. When analysing the mature and immature subpopulation dynamics, while for the simplified model we are able to give explicit formulas for their evolution, for the full model we are unable to do so. The reason for this is that a fixed G1/S transition age results in the dynamics of the mature population being described by a delay differential equation with a constant delay, while considering the G1/S transition age depending on the resource dynamics, the mature subpopulation dynamics is describe by a state-dependent delay differential equation. Although the transient dynamics of the full and simplified models differ, the long-time dynamics of the full model can be showed to converge to the long-time dynamics of a simplified model with the transition G1/S age as the critical G1/S transition age. Therefore we can expect that the oscillatory analysis of the long-time dynamics that we obtained for the simplified model, are also valid for the full model. Therefore we expect that in the full model as the proliferation rate increases, the initial amplitude of oscillations increases but as time goes to infinity, the oscillations dampen.

The non-linear model has some limitations. There is a critical constraint in the biologically realistic parameter regime. The parameters need to be chosen such that $\frac{da_{G1/s}(t)}{dt} < 1$ for all $t \in [0, T]$. Otherwise, cells could go from the S-G2-M phase back to the G1 phase without proliferating (See Eq. (4.41)). In all our simulations the derivative of the transition age is always less than one (see Figure 4.18 ). To avoid restricting the parameter regime, age-structured models with multiple compartments for the cell-cycle phases [27, 42] can be considered.

The numerical study of the discretisation schemes helped us understand how the evolution of the transition G1/S age affected the convergence rate of the discretisation methods. We observed that as the velocity of the transition G1/S increased the

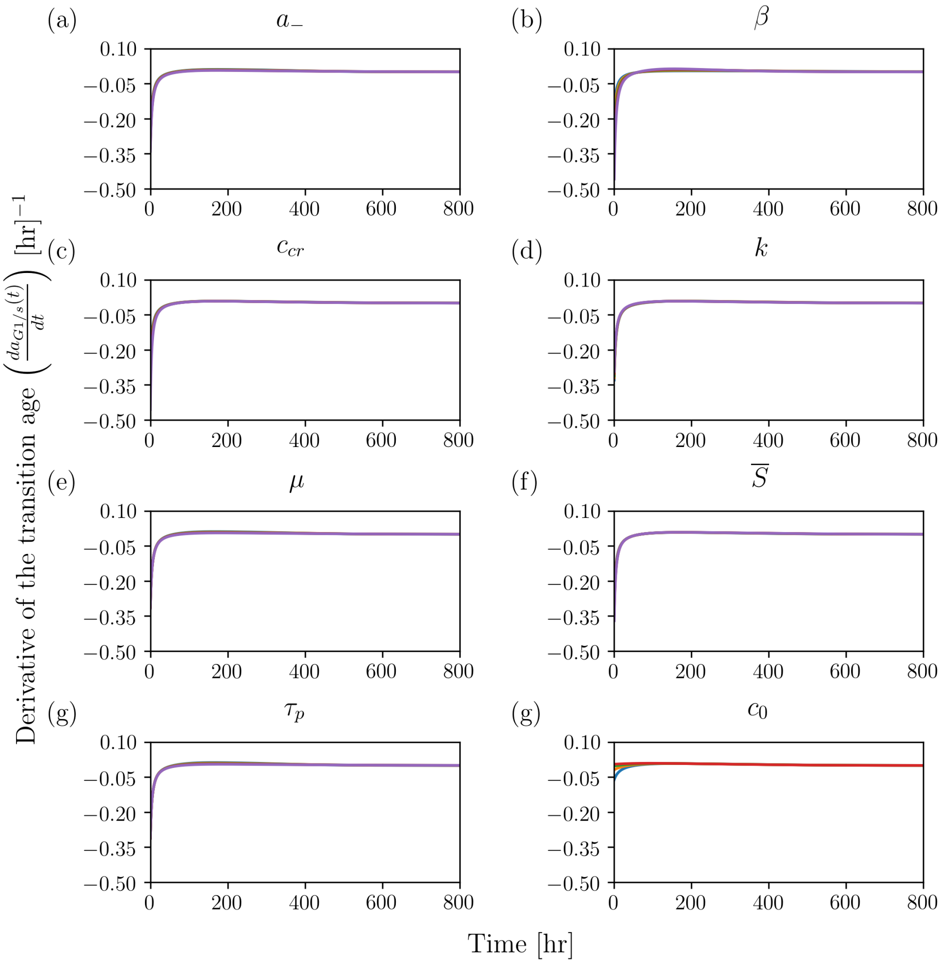


**Figure 4.18:** Plots of the transition age derivative $\frac{da_{G1/s}(t)}{dt}$ for all the simulations of the full model.

convergence estimates worsen. However as we observed in Figure 4.18, the derivative of the G1/S transition age is always smaller than 0.1 and we observed in Sec. 4.5.2 that in the case when the velocity of the transition age is 0.1, the discretisation scheme achieves the theoretical convergence rate values.

In view of the full model, the experimental observations in [118] can be explained as follows: the scratch procedure decreases the cell number in the plate and by replenishing the medium to the same quantity, as customary, the resource concentration is increased, therefore triggering the biphasic dynamics in the per capita growth rate. The predictions of the model could be experimentally tested by modifying the resource concentration in the substrate, examining the plots of per capita growth rate against the total cell population, as was performed in [118], and measuring the resource concentration, as in [27], at the same times as the cell population density is captured. The full model predicts that by increasing the initial resource concentration, the disturbance phase would shorten. Varying the resource concentration has been shown to affect the overall dynamics of the scratch assay in other situations [27].

There are different ways that we can extend this work. The full model was parametrised using estimates calculated in [118]. Their estimates were calculated with respect to the logistic equation and since we are considering a different model, the values are expected to differ. The full model could be parametrised using more appropiate methods as has been done for similar models [27, 79]. We acknowledge that the full model is just one of many possible explanations for the biphasic behaviour observed experimentally in [118]. Mechanical and chemical disturbances have been known to affect cell proliferation in other experimental settings [165]. Models that integrate explicitly biochemical pathways, such as those in [66, 163], can be adapted to investigate whether the biphasic behaviour is a product of signalling events. An inference-based modelling approach could be performed to test multiple hypotheses and assist with the design of experiments that discriminate between feasible alternatives.

## Chapter 5

# A hybrid algorithm for coupling mesoscopic and macroscopic descriptions of the stochastic FKPP model

## 5.1 Introduction

One of the characteristic features of scratch assays is the inhomogeneous cell distribution: large number of cells accumulate far away from the scratch, whereas small number of cells are observed near the edge of the scratch. If the number of cells is large, then modelling each individual cell in a stochastic or agent-based framework is inefficient and a coarse-grained description is preferred, such as partial differential equation models. However certain dynamics, such as extinction, can not be captured in a deterministic description [52].

As mentioned in Section 2.4.5, *hybrid algorithms* couple different modelling frameworks in different parts of the domain and they are used to tackle problems that have some regions in which a specific framework is more appropiate than another. A challenge of hybrid algorithms is to ensure mathematical and biophysical consistency between the modelling frameworks defined in different regions of the domain [148].

In this chapter we introduce a hybrid algorithm that couples the stochastic description of the Fisher-Kolmogorov-Petrovsky-Piscounov (FKPP) model with its mean-field analogue. The stochastic dynamics are described via a reaction-diffusion master equation (RDME) (see Section 2.4.2 for a brief introduction to RDMEs). The mean-field analogue is discretised using finite elements. We estimate the variance of the stochastic system to define the switching point between the two model formulations.

This chapter is organised as follows: in Section 5.2 we introduce the stochastic FKPP model which has been used as a prototype to introduce our hybrid algorithm. We also determine the mean-field equation and the coarse-grained factorial cumulant equation of the stochastic model. In Section 5.3 we introduce the discretisation schemes that are used to solve the coarse-grained equations and then determine their convergence properties. In Section 5.4 we introduce the reference dynamics of the stochastic model. We describe various aspects of the model dynamics, specifically the position of the interface and the velocity of the travelling wave solution. We also investigate how the coarse-grained equations approximate the first and second moments of the stochastic process. We then consider how changes in the carrying capacity and the compartment size affect the dynamics. In Section 5.5 we propose a new hybrid algorithm which we call the "A priori Variance method" (AV method) which estimates the variance at each compartment and uses this estimate to decide which compartments are updated deterministically or stochastically. In Section 5.6 we apply the AV method to the stochastic FKPP model. We perform a statistical analysis to quantify how well the hybrid model approximates the stochastic model. In Section 5.7 we compare our hybrid model with the hybrid model proposed by Spill et al. in [149]. Finally in Section 5.8 we discuss our results.

## 5.2 Stochastic FKPP Model

### 5.2.1 Model

We consider the stochastic FKPP model as a prototypic model to test our hybrid algorithm. The stochastic FKPP model is commonly used to simulate the cell front evolution in scratch assays [40, 121, 194]. Besides describing in-vitro cell dynamics, the stochastic FKPP model has been used to model a variety of problems including propagation of genes in a population [73], epidemic outbreaks [92] and there are many other examples. Following most of the current literature in the field, we choose this model to test our hybrid algorithm since it has travelling wave solutions that are such that in one part of the domain the number of molecules is close to the system size and in another the number of molecules is close to zero. Simulating the stochastic model becomes computationally prohibitive as the system size increases so a hybrid algorithm provides a more efficient method to simulate the dynamics.

The stochastic FKPP model consists of a reaction-diffusion process of a single species, $U$, on a one-dimensional spatial domain. We assume that the single species $U$ represents a cell population. We consider the domain $[0, L]$ and the time period $[0, T]$. The domain is divided into $N$ compartments of size $h = L/N$. The number of cells in the $i$-th compartment $[(i-1)h, ih]$, is denoted by $n_i$, for $i = 1, \ldots, N$. Cell migration is modelled as a stochastic jump process that consists of two sets of reactions:

$$U_i \xrightarrow{d} U_{i+1} \quad \text{for } i = 1, \ldots, N-1. \tag{5.1}$$

$$U_i \xrightarrow{d} U_{i-1} \quad \text{for } i = 2, \ldots, N. \tag{5.2}$$

where the jump rate constant is denoted by $d$. We consider $d = \frac{D}{h^2}$ where $D$ is the diffusion constant of the respective mean-field model and $h$ is the compartment size. We impose reflective boundaries on both sides of the domain $(x = 0, L)$.

In each compartment cell birth and death are modelled by an irreversible chemical reaction:

$$R_1 : U_i \xrightarrow{\lambda} 2U_i, \tag{5.3}$$

and

$$R_2 : U_i + U_i \xrightarrow{\frac{\lambda}{2K}} \emptyset, \tag{5.4}$$

where $\lambda$ is the proliferation rate and $K$ is the carrying capacity.

The stochastic FKPP model consists of the reaction-diffusion process modelled by the reactions (5.1) - (5.4).

Using the formulae given in Section 2.4.4 we determine the stoichiometric matrix, the rate function, the deterministic rate equations and the propensity functions for the reaction system:

- The stoichiometric matrix is the matrix $S \in M_{1\times 2}(\mathbb{Z})$ such that:
$$S = (1 \quad 2). \tag{5.5}$$
- The rate function is:
$$f(\varphi_i) = (\lambda\varphi_i, \frac{\lambda}{2K}\varphi_i^2). \tag{5.6}$$
- The mean-field rate equations are:
$$\frac{\mathrm{d}\varphi_i}{\mathrm{d}t} = S \cdot f(\varphi)^T = \lambda\varphi_i - \frac{\lambda}{K}\varphi_i^2. \tag{5.7}$$
- The propensities are:
$$v_1(n_i) = \lambda n_i, \quad v_2(n_i) = \frac{\lambda}{2hK} n_i(n_i - 1). \tag{5.8}$$

By considering the jump and reaction processes we determine the evolution of the RDME. Let $p(\mathbf{n}, t)$ denote the joint probability of $\mathbf{n} = [n_1, \dots, n_N]$ where $n_i$ denote the population number in the $i$-th compartment for $i = 1, \dots, N$ at time $t$. The corresponding RDME can be written as:

$$\begin{aligned}\frac{\partial p(\mathbf{n},t)}{\partial t} = & \sum_{j=1}^{N-1} d[\mathbb{E}_{j+1}^{-1}\mathbb{E}_j^1 - 1]n_j p(\mathbf{n},t) + \sum_{j=2}^{N} d[\mathbb{E}_{j-1}^{-1}\mathbb{E}_j^1 - 1]n_j p(\mathbf{n},t) \\ & \sum_{j=1}^{N} \lambda[\mathbb{E}_j^{-1} - 1]n_j p(\mathbf{n},t) + \sum_{j=1}^{N} \frac{\lambda}{2Kh}[\mathbb{E}_j^2 - 1]n_j(n_j - 1)p(\mathbf{n},t).\end{aligned} \tag{5.9}$$

We rewrite the RDME to distinguish the diffusion and the local reactions effect:

$$\frac{\partial p(\mathbf{n},t)}{\partial t} = \mathcal{R}[p(\mathbf{n},t)] + \mathcal{D}[p(\mathbf{n},t)], \tag{5.10}$$

where the effect of the diffusion process is given by:

$$\mathcal{D}[p(\mathbf{n},t)] = \sum_{j=1}^{N-1} d[\mathbb{E}_{j+1}^{-1}\mathbb{E}_j^1 - 1]n_j p(\mathbf{n},t) + \sum_{j=2}^{N} d[\mathbb{E}_{j-1}^{-1}\mathbb{E}_j^1 - 1]n_j p(\mathbf{n},t), \tag{5.11}$$

and the effect of the local reactions is given by:

$$\mathcal{R}[p(\mathbf{n},t)] = \sum_{j=1}^{N} \lambda[\mathbb{E}_j^{-1} - 1]n_j p(\mathbf{n},t) + \sum_{j=1}^{N} \frac{\lambda}{2Kh}[\mathbb{E}_j^2 - 1]n_j(n_j - 1)p(\mathbf{n},t). \tag{5.12}$$

### 5.2.2 System size expansion and intrinsic noise approximation

In this subsection we perform a moment closure and an intrinsic noise approximation to determine the mean-field and coarse-grained factorial cumulant based on the theory developed by Van Kampen [215] (see Section 2.4.4).

#### Mean-field equation

To derive the mean-field equation of the stochastic FKPP model, we first determine the evolution of the mean number of cells within each compartment. Let us denote the mean vector as $\langle \mathbf{n} \rangle = (\langle n_1 \rangle, \ldots, \langle n_N \rangle)$ where

$$\langle n_i \rangle = \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} n_i p(\mathbf{n}, t) =: \sum_{\mathbf{n}} n_i p(\mathbf{n}, t), \tag{5.13}$$

gives the mean number of cells in the $i$-th compartment, $i = 1, \ldots, N$.

To determine the evolution of $\langle n_i \rangle$ we multiply Eq. (5.9) by $n_i$ and sum over all possible values of the state vector $\mathbf{n}$. We deduce that for i $\in \{1, \ldots, N\}$, the evolution of the mean number of cells in the $i$-th compartment is given by:

$$\frac{\partial \langle n_i \rangle}{\partial t} = \sum_{\mathbf{n}} n_i \frac{\partial p(\mathbf{n}, t)}{\partial t} = \sum_{\mathbf{n}} n_i \mathcal{D}(\mathbf{n}, t) + \sum_{\mathbf{n}} n_i \mathcal{R}(\mathbf{n}, t) \tag{5.14}$$

where we have exploited Eqs. (5.10)-(5.12).

We start by analysing the diffusion part of Eq. (5.14):

$$\frac{\partial \langle n_i \rangle}{\partial t} = \sum_{\mathbf{n}} n_i \mathcal{D}(\mathbf{n}, t). \tag{5.15}$$

Expanding the sum on the right-hand side of (5.15) and considering all the possible values of the state vector $\mathbf{n}$, we deduce that $\langle n_i \rangle$ satisfies the following system:

$$\frac{\partial \langle n_i \rangle}{\partial t} = d \left( \langle n_{i+1} \rangle + \langle n_{i-1} \rangle - 2 \langle n_i \rangle \right), \quad i = 2, \ldots, N-1, \tag{5.16}$$

$$\frac{\partial \langle n_1 \rangle}{\partial t} = d \left( \langle n_2 \rangle - \langle n_1 \rangle \right), \quad \frac{\partial \langle n_N \rangle}{\partial t} = d \left( \langle n_{N-1} \rangle - \langle n_N \rangle \right). \tag{5.17}$$

Dividing Eq. (5.16) by $h$ and approximating the concentration $u(x,t)$ at $x_i$, the centre of the $i$-th compartment, as $u(x_i, t) \approx \langle n_i(t) \rangle / h$ we obtain

$$\frac{\partial u(x_i, t)}{\partial t} \approx d \left( u(x_i + h, t) - 2u(x_i, t) + u(x_i - h, t) \right). \tag{5.18}$$

By performing a Taylor expansion around $x = x_i$, Eq. (5.18) reduces to

$$\frac{\partial u(x_i, t)}{\partial t} = dh^2 \frac{\partial^2}{\partial x^2} u(x_i, t) + \mathcal{O}(h^4). \tag{5.19}$$

Finally, in the limit as $h \to 0$, given that $d = D/h^2$, the evolution of $u(x,t)$ is then given by a diffusion equation.

A similar argument applied to Eq. (5.17) enables us deduce that the concentration $u(x,t)$ satisfies Neumann boundary conditions.

We now focus on the reaction term in Eq. (5.14). By expanding the sum over all possible values of the state vector $\mathbf{n}$, the reaction term simplifies as follows:

$$\begin{aligned}\sum_{\mathbf{n}} n_i \mathcal{R}(\mathbf{n},t) &= \sum_{n_i=0}^{\infty} \left( \lambda [\mathbb{E}_i^{-1} - 1] n_i p(\mathbf{n},t) + \frac{\lambda}{2Kh} [\mathbb{E}_i^2 - 1] n_i (n_i - 1) p(\mathbf{n},t) \right) \\ &= \lambda \langle n_i \rangle - \frac{\lambda}{hK} \left[ \langle n_i^2 \rangle - \langle n_i \rangle \right]. \end{aligned} \tag{5.20}$$

We notice that the left-hand side of Eq. (5.20) depends on the second moments of $n_i$. Under the moment closure approximation $\langle n_i^2 \rangle - \langle n_i \rangle \approx \langle n_i \rangle^2$ we have:

$$\sum_{\mathbf{n}} n_i \mathcal{R}(\mathbf{n},t) = \lambda \langle n_i \rangle - \frac{\lambda}{hK} \langle n_i \rangle^2 \tag{5.21}$$

The moment close approximation consists of approximating the stochastic process by a Poisson distribution.

Dividing Eq. (5.20) by $h$ and approximating $u(x,t)$ at $x_i$ the centre of the $i$-th compartment, as $u(x_i,t) \approx \langle n_i(t) \rangle / h$, we have

$$\frac{\partial u(x_i,t)}{\partial t} = \lambda u(x_i,t) - \frac{\lambda}{K} u(x_i,t)^2. \tag{5.22}$$

Combining Eqs. (5.19) and (5.22) in the limit as $h \to 0$, we conclude that the mean-field equation for the cell concentration, $u(x,t)$, is:

$$\begin{cases} \frac{\partial u}{\partial t} = D\Delta u + \lambda u \left(1 - \frac{u}{K}\right) & x \in (0,L), t \in (0,T), \\ \frac{\partial u}{\partial x} = 0 & t \in (0,T), \\ u(x,0) = v(x) & x \in [0,L], \end{cases} \tag{5.23}$$

where $v(x)$ is the initial cell distribution.

### Intrinsic noise approximation

We now perform the intrinsic noise approximation of the stochastic FKPP equation, by following the methodology described in Section 2.4.4.

As in the derivation of the mean-field equation, we analyse the diffusion and reaction terms of the master equation separately. We focus first on the reaction term and consider the system size expansion in each compartment. The master equation for the reaction in the $i$-th compartment is given by:

$$\frac{\partial p(n_i,t)}{\partial t} = \mathcal{R}[p(n_i,t)] = \lambda [\mathbb{E}_i^{-1} - 1] n_i p(n_i,t) + \frac{\lambda}{2Kh} [\mathbb{E}_i^2 - 1] n_i (n_i - 1) p(n_i,t). \tag{5.24}$$

Given Eq. (5.24), we consider the system size expansion or Van Kampen ansatz: we assume that the cell number, $n_i$, can be defined in terms of the system size $\Omega$, a deterministic quantity $\phi_i$ and a new random variable $\epsilon_i$ in the following way:

$$n_i = \Omega\phi_i + \Omega^{1/2}\epsilon_i. \tag{5.25}$$

where $\Omega = Kh$.

We summarise the results of the system size expansion in the following theorem.

**Theorem 11.** *By considering the system size expansion* (5.25) *when* $\Omega >> 1$ *and denoting the probability distribution of* $\epsilon_i$ *by* $\Pi_i$*, we deduce the following results:*

- *The evolution of* $\Pi_i$ *is given by:*

$$\frac{\partial \Pi_i}{\partial t} = -A_i(t)\frac{\partial}{\partial \epsilon_i}(\Pi_i \epsilon_i) + \frac{B_i(t)}{2}\frac{\partial^2 \Pi_i}{\partial \epsilon_i^2}. \tag{5.26}$$

*where*

$$A_i(t) := \lambda - 2\lambda\phi_i(t) \quad \text{and} \quad B_i(t) := 2\lambda\phi_i(t)^2 + \lambda\phi_i(t). \tag{5.27}$$

- *The evolution equation for* $\phi_i$ *is given by:*

$$\frac{\partial \phi_i}{\partial t} = -\lambda\phi_i^2 + \lambda\phi_i. \tag{5.28}$$

- *When* $\Omega >> 1$ *we have that:*

$$\frac{n_i}{\Omega} \approx \phi_i. \tag{5.29}$$

*Proof.* The proof of this Theorem can be found in Appendix C.1. □

We now follow the procedure in [188] to extend the linear noise approximation to RDME. For $i, j \in \{1, \ldots, N\}$, the factorial cumulant of $n_i$ and $n_j$ is given by

$$[n_i, n_j] = cov(n_i, n_j) - \delta_{ij}\langle n^i \rangle. \tag{5.30}$$

where $cov(n^i, n^j) = \langle n_i n_j \rangle - \langle n_i \rangle \langle n_j \rangle$.

In order to determine the evolution of the factorial cumulant, we will need to use the following result.

**Lemma 12.** *Given a system size expansion* (5.25) *and a noise term* $\epsilon_i$ *that satisfies the Fokker-Plank equation defined by Eq.* (5.26)*, the covariance has the following dynamics:*

$$\frac{d}{dt}cov(\epsilon_i, \epsilon_j) = [A_i(t) + A_j(t)]cov(\epsilon_i, \epsilon_j) + \delta_{i,j}B_i(t) \tag{5.31}$$

*where* $A_i(t)$ *and* $B_i(t)$ *are given by* (5.27).

*Proof.* The proof of this lemma can be found in [188]. □

We determine the evolution of the factorial cumulant by exploiting Lemma 12.

**Theorem 13.** *Given the system size expansion, Eq.* (5.25)*, and a noise term* $\epsilon_i$ *that satisfies Eq.* (5.26)*, the evolution of the factorial cumulant due to the reactions terms has the following dynamics:*

$$\frac{\mathrm{d}[n_i, n_j]}{\mathrm{d}t} = 2\lambda\left[1 - \phi_i - \phi_j\right][n_i, n_j] + \delta_{ij}\left(2\lambda\Omega\phi_i - \lambda\Omega\left(\phi_i\right)^2\right) \tag{5.32}$$

*Proof.* By substituting Eq. (5.25) in Eq. (5.30) we conclude that the factorial cumulant scales with the system size as:

$$[n_i, n_j] = \Omega cov(\epsilon_i, \epsilon_j) - \delta_{i,j}\langle n_i \rangle \tag{5.33}$$

The evolution of $\langle n_i \rangle$ is determined by the fact that $n_i \approx \Omega\phi_i$ and the corresponding evolution of $\phi_i$ is given by Eq. (5.28):

$$\frac{d\langle n_i \rangle}{dt} = \Omega \frac{d\phi_i}{dt} = -\Omega\lambda\phi_i^2 + \Omega\lambda\phi_i. \tag{5.34}$$

Using Eqs. (5.31) and (5.34) and the fact that we express the factorial cumulant as per Eq. (5.33), we obtain the desired evolution equation for the factorial cumulant. □

To verify formula (5.32), we consider a well-mixed situation for which the cell number is large so the system size expansion is valid. In this case we compare the dynamics of the factorial cumulant given by the simulations with the evolution predicted by Eq. (5.32). The results are reported in Appendix C.2.

The next step in the derivation is to rescale the evolution equation for the factorial cumulant (5.34), i.e. to divide the evolution equation by $h^2$ and derive the evolution in concentration terms. If we divide the expression for the factorial cumulant (5.30) by $h^2$ and take the limit $h \to 0$, then the initial condition of the coarse-grained factorial cumulant should satisfy

$$\lim_{h\to 0} \frac{[n_i, n_j]}{h^2} = \lim_{h\to 0} \frac{cov(n_i, n_j)}{h^2} - \lim_{h\to 0} \frac{\delta_{ij}}{h} \frac{\langle n^i \rangle}{h} = \delta(x_i - x_j)u(x_i, 0) \tag{5.35}$$

where we have used the fact that the covariance of the system is zero at $t = 0$ and that $\frac{\langle n^i(t) \rangle}{h} \to u(x_i, t)$ as $h \to 0$ where $x_i$ is the center of the $i$-th compartment. Since we want to consider a smoother initial condition, we consider instead $u_i = \frac{n^i}{h}$ and its corresponding factorial cumulant: $[u_i, u_j] := [\frac{n^i}{h}, \frac{n^j}{h}]$. We now determine the evolution of the coarse-grained factorial cumulant $w(x_i, x_j, t) = \lim_{h\to 0}[u_i, u_j]$.

**Remark 5.** *We note that calculating the limit of the RDME as h tends to zero is only valid for dimension one. When the dimension is larger than one, the RDME has been shown to not model properly bimolecular reactions [110–112] which are present in our system. To consider our methodology in higher dimensions we need to use other frameworks, like the convergent RDME (CRDME) [111], that converge to a continuum model as the mesh spacing approaches zero.*

**Theorem 14.** *For the stochastic KPP model with RDME given by Eq. (5.9), its coarse-grained factorial cumulant,* $\lim_{h\to 0}[u_i, u_j]$, *converges to the function* $w(x, y, t) : (0, L) \times (0, L) \times (0, T) \to \mathbb{R}$ *that satisfies the following system:*

$$\begin{aligned}\frac{\partial w}{\partial t}(x, y, t) =& D\Delta w(x, y, t) + 2\lambda \left[1 - \frac{u(x, t)}{K} - \frac{u(y, t)}{K}\right] w(x, y, t) \\ &+ \lambda\delta(x - y) \left[u(x, t) + \frac{2u^2(x, t)}{K}\right] + \lambda \mathbb{1}_{x-y=0}(x, y) \left[u - \frac{3u^2}{K}\right]\end{aligned} \tag{5.36}$$

*with Neumann boundary conditions*

$$\frac{\partial w(x, y, t)}{\partial \nu} = 0 \quad \textit{for} \quad (x, y, t) \textit{ in } \partial[0, L]^2 \times [0, T], \tag{5.37}$$

*and initial condition*

$$w(x,y,0) = -\mathbb{1}_{x=y}u(x,0) \quad \textit{for} \quad (x,y) \textit{ in } [0,L]^2, \tag{5.38}$$

*where* $u : \mathbb{R} \times [0,T] \to \mathbb{R}$ *satisfies the system* (5.23).

*Proof.* By rewriting the terms that define $[u_i, u_j]$ and $[n_i, n_j]$ we determine how the two factorial cumulants relate to each other:

$$[u_i, u_j] = \frac{[n_i, n_j]}{h^2} + \frac{\delta_{ij}\langle u_i \rangle}{h} - \delta_{ij}\langle u_i \rangle. \tag{5.39}$$

Using Eqs. (5.32) and (5.39), we deduce that the evolution of $[u_i, u_j]$ is given by:

$$\frac{d[u_i, u_j]}{dt} = 2\lambda \left[1 - \phi_i - \phi_j\right] [u_i, u_j] + \frac{\delta_{ij}}{h}\left(\lambda\Omega\frac{\phi_i}{h} + 2\lambda\Omega\frac{\phi_i^2}{h}\right) + \delta_{ij}\left(\lambda\Omega\frac{\phi_i}{h} - 3\lambda\Omega\frac{\phi_i^2}{h}\right). \tag{5.40}$$

On the other hand using a procedure similar to that used to analyse the effect of the diffusion term in the mean-field equation, it can be shown [215, 216] that the coarse-grained factorial cumulant, $w(x_i, x_j, t) = \lim_{h\to 0}[u_i, u_j](t)$, due to the diffusion satisfies the equation:

$$\frac{\partial w}{\partial t}(x,y,t) = D\Delta_{x,y}w(x,y,t) \tag{5.41}$$

with Neumann boundary conditions. Using Eqs. (5.40) and (5.41) and upscaling the coarse-grained factorial cumulant for the initial conditions we derive the desired evolution equation for the coarse-grained factorial cumulant. □

The estimated variance from the coarse-grained factorial cumulant is given by:

$$var(x,t) = w(x,x) + u(x,t) \tag{5.42}$$

## 5.3 Numerical study of the discretisation of the coarse-grained equations

In order to implement our hybrid model, we must discretise the mean-field equation (Eq. (5.23)) and the coarse-grained factorial cumulant (Eq. (5.36)). In this section we introduce the discretisation schemes for solving both PDEs and analyse their convergence properties.

### 5.3.1 Discretisation schemes

**Mean-field equation**

We focus first on the discretisation of the PDE that describes the evolution of the mean-field concentration, $u(x,t)$, (Eq. (5.23)). We discretise this PDE via the Rothe method [94], discretising in time using the implicit Euler scheme and in space with linear finite elements.

We consider a discretisation of the time and the space domain. Let us consider a partition of the interval $[0,T]$ into $M$ subintervals of length $h_t = \frac{T}{M}$ where $t_n = nh_t$ for $n = 0, 1, \ldots, M$. We denote the space domain $D := (0, L)$. We consider a mesh $\mathcal{T}_h$

of $D$, namely, $D = \bigcup_{j=1}^{N} K_j$ where we refer to $K_j$ as an element of $\mathcal{T}_h$. We assume the mesh is regular (see Sec. 2.3 for more details on regular meshes). We define $h_x$ as the maximum diameter of the elements of the mesh:

$$h_x := \max_{K \in \mathcal{T}_h} h_K.$$

We denote the solution of Eq. (5.23) at $t_n$ by $u^n(x) := u(x, t_n)$. Applying the implicit Euler scheme to the PDE in Eq. (5.23) we obtain:

$$\frac{u^n(x) - u^{n-1}(x)}{h_t} = \left[ D\Delta u^n(x) + \lambda u^n(x) \left( 1 - \frac{u^n(x)}{K} \right) \right]. \tag{5.43}$$

We now discretise Eq. (5.43) by finite elements. We assume that the solution of Eq. (5.43) belongs to $H^1(D)$. To approximate the solution in $H^1(D)$ we consider the finite-dimensional space $V_h$, (see Sec. 2.3 for a description of the functional spaces $H^1(D)$ and $V_h$). We then formulate the variational problem in $V_h$. The weak formulation at each $t_n$ ($n \geq 1$) involves of determining $u^n(x) \in V_h$ such that

$$\left( \frac{u^n(x) - u^{n-1}(x)}{h_t}, v_h(x) \right) + T^u(u^n(x), v_h(x)) = F^u(u^n(x), v_h(x)) \quad \forall v_h(x) \in V_h, \tag{5.44}$$

where

$$T^u(u^n(x), v_h(x)) = D(\nabla u^n(x), \nabla v_h(x))$$

and

$$F^u(u^n(x), v_h(x)) = \lambda(u^n(x), v_h(x)) + \frac{\lambda}{K} \left( (u^n(x))^2, v_h(x) \right).$$

To derive the weak formulation we have exploited the fact that $u^n(x)$ satisfies the relevant Neumann boundary conditions at each time step (see Eq. (5.23) ).

The weak formulation (5.44) is nonlinear so that we can use the Newton method to solve it. We first reformulate the problem so we apply the Newton method. The weak formulation (5.44) is equivalent to finding a function $u^n(x) \in V_h$ at each $t_n$ ($n \geq 1$) such that

$$G(u^n(x)) = 0 \tag{5.45}$$

where

$$\begin{aligned} G(u^n(x)) =& (u^n(x), v_h(x)) + Dh_t(\nabla u^n(x), \nabla v_h(x)) - \lambda h_t(u^n(x), v_h(x)) \\ &+ \frac{\lambda h_t}{K} \left( (u^n(x))^2, v_h(x) \right) - (u^{n-1}, v_h(x)) \quad \forall v_h(x) \in V_h. \end{aligned} \tag{5.46}$$

We consider one of the simplest versions of the Newton method to solve Eq. (5.45) (see [63] for an overview of the different Newton method implementations). We consider the following fixed point iteration:

$$v^{m+1}(x) = v^m(x) + \delta v^m(x) \tag{5.47}$$

where $\delta v^m(x)$ solves the linear problem

$$G'(v^m(x))(\delta v^m(x)) = -G(v^m(x)) \tag{5.48}$$

where $G'(u(x))(\delta u(x))$ is the Gateaux derivative of $G$ in the direction $\delta u(x)$, $v^m(x)$ is the $m$-th Newton's iteration and $v^0(x)$ is the initial estimate of $u^n(x)$.

The Gateaux derivative of the nonlinear functional $G$ (Eq. (5.46)) in the direction $\delta u(x)$ is:

$$\begin{aligned} G'(u(x)(\delta u(x)) =& (\delta u(x), v_h(x)) + Dh_t(\nabla \delta u(x), \nabla v_h(x)) - \lambda h_t(\delta u(x), v_h(x)) \\ &+ \frac{2\lambda h_t}{K}(u(x)\delta u(x), v_h(x)) \quad \forall v_h(x) \in V_h. \end{aligned} \tag{5.49}$$

If the initial Newton iteration is sufficiently close to the function, $u^n(x)$, the Newton iterates can be shown to converge to the function $u^n(x)$ [63]. To apply the Newton method we set a tolerance error, $TOL$, and a maximum number of Newton iterates, $N_{max}$.

At each iteration of the Newton method, we approximate the solution of Eq. (5.48), $\delta v^m$, by its Ritz projection in the space $V_h$:

$$\delta v^m(x) \approx \delta v_h^m(x). \tag{5.50}$$

Let $\{\varphi_0(x), \dots, \varphi_{N_h-1}(x)\}$ denote the nodal basis functions of $V_h$. Then $\delta v_h^m(x)$ is given by:

$$\delta v_h^m(x) = \sum_{j=0}^{N_h-1} \delta V_j^m \varphi_j(x),$$

where $V_j^m$ are real constants for $0 \leq j \leq N_h - 1$.

The variational formulation needs to be satisfied for all $v_h(x) \in V_h$. A necessary and sufficient condition for this to happen is that the basis functions satisfies the variational formulation. For each basis function, $\varphi_i(x)$, the relevant variational problem is:

$$G_i'(v^m)(\delta v_h^m) = -G_i(v^m) \quad \forall i = 0, \dots, N_h - 1$$

where $G_i(u) := G(\varphi_i)(u)$. By considering the variational formulation for all the basis functions, the variational problem reduces to solving a linear system for the coefficients $V_j^m$ for $0 \leq j \leq N_h - 1$. In matrix notation the linear system that needs to be solved is:

$$\boldsymbol{A}^m \boldsymbol{\delta V}^m = \boldsymbol{b}^m,$$

where $\boldsymbol{A}^n$ is an $N_h \times N_h$ matrix whose entries are given by:

$$\boldsymbol{A}_{i,j}^m := (1 - \lambda h_t)(\varphi_i, \varphi_j) + \frac{2\lambda h_t}{K}(v^m \varphi_i, \varphi_j) + Dh_t(\nabla \varphi_i, \nabla \varphi_j),$$

$\boldsymbol{b}^m$ is a $N_h$-dimensional vector with entries given by:

$$\begin{aligned} \boldsymbol{b}_i^m :=& -(v^m, \varphi_i) + Dh_t(\nabla v^m, \nabla \varphi_i) - \lambda h_t(v^m, \varphi_i) \\ &+ \frac{\lambda h_t}{K}\left((v^m)^2, \varphi_i\right) - (u^{n-1}, \varphi_i). \end{aligned} \tag{5.51}$$

and $\boldsymbol{\delta V}^n$ is the $N_h$-dimensional vector solution.

The solution update that uses the discretisation outlined above is described in Algorithm 5.1.

**Algorithm 5.1** Solution update for the mean-field equation

```
Initialise the finite element solution using the initial condition $u^0$. Set $n = 0$.
for $n = 1$ to $n = M - 1$ do
    Update rule for $u^{n-1} \to u^n$ : Consider as the initial iterate of the Newton method: $v_0 = u^{n-1}$. Let $b^0$ denote the $n$-dimensional residual vector, defined by Eq. (5.51).
    while $\|b^m\|_2 > TOL$ and $m < N_{max}$. do
        Compute the Newton update $\delta v^m(x)$ by solving the linear system (5.3.1).
        Consider the Newton iteration $v^{m+1}(x) = v^m(x) + \delta v^m(x)$ and update $m \to m+1$.
    end while
    Update the solution at the $n$-th time point with the final Newton iteration: $u^n(x) = v^{m+1}(x)$.
end for
```

### Coarse-grained factorial cumulant equation

We now discretise the evolution equation of the coarse-grained factorial cumulant (Eq. (5.36)). We first determine how to deal with the indicator and the delta terms. Following [67, 138], we consider the following regularised versions of $\delta(x)$, the delta function, and $\mathbb{1}_{x=x^*}(x)$, the indicator function:

$$\delta(x) \approx \delta^\epsilon(x) = \begin{cases} \frac{1}{\epsilon}\left(1 - \left|\frac{x}{\epsilon}\right|\right), & \text{if } |x| \le \epsilon, \\ 0, & \text{otherwise.} \end{cases} \tag{5.52}$$

$$\mathbb{1}_{x=x^*}(x) \approx \mathbb{1}^\epsilon_{x^*}(x) = \begin{cases} x, & \text{if } |x - x^*| \le \epsilon, \\ 0, & \text{otherwise.} \end{cases} \tag{5.53}$$

where $\epsilon > 0$.

When discretising Eq. (5.36), we use the same time mesh as that used for the mean-field equation. We discretise Eq. (5.36) via the Rothe method, first discretising in time with an implicit Euler scheme and in space with linear finite elements.

We denote the solution of Eq. (5.36) at $t_n$ as $w^n(x,y) := w(x,y,t_n)$. By applying the implicit Euler scheme to Eq. (5.36) while considering the regularised delta and the indicator functions we obtain:

$$\begin{aligned} \frac{w^n(x,y) - w^{n-1}(x,y)}{h_t} =& D\Delta w^n(x,y) + 2\lambda \left[1 - \frac{u^n(x)}{K} - \frac{u^n(y)}{K}\right] w^n(x,y) \\ &+ \lambda\delta^\epsilon(x-y)\left[u^n(x) + \frac{2[u^n(x)]^2}{K}\right] \\ &+ \lambda\mathbb{1}^\epsilon_{x-y=0}(x)\left[u^n(x) - \frac{3[u^n(x)]^2}{K}\right]. \end{aligned} \tag{5.54}$$

To discretise Eq. (5.54) by finite elements, we first assume that its solution is defined in $H^1(D \times D)$. To approximate the solution in $H^1(D \times D)$ we consider the finite-dimensional space $V_h$ corresponding to the domain $D \times D$ (see Sec. 2.3 for a description of the functional spaces $H^1(D \times D)$ and $V_h$). The variational problem at each $t_n$ ($n \geq 1$) consists of determining $w^n(x,y) \in V_h$ that satisfies:

$$\left(\frac{w^n(x,y) - w^{n-1}(x,y)}{h_t}, v_h(x,y)\right) + T^w(w^n(x,y), v_h(x,y)) = F^w(w^n(x,y), v_h(x,y)), \tag{5.55}$$

$\forall v_h(x,y) \in V_h$,
where

$$T^w(u^n(x,y), w_h(x,y)) = D\left(\nabla w^n(x,y), \nabla v_h(x,y)\right), \tag{5.56}$$

and

$$\begin{aligned} F^w(w^n(x,y), v_h(x,y)) =& 2\lambda \left(\left[1 - \frac{u^n(x)}{K} - \frac{u^n(y)}{K}\right] w^n(x,y), v_h(x,y)\right) \\ &+ \lambda \left(\delta^\epsilon(x-y)\left[u^n(x) + \frac{2[u^n(x)]^2}{K}\right], v_h(x,y)\right) \\ &+ \lambda \left(\mathbb{1}^\epsilon_{x-y=0}(x,y)\left[u^n(x) - \frac{3[u^n(x)]^2}{K}\right], v_h(x,y)\right). \end{aligned}$$

By considering the nodal basis functions of the finite element space $V_h$, the weak formulation (5.55) reduces to a linear system.

*The linear system of equations for the basis functions $n = 0, \ldots, N_h$.*

We first approximate the solution of Eq. (5.55) by its Ritz projection in the space $V_h$:

$$w^n(x,y) \approx w_h^n(x,y) \tag{5.57}$$

Let $\{\varphi_0(x,y) \ldots, \varphi_{N_h-1}(x,y)\}$ denote the nodal basis functions of $V_h$, then $w_h^n(x,y)$ is given by:

$$w_h^n(x,y) = \sum_{j=0}^{N_h-1} W_j^n \varphi_j(x,y). \tag{5.58}$$

where $W_j^n$ are real constants for $0 \leq j \leq N_h - 1$.

The variational formulation (5.55) must be satisfied for all $v_h(x,y) \in V_h$. A necessary and sufficient condition for this to happen is that the formulation is satisfied by the basis functions. The variational formulation (5.55) for a basis function is:

$$\left(\frac{w^n(x,y) - w^{n-1}(x,y)}{h_t}, \varphi_i\right) + T^w(w^n(x,y), \varphi_i(x,y)) = F^w(w^n(x,y), \varphi_i(x,y)) \tag{5.59}$$

for all $i = 0, \ldots, N_h$.

By considering the variational problem (5.59) for all basis functions we reduce the problem at each $t_n$ ($n \geq 1$) solving a finite-dimensional linear system for the coefficients $W_j^n$, $0 \leq j \leq N_h - 1$. The matrix formulation of the linear system is the following:

$$\boldsymbol{M}^n \boldsymbol{W}^n = \boldsymbol{Y}^n \tag{5.60}$$

where $\boldsymbol{M}^n$ is an $N_h \times N_h$ matrix whose entries are given by:

$$\begin{aligned}\boldsymbol{M}^n_{i,j} &= (1-2\lambda h_t)\left(\varphi_i(x,y),\varphi_j(x,y)\right) + Dh_t\left(\nabla\varphi_i(x,y),\nabla\varphi_j(x,y)\right)\\ &+ 2\lambda h_t\left(\varphi_i(x,y),\left[\frac{u^n(x)}{K}+\frac{u^n(y)}{K}\right]\varphi_j(x,y)\right),\end{aligned}$$

$\boldsymbol{Y}^n$ is a $N_h$-dimensional vector with entries given by:

$$\begin{aligned}\boldsymbol{Y}^n_i &= \left(\varphi_i(x,y),w^n(x,y)\right) + \lambda h_t\left(\varphi_i(x,y),\delta^{\epsilon}(x-y)\left[u^n(x)+\frac{2[u^n(x)]^2}{K}\right]\varphi_j(x,y)\right)\\ &+ \lambda h_t\left(\varphi_i(x,y),\mathbb{1}^{\epsilon}_{x-y=0}(x,y)\left[u^n(x)-\frac{3[u^n(x)]^2}{K}\right]\varphi_j(x,y)\right),\end{aligned}$$

and $\boldsymbol{W}^n$ is the $N_h$-dimensional vector solution.

We note that in order to solve the equation for the coarse-grained factorial cumulant, we must first determine the values of the mean-field equation.

The solution update of the coarse-grained factorial cumulant evolution equation that use the discretisation scheme outlined above is described in Algorithm 5.2.

**Algorithm 5.2** Solution update for the coarse-grained factorial cumulant

```
1: Determine a time mesh to solve the mean-field equation (Eq. (5.23)) and the coarse-grained factorial cumulant equation (Eq. (5.36)).
2: Initialise the finite element solution using the initial condition $-\mathbb{1}^{\epsilon}_{x-y=0}u^0$. Set $n = 0$.
3: Solve the mean-field equation using algorithm 5.1 and determine $w^n$ for all $1 \leq n \leq M$.
4: for $n = 1$ to $n = M-1$ do
5:     Update rule $w^{n-1} \to w^n$ : Solve the linear system (5.60) to obtain $w^n$.
6: end for
```

*Locally refined mesh*

Since we are only interested in solution values on the diagonal of the domain, we refine locally the spatial mesh in a region close to the diagonal, so it has smaller elements close to the diagonal and coarser elements far away from it. We refine the mesh in this way to obtain an accurate solution on the diagonal and still have an efficient implementation. The procedure that we use to refine the spatial mesh to solve the coarse-grained factorial cumulant is described in Algorithm 5.3 and in Figure 5.1 we illustrate its implementation effect.

**Algorithm 5.3** Mesh refinement algorithm

```
Set an initial mesh with nodes on the diagonal and refine all cells $\chi_1$ times.
Set distance $\chi_2$ and consider
    $$\Phi = \{K \in \mathcal{T}_h \text{ such that } \|c_K - l_D\|_2 < \chi_2\}$$
    where $l_D$ denotes the points of the diagonal and $c_K$ denotes the center of the element $K$.
Set $\chi_3$ : the number of times to refine elements in $\Phi$.
for $n = 1$ to $\chi_3$ do
    Refine the elements in $\Phi$.
end for
Refine all elements in the mesh $\chi_4$ times.
```

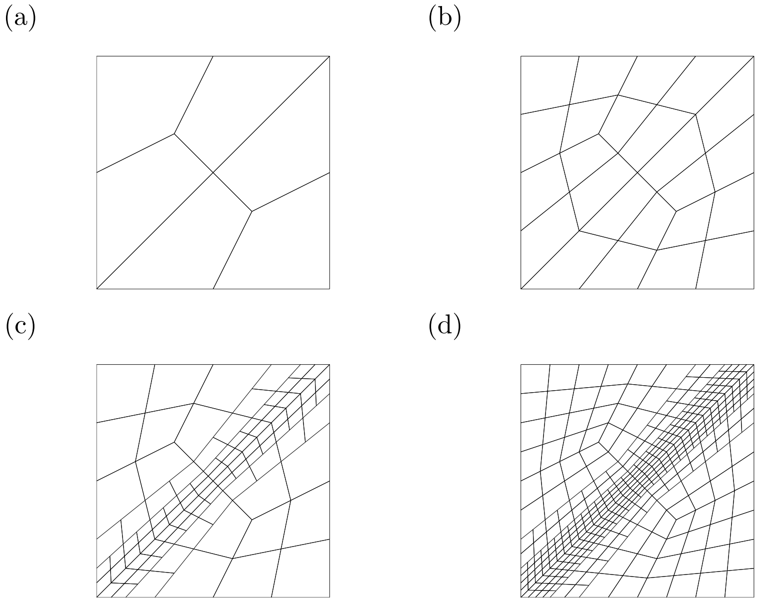


**Figure 5.1:** Series of plots depicting the effect of the steps of the local refinement algorithm 5.3. a) Initial mesh. b) Mesh after refining $\chi_1 = 1$ times. c) Mesh after considering refining $\chi_3$=2 times all the elements that are at a distance $\chi_2$=25 from the diagonal. d) Final mesh after $\chi_4 = 1$ refinements.

In practice we consider a more refined mesh than the final mesh in Figure 5.1. The default parameter values for the local refinement are described in Table 5.1.

| Parameter | Description | Value |
|---|---|---|
| $\chi_1$ | Initial refinement level | 2 |
| $\chi_2$ | Distance from the diagonal that defines the region $\Phi$ | 25 |
| $\chi_3$ | Number of times the local refinement is performed | 2 |
| $\chi_4$ | Refinements performed after local refinement | 4 |

**Table 5.1:** Default parameters value for the local mesh refinement of the domain $(0, L) \times (0, L)$ when solving Eq. (5.36).

### 5.3.2 Convergence analysis

We now analyse the convergence of the discretisation schemes (5.1) and (5.2) that we use to solve the evolution equations of the mean-field and coarse-grained factorial cumulant (Eqs. (5.23) and (5.36)). As for the convergence analysis of the discretisation schemes of the age-structured models, we consider the method of manufactured solutions (see Section 2.3 for a description of the method) to determine the observed convergence rates.

#### Mean-field equation

We first analyse the discretisation scheme for the mean-field equation. For arbitrary initial and boundary conditions, no explicit solutions are known [2]. However, for certain initial and boundary conditions, explicit solutions exist [2]. We consider a known explicit solution of the FKPP equation (see [2]) and modify it for our specific boundary conditions and domain. In [2], the authors considered the non-dimensionalised FKPP equation:

$$\frac{\partial u}{\partial t} = \Delta u + u(1-u), \tag{5.61}$$

for $(x,t) \in \mathbb{R} \times (0,T)$ with boundary conditions:

$$\lim_{x\to-\infty} u(x) = 0 \quad \text{and} \quad \lim_{x\to\infty} u(x) = 1. \tag{5.62}$$

The authors showed that Eq. (5.61), with boundary conditions (5.62), has the following explicit travelling wave solution:

$$u(x,t) = \left[1 + \exp\left(\frac{x-ct}{\sqrt{6}}\right)\right]^{-2} \tag{5.63}$$

where $c = \frac{5}{\sqrt{6}}$.

To use this particular solution in the method of manufactured solutions to test our discretisation scheme. We note that our PDE (5.23) can be non-dimensionalised to Eq. (5.61) via the following change of variables:

$$u = K\hat{u}, \quad x = \sqrt{\frac{D}{\lambda}}\hat{x}, \quad \text{and} \quad t = \frac{1}{\lambda}\hat{t}. \tag{5.64}$$

Then $\hat{u}(\hat{x},\hat{t})$ satisfies the following system:

$$\frac{\partial \hat{u}}{\partial \hat{t}} = \Delta_{\hat{x}\hat{x}}\hat{u} + \hat{u}(1-\hat{u}). \tag{5.65}$$

Let us assume that $\hat{u}(\hat{x},\hat{t})$ is given by (5.63). After dimensionalising again the equation, we deduce that a solution of Eq. (5.23) is given by:

$$u(x,t) = K\hat{u}\left(\frac{x}{\chi},\frac{t}{\tau}\right) = K\left[1 + \exp\left(\sqrt{\frac{\lambda}{6D}}x - \frac{5\lambda}{6}t\right)\right]^{-2}. \tag{5.66}$$

For the travelling wave to be within the domain $(0, L)$, we shift the solution (5.66) in the variable $t$. The explicit solution that we consider for the method of manufactured solutions is then:

$$z(x,t) = K\left[1 + \exp\left(\sqrt{\frac{\lambda}{6D}}x - \frac{5\lambda}{6}t + 20\right)\right]^{-2}. \tag{5.67}$$

It is straightforward to show that $z(x,t)$ satisfies the following PDE:

$$\begin{cases} \frac{\partial u}{\partial t} = D\Delta u + \lambda u\left(1 - \frac{u}{K}\right) & x \in (0,L), t \in (0,T), \\ \frac{\partial u}{\partial x} = \frac{\partial z}{\partial x} & t \in (0,T) \\ u(x,0) = z(x,0) & x \in [0,L]. \end{cases} \tag{5.68}$$

We analyse the performance of the discretisation scheme (5.1) by solving Eq. (5.68) using our discretisation scheme and calculating the error between the numerical solution and the exact solution, $z(x,t)$, at each time step. Let $h_x$, $h_t$ denote the space and time spacing, respectively. For a given simulation we consider the following error measure in space and time:

$$e^z(h_x, h_t) = \|u_h(x,t) - z(x,t)\|_{L^2(D\times(0,T))}. \tag{5.69}$$

We compute the error measure (5.69) by approximating the time integral by a trapezoidal rule and the integration in the space domain $\Omega$ is approximated by a Gaussian-Legendre quadrature.

We use the parameter values stated in Table 5.2 when performing the convergence analysis. We consider $\hat{u}(x,t)$ since a solution of the PDE with $K = 1$ can be transformed to a solution of the PDE with another value for $K$ by rescaling $\hat{u}(x,t) = \frac{u(x)}{K}$.

| Parameter | Description | Value |
|---|---|---|
| $\lambda$ | Proliferation | 1 |
| $K$ | Carrying capacity | 1 |
| $L$ | Domain length | 100 |

**Table 5.2:** Summary of the parameter values for the convergence analysis of the discretisation schemes (5.1) and (5.2) that we use to solve the evolution equations of the mean-field and coarse-grained factorial cumulant (Eqs. (5.23) and (5.36)).

It has been shown that for a modifed FKPP equation solved using the backward Euler method and linear Lagrange finite elements under certain regularity assumptions [59] the error measure satisfies the following inequality:

$$0 \le e^z(h_x, h, t) \le C_z(h_x^2 + h_t). \tag{5.70}$$

for a positive real scalar $C_z$.

*Space discretisation convergence analysis*

We estimate the rate of convergence of the space discretisation by refining the spatial mesh and calculating the error measure $e^z(h_x, h_t)$ at each refinement cycle.

To estimate the space convergence rate, by fixing an end time $T$ and considering a time step, $h_t$, small enough so that $e^z(h_x, h_t)$ is dominated by the spatial error so that the time error can be neglected. We then consider an initial mesh and refine it sequentially: given $h_x$ the spacing of the initial mesh, for each $n$-th refinement cycle we calculate a mesh size of $h_x/2^n$. We calculate the following quantity between each sequential refinement cycle:

$$r_x^z(h_x, h_t) = \frac{\log\left(e^z(h_x, h_t)/e^z(\frac{h_x}{2}, h_t)\right)}{\log 2}. \tag{5.71}$$

The observed convergence rate of the space discretisation is $r_x^z$ as $h_x \to 0$. We will show numerically that the space discretisation convergence rate is of second order by showing that $r_x^z \approx 2$.

In Table 5.3 we summarise the results when analysing the error measure and performing space refinement. We consider an end time of $T = 0.01$ and a time step of 0.001. We consider the spatial mesh to have a mesh size of $200/2^7 \approx 1.56$ and then refine it while calculating the observed convergence rate. In Table 5.3 each row represents a different refinement cycle where we include the time mesh size $h_t$, the spatial mesh size $h_x$, the error measure $e^z(h_x, h_t)$ and the observed space convergence rate $r_x^z(h_x, h_t)$. The observed convergence rate matches the expected one.

| $h_t$ | $h_x$ | $e^z(h_x, h_t)$ | $r_x^z(h_x, h_t)$ |
|---|---|---|---|
| 1.000e-03 | 1.562e+00 | 1.258e-03 | - |
| 1.000e-03 | 7.812e-01 | 3.167e-04 | 1.99 |
| 1.000e-03 | 3.906e-01 | 7.937e-05 | 2.00 |
| 1.000e-03 | 1.953e-01 | 1.990e-05 | 2.00 |
| 1.000e-03 | 9.766e-02 | 5.025e-06 | 1.99 |
| 1.000e-03 | 4.883e-02 | 1.305e-06 | 1.94 |
| 1.000e-03 | 2.441e-02 | 3.769e-07 | 1.79 |

**Table 5.3:** Estimated spatial convergence rates of the discretisation scheme in Algorithm 5.1 when solving Eq. (5.68).

Given the results in Table 5.3, we fix the spatial mesh size $h_x = 200/2^{13} \approx 0.02441$ when solving Eq. (5.68).

*Time discretisation convergence analysis*

For the time convergence analysis we consider an end time of $T = 1$. On the first refinement cycle we fix the initial time step $h_t = 0.5$ and on the $n$-th refinement cycle we consider a time step of $h_t/2^n$. In Table 5.4 we summarise the results of the time refinement convergence analysis. Each row represents a different time refinement cycle where we include the time mesh size $h_t$, the spatial mesh size $h_x$, the error measure $e^z(h_x, h_t)$ and the observed space convergence rate $r_x^z(h_x, h_t)$. The observed convergence rate matches that predicted by (5.70).

| $h_t$ | $h_x$ | $e^z(h_x,h_t)$ | $r_x^z(h_x,h_t)$ |
|---|---|---|---|
| 5.000e-01 | 2.441e-02 | 4.910e-02 | - |
| 2.500e-01 | 2.441e-02 | 1.350e-02 | 1.86 |
| 1.250e-01 | 2.441e-02 | 4.846e-03 | 1.48 |
| 6.250e-02 | 2.441e-02 | 2.026e-03 | 1.26 |
| 3.125e-02 | 2.441e-02 | 9.235e-04 | 1.13 |
| 1.562e-02 | 2.441e-02 | 4.412e-04 | 1.07 |
| 7.812e-03 | 2.441e-02 | 2.164e-04 | 1.03 |

**Table 5.4:** Estimated time convergence rates of the discretisation scheme 5.1 when solving Eq. (5.68).

### Coarse-grained factorial cumulant

To analyse the discretisation of the evolution equation of the factorial cumulant (Eq. (5.36)), we consider the following function:

$$G(x,y,t) = -K\left(\kappa_0 t + 1\right)\exp\left(-\left[\frac{x-y}{\kappa_1}\right]^2\right) \tag{5.72}$$

where $\kappa_0$ and $\kappa_1$ are real scalars.

Following the method of manufactured solutions, we modify the PDE (5.36) so that $G(x,y,t)$ is an explicit solution of the modified system.

The modified system is the following:

$$\begin{aligned}\frac{\partial w}{\partial t}(x,y,t) =& D\Delta w(x,y,t) + 2\lambda\left[1 - \frac{u(x,t)}{K} - \frac{u(y,t)}{K}\right] w(x,y,t) \\ &+ \lambda\delta(x-y)\left[u(x,t) + \frac{2u^2(x,t)}{K}\right] + \lambda\mathbb{1}_{x-y=0}(x,y)\left[u - \frac{3u^2}{K}\right] \\ &+ F(x,y,t) - \lambda\delta(x-y)\left[z(x,t) + \frac{2z^2(x,t)}{K}\right] - \lambda\mathbb{1}_{x-y=0}(x)\left[z - \frac{3z^2}{K}\right]\end{aligned} \tag{5.73}$$

where

$$F(x,y,t) = \frac{\partial G(x,y,t)}{\partial t} - D\Delta G(x,y,t) - 2\lambda\left[1 - \frac{z(x,t)}{K} - \frac{z(y,t)}{K}\right] G(x,y,t), \tag{5.74}$$

and $u(x,t)$ satisfies the problem given by equation (5.68) for which $z(x,t)$ is the explicit solution.

The full modified problem consists of the PDE (5.73), the boundary condition,

$$\frac{\partial w(x,y,t)}{\partial \nu} = \frac{\partial G(x,y,t)}{\partial \nu} \quad \forall x \in \Omega^2 \times [0,T]. \tag{5.75}$$

and the initial condition,

$$w(x,y,0) = G(x,y,0) \quad \forall (x,y) \in \Omega^2. \tag{5.76}$$

In order to solve Eq. (5.73) we need to first solve the equation for the mean-field concentration, $u(x,t)$, Eq. (5.68). To analyse the convergence properties of the Algorithm 5.2 when solving Eq. (5.73), we consider the values of the explicit solution for Eq. (5.68), $z(x,t)$, in the coefficients that correspond to $u(x,t)$ in Eq. (5.73). For a given simulation and refinement level we consider the following error measure in space and time:

$$e^G(h_x, h_t) = \|w_h(x,t) - G(x,t)\|_{L^2(\Omega\times(0,T))}\,. \tag{5.77}$$

*Space discretisation convergence analysis*

For the convergence analysis we consider the parameter values in Table 5.2 and use the parameter values in Table 5.1 to create a locally refined mesh. The only parameter values that remain to be fixed are $\kappa_0 = -0.005$ and $\kappa_1 = 0.4$. When performing the space refinement analysis, we vary $\chi_4$.

In Table 5.5 we summarise the results of the space refinement analysis. We consider an end time $T = 0.01$ and a time step of 0.001. Each row represents a different space refinement cycle where we include the time mesh size, $h_t$, the number of elements, the error measure, $e(h_x, h_t)$, and the observed space convergence rate, $r_x^w(h_x, h_t)$. We do not consider the spatial mesh size since the mesh is locally refined and the elements' size changes through the mesh. We observe that as the refinement levels increase, the space convergence rate converges to the theoretical one.

| $h_t$ | Number of elements | $e^G(h_x, h_t)$ | $r_x^G(h_x, h_t)$ |
|---|---|---|---|
| 1.000e-03 | 696 | 3.811e-02 | - |
| 1.000e-03 | 4097 | 9.919e-03 | 1.94 |
| 1.000e-03 | 11136 | 2.512e-03 | 1.98 |
| 1.000e-03 | 44544 | 6.355e-04 | 1.98 |
| 1.000e-03 | 178176 | 1.761e-04 | 1.85 |
| 1.000e-03 | 712704 | 8.483e-05 | 1.05 |

**Table 5.5:** Estimated space convergence rates of the discretisation scheme defined by Algorithm 5.2 when solving Eq. (5.73).

*Time discretisation convergence analysis*

In Table 5.6 we summarise the results of the time refinement analysis and calculate the observed time convergence rate. The parameters that we use to refine the spatial mesh are listed in Table 5.1. For the time refinement analysis we fix an end time $T = 0.8$ and initial time step $h_t = 0.4$. Each row describes the results of a refinement of the time mesh. We observe that although the observed time convergence rate is close to 1 for the first refinement levels, as the refinement levels increase, the observed convergence rates decrease. This is expected when the space error is not negligible. However, from Table 5.5, we know that this threshold has not been reached so the discretisation scheme has a lower convergence error than the optimal theoretical one.

| $h_t$ | Number of elements | $e^G(h_x, h_t)$ | $r_t^G(h_x, h_t)$ |
|---|---|---|---|
| 4.000e-01 | 44544 | 1.556e+00 | - |
| 2.000e-01 | 44544 | 3.620e-01 | 2.10 |
| 1.000e-01 | 44544 | 1.464e-01 | 1.31 |
| 5.000e-02 | 44544 | 7.251e-02 | 1.01 |
| 2.500e-02 | 44544 | 3.904e-02 | 0.89 |
| 1.250e-02 | 44544 | 2.868e-02 | 0.44 |

**Table 5.6:** Estimated time convergence rates of the discretisation scheme defined by Algorithm 5.2 when solving Eq. (5.73).

## 5.4 Analysis of the stochastic and mean-field behaviour

In this section we analyse how well the coarse-grained equations of the mean-field (Eq. (5.23)) and the estimated variance derived from the coarse-grained factorial cumulant (Eq. (5.36)) approximate the first and second moments of the stochastic model. We refer to the mean-field equation and the estimated variance derived from the coarse-grained factorial cumulant as the deterministic model/equations for the mean and the variance of the reaction-diffusion system.

We first describe the reference dynamics of the stochastic model which we use as a benchmark and for which we introduce the quantities that we will analyse when comparing the hybrid and stochastic models.

### 5.4.1 Reference dynamics

We consider typical dimensionless parameter values that have been used to describe a hybrid algorithm for this stochastic model [149]. The reference values for the parameters in the stochastic model are given in Table 5.7. We fix $\lambda = 1$ and $D = 1$ equal to 1 since it has been observed that for these parameter values, the model exhibits travelling wave solutions for arbitrary values of $K$ [149]. We consider a wave-like initial condition since we are not interested in wave formation in the stochastic model but in the travelling wave behaviour. We consider the following initial condition for the mean-field equation (Eq. (5.23)):

$$v(x) = -\frac{K}{2}\tanh(x - L/4) + \frac{K}{2}. \tag{5.78}$$

The initial condition for the stochastic model is given by:

$$V(x) = round(hv(x)) \tag{5.79}$$

where $round(y)$ rounds $y$ to the nearest integer.

To analyse the dynamics of the deterministic and stochastic models, we compare the cell density, $u(x,t)$, for the deterministic model and the number of molecules in each compartment divided by the compartment size, $U(x,t)/h$, for the stochastic model.

| Parameter | Description | Value |
|---|---|---|
| $\lambda$ | Proliferation rate | 1 |
| $K$ | Carrying capacity | 400 |
| $h$ | Compartment size of the space domain | 0.8 |
| $T$ | End time of simulation | 50 |
| $h_t$ | Time step for the finite element discretisation | 0.05 |
| $L$ | Domain length | 200 |

**Table 5.7:** Summary of parameters for the stochastic FKPP model

Using the parameter values in Table 5.7 and the initial condition for the stochastic model (5.79), we perform 500 simulations of the stochastic model using the Gillespie algorithm (Algorithm 2.1). In Figure 5.2 we plot the mean and standard deviation of the cell density of the stochastic model and the cell density of the deterministic model at six time points. As expected, at time $t = 0$, the mean cell density from the stochastic simulations coincides with the initial condition of the mean-field equation and the standard deviation is zero. We observe that both models have travelling wave solutions with positive velocity. As expected, the mean-field travelling wave moves faster than the stochastic travelling wave solution. For the stochastic model when $t > 0$, the standard deviation is of size 20 for the compartments in which the cell density is close to the carrying capacity, while it is zero for the compartments with no molecules.

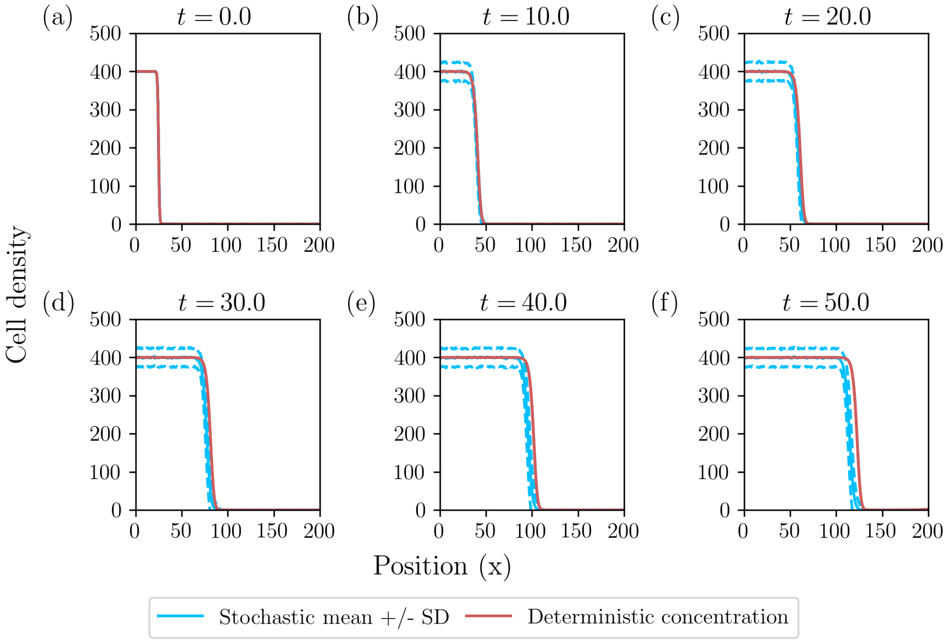


**Figure 5.2:** Series of plots showing the typical behaviour of the mean and standard deviation of the cell density of the stochastic FKPP system (Eq. (5.9)) and the cell density evolution given by the mean-field equation (Eq. (5.23)) at different time points. We consider the parameter values as per Table 5.7 and the initial conditions (5.78) and (5.79) for the deterministic and the stochastic models, respectively.

For a given simulation of the stochastic model (Eq. (5.9)) and the solution of the mean-field equation (Eq. (5.23)), we refer to the smallest spatial position for which the cell density is below the threshold value:

$$0.1K \tag{5.80}$$

as the deterministic and stochastic *interface position*, respectively. In both cases, the interface divides the domain into two subdomains: one where the cell density is close to the carrying capacity, and the other where the number of molecules is much smaller than $K$. In Figure 5.3 (a) we observe the evolution of the mean and standard deviation of the stochastic interface position and the deterministic interface position over time.

To quantify the speed at which the deterministic and stochastic travelling waves move, we consider a linear approximation to the interface position evolution. We refer to the slope of this linear approximation as the *velocity of the travelling wave*. In Figure 5.3 (b) we plot a histogram of the velocities of 500 stochastic simulations. The stochastic velocities have a mean value of 1.85 and a standard deviation of 0.06. We plot a vertical red line at the value of the deterministic travelling wave velocity. Most of the stochastic velocities are smaller, which confirms that the stochastic travelling waves are generally slower than the deterministic travelling wave.

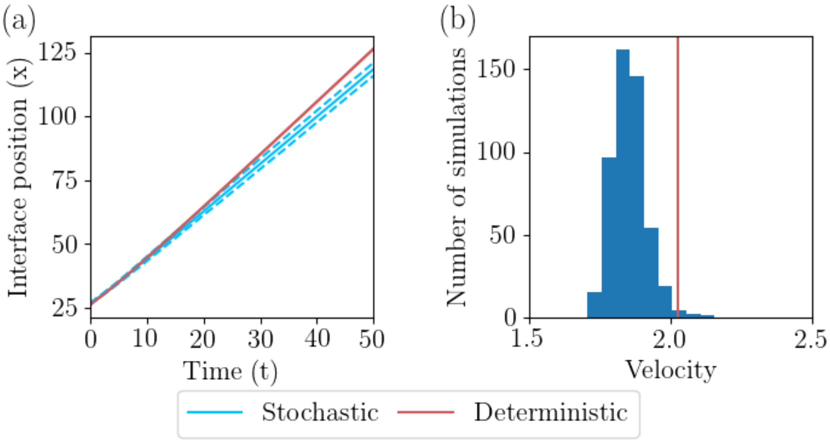


**Figure 5.3:** (a) Evolution of the deterministic interface position and the mean and standard deviation of the stochastic interface position over time. (b) Histogram of the velocities of the stochastic simulations. A red vertical line is plotted at the deterministic velocity value. The velocity of a (deterministic or stochastic) simulation is the slope of the linear approximation to the interface position evolution. We consider the parameter values as per Table 5.7 and the initial conditions (5.78) and (5.79) for the deterministic and the stochastic models, respectively.

We analyse how well the variance derived from the coarse-grained factorial cumulant (Eq. (5.36)) approximates the variance from the stochastic system. We determine the variance from the coarse-grained factorial cumulant using formula (5.42) and we refer to it as the estimated variance. In Figure 5.4 we plot the variance of the stochastic model and the estimated variance at different time points. We consider a log scale for the plots so we can observe all values that the two solutions reach at the different time points considered. However, it is also informative to observe the plot with a smaller y-axis limit upper value so we include the plot using a normal scale for the y-axis in Appendix C.3. In general for $t > 0$, the variance of the compartments in which the cell density is close to the carrying capacity is around 600. We have chosen the time points to distinguish four dynamics:

- At $t = 0$ stochastic and estimated variances are zero, then both variances start increasing, the stochastic variance is lower than the estimated variance but both have a similar travelling wave shape (see Figure 5.4 (a)-(c)).
- At short time, the variances match perfectly and have a travelling wave shape (see Figure 5.4 (d)-(e)).
- At intermediate time, the stochastic and estimated variance have a plateau value for the first compartments and then an abrupt short increase before the variance becomes zero (see Figure 5.4 (f)-(g)).
- For longer time, the stochastic and deterministic variance match for the first part of the domain and then the estimated variance grows abruptly until it reaches a very high value and then slowly decreases. Over time the maximum value of the variance increases. By contrast, after settling at a value of around $\sigma^2 = 600$ for most of the domain, the stochastic variance increases slightly close to the interface and then decreases to zero (see Figure 5.4 (f)-(g)). For the compartments which are close to the solution interface, the estimated variance is higher than the corresponding stochastic variance and the differences increase over time.

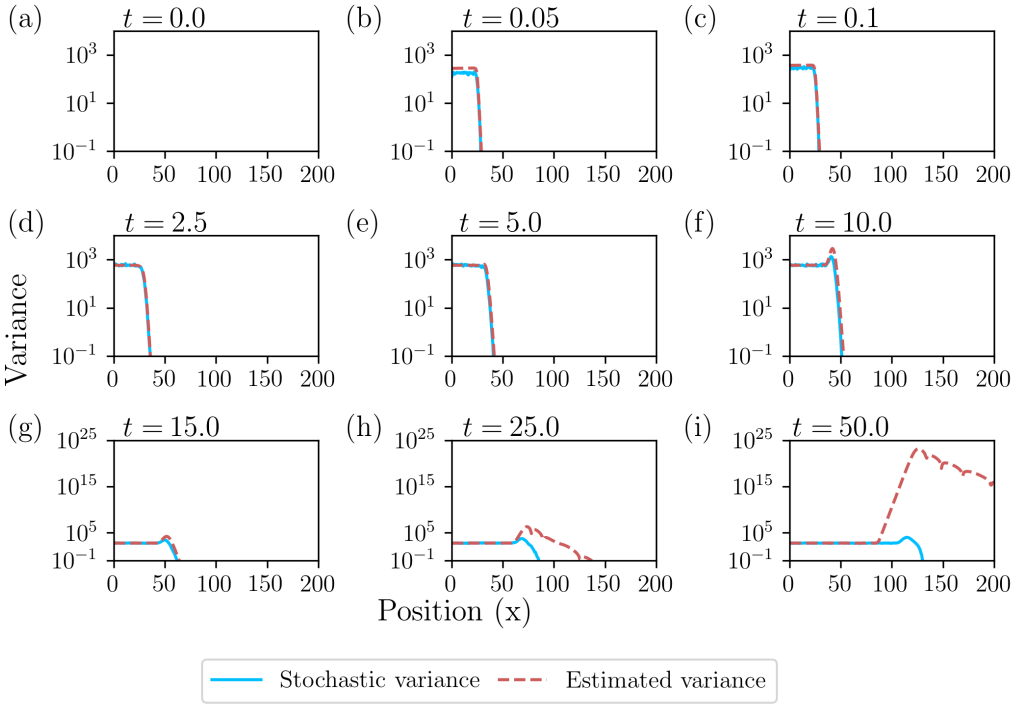


**Figure 5.4:** Plots of the variance of the stochastic model (Eq. (5.9)) and the estimated variance determined from the coarse-grained factorial cumulant (Eq. (5.36)) at different time points. We consider the parameter values as per Table 5.7 and the initial conditions (5.78) and (5.79) for the deterministic and the stochastic models, respectively.

### 5.4.2 Sensitivity analysis

In this section we analyse how the behaviour of the deterministic and stochastic models are affected by the compartment size $h$ and the carrying capacity $K$ of the mean-field system. We consider compartment sizes $h = 0.5$, 0.625 and 0.8 and carrying capacities $K = 50, 100, 200, 600$ and 1000. For each compartment size and carrying capacity we run 500 simulations of the stochastic model (Eq. (5.9)), calculate the solution of the mean-field equation (Eq. (5.23)) and calculate the estimated variance from the coarse-grained factorial cumulant (Eq. (5.36)). The rest of the parameter values are fixed at their default values (see Table 5.7) and initial conditions (5.78) and (5.79) are imposed for the deterministic and stochastic models, respectively.

We first vary the carrying capacity and analyse the mean velocity of the stochastic travelling wave solutions. In Figure 5.5 for each value of the carrying capacity, we plot the mean velocity of 500 simulations of the stochastic model. We also indicate the velocity of the deterministic travelling wave with a horizontal red line.

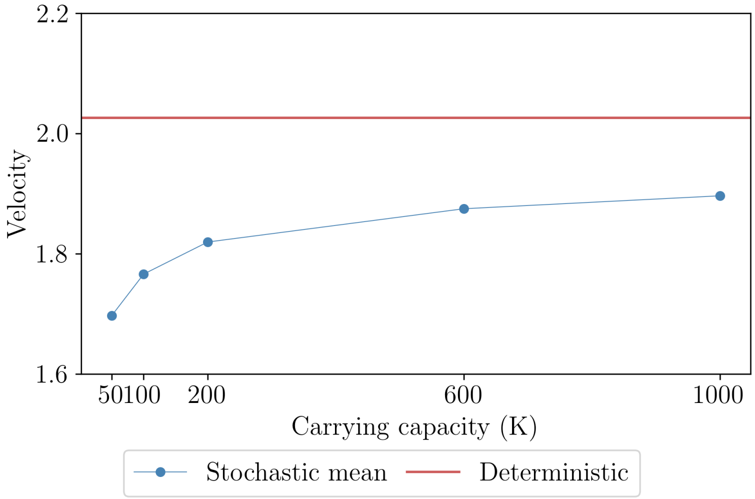


**Figure 5.5:** Sensitivity analysis of the mean velocity of stochastic simulations while varying the carrying capacity. For each carrying capacity considered we plot the mean velocity of 500 stochastic simulations. The velocity of a simulation is the slope of the linear fit of the interface position evolution. The interface position is defined as the smallest compartment in which the cell density is smaller than $0.1K$. We consider the parameter values as per Table 5.7 and the initial conditions (5.78) and (5.79) for the deterministic and the stochastic models, respectively.

To analyse how the estimated and the stochastic variances are affected by the compartment size $h$, we fixed the carrying capacity $K = 400$ and varied the compartment size. For the compartment sizes $h = 0.5$, 0.625 and 0.8 we plot the estimated and stochastic variances over the square of the carrying capacity at different time points in Figure 5.6. We observe that the normalised stochastic variance and estimated variance reach higher values when the compartment size is smaller. Additionally, for $h = 0.625$ and 0.5, we observe that the normalised estimated variance reaches a plateau plateau which is not smooth close to $x = 0$. On the other hand, for $h = 0.8$, there is a smooth plateau at the compartments whose variance is bigger than zero. We therefore consider $h = 0.8$ for the rest of the simulations in the Chapter.

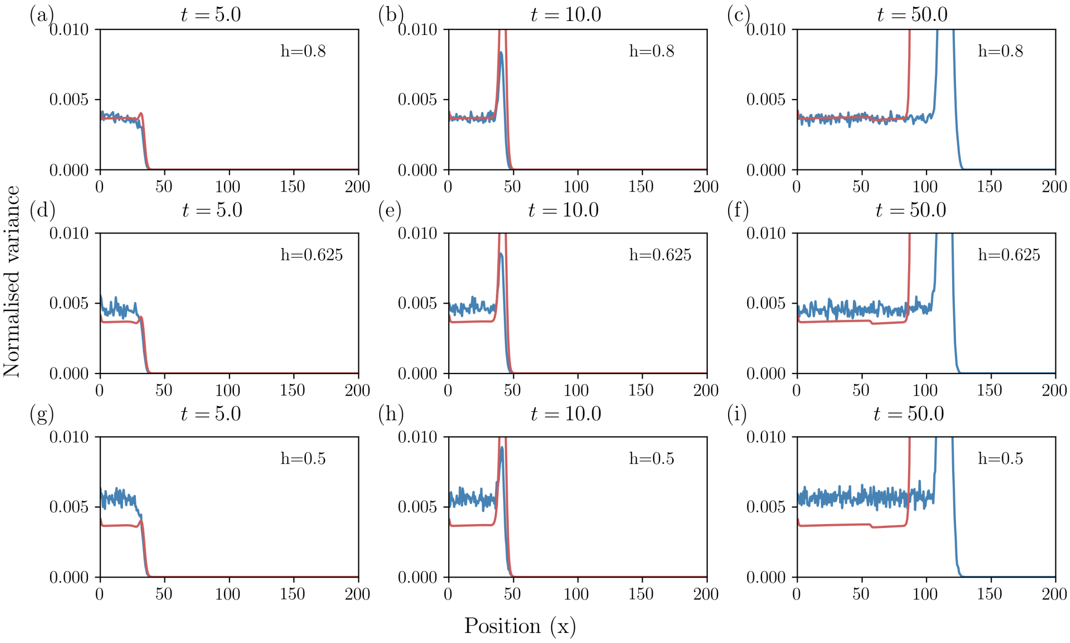


**Figure 5.6:** Evolution of the normalised estimated variance (red line) and stochastic variance (blue line) for fixed carrying capacity $K = 400$ of the macroscopic system and considering the compartment sizes $h = 0.8$, 0.625 and 0.5. The estimated variance is calculated from the coarse-grained factorial cumulant (Eq. (5.36). The variances are normalised by dividing by the carrying capacity squared. Parameter values are as per Table 5.7. For all plots the stochastic normalised variance and the estimated variance is plotted with a blue line coulour and red line coulour respectively.

## 5.5 Proposed hybrid algorithm

In this section we introduce a new hybrid algorithm that extends the algorithm presented in [149]. The main difference between these algorithms is the criterion for the hybrid model to switch from a deterministic to a stochastic framework. We refer to the compartment where the switch between frameworks happen the *model interface*. By contrast in [149] the model interface is defined to be where the cell density passes through a threshold value. If the cell density of a compartment is below the threshold value then the dynamics are modelled stochastically; if the cell density is above that threshold then they are modelled deterministically. Our algorithm defines the model interface by first estimating the variance of the system and then defining a threshold value for the variance. Below we describe our proposed hybrid algorithm in detail.

### 5.5.1 Model interface based on an a priori variance estimator

Our method is predicated on prescribing the model interface based on an estimate of the variance of the stochastic process using the intrinsic noise approximation (Eq. (5.42)). We then use the Van Kampen ansatz to detect the parts of the domain where the deterministic model is appropriate and consequently to identify the location of the model interface.

As described in Section 2.4.4, the Van Kampen ansatz involves approximating the number of molecules in each compartment, $n_i$, by a deterministic and a stochastic variable $\varphi_i$ and $\epsilon_i$, respectively; so that

$$n_i = \Omega\varphi_i + \Omega^{1/2}\epsilon_i \tag{5.81}$$

where $\Omega$ is the system size. If we use this ansatz then we have that the variance satisfies:

$$\frac{var(n_i/h)}{K^2} = \frac{var(\Omega^{1/2}\epsilon_i/h)}{K^2} = \frac{\Omega var(\epsilon_i)/h^2}{\Omega^2/h^2} = \frac{var(\epsilon_i)}{\Omega}. \tag{5.82}$$

Assuming $\Omega >> 1$, we have that $\frac{var(n_i/h)}{K^2} = \frac{var(\epsilon)}{\Omega} << 1$. Motivated by the Van Kampen ansatz and its effects on the variance of $n_i/h$, we define the model interface to be the first compartment in which the following inequality is satisfied:

$$\frac{v(x,t)}{K^2} > \psi \tag{5.83}$$

where $0 < \psi < 1$.

The normalised variance $var(n_i/h)/K^2$ is estimated by solving the PDE for the factorial cumulant (Eq. (5.42)) and then using Eq. (5.36). The value of $\psi$ is set by inspecting the PDE solution and determining the critical value for the normalised variance.

The model interface position is determined *a priori* by solving the PDEs of the mean-field model and the coarse-grained factorial cumulant evolution. We then determine a set of times $t_0 < t_1 < t_2 < \ldots < t_X$ such that $t_0$ is the first time point at which the inequality (5.83) is satisfied, i. e. when the model interface and the two domains first appear. The times $t_i$ ($i > 1$) indicate when the model interface changes position.

We use the stochastic FKPP model to illustrate this procedure. We fix the parameter values as per Table 5.7 and use the initial conditions (5.78) and (5.79) for the deterministic and the stochastic models, respectively. We analyse the evolution of the estimated normalised variance, $w(x_i, x_i, t)/K^2$, and the stochastic variance, $var(n_i/h)/K^2$. In Figure 5.7 we plot the stochastic and the estimated variance over the carrying capacity squared. We observe that the normalised variances are typically less than 0.005 for most of the time points. We observe that for $t = 10$ (Figure 5.7 (f)) the estimated and stochastic normalised variances increase close to the solution interface. Over time, the normalised variance increases close to the interface and reaches higher values and decreases with a slower rate. This behaviour is more pronounced for the estimated normalised variance which, while increasing close to the interface, remains bounded by 0.005 at all simulated time points.

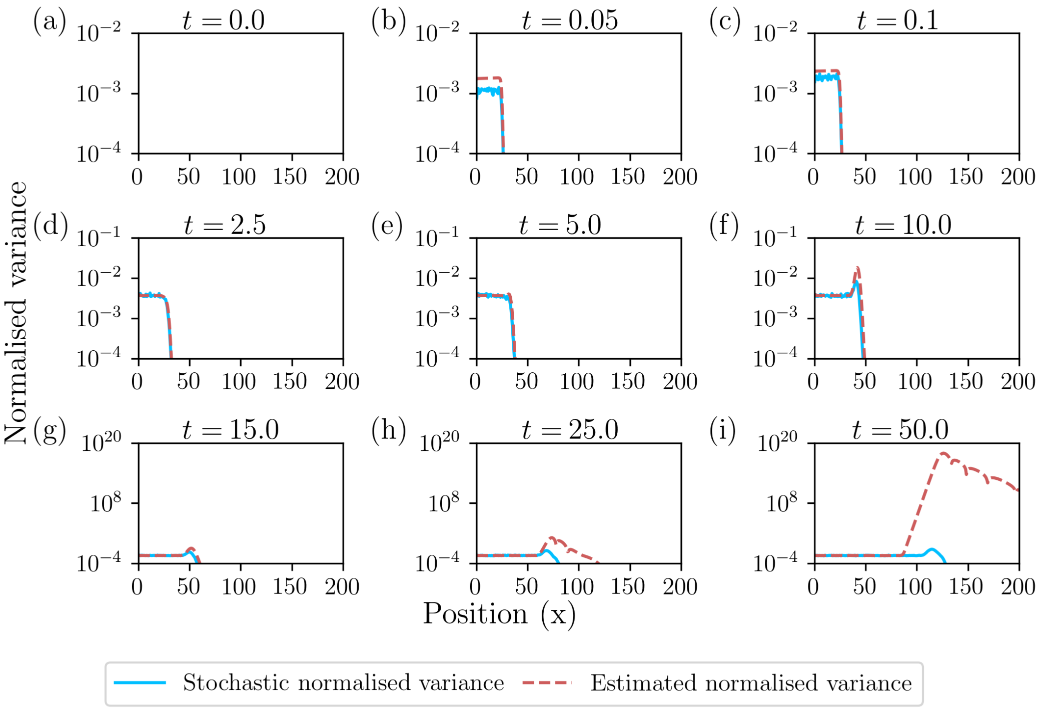


**Figure 5.7:** a) Plots of the estimated variance from the coarse-grained factorial cumulant (Eq. (5.36)) and the stochastic variance divided by the carrying capacity squared for different time points. The y-axis is on a log scale. Parameter values are as per Table 5.7.

In view of these numerical results, we consider the critical value for the model interface estimator to be $\psi = 0.0041$ since this value is an upper bound of the normalised variances for most of the compartments with cell density close to the carrying capacity over time. We then identify the first compartment for which this threshold is reached. Figure 5.7 shows that the normalised estimated variance reaches the threshold at $t^* = 5.15$. In Figure 5.8 we plot the evolution of the model interface.

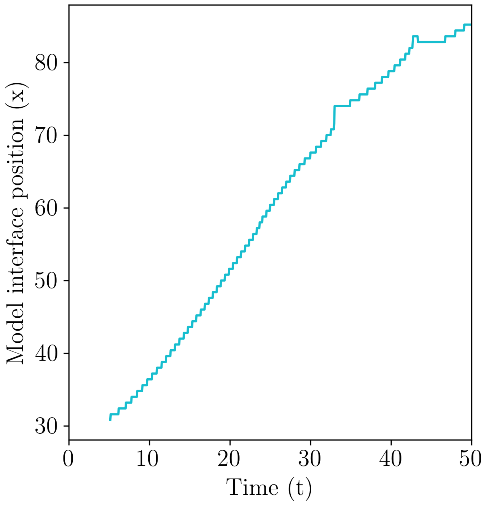


**Figure 5.8:** Time evolution of the model interface. The model interface is the smallest compartment position in which the normalised variance is above the threshold 0.0041. The model interface is the compartment in which, for the hybrid model, the description is switched from stochastic to deterministic. The model interface is initiated at $t^* = 5.15$ since this is the earliest time point for which the threshold is reached.

### 5.5.2 Hybrid algorithm

We now present our hybrid algorithm using the stochastic FKPP model as a prototypical model. We consider this model since it exhibits travelling wave solutions and therefore we expect to have a single model interface. We denote the compartment where the model interface is located as $I$. Given the initial conditions (see Eq. (5.79)) the domain is divided in the following way:

- mean-field subdomain: compartments $j = 1, \dots, I$,
- Stochastic subdomain: compartments $j = I, \dots, N$.

The mean-field and stochastic subdomains denote regions in which the cell density is updated with the mean-field equation or the stochastic algorithm, respectively. Compartment $j = I$ is used in both the mean-field and stochastic formulation.

The hybrid algorithm is described in Algorithm 5.4. Since its main feature is that it is based on estimating the variance, we call it the "*a priori* variance method" or AV method.

**Algorithm 5.4** *A priori* variance method

```
Solve the mean-field evolution Eq. (5.23) and the coarse-grained factorial cumulant evolution equation (Eq. (5.36)) for the time domain $[0, T]$ to determine the times $t_0 < t_1 < t_2 < \ldots < t_X$ at which the model interface changes position.
Set $t = 0$ and initialise the stochastic model in the whole space domain $[0, L]$.
While $t < t_0$, update the model using the Gillespie algorithm (Algorithm 2.1).
At $t = t_0$, initialise the two domains and the compartment interface. Set the total propensity ($\alpha_0$) to be the sum of the propensities for the reactions and the jumps within the stochastic domain and the jumps from $j = I$ and $j = I + 1$.
while $t < T$ do
    Generate $\tau$, the time to the next stochastic reaction using Eq. (2.53).
    Determine which stochastic event happens using condition (2.54).
    Update the concentration at the compartment interface (if changed by the stochastic process). Let $u_I^+$ denote the concentration after the stochastic term.
    Update the concentration in the interface compartment from the deterministic flux between compartments $j = I$ and $j = I - 1$.
        $u_I(t+\tau) = u_I^+(t) + \frac{D\tau}{h}\left(u_I^+(t) - u_{I-1}(t)\right)$        (5.84)
    Update the concentration values in the deterministic domain using the mean-field equation and $\tau$ as the time step.
    Update $t = t + \tau$.
    If $t$ reaches $t_i$, then update the interface and the total propensity.
end while
```

### 5.5.3 Implementation aspects

In Algorithm 5.4 we use finite elements to discretise and update the mean-field equation. In this section we describe the finite element discretisation and the procedure used to obtain the jump rates for the stochastic update.

As described in Algorithm 5.4, the deterministic update is performed using an explicit Euler scheme to discretise the time derivative and finite elements to discretise in space.

We denote the space domain $D_I := (0, x_I)$ where $x_I$ is the location of the model interface. We consider the spatial mesh given by the centers of the compartments of the stochastic model. The spatial mesh consists of $\{x_0, \dots, x_I\}$ where $x_i = h/2 + ih$ for $i = 0, \dots, I-1$. We describe the deterministic update from time $t = \hat{t}$ to time $t = \hat{t} + \tau$ where $\tau$ is determined by Eq. (2.53) according to the hybrid algorithm 5.4 and $\hat{t}$ is a time in $[0, T]$. We denote time $\hat{t}$ as $t_n$ and $\hat{t} + \tau$ as $t_{n+1}$.

Applying the explicit Euler scheme to the mean-field equation (Eq. (5.23)) we obtain:

$$\frac{u^n(x) - u^{n-1}(x)}{\tau} = \left[ D\Delta u^{n-1}(x) + \lambda u^{n-1}(x) \left( 1 - \frac{u^{n-1}(x)}{K} \right) \right]. \tag{5.85}$$

We now discretise Eq. (5.85) using finite elements. We assume the solution of Eq. (5.85) belongs to $H^1(D_I)$. To approximate the solution in $H^1(D_I)$ we consider the finite-dimensional space $V_h$ (see Sec. 2.3 for a description of the functional spaces $H^1(D_I)$ and $V_h$). We then formulate the variational problem in the space $V_h$. The weak formulation at each $t_n$ ($n \geq 1$) consists of determining $u^n(x) \in V_h$ such that

$$\left( \frac{u^n(x) - u^{n-1}(x)}{\tau}, v_h(x) \right) + T^e(u^n(x), v_h(x)) = F^e(u^n(x), v_h(x)) \quad \forall v_h(x) \in V_h, \tag{5.86}$$

where

$$T^e(u^{n-1}(x), v_h(x)) = D(\nabla u^{n-1}(x), \nabla v_h(x)),$$

and

$$F^e(u^{n-1}(x), v_h(x)) = \lambda(u^{n-1}(x), v_h(x)) + \frac{\lambda}{K} \left( (u^{n-1}(x))^2, v_h(x) \right).$$

Unlike the variational problem for the mean-field equation (Eq. (5.44)), Eq. (5.86) is a linear variation problem for $u^n(x)$ so we reduce it to a finite-dimensional linear system.

We approximate the solution $u^n(x)$ by its Ritz projection in the space $V_h$:

$$u^n(x) \approx u^n_h(x).$$

In terms of the nodal basis functions of the space $V_h$, $\{\varphi_1(x), \dots, \varphi_I(x)\}$, $u^n_h(x)$ is given by:

$$u^n_h(x) = \sum_{j=0}^{I-1} U^n_j \varphi_j(x),$$

where $U^n_j$ are real constants for $0 \leq j \leq I-1$.

For the variational problem (5.86) to be satisfied for all $v_h(x) \in V_h$, it is necessary and sufficient that it is satisfied for the basis functions. By considering the variational formulation (5.86) for the basis functions, we reduce the problem to a linear system for the coefficients $U^n_j$, $0 \leq j \leq I-1$. The matrix formulation of the linear system that needs to be solved is the following:

$$\boldsymbol{M}\boldsymbol{U}^n = \boldsymbol{C}^n,$$

where $\boldsymbol{M}$ is the $(I \times I)$ mass matrix whose entries are given by:

$$\boldsymbol{M}_{i,j} := (\varphi_i(x), \varphi_j(x)),$$

$\boldsymbol{C}^n$ is a $I$-dimensional vector with entries given by:

$$\begin{aligned} \boldsymbol{C}_i^n :=& (\varphi_i(x), u^{n-1}(x)) - D\tau(\nabla\varphi_i(x), \nabla u^{n-1}(x)) \\ &+ \tau\lambda(\varphi_i(x), u^{n-1}(x)) - \frac{\tau\lambda}{K}\left(\varphi_i(x), \left(u^{n-1}(x)\right)^2\right), \end{aligned} \tag{5.87}$$

and $\boldsymbol{U}^n$ is the $I$-dimensional solution vector.

## 5.6 Approximation analysis of the hybrid algorithm

In this section we investigate how the hybrid method approximates the stochastic process. We refer to the simulations of the hybrid algorithm 5.4 as simulations of the hybrid model corresponding to the stochastic FKfcPP model. As for the stochastic model, we refer to the hybrid interface, variance and standard deviation of the hybrid simulations.

Following the hybrid algorithm 5.4, we first solve the PDE of the coarse-grained factorial cumulant evolution to obtain an estimate for the stochastic variance evolution. We then obtain the evolution of the model interface (see Figure 5.8). Compartments with a mean-field description form the deterministic domain, and compartments whose cell density is modelled with the stochastic update form the stochastic domain. We perform 500 simulations of the hybrid model using the model interface using the parameter values stated in Table 5.7. In Figure 5.9 we plot the mean cell density of the stochastic and the hybrid model at different time points. The mean evolution of both models matches perfectly.

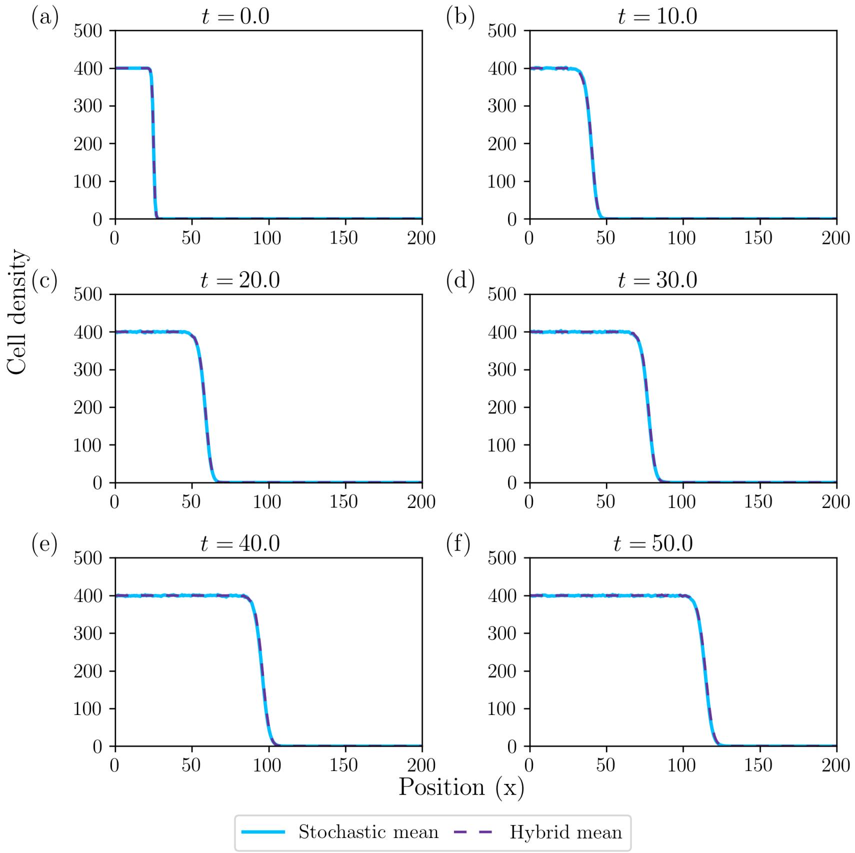


**Figure 5.9:** Series of plots showing the mean evolution of 500 simulations of the stochastic FKPP model (Eq. (5.9)) and the hybrid model (Algorithm 5.4) at different time points. Parameter values are as per Table 5.7. The evolution of the mean cell density appears to match very well.

Regarding the variance, we also observe remarkable agreement between the hybrid approximation and the actual stochastic model. In Figure 5.10 we plot the evolution of the variance at different time points. We note that for small times the hybrid variance matches perfectly the values of the stochastic variance since at early times the hybrid algorithm is identical to the stochastic algorithm (see Figure 5.10 (b)-(e)). Once the model interface has been initiated at $t^* = 5.15$, the hybrid variance of the compartments from the deterministic domain becomes zero and the hybrid variance of the compartments from the stochastic domain approximates well the stochastic variance (see Figure 5.10 (f)-(i)).

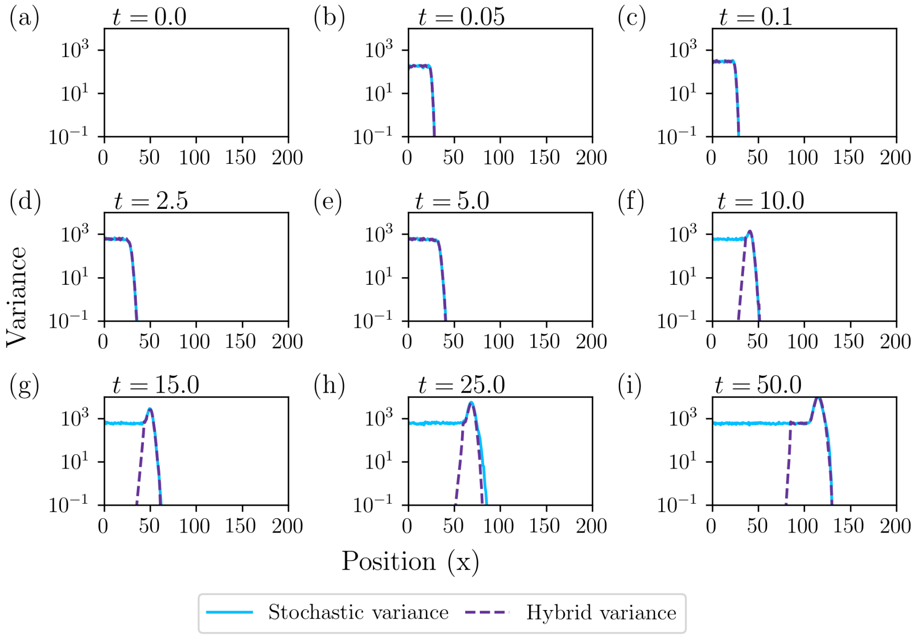


**Figure 5.10:** Hybrid and stochastic variance from 500 simulations at different time points. Parameter values as per Table 5.7.

We also consider the interface positions predicted by the stochastic and hybrid models. In Figure 5.11 we plot the evolution of the interface position of the deterministic model and the mean interface position of the stochastic and hybrid models. We recall that the interface position of a deterministic and stochastic simulation is the smallest spatial position for which the cell density is below the threshold value (5.80). We note that the interface positions of the stochastic and hybrid models are in good agreement while the deterministic interface moves more slowly.

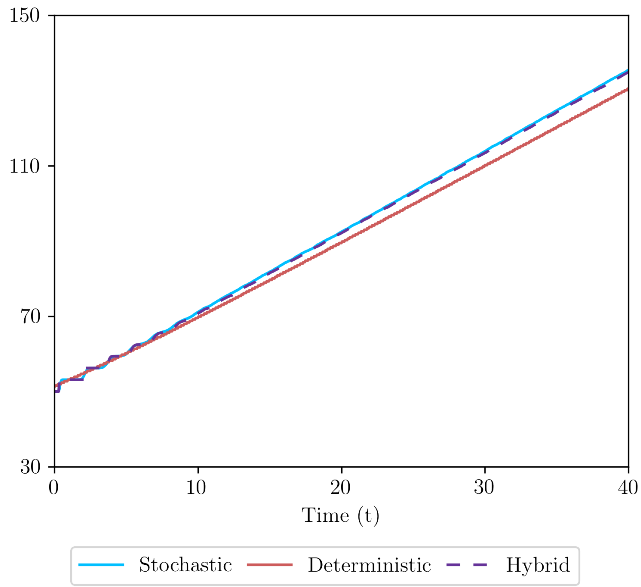


**Figure 5.11:** Plots contrasting the different behaviours of the stochastic and the hybrid model. We plot the evolution of the mean interface position of the stochastic and hybrid models and the deterministic interface evolution.

To further assess the accuracy of the AV method, we compare the statistics of the stochastic and the hybrid models when the model exhibits travelling wave behaviour, we expect the model to have an initial transient behaviour given by the initial conditions and then to exhibit travelling wave behaviour. We consider the model to start exhibiting travelling wave behaviour when the evolution of the deterministic interface position is best described by a linear evolution. We therefore analyse the coefficient of determination of the linear fit to the deterministic interface position evolution while varying the initial time point for making the fit. In Figure 5.12 we plot the coefficient of determination for each initial time point that we consider. We observe that the differences in $R^2$ are in the fourth decimal place and all coefficients of determination are very close to 1. Henceforth we fix the initial time for performing the linear fit as $t = 25$ since it seems that at this time point, the coefficient of determination reaches a plateau value.

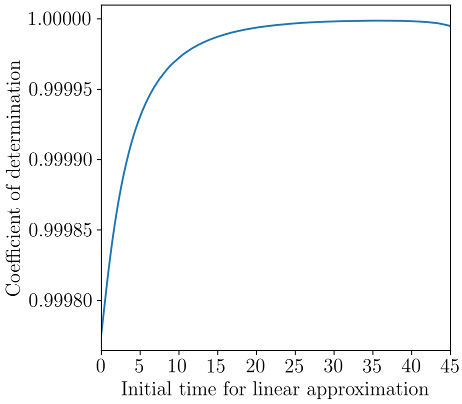


**Figure 5.12:** Coefficient of determination when performing the linear fit of deterministic interface position evolution. For each initial time point $t$, $0 \leq t \leq 45$ we consider the linear fit of the deterministic interface position evolution and calculated the corresponding coefficient of determination. The coeffient of determination seems to be close to 1 for all values of $t$ chosen.

To assess how the hybrid model approximates the stochastic model we perform statistical tests between the distributions of the velocities and the interface positions of the two models. We focus first on comparing the distribution of velocities. We consider as an initial point $t = 25$ and consider a linear fit to the interface evolution of the stochastic and the hybrid simulations. We calculate the set of velocities for each model and perform KS and t-tests between the velocity distributions. We set the significance threshold at $\alpha = 0.05$. In Figure 5.13 (a) we plot the normalised histograms of the two velocity distributions. Both have a similar shape and the mean velocity values are 1.855 (for the stochastic model) and 1.867 (for the hybrid model). In Figure 5.13 (b) we plot the p-value of applying the KS and t-tests to detect significant differences. We have also plotted a horizontal black line to indicate the significance threshold 0.05. If the $p$-value is below the significance threshold, then the null hypothesis is rejected; otherwise, it is not. The KS-test does not detect a significant difference between the distributions while the t-test does. This means that the distributions are similar in shape but their mean values are statistically different. It is not unexpected that the t-test detected a significant difference even though the means are very close: since the number of simulations is 500, the allowed standard error is very small.

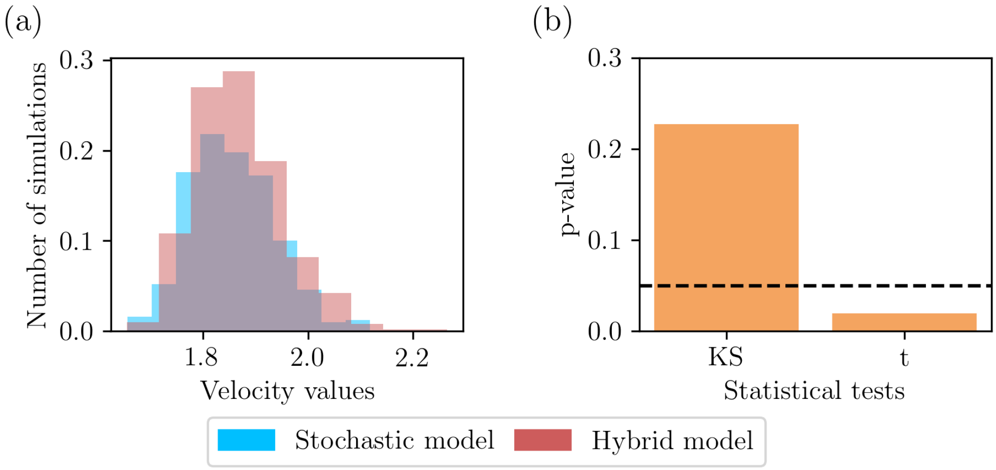


**Figure 5.13:** Plots comparing the velocity distributions of the stochastic and hybrid models. In (a) we plot the normalised velocity distributions for both models. Both distributions are similar in shape and mean values. In (b) we plot the p-value from performing a KS- and t-tests on the velocity between the velocity distributions of the stochastic and the hybrid model. The horizontal black line indicates the significance threshold $\alpha = 0.05$. The KS-test does not detect a significant difference between the distributions while the t-test detects a significant difference. The model velocities in each case are determined by performing a linear fit to the simulations' interface evolution considering as an inital time point $t = 25$.

We now apply the same statistical analysis to the interface position distributions of the hybrid and stochastic models. For times $25 < t < 50$, we perform statistical tests between the hybrid and stochastic interface positions at each time point and then count the number of times the statistical tests detected significant differences. In Figure 5.14 (a) we plot the percentage of times when there is no significant difference between the interface position distributions of the two models. We observe that close to 100 percent of times there is no difference with the KS-test whereas when the t-test is used, 51 percent of times there is a statistically significant difference. In Figure 5.14 (b) we plot the distribution of the interface positions at time $t = 40$ to illustrate the good agreement between the two models.

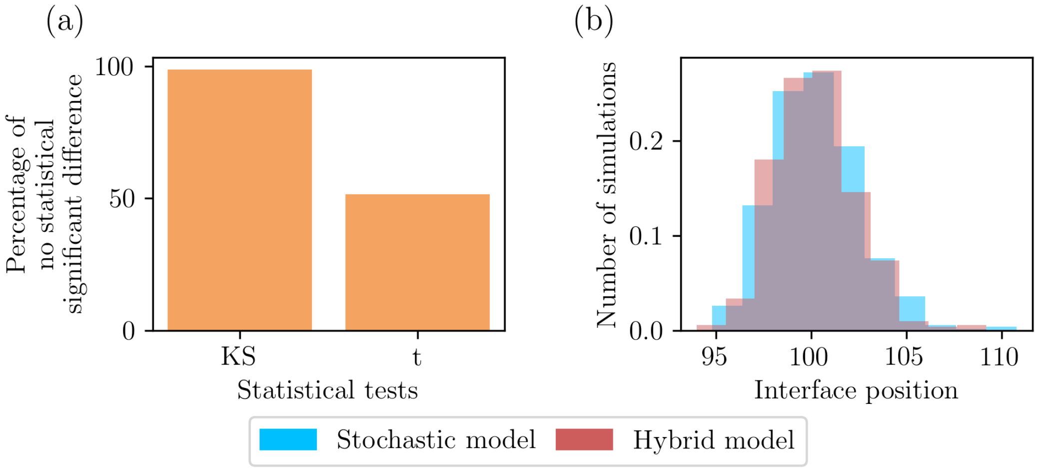


**Figure 5.14:** Plots summarising the statistical analysis of the inteface position distributions of the hybrid and the stochastic simulations. In (a) we plot the percentage of time points at which the statistical test did not find a statistically significant difference between the stochastic and the hybrid interface positions. To exemplify what the interface position distributions look like, in (b) we plot the interface position distributions of the stochastic and the hybrid model.

## 5.7 Comparison of hybrid algorithms

We now compare the performance of our hybrid algorithm with the hybrid algorithm presented by Spill et al. in [149]. Their algorithm defines the model interface as the first compartment in which the cell density is below a threshold value (see Section 2.4.5 for a more detailed description of their model). We implement their algorithm using three thresholds $0.5K, 0.8K$ and, $0.9K$ and compare the hybrid simulations to ours. We refer to their algorithm as the Spill method.

We perform 500 simulations using the stochastic and hybrid algorithms. We then perform statistical tests on the interface position and the velocity distributions of the stochastic model and each hybrid model. We refer to our hybrid algorithm as the AV method and denote by Spill 0.5K, Spill 0.8K and Spill 0.9K, the three different implementations of the Spill method, where we specify the threshold for the model interface.

In Figure 5.15 (a) we plot the mean interface evolution of the stochastic and hybrid models. By inspection, all models appear to match perfectly. If we zoom in on the evolution during the time subdomain $[4.52, 4.66]$ we observe that the interface position evolution of all the models differ by less than 0.2. In Figure 5.15 (b) we plot the velocity distributions of each model. To determine the velocity of each model simulation, we consider the linear fit to the interface evolution starting at $t = 25$. The boxplot corresponding to the stochastic model has a larger number of outliers, but overall the distributions seem to be in good agreement.

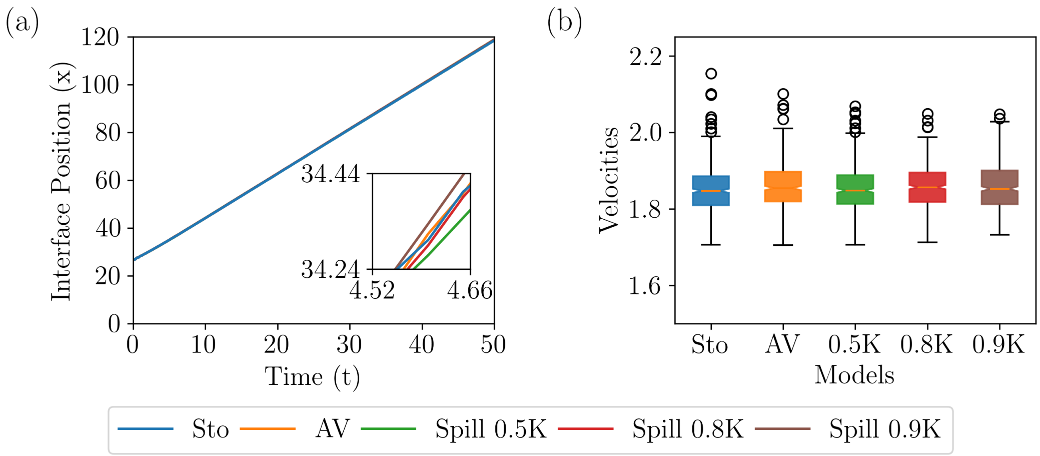


**Figure 5.15:** Series of plots comparing the dynamics of the stochastic and hybrid models. In (a) we plot the mean interface position of the stochastic and the hybrid models.The evolution of the mean interface position of all models is very similar. In (b) we plot the velocity distributions of each model considered. The stochastic model is labelled as "Sto", the Apriori Variance method as "AV" and the implementations of the Spill method are denoted as "$\nu K$" where $\nu$ is the threshold used to define the model interface for each implemtation. The statistics of the models are derived from running 500 simulations with the parameter values as per Table 5.7.

We first compare the velocity distributions of each hybrid model against that of the stochastic model by performing statistical tests. In Figure 5.16 (a) and (b) we plot the $p$-values for KS- and t-tests comparing the velocity distributions of each hybrid model against the velocity distribution of the stochastic model. In both plots we denote with a black horizontal dashed line the significance level $\alpha = 0.05$. If the p-value is below this significance level then we reject the null hypothesis of the statistical tests. The null hypothesis of the KS-test is rejected for all hybrid models and we observe that the largest p-value is for the Spill 0.9K followed by the Spill 0.8K method, then the AV method and finally the Spill 0.5 method. The t-test only detected significant differences for the AV method. As expected, as the threshold increases in the Spill method, its approximation of the stochastic process becomes of better quality. The improvement in the approximation of the Spill method due to its threshold can be seen by the p-value from the statistical tests which increases as the threshold increases (see Figure 5.15).

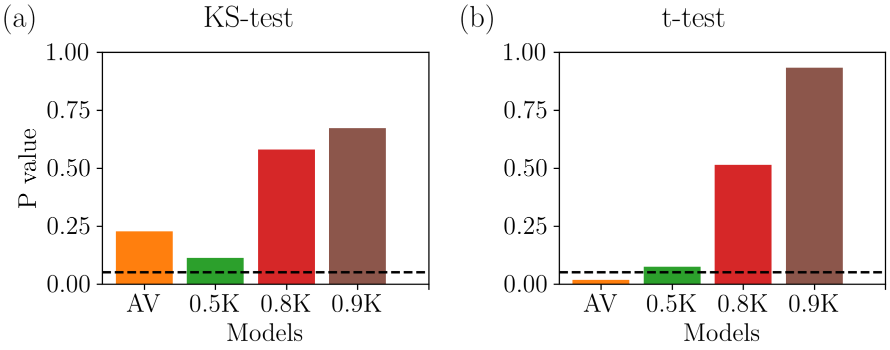


**Figure 5.16:** Statistical results of comparing the velocity distributions of the stochastic model and the hybrid models. The black horizontal dashed line indicates the significance level $\alpha = 0.05$. In (a) and (b) we plot the p-values of performing a KS and a t-test, respectively, between the velocity distribution of the stochastic model and each of the hybrid models considered.

We now compare the interface position distributions of the different models. Since we want to perform this analysis when we have travelling wave dynamics, we consider time points $25 < t < 50$. At each time point we perform a KS-test and a t-test between the interface position distribution of the stochastic model and each hybrid model. We then calculate the percentage of time points where we find no statistical significant difference. In Figure 5.17 (a) and (b) we plot the percentage of time points where the KS and t-test detected no statistical significant difference between the stochastic model and each hybrid model considered. We observe that for the KS-test all the hybrid models have a percentage higher than 98 percent. When considering the t-test, the AV method has the lowest percentage of time points with no statistical significant difference, with 51 percent, while the different implementations of the Spill Method have 100 percent.

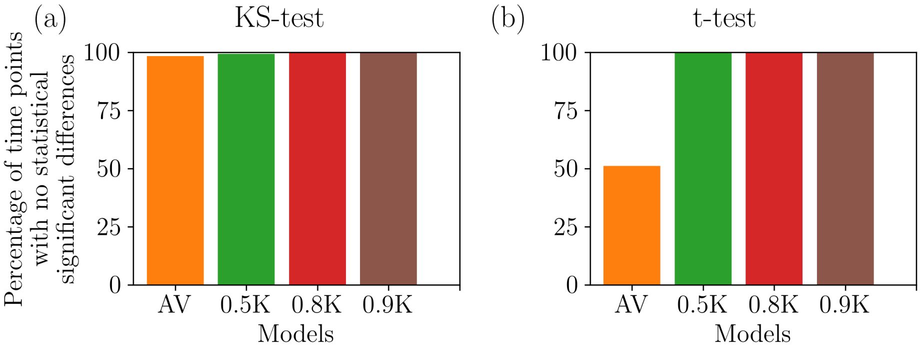


**Figure 5.17:** Statistical results of performing a KS- and a t-test between the interface position distribution of the stochastic model and that of each hybrid model considered. In (a) and (b) we plot the percentage of time points where no statistical significant differences was found between the stochastic and the hybrid interface positions when applying a KS- and a t-test, respectively.

We observe that the model interface location affects how well the hybrid model approximates the stochastic model. In Figure 5.18 we plot the model interface for all the hybrid models considered. As expected for the Spill method, the smaller the threshold for the model interface, the larger the stochastic domain. The model interface from the AV method starts with a smaller stochastic domain than the Spill 0.5K method but after $t = 18$ it has the largest stochastic domain.

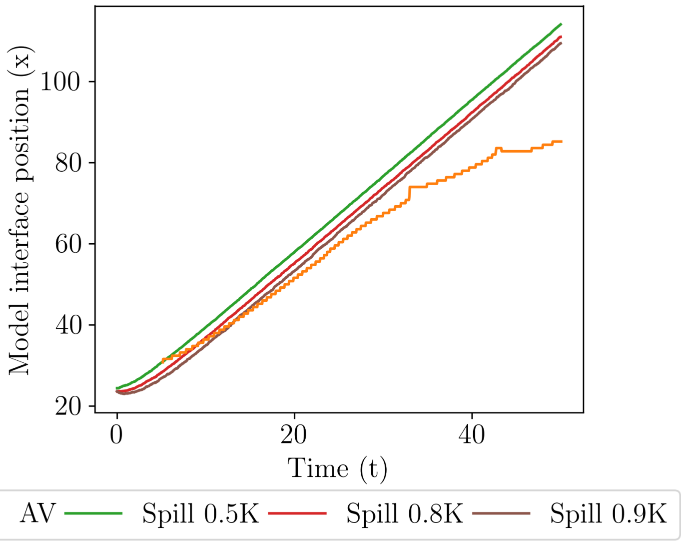


**Figure 5.18:** Model interface evolution of the different hybrid models considered.

To complete our analysis we compare the computational time of running the different models. We consider as initial conditions for the mean-field equation (Eq. (5.23)):

$$v(x) = -\frac{K}{2}\tanh(x - 150) + \frac{K}{2}. \tag{5.88}$$

The initial condition for the stochastic model is given by:

$$V(x) = round(hv(x)). \tag{5.89}$$

We consider $K = 1000$, an end time of $T = 10$ and fix all other parameters at their default values (see Table 5.7).

We consider the initial condition (5.88) since the initial position of the interface position is at $x = 150$ and therefore several compartments have a cell density close to $K = 1000$. In Table 5.8 we compare the computational time needed to perform 10 simulations of the stochastic and hybrid models. Since for the AV method we need to derive the mean-field and the factorial cumulant evolution equations we also include the time taken to solve these equations. The simulations were performed on a Fujitsu ESPRIMO P920 8-core server with 3.40GHz and 15.5 GB RAM. If we consider only the running time of the stochastic and hybrid models, the advantage in performance of the hybrid methods is clear. However, for the AV method the mean-field and the coarse-grained factorial cumulant require additional time to run and significant amount of time is needed to determine the factorial cumulant. For the Spill hybrid model, as the threshold decreases, the running time decreases. Since the stochastic domain is larger for the AV model than the Spill implementations, it makes sense that the simulations take longer.

| Model/Equation | Running time (s) |
|---|---|
| Mean-field | 57 |
| Coarse-grained factorial cumulant | 2206 |
| Stochastic | 102 |
| Spill 0.5 K | 13 |
| Spill 0.8 K | 28 |
| Spill 0.9 K | 42 |
| AV | 52 |

**Table 5.8:** Comparison of the running times of the different models. We include the running time of the mean-field and the coarse-grained factorial cumulant evolution equation since they are needed for the AV method. The running time of the stochastic and the hybrid models correspond to 10 simulations. We consider $K = 1000$, end time $T = 10$ and the rest of the parameter values as per Table 5.7. The initial conditions of the mean-field and the stochastic models are given by (5.88) and (5.89), respectively.

## 5.8 Discussion

Hybrid algorithms typically consider heuristic methods based on the cell density for determining which compartments will be updated deterministically or stochastically [100, 149, 196]. In this chapter we considered instead a natural indicator of when the mean-field approximation is valid: the size of fluctuations. We introduced a new hybrid algorithm that considers the variance of the cell density in each compartment to determine if it is updated deterministically or stochastically. The system's variance is estimated *a priori* the stochastic simulations using Van Kampen's intrinsic noise approximation. We termed this algorithm the "A priori Variance" (AV) method. The aim of this chapter was to examine how well the AV method approximated a prototypical stochastic process and compare its performance to standard heuristic hybrid algorithms.

We applied the AV method to the stochastic Fisher-Kolmogorov-Petrovsky-Piscounov (FKPP) model. We considered this model since it exhibits travelling wave solutions, so at each time point, the model has a region where the cell density is small and the stochastic effects are important, and a region where the cell density is close to the carrying capacity and a mean-field approximation is valid. We first determined the mean-field and coarse-grained factorial cumulant equations for the stochastic FKPP model. We described the discretisation schemes and, through the method of manufactured solutions, we calculated the estimated convergence rates. We then introduced the typical dynamics of the stochastic model and the main features of the solution interface position and the velocity distributions. We then applied the AV method to the stochastic FKPP model. We analysed the approximation of the hybrid algorithm to the stochastic process by performing statistical tests and observed that the distributions differed in the mean but not in shape. Finally, we compared the AV method against the hybrid algorithm introduced by Spill et al. in [149], to which we referred to as the Spill method. We performed statistical tests between the velocities and the interface position of the hybrid simulations and showed that in most of the cases the Spill method was better than ours.

The numerical analysis of the discretisation schemes needs to be further investigated. First, the convergence properties of the discretisation scheme of the coarse-grained factorial cumulant equation need to be further analysed. In this work we estimated the convergence rates of the discretisation scheme using the method of manufactured of solutions. We observed that the estimated time convergence rate was not the optimal one for a discretisation scheme consisting of a backward Euler scheme and linear Lagrange finite elements when solving a regularised system. Since we did not perform an error analysis study, we can not conclude if the observed convergence rate should be the same as the optimal one. An error analysis of the discretisation scheme should be performed to determine the optimal convergence rates. Another aspect that should be analyse is that the coarse-grained factorial cumulant solution blows up at large times (see Figure 5.4). An analysis of the equations needs to be performed to determine if this is a model behaviour or a numerical artificat. There are many directions that could be undertaken to improve the discretisation schemes. The current setup considers a spatial locally refined mesh and the parameters that define the refinement were chosen to be the best from a small subset of parameters that were tested. The space-time mesh could be refined using a more systematic refinement process based on *a posteriori* error estimator approach [6]. We could also analyse the effect of the regularised parameter $\epsilon$ or consider a different discretisation for the delta and the indicator function.

The comparison of our hybrid algorithm with other hybrid algorithms can be extended. We compared the statistical properties of our model against the one developed by Spill et al. in [149]. To our knowledge this is the first time a statistical systematic comparison has been performed between different hybrid algorithms that couple mesoscopic and macroscopic descriptions of a stochastic reaction-diffusion system. An extension of this work could be a comparison of other hybrid algorithms by analysing them in the same systematic way. In this work, by considering different thresholds for the Spill method, we observed that the sooner the model interface appeared, the better the statistical results were (see Figure 5.15 and 5.16). Surprisingly, despite our hybrid algorithm having a larger stochastic domain for the later time points, its statistical performance was inferior to that of the different implementations of the Spill method. A possible way to extend the work in this chapter would be to determine if the following factors decreased the statistical performance of our AV method: the roughness of the model interface evolution, the switch between stochastic to the hybrid formulation during the simulation or the initial phase in which the model interface of the AV method is above the model interface for the different implementations of the Spill method.

An extension of this work would be to extend the hybrid algorithm to a more general class of stochastic reaction-diffusion systems. The presented algorithm considers a single specie, a single model interface in an one-dimensional domain. It appears difficult to generalise the hybrid algorithm to higher spatial dimensions since considering an $n$-dimensional space would involve solving a $2n$-dimensional PDE for the factorial cumulant. However, other methods can be used to estimate the variance that do not rely on solving a PDE for the factorial cumulant evolution. In recent work [147], Lötstedt developed a fast algorithm to approximate the mean and covariances of a RDME model where only a system of ODEs needs to be solved to determine the mean and the covariances, making the extension of the AV method a possibility.

## Chapter 6

# Discussion and conclusions

Mathematical models of biological systems have a great potential to provide insights on the underlying mechanisms of the system. By performing model analysis and simulations of the mathematical model we can systematically analyse the system under perturbations and develop testable predictions. In this work we have introduced different mathematical and statistical models that were able to shed light on different aspects of scratch assays. We discuss the significance and the limitations of the frameworks presented and the future directions of this work.

In Chapter 3 we presented a quantification method that characterises the migration in scratch assays by a set of representative velocities. We tested its performance and compared it to other standard quantification methods by considering in-vitro and in-silico data. The in-silico data consisted of simulations from an agent-based model that simulated the in-vitro process. Through this work we showed the benefits of assessing the quantification method with in-silico data: the agent-based model allowed us to test our quantification method and compare it to standard methods under controlled experimental conditions. We analysed how the quantification methods are affected by smooth and irregular leading edges and different proliferation and migration rates of scratch assays. This methodology can be extended to test how the quantification methods are affected by other biological processes (e.g. cell-cell adhesion and cell mechanics). Ideally new quantification methods should be compared considering benchmarked in-vitro datasets [230]. Assessment of the method can be complemented by considering in-silico data and testing how the quantification method is affected by different cell processes as we did in this work.

The most relevant limitations of the work in Chapter 3 are the high variability of the velocity distributions under the same parameters, the decrease in performance as the proliferation rate increases and the difficulty of extending the method for other geometries. As we mentioned in Chapter 3, Section 3.8 the high variability between the velocity distributions under the same parameters can give rise to false positive classifications as observed in the in-silico data. This weakness can be addressed by characterising the scratch assay dynamics with additional information such as the evolution of the pair-correlation function over the scratch domain [5]. This notion can also help to improve the method to deal properly with high proliferation rates. While the generalisation to other nonlinear wound models is not clear, (e.g. circular model wounds), the proposed method showed that determining the velocity field along the cell front is sufficient to characterise the migration.

In Chapter 4 we analysed an age-structured model with resource-regulated proliferation and showed that the model can account for the biphasic behaviour in the per capita growth rate previously observed experimentally in scratch assays [118]. We described the cell population dynamics in terms of the cell subpopulations which are mature enough to proliferate and those who are not. This allowed us to derive necessary and sufficient conditions under which the model exhibits a logistic type behaviour and two proliferation phases: an initial phase termed "disturbance phase", in which proliferation does not follow a logistic growth, and a "growth phase", where proliferation is approximately logistic. The biphasic behaviour in our model was found to be the result of an initial increase in the fraction of mature cells followed by a decline to their steady-state concentration. The model predicted that by increasing the initial resource concentration, the disturbance phase would shorten. By performing a parameter sensitivity analysis, we showed that our model predictions are robust against parameter perturbations. This was the first model to account for the per capita growth rate biphasic behaviour observed experimentally in [118]. There are four natural extensions of this work: to perform experiments that can validate the model predictions, to improve the model parametrisation, to include spatial effects and to consider alternative explanations for the biphasic dynamics and perform an inference-based modelling approach to test multiple hypotheses and assist with the design of experiments that discriminate between feasible alternatives.

In Chapter 5 we introduced a new hybrid algorithm for coupling the stochastic description of the Fisher-Kolmogorov-Petrovsky-Piscounov (FKPP) model with its mean-field analogue. Typically hybrid algorithms consider heuristic methods based on the cell density for determining which compartments will be updated deterministically or stochastically. In this chapter we instead considered a natural indicator of when the mean-field approximation is valid: the size of the fluctuations in the cell density. We estimated the system variance of the stochastic FKPP model using the intrinsic noise approximation. We defined a threshold on the fluctuation size based on the Van Kampen ansatz to determine the regions where the mean-field is valid and where it is not. We used this threshold to determine the deterministic and stochastic regions, and therefore the model interface: the compartment in which the stochastic and deterministic regions meet. We applied our hybrid algorithm to the stochastic FKPP and analysed how well the hybrid model approximates the stochastic model. Commonly hybrid algorithms are assessed visually and considered to perform well if the plots of the mean evolution and velocity resemble those of the stochastic model [100, 149]. We instead assessed the approximation of the hybrid algorithm by performing statistical tests between the distributions of the interface position and the velocities of the hybrid and the stochastic simulations. Using this approach we then compared systematically the performance of our hybrid model against the hybrid model introduced by Spill et al. in [149]. We showed that the Spill method was superior than our hybrid algorithm. Currently there are no standard comparison methods to assess hybrid algorithms. In this work we have considered the hybrid model to perform well if there are no statistically significant differences between the position interface and the velocity distributions of the hybrid and the stochastic model.

There are different directions to continue the work in Chapter 5. Higher order moment approximations can be used to derive a-priori indicators where the mean-field is appropriate. Defining benchmark problems and a systematic comparison methodology between hybrid models can help understand the strengths and weakness of the different methods and highlight the aspects that new methods need to address.

In Chapter 4 and Chapter 5 we analysed the convergence rates of the discretisation schemes for solving the PDEs in each chapter using the method of manufactured solutions (MMS). While this method allowed us to gauge numerically the performance of the discretisation schemes, a detailed error estimation study of the PDEs is a natural next step of the study.

## 6.1 Outlook

With the new technological advancements that allow us to make high precision imaging and the new experimental techniques that allow us to make perturbations to biological systems under controlled situations, there is a need for robust quantification methods of experimental data, development of hybrid algorithms and analysis of mathematical models to make testable predictions, all topics touched upon in this thesis. By using different mathematical, statistical and computational tools we were able to research different aspects of scratch assays and their modelling approaches. We believe that considering different frameworks (statistical, mathematical and computational methods) and a close collaboration with experimentalists and clinicians will lead to the discovery of new insights on the underlying mechanisms of the biological system under consideration.

## Appendix A

# Migration quantification method

## A.1 Classification performance plots when analysing non- smooth leading edges

We present the results of the classification tests for the three quantification methods when considering non-smooth leading edges (see Section 3.6.4).

In each plot, the colour of the circle at each parameter pair $(p_m, p_p)$ indicates the percentage of times the migration measurements associated with the parameter pair are statistically significantly different from those associated with the focal parameters $\hat{P}$. The focal parameters $\hat{P}$ are indicated by a red circle. The results reveal that the monolayer edge velocimetry method yields a better statistical classification than the other methods. We note also the performance of all three methods declines as the motility rate of the focal parameters $\hat{P}$ increases. For the area method when the proliferation probability is 0.01, the performance is worse than the other two methods but still comparable. When the proliferation probabilities are higher ($\hat{p}_p = 0.05$ and 0.09) the method can not detect statistically signifficant differences between most of the parameter pairs and the focal parameter.

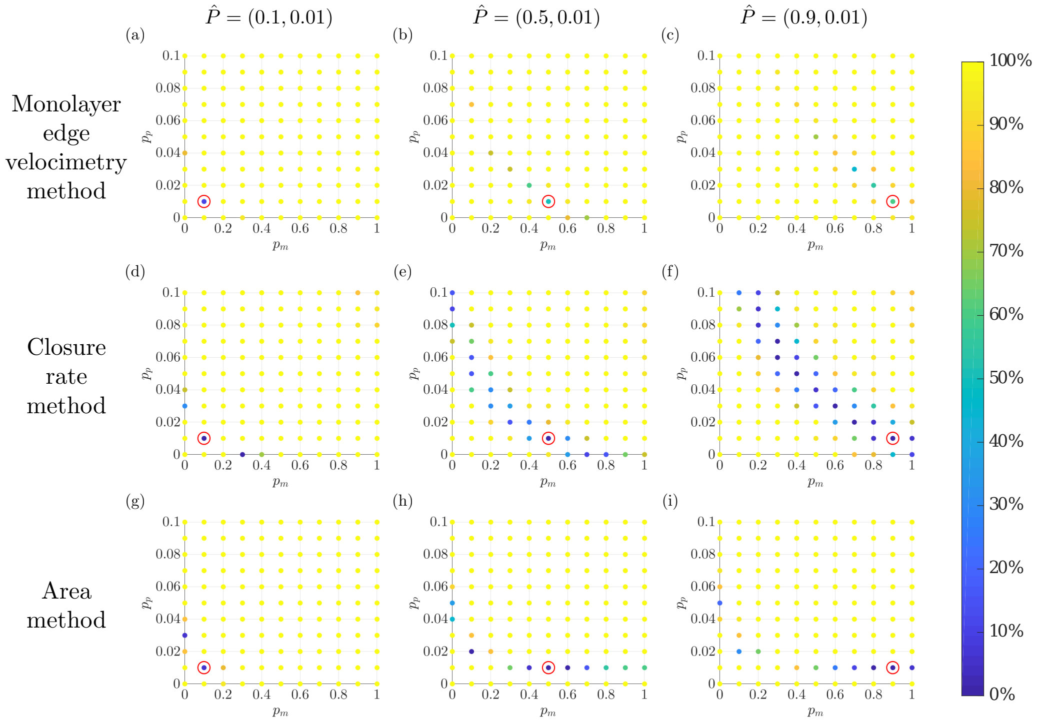


**Figure A.1:** Series of plots showing how the performance of the three quantification methods (when analysing data with non-smooth leading edges) changes as the motility rate of the focal parameters varies with fixed proliferation probability $\hat{p}_p = 0.01$.

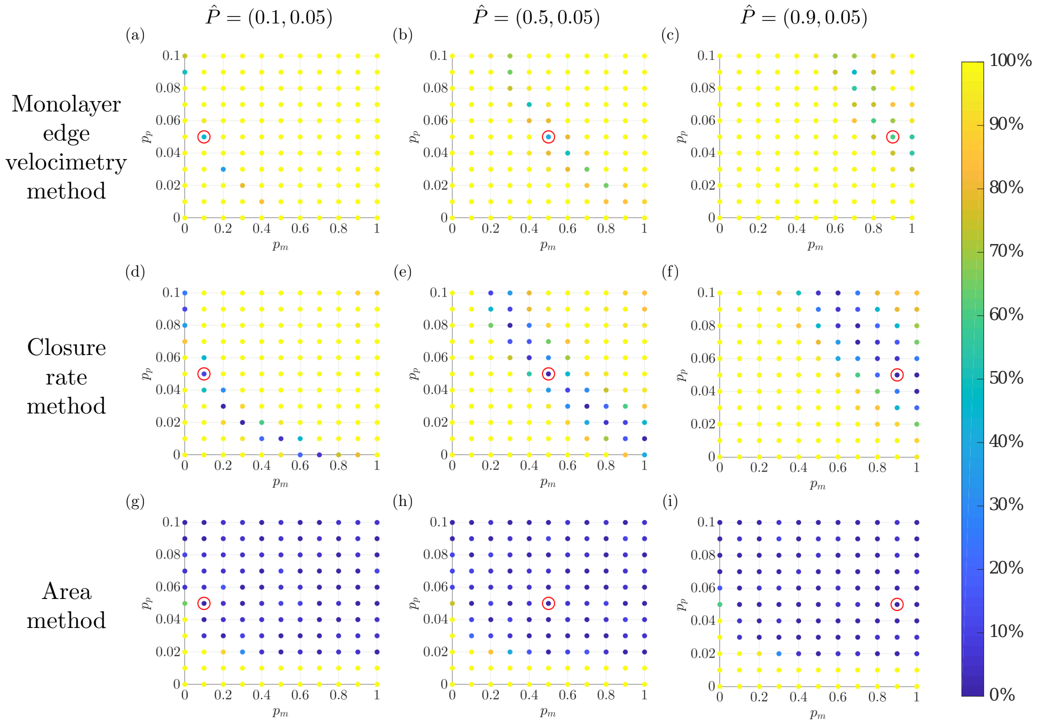


**Figure A.2:** Series of plots showing how the performance of the three quantification methods (when analysing data with non-smooth leading edges) changes as the motility rate of the focal parameters varies with fixed proliferation probability $\hat{p}_p = 0.05$.

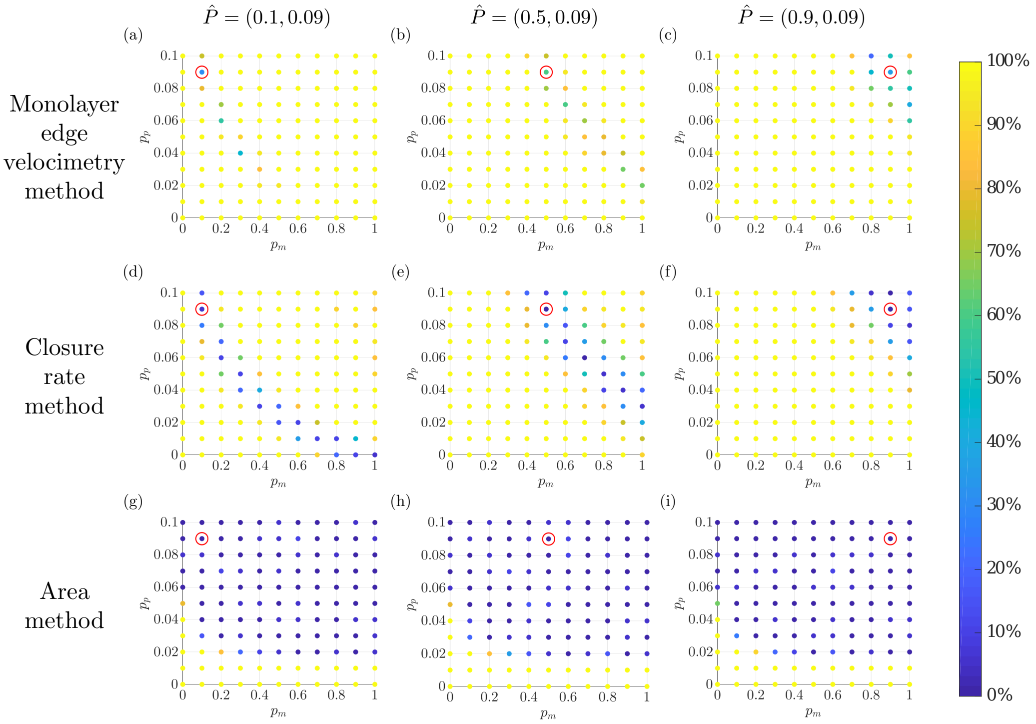


**Figure A.3:** Series of plots showing how the performance of the three quantification methods (when analysing data with non-smooth leading edges) changes as the motility rate of the focal parameters varies with fixed proliferation probability $\hat{p}_p = 0.09$.

# Appendix B

# Age-structured model

## B.1 Additional convergence rate estimation summaries

In this section we include the convergence tables from the convergence analysis of the discretisation schemes of the autonomous age-structured modelv(Eq. (4.68)) that is performed in Section 4.5.2.

### Age convergence analysis

| $h_t$ | $h_a$ | $e(h_a, h_t)$ | $r_a(h_a, h_t)$ |
|---|---|---|---|
| 1.000e-03 | 2.500e+01 | 2.503e-02 | - |
| 1.000e-03 | 1.250e+01 | 7.234e-03 | 1.79 |
| 1.000e-03 | 6.250e+00 | 2.121e-03 | 1.77 |
| 1.000e-03 | 3.125e+00 | 4.892e-04 | 2.12 |
| 1.000e-03 | 1.562e+00 | 1.197e-04 | 2.03 |
| 1.000e-03 | 7.812e-01 | 3.402e-05 | 1.81 |

**Table B.1:** Age convergence rate estimates of the discretisation scheme defined by Algorithm (4.1) when solving the system (4.51) for case 2: $a_0 \approx a_{G1/S}$ and $a_1 = 0$.

| $h_t$ | $h_a$ | $e(h_a, h_t)$ | $r_a(h_a, h_t)$ |
|---|---|---|---|
| 1.000e-03 | 2.500e+01 | 2.223e-02 | - |
| 1.000e-03 | 1.250e+01 | 6.318e-03 | 1.81 |
| 1.000e-03 | 6.250e+00 | 1.649e-03 | 1.94 |
| 1.000e-03 | 3.125e+00 | 4.187e-04 | 1.98 |
| 1.000e-03 | 1.562e+00 | 1.061e-04 | 1.98 |
| 1.000e-03 | 7.812e-01 | 2.841e-05 | 1.90 |

**Table B.2:** Age convergence rate estimates of the discretisation scheme defined by Algorithm (4.1) when solving the system (4.51) for case 3: $a_0 = 7.5$ and $a_1 = 1$.

| $h_t$ | $h_a$ | $e(h_a, h_t)$ | $r_a(h_a, h_t)$ |
|---|---|---|---|
| 1.000e-03 | 2.500e+01 | 2.223e-02 | - |
| 1.000e-03 | 1.250e+01 | 6.321e-03 | 1.81 |
| 1.000e-03 | 6.250e+00 | 1.652e-03 | 1.94 |
| 1.000e-03 | 3.125e+00 | 4.215e-04 | 1.97 |
| 1.000e-03 | 1.562e+00 | 1.089e-04 | 1.95 |
| 1.000e-03 | 7.812e-01 | 3.163e-05 | 1.78 |

**Table B.3:** Age convergence rate estimates of the discretisation scheme defined by Algorithm (4.1) when solving the system (4.51) for case 4: $a_0 = 7.5$ and $a_1 = 2$.

### Time convergence analysis

| $h_t$ | $h_a$ | $e(h_a, h_t)$ | $r_a(h_a, h_t)$ |
|---|---|---|---|
| 2.500e+00 | 2.441e-02 | 9.244e+01 | - |
| 1.250e+00 | 2.441e-02 | 3.883e+01 | 1.25 |
| 6.250e-01 | 2.441e-02 | 1.769e+01 | 1.13 |
| 3.125e-01 | 2.441e-02 | 8.488e+00 | 1.06 |
| 1.562e-01 | 2.441e-02 | 4.169e+00 | 1.03 |

**Table B.4:** Time convergence rate estimates of the discretisation scheme defined by Algorithm (4.1) when solving the system (4.51) for case 2: $a_0 \approx a_{G1/S}$ and $a_1 = 0$.

| $h_t$ | $h_a$ | $e(h_a, h_t)$ | $r_a(h_a, h_t)$ |
|---|---|---|---|
| 2.500e+00 | 2.441e-02 | 8.935e+01 | - |
| 1.250e+00 | 2.441e-02 | 3.731e+01 | 1.26 |
| 6.250e-01 | 2.441e-02 | 1.694e+01 | 1.14 |
| 3.125e-01 | 2.441e-02 | 8.115e+00 | 1.06 |
| 1.562e-01 | 2.441e-02 | 3.982e+00 | 1.03 |

**Table B.5:** Time convergence rate estimates of the discretisation scheme defined by Algorithm (4.1) when solving the system (4.51) for case 3: $a_0 = 7.5$ and $a_1 = 1$.

| $h_t$ | $h_a$ | $e(h_a, h_t)$ | $r_a(h_a, h_t)$ |
|---|---|---|---|
| 2.500e+00 | 2.441e-02 | 8.266e+01 | - |
| 1.250e+00 | 2.441e-02 | 3.335e+01 | 1.31 |
| 6.250e-01 | 2.441e-02 | 1.501e+01 | 1.15 |
| 3.125e-01 | 2.441e-02 | 7.177e+00 | 1.06 |
| 1.562e-01 | 2.441e-02 | 3.521e+00 | 1.03 |

**Table B.6:** Time convergence rate estimates of the discretisation scheme defined by Algorithm (4.1) when solving the system (4.51) for case 3: $a_0 = 7.5$ and $a_1 = 2$.

## B.2 Per capita growth rate plots

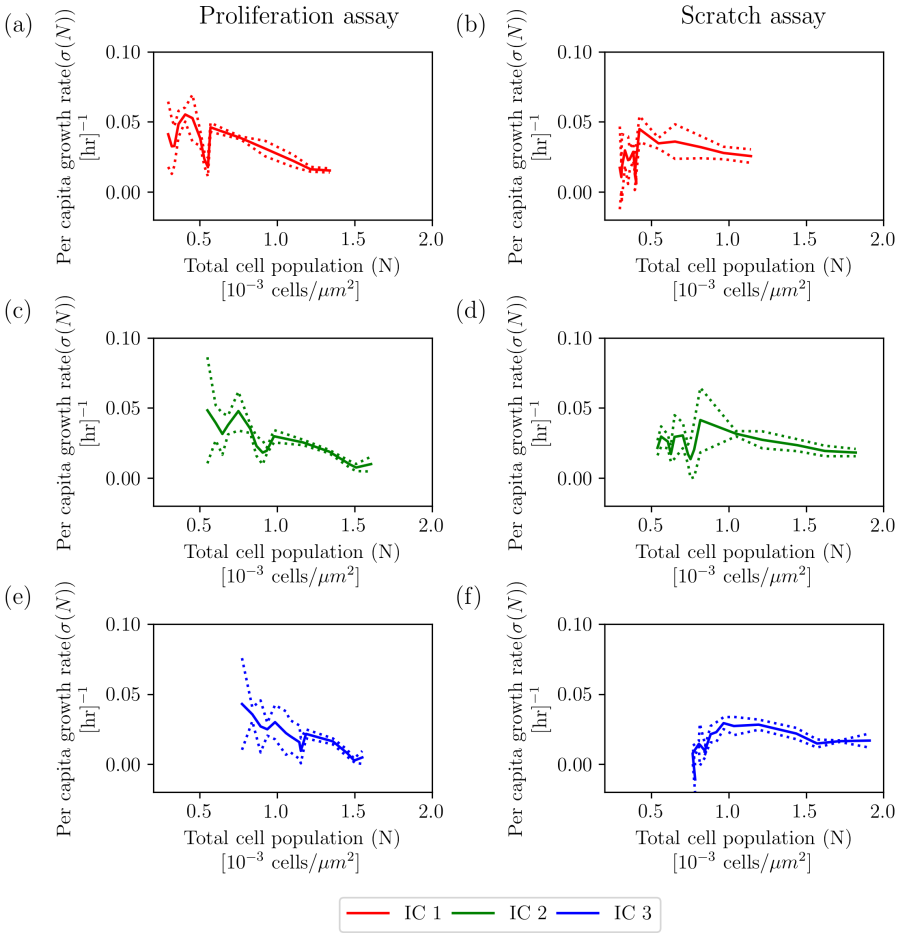


**Figure B.1:** Evolution of the mean and the standard deviation of the per capita growth rate $\sigma(N) = \frac{dN}{dt}\frac{1}{N}$ against the mean cell population, $N$, for the proliferation and scratch assays performed in [118]. The authors consider three replicates for three initial degrees of confluence. We plot the different degrees of confluence corresponding to the proliferation and scratch assays in different rows. We calculate the per capita growth rate in the same way as in [118]. A biphasic trend can be observed in the plots for the scratch assays while a decreasing linear trend can be observed for the proliferation assays.

## B.3 Sensitivity analysis of the full model

We include the plots of the sensitivity analysis of the full age-structured model (Eqs. (4.1) and (4.4)).

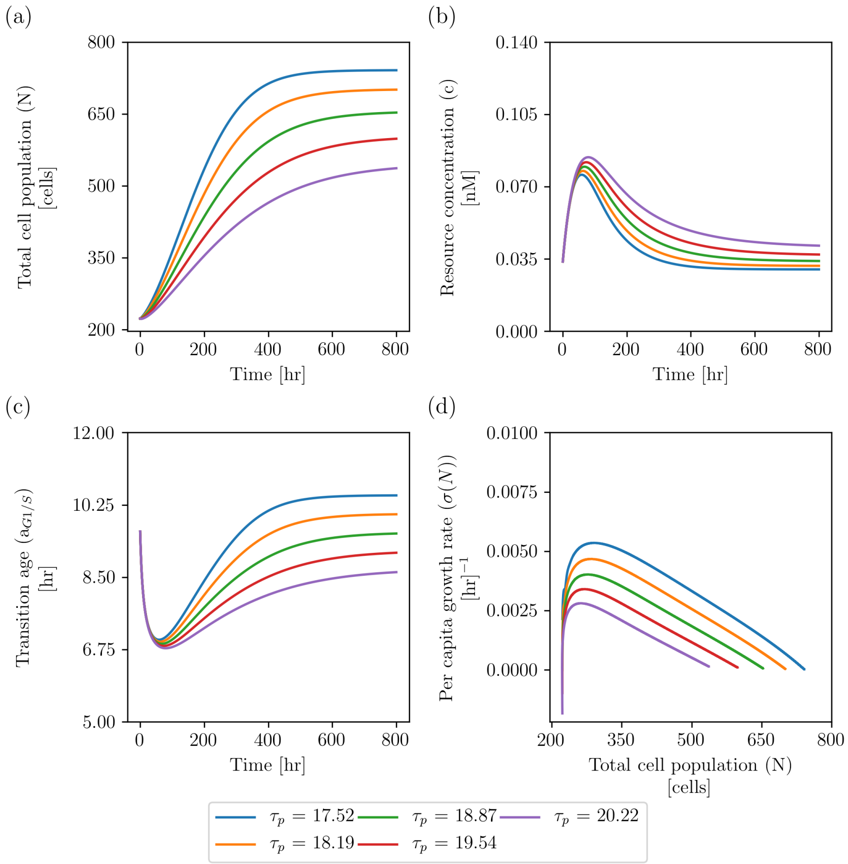


**Figure B.2:** Sensitivity analysis with respect to $\tau_p$, the inverse of the proliferation rate. The values of $\tau_p$ are specified in the legend. All other parameters are fixed at their default values stated in Table 4.5.

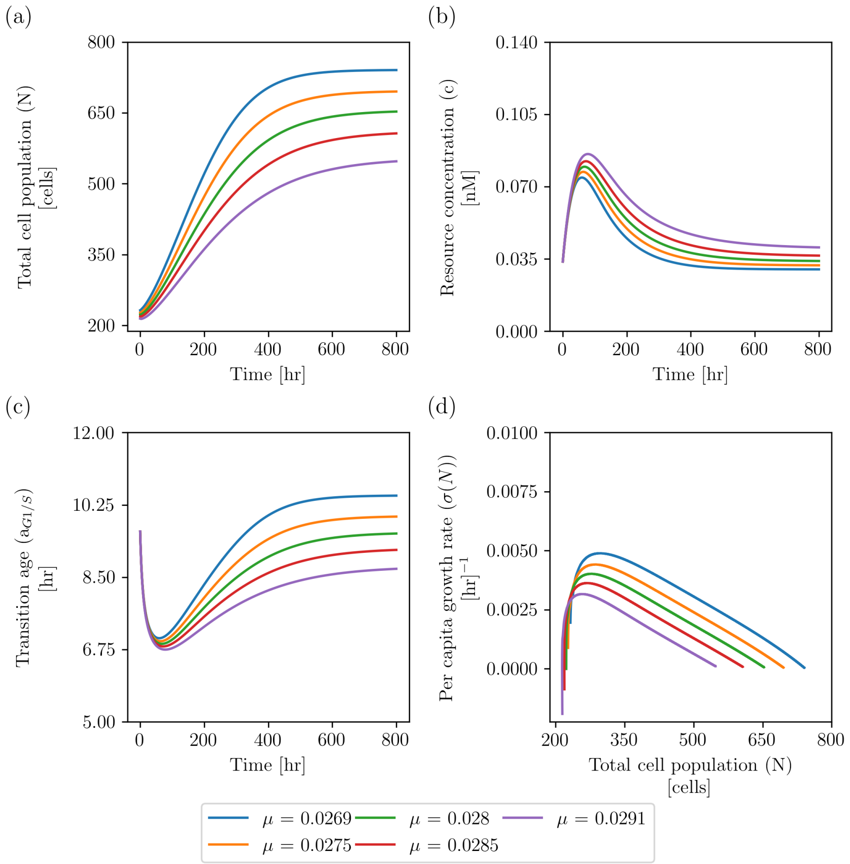


**Figure B.3:** Sensitivity analysis with respect to the death rate $\mu$. The values of $\mu$ are specified in the legend. All other parameters are fixed at their default values stated in Table 4.5.

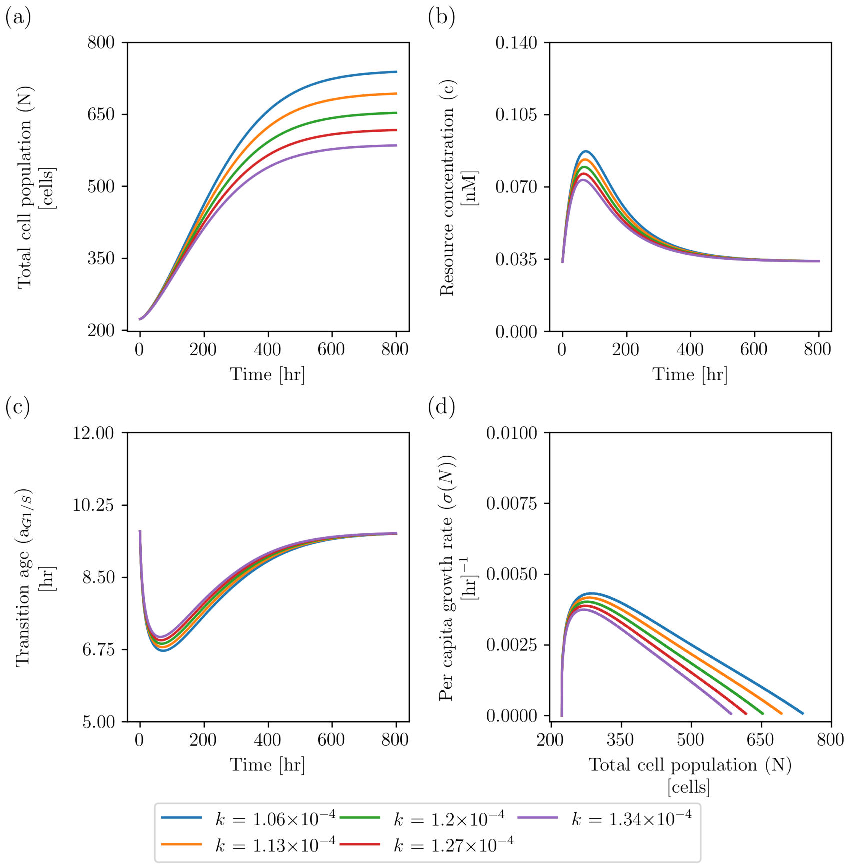


**Figure B.4:** Sensitivity analysis with respect to the resource consumption rate $k$. The values of $k$ are specified in the legend. All other parameters are fixed at their default values stated in Table 4.5.

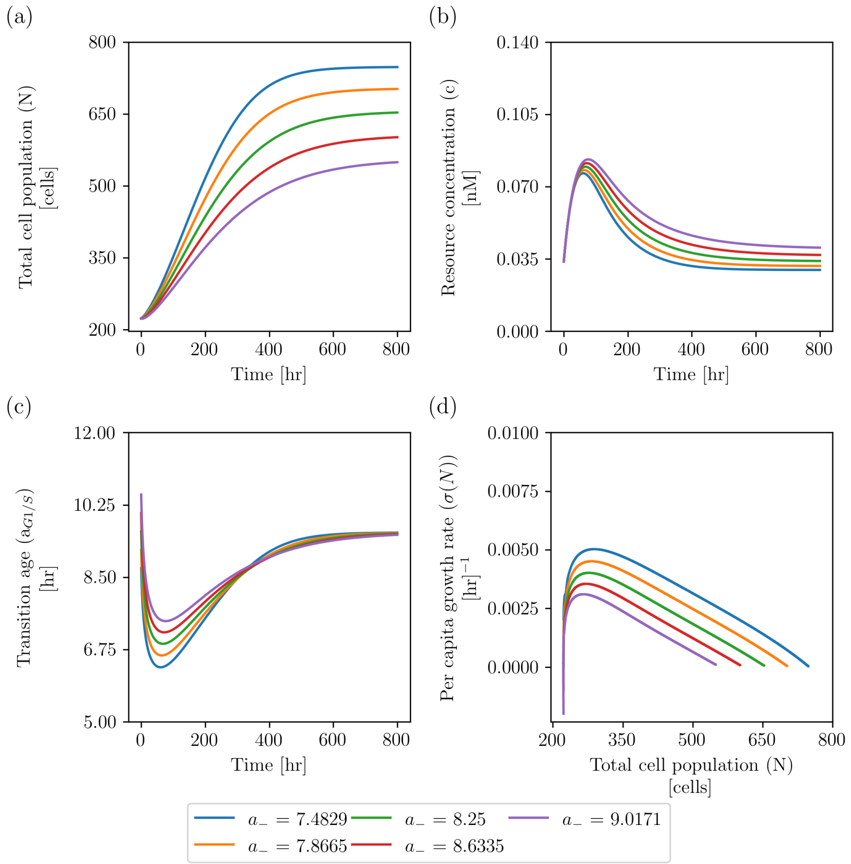


**Figure B.5:** Sensitivity analysis with respect to the scaling constant $a_-$ used to calculate the G1/S transition age (Eq. (4.2)). The values of $a_-$ are specified in the legend. All other parameters are fixed at their default values stated in Table 4.5.

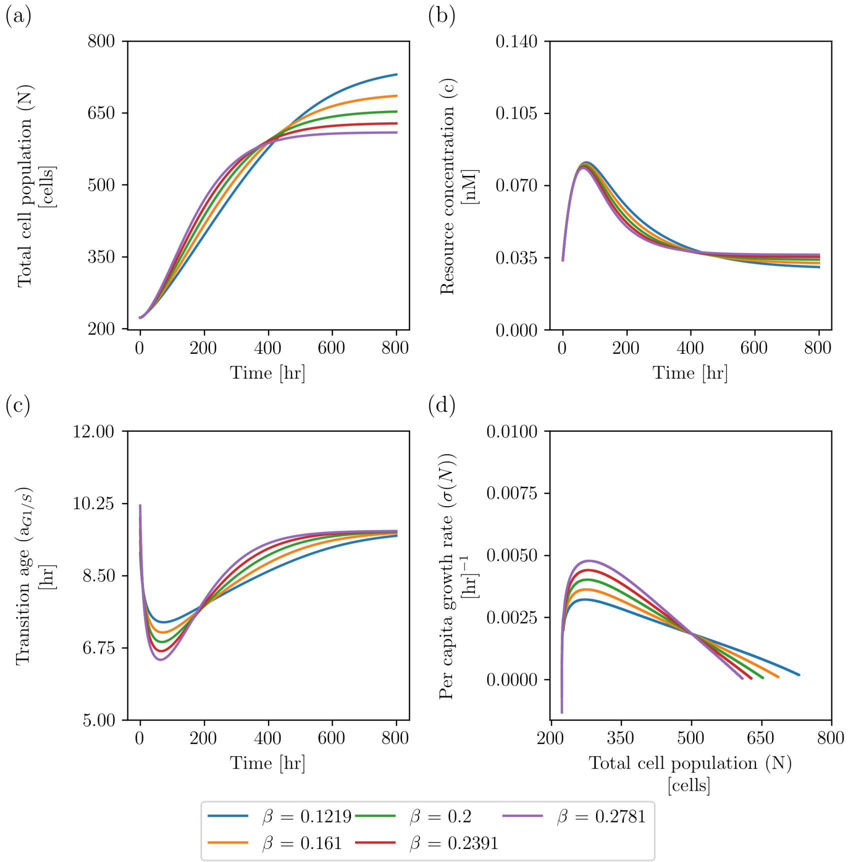


**Figure B.6:** Sensitivity analysis to the scaling constant $a_-$ to the scaling constant $a_-$ used for calculating the G1/S transition age. age (Eq. (4.2)). The values of $\beta$ are specified in the legend. All other parameters are fixed at their default values stated in Table 4.5.

**Table B.7:** Results of the parameter sensitivity analysis. In each row, for each model parameter, we specify the range of values in which the parameter was varied and the mean and standard deviation of the quantities of interest.

| Parameter | Range of values | $a_{min}$ | $c_{max}$ | $\sigma_{max}$ | $t_{cmax}$ | $t_{\sigma max}$ |
|---|---|---|---|---|---|---|
| $\tau_p$ | [17.5200,20.2200] | $6.8874 \pm 0.0811$ | $0.0798 \pm 0.0034$ | $4.0510 \times 10^{-3} \pm 1.0061 \times 10^{-3}$ | $68.6200 \pm 8.1282$ | $67.0400 \pm 7.8812$ |
| $\mu$ | [0.0269,0.0291] | $6.8882 \pm 0.1071$ | $0.0799 \pm 0.0044$ | $4.0227 \times 10^{-3} \pm 6.7155 \times 10^{-4}$ | $68.4200 \pm 7.2355$ | $66.1800 \pm 5.7846$ |
| $k$ | [0.0001,0.0001] | $6.8878 \pm 0.1322$ | $0.0800 \pm 0.0055$ | $4.0242 \times 10^{-3} \pm 2.2681 \times 10^{-4}$ | $68.0200 \pm 3.4332$ | $65.8000 \pm 3.3816$ |
| $a_-$ | [7.4829,9.0171] | $6.8847 \pm 0.4419$ | $0.0797 \pm 0.0027$ | $4.0438 \times 10^{-3} \pm 7.6138 \times 10^{-4}$ | $68.2600 \pm 6.1930$ | $65.6200 \pm 5.4961$ |
| $c_{cr}$ | [0.0202,0.0258] | $6.8855 \pm 0.1584$ | $0.0797 \pm 0.0012$ | $4.0287 \times 10^{-3} \pm 2.7157 \times 10^{-4}$ | $67.9200 \pm 2.4035$ | $65.8000 \pm 3.4322$ |
| $\beta$ | [0.1219,0.2781] | $6.9016 \pm 0.3591$ | $0.0797 \pm 0.0011$ | $4.0091 \times 10^{-3} \pm 6.1631 \times 10^{-4}$ | $68.2200 \pm 3.8739$ | $65.5600 \pm 2.2412$ |

## Appendix C

# Hybrid algorithm

## C.1 System size expansion derivation

In this section we include the proofs from Theorem 11:

*Proof.* We disregard the subscript $i$ during the derivation and consider it again after we derive the linear noise approximation. We assume the Van Kampen ansatz [215], namely, when $\Omega >> 1$ the cell number, $n$ is given by:

$$n = \Omega\phi + \Omega^{1/2}\epsilon, \tag{C.1}$$

where $\Omega$ is the system size, $\phi$ is a deterministic function and $\epsilon$ is a new random variable.

We consider as the system size the carrying capacity multiplied by the compartment size, $\Omega = Kh$. Using the value for $\Omega$ and disregarding the subscript, the master equation (5.24) can be written as follows:

$$\frac{\partial p(n,t)}{\partial t} = \lambda[\mathbb{E}^{-1} - 1]np(n,t) + \frac{\lambda}{2\Omega}[\mathbb{E}^{2} - 1]n(n-1)p(n,t). \tag{C.2}$$

By considering Eq. C.1, the fluctuations of the variable $n$ are now encapsulated in the variable $\epsilon$. We denote the probability distribution of the random variable $\epsilon$ as $\Pi$. We follow now the standard procedure to obtain the Fokker-Plank equation that describes the fluctuations of $\epsilon$ (see Section 5.2.2 ).

We consider the effect of the change of variables (C.1) on the master equation (C.2). We analyse first the effect of the change of variables on the derivatives. If we consider the variable $\epsilon$ as a function of $x = n$, then the effect on the derivative is the following:

$$\left.\frac{\partial}{\partial x}\right|_{x=n} = \Omega^{-1/2}\frac{\partial}{\partial \epsilon}.$$

Therefore the $j$-th derivatives are related by:

$$\left.\frac{\partial^j}{\partial x^j}\right|_{x=n} = \Omega^{-j/2}\frac{\partial^j}{\partial \epsilon^j}. \tag{C.3}$$

The time derivative in (C.2) is taken with constant $x = n$, therefore $\frac{dn}{dt} = 0$. This, together with Eq. (C.1), means

$$\frac{\partial \epsilon}{\partial t} = -\Omega^{1/2}\frac{\partial \phi}{\partial t}. \tag{C.4}$$

We can then describe the derivative in time of the probability distribution in terms of $\Pi$ and $\epsilon$:

$$\frac{\partial p(t)}{\partial t} = \frac{\partial \Pi}{\partial t} + \frac{\partial \Pi}{\partial \epsilon}\frac{\partial \epsilon}{\partial t} = \frac{\partial \Pi}{\partial t} - \Omega^{1/2}\frac{\partial \phi}{\partial t}\frac{\partial \Pi}{\partial \epsilon}. \tag{C.5}$$

Now we see the effect of the change of variables in the operators. Using (C.3) and employing a Taylor expansion we obtain:

$$\begin{aligned}\left[\mathbb{E}^2 - 1\right] f(n) = f(n+2) - f(n) &= \sum_{j=1}^{\infty} \frac{2^j \partial^j f}{j! \partial x^j}(x)\bigg|_{x=n} \\ = \sum_{j=1}^{\infty} \frac{2^j (\Omega)^{-j/2}}{j!} \frac{\partial^j}{\partial \epsilon^j} f(\epsilon)\bigg|_{\epsilon = \Omega^{-1/2}(n - \Omega\phi)} &= \left[\mathbb{E}^{2\Omega^{-1/2}} - 1\right] f(\epsilon). \end{aligned} \tag{C.6}$$

Therefore the functional $\left[\mathbb{E}^2 - 1\right] f(n)$ is equivalent to the functional $\left[\mathbb{E}^{2\Omega^{-1/2}} - 1\right] [f(\epsilon)]$ after the change of variables. Similarly we can see that the functional $\left[\mathbb{E}^{-1} - 1\right] [f(n)]$ is equivalent to $\left[\mathbb{E}^{-\Omega^{-1/2}} - 1\right] [f(\epsilon)]$.

Writing (C.2) in terms of $\epsilon$ after the change of variables and considering Eqs. (C.5) and (C.6), we obtain:

$$\begin{aligned}\frac{\partial \Pi}{\partial t} - \Omega^{1/2}\frac{\partial \phi}{\partial t}\frac{\partial \Pi}{\partial \epsilon} = &\overbrace{\frac{\lambda}{2\Omega}\left[2\Omega^{-1/2}\frac{\partial}{\partial \epsilon}\right][\Omega\phi + \Omega^{1/2}\epsilon][\Omega\phi + \Omega^{1/2}\epsilon - 1]\Pi}^{(C.7.1)} \\ &+ \overbrace{\frac{\lambda}{2\Omega}\left[\frac{(2\Omega^{-1/2})^2}{2}\frac{\partial^2}{\partial \epsilon^2}\right][\Omega\phi + \Omega^{1/2}\epsilon][\Omega\phi + \Omega^{1/2}\epsilon - 1]\Pi}^{(C.7.1)} \\ &+ \overbrace{\frac{\lambda}{2\Omega}O(\Omega^{-3/2})[\Omega\phi + \Omega^{1/2}\epsilon][\Omega\phi + \Omega^{1/2}\epsilon - 1]\Pi}^{(C.7.1)} \\ &\underbrace{+ \lambda\left[-\Omega^{-1/2}\frac{\partial}{\partial \epsilon} + \frac{\Omega^{-1}}{2}\frac{\partial^2}{\partial \epsilon^2} + O(\Omega^{-3/2})\right][\Omega\phi + \Omega^{1/2}\epsilon]\Pi.}_{(C.7.2)}\end{aligned} \tag{C.7}$$

We determine the value of (C.7.1) and (C.7.2) by expanding the operations:

$$\begin{aligned}(C.7.2) &= \lambda\left[-\Omega^{-1/2}\frac{\partial}{\partial \epsilon} + \frac{\Omega^{-1}}{2}\frac{\partial^2}{\partial \epsilon^2} + O(\Omega^{-3/2})\right][\Omega\phi\Pi + \Omega^{1/2}\epsilon\Pi] \\ &= \lambda\left[-\Omega^{1/2}\phi\frac{\partial \Pi}{\partial \epsilon} + \frac{\phi}{2}\frac{\partial^2 \Pi}{\partial \epsilon^2} + O(-\Omega^{-1/2}) - \frac{\partial}{\partial \epsilon}(\Pi\epsilon) + \frac{\Omega^{-1/2}}{2}\frac{\partial^2}{\partial \epsilon^2}(\Pi\epsilon)\right].\end{aligned} \tag{C.8}$$

Hence, $(C.7.1)$ is

$$\begin{aligned}
&= \frac{\lambda}{2\Omega}\left[2\Omega^{-1/2}\frac{\partial}{\partial\epsilon}\right]\left[\Omega^2\phi^2 + 2\Omega^{3/2}\phi\epsilon + \Omega\epsilon^2 - \Omega\phi - \Omega^{1/2}\epsilon\right]\Pi \\
&\quad + \frac{\lambda}{2\Omega}\left[\frac{(2\Omega^{-1/2})^2}{2}\frac{\partial^2}{\partial\epsilon^2}\right]\left[\Omega^2\phi^2 + 2\Omega^{3/2}\phi\epsilon + \Omega\epsilon^2 - \Omega\phi - \Omega^{1/2}\epsilon\right]\Pi \\
&\quad + \frac{\lambda}{2\Omega}O(\Omega^{-3/2})\left[\Omega^2\phi^2 + 2\Omega^{3/2}\phi\epsilon + \Omega\epsilon^2 - \Omega\phi - \Omega^{1/2}\epsilon\right]\Pi \\
&= \lambda\left[\Omega^{-1/2}\frac{\partial}{\partial\epsilon}\right]\left[\Omega\phi^2\Pi + 2\Omega^{1/2}\phi\epsilon\Pi + \epsilon^2\Pi - \phi\Pi - \Omega^{-1/2}\epsilon\Pi\right] \\
&\quad + \lambda\left[\Omega^{-1}\frac{\partial^2}{\partial\epsilon^2} + O(\Omega^{-3/2})\right]\left[\Omega\phi^2\Pi + 2\Omega^{1/2}\phi\epsilon\Pi + \epsilon^2\Pi - \phi\Pi - \Omega^{-1/2}\epsilon\Pi\right] \\
&= \lambda\left[\Omega^{1/2}\phi^2\frac{\partial\Pi}{\partial\epsilon} + \phi^2\frac{\partial^2\Pi}{\partial\epsilon^2} + O(\Omega^{-1/2}) + 2\phi\frac{\partial(\epsilon\Pi)}{\partial\epsilon} + 2\Omega^{-1/2}\phi\frac{\partial^2}{\partial\epsilon^2}(\epsilon\Pi)\right] \\
&\quad + \lambda\left[O(\Omega^{-1}) + \Omega^{-1/2}\frac{\partial}{\partial\epsilon}(\epsilon^2\Pi) + \Omega^{-1}\frac{\partial^2}{\partial\epsilon^2}(\epsilon^2\Pi) + O(\Omega^{-3/2}) - \Omega^{-1/2}\phi\frac{\partial\Pi}{\partial\epsilon}\right] \\
&\quad + \lambda\left[-\Omega^{-1}\phi\frac{\partial^2\Pi}{\partial\epsilon^2} + O(\Omega^{-3/2}) - \Omega^{-1}\frac{\partial}{\partial\epsilon}(\epsilon\Pi) - \Omega^{-3/2}\frac{\partial^2}{\partial\epsilon^2}(\epsilon\Pi) + O(\Omega^{-2})\right].
\end{aligned} \tag{C.9}$$

Using Eqs. (C.8) and (C.9) and keeping only the term of $\Omega$ with order $n \geq 0$, we are left with terms of order $n = 0$ and $1/2$. The terms of order $\Omega^{1/2}$ determine the evolution of $\phi$:

$$\frac{\partial\phi}{\partial t} = -\lambda\phi^2 + \lambda\phi. \tag{C.10}$$

Considering now the terms of order $\Omega^0$, we obtain the following equation:

$$\frac{\partial\Pi}{\partial t} = \left[\frac{2\lambda}{h}\phi - \lambda\right]\frac{\partial}{\partial\epsilon}(\Pi\epsilon) + \left[\frac{\lambda}{h}\phi^2 + \frac{\lambda\phi}{2}\right]\frac{\partial^2\Pi}{\partial\epsilon^2}. \tag{C.11}$$

Typically one uses the following notation:

$$A(t) := \lambda - 2\lambda\phi \qquad B(t) := 2\lambda\phi^2 + \lambda\phi. \tag{C.12}$$

Hence the linear noise approximation is given by

$$\frac{\partial\Pi}{\partial t} = -A\frac{\partial}{\partial\epsilon}(\Pi\epsilon) + \frac{B}{2}\frac{\partial^2\Pi}{\partial\epsilon^2}. \tag{C.13}$$

Using the subscript $i$ in Eqs. (C.10), (C.12) and (C.13) proves the first two equations of Theorem 11. To prove the last equation we divide (C.1) by $\Omega$

$$\frac{n}{\Omega} = \phi + \Omega^{-1/2}\epsilon \approx \epsilon. \tag{C.14}$$

where the last equality is true since $\Omega >> 1$. This proves Eq. (5.29). □

## C.2 Verification of the stochastic FKPP model factorial cumulant derivation

We verify the equation for the uncoarse-grained factorial cumulant evolution of the stochastic FKPP equation, Eq. (5.32). We consider only one compartment and then Eq. (5.32) reduces to:

$$\frac{\mathrm{d}[n^r, n^r]}{\mathrm{d}t} = 2\lambda \left[1 - \frac{2\phi^r}{h}\right] [n^r, n^r] + \left(2\lambda K \phi^r - \frac{\lambda K}{h} (\phi^r)^2\right). \tag{C.15}$$

We note that $\phi^r$ satisfies

$$\frac{\mathrm{d}\phi^r}{\mathrm{d}t} = \lambda \phi^\mu - \frac{\lambda}{h} (\phi^\mu)^2. \tag{C.16}$$

We solve Eq. (C.16) and obtain an explicit formula for $\phi^r$:

$$\varphi^r(t) = \frac{\exp(\lambda t)}{1 + W \exp(\lambda t)/h}, \tag{C.17}$$

where $W = \varphi^r(0)/(1 - \varphi^r(0)/h)$.

**Table C.1:** Summary of parameters for uncoarse-grained simulations

| Parameter | Description | Value |
|---|---|---|
| $K$ | Carrying capacity | 5000 |
| $T$ | End time | 20 |
| $h$ | Compartment size | 1 |
| $\lambda$ | Proliferation rate | 1 |
| $n_0$ | Initial number of molecules | 2000 |

Considering the parameters in Table C.1, we perform 5000 simulations of the stochastic process and compute the mean, standard deviation and factorial cumulant from the stochastic process. We then solve the ODE for the factorial cumulant, and determine the evolution of the standard deviation and the factorial cumulant using the intrinsic noise approximation. In Figure C.1 we plot the results of this comparison. We plot the mean, standard deviation and factorial cumulant according to the stochastic simulations and the intrinsic noise approximation. We observe there is a good match between the three quantities.

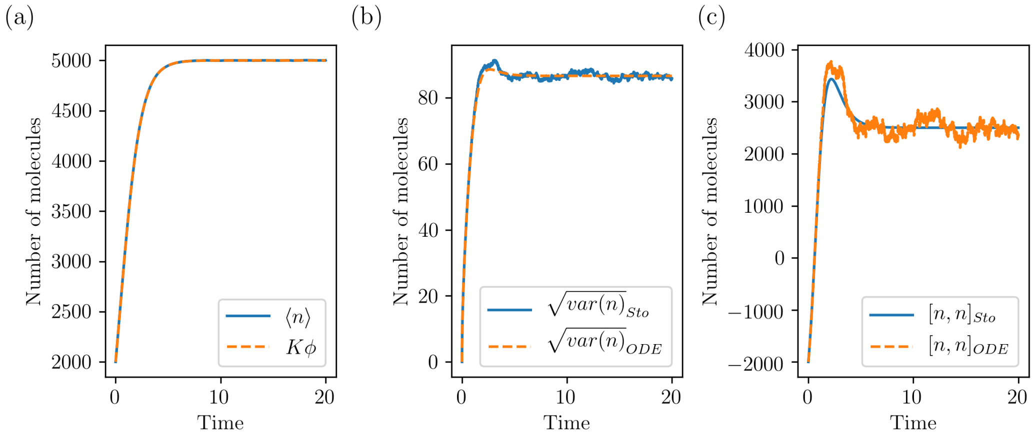


**Figure C.1:** Series of plots depicting the evolution of the uncoarse-grained stochastic FKPP equation in one compartment. We plot the evolution of a) the mean b) the standard deviation, and c) the factorial cumulant according to the stochastic simulations and the formula of the intrinsic noise approximation.

## C.3 Additional plots from the reference dynamics analysis

In this section we include additional plots that showcase the dynamics of the reference dynamics of the stochastic FKPP model.

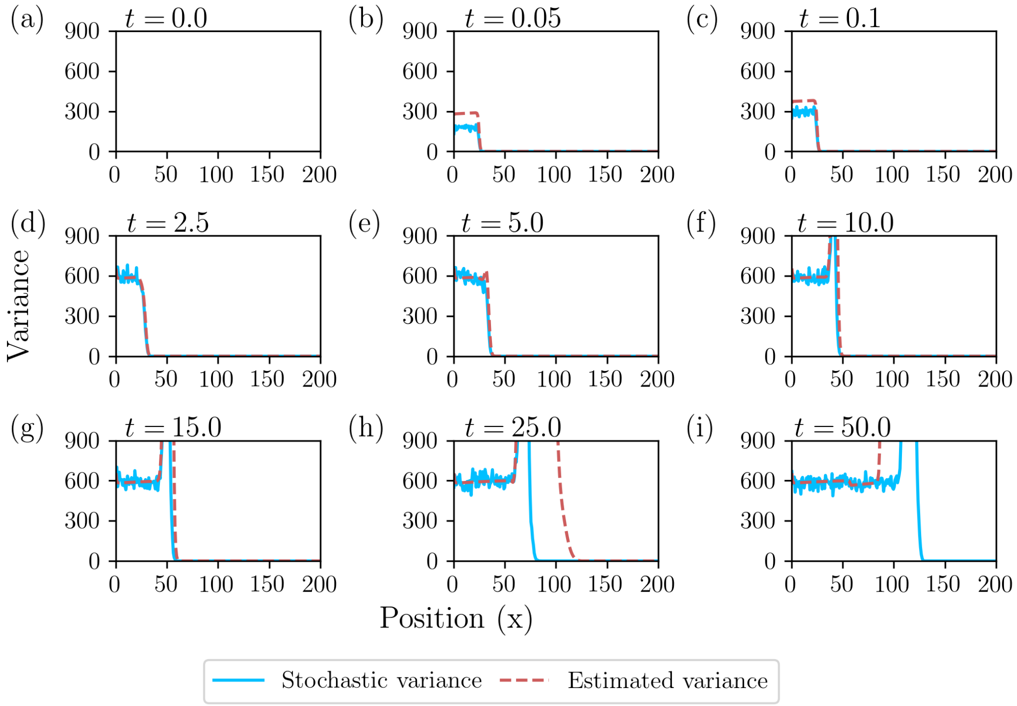


**Figure C.2:** Plots of the standard deviation of the stochastic model (equation (5.9)) and the estimated standard deviation determined from the coarse-grained factorial cumulant (equation (5.36)) at different time points. The closer the compartments are to the solution interface, the higher the standard deviation values. The estimated variance approximates well the stochastic variance in compartments far away from the interface but in the compartments close to the interface, the estimated standard deviation values are much higher.

# Bibliography

[1] L. M. Abia, O. Angulo, and J. C. López-Marcos, “Age-structured population models and their numerical solution”, *Ecological Modelling*, vol. 188, no. 1, pp. 112–136, 2005.

[2] M. J. Ablowitz and A. Zeppetella, “Explicit solutions of Fisher’s equation for a special wave speed”, *Bulletin of Mathematical Biology*, vol. 41, no. 6, pp. 835–840, 1979.

[3] A. Adan, Y. Kiraz, and Y. Baran, “Cell Proliferation and Cytotoxicity Assays.”, *Current pharmaceutical biotechnology*, vol. 17, no. 14, pp. 1213–1221, 2016.

[4] R. P. Agarwal, L. Berezansky, E. Braverman, and A. Domoshnitsky, *Nonoscillation Theory of Functional Differential Equations with Applications*. New York, NY: Springer New York, 2012.

[5] D. Agnew, J. Green, T. Brown, M. Simpson, and B. Binder, “Distinguishing between mechanisms of cell aggregation using pair-correlation functions”, *Journal of Theoretical Biology*, vol. 352, pp. 16–23, 2014.

[6] M. Ainsworth and J. T. Oden, *A posteriori error estimation in finite element analysis*, 1997.

[7] B. Alberts, D. Bray, K. Hopkins, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Essential Cell Biology (4th Edition)*. 2014.

[8] G. An, Q. Mi, J. Dutta-Moscato, and Y. Vodovotz, “Agent-based models in translational systems biology”, *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, vol. 1, no. 2, pp. 159–171, 2009.

[9] D. F. Anderson and T. G. Kurtz, *Stochastic Analysis of Biochemical Systems*. 2015.

[10] J. Arciero and D. Swigon, “Equation-based models of wound healing and collective cell migration”, in *Complex Systems and Computational Biology Approaches to Acute Inflammation*, vol. 9781461480, Springer, 2013, pp. 185–207.

[11] O. Arino, “A survey of structured cell population dynamics”, *Acta Biotheoretica*, vol. 43, no. 1-2, pp. 3–25, 1995.

[12] O. Arino and M. Kimmel, “Comparison of Approaches to Modeling of Cell Population Dynamics”, *SIAM Journal on Applied Mathematics*, vol. 53, no. 5, pp. 1480–1504, 2005.

[13] W. J. Ashby and A. Zijlstra, “Established and novel methods of interrogating two-dimensional cell migration”, *Integrative Biology (United Kingdom)*, vol. 4, no. 11, pp. 1338–1350, 2012.

[14] ——, “Established and novel methods of interrogating two-dimensional cell migration”, *Integrative Biology (United Kingdom)*, vol. 4, no. 11, pp. 1338–1350, 2012.

[15] K. I. Aycock, N. Rebelo, and B. A. Craven, “Method of Manufactured Solutions Code Verification of Elastostatic Solid Mechanics Problems in a Commercial Finite Element Solver”, *arXiv preprint arXiv:1902.07608*, 2019.

[16] A. Bairoch, “The Cellosaurus, a Cell-Line Knowledge Resource”, *Journal of biomolecular techniques : JBT*, vol. 29, no. 2, pp. 25–38, 2018.

[17] S. Banerjee and M. C. Marchetti, “Continuum models of collective cell migration”, *arXiv preprint arXiv:1805.06531*, 2018.

[18] W. Bangerth, R. Hartmann, and G. Kanschat, “Deal.II—A general-purpose object-oriented finite element library”, *ACM Transactions on Mathematical Software*, vol. 33, no. 4, 24–es, 2007.

[19] V. Bansaye and S. Méléard, *Stochastic Models for Structured Populations: Scaling Limits and Long Time Behavior*. 2015, pp. 1–107.

[20] A. Bar-Even, J. Paulsson, N. Maheshri, M. Carmi, E. O’Shea, Y. Pilpel, and N. Barkai, “Noise in protein expression scales with natural protein abundance”, *Nature Genetics*, vol. 38, no. 6, pp. 636–643, 2006.

[21] D. Barik, D. A. Ball, J. Peccoud, and J. J. Tyson, “A Stochastic Model of the Yeast Cell Cycle Reveals Roles for Feedback Regulation in Limiting Cellular Variability”, *PLoS Computational Biology*, vol. 12, no. 12, 2016.

[22] C. Beisbart and N. J. Saam, *Computer Simulation Validation: Fundamental Concepts, Methodological Frameworks, and Philosophical Perspectives*. Springer, 2018.

[23] H. C. Berg and F. Dyson, “Random Walks in Biology”, *Physics Today*, 1987.

[24] Á. Bernabé-García, S. Liarte, J. M. Moraleda, F. J. Nicolás, D. Angosto, C. L. Insausti, C. Ruiz-Cañada, and G. Castellanos, “Amniotic membrane stimulates cell migration by modulating transforming growth factor-$\beta$ signalling”, *Journal of Tissue Engineering and Regenerative Medicine*, vol. 12, no. 3, pp. 808–820, 2017.

[25] D. Bernstein, “Simulating mesoscopic reaction-diffusion systems using the Gillespie algorithm”, *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 71, no. 4, pp. 1–13, 2005.

[26] F. Billy and J. Clairambault, “Designing proliferating cell population models with functional targets for control by anti-cancer drugs”, *Discrete and Continuous Dynamical Systems - Series B*, vol. 18, no. 4, pp. 865–889, 2013.

[27] F. Billy, J. Clairambault, F. Delaunay, C. Feillet, and N. Robert, “Age-structured cell population model to study the influence of growth factors on cell cycle dynamics”, *Mathematical Biosciences and Engineering*, vol. 10, no. 1, pp. 1–17, 2012.

[28] G. Bocharov and K. P. Hadeler, “Structured population models, conservation laws, and delay equations”, *Journal of Differential Equations*, 2000.

[29] P. B. Bochev and J. Choi, “A comparative study of least-squares, SUPG and Galerkin methods for convection problems”, *International Journal of Computational Fluid Dynamics*, vol. 15, no. 2, pp. 127–146, 2001.

[30] M. Bonamente, *Statistics and Analysis of Scientific Data*. Springer, 2017, p. 323.

[31] J. Bourseguin, C. Bonet, E. Renaud, C. Pandiani, M. Boncompagni, S. Giuliano, P. Pawlikowska, H. Karmous-Benailly, R. Ballotti, F. Rosselli, and C. Bertolotto, “FANCD2 functions as a critical factor downstream of MiTF to maintain the proliferation and survival of melanoma cells”, *Scientific Reports*, vol. 6, no. 1, p. 36 539, 2016.

[32] S. Braunewell and S. Bornholdt, “Superstability of the yeast cell-cycle dynamics: Ensuring causality in the presence of biochemical stochasticity”, *Journal of Theoretical Biology*, vol. 245, no. 4, pp. 638–643, 2007.

[33] S. Brenner and R. Scott, *The mathematical theory of finite element methods*. Springer Science & Business Media, 2007, vol. 15.

[34] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Models*. Springer Science & Business Media, 1991.

[35] A. P. Browning, S. W. McCue, and M. J. Simpson, “A Bayesian Computational Approach to Explore the Optimal Duration of a Cell Proliferation Assay”, *Bulletin of Mathematical Biology*, vol. 79, no. 8, pp. 1888–1906, 2017.

[36] E. Burman, “Consistent SUPG-method for transient transport problems: Stability and convergence”, *Computer Methods in Applied Mechanics and Engineering*, vol. 199, no. 17-20, pp. 1114–1123, 2010.

[37] H. Büth, P. Luigi Buttigieg, R. Ostafe, M. Rehders, S. R. Dannenmann, N. Schaschke, H. J. Stark, P. Boukamp, and K. Brix, “Cathepsin B is essential for regeneration of scratch-wounded normal human epidermal keratinocytes”, *European Journal of Cell Biology*, vol. 86, no. 11-12, pp. 747–761, 2007.

[38] A. Q. Cai, K. A. Landman, and B. D. Hughes, “Multi-scale modeling of a wound-healing cell migration assay”, *Journal of Theoretical Biology*, 2007.

[39] ——, “Multi-scale modeling of a wound-healing cell migration assay”, *Journal of Theoretical Biology*, vol. 245, no. 3, pp. 576–594, 2007.

[40] ——, “Multi-scale modeling of a wound-healing cell migration assay”, *Journal of Theoretical Biology*, vol. 245, no. 3, pp. 576–594, 2007.

[41] Canny J., “A Computational Approach to Edge Detection”, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, vol. PAMI-8, Elsevier, 1986, pp. 679–698.

[42] G. S. Chaffey, D. J. Lloyd, A. C. Skeldon, and N. F. Kirkby, “The effect of the G1 - S transition checkpoint on an age structured cell cycle model”, *PLoS ONE*, vol. 9, no. 1, R. A. Calogero, Ed., e83477, 2014.

[43] M. A. J. Chaplain and G. Lolas, “Mathematical Modelling of Cancer Invasion of Tissue: Dynamic Heterogeneity”, *Networks and Heterogeneous Media*, 2006.

[44] K. C. Chen, L. Calzone, A. Csikasz-Nagy, F. R. Cross, B. Novak, and J. J. Tyson, “Integrative analysis of cell cycle control in budding yeast”, *Molecular Biology of the Cell*, 2004.

[45] K. C. Chen, A. Csikasz-Nagy, B. Gyorffy, J. Val, B. Novak, and J. J. Tyson, “Kinetic analysis of a molecular model of the budding yeast cell cycle”, *Molecular Biology of the Cell*, 2000.

[46] J. J. Christiansen and A. K. Rajasekaran, “Reassessing epithelial to mesenchymal transition as a prerequisite for carcinoma invasion and metastasis”, *Cancer Research*, vol. 66, no. 17, pp. 8319–8326, 2006.

[47] J. Clairambault, S. Gaubert, and T. Lepoutre, “Comparison of Perron and Floquet eigenvalues in age structured cell division cycle models”, *Mathematical Modelling of Natural Phenomena*, vol. 4, no. 3, pp. 183–209, 2008.

[48] E. A. Codling, M. J. Plank, and S. Benhamou, *Random walk models in biology*, 2008.

[49] S. M. Cohen and L. B. Ellwein, “Cell proliferation in carcinogenesis”, *Science*, vol. 249, no. 4972, pp. 1007–1011, 1990.

[50] C. C. Cook, A. Kim, S. Terao, A. Gotoh, and M. Higuchi, "Consumption of oxygen: A mitochondrial-generated progression signal of advanced cancer", *Cell Death and Disease*, vol. 3, no. 1, e258, 2012.

[51] G. G. M. Cooper, R. E. R. Hausman, and R. E. R. Hausman, *The cell: A Molecular approach*. ASM press Washington, DC, 2000, vol. 10.

[52] V. Cristini and J. Lowengrub, *Multiscale modeling of cancer: An integrated experimental and mathematical modeling approach*, 1. Cambridge University Press, 2010, vol. 9780521884, pp. 1–278.

[53] R. de la Cruz, P. Guerrero, F. Spill, and T. Alarcón, "Stochastic multi-scale models of competition within heterogeneous cellular populations: Simulation methods and mean-field analysis", *Journal of Theoretical Biology*, vol. 407, pp. 161–183, 2016.

[54] A. Csikász-Nagy, *Computational systems biology of the cell cycle*, 2009.

[55] D. Cunningham and Z. You, "In vitro and in vivo model systems used in prostate cancer research.", *Journal of biological methods*, vol. 2, no. 1, p. 17, 2015.

[56] C. W. Curtis and D. M. Bortz, "Propagation of fronts in the Fisher-Kolmogorov equation with spatially varying diffusion", *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 2012.

[57] P. D. Dale, L Olsen, P. K. Maini, and J. A. Sherratt, "Travelling waves in wound healing", *Applied Math Letters*, vol. 17, no. 5, pp. 575–580, 1995.

[58] P. D. Dale, P. K. Maini, and J. A. Sherratt, "Mathematical modeling of corneal epithelial wound healing", *Mathematical Biosciences*, vol. 124, no. 2, pp. 127–147, 1994.

[59] P Danumjaya and A. K. Pani, "Numerical Methods for the Extended Fisher-Kolmogorov (EFK) Equation", *International Journal of Numerical Analysis and Modeling*, vol. 3, no. 2, pp. 186–210, 2006.

[60] T. Das, K. Safferling, S. Rausch, N. Grabe, H. Boehm, and J. P. Spatz, "A molecular mechanotransduction pathway regulates collective migration of epithelial cells", *Nature Cell Biology*, vol. 17, no. 3, pp. 276–287, 2015.

[61] C. Deroulers, M. Aubert, M. Badoual, and B. Grammaticos, "Modeling tumor cell migration: From microscopic to macroscopic models", *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 2009.

[62] ——, "Modeling tumor cell migration: From microscopic to macroscopic models", *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 2009.

[63] P. Deuflhard, *Newton Methods for Nonlinear Problems: Affine Invariance and Adaptive Algorithms*. 2011.

[64] A. Deutsch and S. Dormann, "Cellular Automaton Modeling of Biological Pattern Formation: Characterization, Applications, and Analysis", in *Modeling and Simulation in Science, Engineering and Technology*, 2005.

[65] O Diekmann, M Gyllenberg, and A. M. D. Roos, "The Dynamics of Physiologically Structured Populations: A Mathematical Framework and Modelling Explorations", *Lecture Notes in Biomathematics*, vol. 68, p. 511, 1986.

[66] P. Domschke, D. Trucu, A. Gerisch, and M. A. Chaplain, "Structured models of cell migration incorporating molecular binding processes", *Journal of Mathematical Biology*, vol. 75, no. 6-7, pp. 1517–1561, 2017.

[67] B. Engquist, A.-K. Tornberg, and R. Tsai, “Discretization of Dirac delta functions in level set methods”, *Journal of Computational Physics*, vol. 207, no. 1, pp. 28–51, 2005.

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

[69] R. Erban and S. J. Chapman, “Stochastic modelling of reaction-diffusion processes: Algorithms for bimolecular reactions”, *Physical Biology*, 2009.

[70] T. Erneux, *Applied Delay Differential Equations*. Springer, 2009, vol. 3, pp. 1–213.

[71] R. Fabbri, F. Gularte, and D. León, “a Statistical Distance Derived From the Kolmogorov-Smirnov Test: Specification, Reference Measures (Benchmarks) and Example Uses”, *arXiv preprint arXiv:1711.00761*, 2017.

[72] G. Fairweather and J. C. López-marcos, “A box method for a nonlinear equation of population dynamics”, *IMA Journal of Numerical Analysis*, vol. 11, no. 4, pp. 525–538, 1991.

[73] R. A. Fisher, “The Wave of Advance of Advantageous Genes”, *Annals of Eugenics*, 1937.

[74] H von Foerster, “Some remarks on changing populations”, *The Kinetics of Cellular Proliferation*, pp. 382–407, 1959.

[75] D. A. Foster, P. Yellen, L. Xu, and M. Saqcena, “Regulation of G1 cell cycle progression: Distinguishing the restriction point from a nutrient-sensing cell growth checkpoint(s)”, *Genes and Cancer*, vol. 1, no. 11, pp. 1124–1131, 2010.

[76] P. Friedl, *Prespecification and plasticity: Shifting mechanisms of cell migration*, 2004.

[77] P. Friedl and D. Gilmour, “Collective cell migration in morphogenesis, regeneration and cancer”, *Nature Reviews Molecular Cell Biology*, vol. 10, no. 7, pp. 445–457, 2009.

[78] P. Gabriel, S. P. Garbett, V. Quaranta, D. R. Tyson, and G. F. Webb, “The contribution of age structure to cell population responses to targeted therapeutics”, *Journal of Theoretical Biology*, vol. 311, pp. 19–27, 2012.

[79] ——, “The contribution of age structure to cell population responses to targeted therapeutics”, *Journal of Theoretical Biology*, vol. 311, pp. 19–27, 2012.

[80] C. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences*, 2009.

[81] E. Gavagnin, M. J. Ford, R. L. Mort, T. Rogers, and C. A. Yates, “The invasion speed of cell migration models with realistic cell cycle time distributions”, *Journal of Theoretical Biology*, pp. 1–19, 2018.

[82] E. Gavagnin and C. A. Yates, “Stochastic and Deterministic Modeling of Cell Migration”, *Handbook of Statistics*, vol. 39, pp. 37–91, 2018.

[83] P. Gerlee, “The model muddle: In search of tumor growth laws”, *Cancer Research*, vol. 73, no. 8, pp. 2407–2411, 2013.

[84] ——, *The model muddle: In search of tumor growth laws*, 2013.

[85] D. T. Gillespie, “A general method for numerically simulating the stochastic time evolution of coupled chemical reactions”, *Journal of Computational Physics*, vol. 22, no. 4, pp. 403–434, 1976.

[86] ——, “The chemical Langevin equation”, *Journal of Chemical Physics*, vol. 113, no. 1, pp. 297–306, 2000.

[87] D. T. Gillespie, A. Hellander, and L. R. Petzold, “Perspective: Stochastic algorithms for chemical kinetics”, *Journal of Chemical Physics*, vol. 138, no. 17, 2013.

[88] D. T. Gillespie, L. R. Petzold, and E. Seitaridou, “Validity conditions for stochastic chemical kinetics in diffusion-limited systems”, *Journal of Chemical Physics*, vol. 140, no. 5, 2014.

[89] I. Gorshkova, D. He, E. Berdyshev, P. Usatuyk, M. Burns, S. Kalari, Y. Zhao, S. Pendyala, J. G. Garcia, N. J. Pyne, D. N. Brindley, and V. Natarajan, “Protein kinase C-$\epsilon$ regulates sphingosine 1-phosphate-mediated migration of human lung endothelial cells through activation of phospholipase D2, protein kinase C-$\zeta$, and Rac1”, *Journal of Biological Chemistry*, vol. 283, no. 17, pp. 11 794–11 806, 2008.

[90] A. Grada, M. Otero-Vinas, F. Prieto-Castrillo, Z. Obagi, and V. Falanga, “Research Techniques Made Simple: Analysis of Collective Cell Migration Using the Wound Healing Assay”, *Journal of Investigative Dermatology*, vol. 137, no. 2, e11–e16, 2017.

[91] C. D. Greenman and T. Chou, “Kinetic theory of age-structured stochastic birth-death processes”, *Physical Review E*, vol. 93, no. 1, 2016.

[92] B. T. Grenfell, O. N. Bjørnstad, and J. Kappey, “Travelling waves and spatial hierarchies in measles epidemics”, *Nature*, vol. 414, no. 6865, pp. 716–723, 2001.

[93] R. Grima, “A study of the accuracy of moment-closure approximations for stochastic chemical kinetics”, *Journal of Chemical Physics*, vol. 136, no. 15, 2012.

[94] C. Grossmann, H.-G. Roos, and M. Stynes, *Numerical Treatment of Partial Differential Equations*. Springer, 2007, vol. 154.

[95] M. E. Gurtin and R. C. Maccamy, “Non-linear age-dependent population dynamics”, *Archive for Rational Mechanics and Analysis*, vol. 54, no. 3, pp. 281–300, 1974.

[96] K. P. Hadeler and G Bocharov, “Where to put delays in population models, in particular in the neutral case”, *Canadian Applied Mathematics Quarterly*, 2003.

[97] A. Haeger, K. Wolf, M. M. Zegers, and P. Friedl, “Collective cell migration: Guidance principles and hierarchies”, *Trends in Cell Biology*, vol. 25, no. 9, pp. 556–566, 2015.

[98] J. F. Hammond and D. M. Bortz, “Analytical solutions to Fisher’s equation with time-variable coefficients”, *Applied Mathematics and Computation*, 2011.

[99] B. Han and J. Wang, “Quantifying robustness and dissipation cost of yeast cell cycle network: The funneled energy landscape perspectives”, *Biophysical Journal*, vol. 92, no. 11, pp. 3755–3763, 2007.

[100] J. U. Harrison and C. A. Yates, “Hybrid algorithm for coupling partial differential equation and compartmentbased dynamics”, *Journal of the Royal Society Interface*, vol. 13, no. 122, p. 20 160 335, 2016.

[101] L. Harvey, B. Arnold, Z. S Lawrence, l. M. Pau, B. David, and D. James, *Molecular Cell Biology (4th edition)*. 2000.

[102] J. Hasenauer, N. Jagiella, S. Hross, and F. J. Theis, “Data-Driven Modelling of Biological Multi-Scale Processes”, *Journal of Coupled Systems and Multiscale Dynamics*, 2015.

[103] J. E. Helmreich, *Statistical Analysis and Data Display: An Intermediate Course with Examples in R (2nd Edition)*, Book Review 2. Springer, 2017, vol. 76.

[104] T. Hillen and H. G. Othmer, “The diffusion limit of transport equations derived from velocity-jump processes”, *SIAM Journal on Applied Mathematics*, 2000.

[105] L. E. Hind, W. J. Vincent, and A. Huttenlocher, *Leading from the Back: The Role of the Uropod in Neutrophil Polarization and Migration*, 2016.

[106] F. Hoppenstedt, *Mathematical Theories of Populations: Demographics, Genetics and Epidemics*. SIAM, 1975.

[107] D. Horstmann, “From 1970 until present : the Keller-Segel model in chemotaxis and its consequences”, *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 2003.

[108] B. P. Ingalls, *Mathematical modeling in systems biology: an introduction*, 07. The MIT Press, 2014, vol. 51, pp. 51–3830–51–3830.

[109] C. Irons, M. J. Plank, and M. J. Simpson, “Lattice-free models of directed cell motility”, *Physica A: Statistical Mechanics and its Applications*, 2016.

[110] S. A. Isaacson, “The reaction-diffusion master equation as an asymptotic approximation of diffusion to a small target”, *SIAM Journal on Applied Mathematics*, 2009.

[111] ——, “A convergent reaction-diffusion master equation”, *Journal of Chemical Physics*, vol. 139, no. 5, 2013.

[112] S. A. Isaacson and D. Isaacson, “Reaction-diffusion master equation, diffusion-limited reactions, and singular potentials”, *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 2009.

[113] S. A. Isaacson and C. S. Peskin, “Incorporating diffusion in complex geometries into stochastic chemical kinetics simulations”, *SIAM Journal on Scientific Computing*, 2006.

[114] H. V. Jain, *Multiscale Models of VEGF-mediated Molecular Signaling Pathways in Intratumoral Angiogenesis*. ProQuest, 2008.

[115] J. Jiang, L. Li, Y. He, and M. Zhao, “Collective cell migration: Implications for wound healing and cancer invasion”, *Burns & Trauma*, vol. 1, no. 1, p. 21, 2015.

[116] W. Jin, K. Y. Lo, S. E. Chou, S. W. McCue, and M. J. Simpson, “The role of initial geometry in experimental models of wound closing”, *Chemical Engineering Science*, vol. 179, pp. 221–226, 2018.

[117] W. Jin, S. W. McCue, and M. J. Simpson, “Extended logistic growth model for heterogeneous populations”, *Journal of Theoretical Biology*, vol. 445, pp. 51–61, 2018.

[118] W. Jin, E. T. Shah, C. J. Penington, S. W. McCue, P. K. Maini, and M. J. Simpson, “Logistic Proliferation of Cells in Scratch Assays is Delayed”, *Bulletin of Mathematical Biology*, vol. 79, no. 5, pp. 1028–1050, 2017.

[119] S. T. Johnston, E. T. Shah, L. K. Chopin, D. L. Sean McElwain, and M. J. Simpson, "Estimating cell diffusivity and cell proliferation rate by interpreting IncuCyte ZOOM$^{\text{TM}}$ assay data using the Fisher-Kolmogorov model", *BMC Systems Biology*, vol. 9, no. 1, p. 38, 2015.

[120] S. T. Johnston, M. J. Simpson, and D. L. McElwain, "How much information can be obtained from tracking the position of the leading edge in a scratch assay?", *Journal of the Royal Society Interface*, vol. 11, no. 97, pp. 20 140 325–20 140 325, 2014.

[121] S. T. Johnston, M. J. Simpson, D. L. McElwain, B. J. Binder, and J. V. Ross, "Interpreting scratch assays using pair density dynamics and approximate Bayesian computation", *Open Biology*, vol. 4, no. 9, pp. 140 097–140 097, 2014.

[122] J. E. Jonkman, J. A. Cathcart, F. Xu, M. E. Bartolini, J. E. Amon, K. M. Stevens, and P. Colarusso, "An introduction to the wound healing assay using live-cell microscopy", *Cell Adhesion and Migration*, vol. 8, no. 5, pp. 440–451, 2014.

[123] H. W. Kang, L. Zheng, and H. G. Othmer, "A new method for choosing the computational cell in stochastic reaction-diffusion systems", *Journal of Mathematical Biology*, vol. 65, no. 6-7, pp. 1017–1099, 2012.

[124] S. Kar, W. T. Baumann, M. R. Paul, and J. J. Tyson, "Exploring the roles of noise in the eukaryotic cell cycle", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 16, pp. 6471–6476, 2009.

[125] E. F. Keller and L. A. Segel, "Initiation of slime mold aggregation viewed as an instability", *Journal of Theoretical Biology*, 1970.

[126] W. O. Kermack and A. G. McKendrick, "Contributions to the mathematical theory of epidemics–III. Further studies of the problem of endemicity. 1933.", *Bulletin of mathematical biology*, vol. 53, no. 1-2, pp. 89–118, 1991.

[127] B. Keyfitz and N. Keyfitz, "The McKendrick partial differential equation and its uses in epidemiology and population study", *Mathematical and Computer Modelling*, vol. 26, no. 6, pp. 1–9, 1997.

[128] E. Khain, M. Katakowski, S. Hopkins, A. Szalad, X. Zheng, F. Jiang, and M. Chopp, "Collective behavior of brain tumor cells: The role of hypoxia", *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 83, no. 3, p. 31 920, 2011.

[129] M. Y. Kim and E. J. Park, "An upwind scheme for a nonlinear model in age-structured population dynamics", *Computers and Mathematics with Applications*, vol. 30, no. 8, pp. 5–17, 1995.

[130] M. Y. Kim and E. J. Park, "Characteristic finite element methods for diffusion epidemic models with age-structured populations", *Applied Mathematics and Computation*, vol. 97, no. 1, pp. 55–70, 1998.

[131] H. Kitano, *Systems biology: A brief overview*, 2002.

[132] E. Klipp, R. Herwig, A. Kowald, C. Wierling, and H. Lehrach, *Systems Biology in Practice*. Weinheim, FRG: Wiley-VCH Verlag GmbH & Co. KGaA, 2005, pp. 1–465.

[133] M. Kot, *Elements of mathematical ecology*, 06. Cambridge University Press, 2013, vol. 39, pp. 39–3372–39–3372.

[134] I. Koutromanos, *Fundamentals of Finite Element Analysis: Linear Finite Element Analysis*. John Wiley & Sons, 2018.

[135] G. Kreitzer, F. Jaulin, and C. Espenel, *Cell Biology Assays: Essential Methods*. Butterworth-Heinemann, 2010.

[136] D. Kuzmin, *A Guide to Numerical Methods for Transport Equations*, 2010.

[137] B. Lee, N. Key, and I. Street, "The McKendrick Partial Differential Equation and its Uses in Epidemiology and Population Study 1 Introduction", *Mathematical and Computer Modelling*, vol. 26, no. 6, pp. 1–15, 1997.

[138] H. G. Lee and J. Kim, "Regularized Dirac delta functions for phase field models", *International Journal for Numerical Methods in Engineering*, vol. 91, no. 3, pp. 269–288, 2012.

[139] H. Levine and W. J. Rappel, "The physics of eukaryotic chemotaxis", *Physics Today*, vol. 66, no. 2, pp. 24–30, 2013.

[140] F. Li, T. Long, Y. Lu, Q. Ouyang, and C. Tang, "The yeast cell-cycle network is robustly designed", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 14, pp. 4781–4786, 2004.

[141] J. Li, "Persistence and extinction in continuous age-structured population models", *Computers and Mathematics with Applications*, vol. 15, no. 6-8, pp. 511–523, 1988.

[142] C.-C. Liang, A. Y. Park, and J.-L. Guan, "In vitro scratch assay: a convenient and inexpensive method for analysis of cell migration in vitro.", *Nature protocols*, vol. 2, no. 2, pp. 329–33, 2007.

[143] S. Liarte, Á. Bernabé-García, D. Armero-Barranco, and F. J. Nicolás, "Microscopy Based Methods for the Assessment of Epithelial Cell Migration During In Vitro Wound Healing", *Journal of Visualized Experiments*, no. 131, e56799, 2018.

[144] S. Lim and P. Kaldis, "Cdks, cyclins and CKIs: roles beyond cell cycle regulation", *Development*, vol. 140, no. 15, pp. 3079–3093, 2013.

[145] M. Lintz, A. Muñoz, and C. A. Reinhart-King, *The Mechanics of Single Cell and Collective Migration of Tumor Cells*, 2017.

[146] J. C. Liu, M. Zacksenhouse, A. Eisen, S. Nofech-Mozes, and E. Zacksenhaus, "Identification of cell proliferation, immune response and cell migration as critical pathways in a prognostic signature for HER2+: ER$\alpha$- breast cancer", *PLoS ONE*, vol. 12, no. 6, e0179223, 2017.

[147] P. Lötstedt, "The Linear Noise Approximation for Spatially Dependent Biochemical Networks", *Bulletin of Mathematical Biology*, vol. 81, no. 8, pp. 2873–2901, 2019.

[148] J. S. Lowengrub, H. B. Frieboes, F. Jin, Y.-L. L. Chuang, X. Li, P. MacKlin, S. M. Wise, and V. Cristini, "Nonlinear modelling of cancer: bridging the gap between cells and tumours", *Nonlinearity*, vol. 23, no. 1, R1–R91, 2010.

[149] P. K. Maini, T. Alarcon, H. Byrne, P. Guerrero, F. Spill, P. Guerrero, T. Alarcon, P. K. Maini, and H. Byrne, "Hybrid approaches for multiple-species stochastic reaction-diffusion models", *Journal of Computational Physics*, vol. 299, pp. 429–445, 2015.

[150] P. K. Maini, D. S. McElwain, and D. I. Leavesley, "Traveling Wave Model to Interpret a Wound-Healing Cell Migration Assay for Human Peritoneal Mesothelial Cells", *Tissue Engineering*, vol. 10, no. 3-4, pp. 475–482, 2004.

[151] T. Maritzen, H. Schachtner, and D. F. Legler, “On the move: endocytic trafficking in cell migration.”, *Cellular and molecular life sciences : CMLS*, vol. 72, no. 11, pp. 2119–34, 2015.

[152] M. Maška, V. Ulman, D. Svoboda, P. Matula, P. Matula, C. Ederra, A. Urbiola, T. España, S. Venkatesan, D. M. Balak, P. Karas, T. Bolcková, M. Štreitová, C. Carthel, S. Coraluppi, N. Harder, K. Rohr, K. E. Magnusson, J. Jaldén, H. M. Blau, O. Dzyubachyk, P. Křížek, G. M. Hagen, D. Pastor-Escuredo, D. Jimenez-Carretero, M. J. Ledesma-Carbayo, A. Muñoz-Barrutia, E. Meijering, M. Kozubek, and C. Ortiz-De-Solorzano, “A benchmark for comparison of cell tracking algorithms”, *Bioinformatics*, vol. 30, no. 11, pp. 1609–1617, 2014.

[153] P. Masuzzo, M. Van Troys, C. Ampe, and L. Martens, “Taking Aim at Moving Targets in Computational Cell Migration”, *Trends in Cell Biology*, vol. 26, no. 2, pp. 88–110, 2016.

[154] O. M. Matsiaka, C. J. Penington, R. E. Baker, and M. J. Simpson, “Discrete and Continuum Approximations for Collective Cell Migration in a Scratch Assay with Cell Size Dynamics”, *Bulletin of Mathematical Biology*, 2018.

[155] D. S. McElwain, J. A. McGovern, M. J. Simpson, P. Haridas, and C. J. Penington, “Quantifying rates of cell migration and cell proliferation in co-culture barrier assays reveals how skin and melanoma cells interact during melanoma spreading and invasion”, *Journal of Theoretical Biology*, vol. 423, pp. 13–25, 2017.

[156] O. Menyhárt, H. Harami-Papp, S. Sukumar, R. Schäfer, L. Magnani, O. de Barrios, and B. Győrffy, “Guidelines for the selection of functional assays to evaluate the hallmarks of cancer”, *Biochimica et Biophysica Acta - Reviews on Cancer*, vol. 1866, no. 2, pp. 300–319, 2016.

[157] J. A. J. Metz and O Diekmann, *The dynamics of physiologically structured populations*. 1986.

[158] J. Metzcar, Y. Wang, R. Heiland, and P. Macklin, “A Review of Cell-Based Computational Modeling in Cancer Biology”, *JCO Clinical Cancer Informatics*, no. 3, pp. 1–13, 2019.

[159] F. Milde, D. Franco, A. Ferrari, V. Kurtcuoglu, D. Poulikakos, and P. Koumoutsakos, “Cell Image Velocimetry (CIV): Boosting the automated quantification of cell migration in wound healing assays”, *Integrative Biology (United Kingdom)*, vol. 4, no. 11, pp. 1437–1447, 2012.

[160] F. Milner and M. Iannelli, *The Basic Approach to Age-Structured Population Dynamics*. 2017, p. 357.

[161] J. D. Murray, *Mathematical Biology I An Introduction (Third Edition)*. 2002.

[162] H. N, Kramer; A, Walzl; C, Unger; M, Rosner; G, Krupitza; M, N. Kramer, A. Walzl, C. Unger, M. Rosner, G. Krupitza, M. Hengstschläger, and H. Dolznig, “In vitro cell migration and invasion assays”, *Mutation Research - Reviews in Mutation Research*, vol. 752, no. 1, pp. 10–24, 2013.

[163] J. T. Nardini and D. M. Bortz, “Investigation of a Structured Fisher’s Equation with Applications in Biochemistry”, *SIAM Journal on Applied Mathematics*, vol. 78, no. 3, pp. 1712–1736, 2018.

[164] B. Novak and J. J. Tyson, “Numerical analysis of a comprehensive model of M-phase control in Xenopus oocyte extracts and intact embryos”, *Journal of Cell Science*, 1993.

[165] S. Nyegaard, B. Christensen, and J. T. Rasmussen, "An optimized method for accurate quantification of cell migration using human small intestine cells", *Metabolic Engineering Communications*, vol. 3, pp. 76–83, 2016.

[166] B. Ortmann, J. Druker, and S. Rocha, "Cell cycle progression in response to oxygen levels", *Cellular and Molecular Life Sciences*, vol. 71, no. 18, pp. 3569–3582, 2014.

[167] J. M. Osborne, A. G. Fletcher, J. M. Pitt-Francis, P. K. Maini, and D. J. Gavaghan, *Comparing individual-based approaches to modelling the self-organization of multicellular tissues*, 2. 2017, vol. 13, pp. 1–34.

[168] H. G. Othmer, S. R. Dunbar, and W. Alt, "Models of dispersal in biological systems", *Journal of Mathematical Biology*, 1988.

[169] ——, "Models of dispersal in biological systems", *Journal of Mathematical Biology*, vol. 26, no. 3, pp. 263–298, 1988.

[170] J. Pahle, *Biochemical simulations: Stochastic, approximate stochastic and hybrid approaches*, 2009.

[171] K. J. Painter, "Modelling cell migration strategies in the extracellular matrix", *Journal of Mathematical Biology*, vol. 58, no. 4-5, pp. 511–543, 2009.

[172] K. J. Painter and J. A. Sherratt, "Modelling the movement of interacting cell populations", *Journal of Theoretical Biology*, vol. 225, no. 3, pp. 327–339, 2003.

[173] C. J. Penington, B. D. Hughes, and K. A. Landman, "Building macroscale models from microscale probabilistic models: A general probabilistic approach for nonlinear diffusion and multispecies phenomena", *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 2011.

[174] C. J. Penington, S. W. McCue, M. J. Simpson, E. T. Shah, L. K. Chopin, W. Jin, E. T. Shah, C. J. Penington, S. W. McCue, L. K. Chopin, M. J. Simpson, E. T. Shah, L. K. Chopin, and W. Jin, "Reproducibility of scratch assays is affected by the initial degree of confluence: Experiments, modelling and model selection", *Journal of Theoretical Biology*, vol. 390, pp. 136–145, 2015.

[175] L. Perko, *Differential equations and dynamical systems*. Springer Science & Business Media, 2013, vol. 7.

[176] B. Perthame, *Transport Equations in Biology*. 2007.

[177] M. J. Plank and M. J. Simpson, *Models of collective cell behaviour with crowding effects: Comparing lattice-based and lattice-free approaches*, 2012.

[178] A. V. Ponce Bobadilla, J. Arévalo, E. Sarró, H. M. Byrne, P. K. Maini, T. Carraro, S. Balocco, A. Meseguer, and T. Alarcón, "In vitro cell migration quantification method for scratch assays", *Journal of the Royal Society Interface*, vol. 16, no. 151, p. 20 180 709, 2019.

[179] A. V. Ponce Bobadilla, T. Carraro, H. M. Byrne, P. K. Maini, and T. Alarcón, "Age Structure Can Account for Delayed Logistic Proliferation of Scratch Assays", *Bulletin of Mathematical Biology*, pp. 1–19, 2019.

[180] E. Ranzato, S. Martinotti, and B. Burlando, "Wound healing properties of jojoba liquid wax: An in vitro study", *Journal of Ethnopharmacology*, vol. 134, no. 2, pp. 443–449, 2011.

[181] P. J. Roache, "Code Verification by the Method of Manufactured Solutions", *Journal of Fluids Engineering*, vol. 124, no. 1, p. 4, 2002.

[182] G. A. Romar, T. S. Kupper, and S. J. Divito, “Research techniques made simple: Techniques to assess cell proliferation”, *Journal of Investigative Dermatology*, vol. 136, no. 1, e1–e7, 2016.

[183] A. Sakaue-Sawano and A. Miyawaki, “Visualizing spatiotemporal dynamics of multicellular cell-cycle progressions with fucci technology”, *Cold Spring Harbor Protocols*, vol. 2014, no. 5, pp. 525–531, 2014.

[184] K. Salari and P. Knupp, “Code Verification by the Method of Manufactured Solutions”, Sandia National Laboratories (SNL), Albuquerque, NM, and Livermore, CA, Tech. Rep., 2000.

[185] U. Savla, L. E. Olson, and C. M. Waters, “Mathematical modeling of airway epithelial wound closure during cyclic mechanical strain”, *Journal of Applied Physiology*, vol. 96, no. 2, pp. 566–574, 2004.

[186] D. Schnoerr, G. Sanguinetti, and R. Grima, “Validity conditions for moment closure approximations in stochastic chemical kinetics”, *Journal of Chemical Physics*, vol. 141, no. 8, 2014.

[187] ——, “Approximation and inference methods for stochastic biochemical kinetics - A tutorial review”, *Journal of Physics A: Mathematical and Theoretical*, vol. 50, no. 9, p. 093 001, 2017.

[188] M. Scott, F. J. Poulin, and H. Tang, “Approximating intrinsic noise in continuous multispecies models”, in *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, The Royal Society, vol. 467, 2011, pp. 718–737.

[189] B. G. Sengers, C. P. Please, and R. O. Oreffo, “Experimental characterization and computational modelling of two-dimensional cell spreading for skeletal regeneration”, *Journal of the Royal Society Interface*, 2007.

[190] L. Seymour, P. J. Brockwell, and R. A. Davis, *Introduction to Time Series and Forecasting.* 440. Springer, 2006, vol. 92, p. 1647.

[191] J. A. Sherratt and J. D. Murray, “Models of epidermal wound healing”, *Proceedings of the Royal Society B: Biological Sciences*, 1990.

[192] G. Shrivakshan and D. Chandrasekar, “A Comparison of various Edge Detection Techniques used in Image Processing”, *International Journal of Computer Science Issues*, vol. 9, no. 5, pp. 269–276, 2012.

[193] M. J. Simpson, W. Jin, S. T. Vittadello, T. A. Tambyah, J. M. Ryan, G. Gunasingh, N. K. Haass, and S. W. McCue, “Stochastic models of cell invasion with fluorescent cell cycle indicators”, *Physica A: Statistical Mechanics and its Applications*, vol. 510, pp. 375–386, 2018.

[194] M. J. Simpson, K. A. Landman, and B. D. Hughes, “Cell invasion with proliferation mechanisms motivated by time-lapse data”, *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 18, pp. 3779–3790, 2010.

[195] M. J. Simpson, K. K. Treloar, B. J. Binder, P. Haridas, K. J. Manton, D. I. Leavesley, D. L. S. McElwain, and R. E. Baker, “Quantifying the roles of cell motility and cell proliferation in a circular barrier assay | Journal of The Royal Society Interface”, *Rsif.Royalsocietypublishing.Org*, vol. 10, no. 82, p. 20 130 007, 2013.

[196] C. A. Smith and C. A. Yates, “Spatially extended hybrid methods: A review”, *Journal of the Royal Society Interface*, vol. 15, no. 139, 2018.

[197] H. Smith, *An Introduction to Delay Differential Equations with Applications to the Life Sciences*, ser. Texts in Applied Mathematics. New York, NY: Springer New York, 2011, vol. 57, p. 178.

[198] J. A. Smith, D. P. Winslow, and P. S. Rudland, "Different growth factors stimulate cell division of rat mammary epithelial, myoepithelial, and stromal cell lines in culture", *Journal of Cellular Physiology*, vol. 119, no. 3, pp. 320–326, 1984.

[199] A. Stéphanou and V. Volpert, "Hybrid Modelling in Biology: a Classification Review", *Mathematical Modelling of Natural Phenomena*, vol. 11, no. 1, pp. 37–48, 2015.

[200] R. Steuer, "Effects of stochasticity in models of the cell cycle: From quantized cycle times to noise-induced oscillations", *Journal of Theoretical Biology*, vol. 228, no. 3, pp. 293–301, 2004.

[201] D. Stichel, A. M. Middleton, B. F. Müller, S. Depner, U. Klingmüller, K. Breuhahn, and F. Matthäus, "An individual-based model for collective cancer cell migration explains speed dynamics and phenotype variability in response to growth factors", *Systems Biology and Applications*, vol. 3, no. 1, 2017.

[202] E. B. Stukalin, I. Aifuwa, J. S. Kim, D. Wirtz, and S. X. Sun, "Age-dependent stochastic models for understanding population fluctuations in continuously cultured cells", *Journal of the Royal Society Interface*, vol. 10, no. 85, 2013.

[203] P. S. Swain, M. B. Elowitz, and E. D. Siggia, "Intrinsic and extrinsic contributions to stochasticity in gene expression", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 20, pp. 12 795–12 800, 2002.

[204] T. Székely and K. Burrage, "Stochastic simulation in systems biology", *Computational and Structural Biotechnology Journal*, vol. 12, no. 20-21, pp. 14–25, 2014.

[205] K. Talkenberger, E. Ada Cavalcanti-Adam, A. Voss-Böhme, and A. Deutsch, "Amoeboid-mesenchymal migration plasticity promotes invasion only in complex heterogeneous microenvironments", *Scientific Reports*, 2017.

[206] W. Y. Tan, *Stochastic models with applications to genetics, cancers, AIDS and other biomedical systems*, ser. Series on Concrete and Applicable Mathematics 4. World Scientific, 2002, vol. 4, xv, 441 p.

[207] G. Topman, O. Sharabani-Yosef, and A. Gefen, "A standardized objective method for continuously measuring the kinematics of cultures covering a mechanically damaged site", *Medical Engineering and Physics*, vol. 34, no. 2, pp. 225–232, 2012.

[208] P. Tracqui, "Biophysical models of tumour growth", *Reports on Progress in Physics*, vol. 72, no. 5, p. 056 701, 2009.

[209] X. Trepat, Z. Chen, and K. Jacobson, "Cell migration", *Comprehensive Physiology*, vol. 2, no. 4, pp. 2369–2392, 2012.

[210] E. Trucco, "Mathematical models for cellular systems the von Foerster equation. Part I", *The Bulletin of Mathematical Biophysics*, vol. 27, no. 3, pp. 285–304, 1965.

[211] ——, "Mathematical models for cellular systems. The von foerster equation. Part II", *The Bulletin of Mathematical Biophysics*, vol. 27, no. 4, pp. 449–471, 1965.

[212] S. Turner, "Using cell potential energy to model the dynamics of adhesive biological cells", *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 2005.

[213] J. J. Tyson and B. Novák, *Models in biology: lessons from modeling regulation of the eukaryotic cell cycle*, 2015.

[214] M. Ullah and O. Wolkenhauer, "Investigating the two-moment characterisation of subcellular biochemical networks", *Journal of Theoretical Biology*, vol. 260, no. 3, pp. 340–352, 2009.

[215] N. G. Van Kampen, "Fluctuations in Continuous Systems", in *Stochastic Processes in Physics and Chemistry*, AIP, vol. 27, AIP, 2007, pp. 363–382.

[216] ——, *Stochastic Processes in Physics and Chemistry*. 2007.

[217] ——, "A Power Series Expansion of the Master Equation", *Canadian Journal of Physics*, vol. 39, no. 4, pp. 551–567, 2011.

[218] P. Van Liedekerke, M. M. Palm, N. Jagiella, and D. Drasdo, "Simulating tissue mechanics with agent-based models: concepts, perspectives and some novel results", *Computational Particle Mechanics*, 2015.

[219] V. Vezhnevets, "{\textquotedblleft} {GrowCut} {\textquotedblright} - Interactive Multi-Label N-D Image Segmentation By Cellular Automata", in *proc. of Graphicon*, Citeseer, vol. 1, 2004, pp. 150–156.

[220] S. T. Vittadello, S. W. McCue, G. Gunasingh, N. K. Haass, and M. J. Simpson, "Mathematical Models for Cell Migration with Real-Time Cell Cycle Dynamics", *Biophysical Journal*, vol. 114, no. 5, pp. 1241–1253, 2018.

[221] S. T. Vittadello, S. W. Mccue, G. Gunasingh, N. K. Haass, and M. J. Simpson, "Mathematical models incorporating a multi-stage cell cycle explain synchronisation in proliferation experiments", *bioRxiv*, p. 557 702, 2019.

[222] P. Waage and C. M. Gulberg, "Studies concerning affinity", *Journal of Chemical Education*, vol. 63, no. 12, p. 1044, 2009.

[223] D. C. Walker, G Hill, S. M. Wood, R. H. Smallwood, and J Southgate, "Agent-Based Computational Modeling of Wounded", *IEEE Transactions on nanobioscience*, vol. 3, no. 3, pp. 153–163, 2004.

[224] D. C. Walker, J. Southgate, G. Hill, M. Holcombe, D. R. Hose, S. M. Wood, S. Mac Neil, and R. H. Smallwood, "The epitheliome: Agent-based modelling of the social behaviour of cells", *BioSystems*, vol. 76, no. 1-3, pp. 89–100, 2004.

[225] M. N. Walter, K. T. Wright, H. R. Fuller, S. MacNeil, and W. E. Johnson, "Mesenchymal stem cell-conditioned medium accelerates skin wound healing: An in vitro study of fibroblast and keratinocyte scratch assays", *Experimental Cell Research*, vol. 316, no. 7, pp. 1271–1281, 2010.

[226] G. F. Webb, *Theory of Nonlinear Age-Dependent Population Dynamics (Chapman & Hall Pure and Applied Mathematics)*. M. Dekker, 1985, p. 294.

[227] R. Wedlich-Soldner and R. Li, "Spontaneous cell polarization: Undermining determinism", *Nature Cell Biology*, vol. 5, no. 4, pp. 267–270, 2003.

[228] N. A. Whalley, S. Walters, and K. Hammond, *Molecular Cell Biology*. Scientific American Books New York, 2018, vol. 2, pp. 37–49.

[229] D. J. Wilkinson, *Stochastic Modelling for Systems Biology, Third Edition*. CRC press, 2018, pp. 1–330.

[230] A. Zaritsky, S. Natan, D. Kaplan, E. Ben-Jacob, and I. Tsarfaty, “Live time-lapse dataset of in vitro wound healing experiments”, *GigaScience*, vol. 4, no. 1, p. 8, 2015.

[231] F. Ziebert, J. Löber, and I. S. Aranson, “Macroscopic Model of Substrate-Based Cell Motility”, in, 2016, pp. 1–67.