Model complexity in biology and bioengineering

Alexander Paul Browning
BMath, MPhil

under the supervision of
Professor Matthew Simpson and Professor Kevin Burrage

Submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy

School of Mathematical Sciences
Faculty of Science
Queensland University of Technology

2021

"This model will be a simplification and an idealization, and consequently a falsification."

— Alan Turing (1952)

Authenticity

In accordance with the requirements of the degree of Doctor of Philosophy at the Queensland University of Technology, I present the following thesis entitled,

Model complexity in biology and bioengineering.

This work was performed under the supervision of Professor Matthew Simpson and Professor Kevin Burrage. I declare that the work submitted in this thesis is my own, except as acknowledged in the text and footnotes, and has not been previously submitted for a degree at the Queensland University of Technology or any other institution.

Signed,

Alexander P Browning
August 27, 2021

Abstract

Mathematical modelling is fast becoming pervasive throughout the life sciences. In biology and bioengineering, mathematical and statistical analysis is the basis for experimental design and interpretation, providing an understanding of biological systems that enables their control and manipulation. Tailoring mathematical and experimental complexity to the biological question of interest is crucial to avoid issues relating to parameter identifiability. While tools to diagnose these issues are well-developed for ordinary differential equation models, there is very little guidance for the stochastic models required to capture fundamental elements of many biological processes.

In the first part of this thesis, we develop data-focused stochastic models to describe collective behaviour in cell biology. Firstly, we apply a stochastic spatial model of cell motility to identify cell-to-cell interactions, representing complex interactions using simple kernel functions that allow for parameter identifiability. Secondly, we model bacteria populations exposed to environmental volatility, describing complex individual-level decision making using control theory. Lastly, we review and develop parameter identifiability analysis tools for stochastic differential equation models in biology.

Quantifying cell behaviour using mathematical modelling enables their control and exploitation in new areas of application. In the second part of this thesis, we incorporate biological systems in new materials. First, we develop and analyse a mathematical model that describes energy induction into a signal-transmitting metamaterial through a biological subsystem. Secondly, we calibrate a mathematical model of tissue growth in artificially constructed scaffolds to answer fundamental biological questions. Here, parameter identifiability and model validation are essential to achieving control over cell behaviour in future medical applications that replace synthetic grafts with bioengineered, patient-derived tissue.

Overall, this thesis develops new methodologies across a range of applications in biology and bioengineering. We conclude by discussing how we enable and inspire future research in mathematical and experimental biology.

Acknowledgements

First, I thank my supervisor and mentor, Professor Matthew Simpson. Mat, I am indebted to you for your encouragement and support, your patience and your humour, your knowledge and your guidance. I thank you for the opportunities you provided throughout my candidature and, of course, your timely and swift feedback. Without your guidance, this research would not have been possible.

Second, I thank my associate supervisor, Professor Kevin Burrage. Kevin, I thank you for helping me think beyond what I thought possible and always encouraging my ideas. I extend my thanks to the researchers with which I have collaborated: Professors Ruth Baker, Michael Plank, Mia Woodruff and Nikolas Haass, Doctors Francis Woodhouse, Wang Jin, David Warne, Oliver Maclaren, Pascal Buenzli, Mark Allenby, Tarunendu Mapder and Chris Baker, and Mr Jesse Sharp and Mr Matthew Lanaro. I thank you all for your feedback, your time and your commitment, your guidance and your wisdom. Thank you for broadening my horizons and showing me new perspectives.

I thank and acknowledge the funding I have received. I thank the Australian Research Council for my scholarship and Matthew Simpson and Kevin Burrage for my top-up. I thank the QUT School of Mathematical Sciences and the Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers for funding to travel. Funding that grew my network and allowed me to share my research.

I thank my fellow PhD students and friends. In particular to Ryan, Jesse, David, Maud, Jin, Nick and Gency, I thank you for your friendship and help, helping me to realise only by collaborating and working together can we do our best science. I thank my family for their love, patience, and continued belief in me.

Last and most importantly, I thank my wife, Raiha. I thank you, Raiha, for your patience, your guidance, your creativity, your support and most of all your love and kindness throughout all these years of study.

List of abbreviations

ABC   Approximate Bayesian computation.

FIM   Fisher information matrix.

IBM   Individual based model.

LNA   Linear noise approximation.

MCMC   Markov-chain Monte Carlo.

MPP   Merton's portfolio problem.

ODE   Ordinary differential equation.

PDE   Partial differential equation.

SDE   Stochastic differential equation.

SMC   Sequential Monte Carlo.

# Contents

Chapter 1

Introduction

1.1 Overview

Mathematical modelling is fast becoming pervasive throughout the life sciences. In biology and bioengineering, mathematical and statistical analysis form the basis for experimental design, interpretation and optimisation. Mathematical models provide an understanding of biological systems that enables their control and manipulation in applications including drug development [1, 2], cancer therapies [3–6] and material science [7–9].

Tailoring model complexity to the availability of experimental data and the biological questions of interest is crucial. Typical modelling goals include predicting future outcomes and estimating parameters that cannot be directly measured. Simply increasing the number of mechanisms captured by a model can often lead to a quantitatively better fit, to the detriment of predictive power and mechanistic insight [10]. We illustrate such issues of *over-fitting* in figure 1.1 by calibrating noisy data generated from the quadratic model $y = x^2$ to a sixth-order polynomial $y = \sum_{i=0}^{6} c_i x^i$. Clearly, the higher-order polynomial—the more complex model—produces an excellent fit, but any attempt to gain mechanistic insight or extrapolate to produce predictions will fail. Parameter estimates and approximate confidence intervals in figure 1.1b reveal a high-level of uncertainty in the calibrated model: our estimate of $c_2 = -10.4$ is only identifiable to within several orders of magnitude above and below the true value of $c_2 = 1$ (95% confidence interval $(-76.7, 55.9)$). If we instead work with a simpler model by constraining $c_i = 0$ for $i > 2$, precision in our estimate of $c_2 = 1.51$ increases (95% confidence interval: $(0.96, 2.06)$), as has the ability of the model to make predictions.

The question of model complexity in mathematical biology is particularly prevalent in the study of collective behaviour, where behaviours of individual cells (or molecules) give rise to complex collective dynamics. Collective dynamics underly many biological processes; from those that are beneficial to human health, such as bone regrowth [11–13] and bacteria in the gut microbiota [14], to pathological processes such as cancer progression [15] and bacterial infections [16]. Similar behaviours are seen in many other phenomena outside biology: in economics, for example, the individual behaviours of investors aggregate into forces that move the stock market [17]. Complicating analysis, phenomena that underly these processes occur across many levels of organisation, from the single cell to the tissue level [18].

While advances in computational modelling for systems biology have recently made whole cell models possible [19], obtaining reliable parameter estimates can require data from hundreds of experiments [20]. In studying these dynamics, model complexity must relate directly to the level of detail available in experimental

(a)

(b)
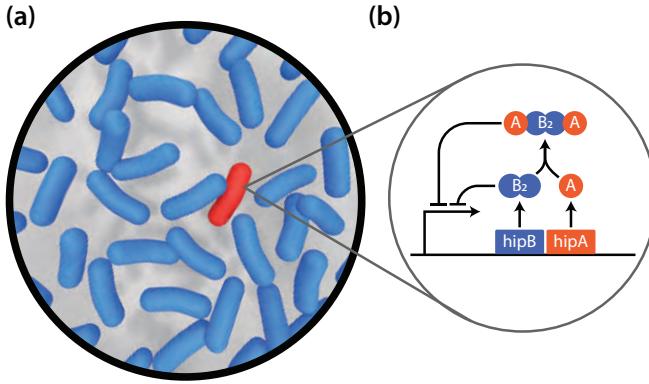
▲ Figure 1.1. (a) Noisy data (red) generated from the quadratic model $y = x^2$ (orange) and calibrated to a sixth order polynomial (blue). (b) Estimates and 95% confidence intervals for each parameter (red) compared to the true values (black).

data. For example, colonies of otherwise identical bacteria invest in a slow-growing subpopulation, called persisters, that ensure the long-term survival of the colony in noisy environments (figure 1.2a). Bacteria achieve this by entering the dormant persister state in response to environmental stimuli. At the molecular level, this process is regulated by gene expression. Feng and colleagues [21] describe the production of the two proteins present in high persistence (hip) *Escherichia coli*, HipA and HipB, encoded by the *hipA* and *hipB* genes, respectively (figure 1.2b). These dynamics give rise to the ordinary differential equation (ODE) model [21]

$$\begin{cases} \dfrac{\mathrm{d}A_T}{\mathrm{d}t} = \dfrac{\alpha\beta}{1 + \left(\frac{B_2 + \theta A_2 B_2}{K_{B_2}}\right)^4} f_1(A_f) - \gamma_1 A_f - f_2(A_f) A_T d, \\[4ex] \dfrac{\mathrm{d}B_T}{\mathrm{d}t} = \dfrac{\alpha\beta}{1 + \left(\frac{B_2 + \theta A_2 B_2}{K_{B_2}}\right)^4} f_1(A_f) N - \gamma_2 B_f - 2\gamma_3 B_2 - f_2(A_f) B_T d. \end{cases} \tag{1.1}$$

Here, A_T and B_T are the total concentrations of HipA and HipB, which are related to the concentrations of the complex $A_2 B_2$ and the free molecules and dimers A_f, B_f, A_2 and B_2 through a non-linear relationship determined by two additional parameters K_B and $K_{A_2 B_2}$. Therefore, these dynamics are governed by a system containing 11 unknown parameters and two unknown Hill functions f_1 and f_2.

Typical experiments of bacterial persistence report noisy measurements of the number of regular and persister cells [16]—a statistic of the *population*, not the molecular dynamics being modelled. While equation (1.1) describes the molecular dynamics that give rise to population-level behaviour, it is not clear whether the parameters in equation (1.1) can be estimated from typically reported experimental data. The complex molecular dynamics give rise to phenotypic switching at the population level, where regular cells enter a dormant persister state with rate u, and revert with rate v. We denote the concentration of regular cells and

◀ Figure 1.2. (a) Bacteria (blue) protect themselves against environmental uncertainty by maintaining a small subpopulation of dormant *persisters* (red). (b) At the molecular level, bacterial persistence is regulated by gene expression [21].

persister cells as $r(t)$ and $p(t)$, respectively. The far simpler system

$$\begin{cases} \dfrac{\mathrm{d}r}{\mathrm{d}t} = -ur + vp, \\[2mm] \dfrac{\mathrm{d}p}{\mathrm{d}t} = ur - vp, \end{cases} \tag{1.2}$$

sacrifices molecular-level detail for parameter identifiability, and is routinely applied to interpret persister dynamics in experimental studies [16, 22]. One drawback of this simpler model is that it does not allow for the adaptive production of persisters, and we address this in this thesis.
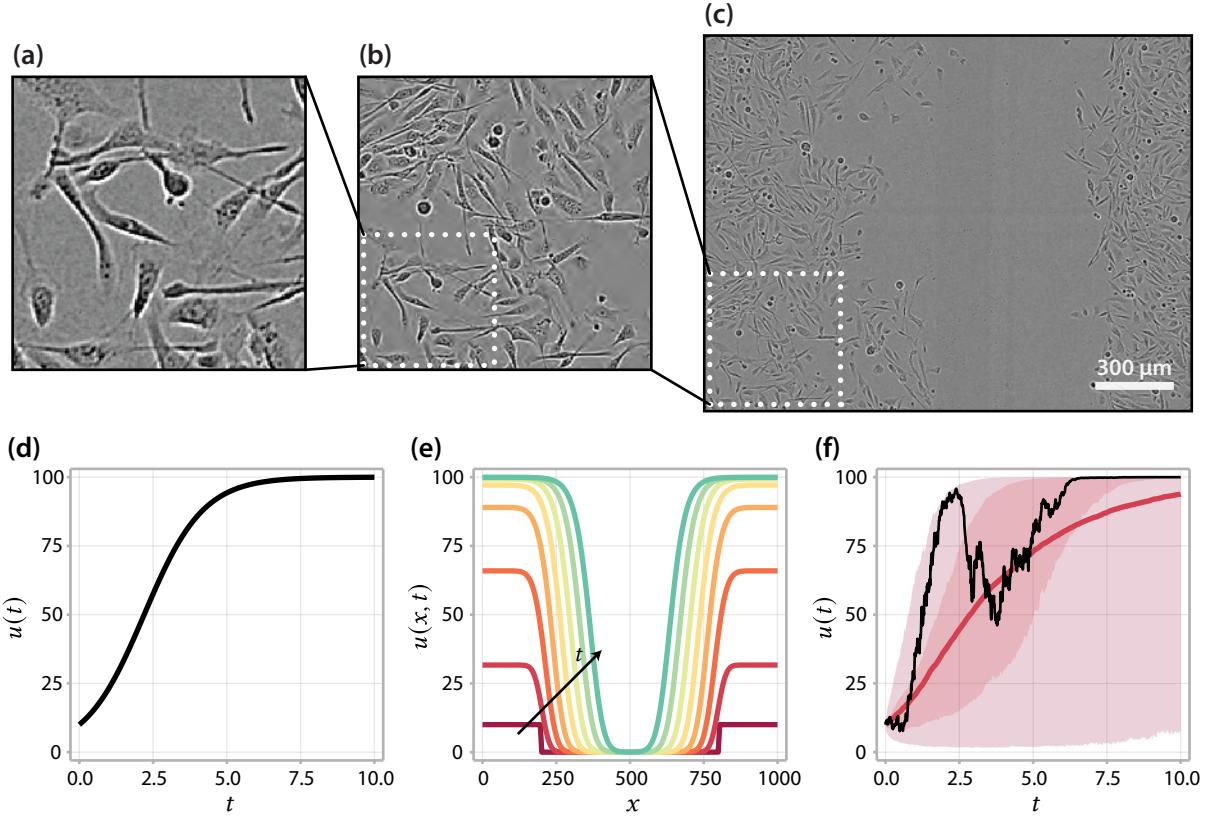
While equation (1.1) and equation (1.2) both describe key elements of bacterial persistence, as *mean-field* models both neglect the variability ubiquitous to many biological processes. In this example, such an approximation is appropriate as a typical bacteria colony contains several million bacteria and each bacterium contains a similarly large number of proteins: the magnitude of noise is typically proportional to $1/\sqrt{N}$, so for $N \gg 1$, noise is comparatively small in magnitude. However, in modelling the collective dynamics of relatively small populations it is often the case that modelling noise provides the information needed to identify parameters [23]. In other cases, variability is a key element in explaining a process; bacterial persistence is thought to have evolved as a strategy for coping with *environmental volatility*, deterministic models simply cannot capture this.

The question of whether to include variability—or further, spatial information—in models of collective behaviour is as much a question of data availability as a trade-off with the computational cost. While parameter identifiability of ODE models can be readily assessed using existing techniques [24–26], such tools have only recently been developed for partial differential equation models [27] and stochastic models [28]. Therefore, it is not always clear at the outset whether the additional computational cost of a stochastic or spatial model will result in additional insight.

Data available from two-dimensional cell-culture experiments, including proliferation and scratch assays [31–35], often provides detailed information about individuals cells across both space and time (figure 1.3). Provided the population is sufficiently large and the spatial distribution of cells is homogeneous (figure 1.3b), spatial information can be neglected and the total population modelled using the logistic growth equation [36]

$$\frac{\mathrm{d}u}{\mathrm{d}t} = \lambda u \left(1 - \frac{u}{K}\right), \tag{1.3}$$

where $u(t)$ represents the population density at time t. We show the solution to equation (1.3) in figure 1.3d.

▲ Figure 1.3. (a–c) A *scratch assay* experiment with PC-3 prostate cancer cells [29, 30] shown at various levels of magnification. (d), (e) and (f) show the solution to equation (1.3), equation (1.4) and equation (1.5), respectively. In all equations, cell proliferation is described using logistic growth. We fix $\lambda = 1$, $\mathscr{D}(u) = 100$, $K = 100$, $u_0 = 10$ and $\sigma = 1$. In (c) a single realisation of the SDE is shown (black) along with the mean, 50% prediction interval, and 95% prediction interval from 1000 realisations. Note that, at $t = 10$, the mean is less than the median due to a negative skew in the distribution of $u(10)$.

This yields a system characterised by two parameters; a net population growth rate, λ, and a carrying capacity, or limiting population, K. Therefore, information in the data relating to cell motility and cell-to-cell interactions cannot be extracted, but such measurements of total population only are comparatively straightforward to extract.

If the data are summarised with statistics that capture spatial information, such as measures of cell location correlation (or *pair correlation*) [34] or measures of local density in a spatially heterogeneous experiment (figure 1.3c), mathematical models can extract information relating to cell movement and cell-to-cell interactions. A spatial analogue of equation (1.3) is the *generalised Fisher-Kolmogorov* model [36]

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left(\mathscr{D}(u)\frac{\partial u}{\partial x}\right) + \lambda u\left(1 - \frac{u}{K}\right), \tag{1.4}$$

with solution shown in figure 1.3e. This model is routinely applied to interpret spatially heterogeneous processes including wound healing [37–39] and tissue growth [13]. Locally, cell proliferation is governed by logistic growth (equation (1.3)), but the model accounts for spatial heterogeneity by describing cell movement with a non-linear diffusion, $\mathscr{D}(u)$. While equation (1.4) describes cell motility, the choice to apply a spatial model requires specification of the unknown functional $\mathscr{D}(u)$. The common choice $\mathscr{D}(u) = D$ (i.e. linear-diffusion) can produce a fit with good agreement with experimental data, however for two-dimensional cell

culture, estimates of D vary significantly with cell density [40]. It is unclear whether motility is stimulated by cell-to-cell interactions in high density regions, or whether it is reduced.

When the number of individuals is small (figure 1.3a), stochastic effects often dominate. One such stochastic analogue of the equation (1.3) is the *stochastic logistic equation* [41],

$$\mathrm{d}u = u\left(1 - \frac{u}{K}\right)(\lambda\mathrm{d}t + \sigma\mathrm{d}W_t). \tag{1.5}$$

Here, the growth rate fluctuates according to a Wiener process, W_t of intensity σ with the property $W_{t+h} - W_t \sim \mathrm{Normal}(0, h)$. Fluctuations cease where $u = 0$ and $u = K$ and are greatest in magnitude at $u = K/2$. Other appropriate stochastic models for the data shown in figure 1.3a include Poisson processes [42] (that neglect space), and individual based models (IBMs) (that include space) [34, 43]. The solution to equation (1.5) in figure 1.3f highlights the importance of modelling stochastic effects in cases where they are significant. Here, it is often assumed that the deterministic, or *mean-field*, model (equation (1.3)) describes the mean of the process. However, a Monte-Carlo estimate of the mean of the stochastic model in figure 1.3f shows a significant difference with the ODE model. The parameters of the ODE model may well be identifiable, but ignoring stochastic effects would, in this case, lead to inaccurate parameter estimates.

Developing mathematical models of an appropriate level of complexity to ensure parameter identifiability allows experimentalists to accurately quantify cell behaviours. This quantification is essential to compare behaviours across experimental conditions and to allow the exploitation and control of biological systems in new areas of application. In medicine, tissues grown in artificial scaffolds are gradually replacing synthetic grafts [11, 44–46]. To develop patient-specific tissue, bio-engineers must understand, control and predict cell proliferation in artificially constructed three-dimensional environments. In materials science, so-called *metamaterials* embed behaviours in their structure [47], allowing signal transmission [48, 49] and logical computation [7, 50] by otherwise passive materials. Currently, a limitation of such materials is the induction of energy. The introduction of biological subsystems can overcome this, by introducing and controlling cell behaviours to introduce energy through chemical hydrolysis [9, 51, 52].

The overarching goal of this thesis is to interpret existing experimental data and enable new experimentation using mathematical models. The fundamental question of whether model parameters are identifiable is routinely overlooked, meaning that many existing mathematical models cannot be parameterised using typical experimental data. We address this question of model complexity and identifiability using a number of approaches. In Chapter 2 and Chapter 3, we summarise complex molecular-level behaviour using control theory (Chapter 3) and kernels, which summarise the net overall effect of cell-to-cell interactions (Chapter 2). In Chapter 4, we explore formal identifiability analysis for stochastic differential equation (SDE) models. In Chapter 5, we develop, simplify, and analyse a mathematical model for a biological sub-system that induces energy into a mechanical metamaterial. Finally, in Chapter 6, we exploit identifiability to investigate the effect of the initial condition and experimental geometry on cell behaviour in 3D printed scaffolds. Our approach in this last chapter is motivated by the importance of the initial condition on mathematical and statistical analysis that we bring to the literature in Chapter 2.

The layout of this thesis by publication is as follows. In Section 1.2 of this chapter, we list and introduce the questions investigated in this thesis. In Section 1.3 we outline the main aims and outcomes of the research, which includes five first-author publications. In Section 1.4 a statement of joint-authorship, including details

of the contributions of the PhD candidate, is provided. These publications are provided verbatim in Chapters 2 to 6. As published research, each chapter stands as an independent piece of work and may be read as such. Finally in Chapter 7 we outline the significant contributions made to the literature by research in this thesis, and outline the opportunities for future work that we enable.

1.2 Research questions

1. Given highly detailed cell-positional information, can we identify the nature of cell-to-cell interactions using a stochastic, spatial, mathematical model?

   Simple two-dimensional cell culture experiments, such as scratch assays, are routinely used to study collective cell behaviour. Typical mean-field mathematical models of scratch assays describe cell motility and proliferation based on assumptions of linear diffusion and logistic growth. While these models can capture results from experiments where the initial density of cells is fixed, it is clear that they cannot describe data initiated across a range of initial densities [40].

   This density-dependent cell behaviour is yet to be explicitly characterised. To address this, we calibrate a stochastic individual-based model to scratch assay data across a range of initial densities. Our approach takes advantage of the vast amount of detailed, cell-positional information available from these experiments; information that existing modelling frameworks cannot incorporate. To determine the nature of cell-to-cell interactions that give rise to the density-dependent behaviour in the experiments, we perform model selection on a hierarchy of models where we systematically and sequentially remove interactions. Our results provide evidence that motility is density-dependent in these experiments. On the other hand, we do not see the effect of crowding on proliferation in these experiments. These results are significant as they are precisely the opposite of the assumptions in standard continuum models.

2. Can we model the complex individual-level behaviours in a population using control theory, to ensure model simplicity?

   Bacteria invest in a slow-growing subpopulation, called persisters, to ensure survival in the face of uncertainty. This *bet-hedging* strategy is remarkably similar to financial hedging, where diversifying an investment portfolio protects against economic uncertainty. This similarity suggests a new theoretical foundation for understanding cellular hedging by unifying the study of biological population dynamics and the mathematics of financial risk management through optimal control theory. By modelling complex individual-level decision making using control theory, we avoid the need to model the complex molecular-level machinery that gives rise to strategic cellular decision making. The complexity of the model, therefore, more closely matches the information available from typical experimental studies.

   Given that environmental uncertainty is fundamental to the emergence of hedging strategies, we describe the behaviour of bacteria in response to several novel models of environmental volatility described by continuous-time stochastic processes. Analytical and simulation results probe the optimal persister strategy, revealing results consistent with experimental observations. Our approach enables new experimental studies that investigate and exploit the response of biological systems to uncertainty.

3. How can we determine whether model parameters in a stochastic model are identifiable?

   Mathematical models are routinely calibrated to experimental data, with goals ranging from building

predictive models to quantifying parameters that cannot be measured. Whether or not reliable parameter estimates are obtainable from the available data can easily be overlooked. Such issues of *parameter identifiability* have ramifications for both the predictive power of a model and the mechanistic insight obtainable. While tools to assess parameter identifiability in deterministic, ODE models are plentiful, there are no commonly-adopted methods for analysing identifiability in stochastic models.

We study two techniques that bring existing tools for identifiability analysis of ODE models to SDE models. Firstly, to assess the structural features of models that lead to non-identifiability, we formulate a system of ODEs that describe the statistical moments of the SDE model. Secondly, we study identifiability in the case of typical experimental data using Markov-chain Monte Carlo. Given the scarcity of available techniques for identifiability analysis of SDE models, we review and discuss alternative approaches and provide an outlook for future research in this area.

4. How can a biological sub-system be incorporated into a metamaterial?

Mechanical metamaterials are designed to enable unique functionalities, but are typically limited by an initial energy state and require an independent energy input to function repeatedly. Our study introduces a theoretical active mechanical metamaterial that incorporates a biological reaction mechanism to overcome this key limitation of passive metamaterials. Our material allows for reversible mechanical signal transmission, where energy is reintroduced by the biologically motivated reaction mechanism. By analysing a coarse-grained continuous analogue of the discrete model, we find that signals can propagate through the material as a travelling wave. Analysis of the continuum model reveals similarities with the well-known FitzHugh-Nagumo system. We also find explicit formulae that approximate the effect of the time scale of the reaction mechanism on the signal transmission speed, which is essential for controlling the material.

5. Can identifiability techniques be applied to estimate and compare parameters in a spatial model of cell growth in artificially constructed scaffolds?

Tissue growth in three-dimensional (3D) printed scaffolds enables exploration and control of cell behaviour in more biologically realistic geometries than that allowed by traditional 2D cell culture. Cell proliferation and migration in these experiments have yet to be explicitly characterised, limiting the ability of experimentalists to determine the effects of various experimental conditions, such as scaffold geometry, on cell behaviour.

We develop a mathematical model to describe tissue growth in melt electro-written scaffolds that comprise thin square pores of various sizes. Calibrating the model to highly detailed temporal measurements of the average cell density, tissue coverage, and tissue geometry reveals the information required to identify parameters. Parameter estimates and uncertainty quantification through profile likelihood analysis reveal consistency in the rate of cell proliferation and steady-state cell density between pore sizes. We find that the Porous-Fisher model can capture features relating to the cell density and tissue coverage but cannot capture geometric features relating to the circularity of the tissue interface. Our analysis identifies two distinct stages of tissue growth, suggests several areas for model refinement, and guides future experimental work.

1.3 Objectives and Outcomes

This thesis comprises two main objectives:

1. Develop data-driven stochastic models of collective behaviour, and explore mathematical techniques that diagnose parameter identifiability in stochastic models.

2. Apply mathematical models to develop and describe biological systems in bioengineering.

Objective (1) is addressed through three publications, provided in Chapters 2 to 4. We address objective (2) through two publications, provided in Chapters 5 to 6.

This thesis is presented by published papers and comprises five peer-reviewed publications, all published in international Q1 journals and provided verbatim in Chapters 2 to 6 [9, 23, 30, 53, 54]. The introduction and conclusion chapters aside, each chapter can be read independently and contains an introduction section (that includes a literature review), a methods section (that outlines the mathematical and statistical models implemented), results and a discussion, which together outline the outcomes obtained from each experimental work, and place the findings in a wider context.

The candidate contributed significantly to all articles and is listed as the first author. Therefore, the work presented here satisfies the requirements by the Queensland University Technology for thesis by publication. The articles (in order of appearance) are:

1. AP Browning, W Jin, MJ Plank, MJ Simpson. 2020 Identifying density-dependent interactions in collective cell behaviour. *Journal of the Royal Society Interface* 17:20200143.
   `doi:10.1098/rsif.2020.0143`

2. AP Browning, JA Sharp, T Mapder, CM Baker, K Burrage, MJ Simpson. 2021 Persistence as an optimal hedging strategy. *Biophysical Journal* 120:133–142.
   `doi:10.1016/j.bpj.2020.11.2260`

3. AP Browning, DJ Warne, K Burrage, RE Baker, MJ Simpson. 2021 Identifiability analysis for stochastic differential equation models in systems biology. *Journal of the Royal Society Interface* 17:2020.11.2260.
   `doi:10.1098/rsif.2020.0652`

4. AP Browning, FG Woodhouse, MJ Simpson. 2019 Reversible signal transmission in an active mechanical metamaterial. *Proceedings of the Royal Society A* 475:20190146.
   `doi:10.1098/rspa.2019.0146`

5. AP Browning, OJ Maclaren, PR Buenzli, M Lanaro, MC Allenby, MA Woodruff, MJ Simpson. 2021 Model-based data analysis of tissue growth in thin 3D printed scaffolds. *Journal of Theoretical Biology* 528:110852.
   `doi:10.1016/j.jtbi.2021.110852`

1.4 Statement of joint authorship

This thesis comprises five publications; four published, and one accepted for publication. In this section, the contributions of the PhD candidate and each co-author are summarised. All co-authors have confirmed their contribution as appears below, and agree to the presentation of these publications in this thesis.

Chapter 2: Identifying density-dependent interactions in collective cell behaviour

AP Browning, W Jin, MJ Plank, MJ Simpson. 2020 Identifying density-dependent interactions in collective cell behaviour. *Journal of the Royal Society Interface* 17:20200143.
`doi:10.1098/rsif.2020.0143`

- **Alexander P Browning (Candidate)** designed the study, implemented code for the individual based model and sequential Monte-Carlo, generated and interpreted results, processed the experimental data, drafted the manuscript, and revised the manuscript during peer-review.

- *Wang Jin* designed the study, processed the experimental data, interpreted the results and provided comments on the manuscript.

- *Michael J Plank* designed the study, interpreted the results and provided comments on the manuscript.

- *Matthew J Simpson* designed the study, supervised the research, interpreted the results, and provided comments on the manuscript.

Chapter 3: Persistence as an optimal hedging strategy

AP Browning, JA Sharp, T Mapder, CM Baker, K Burrage, MJ Simpson. 2021 Persistence as an optimal hedging strategy. *Biophysical Journal* 120:133–142.
`doi:10.1016/j.bpj.2020.11.2260`

- **Alexander P Browning (Candidate)** designed the study, conceived the optimal control model of bacterial persistence, developed and implemented the computational algorithms, generated and interpreted results, drafted the manuscript, and revised the manuscript during peer-review.

- *Jesse A Sharp* designed the study, aided development of the computational algorithms, interpreted the results and provided comments on the manuscript.

- *Tarunendu Mapder* designed the study, aided development of the computational algorithms, interpreted the results and provided comments on the manuscript.

- *Christopher M Baker* interpreted the results and provided comments on the manuscript.

- *Kevin Burrage* designed the study, supervised the research, interpreted the results, and provided comments on the manuscript.

- *Matthew J Simpson* designed the study, supervised the research, interpreted the results, and provided comments on the manuscript.

Chapter 4: Identifiability analysis for stochastic differential equation models in systems biology

AP Browning, DJ Warne, K Burrage, RE Baker, MJ Simpson. 2020 Identifiability analysis for stochastic differential equation models in systems biology. *Journal of the Royal Society Interface* 17:2020112260.
`doi:10.1098/rsif.2020.0652`

- **Alexander P Browning (Candidate)** designed the study, implemented the statistical computational algorithms, derived the moment equations, generated and interpreted results, drafted the manuscript, and revised the manuscript during peer-review.

- *David J Warne* aided in the implementation of the computational algorithms, interpreted the results and provided comments on the manuscript.

- *Kevin Burrage* designed the study, supervised the research, interpreted the results and provided comments on the manuscript.

- *Ruth E Baker* designed the study, supervised the research, interpreted the results and provided comments on the manuscript.

- *Matthew J Simpson* designed the study, supervised the research, interpreted the results, and provided comments on the manuscript.

Chapter 5: Reversible signal transmission in an active mechanical metamaterial

AP Browning, FG Woodhouse, MJ Simpson. 2019 Reversible signal transmission in an active mechanical metamaterial. *Proceedings of the Royal Society A* 475:20190146.
`doi:10.1098/rspa.2019.0146`

- **Alexander P Browning (Candidate)** designed the study, developed and analysed the mathematical model, performed the travelling wave analysis, implemented the numerical algorithms, generated and interpreted results, drafted the manuscript, and revised the manuscript during peer-review.

- *Francis G Woodhouse* designed the study, supervised the research, interpreted the results and provided comments on the manuscript.

- *Matthew J Simpson* designed the study, supervised the research, interpreted the results, and provided comments on the manuscript.

Chapter 6: Model-based data analysis of tissue growth in thin 3D printed scaffolds

AP Browning, OJ Maclaren, PR Buenzli, M Lanaro, MC Allenby, MA Woodruff, MJ Simpson. 2021 Model-based data analysis of tissue growth in thin 3D printed scaffolds. *Journal of Theoretical Biology* 528:110852.
`doi:10.1016/j.jtbi.2021.110852`

- **Alexander P Browning (Candidate)** designed the study, developed and analysed the mathematical model, implemented the statistical techniques, implemented the numerical algorithms, processed the experimental data, generated and interpreted results, drafted the manuscript, and revised the manuscript during peer-review.

- *Oliver J Maclaren* designed the study, developed the mathematical model, interpreted the results and provided comments on the manuscript.

- *Pascal R Buenzli* designed the study, developed the mathematical model, interpreted the results and provided comments on the manuscript.

- *Matthew Lanaro* performed the experiments, processed the experimental data and provided comments on the manuscript.

- *Mark C Allenby* performed the experiments, processed the experimental data and provided comments on the manuscript.

- *Maria A Woodruff* performed the experiments and provided comments on the manuscript.

- *Matthew J Simpson* designed the study, supervised the research, interpreted the results, and provided comments on the manuscript.

Chapter 2

Identifying density-dependent interactions in collective cell behaviour

Scratch assays are routinely used to study collective cell behaviour *in vitro*. Typical experimental protocols do not vary the initial density of cells, and typical mathematical modelling approaches describe cell motility and proliferation based on assumptions of linear diffusion and logistic growth. Jin *et al.* (2016) find that the behaviour of cells in scratch assays is density-dependent, and show that standard modelling approaches cannot simultaneously describe data initiated across a range of initial densities. To address this limitation, we calibrate an individual based model to scratch assay data across a large range of initial densities. Our model allows proliferation, motility, and a direction bias to depend on interactions between neighbouring cells. By considering a hierarchy of models where we systematically and sequentially remove interactions, we perform model selection analysis to identify the minimum interactions required for the model to simultaneously describe data across all initial densities. The calibrated model is able to match the experimental data across all densities using a single parameter distribution, and captures details about the spatial structure of cells. Our results provide strong evidence to suggest that motility is density-dependent in these experiments. On the other hand, we do not see the effect of crowding on proliferation in these experiments. These results are significant as they are precisely the opposite of the assumptions in standard continuum models, such as the Fisher-Kolmogorov equation and its generalisations.

Statement of Contribution of Co-Authors for
Thesis by Published Paper

The authors listed below have certified that:

1. they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
2. they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
3. there are no other authors of the publication according to these criteria;
4. potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit, and
5. they agree to the use of the publication in the student's thesis and its publication on the QUT's ePrints site consistent with any limitations set by publisher requirements.

In the case of this chapter:
Identifying density-dependent interactions in collective cell behaviour
Published 2020 in Journal of the Royal Society Interface
doi:10.1098/rsif.2020.0143

Contributor	Statement of Contribution
Alexander P Browning 4 Aug 2021	Designed the study, implemented code for the individual based model and sequential Monte-Carlo, generated and interpreted results, processed the experimental data, drafted the manuscript, and revised the manuscript during peer-review.
Wang Jin	Designed the study, processed the experimental data, interpreted the results and provided comments on the manuscript.
Michael J Plank	Designed the study, interpreted the results and provided comments on the manuscript.
Matthew J Simpson	Designed the study, supervised the research, interpreted the results, and provided comments on the manuscript.

Principal Supervisor Confirmation
I have sighted email or other correspondence from all Co-authors confirming their certifying authorship.

Matthew J Simpson
Name

Signature

9 August 2021
Date

2.1 Introduction
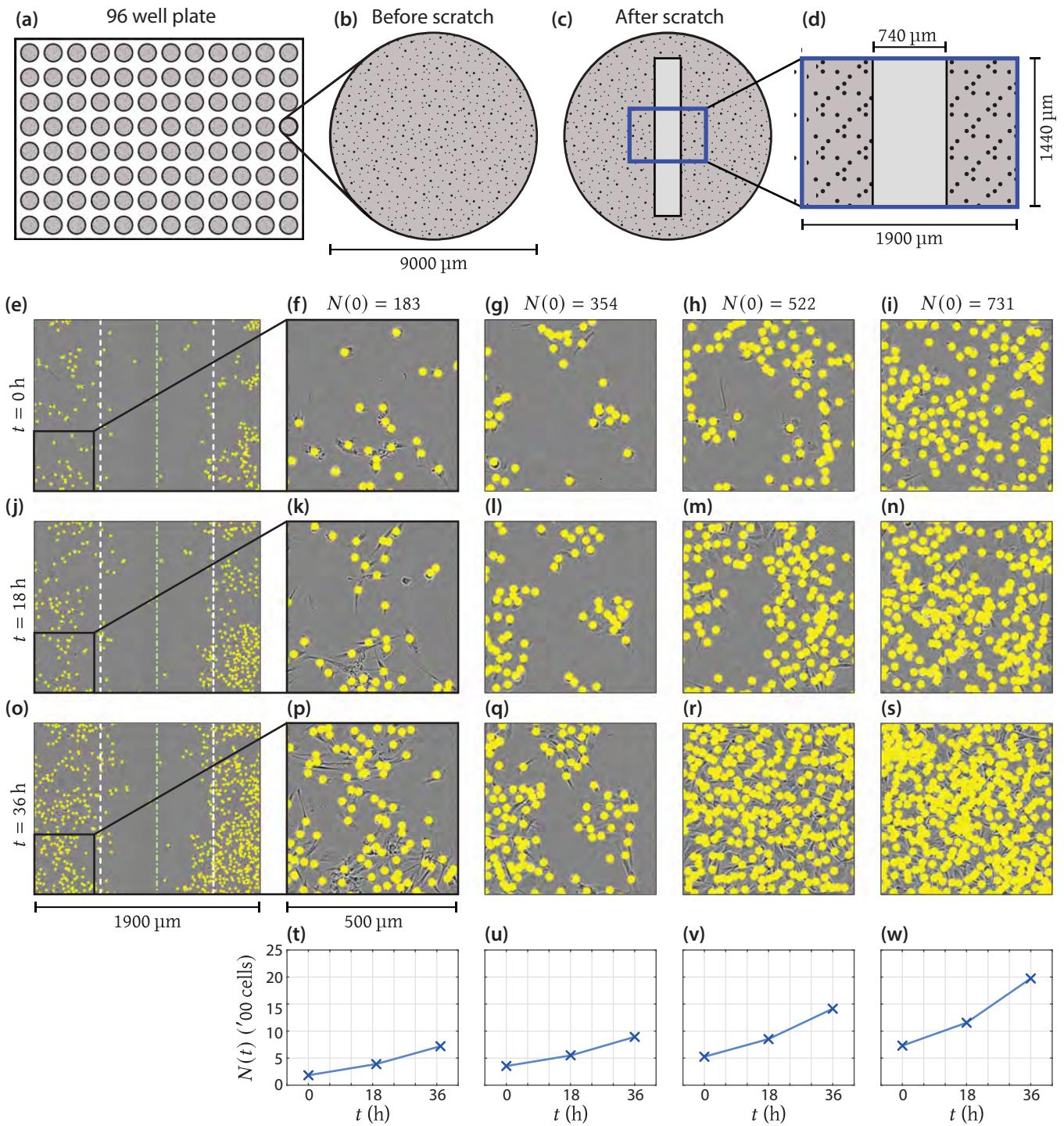
Simple two-dimensional *in vitro* experiments, such as scratch assays, are commonly used to study collective cell behaviour [31–35, 40]. Scratch assays are conducted by placing a uniform monolayer of cells on a two-dimensional substrate and creating an artificial wound, or *scratch*, in the monolayer (figure 2.1a–d) [40]. Typical experimental protocols do not vary the initial density of cells between experiments and, therefore, provide no information on how the initial density affects cell migration or proliferation. In order to study potentially density-dependent cell behaviour, we consider novel scratch assay data where we deliberately vary the initial density of cells between experiments. The variation in the initial cell density in our experiments is large: the initial population in the highest density experiment is greater than the final population in the lowest density experiment.

Logistic growth and linear diffusion are often assumed to be the key mechanisms governing collective cell behaviour in a range of *in vitro* and *in vivo* conditions [36–38, 40, 55–57]. Mean-field mathematical models that incorporate one or both of these mechanisms are routinely used to model tumour spheroids [58]; cells in living tissues [15, 59]; and simple *in vitro* experiments such as scratch [40], migration [60], and proliferation [34] assays. While calibrating these models to experimental data often leads to a good match [37], these models make the standard assumption that the parameters are independent of both initial condition and cell density. For example, the Fisher-Kolmogorov equation [61, 62]

$$\frac{\partial c}{\partial t} = \underbrace{D\frac{\partial^2 c}{\partial x^2}}_{\substack{\text{Density-independent} \\ \text{motility}}} + \underbrace{\lambda c\left(1 - \frac{c}{K}\right)}_{\substack{\text{Contact inhibited} \\ \text{proliferation}}}, \tag{2.1}$$

is commonly used to model scratch assay experiments [40, 63], where $c(x, t)$ is the cell density. Equation (2.1) describes density-independent motility, characterised by a constant diffusivity D; and density-dependent proliferation, characterised by a constant proliferation rate λ and a constant carrying capacity K. Jin *et al.* [40] find that calibrating the solution of equation (2.1) to scratch assay data yields vastly different estimates of D for each initial condition considered. In contrast to equation (2.1), some studies assume that cell motility is density-dependent. However, the way in which this is modelled is inconsistent. For example, Cai *et al.* [64, 65] model motility with a non-linear diffusivity term that decreases with density to simulate crowding. In direct contrast, many other studies model motility using a non-linear diffusivity term that increases with density to simulate contact stimulation [66–68]. These studies all calibrate their respective models to experimental data with a single initial density [64–68]. In contrast, the approach that we take here is biologically significant since we identify the nature of density-dependent interactions using experimental data initiated with a range of initial cell densities.

In this work we describe the cell behaviour with a lattice-free individual based model (IBM) [34, 43, 69]. The IBM represents cells as *agents* that take locations in continuous space, and so we can specify the initial agent locations in the model to precisely match the initial cell locations in the experiments. This choice also allows the model to capture local details—such as spatial structure and clustering—which are neglected by standard continuum modelling approaches [38, 40]. The agents in the IBM proliferate and move, the rates of which we assume depend explicitly on interactions between neighbouring agents. We quantify these interactions with kernels that depend on the distance between pairs of cells. Directional bias is also incorporated so

▲ Figure 2.1. (a–d) Schematic of the experimental geometry. (a) 96 well plate. (b) Each assay was performed by distributing a monolayer of cells in a well of diameter 9000 μm. (c) An artificial wound (light region, not to scale) created within the monolayer of cells. (d) Field-of-view of the experimental data, which is much smaller than each well (not to scale). (e),(j, o) Experimental data for the lowest density experiment (where $N(0) = 183$) at 0 h, 18 h and 36 h, respectively. In (e, j) and (o) the green dash-dot line represents the approximate centre of the scratch at $t = 0$ h; and, the white dashed lines represent the approximate edge of the scratch at $t = 0$ h. Insets in (f, k) and (p) show the lower-left region of respective images in (e, j) and (o). The height and width of the field-of-view in the insets is 500 μm. Subsequent columns show insets for experimental data at increasing densities where $N(0) = 354$, 522 and 731. In each image, the location of each cell is indicated with a yellow marker with diameter $\varphi = 24$ μm (to scale). (t–w) Summary of experimental data in each respective column showing $N(t)$.

that agents are more likely to move either away from, or towards, regions of high density [65, 69]. A key advantage of the IBM is its flexibility: it is trivial to add and remove mechanisms, which we do to study the interactions required for the model to simultaneously match all experiments. Finally, the IBM is stochastic and so naturally describes the variation between experiments.

We aim to identify the nature of interactions which enable the model to simultaneously describe experimental data across a wide range of initial cell densities. This study is the first time scratch assay data initiated across a range of initial cell densities has been calibrated to an IBM. We take a Bayesian approach to parameter estimation [34, 70–72], and identify interactions using model selection [70]. We always force the model to simultaneously match data from all nine experiments. The mathematical model is always initiated using the initial configuration of cells in each experiment, and we compare simulated and experimental data at 18 h and 36 h, the latter which corresponds to the duration of the experiment. The calibrated model is able to replicate the experimental data, and we find evidence to suggest that motility is an increasing function of density, which is contrary to both the common mathematical assumption of linear diffusion and work which assumes motility decreases with density [64, 65]. Experimentation with summary statistics confirms the importance of spatial structure, which is neglected by standard modelling and model calibration approaches.

2.2 Materials and Methods

2.2.1 Experimental methods

Our experimental model of cell migration and proliferation comprises a series of scratch assays using PC-3 prostate cancer cells [29]. We deliberately vary the initial number of cells in each experiment by seeding approximately 8000, 10 000 and 12 000 cells in a 9000 μm diameter well within a 96-well plate (figure 2.1a,b). Cells are grown overnight to create a spatially uniform monolayer before a scratch is created (figure 2.1c). Images of the central 1440×1900 μm of each well are captured over a period of 48 hours after the monolayer is scratched (figure 2.1d). Full details of the experimental methods are provided in [40].

ImageJ [73] is used to determine the approximate coordinates of individual cells in each image, this data is given as supporting material[1]. We exclude the first 12 hours of experimental data from our analysis [34] to ensure that sufficient time has passed so that the cells are migrating and proliferating after the scratch has been made. We then record experimental images and we treat this as the beginning of the experiment, $t = 0$ h. The variability in initial cell number is high: despite an initial seeding density of approximately 8000–12000 cells per well, which corresponds to expected initial number of cells within the field-of-view of 344–516, we find that the initial number of cells within the field-of-view at $t = 0$ h ranges from 183 to 731 (figure 2.1f–i). This variation is also high between experiments of the same seeding density [74], due to the fact that our field-of-view is relatively small so that fluctuations about the expected values are relatively large. We demonstrate this variation in figure 2.1e–s.

2.2.2 Mathematical model

We use a lattice-free individual based model (IBM) [34, 69] which we simulate with the Gillespie algorithm [42]. The model includes density-dependent proliferation and movement, but does not consider death, which is

---

[1]Available at `https://github.com/ap-browning/scratchIBM`

not observed in the experiments. To be consistent with previous experimental observations [65], the model incorporates a bias mechanism so that cells both move, and disperse daughter cells during proliferation, in a direction either towards, or away from, crowded regions. Density dependence is incorporated into the model through cell-to-cell interactions, so that behaviour is dependant upon *local crowding*.

The field-of-view of the experimental data is rectangular, with dimensions $1440 \times 1900 \ \mu$m (figure 2.1d), and we replicate this by using the same geometry in the model. As the well in the tissue culture plate is much larger than this field-of-view, we apply periodic boundary conditions [34] (indicated in blue in figure 2.1c,d). Cells are modelled as *agents* that have a point location but no physical size. In our previous work we find that, on average, these PC-3 prostate cancer cells have an area that corresponds to a disc of diameter $\varphi = 24 \ \mu$m [34]. The interaction mechanisms we model are not based on volume exclusion or hard sphere interactions [43], but rather depend on agent separation in such a way that configurations wherein two agent centres are very close are unlikely. We denote the agent locations $x_n = (x_n, y_n), \ n \in \{1, ..., N(t)\}$, where $N(t)$ denotes the number of agents in the simulation. We specify the initial agent locations in each simulation to match the experimental images at $t = 0$ h.

Directional bias

We quantify crowding by placing a *bias kernel* at the location of each agent to form a *crowding surface*, $B(\mathbf{x})$, as shown in figure 2.2c,d for the configuration of cells in figure 2.2a,b. Mathematically, this is given by

$$B(\mathbf{x}) = \sum_{i=1}^{N(t)} w^{(b)}(\|\mathbf{x} - \mathbf{x}_i\|), \tag{2.2}$$

and describes a measure of local crowding at x, where $w^{(b)}(r)$ is the bias kernel. The contributions of each agent to $B(\mathbf{x})$ depend on the distance between x and the location of the ith agent, x_i, given by $r = \|\mathbf{x} - \mathbf{x}_i\|$. There are many possible choices of kernel [75], however we find that the standard choice of Gaussian leads to a good match with experimental data [34]. In this study, we choose $w^{(b)}(r)$ to be a Gaussian [76] of spread σ with an extremum of γ_b so that
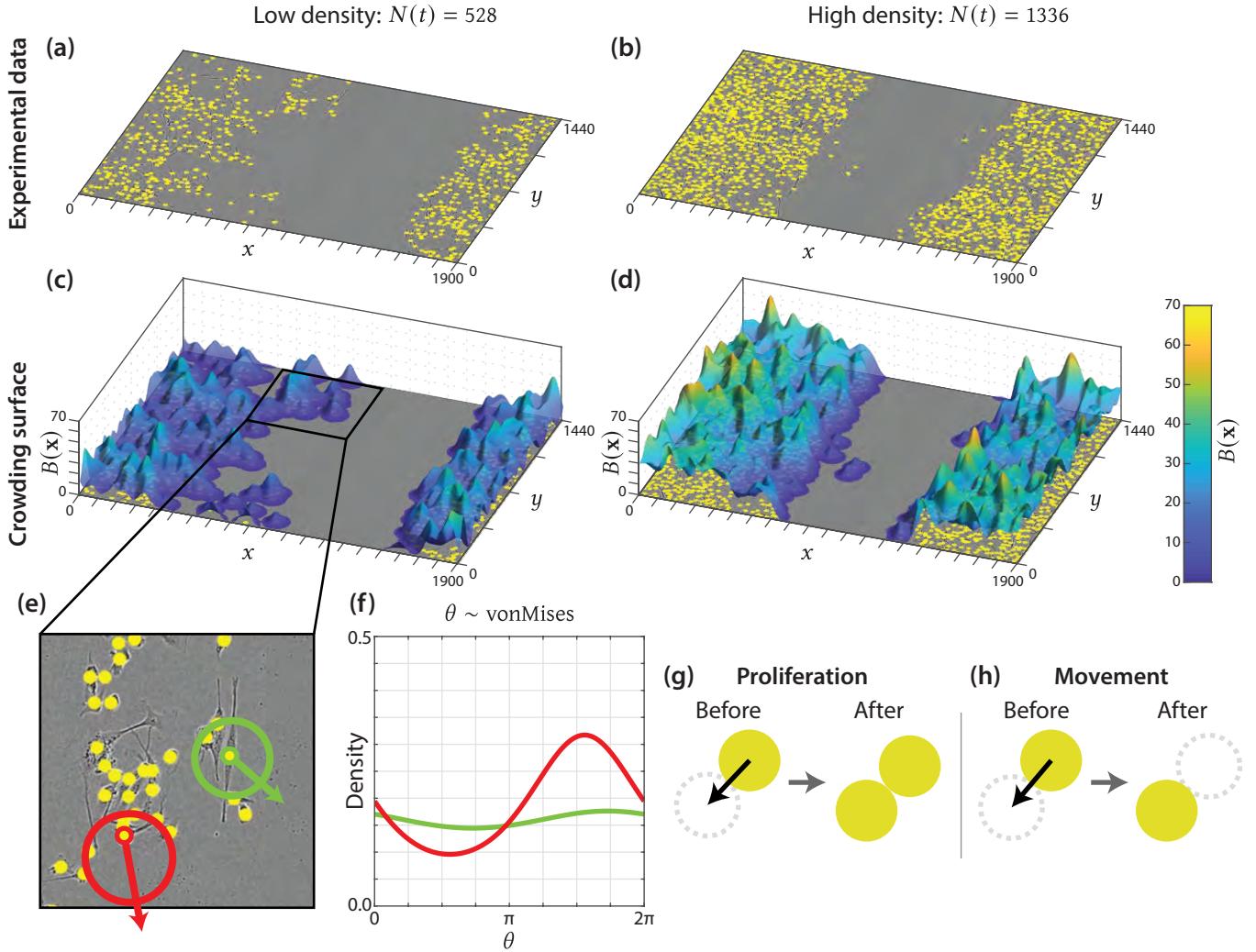
$$w^{(b)}(r) = \gamma_b \exp\left(-\frac{r^2}{2\sigma^2}\right). \tag{2.3}$$

For computational efficiency, we truncate the kernel to zero for $r \geq 3\sigma$ [76]. This truncation means that agents separated by a distance of more than 3σ do not interact. Therefore, $B(\mathrm{x})$ is a measure of *local crowding*.

For $\gamma_b > 0$, agents prefer to move and disperse daughter agents in the direction of steepest descent on the crowding surface, which corresponds to regions of lower density (setting $\gamma_b < 0$ has the opposite effect). This preference depends on the steepness, so that agents close to highly crowded regions are more likely to move and disperse daughter agents in their preferred direction, demonstrated in figure 2.2e,f, where the red agent has a stronger bias strength than the green agent. To do this, we define the bias vector of agent n as

$$\mathrm{B}_n = -\nabla B(\mathbf{x}_n), \tag{2.4}$$

which gives the magnitude and direction of steepest descent. Therefore, B_n is a simple measure of local spatial structure at the location of agent n. The movement and proliferation directions are then sampled

▲ Figure 2.2. (a)–(b) Experimental data at (a) low density; and, (b) high density. The location of each cell is indicated with a yellow marker of diameter $\varphi = 24\ \mu m$ (to scale). (c)–(d) Example crowding surface, where agent locations are taken from experimental data in (a) and (b), respectively. (e) Two-dimension inset of experimental image in (c), showing the bias distribution for two agents in radial coordinates centred at each agent. The off-centeredness of each circle therefore represents the strength of the bias, which is stronger for the red cell than the green cell. (f) The bias distributions in (e) shown in Euclidean coordinates for clarity. (g)–(h) Schematic of cell division (proliferation) and movement events, respectively, where the black arrow indicates the sampled direction of each cell. When an agent proliferates, the daughter cell is placed a distance of φ from the mother cell. When an agent moves, the agent is moved a distance of φ.

from the von Mises distribution [77]

$$\text{von Mises}(\arg(\mathbf{B}_n), \|\mathbf{B}_n\|). \tag{2.5}$$

The expected and most likely direction is, therefore, $\arg(\mathbf{B}_n)$. The direction distribution becomes increasingly concentrated around $\arg(\mathbf{B}_n)$ as $\|\mathbf{B}_n\|$ becomes large, and approaches a uniform distribution on $[0, 2\pi)$ as $\|\mathbf{B}_n\| \to 0$.

We illustrate the directional bias mechanism in figure 2.2c–f. The crowding surface is constructed by placing a Gaussian kernel at the location of each agent (figure 2.2c,d). In figure 2.2e we show the bias distribution and preferred direction for an agent in a low (green) and high (red) density region. For each agent the arrow shows the preferred direction with the corresponding von Mises distribution plotted in

radial coordinates centred at the location of each agent. In figure 2.2f we show these distributions are shown as a function of the angle, $\theta \in [0, 2\pi)$, for clarity.

Proliferation and movement

Cell division (proliferation) and movement events occur according to a Poisson process [78] with density-dependent rates $P_n \geq 0$ and $M_n \geq 0$, respectively. These rates comprise constant intrinsic rates $p > 0$ and $m > 0$, that are modified by interactions with neighbouring agents. These interactions result in a local density dependence, so that agents in high density regions are able to behave differently to solitary agents, or agents in low density regions [69].

We quantify these interactions using kernels, $w^{(\cdot)}(r)$, that depend on the separation distance, $r \geq 0$, between an agent and its neighbours, such that

$$P_n = \max\left(0, p - \sum_{\substack{i \neq n}}^{N(t)} w^{(p)}(\|x_n - x_i\|)\right), \tag{2.6}$$

$$\text{and} \quad M_n = \max\left(0, m - \sum_{\substack{i \neq n}}^{N(t)} w^{(m)}(\|x_n - x_i\|)\right). \tag{2.7}$$

Again, we choose the kernels to be Gaussian [76], with spread σ, so that

$$w^{(p)}(r) = \gamma_p \exp\left(-\frac{r^2}{2\sigma^2}\right), \tag{2.8}$$

$$\text{and} \quad w^{(m)}(r) = \gamma_m \exp\left(-\frac{r^2}{2\sigma^2}\right). \tag{2.9}$$
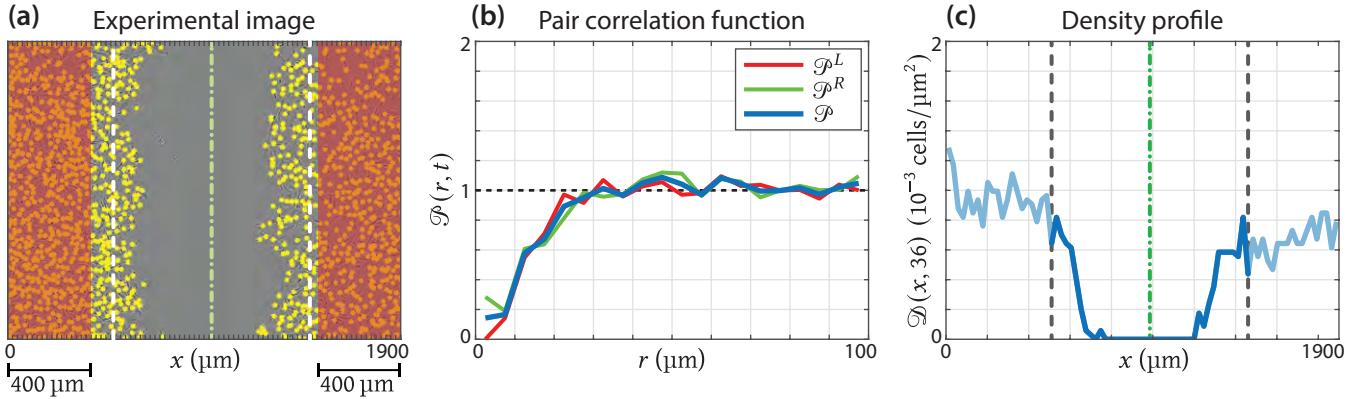
Here, γ_p and γ_m are the extrema of the proliferation and movement kernels, respectively. Setting $\gamma_m < 0$ (or $\gamma_p < 0$) means that crowding increases motility (or proliferation); setting $\gamma_m > 0$ (or $\gamma_p > 0$) means that crowding decreases motility (or proliferation); and, setting $\gamma_m = 0$ (or $\gamma_p = 0$) means that motility (or proliferation) is independent of local density. Again, we truncate the kernels to zero for $r \geq 3\sigma$ [76].

When an agent at x_n proliferates, the daughter agent is dispersed a distance φ (approximately one cell diameter) from x_n, with the direction sampled from the bias distribution for that agent (figure 2.2e,f). This is demonstrated in figure 2.2g. When an agent at x_n moves, it is moved to a location of distance φ from x_n, with the direction sampled from the bias distribution for that agent (figure 2.2e,f). This is demonstrated in figure 2.2h.

2.2.3 Summary statistics

To match model simulations to the experimental data, we record the locations of agents at both $t = 18$ h and $t = 36$ h. We denote the experimental data at both time points from experiment $i \in \{1, ..., 9\}$ as $X_{\text{obs}}^{(i)}$, and simulation data from experiment i as $X_{\text{sim}}^{(i)}$. In this section, we detail how we summarise the high dimensional data X into lower dimensional summary statistics. This allows us to define a distance function, $d(X_{\text{obs}}, X_{\text{sim}})$, that represents the distance between experimental and simulation data.

We aim to capture three key pieces of information in the experiments: (1) the population size; (2) the spatial structure; and, (3) the density profile. The first two pieces of information are related to the first two spatial moments [76], and the last piece of information relates to the wound closure, total population and

(a) Experimental image

(b) Pair correlation function

(c) Density profile

▲ Figure 2.3. (a) Experimental image. The location of each cell is indicated with a yellow marker of diameter $\varphi = 24\ \mu m$ (to scale). (b) Pair correlation function calculated from the distribution in cells in (a). $\mathscr{P}$ is the average of $\mathscr{P}^L$ and $\mathscr{P}^R$, calculated using the agents in the 400 μm to the far-left and far-right (red region) of the experimental data in (a), respectively. Therefore, $\mathscr{P}$ contains information about the spatial distribution of cells, but not about the scratched region. Pair correlation functions for all experimental images are provided in Appendix A (figure A.8 and figure A.10). (c) The density profile, calculated by counting the number of cells in subregions of width $1900/80\ \mu m$ and dividing by the area of each subregion. The sub-regions are indicated as the axis ticks in (a). Only the central 41 bins are used to compare experimental and simulated data. Green dash-dot lines in (a) and (c) indicate the approximate centre of the scratch at $t = 0$ h and dashed lines indicated the approximate boundary of this region at $t = 0$ h.

the spatial distribution of cells. The first spatial moment, the average density, is the number of agents in the population, $N(t)$. The second spatial moment describes the spatial distribution of agents, often characterised by a pair correlation function [33, 69, 76]. In summary, the pair correlation function describes the density of pairs of agents separated by a distance r, relative to the expected density of pairs if the population were uniformly distributed [33]. Since the data is discrete, we define the pair correlation, $\mathscr{P}(j, t)$, $j \in \mathbb{N}$, which describes the relative number of pairs separated by a distances ranging from $(j - 1)\Delta r < r < j\Delta r$, given by

$$\mathscr{P}(j, t) = \frac{LW}{N(t)^2 \pi \Delta r (2j + \Delta r)} \sum_{n=1}^{N(t)} \sum_{\substack{i=1 \\ i \neq n}}^{N(t)} 1_{(j-1)\Delta r \leq \|x_n - x_i\| < j\Delta r}, \tag{2.10}$$

where L and W are length and width, respectively, of the region and 1 is the indicator function. In this study, we choose $\Delta r = 5\ \mu m$, and consider the pair correlation up to a distance of 100 μm such that $j \leq 20$. Smaller values of Δr lead to a noisier pair correlation function, and larger values of Δr smooth the pair correlation, potentially hiding information [79].

In a scratch assay the central region of the experimental field-of-view is initially devoid of agents (figure 2.3a). To account for this, we calculate pair correlation functions for sub-region of width 400 μm in the far-left, and far-right, of the domain (figure 2.3a, indicated in red) denoted $\mathscr{P}^{(L)}(j, t)$ and $\mathscr{P}^{(R)}(j, t)$, respectively. The width of this region is chosen so it does not overlap with the region devoid of cells in the experimental images. We apply periodic boundary conditions on these sub-regions, so that the separation of a pair of agents is the smallest possible distance accounting for the periodic boundary conditions. The pair correlation function that summarises the entire experiment is $\mathscr{P}(j, t) = (\mathscr{P}^{(L)}(j, t) + \mathscr{P}^{(R)}(j, t))/2$ (figure 2.3b). Results in figure 2.3b also confirm our assumption that a typical cell diameter is approximately 24 μm [34]. Results in Appendix A (figure A.8 and figure A.10) show the pair correlation function for all experimental images, where clustering at short distances is observed for earlier time data, for lower cell densities.

The final piece of information, the density profile, describes the wound closure, total population and

	$\boldsymbol{\theta}_k$	Density Dependence
Model 1	$(m, p, \gamma_m, \gamma_p, \gamma_b)$	Proliferation, Motility and Direction
Model 2	$(m, p, \gamma_p, \gamma_b)$	Proliferation and Direction
Model 3	$(m, p, \gamma_m, \gamma_p)$	Proliferation and Motility
Model 4	(m, p, γ_p)	Proliferation only (Fisher-Kolmogorov [61, 62])
Model 5	(m, p)	None (Skellam [80])

▲ Table 2.1. The hierarchy of models considered. The full model (Model 1) contains a parameter governing the: motility rate, m; proliferation rate, p; motility interaction strength, γ_m; proliferation interaction strength, γ_p; and, directional bias strength, γ_b. In subsequent models, we restrict various combinations of the parameters to zero, effectively removing them from the model.

spatial structure. We subdivide the field-of-view in figure 2.3a into 80 vertical sub-regions, each of width $\Delta x = 1900/80 = 23.75 \ \mu$m. We define the density profile $\mathscr{D}(j, t)$ to be the number of agents with an x-coordinate between $(j - 1)\Delta x$ and $j\Delta x$, divided by the area of the sub-region, giving the density. This density profile is shown in figure 2.3c. To avoid capturing excessive noise in our measurement of wound closure, we do not include the entire density profile in the distance metric. Rather, we manually approximate the x-coordinate of the centre of the scratch at $t = 0$ h for each experiment, denoting $I_{\text{mid}}^{(i)}$ as the bin index of the centre the scratch in experiment i. We include the central 41-subregions which, in effect, surround the initially scratched region of each experiment. This region is indicated in figure 2.3c and avoids the fluctuations in density outside this region.

The distance metric, $d(X_{\text{obs}}, X_{\text{sim}})$, is given by

$$d(X_{\text{obs}}, X_{\text{sim}}) = \sum_{t \in \{18, 30\}} \left(\frac{[N_{\text{sim}}(t) - N_{\text{obs}}(t)]^2}{N_{\text{obs}}(t)^2} \right.$$
$$+ \frac{\sum_{j=1}^{20} [\mathscr{P}_{\text{sim}}(j, t) - \mathscr{P}_{\text{obs}}(j, t)]^2}{\sum_{j=1}^{20} \mathscr{P}_{\text{obs}}(j, t)^2} \tag{2.11}$$
$$\left. + \frac{\sum_{j=I_{\text{mid}}-20}^{I_{\text{mid}}+20} [\mathscr{D}_{\text{sim}}(j, t) - \mathscr{D}_{\text{obs}}(j, t)]^2}{\sum_{j=I_{\text{mid}}-20}^{I_{\text{mid}}+20} \mathscr{D}_{\text{obs}}(j, t)^2} \right),$$

and includes information from all three summary statistics, at $t = 18$ h and $t = 30$ h. Therefore, $d(X_{\text{obs}}, X_{\text{sim}})$ is the relative square error of the simulation from the experiment. For $\mathscr{P}$ and $\mathscr{D}$, the contributions to $d(X_{\text{obs}}, X_{\text{sim}})$ approximate the relative square error in the integral of each summary statistic, given the spatial discretisation we have applied to each.

2.2.4 Approximate Bayesian computation and model selection

We consider a hierarchy of models. The full model, which we denote as Model 1, contains the five unknown parameters $\boldsymbol{\theta}_1 = (m, p, \gamma_m, \gamma_p, \gamma_b)$. Models 2 to 5 are subsets of the full model, where we progressively restrict various combinations of the interaction strength parameters γ_m, γ_p and γ_b to be zero, effectively removing them from the model. We summarise these five models in table A.2, where we denote $\boldsymbol{\theta}_k$ as the unknown parameter combination for Model k.

We treat the unknown parameters in each model as a random variable, $\boldsymbol{\theta}$. In the absence of experimental

observations, our knowledge of $\boldsymbol{\theta}$ is characterised by specified prior distributions. When included in the model, the priors were chosen to be independent and are as follows: $\pi(m) = U(0, 10)\ \mathrm{h}^{-1}$; $\pi(p) = U(0.02, 0.05)\ \mathrm{h}^{-1}$; $\pi(\gamma_m) = U(-2, 2)\ \mathrm{h}^{-1}$; $\pi(\gamma_p) = U(0, 0.02)\ \mathrm{h}^{-1}$; and $\pi(\gamma_b) = U(0, 100)\ \mu m$. In this context, $\pi(\cdot)$ represents a probability distribution. In Appendix A, we show that widening these priors has negligible effect on the results. Initially, we also treat σ as an unknown parameter where $\pi(\sigma) = U(2, 30)\ \mu m$. This initial analysis provides strong evidence for the value of σ, so we set $\sigma = \varphi/2 = 12\ \mu m$ to decrease the dimensionality of the parameter space. In Appendix A, we also investigate $\sigma = \varphi = 24\ \mu m$, since this is a natural choice in a lattice-based framework where the migration distance and dispersal distance are also the same as the average agent diameter. We apply approximate Bayesian computation (ABC) [34, 59, 70, 72] to update our knowledge of the parameters using experimental observations, $\mathcal{X}_{\mathrm{obs}}$, from all nine experiments, to produce posterior distributions, $\pi(\boldsymbol{\theta}|\mathcal{X}_{\mathrm{obs}})$. Since this model is known to be computationally expensive [34] and we have a high-dimensional parameter space, we apply an ABC method based on sequential Monte-Carlo (SMC) [70, 72, 81].

In this study, we aim to find parameter combinations that simultaneously match all nine experimental data sets, such that $\mathcal{X}_{\mathrm{obs}} = \{\mathrm{X}_{\mathrm{obs}}^{(i)}\}_{i=1}^{9}$. For each prior sample in the ABC rejection algorithm we simulate a model realisation using each experimental initial condition, to obtain $\mathcal{X}_{\mathrm{sim}} = \{\mathrm{X}_{\mathrm{sim}}^{(i)}\}_{i=1}^{9}$. We then compare observed data, $\mathcal{X}_{\mathrm{obs}}$, to simulated data, $\mathcal{X}_{\mathrm{sim}}$, using the discrepancy measure

$$\rho(\mathcal{X}_{\mathrm{obs}}, \mathcal{X}_{\mathrm{sim}}) = \sum_{i=1}^{9} d\left(\mathrm{X}_{\mathrm{obs}}^{(i)}, \mathrm{X}_{\mathrm{sim}}^{(i)}\right), \tag{2.12}$$

where $d(\cdot, \cdot)$ is given in equation (A.1). In ABC techniques, we accept a proposal as a posterior sample if $\rho(\mathcal{X}_{\mathrm{obs}}, \mathcal{X}_{\mathrm{sim}}) < \varepsilon$ for some threshold ε. As $d(\cdot, \cdot) \geq 0$, the sum in equation (2.12) is non-decreasing in i. We therefore implement early rejection [82] by sequentially producing model realisations for $i \in \{1, ..., 9\}$. If, at any time, the partial sum up to a value i exceeds the threshold ε, we immediately reject the sample. In practice, this saves considerable computation time by reducing the number of times the model must be simulated using high-density initial conditions.

The principle behind ABC SMC is to propagate a series of prior samples, called *particles*, through a sequence of distributions $\pi(\boldsymbol{\theta}|\rho(\mathcal{X}_{\mathrm{obs}}, \mathcal{X}_{\mathrm{sim}}) < \varepsilon_u), u = \{1, ..., U\}$ [70, 72, 81]. The thresholds ε_u satisfy $\varepsilon_u > \varepsilon_{u+1}$, so that the distribution gradually evolves to the target distribution $\pi(\boldsymbol{\theta}|\rho(\mathcal{X}_{\mathrm{obs}}, \mathcal{X}_{\mathrm{sim}}) < \varepsilon_U) \approx \pi(\boldsymbol{\theta}|\mathcal{X}_{\mathrm{obs}})$. To obtain a sequence of thresholds, and an estimate of the smallest discrepancy possible in all models, we first perform a pilot run using ABC rejection [34, 72] with Model 1 (Appendix A, Section A.2). From 100,000 prior samples, this provides an estimate of the probabilities $\Pr(\rho(\mathcal{X}_{\mathrm{obs}}, \mathcal{X}_{\mathrm{sim}}) < \varepsilon_u)$, given $\boldsymbol{\theta}$ is simulated from the prior. We choose the sequence $\{\varepsilon_u\}_{u=1}^{U}$ by examining a quantile plot (Appendix A, Section A.5). We choose ε_U to corresponds to an acceptance rate of approximately 1% under ABC rejection. The sequence of discrepancies, and details of the ABC rejection and SMC algorithms are given in Appendix A.

We follow the ABC SMC algorithm of Toni *et al.* [70] to perform parameter inference and model selection. Under this algorithm, we place a prior distribution on the model index, $\pi(M_k)$, which we choose to be a discrete uniform distribution so that each model is equiprobable. ABC SMC is then used to estimate the posterior probability of each model, $\pi(M_k|\mathcal{X}_{\mathrm{obs}})$. We detail this algorithm in Appendix A. A key feature of this technique is to implicitly penalise models with a higher number of parameters. We compare models by computing the Bayes factor, $\mathcal{B}_k$ [83], which describe the *evidence* in favour of Model k over the full model,

Model 1. As a uniform prior is placed on the model index, the Bayes factor is given by

$$\mathcal{B}_k = \frac{\pi(M_k|\mathcal{X}_{\text{obs}})}{\pi(M_1|\mathcal{X}_{\text{obs}})}. \tag{2.13}$$
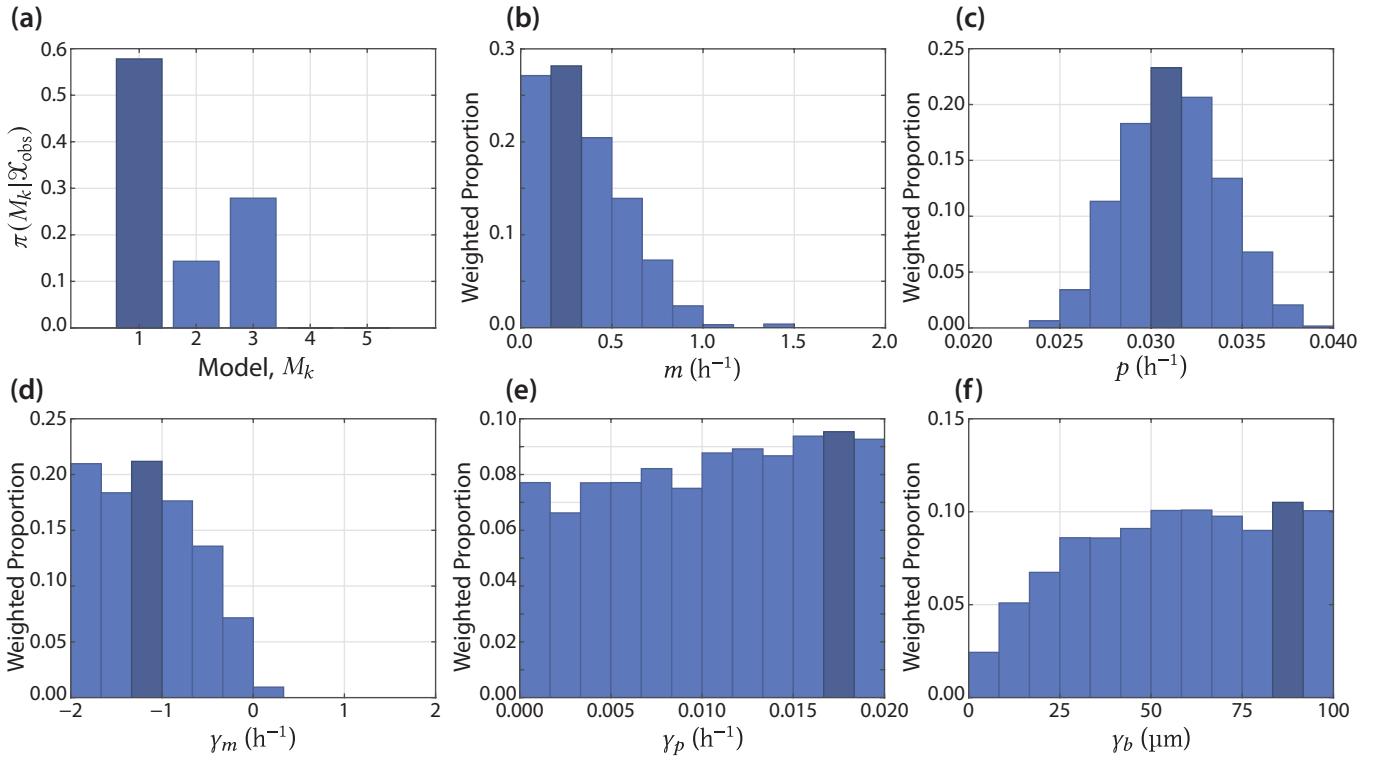
Here, $\pi(M_k|\mathcal{X}_{\text{obs}})$ denotes the marginal posterior density of M_k (Model k). A value $\mathcal{B}_k > 1$ indicates evidence in favour of Model k compared to the full model, and vice-versa for $\mathcal{B}_k < 1$. The Bayes factor is therefore simply the ratio of the posterior density for Models k and 1, and provides evidence to compare models in a similar way to that used in frequentist hypothesis testing.

2.3 Results and Discussion

Common mean-field models, such as the Fisher-Kolmogorov equation [62] and its generalisations, are not able to simultaneously describe collective cell behaviour in scratch assay experiments across a range of initial densities [40]. This suggests density-dependent behaviour in these experiments. Our model allows interactions between cells to affect proliferation, movement and direction. To identify the importance of each of these interactions, we simultaneously calibrate our model to nine scratch assay experiments which we initiate across a wide range of initial densities. We always initiate the IBM using the initial configuration of cells in the experiments and perform inference using data at an intermediate time point, $t = 18$ h, and at the conclusion of the experiment, $t = 36$ h. In a set of preliminary results (not shown) we only included data at the last time point, $t = 36$ h, and the inclusion of the intermediate time point made negligible difference to the results. Therefore, we do not expect the results to change significantly should more than two time points be considered.

Our first result is to identify the distance over which these interactions occur. We quantify interactions using Gaussian kernels dependent on the distance between pairs of agents [76], and characterised by a spread parameter σ (equations (2.3), (2.8) and (2.9)). The interaction between a pair of agents separated by more than approximately 3σ is, therefore, negligible. We expect σ to be of the same order of magnitude as $\varphi = 24\,\mu$m, which is the approximate cell diameter [34]. We perform ABC rejection where σ is sampled from the prior $U(2, 30)$ (Appendix A, Section A.4.1). These results suggest that $\sigma \approx \varphi/2 = 12\,\mu$m, and we fix this for the rest of the study to reduce the number of unknown parameters. This result suggests that interactions between cells occurs over a relatively short distance, since the model predicts interactions between cells separated by more than $3\sigma = 36\,\mu$m is negligible.

One of the most important aspects of the lattice-free IBM is its ability to describe, in fine detail, the spatial structure of cells in the experiments, which we quantify using the pair correlation function. In contrast, mean-field models consider only average properties of the cell population [63] and lattice-based methods [43, 71] are not able to precisely capture the initial agent configuration from the experiments. Lattice-based methods also, by definition, constrain the separation of agents to take discrete values, and typically agents in these models cannot lie closer than one cell diameter. The pair correlation describes the probability of finding pairs of agents separated by each distance, and hence can provide information about the effect of interactions on the dynamics. To show this, we repeat ABC rejection but exclude the pair correlation function from the distance metric (Appendix A, Section A.4.3). These results show that the posterior distributions change significantly in this case, verifying that the pair-correlation function contains

(a) Posterior for the model index, $\pi(M_k|\mathcal{X}_{\mathrm{obs}})$, showing that Model 1 (the full model) is the posterior mode. **(b–f)** Marginal posterior distributions for each parameter in Model 1, shown as weighted histograms. In all cases, the posterior mode is indicated in dark blue.

a significant amount of information about these interactions.

To quantitatively determine the importance of each interaction, we consider a hierarchy of models where we successively set interaction strength parameters (γ_m, γ_p and γ_b) to zero to remove the corresponding interaction from the model. We use the model selection algorithm of Toni *et al.* [70], and compare the evidence in favour of each model over the full model (Model 1) using Bayes factors [70, 83]. We show the posterior density for each model in figure 2.4a, and summarise the Bayes factors and evidence in table A.3. Overall, we find that Model 1 has the highest posterior density (figure 2.4a). We find positive evidence in favour of Model 1 over Model 2 (where $\gamma_m = 0$ and so motility is density-independent); and weak evidence in favour of Model 1 over Model 3 (where $\gamma_b = 0$ and so there is no directional bias). Importantly, we find that Models 4 and 5, where $\gamma_m, \gamma_b = 0$ and $\gamma_m, \gamma_p, \gamma_b = 0$, respectively, cannot match the experimental data ($\mathscr{B}_4 = \mathscr{B}_5 = 0$). Contrary to assumptions that are commonly made in models such as the Fisher-Kolmogorov equation, these results provide evidence to suggest that motility is density-dependent, as either a density dependent movement rate must be included (Models 1 and 3) or a directional bias (Models 1 and 2).

|	Bayes factor, $\mathscr{B}_k$	Evidence in favour of Model 1
Model 1	1.000	–
Model 2	0.248	Positive
Model 3	0.482	Weak
Model 4	0.000	Very Strong
Model 5	0.000	Very Strong
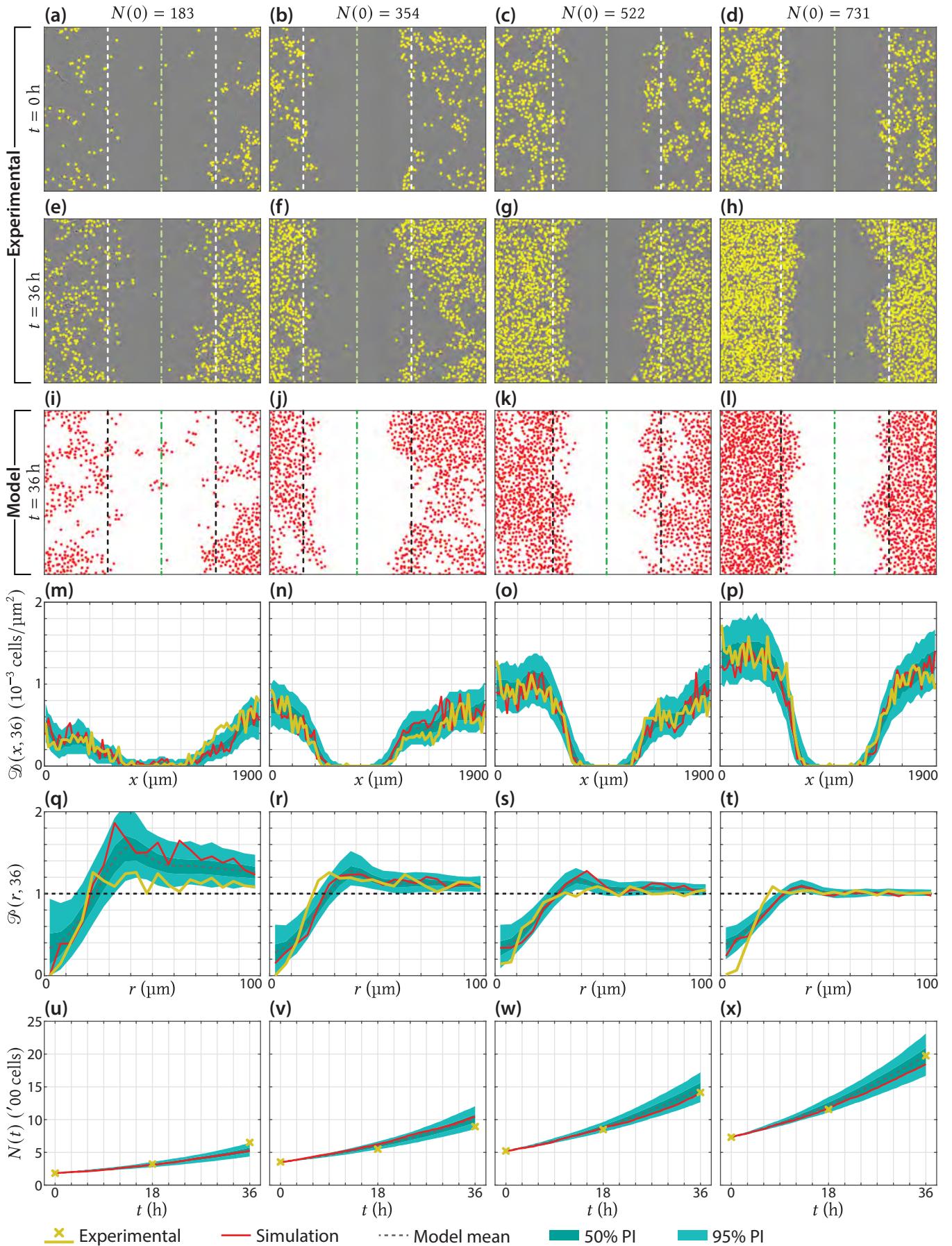
◀ Table 2.2. Bayes factor for each model, which describes the evidence in favour of Model 1 over Model k. A Bayes factor close to 1 indicates limited evidence in favour of Model 1 over Model k, and a Bayes factor close to 0 indicates very strong evidence in favour of Model 1 over Model k [70].

We now focus on results for the full model (Model 1), which has the highest posterior density. In figure 2.4b–f we show marginal posterior distributions for each parameter in Model 1, and in figure 2.5 we compare the experimental data from four of the nine experiments to the calibrated model (in Appendix A.4.3, we show these results for all nine experiments). Overall, we find an excellent match between the model and experimental data, which has not been seen across a range of initial densities for this kind of experimental data. In addition to matching the density profile (figure 2.5m–p) and population (figure 2.5u–x), we find that the calibrated IBM is able to capture information about the spatial structure of cells, specifically, the pair correlation function (figure 2.5q–t). We perform a posterior predictive check for each summary statistic by producing 50% and 95% prediction intervals (PI) that characterise both the parameter uncertainty and stochasticity described by the model. The summary statistics produced from the experimental data almost always lie completely within the 95% PI, further indicating that the calibrated model is consistent with the experimental data across the range of initial densities. While we have not presented these results for Models 2 and 3, which have non-zero posterior density, the nature of ABC means that all accepted samples lie a similar distance to the experimental data.

Results in figure 2.4d suggest that $\gamma_m < 0$, so that crowding *increases* motility. This is consistent with mean-field models such as the porous Fisher equation [84] where the diffusivity increases with local density, but contrasts to other non-linear diffusion models where cell motility decreases with crowding [85]. This observation also explains why model realisations with small values of the motility rate, m, are able to match the data (this is seen in figure 2.4b), since a value $\gamma_m < 0$ allows motility in crowded regions if $m \ll 1$. Interestingly, these results are less clear in the case where the pair correlation function is neglected (Appendix A, Section A.4.3), which highlights the importance of considering spatial structure when studying these interactions. The increase of motility due to crowding may correspond to mechanical interactions, such as volume exclusion, in regions of very high cell density. It is trivial to add mechanisms to the IBM, and future work may examine γ_m in the case volume exclusion [43], or other kinds of mechanical interactions [86–88], are included as additional mechanisms. Alternatively, the inclusion of non-monotonic interaction kernels [89] may allow movement to increase for agents close together, and decrease in crowded regions.

An interesting result is that the directional bias is included in the models with the highest posterior density (Models 1 and 3), but examining the marginal posterior for γ_b (figure 2.4f), we see that the strength of this bias may not be identifiable: the posterior distribution is relatively flat without a clear mode. These results might suggest that, past a certain point, increasing the strength of the directional bias has negligible effect. We verify these observations Appendix A.4.3. To obtain more information about the strength of the directional bias, more detailed data, such as time-lapse cell tracking data, may be required [88].

Results in figure 2.4e indicate that the proliferation interaction strength parameter, γ_p, appears to be unidentifiable [27], since the posterior distribution contains no well defined maxima. Figure 2.5u–x shows that population growth in both the experiments and calibrated model appears to be exponential, so we do not see crowding effects on proliferation in these experiments [71, 90]. We verify this by performing model selection with three additional models (Models 6–8) that respectively correspond to Models 1–3 with $\gamma_p = 0$ (Appendix A, Section A.6). These additional results show that the distributions for Models 6–8 are similar to those for Models 1–3 and confirm that crowding effects on proliferation, such as contact inhibition, are simply not seen in these experiments. Early-time data often illustrates exponential growth for a variety of growth laws and experiments must be conducted over a longer period of time to identify the appropriate

▲ **Figure 2.5.** Comparison between experimental data and model estimate. Dashed lines in (a–l) give the scratched region that was used to compare the observed and simulated density profile, where green dash-dot lines give the approximate centre of the scratch at $t = 0$ h. Simulation results in (i–x) refer to a single realisation of the IBM using the parameter combination that gave the smallest discrepancy. In (m–x) the model mean, 50% prediction interval (PI) and 95% PI were produced using 500 realisations of the IBM using parameters resampled from the posterior distribution. Results for all nine experiments are provided in Appendix A.8.

growth function [71].

2.4 Conclusion

The ability of common mean-field models, such as the Fisher-Kolmogorov equation and its generalisations, to match experimental data across a range of densities is rarely tested as typical experimental protocols do not vary the initial number of cells. These models typically assume either or both density-dependent proliferation and density-independent motility [36–38, 40, 55–57]. By modelling density-dependent interactions which affect motility, proliferation and directional bias, we calibrate a mathematical model that simultaneously describes scratch assay data across a range of densities. Using model selection, we quantitatively assess which interactions are most important. Our results provide an indication of how density affects the behaviour cells in our experiments. In opposition to common modelling assumptions [37], proliferation appears to be unaffected as density increases whereas cells become more motile as density increases. This information provides a hint about the kind of partial differential equation that is most appropriate, such as

$$\frac{\partial c}{\partial t} = \underbrace{\frac{\partial}{\partial x}\left(D(c)\frac{\partial c}{\partial x}\right)}_{\substack{\text{Contact stimulated} \\ \text{motility}}} + \underbrace{\lambda c,}_{\substack{\text{Density-independent} \\ \text{proliferation}}} \qquad D'(c) > 0. \tag{2.14}$$

may be a more appropriate mean-field model for this kind of experimental data. Our findings agree with some areas of the literature [66–68], but contrast with studies that suggest density decreases motility [64, 65], or that motility is independent of density [40, 63]. Our finding that exponential growth describes the data is consistent with observations by Vittadello *et al.* [91], who point out that a loss of contact inhibition is a hallmark of cancer [92]. Despite this, the assumption of logistic growth is common in the mathematical modelling literature [74, 93, 94]. Our study demonstrates that exponential and logistic growth are not always distinguishable from typical experimental data [71]. Parameter identifiability [27] should be considered when calibrating logistic growth models to scratch assay data.

We study collective cell behaviour using a mathematical model which incorporates density-dependent interactions affecting proliferation, motility and directional bias. Applying SMC, which penalises models with high dimensionality of the unknown parameters, our study suggests the minimal model required to match the experimental data. Two of the primary advantages of our IBM approach is the ability to precisely replicate the experimental initial condition; and, the ease of which new mechanisms can be incorporated into, and removed from, the model. Our approach can, therefore, be applied to quantify experimental evidence for more complex mechanisms including chemotaxis [59, 95], mechanotaxis [96], and generalised growth laws [93], as well as comparing more complicated choices of interaction kernel [75]. Cell aspect ratio [67] can be incorporated into the model using asymmetric choices of interaction kernels, however would require more detailed experimental data, such as that provided by machine vision. We do not pursue such extensions here since we find that our simpler modelling framework already provides a good match to experimental data across a range of densities.

Chapter 3

Persistence as an optimal hedging strategy

Published 2021 in *Biophysical Journal*.

Bacteria invest in a slow-growing subpopulation, called persisters, to ensure survival in the face of uncertainty. This hedging strategy is remarkably similar to financial hedging, where diversifying an investment portfolio protects against economic uncertainty. We provide a new theoretical foundation for understanding cellular hedging by unifying the study of biological population dynamics and the mathematics of financial risk management through optimal control theory. Motivated by the widely accepted role of volatility in the emergence of persistence, we consider several novel models of environmental volatility described by continuous-time stochastic processes. This allows us to study an emergent cellular hedging strategy that maximises the expected per-capita growth rate of the population. Analytical and simulation results probe the optimal persister strategy, revealing results that are consistent with experimental observations and suggest at new opportunities for experimental investigation and design. Overall, we provide a new way of conceptualising and modelling cellular decision-making in volatile environments by explicitly unifying theory from mathematical biology and finance.

Statement of Contribution of Co-Authors for
Thesis by Published Paper

The authors listed below have certified that:

1. they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
2. they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
3. there are no other authors of the publication according to these criteria;
4. potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit, and
5. they agree to the use of the publication in the student's thesis and its publication on the QUT's ePrints site consistent with any limitations set by publisher requirements.

In the case of this chapter:
Persistence as an optimal hedging strategy
Published 2021 in Biophysical Journal
doi:10.1016/j.bpj.2020.11.2260

Contributor	Statement of Contribution
Alexander P Browning 4 Aug 2021	Designed the study, conceived the optimal control model of bacterial persistence, developed and implemented the computational algorithms, generated and interpreted results, drafted the manuscript, and revised the manuscript during peer-review.
Jesse A Sharp	Designed the study, aided development of the computational algorithms, interpreted the results and provided comments on the manuscript.
Tarunendu Mapder	Designed the study, aided development of the computational algorithms, interpreted the results and provided comments on the manuscript.
Christopher M Baker	Interpreted the results and provided comments on the manuscript.
Kevin Burrage	Designed the study, supervised the research, interpreted the results, and provided comments on the manuscript.
Matthew J Simpson	Designed the study, supervised the research, interpreted the results, and provided comments on the manuscript.

Principal Supervisor Confirmation

I have sighted email or other correspondence from all Co-authors confirming their certifying authorship.

Matthew J Simpson
Name

Signature

9 August 2021
Date

3.1 Introduction

Shortly after the clinical introduction of penicillin, Bigger noticed a resistance in a small subpopulation of *Staphylococcal pyogenes* [97]. The resistant cells, termed *persisters*, are a genetically identical, slow-growing, phenotypic variant. Bacteria invest in persisters to ensure survival: persisters are less proliferative than regular cells in a nutrient rich environment but can withstand adversity [98]. These strategies are akin to financial hedging, where diversifying an investment portfolio protects against economic uncertainty [17, 99–101]. This similarity is widely acknowledged: parts of the biology and ecology literature refer to this phenomena as *bet-hedging* [98, 100, 102–104]. In this work, we introduce the concept of *cellular hedging*, where we explicitly model bacterial persistence using techniques from mathematical finance, including stochastic differential equations (SDEs) and stochastic optimal control theory.

Bacterial persistence poses a significant clinical challenge and is highly advantageous to bacteria. For example, persisters are less sensitive to antibiotic treatment [16, 105–108], are undetectable in routine clinical tests [109], and are thought to be responsible for the formation [110] and incurability [111] of many infections. Antimicrobial treatments will, therefore, benefit from an understanding of how persisters arise and function [112]. The similarity between bacterial persistence and financial hedging strategies suggests a new pathway to investigate persister dynamics. In this study, we provide a novel, quantitive understanding of persister strategies using techniques from financial mathematics and stochastic optimal control theory. Persister production is known to depend on the environment [16, 113, 114] and our mathematical modelling approach allows us to unearth how various optimal persister strategies depend upon environmental volatility.

Persisters can be revealed experimentally by disinfecting a population of *Escherichia coli* (*E. coli*) with ampicillin, which targets proliferative cells [16]. The initially rapid reduction in population size eventually slows, revealing a small subpopulation that is less sensitive to the treatment. Experimental investigations are complicated by the extreme scarcity of persister cells: typically less than 1 in 10^5 cells are persisters in wild type *E. coli* [115]. For subpopulations of this size, stochastic effects are significant [22, 116].

In producing persisters, bacteria allocate resources to hedge against environmental volatility for the purpose of survival. We assume that these processes occur in much the same way that a financial investor hedges a portfolio to protect from, and take advantage of, economic volatility [117]. The archetypal example of portfolio diversification in finance is Merton's portfolio problem (MPP) [17]. Here, an investor allocates a fraction of their wealth in a high-yield volatile asset, such as stocks; and a low-yield stable asset, such as government bonds. Analogously, we suppose that bacteria have evolved mechanisms to regulate an allocation of their total population as proliferative although susceptible, and some as persisters. In the finance problem, market volatility is modelled such that the underlying price of each asset can be described as an Itô SDE driven by Wiener noise [17, 118], similar to models of noisy exponential growth proposed for biological problems [119]. More complicated models of market volatility have been extensively explored in the finance literature, such as those that incorporate Poisson jump noise to describe market shocks [120]. In Merton's original work, stochastic optimal control theory [118] is used to maximise the investors wealth by modelling the proportion allocated to each asset as a *control*. Merton revealed that it is not advantageous to possess the low-yield asset in the absence of uncertainty, or when the growth of the high-yield asset is deterministic. MPP revolutionised the field of mathematical finance, and many of the ideas in MPP formed the basis of Merton's later work on options pricing with Black and Scholes [121, 122] that led to the the 1997 Nobel Prize

in Economics.

Current mathematical models of persistence typically describe environmental volatility with a growth rate that transitions through finitely many states (such as a growth and stress state). Transitions between these states are assumed to occur either periodically [22, 103, 123] or stochastically [99, 100, 124]. Many existing studies maximise some measure of the long-term growth rate of the population and probe the environmental conditions under which persistence is advantageous using, for example, Lyapunov exponents [124]. A limitation of these existing modelling approaches is their inability to capture more general models of environmental volatility. We leverage an explicit connection with mathematical finance to study optimal cellular hedging strategies using stochastic optimal control theory [118, 125]. This approach has several advantages. First, we can probe cellular hedging strategies under more complex models of environmental volatility. Second, we can model the emergence of an environment-dependent persister strategy.
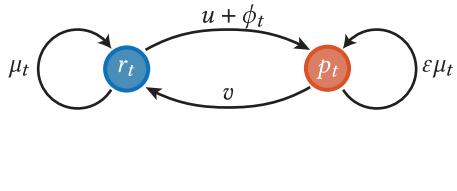
In our study, we first model persistence in a population where the growth rate is subject to continuous stochastic fluctuations in the form of Wiener noise [116, 118, 126]. We posit that, in many cases, this is a more appropriate model of environmental volatility than the discrete transitions, or shocks, that are currently commonplace in mathematical models of persistence [22, 103, 123, 124]. Through this model, we draw a direct connection with mathematical finance and demonstrate how MPP can be applied directly to the biological problem. Next, we expand on this simplistic model through an environment-dependent hedging strategy under more complex models of environmental volatility. We include in our analysis a model that uses Poisson jump noise [125], to capture the aforementioned existing models of environmental volatility [22, 99, 100, 103, 124] where transitions in the growth rate occur due to, for example, shocks. Our modelling framework is readily extensible to any form of environmental volatility that can be described using an Itô SDE.

Our goal is to develop a new framework for studying cellular hedging in response to environmental volatility by unifying the study of biological population dynamics with techniques financial mathematics. It is for this reason we use the term *cellular hedging*, instead of the term *bet-hedging* [98, 100, 102–104]. Despite the established importance of volatility in elucidating bacterial persistence, we find there is currently a scarcity of methods available to describe cellular hedging strategies under environmental volatility. Our mathematical framework demonstrates the importance of considering more complex models of environmental volatility, and we lay theory to complement future experimental studies that probe emergent persister strategies in response to continuously varying stochastic environments. Our new model and approach leads to mathematical results that are consistent with observations from several existing experimental studies, and provides new insights into bacteria dynamics in the face of environmental volatility.

3.2 Methods

3.2.1 Stochastic model of bacteria growth

We describe population-level bacteria dynamics in a volatile environment with a system of SDEs driven by Wiener noise [116, 119, 126]. This choice of model can be thought of as a bridge between discrete Markov models [22, 117], and deterministic ordinary differential equation (ODE) models [16]. In addition, our population-level model of bacterial persistence complements stochastic gene expression models that describe the regulation of persister strategies [101, 127]. Further, our modelling approach can be generalised to can capture many other

◄ Figure 3.1. Schematic of the bacteria growth model with regular cells, r_t, and persisters, p_t. Switching from persisters to regular cells is constant, v. Switching from regular cells to persisters is taken to be the sum of a constant rate, u, and a variable rate, ϕ_t, that depends upon environmental volatility.

forms of environmental noise. For example, we model the dynamics of bacteria in a volatile environment by coupling the growth rate to a stochastic process representing the environment [128].

We model exponential growth in a population composed of regular (non-persister) cells, r_t, and persisters, p_t. Here, a subscript t indicates that each stochastic processes depends on time. The net growth rate of regular cells is a stochastic process with expectation μ_t and amplitude σ. The quiescence of persisters manifests as slow metabolic activity, so we assume the expected growth rate of persisters is some small proportion, $\varepsilon \ll 1$, of regular cells. The net persister growth rate is, therefore, a stochastic process with expectation $\varepsilon\mu_t$ and amplitude $\eta = \varepsilon\sigma$. In the absence of subpopulation switching, these dynamics give rise to the system of differential equations [119, 126]

$$\frac{\mathrm{d}r_t}{\mathrm{d}t} = \left(\mu_t + \sigma\xi_t^{(1)}\right) r_t, \tag{3.1a}$$

$$\frac{\mathrm{d}p_t}{\mathrm{d}t} = \left(\varepsilon\mu_t + \eta\xi_t^{(2)}\right) p_t. \tag{3.1b}$$

Here, $\xi_t^{(1)}$ and $\xi_t^{(2)}$ are independent Gaussian white noise processes. The expected growth rate of regular cells is given by μ_t, which may be constant, or itself a stochastic process. We interpret Equation 3.1 in the Itô sense for two reasons. First, the properties of the Itô integral make it well suited to models in population biology [119, 129]. Second, Itô SDEs are widely applied in mathematical finance, including in MPP [17], where the property $\mathbb{E}\left[\int x_t \, \mathrm{d}W_t\right] = 0$, in the Itô sense, is vital.

The current literature classifies two types persister production in a population: variable and environment dependent, commonly referred to as Type I; or constant, commonly referred to as Type II [16] (figure 3.1). It is understood that constant production of persisters, which results in an approximately constant proportion in a growing population, may be regulated by stochastic gene fluctuation on a single cell level [130–132]. In addition, a growth feedback mechanism [21] — possibly regulated by quorum sensing [133] and intercellular signalling [134] — may enable cells to respond and vary the persister production rate. In our model, regular cells switch to persisters at a rate of $u + \phi_t$, where: $u \geq 0$ is the constant rate [16]; and $\phi_t \geq 0$ is the variable, environment-dependent, rate. Persisters revert to regular cells at a constant rate $v \geq 0$. Our model can, therefore, describe both Type I and Type II strategies.

Incorporating subpopulation switching into the growth equations (equations (3.1)) yields

$$\mathrm{d}r_t = \mu_t r_t \, \mathrm{d}t + \sigma r_t \, \mathrm{d}W_t^{(1)} - \left[(u + \phi_t)r_t - vp_t\right] \mathrm{d}t, \tag{3.2a}$$

$$\mathrm{d}p_t = \underbrace{\varepsilon\mu_t p_t \, \mathrm{d}t + \eta p_t \, \mathrm{d}W_t^{(2)}}_{\text{Subpopulation growth}} + \underbrace{\left[(u + \phi_t)r_t - vp_t\right] \mathrm{d}t}_{\text{Phenotype switching}}. \tag{3.2b}$$

Such a system of equations is said to be in the form of multiplicative Wiener noise [125]. We note that

in removing variability from the model by setting $\sigma = \eta = 0$, we recover the ODE model of Balaban *et al.* [16]. Many other choices for a stochastic population dynamics model exist [116], such as models that consider intrinsic noise caused by subpopulation switching [135]. However, the focus of the current work is on fluctuations in the environment and growth rate, which we assume to be independent of the population size.

We find it natural to consider the variable transformation

$$n_t = r_t + p_t, \quad \theta_t = \frac{p_t}{n_t}, \tag{3.3}$$

such that n_t represents the population size and θ_t represents the proportion of persisters. Following Itô's lemma [118], the transformed state equations are

$$\mathrm{d}n_t = \big[1 - \theta_t(1 - \varepsilon)\big]\mu_t n_t \,\mathrm{d}t \tag{3.4a}$$
$$+ \sigma n_t(1 - \theta_t) \,\mathrm{d}W_t^{(1)} + \eta n_t \theta_t \,\mathrm{d}W_t^{(2)},$$
$$\mathrm{d}\theta_t = \Big[(1 - \theta_t)(u + \phi_t) - \theta_t v - (1 - \theta_t)\theta_t\big((1 - \varepsilon)\mu_t \tag{3.4b}$$
$$- \sigma^2(1 - \theta_t) + \eta^2\theta_t\big)\Big] \,\mathrm{d}t$$
$$- (1 - \theta_t)\theta_t \sigma \,\mathrm{d}W_t^{(1)} + \eta(1 - \theta_t)\theta_t \,\mathrm{d}W_t^{(2)},$$

revealing that the dynamics of the persister proportion, θ_t, are independent of the population size, n_t.

Stochastic model environmental volatility

We model environmental volatility by assuming $\mu_t = m(\zeta_t)$, where ζ_t is a stochastic process that represents a volatile environment. There are many appropriate choices for $m(\cdot)$ and ζ_t, but we focus our analysis on three environments, samples paths of each shown in figure 3.2a–c. These are

1. *Constant.* (figure 3.2a) We set $m(\zeta) \in \{\mu_G, \mu_S\}$, such that μ_t is constant. Here μ_G represents a colony during growth; and μ_S represents a colony under stress or antimicrobial treatment. For numerical results, we choose $\mu_G = 2\,\mathrm{h}^{-1}$ and $\mu_S = -2\,\mathrm{h}^{-1}$ to match experimental data for *E. coli* during growth and ampicillin treatment [16].

2. *Monod.* (figure 3.2b) Monod kinetics are commonly used to model the growth of bacteria [136–138], and feature dynamics with an asymptotic upper bound on the growth rate, but no lower bound. We describe this environment by a mean-reverting Ornstein-Uhlenbeck process [118] and couple the growth rate using a Monod equation [136],

$$m(\zeta_t) = \frac{\mu_{\max}\zeta_t}{K_\zeta + \zeta_t} - \delta, \tag{3.5a}$$
$$\mathrm{d}\zeta_t = \gamma(K_\zeta - \zeta_t) \,\mathrm{d}t + \kappa \,\mathrm{d}W_t^{(3)}, \tag{3.5b}$$

where $W_t^{(3)}$ is a Wiener process independent of both $W_t^{(1)}$ and $W_t^{(2)}$. Here, ζ_t experiences fluctuations proportional to κ and a reversion force to the state $\zeta_t = K_\zeta$ of strength proportional to γ. The Monod

coupling features an asymptotic upper-bound on the growth rate of $\mu_t = \mu_{max}$ and no lower bound. Unfavourable environmental changes have, therefore, a larger effect on the growth rate than those that are favorable, and the growth rate experiences reversion to a growth rate $\mu_t = \mu_{max}/2 - \delta$, where δ represents the natural death rate. For numerical results, we choose $\gamma = 0.1\,\text{h}^{-1}$, $K_\zeta = 1$, $\zeta_0 = 1$, $\kappa = 0.3$, $\mu_{max} = 8\,\text{h}^{-1}$ and $\delta = 2\,\text{h}^{-1}$. For this choice of parameters the growth rate experiences reversion to a growth rate of $\mu_G = 2\,\text{h}^{-1}$, which corresponds to the constant environment. To deal with the discontinuity at $\zeta = -K_\zeta$, we truncate equation (3.5a) so that $m(\zeta) = m(-0.5)$ for $\zeta < -0.5$.

3. *Poisson.* (figure 3.2c) An existing class of mathematical models describe environmental uncertainty using alternating periods of growth and stress [22, 99, 100, 103, 124]. We reproduce this type of environment in a stochastic model by assuming that the environment switches between growth and stress according to a Poisson process. We consider that $m(\zeta) = \mu_G$ for $\zeta \geq 0$; $m(\zeta) = \mu_S$ for $\zeta < 0$; and model the environment as the jump or telegraph process [125]

$$\mathrm{d}\zeta_t = \rho(\zeta_t)\,\mathrm{d}P_t\,, \quad \mathrm{d}P_t \sim \mathrm{Po}(\mathrm{d}t\,\lambda(\zeta_t)), \tag{3.6a}$$

$$\text{where } \rho(\zeta_t) = \begin{cases} -2, & \zeta_t = +1, \\ +2, & \zeta_t = -1, \end{cases} \tag{3.6b}$$
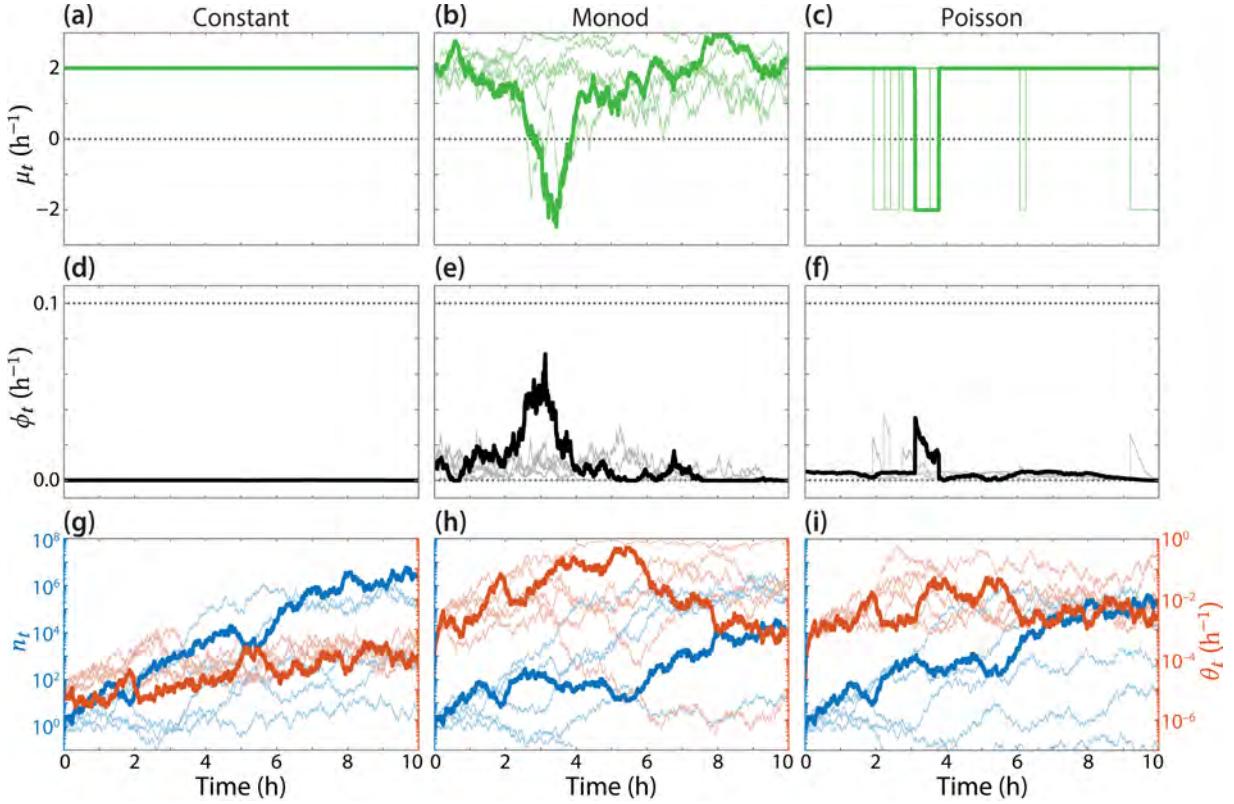
$$\text{and } \lambda(\zeta_t) = \begin{cases} 1/\tau_G, & \zeta_t = +1, \\ 1/\tau_S, & \zeta_t = -1. \end{cases} \tag{3.6c}$$

Here, P_t is a Poisson process with intensity $\lambda(\zeta_t)$. This formulation allows for a mean time of τ_G in the growth phase, where $\mu_t = \mu_G$, and a mean time of τ_S in the stress phase, where $\mu_t = \mu_S$. For numerical results in this study, we choose $\tau_G = 9.5\,\text{h}$, $\tau_S = 0.5\,\text{h}$, $\zeta_0 = 1$, $\mu_G = 2\,\text{h}^{-1}$ and $\mu_S = -2\,\text{h}^{-1}$. This model of environmental volatility can be readily extended to capture more general forms of Poisson noise leading to discontinuous fluctuations in the environment due to, for example, shocks.

These choices represent environments in which changes happen gradually (Monod environment) or abruptly (Poisson environment), as demonstrated in figure 3.2b and figure 3.2c, respectively. For comparative purposes, we demonstrate results using realisations of each environment that are similar, and choose the initial condition for all environments to correspond to the initial growth rate $\mu_0 = \mu_G = 2\,\text{h}^{-1}$. In Appendix B.2, we explore results for two other models of environmental volatility: (4) an Ornstein-Uhlenbeck process without a Monod coupling (figure B.1d) and, (5) a Duffing oscillator (figure B.1e).

3.2.2 Optimal persister production strategies

Our ideas of *cellular hedging* suggest that a cell population has developed a persister production strategy $\mathcal{S} = \{u, v, \phi_t\}$ that is optimal in some way. Mathematically, we define optimality as the strategy that maximises some *fitness measure*, which we now construct. We assume that there is no explicit cost to producing persisters with constant rates u and v; but that there is a quadratic running cost to produce persisters with a variable rate ϕ_t. The cost of applying a non-zero ϕ_t accounts for the sensing mechanisms that cells must use to respond to the environment.

▲ Figure 3.2. Optimal persister production in cell populations under various types of environment. (a–c) Growth rates, μ_t, sampled from each environment to simulate bacteria growth in the rest of each column. (d–f) The optimal variable persister production rate, ϕ_t^*. (g–i) The population, n_t, in multiples of the original population (blue, left scale) and persister proportion, θ_t, (red, right scale), for a single realisation of the model. The seeds used to generate the Wiener processes $W_t^{(1)}$ and $W_t^{(2)}$ are identical for all environments. Also shown are five additional, independent, realisations (semi-transparent).

We choose a fitness measure, commonly referred to as a *payoff* in optimal control theory, as

$$J_S = \mathbb{E}\left[\underbrace{\int_0^T \alpha \phi_t^2 \, dt}_{\text{Sensing}} + \underbrace{\log_e(n_T)}_{\text{Growth}} \right], \quad \alpha < 0. \tag{3.7}$$

Here T denotes a terminal time, so that the maximisation is carried out on the interval $t \in [0, T]$; $\alpha < 0$ characterises a trade-off between growth and operating the sensing mechanisms required to vary the persister production rate; and the expectation is taken with respect to the stochastic processes governing subpopulation growth, n_t, and the environment, ζ_t. In this work, we consider a finite terminal time since we model an exponentially growing population. It is not obvious how to incorporate stationary phase dynamics into the stochastic environment model. Future work is needed to explore the infinite time-horizon problem in conjunction with environmental volatility and a logistic growth term [119]. We interpret the terminal time, T, as either the duration of the growth phase of the population [16], or the duration of an experiment. Maximising the logarithmic term in equation (3.7) can be interpreted as cells maximising their per-capita growth rate over the interval $t \in [0, T]$, since

$$\log_e(n_T) = \log_e(n_0) + \int_0^T \frac{1}{n_t} \frac{\mathrm{d}n_t}{\mathrm{d}t} \, \mathrm{d}t \,,$$

and $\log_e(n_0)$ is constant [139]. Our choice of fitness measure corresponds to evolutionary mechanisms that perpetuate highly productive populations [22], meaning cells that implement the optimal strategy carry an evolutionary advantage over those that do not. There are many other choices of fitness, such as resource allocation based on a maximum entropy principle [140]. We expect our methodology to carry across to these more complicated choices of fitness measure.

Together with the state equations (equation (3.4)), maximising equation (3.7) corresponds to an optimal control problem. We apply Hamilton-Jacobi-Bellman (HJB) optimal control theory [125, 141, 142] to reformulate the optimal control problem as a partial differential equation (PDE) problem. We now review essential elements of HJB optimal control theory.

Review of optimal control theory

A stochastic control problem may be stated as

$$\max_{\omega_t \in \mathcal{U}} \underbrace{\mathbb{E}\left(\int_0^T \mathscr{L}(\mathbf{x}_t, \omega_t, t) \, \mathrm{d}t + \Phi(\mathbf{x}_T) \right)}_{\text{Payoff}}, \tag{3.8}$$

where ω_t is a time-dependent *control* from the set of allowable controls, $\mathcal{U}$. The payoff contains terms $\mathscr{L}(\mathrm{x}_t, \omega_t, t)$, which represents cumulative value (in our case, the sensing mechanisms), and $\Phi(\mathbf{x}_T)$, which represents the terminal value (in our case, the final population size). For a stochastic problem, the payoff involves an expectation, which is taken with respect to the stochastic process x_t, representing the system state. The goal is to find the so-called *optimal control*, denoted ω_t^*, that maximises the payoff.

The state equations are given by the three-dimensional stochastic process, $\mathbf{x}_t = (n_t, \theta_t, \zeta_t)$, governed by the three-dimensional Itô SDE

$$\mathrm{d}\mathbf{x}_t = \mathbf{M}(\mathrm{x}_t, \omega_t, t) \, \mathrm{d}t + \Sigma(\mathrm{x}_t, \omega_t, t) \, \mathrm{d}\mathrm{W}_t \,, \tag{3.9}$$

where $\mathbf{M} \in \mathbb{R}^3$ is a vector valued function containing the drift for each state; $\mathrm{d}\mathbf{W}_t \in \mathbb{R}^3$ is a Wiener process of the same dimension as $\mathbf{x}_t$ with independent constituents; and, $\Sigma\Sigma^{\mathrm{tr}} \in \mathbb{R}^{3 \times 3}$ describes the covariance of the Wiener process (superscript $^{\mathrm{tr}}$ denotes the matrix transpose). In our study, $\mathbf{x}_t$ comprises the stochastic processes governing the growth of the bacteria population, (n_t, θ_t), in addition to the independent stochastic process for the environment, ζ_t. We also study a form of $\mathbf{x}_t$ that includes Poisson jump noise, and we include details of optimal control theory in this case in Appendix B.4.2.

To solve the optimal control problem we define a *value function*, $V(\mathbf{x}, s)$, as the optimal payoff obtainable were the system to be at state $\mathbf{x}$ at time s. That is,

$$V(\mathbf{x}, s) = \max_{\omega_t \in \mathcal{U}} \mathbb{E}\left(\int_s^T \mathscr{L}(\mathbf{x}_t, \omega_t, t) \, \mathrm{d}t + \Phi(\mathbf{x}_T) \right), \tag{3.10}$$

subject to $\mathbf{x}_s = \mathbf{x}$.

HJB optimal control theory then describes the time-evolution of V as the solution of a partial differential equation (PDE) [141, 142], given by

$$0 = \max_{\omega_t \in \mathcal{U}} \left(\mathcal{A}V + \mathcal{L}(\mathbf{x}_t, \omega_t, t) \right), \tag{3.11}$$

where $\mathcal{A}$ is the stochastic generator for the process governing $\mathbf{x}_t$. For the three-dimensional system given by equation (3.9), the stochastic generator is given by

$$\mathcal{A}V = \frac{\partial V}{\partial s} + \sum_{i=1}^{3} \left(\mathbf{M}_i \frac{\partial V}{\partial x_i} + \frac{1}{2} \sum_{j=1}^{3} (\Sigma \Sigma^{\mathrm{tr}})_{ij} \frac{\partial^2 V}{\partial x_i \partial x_j} \right) \tag{3.12}$$

where $\mathbf{M}_i$ denotes the ith element of $\mathbf{M}(\mathbf{x}_t, \omega_t, t)$, and $(\Sigma \Sigma^{\mathrm{tr}})_{ij}$ denotes the element in the ith row and jth column of $\Sigma \Sigma^{\mathrm{tr}}$, for $\Sigma = \Sigma(\mathbf{x}_t, \omega_t, t)$. Therefore, mixed partial derivatives are only included in equation (3.12) in the case where the stochastic processes are correlated. The idea behind HJB optimal control theory is that $V(\mathbf{x}, s)$ is known at time $s = T$. Therefore, equation (3.11) is coupled to the terminal condition $V(\mathbf{x}, T) = \Phi(\mathbf{x}_T)$, and solved backwards in time [143].

In certain cases, the argmax($\mathcal{A}V + \mathcal{L}$) term in equation (3.11) may be found analytically through differentiation. This yields both an expression for the optimal control, ω_t^*; and a three-dimensional, non-linear, PDE coupled to a terminal condition. In the case of MPP, where the volatility appears only as geometric Brownian motion in the state equations for (n_t, θ_t), and not in an independent equation for the environment, equation (3.11) has an analytical solution [17]. In most cases, however, we are required to solve equation (3.11) numerically. To obtain an optimal trajectory starting at state $\mathbf{x}_0$, we solve the state equations (equation (3.9)), coupled to the solution of the HJB PDE (equation (3.11)), forward in time using the Euler-Maruyama algorithm [144]. Full details of the numerical techniques are provided in Appendix B.5.

Numerical methods to solve the SDEs

We integrate the SDE models using the Euler-Maruyama algorithm [144] with time step $h = 2 \times 10^{-3}$. To compare results between environments, we fix the seeds used to generate the Wiener processes $W_t^{(1)}$ and $W_t^{(2)}$ when simulating different models. Full details of the numerical method used to solve the SDEs are given in Appendix B.5.

Constant persister production

We first examine a population that can only produce persisters at a constant rate, so we fix the variable rate, $\phi_t = 0$. In this case, we only assume the constant environment so $\mu_t = \mu_G$.

As u and v only appear in the optimal control problem linearly, the solution is not finite unless a bound is enforced on u and v [4]. The unbounded problem corresponds to a cell population that is able to move itself instantaneously to anywhere in the state space. To address this, we assume that the constant switching rates u and v are chosen by the population to control the steady-state persister proportion, which we denote $\hat{\theta}_t \in [0, 1]$. Taking the drift term in the state equation for θ_t (equation (3.4b)) to be zero, we see that u and v are related to $\hat{\theta}_t$ by

$$u = \hat{\theta}_t \left(\frac{v}{1 - \hat{\theta}_t} + \eta^2 \hat{\theta}_t - (1 - \hat{\theta}_t)\sigma^2 + (1 - \varepsilon)\mu_G \right). \tag{3.13}$$

At this point, we note that equation (3.13) is only consistent if $\hat{\theta}_t$ is constant, and we address this shortly.

Allowing $\hat{\theta}_t$ to be a control, the state equation for n_t (equation (3.4a)) becomes

$$\frac{\mathrm{d}n_t}{n_t} = (1 - \hat{\theta}_t(1 - \varepsilon))\mu_G \, \mathrm{d}t + \sqrt{(1 - \hat{\theta}_t)^2\sigma^2 + \hat{\theta}_t^2\eta^2} \, \mathrm{d}W_t, \tag{3.14}$$

since $a \, \mathrm{d}W_t^{(1)} + b \, \mathrm{d}W_t^{(2)}$ can be considered as $\sqrt{a^2 + b^2} \, \mathrm{d}W_t$. Assuming that $\hat{\theta}_t$ is chosen to maximise the fitness measure (equation (3.7)) reveals a problem closely related to MPP [17]. The problems differ intrinsically in that an investor is able to reallocate assets within a financial portfolio instantaneously [117] and without cost — though variations of MPP address this — whereas the cell colony comprises finitely many cells that act heterogeneously using finite switching rates to produce persisters.

We apply stochastic optimal control theory [118, 125] and exploit the similarity between this formulation of the persister problem and MPP [17] to obtain an analytical solution, in which the optimal control, denoted $\hat{\theta}_t^* \in [0, 1]$, is given by

$$\hat{\theta}_t^* = \max\left[0, \min\left(1, \frac{\sigma^2 - (1 - \varepsilon)\mu_G}{\eta^2 + \sigma^2}\right)\right]. \tag{3.15}$$

Full details of this analytical solution are given in Appendix B.3. As is the case with MPP, $\hat{\theta}_t^*$ is constant, so the strategy is independent of the current time, the terminal time, and the system state. Solving equations (3.13) and (B.8) simultaneously gives $\hat{\theta}_t^* = 1.19 \times 10^{-5}$ with $\sigma = 1.414222$ which matches experimental observations for wild type *E. coli* bacteria where $\mu_G = 2 \, \mathrm{h}^{-1}$, $\varepsilon = 0$, $\eta = \varepsilon\sigma = 0$, $u = 1.2 \times 10^{-6} \, \mathrm{h}^{-1}$ and $v = 0.1 \, \mathrm{h}^{-1}$ [16]. Here, we note that $\sigma^2 \approx \mu_G$, since $\hat{\theta}_t^* \ll 1$.

For numerical results in the rest of this study, we fix σ, ε, η, u and v to the aforementioned values, and set the initial persister proportion $\theta_0 = 1.19 \times 10^{-5}$. We set $n_0 = 1$, such that the population is measured relative to the initial population. In practise, these parameters would be obtained by calibrating the stochastic model to experimental data, where we expect σ, and, therefore, u and v, to depend upon the stochastic environment induced by the experimental conditions. Further, setting $\sigma = \eta = 0$ would recover a model in which ζ_t is the only source of environmental volatility.

Variable persister production

We now consider a cell colony that is able vary persister production in response to their environment. In this case, $\phi_t \in \mathbb{R}^+$, can be thought of as a Markovian control [118], or a control that is chosen based on information that includes the current state and time, but does not carry a memory about the past state. For numerical stability, we place an upper-bound on the control such that $\phi_t \le 0.1 \, \mathrm{h}^{-1}$.

To solve the control problem, we define the value function, V, by

$$V(x, y, z, s) = \max_{\phi_t} \mathbb{E}\left[\int_s^T \alpha\phi_t^2 \, \mathrm{d}t + \log_e(n_T)\right], \tag{3.16a}$$

$$\text{where} \quad (n_s, \theta_s, \zeta_s) = (x, y, z). \tag{3.16b}$$

HJB optimal control theory describes V with a PDE, expressed in this case as

$$0 = \max_{\phi_t}\left(\alpha\phi_t^2 + \frac{\partial V}{\partial y}(1 - y)\phi_t + \mathscr{H}_\star\right), \tag{3.17}$$

where $\mathcal{H}_{\star}$ represents a collection of terms independent of the control, ϕ_t. As equation (3.17) is quadratic in ϕ_t, we can carry out the maximisation by setting the derivative to zero to find the optimal control, denoted $\phi_t^* \in [0, 0.1]$, as

$$\phi_t^* = \max \left[0, \min \left(0.1, -\frac{1}{2\alpha} \frac{\partial V}{\partial y}(1-y) \right) \right]. \tag{3.18}$$

Here, $y = \theta_t$ is the current persister proportion. Since $y \in [0, 1]$, an interpretation of equation (3.18) is that the population uses ϕ_t^* to steer the population toward the optimal persister proportion, where $\partial V/\partial y = 0$. As environmental triggers only create persisters (we assume the quiescent state of persisters means they are unable to react to the environment), $\phi_t^* > 0$, and so ϕ_t^* is only active if the current proportion is less than the optimal proportion. Finally, $\phi_t^* \to 0$ as $\alpha \to -\infty$, which represents the sensing mechanisms becoming prohibitively expensive.

Substituting equation (3.18) into the HJB equation leads to a non-linear PDE for V that must be solved backward in time. Full details of this equation are given in Appendix B.4. In summary, the variable transformation $x \to \log_e(x)$ removes all terms containing the independent variable x (representing the current population size) from the equation. Therefore, we find that the ansatz
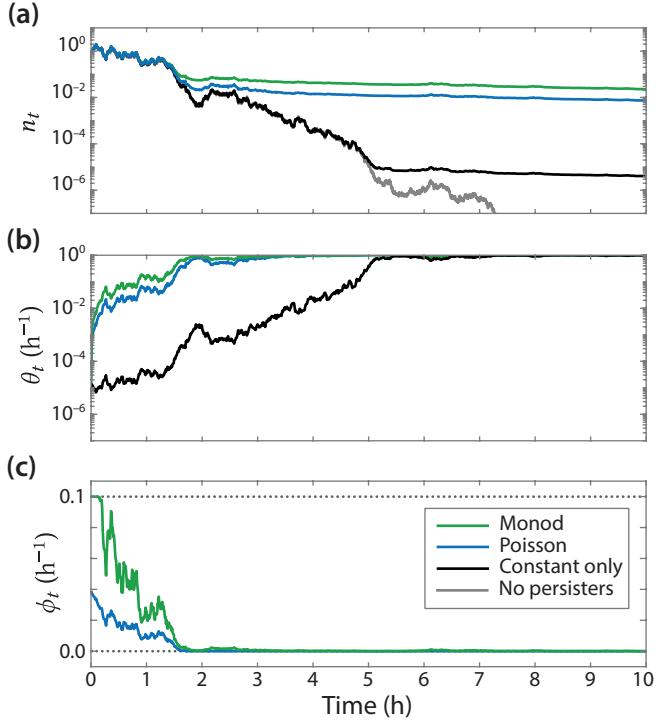
$$V(x, y, z, s) = \Psi(y, z, s) + \log_e(x), \tag{3.19}$$

is consistent with the system. This result reveals that the optimal control (equation (3.18)) is independent of the population size since $\partial V/\partial y = \partial \Psi/\partial y$. This gives an optimal strategy of the form $\phi_t^* = \phi^*(t, \theta_t, \zeta_t)$. These results suggest that bacteria can implement an optimal variable persister strategy using only information about the proportion of persisters (through, for example, quorum-sensing [133] or intercellular signalling [134]), and the environment (through, for example, a growth feedback mechanism [21]). Therefore, the persister strategy is dependent upon stochastic fluctuations in both the proportion of persisters and the environment.

In figure 3.2 we show dynamics for cell colonies in all three environments for $\alpha = -100\,\mathrm{h}$. For these results, we solve for Ψ numerically. As expected, we find that unfavorable environments trigger variable persister production (figure 3.2e,f), without any additional persister production for the constant environment (figure 3.2d). This suggests that a constant persister production strategy is sufficient for an environment with a constant expected growth rate. On the other hand, additional persister production is seen in response to environmental cues for more volatile environments (figure 3.2e,f). An interesting result in figure 3.2e is that the peak variable production rate under the Monod environment lies before the minimum growth rate. This suggests that the model implicitly incorporates mechanisms that allow cells to respond to features of the stochastic environment, to the extent that they may be able to anticipate environmental changes based upon their own current growth rate.

Behaviour under unfamiliar environments and antimicrobial treatment

By assuming that cells monitor their growth rate, μ_t, we can explore how bacteria that are specialised to one environment (i.e. implement the optimal persister strategy) behave under an unfamiliar environment. We couple the growth rate to the optimal control by considering $\phi_t^* = \phi^*(\theta_t, \zeta_t)$ where $\zeta_t = m^{-1}(\cdot)$ denotes the inverse of the growth rate coupling function $m(\zeta)$. In other words, the cells measure the current environment

◄ Figure 3.3. Single realisations of the model showing the behaviour of a cell population under antibiotic treatment. (a) Shows the population, n_t, in multiples of the original population; (b) shows the persister proportion, θ_t; and, (c) shows the variable persister production rate, ϕ_t^*. In all cases, the cells experience a growth rate $\mu_t = -2\,\mathrm{h}^{-1}$ and implement: the variable strategy optimal under either the Monod environment (green) or the Poisson environment (red); constant switching only optimal under the constant environment (black); or do not produce persisters (grey).

state using the growth rate.

Persisters are revealed experimentally by exposing the population to antibiotics and monitoring the population size [16]. We simulate this process to examine how bacteria specialised to each environment behave when exposed to a constant expected decay rate. Simulation results in figure 3.3a, for a population that can only produce persisters at a constant rate, are similar to experimental results for wild-type *E. coli* [16]. In comparison, we find that colonies that implement a strategy optimal under a more volatile environment, such as the Poisson and Monod environments, produce persisters using the variable rate (figure 3.3c) reach saturation of persisters more quickly (figure 3.3b) and have a higher long term population (figure 3.3a). These qualitative observations suggest that high persistence bacteria strains [16, 22] could arise in response to highly volatile environments. In Appendix B.2, we apply the same methodology to simulate bacteria specialised to each environment under other, unfamiliar, environments. These results demonstrate the potential of our work to model, for example, how an optimal persister strategy interacts with therapeutic interventions such as drug sequencing [145, 146].

3.3 Discussion

We describe persister dynamics with a stochastic model and introduce the idea of *cellular hedging* to study an optimal persister production strategy. By applying cross-disciplinary ideas from mathematical finance to the persister problem, we provide several analytical and simulation results that elucidate bacterial persistence under several novel models of environmental volatility.

Our results conform to the consensus view that it is detrimental to produce persisters in the absence of volatility [22, 104, 117, 147]. In examining the optimal strategy for constant persister production, we reveal the proportion of persisters that must be maintained to maximise the expected growth of the population (equation (B.8)). Removing volatility from the model by setting $\sigma = 0$ recovers the deterministic model of

Balaban *et al.* [16] and reveals an optimal persister proportion of zero. Therefore, our model predicts that producing persisters in the absence of environmental volatility is suboptimal and, therefore, leads to a lower expected per-capita growth rate. In contrast, the straightforward assumption of a growth rate driven by Wiener noise reveals a small proportion of persisters (approximately 1 in 10^5) — consistent with experimental observations of Balaban *et al.* [16] — that maximises the expected per-capita growth rate. Our results contrast with previous mathematical studies that assume external factors, such as antibiotic treatment, to explain persistence [22].

Compared with hedging strategies in finance, that often rely on complex, market-level information, a cellular hedging strategy must be limited in complexity. A key result of the direct comparison between bacterial persistence and MPP is to reveal a constant optimal proportion of persisters that should be maintained in an environment with a constant expected growth rate. This result is significant for the persister problem as, unlike in MPP, the cell population cannot directly control the proportion of persisters. This key difference between the biological and financial hedging problems is widely understood in the literature [117]. However, we show how a population can maintain a constant *expected* proportion of persisters, regulated by constant switching propensities, u and v.

The mechanisms that enable cells to implement environment-dependent cellular hedging strategies are unknown [148], but must come at a cost to the cells—in terms of additional genetic machinery—and rely on the limited information available to individual bacteria. Our study reveals an optimal variable persister production strategy that depends only on the time, persister proportion and the environment, but not on the the population size. Additional results in Appendix B.6 (figure B.6) indicate that these strategies become time-independent far from the terminal time, T. If an infinite terminal time problem was studied, necessarily with a growth model that incorporated crowding effects (such as logistic growth), we expect the optimal strategy to be independent of time. Furthermore, the optimal environment-dependent strategy revealed in equation (3.18) is not complex. Analysis of equation (3.18) suggests that bacteria increase persister production when the persister proportion for a given growth rate is less than optimal. This kind of information could be genetically encoded as an evolutionary adaptation to a particular environment. The parameter α, therefore, specifies the rate at which the population is able to respond to changes in the environment: if variable switching is expensive ($\alpha \gg 1$), the population will respond more slowly than if variable switching is relatively cheap. These results are surprising as optimal control theory only provides the strategy that maximises the payoff, and does not necessarily enforce any level of complexity in the cellular hedging strategy.

In figure 3.3, we simulate a persister producing bacteria colony under antibiotic treatment, demonstrating the distinct advantage persisters afford bacteria [16,109,112]. Our stochastic model of persister production can be used to improve the efficacy of antimicrobial therapies using optimal control from an optimal treatment perspective [4,149–151]. Furthermore, recent experimental and mathematical work examines the potential of so-called *evolutionarily informed therapy* or *drug sequencing* [145, 146, 152], where a sequence of drugs is administered to sequentially induce susceptibility and overcome drug resistance. As demonstrated in figure 3.3 and figure B.2, our model allows simulation of how a population specialised to one environment could behave under another. Therefore, our framework may naturally extend to predict how a persister strategy optimal under one type of drug behaves under another. Our results already suggest new ways to design treatment strategies. Additional results (figure B.6) suggest that, for all environments we model,

persister production decreases monotonically with the growth rate. Temporarily exposing the cell colony to favorable conditions will decrease variable persister production, lowering the proportion of persisters, potentially making the population more susceptible to antibiotics. This is consistent with existing experimental observations where exposing a colony to a fresh growth medium before applying antibiotics decreases persistence [153], and is a simplistic example of drug sequencing.

We focus on a large, exponentially growing population in a homogeneous, temporally fluctuating, environment. Our framework is readily extensible to more complex growth models (such as logistic growth); heterogeneous populations containing more than two phenotypes; and the inclusion of demographic noise. We have not, however, considered spatial effects such as those corresponding to a spatially fluctuating environment or local crowding effects. A stochastic partial differential equation (SPDE) model would also allow for direct modelling of the chemical signals that regulate quorum sensing [154] allowing for population-level cooperation, however applying optimal control to these SPDE models is not straightforward. Further work is needed to elucidate the effect of demographic or sensor noise on a population's ability to regulate the optimal persister proportion through subpopulation switching [117, 155]. In large exponentially growing populations, we expect deterministic switching rates are an appropriate model of subpopulation switching. For small populations, our optimal control approach can be applied to study the effects of demographic noise by incorporating intrinsic noise into the stochastic growth model through the chemical Langevin equation.

Rapid evolution of cellular hedging strategies is experimentally reproducible [156]: Van den Bergh *et al.* [109] found that exposing *E. coli* to daily antibiotic treatments increased survival by between three and 300-fold after just three treatments, and Rodriguez-Beltran *et al.* [157] study induced evolutionary changes in bacteria behaviour through repeated application of antibiotics. Evolutionary adaption of organisms such as bacteria within a *fitness landscape* is an active area of research [110], with an understanding that stochasticity plays a vital role [146]. Financial mathematics techniques and cellular hedging can also be applied to study the evolutionary process as it occurs. For example, policy adjustment models [158] quantify the cost of strategy change, and can be compared to mutation costs in a rapidly-evolving cell colony. Our framework can be directly applied in this context to model the adaption of a simple hedging strategy to an unfamiliar environment by modelling the time-derivative of the constant switching rates as a control.

Our analysis suggests important experimental avenues to further elucidate bacterial persistence. Alternating periods of growth and stress is still a common model of environmental volatility both [157] and in theoretical studies [104]. We provide a new theoretical foundation for studying any Markovian model of environmental volatility. Additional results in Appendix B.2 (figure B.2) show a diversity in optimal responses when a strategy optimal under one type of environment reacts to another. An experimental study where bacteria are repeatedly exposed to a known volatile environment, potentially based upon the novel models of environmental volatility that we study, will provide insight into how bacteria adapt to a form of uncertainty that can be quantified. Our modelling framework can then predict both the emergent strategy, and how the population might behave when exposed to interventions such as antibiotics. Furthermore, sensitivity analysis on parameters in the environmental volatility model can aid experimental design by revealing what features of an environment have the largest effect on any optimal strategy. We expect, for instance, different responses to environments where changes happen continuously (for example, the Monod environment), compared to where changes are due to shocks (for example, the Poisson environment).
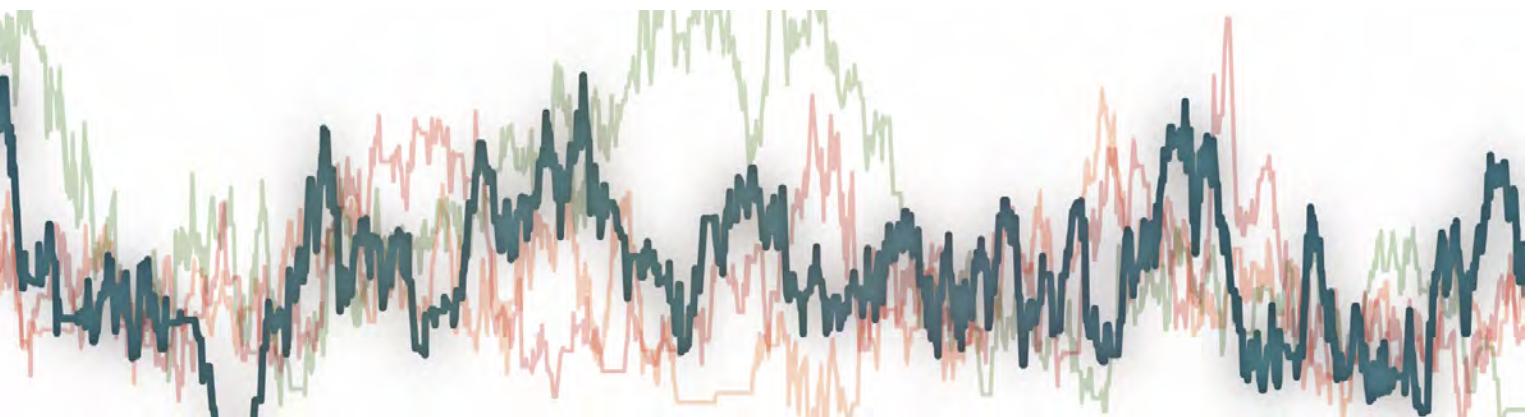
3.4 Conclusion

We provide new insight into cellular hedging through an explicit connection between the study of biological population dynamics and the mathematics of financial risk management. We present a new stochastic model of bacteria growth in a volatile environment and apply optimal control theory to probe the persister strategy that maximises the per-capita growth rate. A fundamental result of our study is to provide a solid theoretical understanding of why persistence is only advantageous in the presence of environmental volatility, demonstrating that the study of bacterial persistence must take a stochastic perspective. Our model of cellular hedging has clinical significance and can be applied to improve the efficacy of antimicrobial therapies. Furthermore, the framework we develop can be applied to more complex models of bacterial population dynamics, and offers an opportunity for future generalisations to explore cellular decision making in a broader context, including in cases where sensor or demographic noise are significant. Many seemingly complex cellular phenomena from bet-hedging in cancers [102, 159, 160], herpes viruses [161] and HIV [162] to decision making in the epithelial-mesenchymal transition [163] can be modelled using ideas from cellular hedging and by furthering the unification of mathematical finance and biology.

Chapter 4

Identifiability analysis for stochastic differential equation models in systems biology

Mathematical models are routinely calibrated to experimental data, with goals ranging from building predictive models to quantifying parameters that cannot be measured. Whether or not reliable parameter estimates are obtainable from the available data can easily be overlooked. Such issues of *parameter identifiability* have important ramifications for both the predictive power of a model, and the mechanistic insight that can be obtained. Identifiability analysis is well-established for deterministic, ordinary differential equation (ODE) models, but there are no commonly-adopted methods for analysing identifiability in stochastic models. We provide an accessible introduction to identifiability analysis and demonstrate how existing ideas for analysis of ODE models can be applied to stochastic differential equation (SDE) models through four practical case studies. To assess *structural identifiability*, we study ODEs that describe the statistical moments of the stochastic process using open-source software tools. Using practically-motivated synthetic data and Markov-chain Monte Carlo (MCMC) methods, we assess parameter identifiability in the context of available data. Our analysis shows that SDE models can often extract more information about parameters than deterministic descriptions. All code used to perform the analysis is available on Github.

Statement of Contribution of Co-Authors for
Thesis by Published Paper

The authors listed below have certified that:

1. they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
2. they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
3. there are no other authors of the publication according to these criteria;
4. potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit, and
5. they agree to the use of the publication in the student's thesis and its publication on the QUT's ePrints site consistent with any limitations set by publisher requirements.

In the case of this chapter:

Identifiability analysis for stochastic differential equation models in systems biology
Published 2020 in Journal of the Royal Society Interface
doi:10.1098/rsif.2020.0652

Contributor	Statement of Contribution
Alexander P Browning 4 Aug 2021	Designed the study, implemented the statistical computational algorithms, derived the moment equations, generated and interpreted results, drafted the manuscript, and revised the manuscript during peer-review.
David J Warne	Aided in the implementation of the computational algorithms, interpreted the results and provided comments on the manuscript.
Kevin Burrage	Designed the study, supervised the research, interpreted the results and provided comments on the manuscript.
Ruth E Baker	Designed the study, supervised the research, interpreted the results and provided comments on the manuscript.
Matthew J Simpson	Designed the study, supervised the research, interpreted the results, and provided comments on the manuscript.

| **Principal Supervisor Confirmation** |
| --- |
| I have sighted email or other correspondence from all Co-authors confirming their certifying authorship. |

Matthew J Simpson
Name

Signature

9 August 2021
Date

4.1 Introduction

Stochastic mathematical models are rapidly becoming an essential tool for interpreting biological phenomena [164–170]. These models are necessitated, in part, by increasing experimental interest in capturing finer-scale time-series observations [171–175] as well as spatial information [30, 176–180] rather than coarse-scale deterministic trends (figure 4.1). As computational inference techniques for stochastic models have improved [70, 181–184], a fundamental question that often remains overlooked is whether or not model parameters can be confidently estimated from the available data. Drug development, for example, often relies on the quantification of cell growth rates from a *proliferation assay* (figure 4.1a–d) [185]. If a mean-field model is applied to interpret the most frequently reported observation—cell count data—only the *net* growth rate is identifiable, not the proliferation and death rates [186, 187]. Establishing the *identifiability* of model parameters is critical as predictions, and parameter estimates, from a non-identifiable model may be unreliable [24, 188–190], with further analysis required to quantify prediction uncertainty in non-identifiable models [191–193]. Identifiability should always, therefore, be established before parameter estimation is attempted. Such identifiability analysis is well-established for deterministic ordinary differential equation (ODE) models [24, 194–201], but there is a scarcity of methods available for the stochastic models that are becoming increasingly important.

Stochasticity is fundamental to many processes [53, 165, 202–207]. Diabetic patients, for example, rely on the rapid interpretation of highly volatile blood glucose measurements to determine insulin input (figure 4.1f) [208, 209]. Data from the COVID-19 pandemic [164] is also volatile (figure 4.1e), and inferences of epidemic data must often be drawn from a single, stochastic, time-series. Finally, for systems at equilibrium in the mean-field, such as ion-channel data, models that account for system noise are required to establish parameters [210, 211]. Stochastic differential equation (SDE) models of the Itô form are widely applied in systems biology to describe stochastic phenomena [119, 212–214]. SDE models can describe intrinsic noise in, for example, gene expression [165, 172, 184] or a bio-chemical reaction network [215]; extrinsic noise describing volatility in the environment [53, 119, 126, 216]; and model approximations and unknown effects in so-called *grey-box* models [217, 218]. Explicitly modelling this variability in biological systems can often capture more information about a process than a deterministic model is able to [219–222]. Further, SDE models can account for the correlations inherent to time-series data and account for noise that might otherwise obscure parameters. We demonstrate how to establish parameter identifiability for SDE models that encode information about the intrinsic noise of the process [219]. Our focus is on SDE state-space models that can be formulated through the chemical Langevin equation (CLE), although our analysis is applicable to any SDE of the Itô form. While simulation [72], inference [223, 224], and identifiability analysis [225] can, for many stochastic systems, be conducted for discrete Markov models, SDE approximations can offer a significant computational advantage and have a long and extensive history of use in the systems biology literature. Further, use of reflected SDEs [226] can guarantee good agreement with their discrete counterparts at boundaries [226, 227].

A prerequisite for parameter estimation is that model parameters be *structurally identifiable* [2, 24, 194–196, 230, 231]. Structural identifiability refers to the question of whether a parameter can be identified given an infinite amount of noise-free data. A state-space model is said to be structurally identifiable if distinct values of the parameters imply distinct observed model outputs (or in the case of a stochastic model, distinct observed

▲ Figure 4.1. (*a–d*) Cell proliferation and death observed *in vitro* over 36 hours in a proliferation assay [34]. Each snapshot has a field-of-view of 1440 × 1440 μm and the location of each cell is indicated with a yellow marker. (*e*) Data from the early stages of the coronavirus pandemic comprising the observed number of (*i*) infected individuals, deaths, and (*ii*) daily new case count in Australia [228]. (*f*) Continuous glucose monitoring data from a single individual over three consecutive days [229].
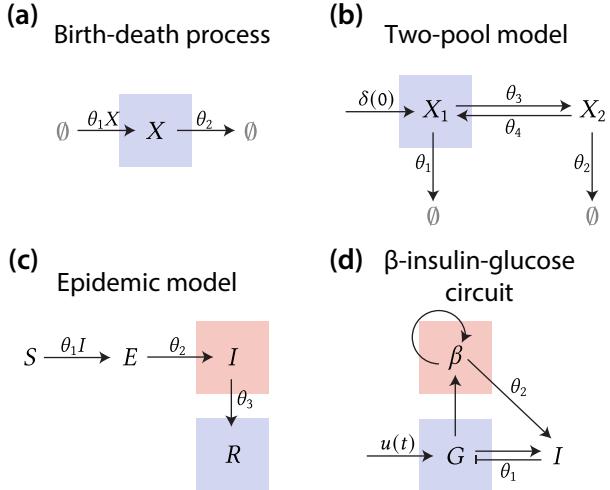
output distributions [232]), and vice versa [233–235]. Techniques such as differential algebra [198, 236, 237] and transfer function approaches [194, 195] can establish structural identifiability in ODE models. These approaches are also used to establish identifiable relationships between parameters [195, 238]—for example, the net growth rate in a proliferation assay—which can aid model design and model reduction [238–240]. Many of these techniques have accessible implementations in symbolic computation packages [198, 241–244], meaning structural identifiability analysis does not require a detailed understanding of the, often complex, underlying mathematical analysis [198].

When experimental data is considered, a more useful question is that of *practical identifiability* or *estimability* [24, 235, 241, 245]. That is, can parameters in the model be accurately estimated given a finite amount of noisy experimental data? This kind of analysis is routinely used in the field of experimental design to assess the nature of data required to adequately identify biophysical parameters [27, 90, 189, 210, 213, 246]. Practical identifiability is established in conjunction with an inference technique, such as profile or maximum likelihood [27, 247–249] or Markov-chain Monte-Carlo (MCMC) [189, 210]. These techniques provide information about the flatness (or otherwise) of the likelihood function—in the Bayesian case, the posterior distribution—that describes knowledge about the parameters after the experimental data is taken into consideration. For deterministic and simple stochastic models, this information can be obtained directly from the Fisher information matrix [249, 250]. Compared with structural identifiability, which is a property of the model, practical identifiability is more nuanced and additionally dependent upon prior knowledge; the experimental data; and consequentially, the experiment itself [189, 241]. For example, should the model and data provide no more information about a parameter than that already established in

previous studies, the parameter may be classified as practically non-identifiable from the data and model in question. For this reason, we take a Bayesian approach to parameter estimation and encode existing knowledge about the parameters in a *prior distribution*. Therefore, we classify a model parameter as practically non-identifiable if it cannot be uniquely established within a level of confidence not already established from prior knowledge [24, 189]. This question of practical identifiability has not yet been demonstrated for SDE models in systems biology.

Computational inference for stochastic models is a significant challenge [183]. Unlike approaches to parameter estimation for deterministic models, the likelihood function for a realistic stochastic model is, generally, intractable [183]. Techniques based on approximations, such as a linear noise approximation [249] or approximate Bayesian computation [70, 251–255], are available for SDEs but are, naturally, approximations. Pseudo-marginal methods [256, 257], developed relatively recently, are computationally costly, but provide an unbiased estimate of the true likelihood function for partially observed time-series described by non-linear stochastic models. In this chapter, we utilise a pseudo-marginal MCMC approach, where we estimate the likelihood with a particle filter, which we refer to as particle MCMC [258–260]. There are many excellent articles and reviews of inference for stochastic models in systems biology [70, 183, 260, 261], so we do not focus on the details our out implementation here. Despite the established importance of identifiability, it is all too common in parts of the inference literature to draw the standard assumption that the model parameters are identifiable: we note that all the aforementioned review articles make no mention of identifiability. The computational cost of inference for stochastic models, in itself, motivates us to consider identifiability. For example, identifiability can guide model selection: if both a deterministic and stochastic description of a process are practically non-identifiable, the cheaper deterministic model may, in some cases, be adequate for parameter estimation. Where structural non-identifiability is detected, practical non-identifiability necessarily follows and does not need to be established separately.

The focus of this chapter is to provide an accessible guide to establishing identifiability in SDE models in biology. To do this, we analyse identifiability in SDE descriptions of four case study models, shown in figure 4.2. The simplest model we consider is a birth-death process (figure 4.2a) that is routinely used to describe cell proliferation and death in a range of *in vitro* and *in vivo* biological systems, such as that shown in figure 4.1a. We demonstrate that, from cell count-data, the cell proliferation and death rates are structurally non-identifiable for a routinely employed ODE model, but can be identified for an SDE model. Next, we consider two multi-state models where only partial observations of the system are available. First, a two-pool model (figure 4.2b) that can describe, for example, the decay of human cholesterol whilst it transfers between two organs [195, 262]. We assume that data from the two-pool model comprises several time-series observations of the substance concentration in a single pool. Second, an epidemic model (figure 4.2c) [263–265] describes individuals infected due to interactions between susceptible and infectious individuals. We model a testing procedure such that unknown proportions of the number of infectious and recovered individuals are observed, and inferences are drawn from a single time-series. The last model we consider is a non-linear SDE model for insulin regulation by β-cells (figure 4.2d) [266, 267]. This type of model can describe the volatility associated with data from a continuous glucose monitoring device (figure 4.1f) [229]. The equivalent ODE description of the β-insulin-glucose circuit is not structurally or practically identifiable [268], and we demonstrate how the analysis for the ODE description can inform a parameter transformation to aid identifiability analysis for the SDE model.

(a) Birth-death process

(b) Two-pool model

◄ Figure 4.2. We demonstrate identifiability in an SDE CLE description of four models: (*a*) a birth-death process; (*b*) a two-pool model; (*c*) an epidemic model; and (*d*) a β-insulin-glucose circuit. The coloured boxes indicate the observed quantity, which is coupled to a noisy observation process.

(c) Epidemic model

(d) β-insulin-glucose circuit

We demonstrate two main approaches to assess identifiability in SDE models. First, we assess structural identifiability through a surrogate model, taken to be a system of ODEs that describe the time-evolution of the statistical moments of the SDE [269–273]. This allows us to apply the established open-source structural identifiability software package DAISY (written for the freeware REDUCE software) to the SDE models through the moment equations. We repeat this analysis in the more recent open-source software package GenSSI2 [25, 274], written for MATLAB, which can be more efficient for non-linear systems. We interpret these results as a proxy for identifiability of the SDE model itself. While this approach is not always conclusive, it can provide a rapid preliminary screening tool and allows direct comparison of identifiability for an SDE model, which contains information about the mean, variance and higher moments; to identifiability for a corresponding ODE model that is typically assumed to describe an approximation of the mean. We only apply this approach where an exact system of moment equations can be derived, which occurs when the reaction rates are polynomial. For more complex stochastic models containing terms such as Hill functions, as found in the β-insulin-glucose circuit model, an exact system of moment equations cannot be derived, we do not apply the moment dynamics approach in this case. Second, we assess practical identifiability using the full SDE models through MCMC [189, 210], first demonstrating how practical identifiability can be cheaply established from a naïve proposal kernel. To compute credible intervals for each parameter, and visualise potential correlations between parameters, we produce results using a tuned proposal kernel where we can be more certain of convergence.

The outline of this chapter is as follows. In Section 4.2, we establish the types of SDE models and observation processes that we consider, and then outline the techniques used to generate synthetic data. Following this, in Section 4.2.2, we summarise moment closure techniques for SDEs and describe how we implement the software tools DAISY and GenSSI to assess for structural identifiability. Next, in Section 4.2.3, we provide a brief overview of our implementation of the particle MCMC algorithm. Full details of particle MCMC for SDE models can be found in the existing literature [260, 261] and as supporting material. In Section 4.3, we use these tools to assess identifiability using an SDE description of four models. In Sections 4.4 and 4.5, we discuss our results and provide an outlook on the future of identifiability for stochastic models in biology. Specifically, we discuss alternative approaches, including those based upon approximations, to perform practical identifiability analysis in light of the computational challenges of working with SDE models. To aid in the accessibility of the techniques we review, we provide our MCMC code in the form of a module[1] for the

open-source, high-performance `Julia` programming language [275].

4.2 Mathematical techniques

In this section, we outline the mathematical and statistical techniques we use to perform identifiability analysis. Full details of all algorithms used are provided as supporting material.

4.2.1 Stochastic models in biology

We consider Itô SDE state space models of the form

$$dX_t = \boldsymbol{\alpha}(X_t, t; \boldsymbol{\theta}) \, dt + \boldsymbol{\sigma}(X_t, t; \boldsymbol{\theta}) \, dW_t, \tag{4.1}$$

$$Y_t \sim g(Y_t | X_t, t; \boldsymbol{\theta}). \tag{4.2}$$

Here, the state is described by $X_t = (X_{1,t}, X_{2,t}, \ldots, X_{N,t}) \in \mathbb{R}^N$, $W_t \in \mathbb{R}^Q$ is a Q-dimensional Wiener process with independent components; $\boldsymbol{\alpha}(\cdot)$ maps to an N-dimensional vector; and $\boldsymbol{\sigma}(\cdot)$ maps to an $N \times Q$ matrix. The observables, $Y_t = (Y_{1,t}, Y_{2,t}, \ldots, Y_{M,t}) \in \mathbb{R}^M$, are connected to the state variables according to an observation process with probability density function $g(Y_t | X_t, t; \boldsymbol{\theta})$. We consider several forms of observation function, including partial observations of the state with both additive and multiplicative Gaussian noise with unknown variance σ_{err}^2. In equations (C.2) and (4.2), $\boldsymbol{\theta}$ is a vector of unknown parameters to be determined through inference. In this chapter, all variables and parameters are dimensionless.

The focus of this chapter is on Itô SDE models that are formulated through the CLE description of a system of bio-chemical reactions [215, 276, 277]. Therefore, additional information about rate parameters is encoded in the noise of the process. The first three models we consider (figure 4.2a–c) can be expressed directly as a network of reactions. As the β-insulin-glucose circuit model (figure 4.2d) involves state variables modelled as concentrations, not individual counts, we derive a stochastic description from the CLE but scale the noise term in proportion to the concentration of each species.
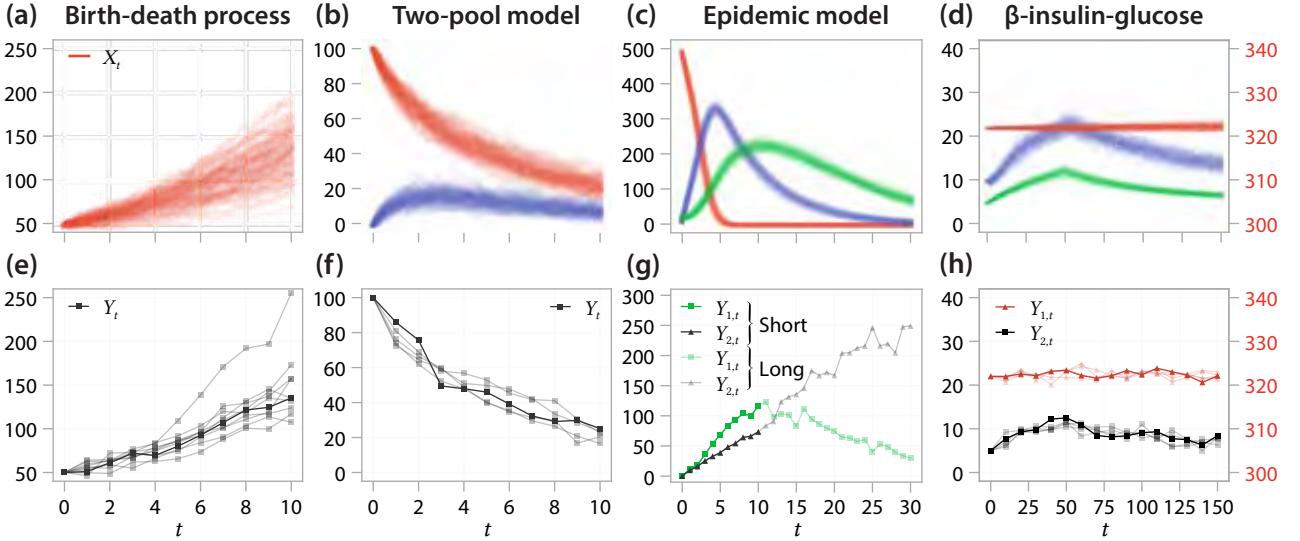
In summary, a bio-chemical reaction network comprises N species, $X_1, X_2, \ldots, X_N$, that interact through Q reactions [28, 72, 278]. The population of each species is given by $X_t = (X_{1,t}, X_{2,t}, \ldots, X_{N,t}) \in \mathbb{R}^N$. By the law of mass action [215, 279], each reaction occurs with a rate described by a *propensity function*, $a_k(X_t, t; \boldsymbol{\theta})$, which is equal to the product of the reactants and the rate constant. The net effect of the kth reaction is described by the stoichiometry $\boldsymbol{\nu}_k$ such that, should reaction k occur in $[t, t + dt)$,

$$X_{t+dt} = X_t + \boldsymbol{\nu}_k. \tag{4.3}$$

For bio-chemical reaction networks without an explicit time-dependent input, the propensity functions will be independent of t and the system can be simulated exactly using an event-driven stochastic simulation algorithm (SSA) [168, 279–281]. The principle behind an exact SSA is that reactions can be modelled by an inhomogeneous Poisson process. The time interval between reactions, Δt, is exponentially distributed such that

$$\Delta t \sim \text{Exp}\left(\sum_{k=1}^{Q} a_k(X_t; \boldsymbol{\theta})\right). \tag{4.4}$$

---

[1]Code available on Github at `https://github.com/ap-browning/SDE-Identifiability`

▲ **Figure 4.3.** (*a–d*) 100 example realisations of each model, produced using: (*a–c*) the SSA; and, (*d*) the SDE. (*e–h*) Synthetic data used for practical identifiability analysis. Synthetic data comprises noisy observations of the (*e*) full and (*f–h*) partial state. In (*e*,*f*,*h*), experimental replicates used simultaneously for parameterisation are shown semi-transparent, with the first replicate fully opaque. For the epidemic model, both short-time (opaque) and long-time (semi-transparent) data are considered separately. In both cases of the epidemic model, an unknown proportion of the number of infected individuals (green), and the recovered individuals (black), is observed. In (*d*,*h*), the β cell concentration, β_t (and the measured concentration $Y_{1,t}$) is shown on the right axis.

A single reaction occurs at each time-step; the kth reaction occurs with probability proportional to $a_k(X_t; \theta)$. A typical implementation of the SSA first samples a time-step using equation (4.4); then samples the next event to occur; and finally updates the state. Full details of our implementation of an SSA are given as supporting material, and the reader is directed to [278] for a comprehensive review of simulation algorithms for bio-chemical reaction networks. We generate synthetic data for the first three models, for which the propensity functions are independent of t, using the SSA. In figure 4.3a–c we show 100 realisations of the SSA for the birth-death process, two-pool model and epidemic model, respectively.

When the population of each species is large and reactions sufficiently frequent, the dynamics of a bio-chemical reaction network can be approximated using the CLE [186,276,278]. Such an approximation is widely applied in systems biology [282, 283], and it is often necessary as the SSA quickly becomes computationally expensive as the populations become large and reactions are frequent enough [284]. The CLE is an Itô SDE of the form

$$\mathrm{d}X_t = \underbrace{\sum_{k=1}^{Q} \boldsymbol{\nu}_k a_k(X_t, t; \boldsymbol{\theta})\,\mathrm{d}t}_{\boldsymbol{\alpha}(X_t,t;\boldsymbol{\theta})} + \underbrace{\sum_{k=1}^{Q} \boldsymbol{\nu}_k \sqrt{a_k(X_t, t; \boldsymbol{\theta})}\mathrm{d}W_{k,t}}_{\boldsymbol{\sigma}(X_t,t;\boldsymbol{\theta})\mathrm{d}W_t}. \tag{4.5}$$

Here $W_t = (W_{1,t}, W_{2,t}, \ldots, W_{Q,t})$ is a Q-dimensional Wiener process with independent components. In this study, we derive an SDE description for each model using the CLE, and we calibrate this SDE to the synthetic data to approximate the parameters in each model. For the first three models, where data is generated using the SSA, not the SDE, this means that identifiability analysis is conducted in such a way that model misspecification could potentially arise. This pragmatically mirrors experimental data, where any model (including an ODE and SDE description) is an approximation. The forward simulation for each SDE is approximated using the Euler-Maruyama algorithm [285], where we apply reflected SDEs to ensure

positivity [226]. Full details of the numerical algorithm are given as supporting material.

4.2.2 Moment dynamics

To enable the application of established methods for structural identifiability analysis to SDE models, we formulate a system of ODEs that describe the statistical moments of the random variable $X_t \in \mathbb{R}^N$. We denote $m_{i_1 i_2 \dots i_d}(t)$ as a raw moment of X_t, such that [270–272, 277]

$$m_{i_1 i_2 \dots i_N}(t) = \left\langle \prod_{j=1}^{N} X_{j,t}^{i_j} \right\rangle, \tag{4.6}$$

where $\langle \cdot \rangle$ indicates the expectation taken with respect to the probability measure of the random variable X_t. Here, $J = \sum_{j=1}^{N} i_j$ is the *order* of the moment. For example, the first order moments correspond to the mean of each dimension of X_t, the second order moments relate to the variances and covariances, and so forth.

We apply the software packages DAISY [198] and GenSSI2 [274] to establish structural identifiability of the resultant system of moment equations. The software package takes a system of ODEs describing the state equations—in our case, the moment equations—in addition to an explicit algebraic relationship between the observables and the state. We, therefore, provide the moments of the observables, Y_t, in the noise-free limit, which we denote

$$n_{i_1 i_2 \dots i_M}(t) = \lim_{\sigma_{\text{err}}^2 \to 0} \left\langle \prod_{j=1}^{M} Y_{j,t}^{i_j} \right\rangle. \tag{4.7}$$

In many cases, the observation distribution, $g(Y_t|X_t, t; \boldsymbol{\theta})$, will depend upon the unknown parameters, $\boldsymbol{\theta}$, if, for example, an unknown proportion of the state is observed. This is captured in the structural identifiability analysis as the equations derived for the observed moments, n, may depend on $\boldsymbol{\theta}$. We provide well commented input and output obtained using DAISY on Github as supporting material.

An expression for the time derivative of each moment can be found using Itô's lemma (supplementary material). When each component of $\boldsymbol{\sigma}^T \boldsymbol{\sigma}$ is an analytic function, which occurs when all the propensity functions in the bio-chemical reaction network are also analytic functions, we obtain [286]

$$\begin{aligned}
\frac{\mathrm{d}m_{i_1 i_2 \dots i_N}(t)}{\mathrm{d}t} = & \left\langle \boldsymbol{\alpha}(X_t, t; \boldsymbol{\theta}) \cdot \boldsymbol{\nabla} \left(\prod_{j=1}^{N} X_{j,t}^{i_j} \right) \right. \\
& \left. + \frac{1}{2} \mathrm{Tr} \left(\boldsymbol{\sigma}^T(X_t, t; \boldsymbol{\theta}) \mathrm{H} \left(\prod_{j=1}^{N} X_j^{i_j} \right) \boldsymbol{\sigma}(X_t, t; \boldsymbol{\theta}) \right) \right\rangle,
\end{aligned} \tag{4.8}$$

where $\mathrm{H}(\cdot)$ denotes the $N \times N$ Hessian matrix of its argument and $\boldsymbol{\nabla} = \left(\frac{\partial}{\partial X_1}, \frac{\partial}{\partial X_2}, \dots, \frac{\partial}{\partial X_N} \right)$. In the case that $N = 1$, equation (4.8) reduces to

$$\frac{\mathrm{d}m_i(t)}{\mathrm{d}t} = \left\langle i X_t^{i-1} \alpha(X_t, t; \boldsymbol{\theta}) + i(i-1) X_t^{i-2} \frac{\sigma^2(X_t, t; \boldsymbol{\theta})}{2} \right\rangle,$$

where α and σ are now scalar functions.

When each component of $\boldsymbol{\alpha}$ and $\boldsymbol{\sigma}^T \boldsymbol{\sigma}$ are polynomials in X_t, the expectation in equation (4.8) can be carried through to replace powers of X_t with appropriate moments. This, in general, provides an infinite

system ODEs that *exactly* describe the time evolution of the moments. In practice, we consider a finite system of moments, up to and including moments of order J. We express this now finite system of ODEs as

$$\frac{dm_{\leq J}(t)}{dt} = f_{\leq J}(m_{\leq J}(t), m_{>J}(t)),$$

(4.9)

where $m_{\leq J}(t)$ is a vector containing all the moments up to, and including, order J; and $m_{>J}(t)$ is a vector containing all moments of order $J + 1$ and above. In the case that $f_{\leq J}(\cdot)$ depends only on moments up to order J, the system is said to be *closed* at order J. That is, the infinite system of equations can be truncated at order J and solved directly to obtain an exact solution for the moments. This is the case if $\boldsymbol{\alpha}$ and $\boldsymbol{\sigma}^T \boldsymbol{\sigma}$ are linear in X_t, which occurs in SDEs derived from the CLE if each propensity is linear in X_t, as is the case for the first two models we consider (figure 4.2a,b).

For more complicated models, including the epidemic model (figure 4.2c), the system will not, in general, be closed. We must, therefore, apply a *moment closure* approximation to express moments of order higher than J in terms of lower order moments [205]. Moment closures typically make an *a priori* assumption about the distribution of the random variable X_t. For example, assuming components of X_t are independent or normally distributed is a common approach. In this chapter, we consider three common moment closures: (1) a mean-field closure [271]; (2) a pairwise closure [271]; and (3) a Gaussian closure [270].

The *mean-field closure* we consider makes the approximation

$$m_{i_1 i_2 \ldots i_N} \approx \prod_{j=1}^{N} \left\langle X_{j,t}^{i_j} \right\rangle.$$

(4.10)

This closure is derived from the assumption that components of X_t are weakly correlated [271] and is also referred to as the *covariance closure* [287]. In the case a closure is drawn at $J = 1$, the mean-field closure often corresponds to an ODE description of the process. For our analysis of the epidemic model, we find that the mean-field closure behaves poorly, suggesting that an assumption that the components of X_t are independent may not be appropriate (Appendix C, figure C.1).

While the mean-field closure is commonly drawn at order $J = 1$, it is more common for the *pair-approximation closure* to be applied for second and higher order closures [271]. The pair-approximation closure assumes that a third order moment can be expressed as

$$\left\langle X_{a,t} X_{b,t} X_{c,t} \right\rangle \approx \frac{\left\langle X_{a,t} X_{b,t} \right\rangle \left\langle X_{b,t} X_{c,t} \right\rangle}{\left\langle X_{b,t} \right\rangle}, \quad \left\langle X_{b,t} \right\rangle \neq 0.$$

(4.11)

The *Gaussian closure* approximates higher order moments to match those of the normal distribution, and gives a closure in terms of the mean and covariances. Higher order moments can be approximated with [270, 288]

$$\hat{m}_{i_1 i_2 \ldots i_N}(t) \approx \begin{cases} 0, & \text{if } J = \sum_{j=1}^{N} i_j \text{ is odd}, \\ \sum_s \prod_{(j,k) \in I_s} \text{Cov}(X_{j,t} X_{k,t}), & \text{otherwise}. \end{cases}$$

(4.12)

Here, $\hat{m}_{i_1 i_2 \ldots i_N}(t) = \left\langle \prod_{j=1}^{N} (X_{j,t} - \left\langle X_{j,t} \right\rangle)^{i_j} \right\rangle$ denotes a central moment; $\text{Cov}(X_{j,t} X_{k,t})$ denotes the

covariance between $X_{j,t}$ and $X_{k,t}$; and I_s are the sets formed by partitioning the set

$$\{\underbrace{1, 1, \ldots, 1}_{i_1}, \ldots, \underbrace{N, N, \ldots, N}_{i_N}\}$$

into unordered pairs, where s is the number of sets. The raw moments, $m_{i_1 i_2 \ldots i_N}(t)$ can then be solved from the expressions for the central moments obtained from equation (4.12). For a practical example of the Gaussian closure, see [270].

Other choices of moment closure are routinely used in systems biology, such as those based upon a multivariate lognormal distribution [270] or a derivative matching scheme [289]. However, more complex closures add further complexity to the moment equations, which is a significant computational disadvantage for automated assessment of structural identifiability in software packages such as DAISY and GenSSI2. Furthermore, an *approximate* system of moment equations (which must then also be closed) could be obtained by applying a series expansion approximation, or an approximation similar to the mean-field closure, to systems containing non-polynomial analytic functions; this is the case for the fourth model we consider (figure 4.2d). We do not consider the moment dynamics approach for such non-polynomial models in this chapter as there are difficulties in determining how approximations of different types of non-linearities impinge on identifiability. This is beyond the scope of this chapter.

4.2.3 Inference with MCMC

We take a Bayesian approach to parameter estimation to update our knowledge about the parameters, $\boldsymbol{\theta}$, from a set of observations, $\mathscr{D}$, using the likelihood function, $\mathscr{L}$, such that [290]

$$p(\boldsymbol{\theta}|\mathscr{D}) \propto \mathscr{L}(\mathscr{D}|\boldsymbol{\theta})p(\boldsymbol{\theta}). \tag{4.13}$$

Here, $p(\boldsymbol{\theta})$ is the *prior distribution*, and represents our knowledge of $\boldsymbol{\theta}$ before consideration of the observations $\mathscr{D}$. The prior distribution may encode information from, for example, previous experiments, established knowledge, or physical restrictions on the parameters. In the context of practical identifiability, our goal is to significantly increase our understanding of $\boldsymbol{\theta}$ from our prior knowledge. We specify $p(\boldsymbol{\theta})$ to be a truncated uniform distribution: all parameters within a specified region of realistic parameter values (the support) are considered equally likely [189]. An advantage of a uniform prior in the context of identifiability is that the posterior corresponds to the truncated likelihood function, and, therefore, high density regions of the posterior correspond to regions of high likelihood. Further, should an improper, unbounded uniform prior be considered, the posterior will be directly proportional to the likelihood. Thus, our methodology can also be applied to assess parameter identifiability using a purely likelihood-based approach.

We use an MCMC technique, based on the Metropolis-Hastings algorithm, to sample from the posterior distribution [290–293]. The principle behind MCMC in Bayesian inference is to construct a Markov chain, $\{\boldsymbol{\theta}_m\}_{m \geq 0}$, with a stationary distribution equal to $p(\boldsymbol{\theta}|\mathscr{D})$. We make a standard choice to initiate the chain from a prior sample, $\boldsymbol{\theta}_0 \sim p(\boldsymbol{\theta})$. At each iteration of the algorithm, a new state is proposed, $\boldsymbol{\theta}^* \sim q(\boldsymbol{\theta}^*|\boldsymbol{\theta}_m)$,

where q is termed the *proposal kernel*. The proposal is accepted, $\theta_{m+1} \leftarrow \theta^*$, with probability

$$\alpha_{\mathrm{MH}}(\theta^*, \theta_m) = \min\left(1, \frac{q(\theta_m|\theta^*)p(\theta^*)\mathscr{L}(\mathscr{D}|\theta^*)}{q(\theta^*|\theta_m)p(\theta_m)\mathscr{L}(\mathscr{D}|\theta_m)}\right), \tag{4.14}$$

else the proposal is rejected, $\theta_{m+1} \leftarrow \theta_m$. Full details of our implementation are provided in Appendix C. In this chapter, we use a multivariate normal proposal so that $q(\theta_m|\theta^*) = q(\theta^*|\theta_m)$. An interpretation of the Metropolis choice of acceptance probability, equation (4.14), where the proposal is normal and, therefore, symmetric, is that proposals that increase the posterior density are always accepted, whereas proposals that decrease the posterior density are accepted with some reduced probability [189].

We refer to the first set of MCMC chains for each problem as *pilot chains* [294]. The proposal distribution for each pilot chain is set to be a multivariate normal distribution with independent components and variances equal to one-tenth the corresponding prior variance for each parameter, a typical choice. We always produce four pilot chains, each of 10,000 iterations, which we find to be sufficient to indicate identifiability for our models. These pilot chains are then used to *tune* the MCMC proposal kernel [295]. We then produce four *tuned chains*, which can be reliably used to estimate credible intervals and other features of the posterior distribution. The proposal distribution for each tuned chain is chosen to be multivariate normal, with covariance given by [294]

$$\Sigma_{\mathrm{opt}} = \frac{2.38^2}{\dim(\theta)}\hat{\Sigma}. \tag{4.15}$$

Here, $\dim(\theta)$ is the number of unknown parameters, and $\hat{\Sigma}$ is the covariance matrix for the pooled samples from the four pilot chains (a total of 28,000 samples after 3,000 samples are discarded as burn-in from each pilot chain). To assess convergence, we calculate the commonly used $\hat{R}$ [296] and n_{eff} (effective sample size) [290] diagnostics. In summary, $\hat{R}$ measures the ratio of between-chain and within-chain variance; and n_{eff} measures the effective number of independent samples drawn from the posterior. To draw reliable inferences, Gelman *et al.* [290] suggest ensuring that $\hat{R} < 1.1$. Full details of these convergence statistics are available in [290].

The primary challenge with performing inference for SDE models, with time-series data, is computing the likelihood function. In this chapter, we consider synthetic data from E independent experiments, each with N_E time-series observations. The data are denoted

$$\mathscr{D} = \left\{\{t_{n,i}, \mathrm{Y}_{\mathrm{obs}}^{n,i}\}_{n=1}^{N_E}\right\}_{i=1}^{E}, \tag{4.16}$$

and correspond to the likelihood function

$$\mathscr{L}(\mathscr{D}|\theta) = \prod_{i=1}^{E}\prod_{n=1}^{N_E} p(\mathrm{Y}_{\mathrm{obs}}^{n,i}|\mathrm{Y}_{\mathrm{obs}}^{1,i}, \dots, \mathrm{Y}_{\mathrm{obs}}^{n-1,i}). \tag{4.17}$$

In most cases, the likelihood for noisy time-series data modelled by an SDE will be intractable [260]. This contrasts with data modelled by a deterministic model, which are typically assumed to be independent and normally distributed about the model output [189]. Likelihood free methods, such as ABC [70, 254] and pseudo-marginal approaches [257], are routinely used in systems biology to calibrate complex stochastic models to experimental data by approximating equation (4.17). In this study, we apply a pseudo-marginal

approach based on a bootstrap particle filter to approximate the likelihood and calibrate each SDE model to synthetic experimental data [260]. In summary, the bootstrap particle filter approximates equation (4.17) by

$$\hat{\mathscr{L}}(\mathscr{D}|\boldsymbol{\theta}) = \prod_{i=1}^{E} \prod_{n=1}^{N_E} \frac{1}{R} \sum_{r=1}^{R} g(Y_{\text{obs}}^{n,i}|X_{t_n}^{i,r}, t; \boldsymbol{\theta}). \qquad (4.18)$$

Here, the observation probability density, g (equation (4.2)), is averaged over R samples from the SDE, $X_{t_n}^{i,r}|X_{t_{n-1}}^{i,r}$ to approximate the likelihood. The bootstrap particle filter then resamples from the set of weighted samples, $\{(g(Y_{\text{obs}}^{n,i}|X_{t_n}^{i,r}), X_{t_n}^{i,r})\}_{r=1}^{R}$, at each time-step to form the starting locations for each SDE sample to sample forward to t_{n+1}. This process is repeated for each independent experiment, and the result is an unbiased Monte Carlo estimate of the likelihood function, $\hat{\mathscr{L}}(\mathscr{D}|\boldsymbol{\theta})$, that replaces $\mathscr{L}$ in the Metropolis acceptance probability (equation (4.14)). Full details of the particle MCMC algorithm, including an implementation for an ODE model used in one case study, are provided in Appendix C, and for further information the reader is directed to [260, 261].

4.3 Case studies
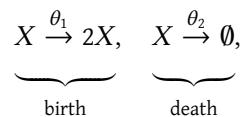
Using the moment equations and MCMC, we provide a practical guide for assessing parameter identifiability in SDE CLE models through four case studies. We generate synthetic data for each model using the SSA when the propensity functions are time-independent (the birth-death process, two-pool model and epidemic model), and the corresponding CLE when the propensity functions are time-dependent (the β-insulin-glucose circuit). In practice, we would first assess practical identifiability using the experimental data available. However, working with synthetic data provides the means to evaluate the effect of different experiment designs, and observation protocols, on practical identifiability. Our focus is on data comprising partial observations of the process that realistically captures potential experimental data.

4.3.1 Birth-death process

The first model we consider is a birth-death process (figure 4.2a). The birth-death processes can describe, for example, the growth of a well-mixed cell population where individuals proliferate and die according to rates θ_1 and θ_2, respectively. We consider practical identifiability for synthetic data comprising noisy measurements of the cell count at 10 equally spaced times in 10 identically prepared experiments. Such data are typical for *in vitro* cell proliferation experiments [185, 297], an example of which is shown in figure 4.1a–d.

Model formulation and moment equations

The birth-death process can be expressed as the bio-chemical reaction network

$$\underbrace{X \xrightarrow{\theta_1} 2X}_{\text{birth}}, \quad \underbrace{X \xrightarrow{\theta_2} \emptyset}_{\text{death}},$$

with stoichiometries $\nu_1 = 1$ and $\nu_2 = -1$; and propensities $a_1(X_t) = \theta_1 X_t$ and $a_2(X_t) = \theta_2 X_t$. Here, we denote X_t as the number of individuals in the population. The observed number of individuals, Y_t, is described

by the noise model

$$Y_t = \xi_t X_t, \quad \xi_t \sim \mathcal{N}(1, \sigma_{\mathrm{err}}^2). \tag{4.19}$$

Here, we consider a noise process that scales with the total population, that is, multiplicative Gaussian noise. We show 100 realisations of the SSA for the birth-death process in figure 4.3a, and the synthetic data used for practical identifiability analysis in figure 4.3e. The data are generated using the initial condition $X_0 = 50$ and target parameter values $\theta_1 = 0.2$, $\theta_2 = 0.1$ and $\sigma_{\mathrm{err}} = 0.05$. Here, $\sigma_{\mathrm{err}} \ll 1$, which ensures that Y_t remains positive.

The CLE for the birth-death process is

$$\mathrm{d}X_t = (\theta_1 - \theta_2)X_t\, \mathrm{d}t + \sqrt{(\theta_1 + \theta_2)X_t}\, \mathrm{d}W_t, \tag{4.20}$$

and the first and second order moment equations are

$$\begin{aligned}
\frac{\mathrm{d}m_1}{\mathrm{d}t} &= (\theta_1 - \theta_2)m_1, \\
\frac{\mathrm{d}m_2}{\mathrm{d}t} &= 2(\theta_1 - \theta_2)m_2 + (\theta_1 + \theta_2)m_1.
\end{aligned} \tag{4.21}$$

The moment equations for the SDE description of the birth-death model above are identical to the moment equations for the discrete Markov model that we simulate using the SSA [41]. The moments of the observable (in the noise-free limit) are given to second order by $n_1 = m_1$ and $n_2 = m_2$. As $\alpha(\cdot)$ and $\sigma^2(\cdot)$ are linear in X_t, the moment equations of the birth-death process are closed at every order and so equations (4.21) are exact. Further, we note that the common mean-field model for the birth-death process,

$$\frac{\mathrm{d}\widetilde{X}}{\mathrm{d}t} = (\theta_1 - \theta_2)\widetilde{X}, \tag{4.22}$$

corresponds to the first moment, and describes the average behaviour of X_t. The solution to equation (4.22) is
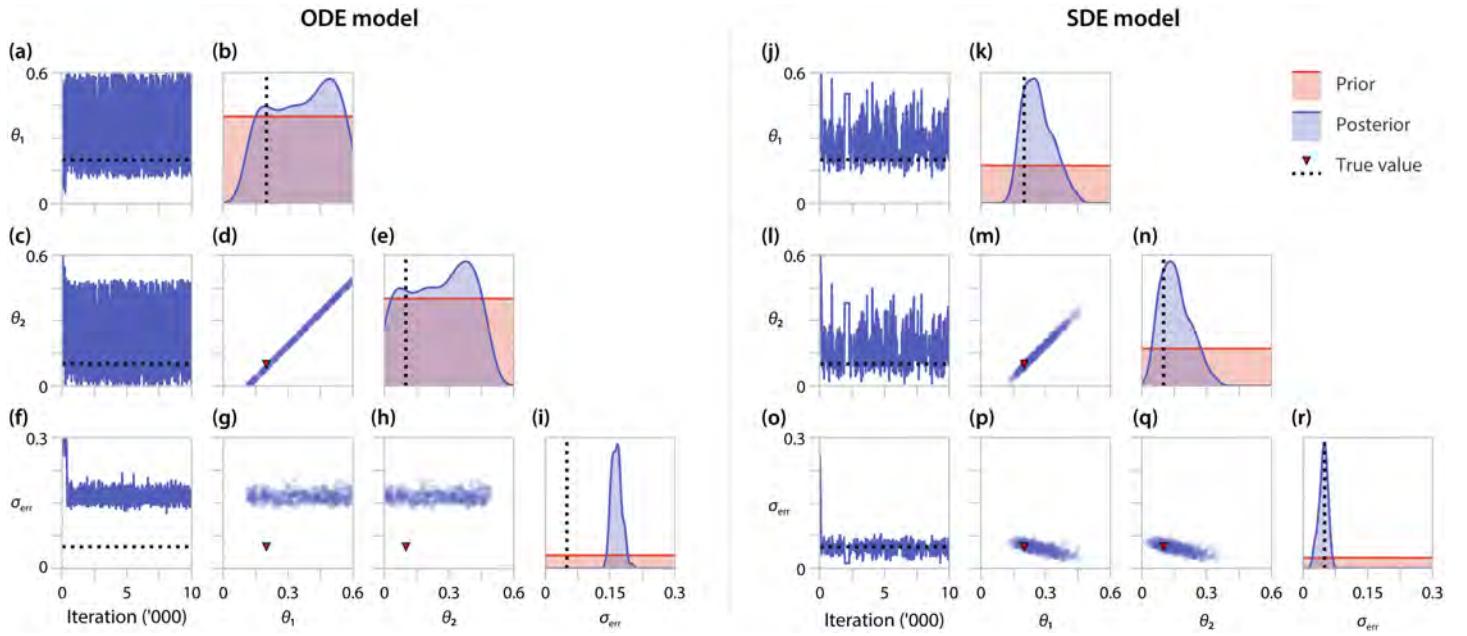
$$\widetilde{X}(t) = \widetilde{X}(0) \exp\left\{(\theta_1 - \theta_2)t\right\}. \tag{4.23}$$

Here, the population, $\widetilde{X}(t)$, undergoes exponential growth with a net-growth rate of $\theta_1 - \theta_2$. Therefore, intuitively, it is not possible to identify θ_1 and θ_2 if only average growth behaviour is observed [186].

Structural identifiability

We first assess structural identifiability of the moment equations in DAISY [198]. If only the first moment, n_1, is observed, the system is structurally non-identifiable, meaning the model parameters cannot be uniquely estimated with any amount of data. However, the system becomes structurally identifiable if n_2 is also observed. As the moment equations are closed at every order, and therefore exact, this analysis indicates that the ODE model (equation (4.22), corresponding to the first moment equation) is structurally non-identifiable, while the SDE model is structurally identifiable.

These structural identifiability results can be intuitively understood through re-parameterisation [200]. The first moment equation (or the ODE model) can be re-parameterised with $\tilde{\theta}_1 = \theta_1 - \theta_2$ where $\tilde{\theta}_1$ is the sole parameter in the model. Therefore, for a fixed $\tilde{\theta}_1$, all values on the line $\theta_1 = \tilde{\theta}_1 + \theta_2$ produce

▲ Figure 4.4. MCMC results for (*a–i*) an ODE and (*j–r*) an SDE description of the birth-death process. (*a,c,f*) and (*j,l,o*) show trace plots for the ODE and SDE models, respectively. Kernel density estimates of the posterior for each parameter ((*b,e,i*) and (*k,n,r*)), and bivariate scatter plots ((*d,g,h*) and (*m,p,q*)), are produced by thinning the MCMC chains by using every 100th sample from four independent MCMC chains, after burn-in.

indistinguishable behaviour in the first moment, m_1, and hence in the observation, n_1. On the other hand, when re-parameterised the second moment equation contains a second, linearly independent, parameter $\tilde{\theta}_2 = \theta_1 + \theta_2$. For the birth-death process, the second moment provides enough additional information to uniquely identify both parameters θ_1 and θ_2, provided enough data is available. Thus, the birth-death process is structurally identifiable from the first two moments.

Practical identifiability

We assess practical identifiability of the parameter vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \sigma_{\mathrm{err}})$ for the ODE and SDE models using MCMC. We place independent uniform priors on each parameter so that $p(\theta_1) = p(\theta_2) = \mathcal{U}(0, 0.6)$ and $p(\sigma_{\mathrm{err}}) = \mathcal{U}(0, 0.3)$. If prior knowledge about the population (i.e., the cell line) is available, perhaps based upon previously conducted experiments, this can be incorporated into the analysis through an informative prior. For example, upper bounds that define reasonable values for biological parameters are routinely applied in this context [27].

In figure 4.4a–i, we show MCMC results for the birth-death process using the ODE model. Based on the structural identifiability results, we expect the likelihood (and for a uniform prior, the posterior density) to be constant along the identifiable parameter combination $\tilde{\theta}_1 = \theta_1 - \theta_2$, and we see this in figure 4.4d. These results also suggest that, should one of θ_1 or θ_2 be known (for example, if the cells are treated with an anti-proliferative drug that enforces $\theta_2 = 0$ [298]) the other be identifiable. However, lower and upper bounds for θ_1 and θ_2, respectively, are able to be established as a direct consequence of the prior assumption that all parameters are strictly positive. Examination of univariate credible intervals, shown in table 4.1, reveals that each parameter cannot individually be identified within 3–4 orders of magnitude, a hallmark of non-identifiability [189]. We note that σ_{err} is practically identifiable (figure 4.4i, 95% CrI: (0.1448, 0.1907))

		ODE			SDE		
	True	95% CrI	$\hat{R}$	S_{eff}	95% CrI	$\hat{R}$	S_{eff}
θ_1	0.2	(0.1276,0.5891)	1.00056	2292	(0.1609,0.4059)	1.01068	104
θ_2	0.1	(0.0130,0.4744)	1.00056	2300	(0.0477,0.3016)	1.01107	107
σ_{err}	0.05	(0.1448,0.1907)	1.00242	2254	(0.0270,0.0667)	1.00079	364

▲ Table 4.1. 95% credible intervals, and diagnostics, for the parameter estimates for the birth-death process. Credible intervals are approximated using the MCMC quantiles after burn-in.

from the ODE model, however it will always be overestimated as the observation model for the ODE model must also account for the intrinsic noise of the process.
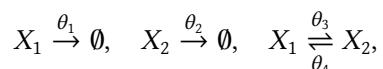
We repeat the analysis for the SDE model, results of which are shown in figure 4.4j–r. For the prior support chosen, both θ_1 and θ_2 are practically identifiable, as seen in figure 4.4k,n. Further, 95% credible intervals identify each parameter within a single order of magnitude (table 4.1). While structural identifiability analysis revealed that the SDE model is identifiable in the limit of infinite, noise-free data, it is not necessarily so for data with a realistic signal-to-noise ratio, characterised by the noise model parameter σ_{err}. In our case, if prior knowledge provided an upper bound for θ_1 and θ_2 at, for example, 0.3, conclusions of practical identifiability may be analogous to those of the ODE model. We see this in table 4.1, where the upper bounds of the credible intervals for θ_1 and θ_2 extend beyond 0.3. This is also evident from both the bivariate scatter plot (figure 4.4m) and MCMC trace plots (figure 4.4j,l), where posterior samples above 0.3 are regularly drawn for both θ_1 and θ_2. As the SDE explicitly accounts for intrinsic noise, σ_{err} is identifiable with estimates close to the true value, in contrast to results from the ODE model.

4.3.2 Two-pool model

Next, we consider partial observations of a process governed by a two-pool model, describing the decay of a substance that is able to transfer between two pools (figure 4.2b). Identifiability of a two-pool model was first examined in the fundamental study of Bellman and Åström [194] as they introduced the concept of structural identifiability. The model can represent, for example, human cholesterol distribution dispersed through two-pools (for example, two organs), where measurements are taken from a tracer in the first pool [262]. Bellman [194] and later Cobelli [195] show that, for an ODE model, the pool transfer and decay rates are not structurally identifiable. We consider practical identifiability for synthetic data comprising noisy measurements of the first pool at 10 equally spaced time points in five identically prepared experiments. Although measurements of the second pool are not taken, we assume, for demonstration purposes, that the initial concentration in each pool is zero before a known amount is introduced to the first pool, thus the full initial condition is known. In practice, the initial condition may also depend on a set of unknown parameters, and we focus on this with the epidemic model.

Model formulation and moment equations

The two-pool model can be expressed as the bio-chemical reaction network

$$X_1 \xrightarrow{\theta_1} \emptyset, \quad X_2 \xrightarrow{\theta_2} \emptyset, \quad X_1 \underset{\theta_4}{\overset{\theta_3}{\rightleftharpoons}} X_2,$$

with stoichiometries $\boldsymbol{\nu}_1 = (-1, 0)^T$, $\boldsymbol{\nu}_2 = (0, -1)^T$, $\boldsymbol{\nu}_3 = (-1, 1)^T$ and $\boldsymbol{\nu}_4 = (1, -1)^T$; and propensities $a_1(X_t) = \theta_1 X_1$, $a_2(X_t) = \theta_2 X_2$, $a_3(X_t) = \theta_3 X_1$ and $a_4(X_t) = \theta_4 X_2$. Here, we denote $X_t = (X_{1,t}, X_{2,t})^T$ as the concentration of cholesterol in the first and second pools, respectively. The observed concentration, Y_t, is described by the noise model

$$Y_t = X_{1,t} + \xi_t, \quad \xi_t \sim \mathcal{N}(0, \sigma_{\text{err}}^2), \tag{4.24}$$

in which we consider that the data are subject to measurement error in the form of additive Gaussian noise [172, 258, 299]. We show 100 realisations of the SSA for the two-pool model in figure 4.3b, and the synthetic data used for practical identifiability analysis in figure 4.3f. The data are generated using the initial condition $X_0 = (100, 0)^T$ and target parameter values $\theta_1 = 0.1$, $\theta_2 = 0.2$, $\theta_3 = 0.2$, $\theta_4 = 0.5$ and $\sigma_{\text{err}} = 2$. Here, we note that $\sigma_{\text{err}} \ll X_t$ (figure 4.3c), which ensures $Y_t > 0$.

The CLE for the two-pool model is

$$
\begin{aligned}
dX_t = &\begin{pmatrix} \theta_4 X_{2,t} - (\theta_1 + \theta_3) X_{1,t} \\ \theta_3 X_{1,t} - (\theta_2 + \theta_4) X_{2,t} \end{pmatrix} dt \\
&+ \begin{pmatrix} -\sqrt{\theta_1 X_{1,t}} & 0 & -\sqrt{\theta_3 X_{1,t}} & \sqrt{\theta_4 X_{2,t}} \\ 0 & -\sqrt{\theta_2 X_{2,t}} & \sqrt{\theta_3 X_{1,t}} & -\sqrt{\theta_4 X_{2,t}} \end{pmatrix} dW_t,
\end{aligned}
\tag{4.25}
$$

and the moment equations are given to second order by
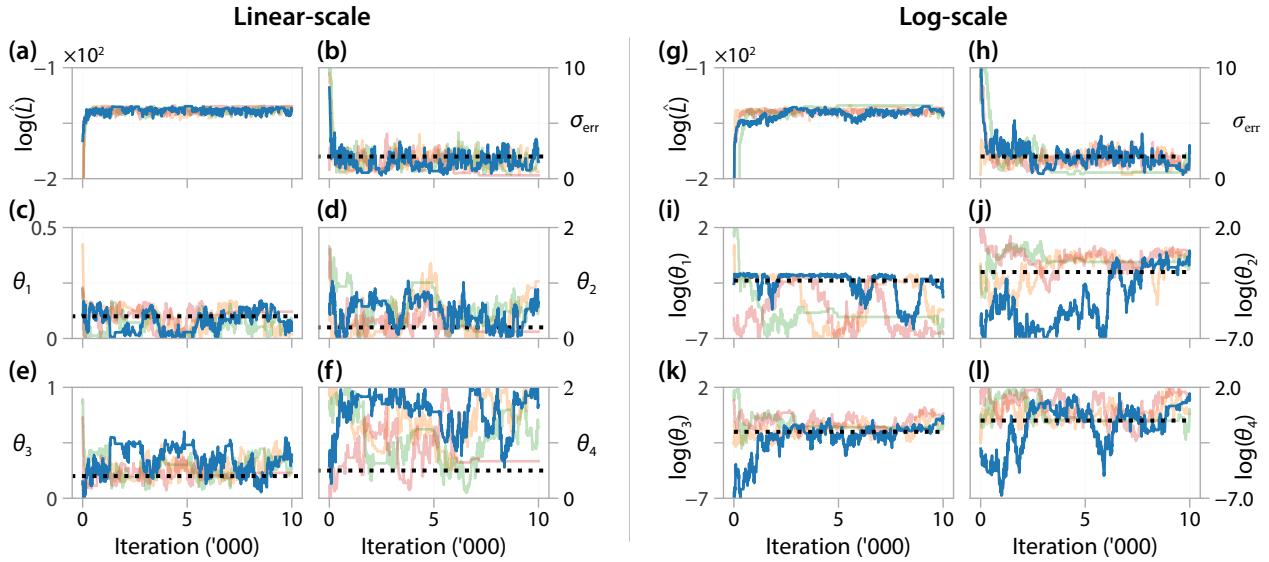
$$
\begin{aligned}
\frac{dm_{10}}{dt} &= \theta_4 m_{01} - (\theta_1 + \theta_3) m_{10}, \\
\frac{dm_{01}}{dt} &= \theta_3 m_{10} - (\theta_2 + \theta_4) m_{01}, \\
\frac{dm_{20}}{dt} &= \theta_4 (m_{01} + 2m_{11}) + (\theta_1 + \theta_3)(m_{10} - 2m_{20}), \\
\frac{dm_{02}}{dt} &= \theta_3 (m_{10} + 2m_{11}) + (\theta_2 + \theta_4)(m_{01} - 2m_{02}), \\
\frac{dm_{11}}{dt} &= -(\theta_1 + \theta_2) m_{11} - \theta_4 (m_{01} - m_{02} + m_{11}) \\
&\quad - \theta_3 (m_{10} + m_{11} - m_{20}).
\end{aligned}
\tag{4.26}
$$

The moments of the observed cholesterol concentration are given in the noise-free limit by $n_1 = m_{10}$ and $n_2 = m_{20}$. As with the birth-death process, all elements of $\boldsymbol{\alpha}(\cdot)$ and $\boldsymbol{\sigma}(\cdot)\boldsymbol{\sigma}(\cdot)^T$ are linear in X_t, so the moment equations are closed at every order and, therefore, exact.

Structural identifiability

The two-pool model provides an archetypical example of structural non-identifiability in an ODE model [194, 195]. Unless a restriction is placed on one of the parameters (for example, if decay of the substance can only occur from the first pool so $\theta_2 = 0$), the model parameters are structurally non-identifiable: many parameter combinations give identical behaviour in the ODE model. Therefore, the model parameters cannot be uniquely determined from any amount of noise-free experimental data if observations are made from only the first pool.

We assess structural identifiability of an SDE description of the two-pool SDE model using DAISY with

▲ **Figure 4.5.** Pilot MCMC trace plots, and log-likelihood estimates, of four chains for the two pool SDE model on with (*a–f*) untransformed parameters; and (*g–l*) transformed parameters. Priors for each parameter are uniform with support corresponding to the respective axis limits. The target parameters, used to generate synthetic data, are indicated (black dashed line).

the system of moment equations up to second order (equation (4.26)). While the ODE model is structurally non-identifiable, the SDE model is structurally *locally* identifiable: the parameters identifiable, but non-uniquely. For the parameter values considered, which are rates and, by definition, positive, DAISY indicates that the parameters are uniquely identifiable. Therefore, in the limit of infinite, noise-free data, the model parameters can be determined from an SDE description of the two-pool model.

Practical identifiability

To assess practical identifiability of the two-pool model, we apply MCMC to infer $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4, \sigma_{\mathrm{err}})$. Initially, independent uniform priors are chosen such that $p(\theta_1) = \mathcal{U}(0, 0.5)$, $p(\theta_2) = \mathcal{U}(0, 2)$, $p(\theta_3) = \mathcal{U}(0, 1)$, $p(\theta_4) = \mathcal{U}(0, 0.5)$, and $p(\sigma_{\mathrm{err}}) = \mathcal{U}(0, 10)$. The support of each prior is chosen to cover a range of magnitudes over the target parameter values. Results from four independent pilot chains, each initiated at a random sample from the prior, are shown in figure 4.5a–f. In figure 4.5a we see that the log-likelihood estimate rapidly stabilises, indicating that the chain has moved to a high-likelihood region of the parameter space. Results for σ_{err} and θ_3 also rapidly stabilise, indicating that these parameters are practically identifiable [189]. Results for the remaining three kinetic rate parameters in figure 4.5c,d,f indicate that θ_1, θ_2 and θ_4 are practically non-identifiable. In particular, chains for θ_1 and θ_2 spend a non-negligible time near zero, indicating that the model may be indistinguishable (using the available data) from a model where removal only occurs from a single pool.

We next repeat the analysis using MCMC to infer $\boldsymbol{\theta}_* = (\log \theta_1, \log \theta_2, \log \theta_3, \log \theta_4, \sigma_{\mathrm{err}})$. Inferring the logarithm of rate parameters will provide more detailed information about the magnitude of rate parameters potentially close to zero [260]. This transformation provides an excellent example of why even a uniform prior is informative, since a uniform prior placed on the linear-scale is not uniform on the log-scale. A uniform prior on the linear-scale makes parameters of a smaller magnitude less likely than a larger magnitude. The priors are again chosen to be independent and uniform (on the log-scale), such that $p(\log \theta_i) = \mathcal{U}(-7, 2)$

	True	95% CrI	$\hat{R}$	S_{eff}
θ_1	0.1	(0.0042,0.1503)	1.0024	510
θ_2	0.2	(0.0307,1.0699)	1.0014	456
θ_3	0.2	(0.1356,0.4857)	1.0023	515
θ_4	0.5	(0.4372,1.9585)	1.0004	741
σ_{err}	2.0	(0.5715,2.8773)	1.0089	409

◀ Table 4.2. 95% credible intervals, and diagnostics, for the parameter estimates (on the linear-scale) for the two-pool model. Credible intervals are approximated using the MCMC quantiles after burn-in.
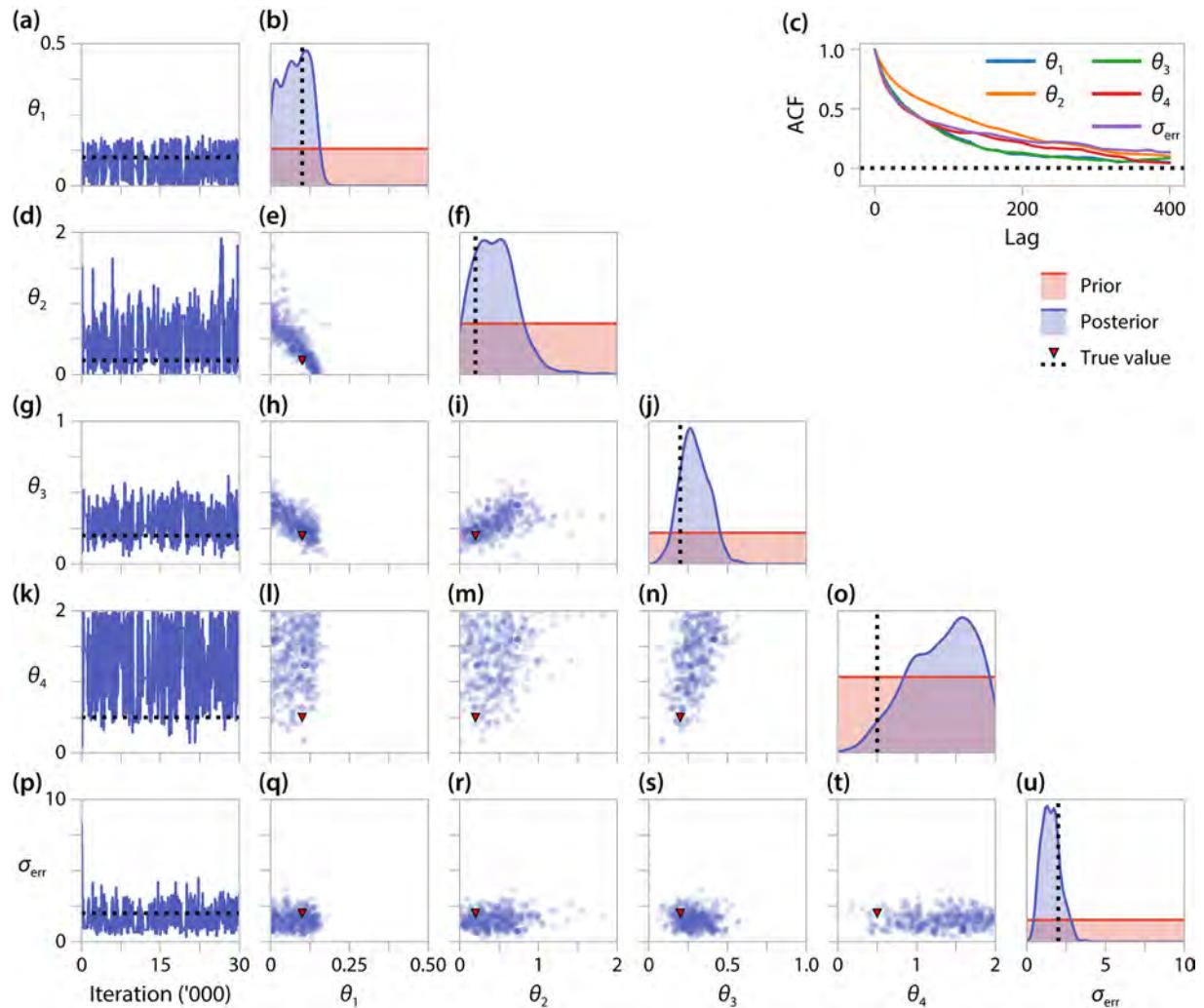
for all i and $p(\sigma_{\text{err}}) = \mathcal{U}(0, 10)$ as before. The support of each prior is chosen, again, to cover a range of magnitudes above and below that of the target parameter values. Results in figure 4.5k confirm that θ_3 is practically identifiable, while θ_2 and θ_4 are practically non-identifiable. From results in figure 4.5l we term θ_4 *one-sided identifiable*: the parameter has an identifiable lower bound, and is distinguishable from zero.

To visualise correlations between inferred parameters, we tune the proposal kernel (equation (4.15)) and run the MCMC algorithm for 30,000 iterations, results are shown in figure 4.6 and table 4.2. If only the univariate marginal distributions are considered, all parameters except for θ_4 may be classified as practically identifiable. However, our analysis shows that θ_1 and θ_2 are distinguishable only within a large range of magnitudes. A strong correlation is seen between θ_1 and θ_2, indicating that the total substance exit rate, $\theta_1 + \theta_2$, may be practically identifiable. If one of θ_1 or θ_2 were known in advance, perhaps based on past experimental knowledge, the other may become practically identifiable. Further, results from the tuned chains verify that θ_3 is practically identifiable (95% CrI (0.1356,0.4857)) and θ_4 is distinguishable from zero.

4.3.3 Epidemic model

Here, we consider a four-compartment epidemic model — the SEIR model [263–265, 300] (figure 4.2c). In this model, susceptible individuals, S, are infected due to interactions with infectious individuals, I, and undergo an unknown period of time during which they have been exposed, E, but are not themselves infectious. Infectious individuals either recover or are removed from the total population, R. A noisy unknown proportion, ξ, with mean μ_{obs}, of the number of infectious and recovered individuals is monitored. This captures a testing regime where not all infectious or recovered individuals are tested. We supplement these results by considering a scenario where the same unknown proportion of the exposed individuals is also monitored during the early part of the epidemic.
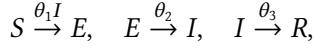
The kind of data available for the epidemic model differs significantly from that for the experiment-based models we have considered thus far: we are interested in a practical identifiability problem where data from only a single time-series is available, which mirrors data available from an actual epidemic [301]. We first consider practical identifiability using data from the early part of the epidemic, before the number of cases is observed to decrease. Next, these results are compared to a case where data further through the course of the epidemic is considered (figure 4.3g). Initially, 10 infected individuals and 10 recovered individuals are detected. For simplicity we assume there is no noise in these initial observations, so the number of infected and recovered individuals is given by $10/\mu_{\text{obs}}$. An unknown number of individuals, E_0, are initially exposed. In our analysis, we assume that E_0 is not of direct interest, and we class it a nuisance parameter.

▲ **Figure 4.6.** Tuned MCMC results for the two-pool model with a parameters on the linear-scale. The left-most column shows an MCMC trace from a single chain. Kernel density estimates of the marginal posterior for each parameter and bivariate scatter plots are produced using every 300th sample from four independent MCMC chains, after burn-in. The autocorrelation function for a single chain is shown in (*c*), indicating that every 300th sample is approximately independent.

Model formulation and moment equations

The SEIR model can be represented by the following bio-chemical reactions

$$S \xrightarrow{\theta_1 I} E, \quad E \xrightarrow{\theta_2} I, \quad I \xrightarrow{\theta_3} R,$$

with stoichiometries $\boldsymbol{\nu}_1 = (-1, 1, 0, 0)^T$, $\boldsymbol{\nu}_2 = (0, -1, 1, 0)^T$ and $\boldsymbol{\nu}_3 = (0, 0, -1, 1)^T$; and propensities $a_1(X_t) = \theta_1 S_t I_t$, $a_2(X_t) = \theta_2 E_t$ and $a_3(X_t) = \theta_3 I_t$. Here, we denote $X_t = (S_t, E_t, I_t, R_t)^T$ as the number of individuals in each compartment. Two observations are made,

$$Y_{1,t} = \xi_{1,t} I_t, \qquad\qquad \xi_{1,t} \sim \mathcal{N}(\mu_{\text{obs}}, \sigma_{\text{err}}^2), \qquad\qquad (4.27)$$

$$Y_{2,t} = \xi_{2,t} R_t, \qquad\qquad \xi_{2,t} \sim \mathcal{N}(\mu_{\text{obs}}, \sigma_{\text{err}}^2). \qquad\qquad (4.28)$$

Here, $Y_{1,t}$ and $Y_{2,t}$ describe the observed number of infected individuals and recovered individuals, respectively. We further assume that μ_{obs}, the average observed proportion; and σ_{err}, the observation error, are unknown and must be estimated. We show 100 realisations of the SSA for the epidemic model in figure 4.3c, and synthetic data used for practical identifiability analysis in figure 4.3g. The data are generated using the initial condition $X_0 = (500 - E_0, E_0, 10/\mu_{\text{obs}}, 10/\mu_{\text{obs}})^T$ and target parameter values $\theta_1 = 0.01$, $\theta_2 = 0.2$, $\theta_3 = 0.1$, $E_0 = 20$, $\mu_{\text{obs}} = 0.5$ and $\sigma_{\text{err}} = 0.05$. Here, we note that $\sigma_{\text{err}} \ll \mu_{\text{obs}}$, ensuring that $Y_{1,t}$ and $Y_{2,t}$ remain positive.

The moment equations differ from the previous two models considered in that they are not closed. Therefore, the first order moment equations are not equivalent to those for the corresponding ODE model [190], unless a mean-field closure is drawn at first order. To make progress, we close the moment equations after second order to form an approximate system of moment equations for the first two moments. We give the system of 14 moment equations, under all three moment closures considered, in Appendix C. The moments of the observation variables are given in the noise-free limit by

$$n_{ij}(t) = \mu_{\text{obs}}^{i+j} m_{00ij}(t), \quad i + j \leq 2. \qquad\qquad (4.29)$$

In Appendix C, we produce numerical solutions to the moment equations for the epidemic model for each closure considered (figure C.1). All closures predict visually identical behaviour at first order, and the pair-approximation and Gaussian closures are in agreement at second order. For the target parameters we consider, the mean-field closure does not agree at second order with the more advanced closures. Whereas a numerical solution to the moment equations for the pair-approximation and Gaussian closures is readily obtainable from a standard solver in Julia [302], the mean-field closure required a positivity-preserving Patankar-type method [303] to avoid blow up.

Structural identifiability

We assess structural identifiability of the approximate system of moment equations in DAISY and GenSSI2, results are shown in table 4.3. The ODE model, equivalent to a mean-field closure (equation (4.10)) drawn after the first moment, is structurally non-identifiable. The second-order systems, for all closures, are structurally identifiable (table 4.3). As the second-order systems are approximate, this analysis is not conclusive for the

Model	Structure identifiability	Runtime	
		DAISY	GenSSI2
ODE	Non-Identifiable	5 seconds	5 seconds
SDE (mean-field closure)	Identifiable	5 minutes	2 seconds
SDE (pair-wise closure)	Identifiable	16 hours	2 seconds
SDE (Gaussian closure)	Identifiable	7 hours	2 seconds

▲ Table 4.3. Structural identifiability of the partially observed SEIR model assessed in DAISY and GenSSI2. Structural identifiability of the SDE is assessed using each closure method for third and higher order moments. Note that the ODE model is equivalent to the SDE model with a mean-field closure for second and higher order moments. Runtimes correspond to a 3.7GHz quad-core i7 desktop machine running Windows 10.

SDE. However, we can conclude that if the mean and variance of the epidemic model (the first two moments) are modelled using the system of moment equations, and data is available accordingly, the parameters are identifiable in the limit of infinite, noise-free data. We highlight the computational cost in DAISY of introducing complexity into the moment equations through the closure methods. The pair-wise closure, equation (4.11), which introduces a quotient, and the Gaussian closure, equation (4.12), which introduces a cubic, take significantly longer using DAISY to assess than the mean-field closure, equation (4.10), yet give the same result. However, unlike MCMC, we note that structural identifiability results are deterministic, and independent of user choices such as prior, number of particles, and generated or real synthetic data.
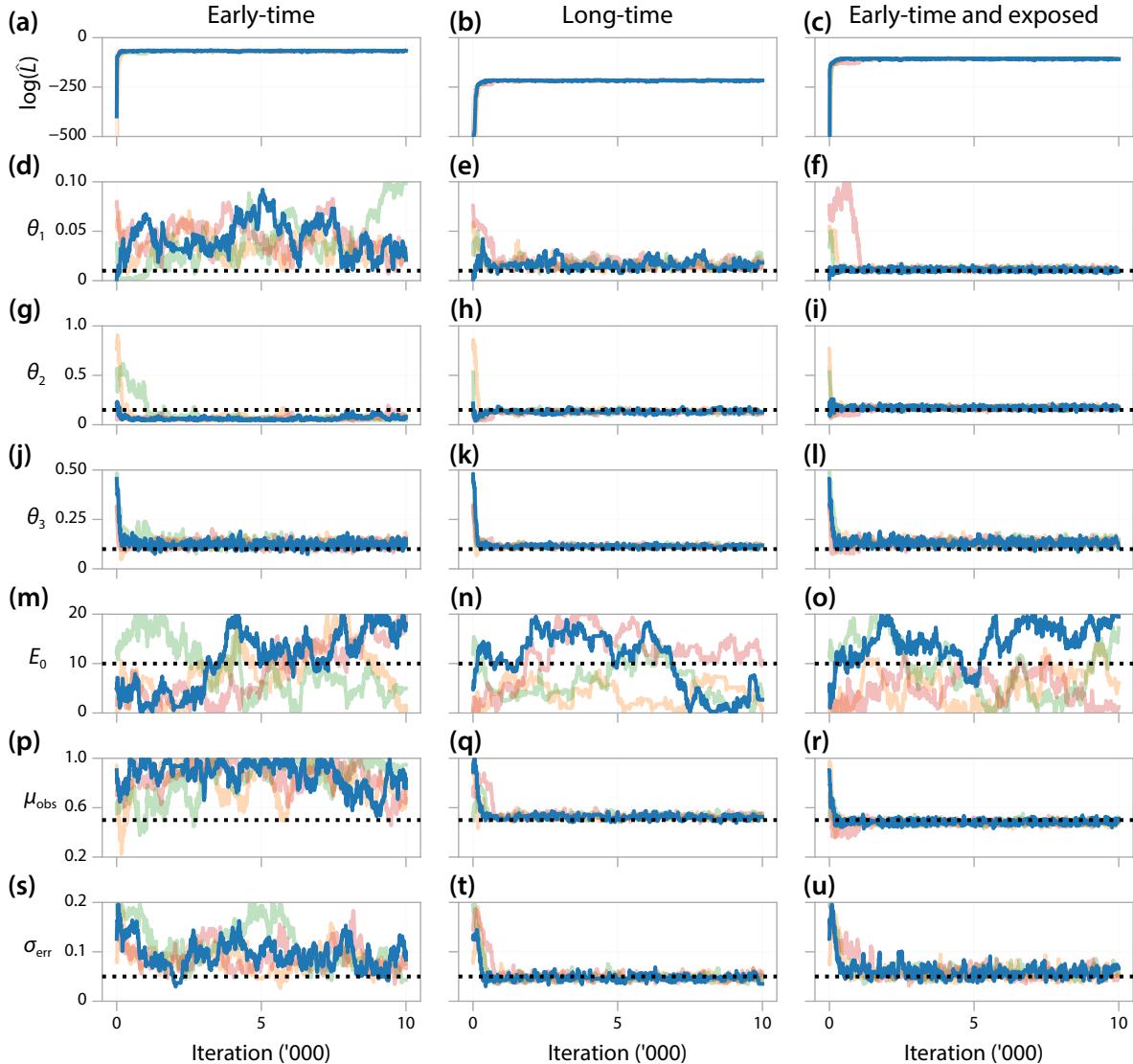
Practical identifiability

We assess practical identifiability of the epidemic model using MCMC to infer $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, E_0, \mu_{\text{obs}}, \sigma_{\text{err}})$. Independent uniform priors are placed on each parameter so that $p(\theta_1) = \mathcal{U}(0, 0.1)$, $p(\theta_2) = \mathcal{U}(0, 1)$, $p(\theta_3) = \mathcal{U}(0, 0.5)$, $p(E_0) = \mathcal{U}(0, 20)$, $p(\mu_{\text{obs}}) = \mathcal{U}(0.2, 1)$ and $p(\sigma_{\text{err}}) = \mathcal{U}(0, 0.2)$. Results are shown in figure 4.7, where we initiate each chain at the same location for all forms of data we consider.
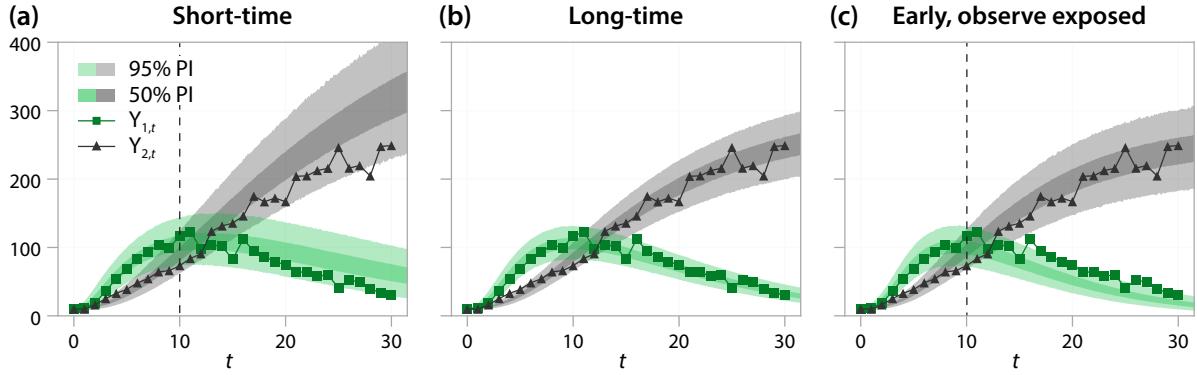
First, we assess identifiability when only early-time data is available. The log-likelihood estimate rapidly stabilises (figure 4.7a), indicating that the chains have moved to a high-likelihood region of the parameter space [189]. Results for θ_3, the recovery rate, also stabilise, indicating that θ_3 is structurally identifiable. Eventually, we see the estimate for θ_2 stabilises in all chains, however they under-estimate the target value, although proposals equal to and greater than the target value $\theta_2 = 0.2$ are occasionally accepted. To compensate, the estimate of θ_1 stabilises, and covers a region an order of magnitude greater than the target ($\theta_1 = 0.01$). Therefore, although θ_1 is practically identifiable to a large, but finite, range of values, we classify θ_1 as non-identifiable from the short-time data. Estimates for E_0 and μ_{obs} in figure 4.7m,p do not stabilise, and are practically non-identifiable.

Next, we consider a scenario where long-time data are available, such that the number of infected individuals is observed to eventually decrease. The log-likelihood estimate (figure 4.7b) and chains for all parameters, except E_0, are observed to stabilise, indicating that all parameters of interest are now practically identifiable. We supplement these results by considering a third scenario, where only early-time data are available, but the same unknown proportion of the number of exposed individuals is also monitored. As with the long-time data, all parameters of interest are now practically identifiable.

We perform a posterior predictive check [290] of the epidemic model to compare the model prediction—

▲ Figure 4.7. Pilot MCMC trace plots, and log-likelihood estimate, of four chains for the epidemic model. We consider data comprising noisy observations of an unknown proportion of the number of infected and recovered individuals during the early part of the epidemic (first column) and throughout the epidemic (second column). We supplement these results by considering the case we are also able to observe the same unknown proportion of the number of exposed individuals during the early part of the epidemic (third column). Priors for each parameter are uniform with support corresponding to the respective axis limits. The target parameter set, used to generate synthetic data, are indicated (black dashed line).
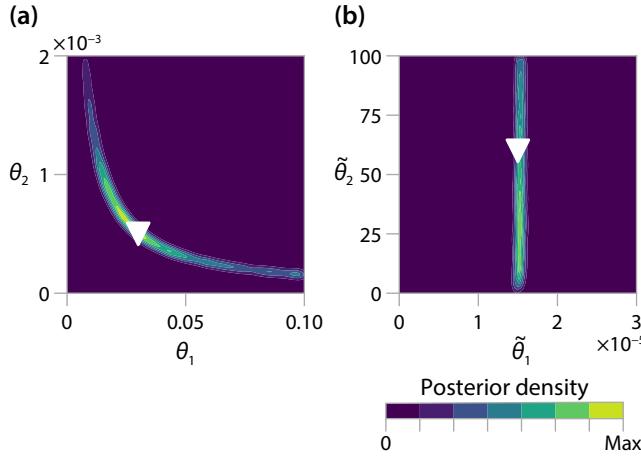
▲ Figure 4.8. Posterior predictive distribution for the epidemic model using (*a*) short-time data; (*b*) long-time data; and (*c*) short-time data where observations are also made of the number of exposed individuals. In (*a,c*), the dashed line indicates the last observation point used for inference. The first 3,000 samples from each pilot chain is discarded as burn-in. We resample 10,000 parameter combinations (with replacement) and solve the SDE model to estimate posterior predictive intervals (PIs). Shown are 50% (darker) and 95% (lighted) prediction intervals computed from the quantiles of the posterior predictive distribution.

which accounts for parameter uncertainty, intrinsic noise and observation error—to the synthetic data used for inference. We discard the first 3,000 samples from each pilot chain as burn-in, and resample 10,000 parameter combinations for each data type considered. Results in figure 4.8 show that, in all cases, the model predictions are in agreement with the full time-course (although, we note, the long-time data is only used to calibrate parameters in figure 4.8b). Results in figure 4.8a, for the short–time data, highlight how practical non-identifiability affects model predictions. These results predict an epidemic size at $t = 30$ is noticeably wider and higher than those for the data types where θ_1 is practically identifiable. Further, the lower 95% credible interval for the observed number of infected individuals reduces much faster than that predicted by the other data types.

4.3.4 β-insulin-glucose circuit

Finally, we consider a non-linear model of glucose homeostasis, the β-insulin-glucose circuit [266, 267] (figure 4.2d). Parameterising mathematical models of glucose homeostasis is important for the development of patient-specific insulin delivery for type 1 diabetics [208]. Time-series data of blood glucose concentration is available from continuous glucose monitoring sensors, a critical component of type 1 diabetes management [209, 229], an example of which is shown in figure 4.1f. The model describes the regulation of blood plasma glucose by insulin secreted by pancreatic β cells. Glucose is introduced into the system through a base production plus a meal intake, $u(t)$, and decays linearly according to the insulin concentration. Insulin is secreted by β cells at a rate given by a non-linear Hill function [267]. β cells are produced and decay in a non-response to the glucose concentration. We consider identifiability for synthetic data comprising noisy measurements of the β cell and glucose concentrations, but not the insulin concentration. The data consists

(a)

(b)

◀ Figure 4.9. Kernel density of the bivariate marginal posterior distribution of the biophysical parameters in the β-insulin-glucose circuit, using the ODE and 100,000 pilot MCMC iterations (the first 3,000 are discarded as burn-in. (*a*) The posterior for the untransformed parameters, (θ_1, θ_2) shows non-identifiability. (*b*) The posterior for the transformed parameters, $(\tilde{\theta}_1, \tilde{\theta}_2)$, demonstrates that $\tilde{\theta}_1 = \theta_1 \theta_2$ is identifiable, but $\tilde{\theta}_2 = \theta_1 / \theta_2$ is not.

of five independent experiments, each comprising 15 time-series observations following a meal intake. We only consider inference for two biophysical parameters: θ_1, the insulin secretion rate; and θ_2, the insulin sensitivity. The non-linearities in the model mean that the moment equation approach is not available, and inference using MCMC is computationally expensive. We demonstrate how structural identifiability analysis of the corresponding ODE system [26] can guide analysis of the SDE system and alleviate some of the computational challenges.
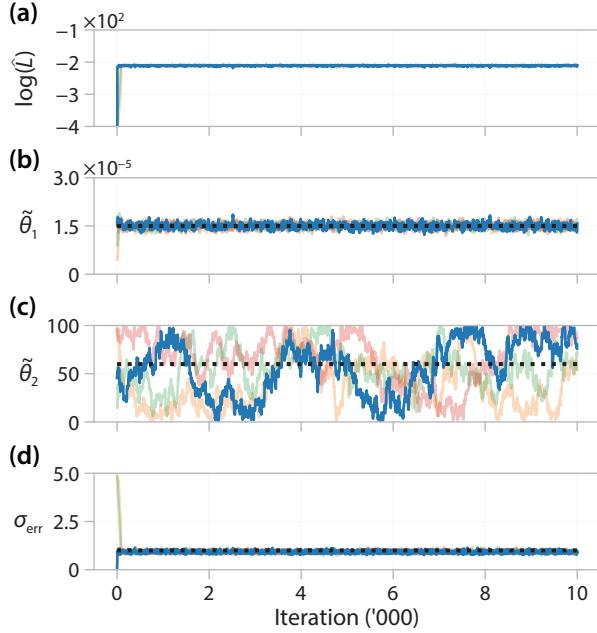
Model formulation

We consider a stochastic analogue of the model presented by Karin *et al.* [267]. Denoting $X_t = (\beta_t, I_t, G_t)^T$ as the concentrations of β cells, insulin and glucose, respectively, the propensity functions and corresponding stoichiometries are given by

$$
\begin{aligned}
a_1(X_t, t) &= \beta(t)\lambda_+(G_t), & \nu_1 &= (1, 0, 0)^T, \\
a_2(X_t, t) &= I_t\lambda_-(G_t), & \nu_2 &= (-1, 0, 0)^T, \\
a_3(X_t, t) &= \theta_1 \beta_t \rho(G_t), & \nu_3 &= (0, 1, 0)^T, \\
a_4(X_t, t) &= \gamma I_t, & \nu_4 &= (0, -1, 0)^T, \\
a_5(X_t, t) &= u_0, & \nu_5 &= (0, 0, 1)^T, \\
a_6(X_t, t) &= u(t), & \nu_6 &= (0, 0, 1)^T, \\
a_7(X_t, t) &= cG_t, & \nu_7 &= (0, 0, -1)^T, \\
a_8(X_t, t) &= \theta_2 I_t G_t, & \nu_8 &= (0, 0, -1)^T,
\end{aligned}
$$

where

$$
\lambda_+(G_t) = \frac{\mu_+}{1 + \left(\dfrac{8.6}{G_t}\right)^{1.7}}, \qquad \lambda_-(G_t) = \frac{\mu_-}{1 + \left(\dfrac{G_t}{4.8}\right)^{8.5}},
$$

$$
\rho(G_t) = \frac{G_t^2}{\eta^2 + G_t^2}, \qquad u(t) = \begin{cases} 0.2, & t < 50, \\ 0, & t \geq 50. \end{cases}
$$

(a) $\log(\hat{L})$

(b) $\tilde{\theta}_1$

(c) $\tilde{\theta}_2$

(d) σ_{err}

◄ Figure 4.10. Pilot MCMC trace plots, and log likelihood estimate, of four chains for the β-insulin-glucose circuit in the transformed parameter space. The likelihood quickly stabilises, but estimates for $\tilde{\theta}_2$ do not, indicating practical non-identifiability. Priors for each parameter are uniform with support corresponding to the respective axis limits. The target parameter set, used to generate synthetic data, are indicated (black dashed line).

Since β_t, I_t and G_t denote the concentrations of each substance, and not the population counts, we scale the diffusion term in the CLE to represent the relative concentrations of each substance [226]. Denoting N_β, N_I and N_G the relative concentration of β cells, insulin and glucose, respectively, we write

$$dX_t = \sum_{k=1}^{8} \nu_k a_k(X_t, t; \boldsymbol{\theta})dt + \text{diag}\left(\frac{1}{\sqrt{N_\beta}}, \frac{1}{\sqrt{N_I}}, \frac{1}{\sqrt{N_G}}\right) \sum_{k=1}^{8} \nu_k \sqrt{a_k(X_t, t; \boldsymbol{\theta})}dW_{k,t}. \qquad (4.30)$$
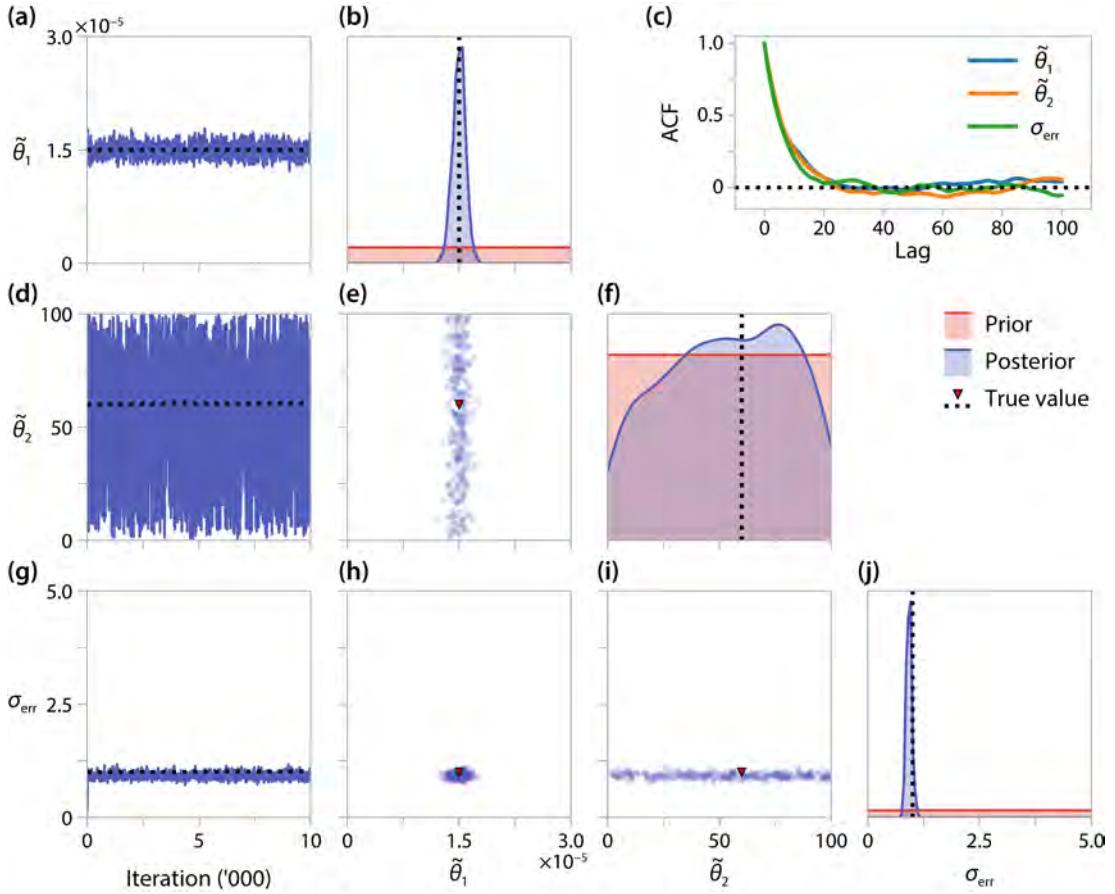
Two observations are made,

$$Y_{1,t} = \beta_t + \xi_{1,t}, \quad \xi_{1,t} \sim \mathcal{N}(0, \sigma_{\text{err}}^2),$$
$$Y_{2,t} = G_t + \xi_{2,t}, \quad \xi_{2,t} \sim \mathcal{N}(0, \sigma_{\text{err}}^2),$$

such that $Y_{1,t}$ and $Y_{2,t}$ are the observed β cell and glucose concentrations, respectively. We show 100 realisations of the SSA for the β-insulin-glucose circuit in figure 4.3d, and the synthetic data used for practical identifiability analysis in figure 4.3h. The data are generated using the initial condition $X_0 = (322, 10, 5)^T$ with fixed parameters, $\mu_+ = 0.021/(24 \times 60)$, $\mu_- = 0.025/(24 \times 60)$, $\eta = 7.85$, $\gamma = 0.3$, $u_0 = 1/30$, $c = 10^{-3}$, $N_\beta = 1$, $N_I = N_G = 20$, and target parameters $\theta_1 = 0.03$, $\theta_2 = 0.0005$ and $\sigma_{\text{err}} = 1$ [267]. Here, we note $\sigma_{\text{err}} \ll \beta_t, G_t$ (figure 4.3d), which ensures that $Y_{1,t}$ and $Y_{2,t}$ remain positive.

Parameter transform

Villaverde *et al.* [304] study structural identifiability of the corresponding ODE model using differential geometry. In the ODE model, θ_1 and θ_2 are structurally non-identifiable, unless the insulin concentration is also observed or one of these two parameters is known. We demonstrate this using MCMC in figure 4.8a, where the marginal posterior for (θ_1, θ_2) covers a hyperbolic region of the parameter space of equal posterior density. In the ODE model, the product $\theta_1\theta_2$ is structurally identifiable. To demonstrate this, we perform MCMC on the ODE model with transformed variables $\tilde{\theta}_1 = \theta_1\theta_2$ and $\tilde{\theta}_2 = \theta_1/\theta_2$, results shown in figure 4.8b.

▲ **Figure 4.11.** Tuned MCMC results for the β-insulin-glucose circuit in the transformed parameter space. The left-most column shows an MCMC trace from a single chain. Kernel density estimates of the marginal posterior for each parameter and bivariate scatter plots are produced using every 100th sample from four independent MCMC chains, after burn-in. The autocorrelation function for a single chain is shown in (c), indicating that every 20th sample is approximately independent.

These results also show how inefficient a naïve MCMC proposal can be when correlations between posterior parameters are non-linear. Structural identifiability analysis [304] indicates that the hyperbolic region defined by $\tilde{\theta}_1 = \theta_1 \theta_2$ (for a fixed $\tilde{\theta}_1$) produces indistinguishable behaviour, corresponding to a flat posterior when a uniform prior is applied. Despite this, the tail regions in figure 4.8a are rarely sampled, which could give the impression that the parameters are practically identifiable.

As the propensity functions for the β-insulin-glucose circuit model contain non-polynomial functions, we cannot produce an exact expression for the moment equations. Therefore, we only study practical identifiability using MCMC, and do not consider structural identifiability of the SDE for the β-insulin-glucose circuit using the moment equations. Motivated by the structural identifiability analysis of the ODE model, we use MCMC to infer $\boldsymbol{\theta} = (\tilde{\theta}_1, \tilde{\theta}_2, \sigma_{\text{err}})$, where we only consider the transformed variables $\tilde{\theta}_1 = \theta_1 \theta_2$ and $\tilde{\theta}_2 = \theta_1 / \theta_2$.

Practical identifiability

We show MCMC results from four pilot chains in figure 4.10. The log-likelihood estimate rapidly stabilises (figure 4.10a), as do results for $\tilde{\theta}_1$ and σ_{err} (figure 4.10b,d). As with the ODE model, $\tilde{\theta}_1$ is practically identifiable, but $\tilde{\theta}_2$ is not. To visualise possible correlations between inferred parameters, we tune the proposal kernel

(equation (4.15)) and run the MCMC algorithm for 10,000 iterations. The univariate marginal distributions, and MCMC trace plots, show that $\tilde{\theta}_1$ (95% CrI: $(1.34, 1.67) \times 10^{-5}$) and σ_{err} (95% CrI: $(0.812, 1.049)$) are practically identifiable, whereas $\tilde{\theta}_2$ is not (95% CrI: $(8.21, 97.79)$). No large correlations are seen between the parameters ($\rho(\tilde{\theta}_1, \tilde{\theta}_2) = 0.10$), and θ_2 is clearly practically non-identifiable as samples cover the entire range of the prior.

4.4 Discussion

Mathematical models are routinely calibrated to experimental data, with goals ranging from building a predictive model to quantifying biophysical parameters that cannot be directly measured. Much of the usefulness of calibrated models hinges on an assumption that model parameters are identifiable. Heavily over-parameterised models, with large numbers of practically non-identifiable parameters, are often referred to as *sloppy* in the systems biology literature [231,305–308]. Worryingly, these issues of parameter identifiability can often go undetected: models with non-identifiable parameters can still match experimental data (figure 4.8), but may have poor predictive power and provide little or no mechanistic insight [24]. Identifiability analysis is well-developed for deterministic ODE models, but there is little guidance in the literature to conducting such analysis for the stochastic models that are often vital for interpreting complex experimental data. In this chapter, we demonstrate how existing techniques can be applied to assess both structural and practical identifiability of SDE models in biology.

4.4.1 Moment dynamics approach

We demonstrate how existing ODE identifiability techniques can be applied directly to stochastic problems by formulating a system of moment equations. In the birth-death process and the two-pool model, the derived moment equations are closed and, therefore, exactly describe the time-evolution of the moments of the SDE. In these two case studies we find that the moment equations are structurally identifiable. This implies structural identifiability of the corresponding SDE model, and parameters can be uniquely estimated in the limit of infinite, noise-free data. For an SDE model this implies an infinite number of *observation-noise* free trajectories of the SDE, since the variability, which relates to higher-order moments, contains information. While we find that the two-pool model of cholesterol distribution is not practically identifiable, establishing structural identifiability is useful as it suggests to the practitioner that the observation process (i.e. observe cholesterol in the first pool) is sufficient, in principle, to fully parameterise the model.

For the epidemic model, the moment equations are not closed, so we study structural identifiability through an approximate system of second-order moment equations. The idea of studying identifiability through an approximate system was first suggested by Pohjanpalo [233], who studies identifiability of ODE systems through a power series expansion. The closed system of moment equations suggest the epidemic SDE model could be structurally identifiable, and these results agree, in our case, with practical identifiability detected using MCMC. More research is needed to establish how identifiability is affected when closing, or truncating, a system of moment equations. For example, if information required to identify model parameters is contained in third or higher order moments, results suggesting that a model is practically non-identifiable from a second order closure will not be indicative of non-identifiability in the SDE model. Furthermore, if structural identifiability differs between moment closures, such a preliminary screening tool needs to be interpreted with caution. If this were the case, a conclusion of structural identifiability is indicative of the model under a

particular closure. Recent work suggests that a finite number of moments often contain the information required to identify parameters [309], even for a bimodal distribution and if a closure is applied [310].

Due to the computational constraints placed on analysing structural identifiability of non-polynomial ODE models, we do not attempt to apply the moment dynamics approach to the stochastic β-insulin-glucose circuit model. However, for many models, a mean-field closure corresponds to an ODE description of the system, and studying identifiability of this ODE model can aid practical identifiability analysis of the corresponding SDE. In our case, the corresponding ODE model is structurally non-identifiable due to a hyperbolic relationship between the two parameters of interest: for a fixed $\theta_1\theta_2$, model outputs are indistinguishable [304]. The question of whether an SDE description can provide enough information to practically identify θ_1 and θ_2 can be answered through MCMC, however simple variants of MCMC can struggle when correlations between parameters are strong and non-linear. Therefore, we work in a transformed parameter space where, for the ODE model, $\tilde{\theta}_1 = \theta_1\theta_2$ is identifiable but $\tilde{\theta}_2$ is not (figure 4.8). This analysis provides a better sense of whether the SDE model captures enough information to identify the parameters, and provides more robust results that are less dependent upon choices made in the MCMC algorithm.

4.4.2 Particle Markov-chain Monte Carlo

We demonstrate practical identifiability by calibrating each model to synthetic data using particle MCMC. We observe the MCMC chains to stabilise in a region of high posterior density, after which time transitions produce samples from the posterior distribution [189]. By visualising MCMC trace plots, we see that estimates of practically identifiable parameters also stabilise, but those of practically non-identifiable parameters do not. These results also demonstrate that, although estimates made of practically identifiable parameters are precise (that is, within a reasonable level of confidence), they are not necessarily accurate. For example, in figure 4.7g, the rate at which exposed individuals become infectious is practically identifiable, but it is underestimated compared to the target value, which could hint at model misspecification.

Given that particle MCMC is computationally expensive, our implementation of a standard technique to detect identifiability from pilot chains carries several advantages. First, pilot chains are regularly generated in the early stages of many inference procedures to establish efficient proposal kernels. Practical identifiability can, therefore, be established as part of an existing workflow. Second, more sophisticated methods are by their very nature more difficult to implement and dependent on practitioner choices, which could obscure results and require more algorithmic experimentation. In comparison, we take an automated approach: aside from the model and choice of prior, the procedure to perform MCMC for each model we consider is identical. Once identifiability is established, the computational cost of MCMC can be alleviated to some extent by adopting a more efficient inference technique. For example, adaptive MCMC [311], sequential Monte-Carlo [312], multi-level methods [313–315], sub-sampling techniques [316] and model-based proposal methods [317] provide significant performance improvements over the standard technique we employ. Further, we expect applying higher-order SDE simulation algorithms, such as a Runge-Kutta method [318], or considering GPU approaches to particle MCMC [319], to improve performance.

As we calibrate to synthetic data for the purpose of a didactic demonstration, we take a pragmatic approach by treating the true values as unknown. Hence, we initiate each chain as a random sample from the prior distribution. This involves a burn-in phase before the MCMC chain settles in an area of high posterior density. For computationally expensive models, such as those found in the cardiac modelling literature [320], synthetic

data can be used with pilot chains initiated at the target values. If models have have already been calibrated to experimental data using, for example, maximum likelihood estimation, the chain can be initiated at the calibrated values. MCMC then, relatively cheaply, provides information about the posterior distribution about this point, akin to the Fisher information for models where it can be calculated [201].

MCMC can be applied to detect identifiability for any stochastic model provided an approximation to the likelihood is available. Recent developments to particle MCMC have seen its adoption for more complicated SDE models, such as SDE mixed effects models [321]. For systems with relatively small populations, it may be more appropriate to work directly with an SSA with, for example, a tau-leap method [212, 276]. Alternative approximations to the likelihood, such as those employed by ABC, may be necessary if model complexity requires; for example, should the model include spatial effects [30]. A major drawback of ABC in the context of identifiability is that one must typically decide *a priori* which features of the model and data to match. Common applications of ABC for SDE models match the mean and variance of system [261] or the mean square error between simulations and data [322]. Estimating the likelihood directly, as particle MCMC does, is advantageous when assessing identifiability as it is not clear *a priori* which features of the data and model are significant. For example, some systems might contain the information required for identifiability in higher-order moments or auto-correlations between time-series observations. If ABC is used, a variant that preserves features of the model distribution might be desirable [323].

4.4.3 Modelling noise

In contrast to many studies of identifiability analysis for ODEs, we do not pre-specify parameters in the observation distribution. In a deterministic modelling framework, it is common to assume that all the variability in the data is uncorrelated and sourced from the observation process [27, 201, 324]. Therefore, for an ODE model, the observation parameters can be reliably estimated using the pooled sample variance. For inference on the birth-death ODE model (figure 4.4), we see that, because the observation variance must now also account for intrinsic noise, the identified value of σ_{err} is significantly larger than the target value. For an ODE model with additive homoscedastic Gaussian noise, the posterior mode (in the case of an improper uniform prior), maximum likelihood estimate and least-squares estimate are identical and are independent of the choice of the observed variance. For an SDE model, this is not the case as the intrinsic component of the noise is also modelled implicitly. Therefore, pre-specifying the observation variance could lead to biased estimates and obscure parameter identifiability. We account for this by treating the observation distribution variance as a nuisance parameter that we infer using MCMC, finding it to be identifiable in every case-study considered.

We have focussed our analysis on SDEs derived through the CLE, where the intrinsic noise can provide more information about the process. However, for large populations, the information contained in higher-order moments dissipates: to leading order, $\langle X^2 \rangle \to \langle X \rangle^2$ as $X \to \infty$. We see this in the epidemic model (figure 4.3c), where the variance is small compared to the scale of the mean. This loss of information in higher-order moments will not be detected by structural identifiability analysis of the moment equations, which is independent of the relative sizes of each moment. As populations become large, the information tends towards that obtained from the equivalent ODE system: this is the assumption behind many mean-field models. There are, however, many other models that contain sources of variability in their own right. For example, Mummert and Otunuga [216] study identifiability of an epidemic model where the infection

rate varies according to a white noise process. Other external effects, such as seasonal effects, are often incorporated into epidemic models [325, 326]. In other systems, extrinsic noise describing, for example, the environment, forms a core part of the process and is described by an SDE independent of the population size [53]. Grey-box models use a diffusion term to characterise uncertain physiological effects [218] that could obscure inference, rather than contain information. Making high-level assumptions about which noise process contains information can help with some of the computational challenges by formulating hybrid models containing a mixture of ODEs and SDEs. Particle MCMC carries across, trivially, to any Itô SDE, and the moment equation approach can be applied provided a system of moments be constructed. We have not considered identifiability of SDE models containing non-diffusion noise, such as coloured noise or jump noise. These models lend themselves to different inference techniques, such as forms of rejection sampling [327].

4.4.4 Approaches to computational challenges

The primary computational cost of working with SDE models stems from the need to simulate a suite of trajectories at each iteration of the particle MCMC algorithm. This cost increases not only with the dimensionality of the problem (as for deterministic models) but also with the amount of data, since the number of particles required for an unbiased likelihood estimate increases with the sample size [186, 257]. We see this, in particular, when conducting practical identifiability analysis of the two-pool and β-insulin-glucose models. These issues have important ramifications for identifiability, as it may not always be feasible to increase the amount of experimental data to rectify practical non-identifiability. Working with a surrogate model, such as a system of moment equations, can help alleviate some of these challenges. For example, establishing structural identifiability—which is requisite for parameter estimation [24]—indicates that the computational investment is worthwhile. Furthermore, such surrogate models can also form a computationally efficient alternative for assessing practical identifiability and performing inference [249, 310, 328], while still capturing more information than a purely deterministic description. The linear noise approximation (LNA) [329] captures information about the mean and variance of a process through a linear SDE. Although the LNA neglects higher order behaviour, a tractable likelihood function allows practical identifiability analysis directly through the Fisher information matrix (in the frequentist context) or through likelihood-based MCMC techniques (in the Bayesian context). As there is an extensive body of work analysing inference and identifiability using the LNA [186, 249, 330], we do not consider this approach here. However, conclusions of practical identifiability analysis drawn through MCMC are independent of whether the likelihood is calculated analytically (as through the LNA) or estimated through the a particle filter approach.

A large class of high-dimensional stochastic models lend themselves to structural identifiability analysis through moment equations. For example, CLE descriptions of multi-state ion-channel systems [226] and cascades with many bimolecular reactions [331], can be analysed in terms of a surrogate model using moment equations. This approach can be used because the propensity functions are often polynomial. However, the systems of moment equations are often infinite and require a moment closure approximation to facilitate this analysis. When a moment closure assumption is required, we find that newer software, such as `GenSSI2` in `Matlab`, is computationally advantageous over `DAISY` (table 4.3). Further, for larger systems, it may not be necessary to consider a full system of second-order moments: a closed system that neglects covariances is also a potentially useful surrogate model.

Many stochastic problems are both computationally expensive to assess using particle MCMC and may

not directly permit moment dynamics analysis. The β-insulin-glucose model comprises eight reactions and takes approximately 30 hours to perform 10,000 iterations of a pilot chain[2]. These issues are magnified by the quantity of data we consider in this example: five time-series of 15 observations each. Non-polynomial propensity functions mean that an exact expression for the moment equations cannot be derived, so it may not be possible to pre-detect structural non-identifiability through the moment dynamics approach. Fortunately, other approaches, such as those that use polynomial chaos [332], Gaussian processes [333], or the LNA [249] can provide alternative means of deriving surrogate models. This kind of approximation is already routine in the field of uncertainty quantification, which has deep connections to identifiability [334, 335]. In the future, many of these ideas could allow tractable structural and practical identifiability analysis of large systems of SDEs and, by extension, analysis of spatial problems described by stochastic partial differential equations (SPDEs).

The computational cost of MCMC, in particular for stochastic models with many parameters, has spurred the development of alternatives to explore and exploit the geometry of the likelihood near parameter estimates. The concept of *information geometry* [307, 336] generalises Fisher information and can be applied to detect identifiability through local information [337], and improve the performance of MCMC algorithms [338]. For SDE models in particular, variational Bayesian techniques provide an efficient alternative to MCMC for parameter estimation [339]. In many cases, mathematical models are calibrated to experimental data to establish the value of a biophysical parameter, not to fully parameterise a model. Profile likelihood [248, 340] is widely applied to assess identifiability in ODE models by maximising out parameters that are not of direct interest to reduce the dimensionality of the analysis. Since the bootstrap particle filter that we employ estimates the likelihood function, profile likelihood could be applied to SDE problems.

4.5 Conclusion

It is essential to consider identifiability when performing inference. Yet, there is a scarcity of methods available for assessing identifiability of the stochastic models that are becoming increasingly important. We have provided, through this chapter, an introduction to identifiability and a guide for performing identifiability analysis of SDE models in systems biology. By formulating a system of moment equations, we show how existing techniques for structural identifiability analysis of ODE models can be applied directly to SDE models [24, 26, 194, 195]. Through synthetic data and particle MCMC, we have demonstrated how to establish practical identifiability for SDE models from data [189, 210].
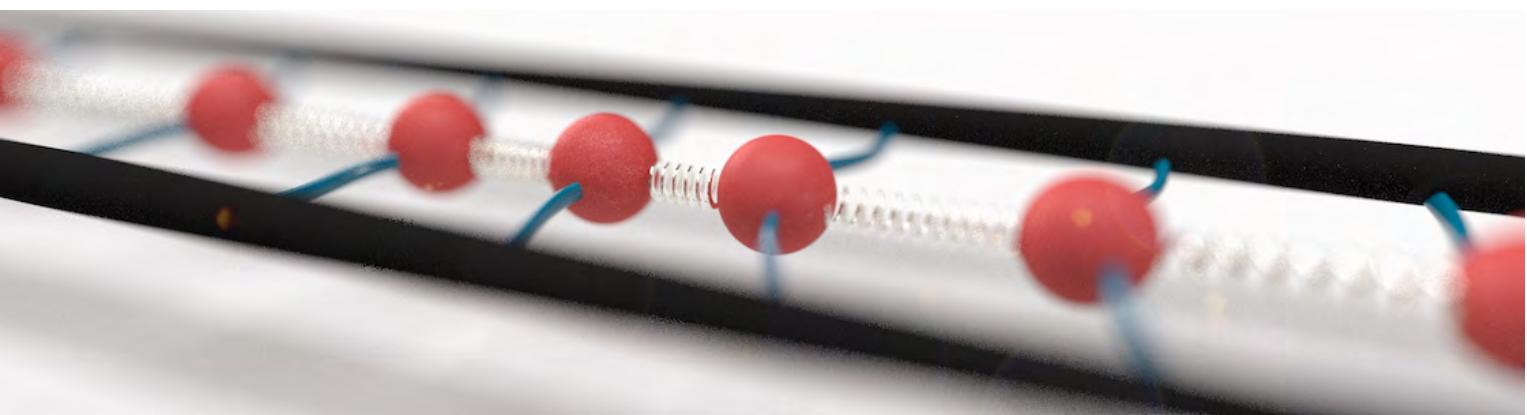
The analysis we demonstrate is critical for tailoring model complexity to the available data [24]. When a structurally identifiable model is found to be practically non-identifiable, identifiability analysis can guide experiment design to discern the quality and quantity of data required to estimate model parameters [341]. On the other hand, models found to be structurally non-identifiable should be re-parameterised, reduced in complexity, or changed [245, 342]. Moving from an ODE to an SDE model can often provide enough information to render an otherwise structurally non-identifiable parameter identifiable: we demonstrate this with the birth-death process. As increasing computing power facilitates inference of complex stochastic models, we expect identifiability to become ever more relevant.

---

[2]Runtimes for all results produced are available on Github at `https://github.com/ap-browning/SDE-Identifiability`

Chapter 5

Reversible signal transmission in an active mechanical metamaterial

Mechanical metamaterials are designed to enable unique functionalities, but are typically limited by an initial energy state and require an independent energy input to function repeatedly. Our study introduces a theoretical active mechanical metamaterial that incorporates a biological reaction mechanism to overcome this key limitation of passive metamaterials. Our material allows for reversible mechanical signal transmission, where energy is reintroduced by the biologically motivated reaction mechanism. By analysing a coarse-grained continuous analogue of the discrete model, we find that signals can be propagated through the material by a travelling wave. Analysis of the continuum model provides the region of the parameter space that allows signal transmission, and reveals similarities with the well-known FitzHugh–Nagumo system. We also find explicit formulae that approximate the effect of the time scale of the reaction mechanism on the signal transmission speed, which is essential for controlling the material.

Statement of Contribution of Co-Authors for Thesis by Published Paper

The authors listed below have certified that:

1. they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
2. they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
3. there are no other authors of the publication according to these criteria;
4. potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit, and
5. they agree to the use of the publication in the student's thesis and its publication on the QUT's ePrints site consistent with any limitations set by publisher requirements.

In the case of this chapter:
Reversible signal transmission in an active mechanical metamaterial
Published 2019 in Proceedings of the Royal Society A
doi:10.1098/rspa.2019.0146

Contributor	Statement of Contribution
Alexander P Browning 4 Aug 2021	Designed the study, developed and analysed the mathematical model, performed the travelling wave analysis, implemented the numerical algorithms, generated and interpreted results, drafted the manuscript, and revised the manuscript during peer-review.
Francis G Woodhouse	Designed the study, supervised the research, interpreted the results and provided comments on the manuscript.
Matthew J Simpson	Designed the study, supervised the research, interpreted the results, and provided comments on the manuscript.

Principal Supervisor Confirmation

I have sighted email or other correspondence from all Co-authors confirming their certifying authorship.

| Matthew J Simpson		9 August 2024
Name	Signature	Date

5.1 Introduction

Mechanical metamaterials are artificially constructed and have mechanical properties defined by their structure [47]. Simple metamaterials consist of a one, two, or three-dimensional array of elements connected by links [47, 343, 344] that may be elastic [7, 345–347], magnetic [348, 349] or electrostatic [345]. Mechanical metamaterials are highly tuneable [350–352] and by altering the structure of these elements, and the properties of the links, materials have been developed that selectively transmit signals [48, 49], behave as logic gates [7, 50] or buckle after the application of external stimulus [343]. There are many recent studies that experimentally realise simple mechanical metamaterials [49, 346, 353–356]. An advantage of these designs is that they are often well suited to utilise three-dimensional printing technology [7, 344, 353, 355, 357], however a common theme among existing metamaterials is that they generally require an external source of energy be provided in order to power their functions [7, 358]. Many existing technologies can be thought of as *static* or *inactive* in the sense that they are limited by a fixed initial energy state, and are only able to respond to a finite number of stimuli before the manual introduction of external energy.

Recent mathematical and experimental work examines properties of a class of one-dimensional bistable metamaterials [7, 345]. These systems comprise elements, each consisting of a mass connected to an external wall by a set of elastic elements that produce a bistable elastic potential. Individual elements are arranged in a one-dimensional lattice and interconnected by linear springs (figure 5.1a). These elements may be tuned so that the elastic potential energy function is asymmetric, resulting in both a high and low potential energy stable configuration for each element (figure 5.1b). The system can therefore be designed so that an external stimulus, for example the change of a single node from the high to low potential energy stable state, can trigger a change in element configuration through the entire lattice [7]. This change is the transmission of a mechanical signal powered by stored elastic potential energy.

A limitation of this mechanical regime is that the system must receive an external energy input before the transmission of an additional signal [7, 344]. Stiffness grading has been shown to overcome this limitation by exploiting a symmetric potential function [344], however these techniques may not allow the propagation of waves in systems with nonzero damping. The material in its current state is reset by manually moving each element back into its high potential energy configuration [7]. Our study introduces a theoretical biologically inspired mechanism that automatically resets each element to a high potential energy state, allowing the transmission of further signals. Many recent studies introduce the idea of manipulating biological subsystems in materials [359–361], or discuss behaviours that arise in *active matter* systems, where biological systems exert mechanical forces [362]. Some systems are biologically inspired [358, 360, 361] where properties of the metamaterial are designed to mimic a biological phenomenon, and some systems exploit the properties of biological subsystems to produce new behaviours in the material [359]. One possibility for our mechanism is to exploit actin filaments in eukaryotic cells [363–366] to convert energy provided to the cells as nutrients through chemical hydrolysis [51] into mechanical energy which can reset the bistable elements to a high potential energy configuration. The application of actin filaments in nanotechnology is well studied [365] and their exploitation in metamaterials has been previously suggested [51, 366].

The biological reaction mechanism we introduce is designed to react to changes in the displacement of individual elements and respond by inducing elastic potential energy back into each element. Mathematically, the effect of this process is to reset the potential function so that each element eventually reverts to a high

▲ Figure 5.1. (a) Schematic of the one-dimensional metamaterial [7]. Each element (red), of mass m, has a natural spacing of Δ. Each mass is connected to an external wall by a bistable elastic element (green) and to neighbouring elements by linear elastic springs (blue). The inset shows both stable states for each element, $u_i(t) = \pm\delta$ where i is the mass index. (b) An example of bistable potential function (red) and its derivative (blue) for $a_i = -0.5$ and $\delta = 1$ in the inactive model, or the active model at $t = 0$. (b–d) The effect of the reaction mechanism, which resets the potential function. Note that (d) corresponds to a reflection of (b) about $u_i = 0$. For each plot, the stable steady states (dashed grey) and unstable steady state (dotted grey) are shown. (e, f) Signal propagation through this material showing (e) the displacement of each element, $u_i(t)$, and (f) the biological reaction, $a_i(t)$. The signal was initiated by moving the first element from a displacement of $-\delta$ to δ at $t = 0$, and was retransmitted by moving the last element from a displacement of δ to $-\delta$ at $t = 400$. We note the original signal could also be initiated from the right boundary by moving the last element from a displacement of $-\delta$ to δ or indeed anywhere in the domain. Parameters used are $m = 1\,\mathrm{g}$, $k = 1\,\mathrm{g\,m\,s^{-2}}$, $\gamma = 1\,\mathrm{g\,s^{-2}}$, $\Delta = 0.002\,\mathrm{m}$, $\delta = 1\,\mathrm{m}$, $\varepsilon = 0.01\,\mathrm{s^{-1}}$, $\eta = 2$, $v = 1\,\mathrm{g\,m^{-2}\,s^{-2}}$ and $N = 101$ masses.

potential energy state, as shown in figure 5.1b-d. We present a mathematical characterisation of this reaction mechanism and explore how the timescale of the reaction mechanism affects the ability of the system to transmit signals. We find that signal transmission through a coarse-grained description of the material takes the form of a travelling wave. Using a travelling wave model, we find explicit formulae that bound the parameter space for which signal transmission can occur, and approximate the effect the timescale of the reaction mechanism has on the signal transmission speed. We lay the foundation for future work on this system where the metamaterial can be tuned to produce useful new behaviours. The results we provide quantify the trade-off between the signal transmission speed and the timescale of the biological response, which are essential for control and tuning of the material. For clarity, throughout this work we refer to the system without the reaction mechanism as the *inactive* system, and the system that includes the biological reaction mechanism we introduce as the *active* system.

In Section 5.2 we present a mathematical model that describes the discrete active mechanical system. Following this, we take a continuous limit of the discrete model [345] with which we qualitatively explore the effect of the reaction mechanism on the ability of the system to transmit mechanical signals. We find evidence of travelling waves in the continuous model, where the wavespeed corresponds to the signal transmission speed. In Section 5.3 we solve for the wavespeed and shape in the case where the reaction mechanism is excluded. This analysis is then extended to explore the effect that the reaction mechanism has on the wave speed and shape by taking a singular perturbation expansion (Section 5.4) and applying an energy conservation argument (Section 5.5). Finally, in Section 5.6 we discuss and summarise our results, and outline future work involving our active metamaterial.

5.2 Mathematical model

The metamaterial presented by Raney *et al.* [7] consists of N bistable elements of mass m, interconnected by $N - 1$ linear springs of stiffness k, and to two external walls by a pair of elastic elements, with a separation of Δ. A schematic of this physical system is shown in figure 5.1a. Denoting the displacement of the ith mass relative to the mean of its two steady states as $u_i(t)$, the state of the discrete system is governed by

$$
\begin{aligned}
0 &= m\frac{\mathrm{d}^2 u_i}{\mathrm{d}t^2} - k(u_{i+1} - u_i - \Delta) + \gamma\frac{\mathrm{d}u_i}{\mathrm{d}t} + \frac{\partial V}{\partial u_i}, & i &= 1, \\
0 &= m\frac{\mathrm{d}^2 u_i}{\mathrm{d}t^2} - k(u_{i+1} - 2u_i + u_{i-1}) + \gamma\frac{\mathrm{d}u_i}{\mathrm{d}t} + \frac{\partial V}{\partial u_i}, & i &= 2, ..., N-1, \\
0 &= m\frac{\mathrm{d}^2 u_i}{\mathrm{d}t^2} - k(u_i - u_{i-1} + \Delta) + \gamma\frac{\mathrm{d}u_i}{\mathrm{d}t} + \frac{\partial V}{\partial u_i}, & i &= N,
\end{aligned}
\tag{5.1}
$$

where γ is a damping parameter and $V(u_i, a_i)$ describes the potential energy of the bistable elements that connect each mass to the external wall [7, 345].

In this study, we choose $V(u_i, a_i)$ to be a quartic [345] defined in terms of its derivative,

$$
\frac{\partial V}{\partial u_i} = v(u_i^2 - \delta^2)(u_i - a_i),
\tag{5.2}
$$

where v describes the stiffness of the bistable elements and relates to the size of the energy gap between high and low potential energy configurations, $u_i = \pm\delta$ are the stable fixed points of $V(u_i, a_i)$ and $u_i = a_i \in [-\delta, \delta]$

is the unstable fixed point which governs the symmetry of $V(u_i, a_i)$ (figure 5.1b). For this choice of potential energy function, $u_i = \text{sign}(a_i)\delta$ corresponds to an element in the high potential energy configuration and $u_i = -\text{sign}(a_i)\delta$ corresponds to an element in the low potential energy configuration (figure 5.1b).

In our study we allow the symmetry of the potential function to vary in reaction to changes in the displacement, $u_i(t)$, by allowing $a_i = a_i(t)$ and enforcing

$$\frac{\mathrm{d}a_i}{\mathrm{d}t} = \epsilon\left(\frac{u_i}{\eta} - a_i\right). \tag{5.3}$$

Physically, a_i represents a biological subsystem that receives energy from external sources and induces it into the material at a rate proportional to ϵ. The parameter $\eta > 1$ determines the extrema of the reaction parameter so that $a_i(t) \in [-\delta/\eta, \delta/\eta]$. This implementation means that if a user transmits a signal through the material by changing the displacement of a node (figure 5.1e), and waits a period of time of $\mathbb{O}(\epsilon^{-1})$ after the signal reaches the end of the domain, all elements of the system will have reverted to their high potential energy states. This resetting process of $a_i(t)$, and by extension $V(u_i, a_i)$, is shown for a single element figure 5.1b-d, and throughout the material in figure 5.1f. The discrete system is similar to other fast-slow bistable systems, such as the FitzHugh-Nagumo model [367–369]. In this context, we consider that the displacement function, $u_i(t)$, undergoes an excitable excursion in phase space in response to an external stimulus, and the variable representing the biological response, $a_i(t)$, behaves as a linear recovery variable. In addition to the results in figure 5.1e, we reproduce results in the Appendix D in the case the second signal is initiated too early, so propagation can't occur.

If the length of the material is large relative to the separation of each mass, Δ, we can describe the material with a coarse grained continuous model [370]. To derive a continuous description of the system described by equations (5.1), we consider a material of fixed length, $L = (N-1)\Delta$, and take the limit $N \to \infty$ so that $\Delta \to 0$. Following this, we define field functions $u(x, t)$ and $a(x, t)$ that describe $u_i(t)$ and $a_i(t)$, respectively, for $x = (i-1)\Delta \in (0, L)$. When taking a continuous limit of the discrete system we require that the macroscopic quantities in the discrete model remain $\mathbb{O}(1)$ for physical reasons [345]. To do this, we replace unit quantities m, v and γ with density quantities, $\rho = m/\Delta$, $\hat{v} = v/\Delta$ and $\hat{\gamma} = \gamma/\Delta$, and scale the connecting spring force, $\hat{k} = \Delta k$.
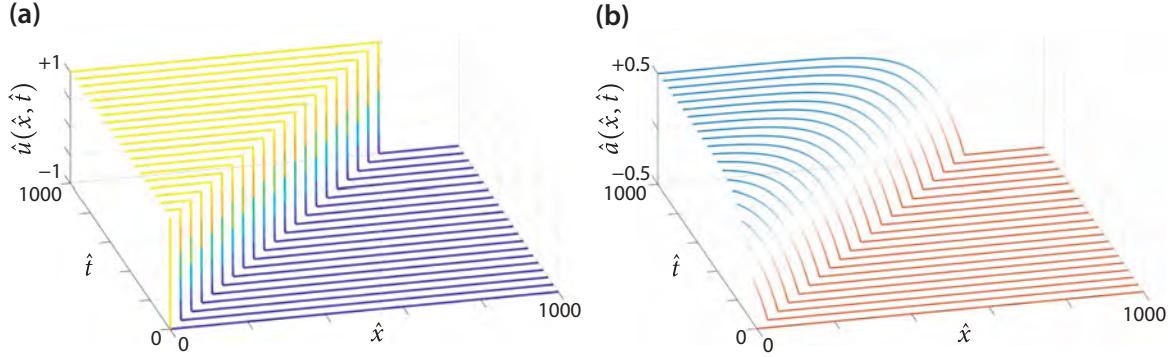
Dividing equation (5.1) by Δ and taking the limit $\Delta \to 0$ results in the continuous model,

$$0 = \rho\frac{\partial^2 u}{\partial t^2} - \hat{k}\frac{\partial^2 u}{\partial x^2} + \hat{\gamma}\frac{\partial u}{\partial t} + \hat{v}(u^2 - \delta)(u - a) = 0, \tag{5.4}$$

$$0 = \frac{\partial a}{\partial t} - \epsilon\left(\frac{u}{\eta} - a\right). \tag{5.5}$$

A no flux boundary condition $\partial u/\partial x = 0$ is applied at $x = 0$ and $x = L$.

A key aspect of this study is to investigate the signal transmission speed through the parameter space, particularly as ϵ increases. We therefore non-dimensionalise equations (5.4) and (5.5) by scaling $t = T\hat{t}$, $x = X\hat{x}$, $u = U\hat{u}$ and $a = A\hat{a}$, where hat notation represents dimensionless variables. Choosing $U = A = \delta$,

(a)

(b)

▲ Figure 5.2. Numerical solutions to the non-dimensional continuous model (equations (5.6) and (5.7)) showing (a) the displacement field, $\hat{u}(\hat{x}, \hat{t})$; and, (b) the biological reaction, $\hat{a}(\hat{x}, \hat{t})$. Parameters used are $v = 4$, $\eta = 2$ and $\kappa = 0.01$. The signal is initiated using the initial condition described by equations (5.8) and (5.9) with $Q = 1$.

$T^2 = \rho/\hat{\gamma}$ and $X^2 = \hat{k}T^2/\rho$, gives

$$0 = \frac{\partial^2 \hat{u}}{\partial \hat{t}^2} - \frac{\partial^2 \hat{u}}{\partial \hat{x}^2} + \frac{\partial \hat{u}}{\partial \hat{t}} + v(\hat{u}^2 - 1)\left(\frac{\hat{u}}{\eta} - \hat{a}\right), \tag{5.6}$$

$$0 = \frac{\partial \hat{a}}{\partial \hat{t}} - \kappa\left(\frac{\hat{u}}{\eta} - \hat{a}\right), \tag{5.7}$$

where $\kappa = T\epsilon \geq 0$, $v = \hat{\vartheta}\delta^2\rho/\hat{\gamma}^2 > 0$, $\eta > 1$ and $\hat{x} \in (0, \hat{L})$. The behaviour of the system can now be studied through the three-dimensional parameter space (v, η, κ) where: $v > 0$ is the relative strength of the potential function; $\eta > 1$ describes the steady state locations of $\hat{a}(\hat{x}, \hat{t})$; and $\kappa \geq 0$ is the relative timescale of the reaction mechanism. In this non-dimensional regime, the stable states are located at $\hat{u} = \pm 1$ where the high potential energy state is always given by sign$(\hat{a})$.

Previous studies provide evidence to suggest that the transmitted energy is input-independent [8], so we do not expect the initial condition to affect the transmission speed in the centre of the domain. A signal is initiated in the continuous model using a Heaviside initial condition for the displacement where

$$\hat{u}(\hat{x}, 0) = \begin{cases} 1, & 0 \leq \hat{x} < Q, \\ -1, & Q \leq \hat{x} \leq \hat{L}, \end{cases} \tag{5.8}$$

and the biological response is kept as it was before the signal was initiated by setting

$$\hat{a}(\hat{x}, 0) = -1/\eta. \tag{5.9}$$

In figure 5.2a numerical solutions to the continuous model show that the transition of the displacement variable, $\hat{u}(\hat{x}, \hat{t})$, is carried by a wave which appears to approach a constant shape and speed. figure 5.2b demonstrates the slow biological response where $\hat{a}(\hat{x}, \hat{t})$ also undergoes a slow transition in response to changes in $\hat{u}(\hat{x}, \hat{t})$. Results in figure 5.3 illustrate the dependence of the transmission speed on the timescale of the response, κ. These results indicate a negative monotonic relationship between κ and the transmission speed. Full details of the numerical scheme used to solve the continuous model are provided in Appendix D.

5.3 Travelling wave model

Figure 5.2 suggests that signals are propagated through the system by waves which appear to approach a constant shape and speed. Motivated by this, we now look for a travelling wave solution to the continuous model by extending the domain to represent a material of infinite length, so that $\hat{x} \in \mathbb{R}^+$ [36, 344, 345, 368, 369, 371–376]. We define the wavespeed, c, and without loss of generality enforce $c > 0$ by investigating travelling waves that are initiated at the left boundary and move in only the positive $\hat{x}$ direction. In reality, we expect symmetry in these solutions as a travelling wave may also travel in the negative $\hat{x}$ direction if the signal is initiated in the centre of the domain.
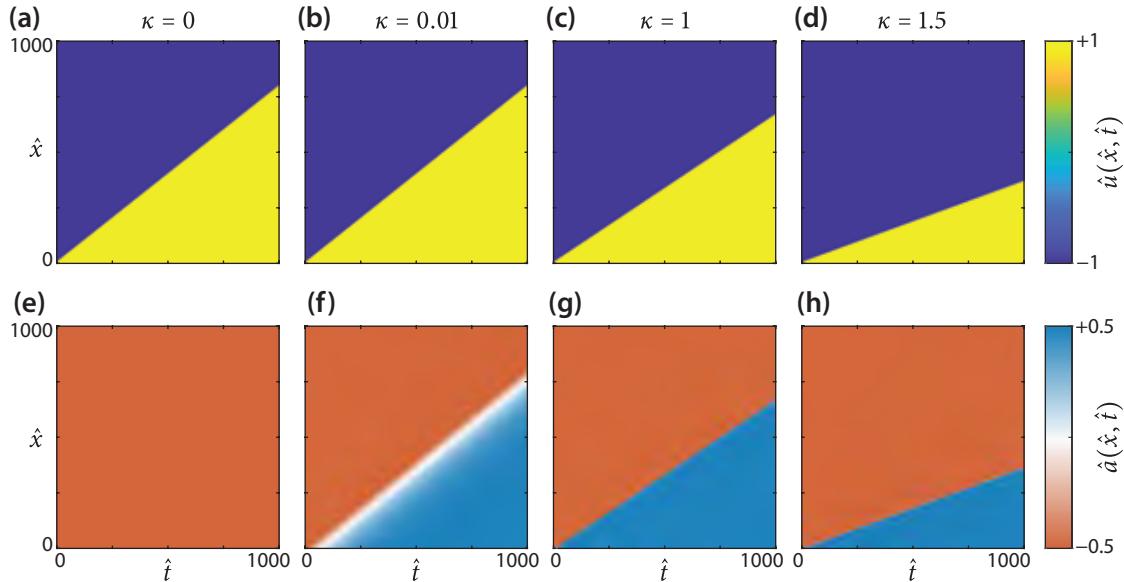
Substituting travelling wave variables $f(z) = \hat{u}(\hat{x}, \hat{t})$ and $h(z) = \hat{a}(\hat{x}, \hat{t})$, where $z = \hat{x} - c\hat{t} \in \mathbb{R}$, into equations (5.6) and (5.7), and dividing by $c^2 - 1$, gives the travelling wave model:

$$0 = \frac{\mathrm{d}^2 f}{\mathrm{d}z^2} + \frac{c}{1 - c^2} \frac{\mathrm{d}f}{\mathrm{d}z} - \frac{\nu}{1 - c^2}(f^2 - 1)(f - h), \quad f(\pm\infty) = \mp 1, \tag{5.10}$$
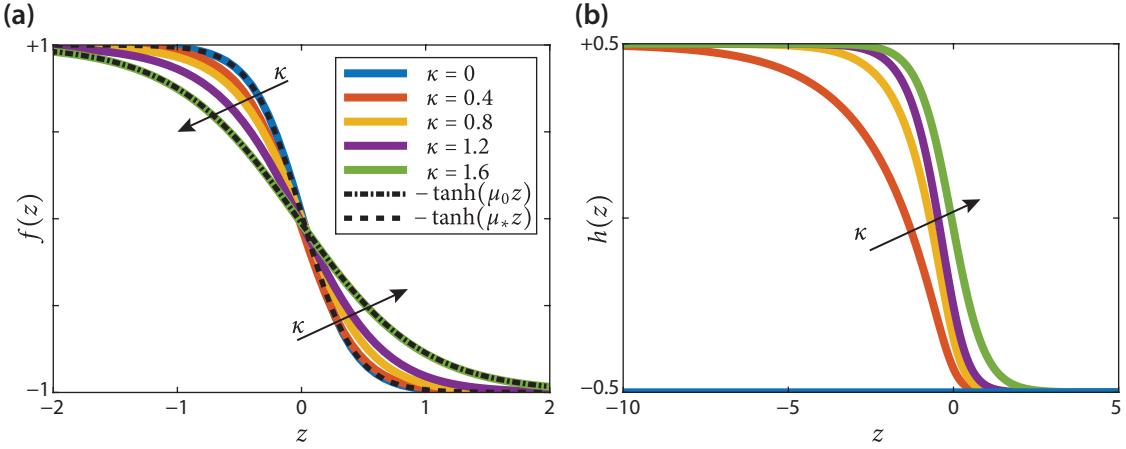
$$0 = \frac{\mathrm{d}h}{\mathrm{d}z} + \frac{\kappa}{c}\left(\frac{f}{\eta} - h\right), \quad h(\infty) = -\frac{1}{\eta}, \tag{5.11}$$

which, in physical coordinates, corresponds to a wave profile connecting $\hat{u} = -1$ to $\hat{u} = 1$ as $t \to \infty$.

In figure 5.3 numerical solutions of the continuous model show that increasing the response speed κ typically reduces the wavespeed from a maximum which occurs when $\kappa = 0$. We therefore denote a *fast wave* as a travelling wave solution at, or near, $\kappa = 0$, and a *slow wave* as a travelling wave solution in the limit, or near, $c = 0$. Removing the active component of the system from the model by setting $\kappa = 0$ (and defining c_0

▲ **Figure 5.3.** The solution to the continuous model, equation (5.6), shown as kymographs where colouring represents (a–d) the displacement function, $\hat{u}(\hat{x}, \hat{t})$; and, (e–h) the biological response, $\hat{a}(\hat{x}, \hat{t})$, for increasing values of κ. Parameter values used are $\nu = 4$ and $\eta = 2$. Each signal is initiated using the initial condition described by equations (5.8) and (5.9) with $Q = 1$.

▲ Figure 5.4. Approximate numerical solutions to the travelling wave model, formulated from numerical solutions to the continuous model, for (a) the displacement function, $f(z)$; and, (b) the biological reaction, $h(z)$. Results are shown for various values of the response parameter, κ, with $\nu = 4$ and $\eta = 2$. Additionally, (a) shows the exact solutions to the travelling wave model, equations (5.10) and (5.11), for $\kappa = 0$ (black dashed-dot) and $\kappa = \kappa_* = 1.6$ (black dashed).

as the corresponding wavespeed) gives:

$$0 = \frac{\mathrm{d}^2 f_0}{\mathrm{d}z^2} + \frac{c_0}{1 - c_0^2} \frac{\mathrm{d}f_0}{\mathrm{d}z} - \frac{\nu}{1 - c_0^2}(f_0^2 - 1)\left(f_0 + \frac{1}{\eta}\right), \quad f_0(\pm\infty) = \mp 1, \tag{5.12}$$

$$0 = \frac{\mathrm{d}h_0}{\mathrm{d}z}, \quad h_0(\infty) = -\frac{1}{\eta}, \tag{5.13}$$

where $f_0(z)$ will correspond to the solution to the fast wave. Under certain parameter transformations, equation (5.12) is analogous to the well-studied bistable equation, that arises from analysis of the FitzHugh-Nagumo model [367–369, 372]. The solution of equations (5.12) and (5.13) is therefore

$$f_0(z) = -\tanh(\mu_0 z), \tag{5.14}$$

$$h_0(z) = -\frac{1}{\eta}, \tag{5.15}$$
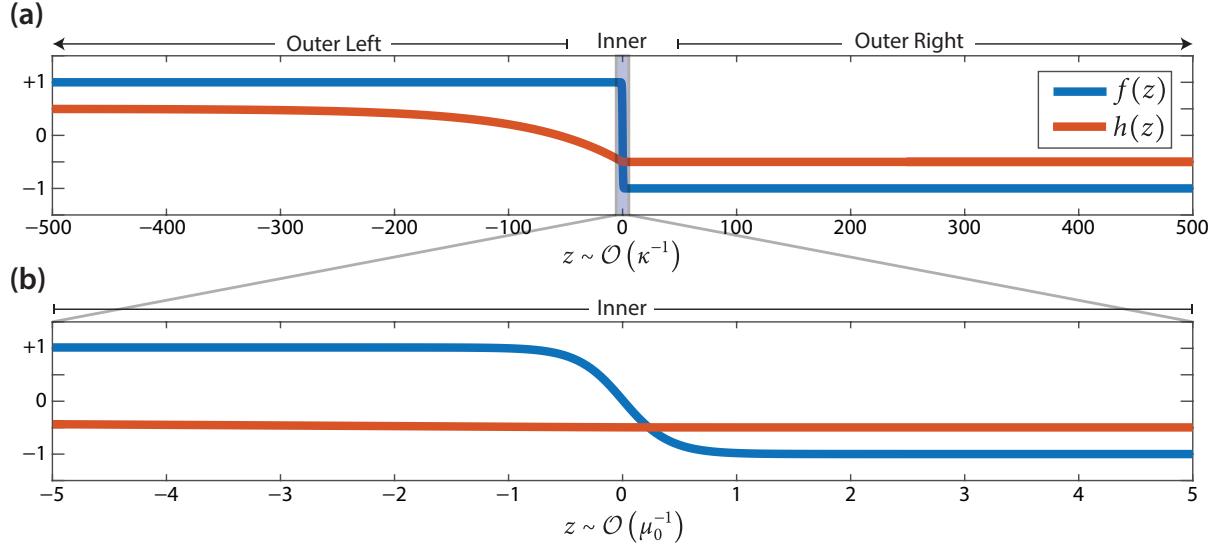
where

$$\mu_0 = \sqrt{\frac{\nu^2}{\eta^2} + \frac{\nu}{2}} \quad \text{and} \quad c_0 = \sqrt{\frac{1}{\frac{\eta^2}{2\nu} + 1}}. \tag{5.16}$$

For physically realistic parameter combinations, equation (5.16) suggests that $c_0 < 1$.

5.4 Perturbation solution for the fast wave

An exact solution in the limiting case $\kappa = 0$ (equation (5.14)) allows the formulation of a perturbation solution for $0 < \kappa \ll 1$ [372, 373, 377]. Figure 5.5 shows an approximation to the wave profiles $f(z)$ and $h(z)$ for $\kappa = 0.01$. A fast transition region is seen in $f(z)$ around $z = 0$ (figure 5.5a) suggesting behaviour necessitating a singular perturbation analysis [373, 377]. For $z \sim \mathcal{O}(\mu_0^{-1})$ the solution appears to match a solution for $\kappa = 0$, since, at this scale, $h(z)$ is approximately constant (figure 5.5b). In figure 5.5a we show

(a)

(b)

▲ Figure 5.5. Approximation of the travelling wave profile using a solution to the continuous model for (a) the displacement function, $f(z)$; and, (b) the biological reaction, $h(z)$. Parameters used are $v = 4$, $\eta = 2$ and $\kappa = 0.01$. In (a) the horizontal axis is scaled to show both outer regions where $z \sim \mathcal{O}(\kappa^{-1})$, which is the rate of change of $h(z)$. In (b) the horizontal axis is scaled to show the inner region where $z \sim \mathcal{O}(\mu_0^{-1})$, which is the rate of change of $f(z)$.

that $h(z)$ is not constant but rather a slow reaction of $\mathcal{O}(\kappa^{-1})$ which transitions $h(z) = -1/\eta$ to $h(z) = 1/\eta$ as $z \to -\infty$. This observation further suggests a singular perturbation analysis, since the behaviour of the response mechanism $h(z)$ varies significantly for $0 < \kappa \ll 1$ compared to $\kappa = 0$.

We propose a three-part perturbation solution about $\kappa = 0$ and define independent variables $z \sim \mathcal{O}(\mu_0^{-1})$ to correspond to an *inner region*, and $Z = \kappa z$ for $z \sim \mathcal{O}(\kappa^{-1})$, to correspond to two *outer regions*. These three regions are shown in figure 5.5a. This regime requires the fast process, which occurs in the inner region, to be much faster than the slow process, which occurs in the outer region. That is, we require $\kappa \ll \mu_0$.

Our aim in looking for a perturbation solution is to determine the effect of small perturbations in κ on c. To do this, we pose a perturbation expansion for the wavespeed through the entire domain,

$$c = c_0 + c_1 \kappa + \mathcal{O}(\kappa^2), \tag{5.17}$$

where c_0 is given by equation (5.16).

In the inner region, we pose a perturbation solution of the form

$$f(z) = f_0(z) + f_1(z)\kappa + \mathcal{O}(\kappa^2), \tag{5.18}$$
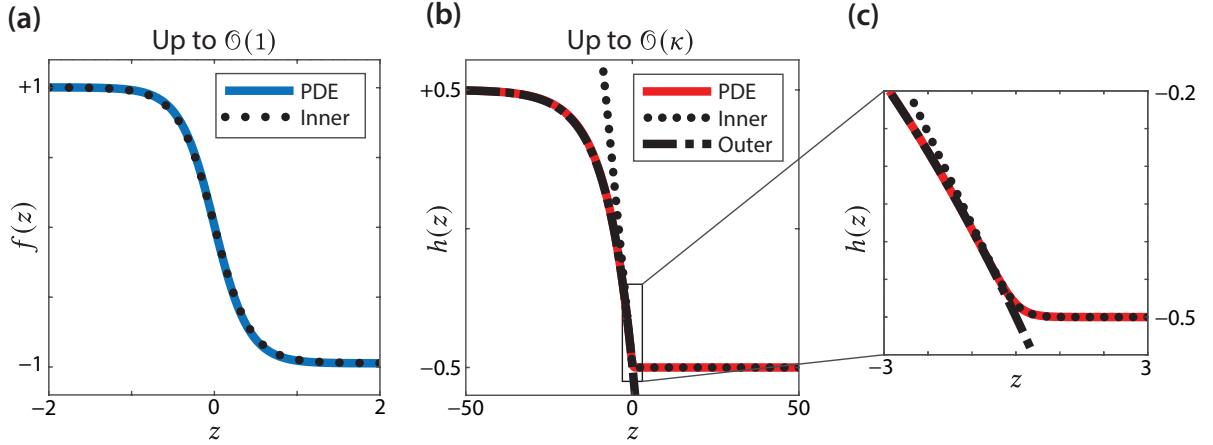
$$h(z) = h_0(z) + h_1(z)\kappa + \mathcal{O}(\kappa^2), \tag{5.19}$$

where $f_0(z)$ and $h_0(z)$ correspond to the shape of the fast wave at $\kappa = 0$, given by equation (5.14). Full details of the perturbation solution in the inner region are given in Appendix D. In summary, the solution is given by

$$f(z) = -\tanh(\mu_0 z) + f_1(z)\kappa + \mathcal{O}(\kappa^2), \tag{5.20}$$

$$h(z) = -\frac{1}{\eta} + \frac{1}{c_0\mu_0\eta}\left(\log(\cosh(\mu_0 z)) - \mu_0 z + \log(2)\right)\kappa + \mathcal{O}(\kappa^2), \tag{5.21}$$

where $f_1(z)$ and c_1 are defined by the solution of a second-order boundary value problem, which can be

▲ Figure 5.6. Comparison of the wave shape from a numerical solution to continuous model (coloured solid curves), to the perturbation solution (black dotted and dotted-dashed curves) for (a) the displacement function, $f(z)$, to $\mathcal{O}(1)$ and (b) the reaction mechanism, $h(z)$, to $\mathcal{O}(\kappa)$. The inset (c) shows the crossover between the inner and outer perturbation solutions to $h(z)$ in further detail, showing an excellent match to the wave shape from the continuous model, and between inner and outer solutions.

solved numerically. Full details of this numerical scheme are given in Appendix D. These solutions are shown in figure 5.6, and the wavespeed correction, c_1, is summarised for various parameter combinations in table D.2 in Appendix D.

In the outer region we denote solutions to the slow system using uppercase variables, $F(Z)$ and $H(Z)$. We expect the outer solution to apply for $Z \sim \mathcal{O}(1) \Rightarrow z \sim \mathcal{O}(\kappa^{-1})$ and as $z, Z \to \infty$. In the outer region, equation (5.10) becomes

$$0 = \kappa^2 \frac{d^2 F}{dZ^2} + \frac{c\kappa}{1-c^2} \frac{dF}{dZ} - \frac{\nu}{1-c^2}(F^2 - 1)(F - H), \quad F(\pm\infty) = \mp 1, \tag{5.22}$$

which has the solution

$$F(Z) = -\text{sign}(Z). \tag{5.23}$$

This agrees with the sharp transition region seen at this scale in figure 5.5a. Substituting equation (5.23) into equation (5.11) we see that

$$0 = \frac{dH}{dZ} - \frac{1}{c}\left(\frac{\text{sign}(Z)}{\eta} + H\right), \quad H(\infty) = -\frac{1}{\eta}. \tag{5.24}$$

Full details of perturbation solution in the outer region are given in Appendix D. In summary, the solution to equation (5.24) is given by

$$H(Z) = \begin{cases} \dfrac{1}{\eta} - \dfrac{2}{\eta}\exp\left(\dfrac{Z}{c_0}\right) - \dfrac{2c_1}{\eta c_0^2}Z\exp\left(\dfrac{Z}{c_0}\right)\kappa + \mathcal{O}(\kappa^2), & Z < 0, \\ -\dfrac{1}{\eta}, & Z > 0. \end{cases} \tag{5.25}$$

Figure 5.6b-c shows a comparison between solutions for the reaction mechanism in the inner and outer regions (given by equations (5.20) and (5.25), respectively), to an approximation of the wave shape formed form the numerical solution to the continuous model. We see an excellent match between both the continuous

model and the perturbation solutions, as well as between the inner and outer regions of the perturbation solution. These results provide excellent information about the behaviour of $h(z)$ for $0 < \kappa \ll 1$, and are particularly important as the behaviour for $0 < \kappa \ll 1$ varies significantly from the behaviour at $\kappa = 0$.

5.5 Energy transport

To determine information about the slow wave, and to obtain more information about the fast wave, we follow Nadkarni *et al.* [345, 375] to derive an integrability condition – that is, a necessary condition for existence of a solution with given boundary conditions – to investigate the transported energy. Multiplying the travelling wave model, given by equation (5.10), by $\mathrm{d}f/\mathrm{d}z$ and integrating gives

$$0 = \int_0^\infty \frac{\mathrm{d}^2 f}{\mathrm{d}z^2} \frac{\mathrm{d}f}{\mathrm{d}z}\, \mathrm{d}z + \frac{c}{1-c^2} \int_0^\infty \left(\frac{\mathrm{d}f}{\mathrm{d}z}\right)^2 \mathrm{d}z - \frac{v}{1-c^2} \int_0^\infty (f^2 - 1)(f - h)\frac{\mathrm{d}f}{\mathrm{d}z}\, \mathrm{d}z. \tag{5.26}$$

For a parameter regime where the transition wave exists, the velocity will vanish in the far field so that $f \to \mp 1$ and $\mathrm{d}f/\mathrm{d}z \to 0$ as $z \to \pm\infty$. Therefore some components of equation (5.26) vanish:

$$\int_0^\infty \frac{\mathrm{d}^2 f}{\mathrm{d}z^2}\frac{\mathrm{d}f}{\mathrm{d}z}\, \mathrm{d}z = \left[\frac{1}{2}\left(\frac{\mathrm{d}f}{\mathrm{d}z}\right)^2\right]_{-\infty}^{\infty} = 0,$$

$$\text{and } \int_0^\infty (f^2 - 1)(f - h)\frac{\mathrm{d}f}{\mathrm{d}z}\, \mathrm{d}z = -\int_0^\infty \frac{\mathrm{d}f}{\mathrm{d}z} h\, \mathrm{d}z.$$

Under the assumption $|c| \neq 1$, equation (5.26) becomes an *integrability condition*,

$$c\int_0^\infty \left(\frac{\mathrm{d}f}{\mathrm{d}z}\right)^2 \mathrm{d}z = -v \int_0^\infty \frac{\mathrm{d}f}{\mathrm{d}z} h\, \mathrm{d}z. \tag{5.27}$$

It is useful to note that, for $f = -\tanh(\sigma z)$, for some σ, which occurs as $\kappa \to 0$ with $\sigma = \mu_0$,

$$(f^2 - 1)\frac{\mathrm{d}f}{\mathrm{d}z} h \propto \mathrm{sech}^4(\sigma z) h. \tag{5.28}$$

This suggests that only the component of $h(z)$ near $z = 0$ is important in gaining any approximation from the integrability condition, provided $f(z)$ has the form of a hyperbolic tangent function. Since a travelling wave connecting $f(z) = 1$ to $f(z) = -1$ as $z \to \infty$ will always have a sigmoidal form (figure 5.4), we expect this observation to apply for all regions of the parameter space where a travelling wave exists.

In the inactive model, where $\kappa = 0$, Nadkarni *et al.* [345] show that the integrability condition reduces to

$$E_k = \frac{c\Delta V}{2\gamma}, \tag{5.29}$$

where E_k represents the total kinetic energy per density transported by the transition wave, and ΔV represents an *energy gap* or difference the potential energy between the high and the low potential energy states. This result can be used to find an upper bound for c for $\kappa > 0$: Since $|h(t)| \leq 1/\eta$, and ΔV is a monotonically decreasing function of $h(t)$, the available kinetic energy in the system is always bounded above by that which occurs when $h(t) \equiv 1/\eta$, which occurs for $\kappa = 0$. This suggests that $c(\kappa) \leq c_0 < 1$ and substantiates numerical evidence seen in figure 5.3 which suggests a decreasing monotonic relationship

between c and κ.

In the following Sections we apply the integrability condition to obtain approximations to $c(\kappa)$ while holding v and η constant. We also find the region of the parameter space that allows signal transmission. The advantage of these approximations is that they avoid numerical solutions to the continuous model to approximate the wavespeed. Numerical solutions to this model are not computationally inexpensive and it is generally difficult to obtain and verify the results.

5.5.1 Energy transport in the fast wave

For $\kappa \ll 1$ it is reasonable to assume that $f \approx -\tanh(\mu z)$ where μ depends on κ, and $\mu \to \mu_0$ as $\kappa \to 0$. By assuming $f(z)$ has a similar form to $f_0(z)$ for $\kappa \ll 1$, it is reasonable to use the perturbation solution (equation (5.21)) as an approximation for $h(z)$. Since the integrability condition (equation (5.27)) depends only on the component of $h(z)$ near $z = 0$, we use only equation (5.21). Allowing $\mu_0 = \mu(0)$ and $c_0 = c(0)$ where $\mu = \mu(\kappa)$ and $c = c(\kappa)$ depend on κ, we assume that

$$h(z) \approx -\frac{1}{\eta} + \frac{\kappa}{c\mu\eta}\Big(\log(\cosh(\mu z)) - \mu z + \log(2)\Big), \quad \kappa \ll 1. \tag{5.30}$$

Equation (5.30) corresponds to the *smoothed* piecewise-defined function where growth is equal to $2\kappa/(c\eta)$ for $z < 0$, and zero for $z > 0$. This function corresponds exactly to an $\mathscr{O}(1)$ approximation to $h(z)$ which uses $f(z) = -\text{sign}(z)$. Substituting equation (5.30) into the integrability condition (equation (5.27)) provides the relationship

$$\frac{4c^2\mu^2}{3} = \frac{2v(6c\mu - 5\kappa)}{9\eta}. \tag{5.31}$$

This result does not allow c and μ to be determined independently, so we consider a far field expansion of the travelling wave [371], that is, an expansion about $\phi \approx 0$, where $\phi = e^{-\mu z}$:

$$f(z) = -\tanh(\mu z) = \frac{e^{-2\mu z} - 1}{e^{-2\mu z} + 1} = \frac{\phi^2 - 1}{\phi^2 + 1} = -1 + 2\phi^2 + \mathscr{O}(\phi^4). \tag{5.32}$$

Substituting $\phi = e^{-\mu z}$ and equation (5.32) into the travelling wave equation for $h(z)$ (equation (5.11)) gives

$$\begin{aligned}
f(z) &\sim -1 + 2e^{-2\mu z}, \quad |z^{-1}| \ll 1, \\
h(z) &\sim -\frac{1}{\eta} + \frac{2\kappa}{\eta(\kappa + 2c\mu)}e^{-2\mu z}, \quad |z^{-1}| \ll 1.
\end{aligned} \tag{5.33}$$

Substituting equation (5.33) into equation (5.10) provides a far-field relationship,

$$0 = \frac{v}{\eta} - c\mu - v + 2\mu^2(1 - c^2). \tag{5.34}$$

To obtain a useful analytical expression for c and μ, we pose a perturbation solution to equations (5.31) and (5.34) around $\kappa = 0$, such that

$$c = c_0 + \tilde{c}_1\kappa + \mathscr{O}(\kappa^2), \tag{5.35}$$

$$\mu = \mu_0 + \tilde{\mu}_1\kappa + \mathscr{O}(\kappa^2), \tag{5.36}$$

where c_0 and μ_0 are given by equation (5.16). We note that $\tilde{c}_1$ approximates c_1, the gradient of $c(\kappa)$ at $\kappa = 0$, where c_1 is defined exactly by the solution of the perturbation problem. Substitution into equations (5.31) and (5.34) gives

$$\tilde{c}_1 = \frac{5\nu\left(4c_0^2\mu_0 + c_0 - 4\mu_0\right)}{24\mu_0^2\left(2c_0\eta\mu_0 - \nu\right)}. \tag{5.37}$$

We compare this estimate of c_1 from that calculated numerically by solving the boundary value problem that comes from the singular perturbation expansion, and that estimated from the continuous model, in table D.2 (Appendix D). In figure 5.7, we show that, for $\kappa \ll 1$,

$$c(\kappa) \sim c_0 + \tilde{c}_1\kappa, \tag{5.38}$$

matches numerical results for $c(\kappa)$.

5.5.2 Energy transport in the slow wave

We denote $f_*(z)$ and $h_*(z)$ as the shape of the slow wave, which occurs as $c \to 0$, and $\kappa \to \kappa_*$, where κ_* is yet to be determined. In addition, we expect the curve $c(\kappa)$ to be perpendicular to the κ axis at κ_* to maintain continuity in the symmetry of the problem where solutions with a negative wavespeed are equally valid. To determine a governing equation for the slow wave, we take $c \to 0$, so that equations (5.10) and (5.11) become

$$0 = \frac{d^2 f_*}{dz^2} - \nu\left(1 - \frac{1}{\eta}\right)f_*(f_*^2 - 1), \quad f_*(\pm\infty) = \mp 1, \tag{5.39}$$

$$0 = \frac{f_*}{\eta} - h_*, \tag{5.40}$$

which have the analytical solution

$$f_*(z) = -\tanh\left(\mu_* z\right),$$
$$h_*(z) = -\frac{1}{\eta}\tanh\left(\mu_* z\right), \tag{5.41}$$

where

$$\mu_* = \sqrt{\frac{\nu}{2}\left(1 - \frac{1}{\eta}\right)}. \tag{5.42}$$
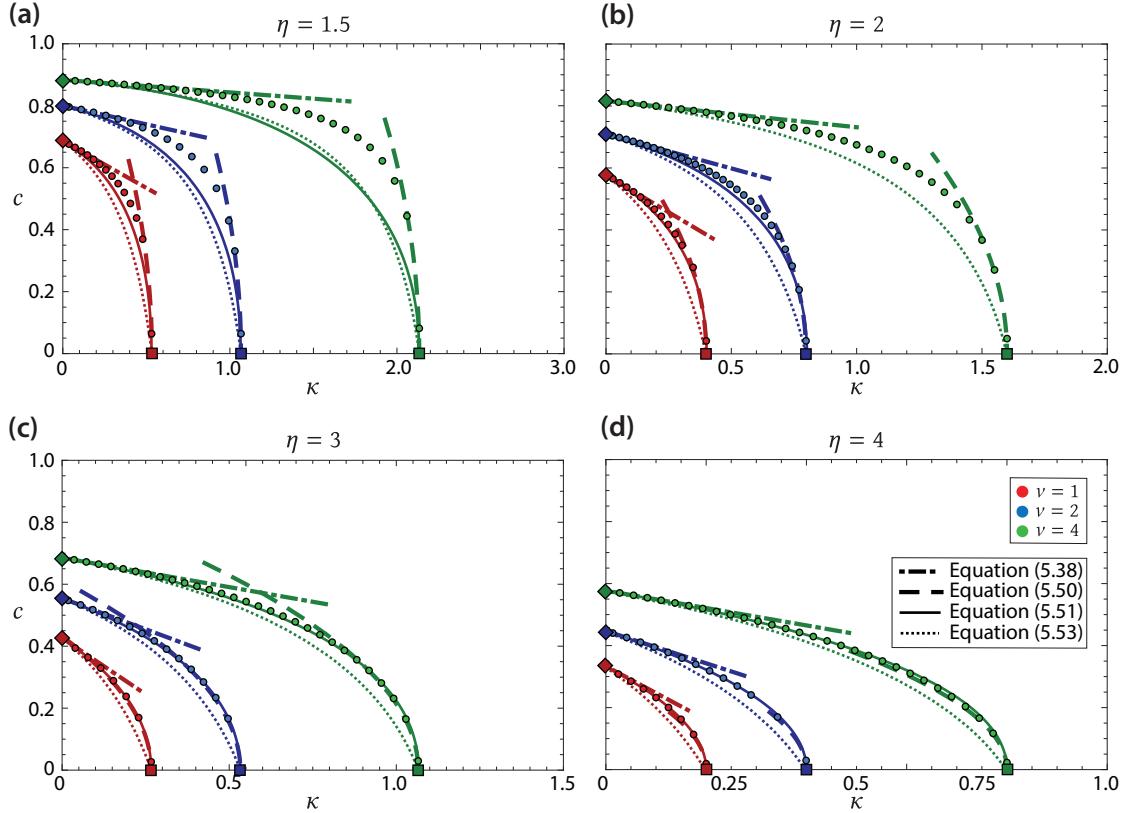
Direct substitution of $f_*(z)$ and $h_*(z)$ into the integrability condition (equation (5.27)) causes all terms to vanish, so higher-order behaviour of $\mathcal{O}(c)$ as $c \to 0$ is important. To allow for this, we note that equation (5.11) can be solved, for all κ, to give

$$h(z) = \frac{\kappa}{c\eta}\int_z^\infty f(s)\exp\left[\frac{\kappa}{c}(z - s)\right]\,ds. \tag{5.43}$$

Equation (5.43) shows that $h(z)$ depends the product of $f(s)$ and a function that decays rapidly as z moves away from s. Expanding $f(s)$ in a Taylor series about z gives

$$h(z) \approx \frac{\kappa}{c\eta}\int_z^\infty\left[\sum_{n=0}^\infty\frac{(s-z)^n}{n!}\frac{d^n f}{dz^n}\right]\exp\left(\frac{\kappa}{c}(z-s)\right)\,ds = \frac{1}{\eta}\sum_{n=0}^\infty\frac{c^n}{\kappa^n}\frac{d^n f}{dz^n}. \tag{5.44}$$

Assuming that $f \sim \tanh(\mu(\kappa)z)$ provided $c \ll 1$, where $\mu(\kappa_*) = \mu_*$, and truncating the infinite series

▲ Figure 5.7. Comparison between the wavespeed as a function of κ, from the continuous model (circles) to: the analytical expression at $\kappa = 0$, given by equation (5.16) (diamonds); the analytical expression for κ_* at $c = 0$, given by equation (5.46) (squares); the analytical approximation given by equation (5.38), which applies for $\kappa \ll 1$ (dashed-dotted curve); the analytical approximation given by equation (5.50), which applies for $c \ll 1$ (dashed-curve); the combined approximation given by equation (5.51) which applies everywhere (dotted curve); and, the approximation given by equation (5.53) which applies everywhere (solid curve).

given by equation (5.44) after $n = 3$, the integrability condition (equation (5.27)) gives the relationship

$$\frac{4\kappa^3}{3} = \frac{16\nu \left(7\kappa^2 - 8c^2\mu^2\right)}{105\eta}.$$ (5.45)

For η and ν fixed, $c \to 0$ only as $\kappa \to \kappa_*$. Substituting $c = 0$ into equation (5.45) gives

$$\kappa_* = \frac{4}{5}\frac{\nu}{\eta}.$$ (5.46)

Under the assumption that $c(\kappa)$ is monotonically decreasing, the result in equation (5.46) provides an analytical expression for the region of the parameter space where we expect signal transmission. In figure 5.7 we show that this expression matches the numerical results. Interestingly, equation (5.46) depends only on the ratio of the other parameters, ν/η. The scales of the horizontal axes for figure pairs figure 5.7a,c and figure 5.7b,d have been chosen to highlight this.

To gain information about the shape of $c(\kappa)$ near $c = 0$, we pose a perturbation solution around $\kappa = \kappa_*$ to

the system governed by equation (5.45) and the far-field relationship, equation (5.34), where

$$c = c_* + c_{\frac{1}{2}} \sqrt{\kappa_* - \kappa} + \mathcal{O}\left(\kappa_* - \kappa\right), \tag{5.47}$$

$$\mu = \mu_* + \mu_{\frac{1}{2}} \sqrt{\kappa_* - \kappa} + \mathcal{O}\left(\kappa_* - \kappa\right), \tag{5.48}$$

where μ_*, given by equation (5.42), and $c_* = 0$ apply at $\kappa = \kappa_*$. Substitution of these expansions into equations (5.34) and (5.45) gives

$$c_{\frac{1}{2}} = \sqrt{\frac{7}{5(\eta - 1)}}, \tag{5.49}$$

so that, for $\kappa \approx \kappa_*$,

$$c(\kappa) \sim \sqrt{\frac{7(\kappa_* - \kappa)}{5(\eta - 1)}}. \tag{5.50}$$

Figure 5.7 shows that equation (5.50) matches numerical results for $c(\kappa)$ for a surprisingly wide range of $\kappa < \kappa_*$, especially for larger η. This result is particularly important as we have not constructed a full perturbation solution to the travelling wave model about $c = 0$.

5.5.3 Combined approximation

We can combine the approximations to $c(\kappa)$ from the slow and the fast wave to obtain a curve that behaves like equation (5.38) for $\kappa \to 0$ and like equation (5.50) as $\kappa \to \kappa_*$. To do this, we propose a form like

$$c(\kappa) \approx \alpha_1 + \alpha_2 \kappa + \alpha_3 \sqrt{\kappa_* - \kappa}, \tag{5.51}$$

where we choose

$$\alpha_1 = -(c_0 + 2\tilde{c}_1 \kappa_*), \quad \alpha_2 = \frac{c_0}{\kappa_*} + 2\tilde{c}_1, \quad \alpha_3 = \frac{2(c_0 + \tilde{c}_1 \kappa_*)}{\sqrt{\kappa_*}} \tag{5.52}$$

where c_0 is given by equation (5.16); $\tilde{c}_1$ is given by equation (5.37); and, κ_* is given by equation (5.46). In figure 5.7 we show that this combined approximation provides a reasonable approximation to $c(\kappa)$ for $\kappa \in [0, \kappa_*]$.

5.5.4 Whole domain ansatz

Results in equations (5.14) and (5.41) show that $f(z)$ is described exactly by a hyperbolic tangent at both $\kappa = 0$ and $\kappa = \kappa_*$, and figure 5.4a suggests that $f(z)$ remains sigmoidal. Therefore, it may be reasonable to approximate $f(z) \approx \tanh(\mu(\kappa)z)$ for $\kappa \in [0, \kappa_*]$, where $\mu(\kappa)$ depends on κ. Additionally assuming that $\mu(\kappa)$ decays linearly from $\mu = \mu_0$ at $\kappa = 0$ to $\mu = \mu_*$ at $\kappa = \kappa_*$, we obtain an approximation to $f(z)$ for all κ:

$$f(z) \sim -\tanh\left(\mu(\kappa)z\right), \quad \mu(\kappa) = \frac{\mu_* - \mu_0}{\kappa_*}\kappa + \mu_0. \tag{5.53}$$

Substituting the approximation for $\mu(\kappa)$ given by equation (5.53) into the far-field matching condition (equation (5.34)) gives an approximation to $c(\kappa)$ which can apply for $\kappa \in [0, \kappa_*]$. We show that this approximation is reasonably accurate throughout $\kappa \in [0, \kappa_*]$ in figure 5.7, however it is clear from numerical results in figure 5.4, when $\kappa = 0.4$, that the solution does not have the symmetry of a hyperbolic tangent function.

5.6 Discussion and Conclusion

Currently, the inactive metamaterial described mathematically by Nadkarni *et al.* [345] and experimentally realised by Raney *et al.* [7] is able to transmit mechanical signals by the release of stored potential energy. A limitation of this design is that mechanical energy must be manually introduced into the system before additional signals can be transmitted. Our study presents a novel biologically inspired metamaterial that incorporates a theoretical biological mechanism that harvests energy to reset the system to a high potential energy state, allowing the transmission of additional signals. Energy may be induced into the active metamaterial through a biological process, such as actin filaments in eukaryotic cells [363–365]. That said, our analysis does not necessarily require this mechanism to have a biological origin: the reaction mechanism may also represent a mechanical system where energy is added through other electrochemical [378], photovoltaic, thermodynamic [357] or pneumatic [358] subsystems.

By finding evidence of travelling wave solutions, we are able to analyse limiting behaviour describing the signal transmission speed and wave shape. We provide a detailed analysis to qualitatively and quantitatively understand the effect of our reaction mechanism on signal transmission abilities of the material. Our main results consist of a set of analytical approximations that quantify the signal transmission speed as a function of the parameters which describe the physical properties of the material. Results in figure 5.7 show that the approximation we develop to apply through the whole domain, given by equation (5.51), provides an excellent match to the numerical results, particularly for large η. In addition, our approximation for the wavespeed near the slow wave, given by equation (5.50), is able to provide excellent information about the shape of $c(\kappa)$ as $c \to 0$, which we find is difficult to obtain numerically. This approximation is also able to provide a region of the parameter space for which signal transmission can occur, given by $\kappa < 4\nu/(5\eta)$ (equation (5.46)). This understanding of the effect of our mechanism on the signal transmission speed is useful as it allows our active metamaterial to be tuned to produce desirable new behaviours. For example, our results allow quantification of the trade-off between signal transmission speed and the response time, which is essential for controlling the material. These insights are also essential for building a material containing a biological mechanism that induces energy into the system. Decreasing ν and η in the same proportion increases the transmission speed at the cost of increased sensitivity to noise-induced misfiring, but may be essential if the energy budget is small.

A key aspect of our study is to follow Nadkarni *et al.* [345] by representing the bistable potential energy function as a quartic (equation (5.2)). This approach leads us to obtain numerous analytical approximations that characterise the effect of the biological mechanism on the transmission speed which, although qualitatively reliable, may not always be quantitatively appropriate for particular systems [7]. In fact, the analytical expression for the transmission speed is a result of the similarity between our model and the well-studied bistable equation [372]. These choices mean that our system has mechanical and algebraic properties that are similar to other bistable systems, such as the FitzHugh-Nagumo model [367, 369]. That said, we do not assume that the timescale of the response is significantly slower than the timescale of the excitement, as is often the case in analysis of such models. Indeed, our aim is to develop an intelligent biomechanical material that has tuneable properties. In some sense, it is desirable that the response is as fast as possible to allow for a short period of time between signal reception and retransmission. Future work may examine the role of heterogeneities in the properties of the material [86, 344]. Such features could allow the material to

selectively transmit signals by creating energy barriers that interact with signals of certain properties [48] .
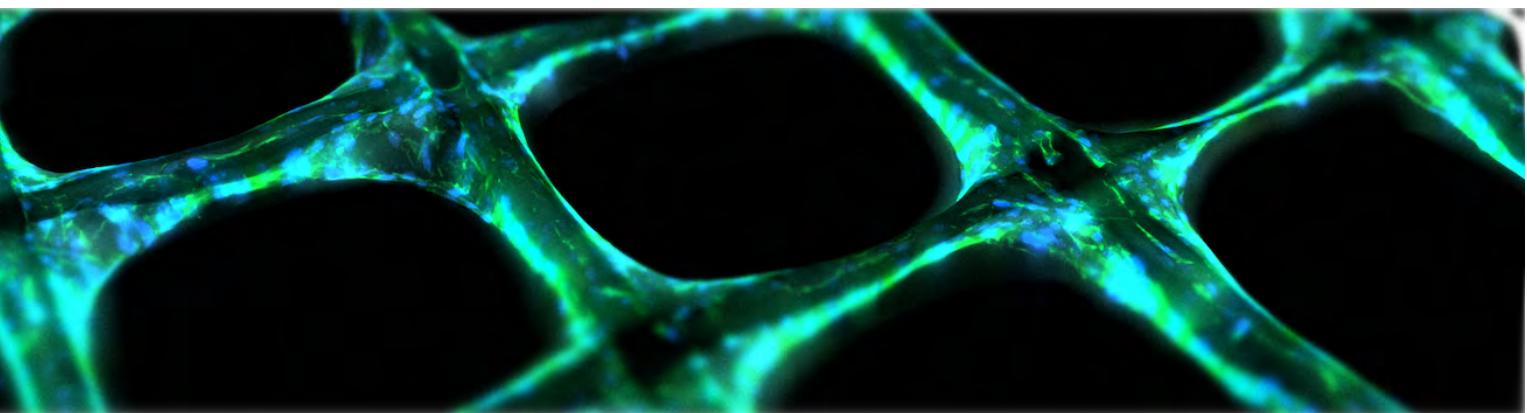
The travelling wave analysis we conduct assumes a material of infinite length over a large period of time. However, applications of our material will a have finite length and may have properties not suitable for a continuum model. For example, in a material where the spacing between elements is not significantly different to the length of the material, a discrete travelling wave analysis may be more appropriate. The discrete problem is known to be substantially more difficult that the continuous problem [372], so the limiting transmission speed our analysis provides may still be useful. Furthermore, the inclusion of our biologically inspired mechanism can be incorporated into passive metamaterials of higher dimensions to enable new behaviours and the travelling wave analysis can be extended to investigate two-dimensional signal propagation. In Appendix D, we produce results which show the transmission of concurrent signals (figures D.1 and D.2) and interacting signals initiated from both ends of the material (figure D.3). Further analysis is needed to examine the behaviour of these types of interacting waves [379] and the material's ability transmit oscillatory or concurrent signals.

To conclude, we have presented a novel, biologically inspired, active metamaterial that can reversibility transmit mechanical signals. This work provides an analytical expression that describes the mechanical properties of the material required for signal transmission. We also provide numerous approximations that quantify the effect of the mechanical properties, and the timescale of the biological response, on the transmission speed. This work demonstrates how a new class of biologically inspired metamaterials are able to produce useful new functionalities. The type of analysis we present is invaluable for tuning and controlling the active metamaterial.

Chapter 6

Model-based data analysis of tissue growth in thin 3D printed scaffolds

Published 2021 in *Journal of Theoretical Biology*.

Tissue growth in three-dimensional (3D) printed scaffolds enables exploration and control of cell behaviour in more biologically realistic geometries than that allowed by traditional 2D cell culture. Cell proliferation and migration in these experiments have yet to be explicitly characterised, limiting the ability of experimentalists to determine the effects of various experimental conditions, such as scaffold geometry, on cell behaviour. We consider tissue growth by osteoblastic cells in melt electro-written scaffolds that comprise thin square pores with sizes that were deliberately increased between experiments. We collect highly detailed temporal measurements of the average cell density, tissue coverage, and tissue geometry. To quantify tissue growth in terms of the underlying cell proliferation and migration processes, we introduce and calibrate a mechanistic mathematical model based on the Porous-Fisher reaction-diffusion equation. Parameter estimates and uncertainty quantification through profile likelihood analysis reveal consistency in the rate of cell proliferation and steady-state cell density between pore sizes. This analysis also serves as an important model verification tool: while the use of reaction-diffusion models in biology is widespread, the appropriateness of these models to describe tissue growth in 3D scaffolds has yet to be explored. We find that the Porous-Fisher model is able to capture features relating to the cell density and tissue coverage, but is not able to capture geometric features relating to the circularity of the tissue interface. Our analysis identifies two distinct stages of tissue growth, suggests several areas for model refinement, and provides guidance for future experimental work that explores tissue growth in 3D printed scaffolds.

Statement of Contribution of Co-Authors for
Thesis by Published Paper

The authors listed below have certified that:

1. they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
2. they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
3. there are no other authors of the publication according to these criteria;
4. potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit, and
5. they agree to the use of the publication in the student's thesis and its publication on the QUT's ePrints site consistent with any limitations set by publisher requirements.

In the case of this chapter:

Model-based data analysis of tissue growth in thin 3D printed scaffolds
Published 2021 in Journal of Theoretical Biology
doi:10.1016/j.jtbi.2021.110852

Contributor	Statement of Contribution
Alexander P Browning 4 Aug 2021	Designed the study, developed and analysed the mathematical model, implemented the statistical techniques, implemented the numerical algorithms, processed the experimental data, generated and interpreted results, drafted the manuscript, and revised the manuscript during peer-review.
Oliver J Maclaren	Designed the study, developed the mathematical model, interpreted the results and provided comments on the manuscript.
Pascal R Buenzli	Designed the study, developed the mathematical model, interpreted the results and provided comments on the manuscript.
Matthew Lanaro	Performed the experiments, processed the experimental data and provided comments on the manuscript.
Mark C Allenby	Performed the experiments, processed the experimental data and provided comments on the manuscript.
Maria A Woodruff	Performed the experiments and provided comments on the manuscript.
Matthew J Simpson	Designed the study, supervised the research, interpreted the results, and provided comments on the manuscript.

Principal Supervisor Confirmation

I have sighted email or other correspondence from all Co-authors confirming their certifying authorship.

Matthew J Simpson
Name

Signature

9 August 2021
Date

6.1 Introduction

Cell culture scaffolds provide biomimetic experimental models to explore tissue growth in essential biological processes such as bone remodelling and development [12, 44, 380]. Achieving control over tissue growth through these scaffolds has clinical applications such as replacing synthetic grafts with artificially regenerated tissues [11, 44–46]. Three-dimensional (3D) printing technology [381–384] enables precise control of scaffold geometry, including the size and shape of the pores that comprise each scaffold. Despite these technological advances, the effects of scaffold geometry on scaffold-level properties of tissue growth, such as the time for tissue to close or *bridge* scaffold pores, and individual-level properties, such as cell proliferation and migration rates, are yet to be explicitly understood.

A preference to move away from traditional *in vitro* 2D culture and towards mimicking biological features, such as the bone micro-environment, through 3D scaffolds has been aided by more accurate 3D printing processes [381, 382]. Technologies based on melt electrowriting [383, 384] enable precise control of scaffold geometry, ensuring consistency and reproducibility, whilst printing fibres that are significantly smaller than traditional 3D printing processes enabling us to create more biologically relevant structures. There is a significant body of research that guides the material and physical properties of scaffold construction, but a comparative scarcity on the influence of scaffold architecture on cell and tissue behaviour. In fact, several recent studies suggest that tissue growth in pore infilling experiments is strongly curvature controlled [385–388], which implies that pore shape and size play a significant role in tissue growth [386, 388] since the average curvature of a pore is a function of its size [13].

In this work, we consider tissue growth by osteoblastic cells in a 3D printed scaffold formed of thin square pores with depth $\approx 100\,\mu\text{m}$ and side lengths ranging from 300 to $600\,\mu\text{m}$ (figure 6.1a–d). This thin geometry means that we can approximate the three-dimensional tissue growth as a depth-averaged two-dimensional phenomena [389]. Initially located only on the scaffold fibres, cells migrate and proliferate to form new tissue that bridges each pore over an experimental duration of 28 days (figure 6.1e–h). Scaffolds are systematically harvested and stained to obtain fluorescent microscopy images that provide highly detailed information about the pore bridging progress (which we measure as the proportion of the pore containing tissue) and the cell density within each pore. The variability in pore bridging we see in the experimental data is striking: smaller pores appear, on average, to bridge at earlier times (figure 6.1e–f) [13]; and some, but not all, larger pores are bridged at the conclusion of the experiment (figure 6.1g–h). While we expect larger pores—which require the development of a larger amount of tissue and cells to migrate a greater distance—to bridge more slowly [13], it is unclear whether there are also changes in cell behaviour between pore sizes. We aim to determine whether there are fundamental differences in cell proliferation and migration between different pore sizes while demonstrating the value of collecting experimental data relating to both tissue coverage and cell population.

To disentangle the effects of cell proliferation and migration on tissue growth, we perform model-based data analysis using a deterministic, continuum, process model [390]. Existing continuum models of tissue growth within porous scaffolds typically neglect information relating to properties such as cell count or density [391, 392]. Instead, the time-evolution of tissue interfaces are described using techniques ranging from continuum mechanics [12, 392, 393] to curvature flow [385, 387, 394–398]. While these models often provide good agreement with geometric features in experimental data, they yield parameter estimates that are purely
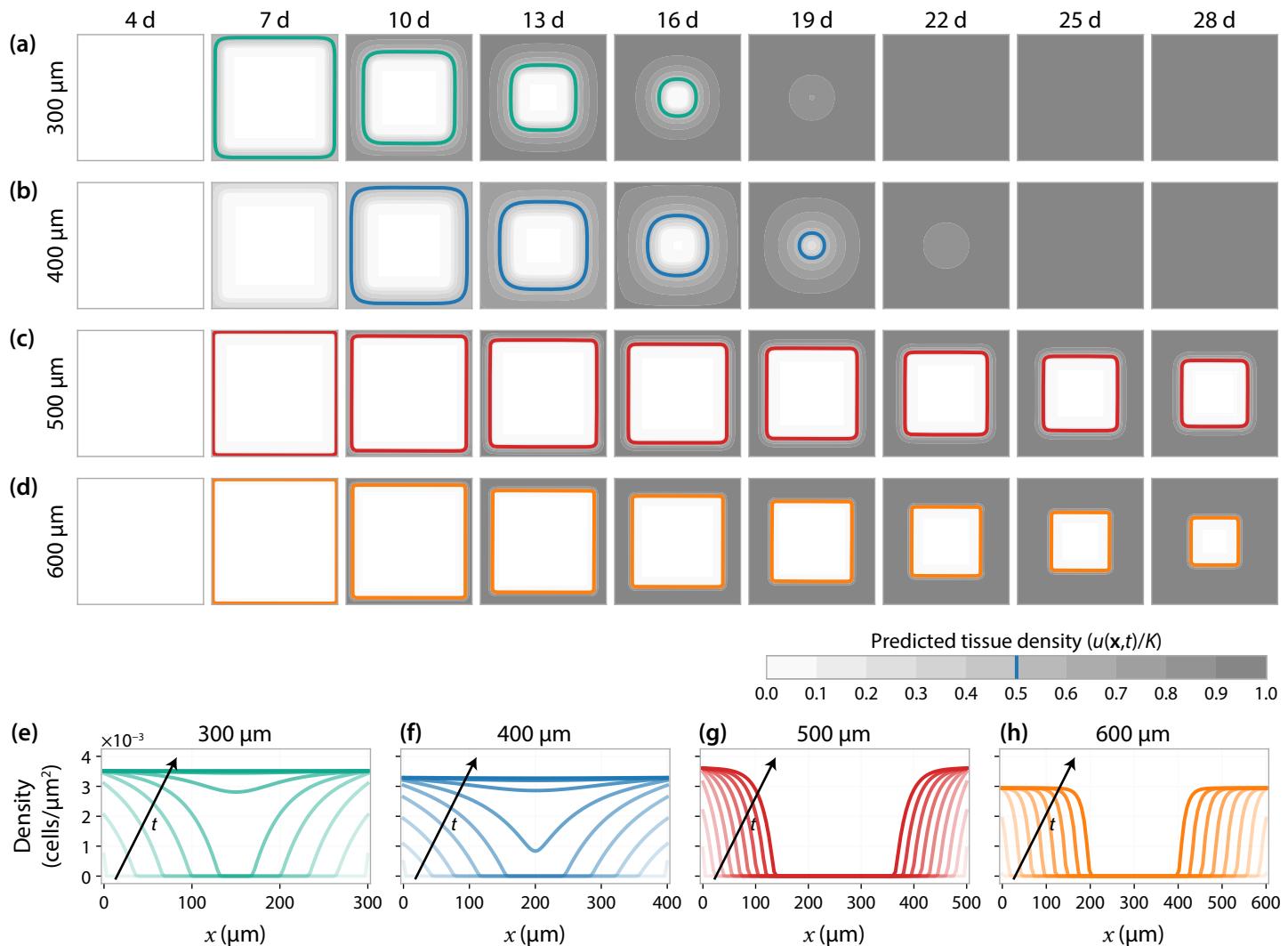
▲ Figure 6.1. Scaffold geometry and experimental data. (a–e) Scaffolds comprise a grid of square pores with lengths ranging from 300 to 600 µm. (a) 3D graphical render of the scaffold structure. (b–e) An 800 × 800 µm DIC image taken from the central region of a scaffold for each pore size. (f–i) Composite fluorescence microscopy images of pore bridging experiments. Cell nuclei, stained with DAPI, are shown in the blue channel; tissue and cytoskeleton, stained with phalloidin, are indicated in the green channel. Scale varies between pore sizes, but is identical within a pore size and is indicated in day 28 images. It is important to note that scaffolds are fixed to obtain images: data from successive time-points are independent experiments.

phenomenological and lack a clear biophysical interpretation. We describe pore bridging using a relatively simple two-dimensional reaction-diffusion equation, often referred to as the Porous-Fisher model [13, 68]. This choice naturally accounts for density-dependent behaviour expected in these experiments: contact inhibition limits cell proliferation in high-density regions, and contact stimulates cell migration, leading to co-operative tissue growth that is limited in regions of low cell density.
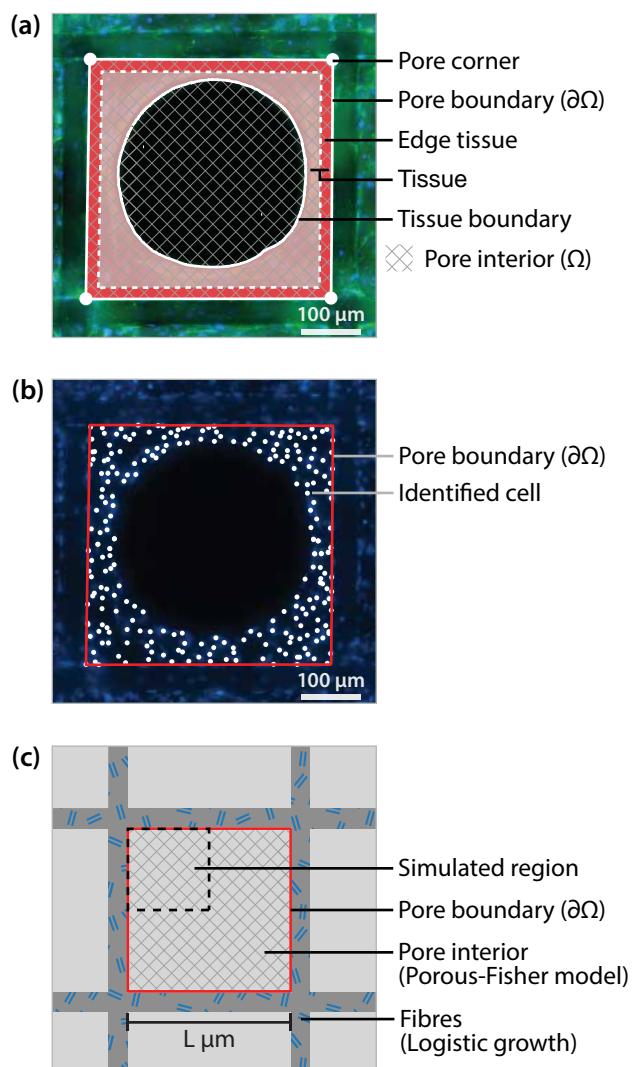
We take a summary statistic and likelihood-based approach to parameter inference [15] to identify parameters that characterise cell behaviour both individually within each pore size, and across all pore sizes simultaneously. In comparison to our previous work [13], we consider a temporal dataset that includes information about both cell density and bridging progress. To quantify the uncertainty associated with parameter estimates—which may be non-identifiable from the available information in the experimental data [243, 268]—we perform profile likelihood analysis [399, 400], which facilitates the computation of approximate confidence intervals [24, 27, 401]. We compare parameter estimates that quantify cell proliferation and migration rates across pore sizes to determine whether pore size, and by extension, curvature, influence cell behaviour. For example, if pore size and, by extension, curvature, play a significant role in cell proliferation, we would expect the estimates of the cell proliferation rate to vary significantly between pore sizes.

Compared to models of 2D culture, which are well developed and routinely applied in experimental design [66, 246, 390, 402–404], there is limited data-based modelling guidance for tissue growth within 3D scaffolds. Development and verification of mechanistic models for pore bridging is essential: models can guide engineering design choices in scaffold construction to optimise and control tissue growth [405]. Despite the widespread application of reaction-diffusion models in collective cell behaviour [36–38, 66] and biology more broadly [18, 406–408], their suitability to describe geometrically-induced phenomena—such as that arising from corners and the relatively small, constrained, domain in our experiments—remains largely unexplored. Qualitatively, the Porous-Fisher model produces results that capture key behaviours in the experimental data; namely both an increase in cell density over the duration of the experiment, and sharp-fronted tissue growth that bridges each pore (figure 6.2). A key focus of our work is to further verify the appropriateness of the Porous-Fisher model by comparing features not used for calibration to model predictions, and comparing parameter estimate and model behaviour across pore sizes. Given that tissue growth is thought to be curvature controlled [386], we focus on comparing geometric features in the data, such as circularity, to model predictions. Comparing parameter estimates and model predictions across pore sizes is crucial for model verification: if only a single experimental condition is considered, the model might appear to match the experimental data but be incapable of matching data across multiple experimental conditions without significantly varying the parameters [30, 403]. Through this analysis, we identify several avenues for both future experimentation and model refinement.

The outline of the work is as follows. We first describe the experimental model and methods used to summarise the data (Section 6.2.1). The data are available on GitHub as supplementary material. In Section 6.2.2, we describe a two-component mathematical model comprising both a deterministic process model that describes pore bridging dynamics and a probabilistic observation process that connects model predictions to noisy experimental observations. Subsequently, we outline the techniques used to obtain maximum likelihood estimates and likelihood profiles (Section 6.2.3). We present and discuss the results in Section 6.3 before outlining future experimental and mathematical modelling recommendations in Section 7. Code to reproduce all results are provided in the high-performance, open-source, Julia language on GitHub.

▲ Figure 6.2. Model simulated tissue growth. (a–d) Model simulation using the maximum likelihood estimates for each pore size, where information relating to cell density and tissue coverage are included in the likelihood. Parameters are given in table 6.1. The coloured curves show the boundary of the ECM, taken to be $\tau = 0.5$ (50%) of carrying capacity, K. Shown in greyscale is the density as a proportion of carrying capacity. (e–h) Cell density profiles, as a function of x, taken at across the centre of each pore at $y = L/2$. Shown is a profile for each value of t in (a–d); arrows and increasing opacity indicate direction of increasing t.

◄ Figure 6.3. Data processing technique and experimental domain. Example data summarisation for a 400 μm pore at day 14, indicated in figure 6.1g. (a) The pore boundary and tissue identified using the semi-automated data processing approach. Also shown is the region classified as edge tissue. In the model, we denote the boundary of the pore $\partial\Omega$, and the interior of the pore Ω. (b) DAPI image, showing cell nuclei, with the pore boundary and cell locations superimposed. (c) Schematic of the mathematical domain. Tissue growth in the pore interior is described using the Porous-Fisher reaction diffusion model, and tissue growth on the fibres is described using logistic growth.

6.2 Methods

6.2.1 Pore bridging experiments

Polycaprolactone fibres of diameter 50 μm are fabricated into a two-layer scaffold of size 7×7 mm through melt electrospinning. The resultant scaffold has an overall thickness of approximately 100 μm (two fibre layers) and comprises square shaped pores of lengths 300, 400, 500 and 600 μm (figure 6.1a–d). Prior to cell seeding, scaffolds are sterilised and incubated in 5% CO_2 overnight.

Murine calvarial osteoblastic cells (MC3T3-E1) [409] are cultured in α-MEM, 10% fetal bovine serum, and 1% penicillin-streptomycin (Thermo Fisher). Scaffolds are placed on top of non-adherent 2% agarose to prevent cell-to-plate attachment within a 48-well plate. Cells are detached using 0.05% trypsin and seeded at 7500 cells in 250 μL media onto each scaffold within a 48-well plate (Nunc, Thermo Fisher). Cells are allowed 4 h to attach to each scaffold before an additional 250 μL of media is added. Cell-seeded scaffolds are cultured in a humidified environment at 37 °C in 5% CO_2 for 28 days. Media is changed every 2–3 days from day 1 to 14, every 1–2 days from day 15 to 21, then every day from day 22 to 28. Cell viability is assessed at day 10, 14 and 28 using calcein AM (to stain live cells) and ethidium homodimer (to stain dead cells).

Cell-seeded scaffolds are fixed with 4% paraformaldehye at day 4, 7, 10, 14, 18, and 28. Replicates are

stained with both DAPI and Alexa Fluor™ 488 Phalloidin (Thermo Fisher), which stain cell nuclei and actin filaments, respectively. Fluorescent microscopy (Leica AF6000 LX) is used to capture high resolution images of the centre of each scaffold. To accurately identify scaffold geometry, a differential interference contrast (DIC) image is also captured. Fixation, staining and microscopy are repeated across two or three replicates for each pore size and time point. Each experimental replicate yields information about 9 to 12 pores, providing tissue growth data across days 4 to 28 from 618 pores in total. In figure 6.1e–h we show composite images of four pores for each pore size, for each time point.

Data summarisation

The tissue growth data are processed in a semi-automated fashion using MATLAB [410] to obtain information about the cell population and bridging progress in each pore (figure 6.3). First, the four corners of each pore are identified manually from the DIC image and thresholding is applied to the phalloidin image to establish the region in each pore containing tissue (figure 6.3a). Next, the location of the cells within each pore are identified using the thresholded DAPI image, which colours cell nuclei (figure 6.3b). Data are discarded for pores that are not deemed sufficiently regular in shape, or for which accurate measurements cannot be taken.

We summarise the experimental data obtained from each pore with four summary statistics, denoting $y_i^{L,t,j}$ the jth observation of the ith summary statistic at time t for a pore of size length L. These are as follows.

1. Average cell density:
$$y_1^{L,t,j} = \frac{\text{Cell count in pore}}{\text{Area of pore}}, \quad 0 \leq y_1^{L,t,j} < \infty. \tag{6.1}$$

2. Coverage:
$$y_2^{L,t,j} = \frac{\text{Area of tissue}}{\text{Area of pore}}, \quad 0 \leq y_2^{L,t,j} \leq 1. \tag{6.2}$$

3. Edge density:
$$y_3^{L,t,j} = \frac{\text{Cell count on edge tissue}}{\text{Area of edge tissue}}, 0 \leq y_3^{L,t,j} < \infty. \tag{6.3}$$

Here, we define edge tissue as tissue located within approximately $20\,\mu\text{m}$ of the pore boundary (figure 6.3a).

4. Circularity:
$$\tilde{y}_4^{L,t,j} = \frac{4\pi \times \text{Area of tissue void}}{(\text{Perimeter of tissue void})^2}, \quad \pi/4 \lessapprox \tilde{y}_4^{L,t,j} \lessapprox 1. \tag{6.4}$$

$$y_4^{L,t,j} = \frac{\tilde{y}_4^{L,t,j} - 1}{1 - \pi/4} + 1, \quad 0 \lessapprox y_4^{L,t,j} \lessapprox 1. \tag{6.5}$$

Here, $\tilde{y}_4^{L,t,j}$ represents the standard measure of circularity or roundness [410], which approaches unity as the tissue void approaches a perfect circle. For a square shape, $\tilde{y}_4^{L,t,j} = \pi/4$. Since our experiments consider a scaffold that is approximately square, we normalise $\tilde{y}_4^{L,t,j}$ to obtain $y_4^{L,t,j}$ that still tends to unity as the tissue void approaches a perfect circle, but tends to zero as the tissue void approaches a

square. To smooth out small-scale irregularities in the identified tissue shape, the convex hull of the largest contiguous tissue void is used to calculate the circularity [410].

6.2.2 Mathematical model

We interpret the pore bridging experiments with a deterministic spatio-temporal process model that aims to capture the key biological processes involved in tissue growth. To account for variability in the experimental data, we model experimental observations as normally distributed about predictions made through the process model [189, 411, 412]. In this section, we describe the process model and the probabilistic observation process used for analysis. A schematic of the mathematical domain is provided in figure 6.3c.

Process model

The substrate of the wells containing the scaffolds is coated with non-adherent agarose that does not allow cell attachment, while cells are initially adhered to the scaffold fibres. Cells, therefore, cannot move freely into the pore void. Rather, cells work together to bridge the pore through interconnecting material such as extracellular matrix and intracellular actin filaments. Therefore, traditional models of cell migration based upon linear diffusion, which do not capture the sharp tissue boundary seen in the experimental data (figure 6.1e–h), are inappropriate.

In this work, we assume that cells move at a rate proportional to their own density and proliferate logistically to a maximum density of K, which we model with the Porous-Fisher [36] equation, given by

$$\frac{\partial u}{\partial t} = D\nabla \cdot \left[\left(\frac{u}{K} \right) \nabla u \right] + \lambda u \left(1 - \frac{u}{K} \right), \quad x \in \Omega. \tag{6.6}$$

The vertical depth of the pores is small compared to the horizontal length scale (figure 6.1a), and we observe that cells form a thin horizontal layer of tissue, approximately one cell thick. Therefore, we integrate out the vertical dimension [389] so that $x = (x, y)$ and $u(x, t)$ is a depth-averaged density, which we refer to as the cell density. In equation (6.6), Ω the interior of the pore (figure 6.3a) and $\theta = (D, \lambda, K)$ are parameters that relate to the diffusivity, proliferation rate, and carrying capacity, respectively.

The pore is surrounded by a fibre on which cells are initially placed approximately uniformly so that, on the fibre, u is homogeneous, $u > 0$, and $\nabla u = 0$. We assume that both the proliferation rate and maximum packing density of cells on the fibre is the same as in the pore interior and, therefore, assume that cell growth on the fibres is logistic (figure 6.3c). Substituting $\nabla u = 0$ into equation (6.6) recovers a time-dependent Dirichlet boundary condition on the edge of the pore

$$\frac{\mathrm{d}u}{\mathrm{d}t} = \lambda u \left(1 - \frac{u}{K} \right), \quad x \in \partial\Omega, \tag{6.7}$$

where $\partial\Omega$ represents the pore boundary (figure 6.3a). In Appendix E.8, we investigate the application of homogeneous Neumann (zero-flux) boundary conditions which can be derived through a symmetry argument from the migration of adjoining pores. In summary, we find that such a boundary condition leads to behaviour inconsistent with the experimental observations.

Initially, cells appear distributed exclusively on the fibre, and not in the interior of the pore. It is not until after $t_0 = 4$ d that cells visibly start the pore bridging process (figure 6.1a–d). We, therefore, assume that at

$t_0 = 4$ d, cells are distributed around the pore boundary (i.e., on the fibre) with an initial density u_0, which we assume to be unknown and, therefore, estimate for each pore size. The initial condition is given by

$$u(\mathbf{x}, t_0) = \begin{cases} u_0, & \mathbf{x} \in \partial\Omega, \\ 0, & \mathbf{x} \in \Omega. \end{cases} \tag{6.8}$$

We solve equation (6.6)–(6.8) using a finite difference scheme based upon a discretisation with 101×101 mesh points for each pore size. Due to the symmetry of the problem, we only solve equation (6.6)–(6.8) on a quarter-domain (figure 6.3c). To integrate the resultant system of ordinary differential equations, we apply the standard `Tsit5` routine in Julia [275, 302]. Full details and verification that the spatial discretisation is sufficient are available in the supplementary material.

Observation process

Whereas output from the mathematical model is deterministic and comprises the cell density, $u(\mathbf{x}, t)$, as a function of space and time, the experimental observations comprise noisy observations of four summary statistics, $\mathbf{y}^{L,t,j} = (y_1^{L,t,j}, y_2^{L,t,j}, y_3^{L,t,j}, y_4^{L,t,j})$. To compare model realisations to experimental observations, we define functions that map $u(\mathbf{x}, t)$ to summary statistics that correspond to those that summarise the experimental data. These functions are as follows.

1. Average cell density:

$$\mu_1(t) = \frac{1}{L^2} \iint_\Omega u(\mathbf{x}, t) \, d\mathbf{x}, \quad 0 \leq \mu_1(t) \leq K. \tag{6.9}$$

We approximate the integral in equation (6.9) numerically using the trapezoid rule.

2. Coverage:

$$\mu_2(t) = 1 - \frac{A_{\text{void}}(u(\mathbf{x}, t); \tau K)}{L^2}, \quad 0 \leq \mu_2(t) \leq 1. \tag{6.10}$$

Here, τ represents a proportion of maximum cell density, K, at which tissue becomes visible, so that in regions where $u(t, x, y) > \tau K$, cells are considered part of the observed newly formed tissue and $A_{\text{void}}(u(\mathbf{x}, t); \tau K)$ is the area of the tissue void. In this work, we fix $\tau = 0.5$, so that the tissue boundary in the model is assumed to be where the density is 50% of the maximum [13]. To calculate the area of the tissue void, $A_{\text{void}}(u(\mathbf{x}, t); \tau K)$, we apply an interpolation method to approximate the tissue boundary (supplementary material). This approach ensures that $\mu_2(t)$ remains a continuous function in the parameter space, which is desirable for computational inference.

3. Edge density:

$$\mu_3(t) = u(\mathbf{x}_b, t), \quad 0 \leq \mu_3(t) \leq K. \tag{6.11}$$

Here, $\mathbf{x}_b$ is any point on the pore boundary (the modelled cell density is homogeneous on the pore boundary); we set $\mathbf{x}_b = (0, 0)$.

4. Circularity:

$$\tilde{\mu}_4(t) = \frac{4\pi A_{\text{void}}\big(u(\mathbf{x}, t); \tau K\big)}{P^2_{\text{void}}\big(u(\mathbf{x}, t); \tau K\big)}, \qquad \pi/4 \leq \tilde{\mu}_4(t) \leq 1, \qquad (6.12)$$

$$\mu_4(t) = \frac{\tilde{\mu}_4 - 1}{1 - \pi/4} + 1, \qquad 0 \leq \mu_4(t) \leq 1. \qquad (6.13)$$

Here, $P_{\text{void}}\big(u(\mathbf{x}, t); \tau K\big)$ is an interpolated approximation of the perimeter of the tissue boundary (supplementary material). As for $\tilde{y}_4^{L,t,j}$, we normalise $\tilde{\mu}_4(t)$ to obtain $0 \leq \mu_4(t) \leq 1$ (equation (6.5)). For simulations where the coverage exceeds 0.99, we set $\mu_4(t) = 1$ for convenience.

To account for biological noise and measurement error, we assume that model realisations describe the *expected behaviour* and that observations of the summary statistics are independent and normally distributed [411]. Therefore,

$$y_i^{L,t,j} \sim \text{Normal}\bigg(\mu_i(t; L, \boldsymbol{\theta}), \sigma_i^2\big(\mu_i(t; L, \boldsymbol{\theta})\big)\bigg). \qquad (6.14)$$

Here, we write $\mu_i(t) = \mu_i(t; L, \boldsymbol{\theta})$ to emphasise the dependence of model realisations on the pore size, L, and set of unknown parameters, $\boldsymbol{\theta} = (D, \lambda, K, u_0)$. We observe in figure 6.4 that the variability in the experimental data varies significantly between both summary statistics and observation times. Therefore, we pre-estimate a variance function, $\sigma_i(\mu_i)$ as a function of the mean [411]. Here, we take $\sigma_i(\cdot)$ to be a quadratic, with intercept of 10% of the maximum standard deviation observed for the summary statistic (supplementary material).

6.2.3 Inference
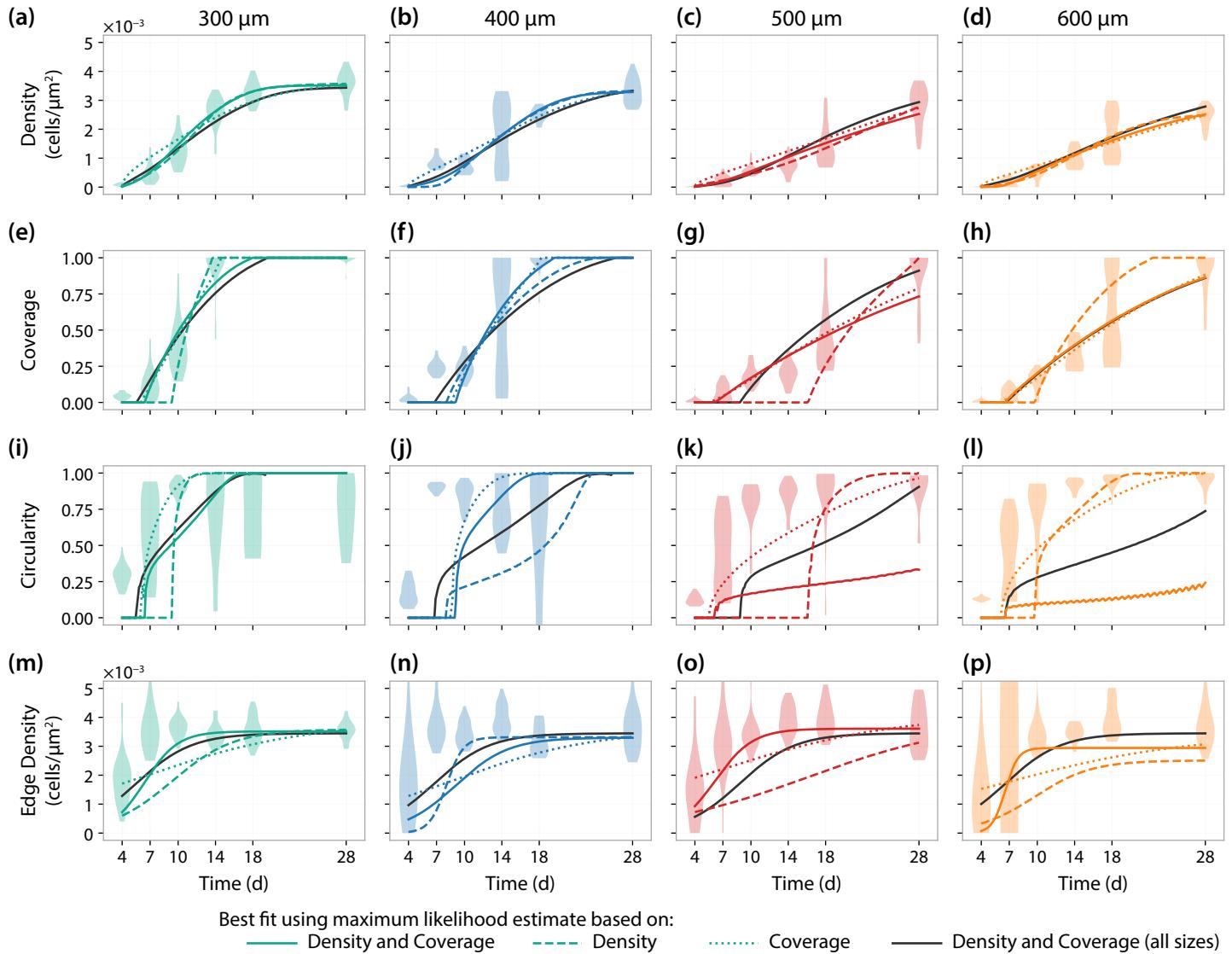
We take a summary statistic, likelihood-based, approach to inference and sensitivity analysis. Given a set of observations from pores of size L, $\text{Y}^L = \{\text{y}^{L,t,j}\}_{j,t}$, the log-likelihood function is given by

$$\ell(\boldsymbol{\theta}; \text{Y}^L, L) = \sum_{t \in \mathcal{T}} \sum_j \sum_{i \in \mathcal{S}} \log \phi\bigg(y_i^{L,t,j}; \mu_i(t; L, \boldsymbol{\theta}), \sigma_i^2\big(\mu_i(t; L, \boldsymbol{\theta})\big)\bigg), \qquad (6.15)$$

where $\mathcal{T} = \{7, 10, 14, 18, 28\}$ is the set of observation times ($t_0 = 4\,\text{d}$ is excluded from the analysis); $\mathcal{S} \subseteq \{1, 2, 3, 4\}$ is the set of summary statistics included in the analysis; and $\phi(x; \mu, \sigma^2)$ is the normal density function.

Parameter bounds

The set of unknown parameters, $\boldsymbol{\theta} = (D, \lambda, K, u_0)$, carry a physical interpretation so we can formulate realistic parameter bounds. The doubling time of MC3T3-E1 osteoblast cells in two-dimensional culture is approximately $15\,\text{h}$ [409], which corresponds to a proliferation rate of approximately $\lambda \approx 1.1\,\text{d}^{-1}$. Analysis based upon the overall bridging time of MC3T3-E1 osteoblast cells suggests D carries a magnitude of approximately $100\,\mu\text{m}^2\,\text{d}^{-1}$ [13]. Results in figure 6.4a,b suggest that cell density is bounded above by approximately $4 \times 10^{-3}\,\text{cells}\,\mu\text{m}^{-1}$, which corresponds to a packing density where a monolayer of cells occupy the same amount of space as a disk with diameter of approximately $18\,\mu\text{m}$. Based on these values, we choose conservative bounds such that

▲ Figure 6.4. **Processed experimental data and model fits.** Experimental data and model fit showing (a–d) the density, (e–h) the coverage, (i–l) the circularity, and (m–p) the edge density. Violin plots show the experimental data. In each case, a model prediction is shown based on the maximum likelihood estimate that includes information relating to the cell density (dashed colour); cell density and tissue coverage (solid colour); tissue coverage with day 28 density measurements (dotted colour); and cell density and tissue coverage from all pore sizes (solid grey).

$$
\begin{aligned}
10 &\le D \le 2000\,\mu\text{m}^2\,\text{d}^{-1}, \\
1 \times 10^{-2} &\le \lambda \le 2\,\text{d}^{-1}, \\
2 \times 10^{-3} &\le K \le 5 \times 10^{-3}\,\text{cells}\,\mu\text{m}^{-2}, \\
1 \times 10^{-5} &\le u_0 \le 2 \times 10^{-3}\,\text{cells}\,\mu\text{m}^{-2}.
\end{aligned}
\tag{6.16}
$$

Maximum likelihood estimation

We apply maximum likelihood estimation [413] to obtain a best fit parameter combination, $\hat{\boldsymbol{\theta}}^L$, for each pore size. The maximum likelihood estimate (MLE) is given by

$$
\hat{\boldsymbol{\theta}}^L = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, \ell(\boldsymbol{\theta}; \mathrm{Y}^L, L),
\tag{6.17}
$$

subject to the bounds given in equation (6.16).

To compute a numerical approximate the solution to equation (6.17), we employ both a global and local optimisation routine from the open-source `NLopt` optimisation library [414]. First, we run a global optimisation routine, based on the `DIRECT` algorithm [415], for a fixed amount of time (chosen to be 6 hours). This approach avoids the need to specify an initial guess of $\boldsymbol{\theta}$ for the optimisation routine. We then use the output from the global optimisation routine as the initial guess in a the local optimisation algorithm `BOBYQA` [416]. We look for a maximum with absolute threshold of 10^{-4}, several orders of magnitude below the threshold of 1.92 for an approximate univariate 95% confidence interval from a normalised-log-profile-likelihood [413].

Profile likelihood analysis

While point estimates provide a means of assessing the ability of the model to capture features in the data, we are interested in establishing parameter uncertainties and comparing estimates across pore sizes. To do this, we profile the log-likelihood function for each parameter [24, 401].

First, we partition the parameter space into a parameter (or group of parameters) of interest, ψ, and nuisance parameters, $\boldsymbol{\gamma}$, such that $\boldsymbol{\theta} = (\psi, \boldsymbol{\gamma})$. The profile log-likelihood for the parameter ψ is given by

$$
\ell_p(\psi; \mathrm{Y}^L, L) = \sup_{\boldsymbol{\lambda}} \ell(\psi, \boldsymbol{\lambda}; \mathrm{Y}^L, L).
\tag{6.18}
$$

For example, to profile the diffusivity, we would specify $\psi = D$ and $\boldsymbol{\gamma} = (\lambda, K, u_0)$. To obtain a value of $\ell_p(D; \mathrm{Y}^L, L)$, we maximise the log-likelihood function in the case that D is fixed.

Likelihood-based confidence intervals can be defined from the profile log-likelihood by an asymptotic approximation using the chi-squared distribution, for sufficiently regular problems [413, 417]. 95% confidence intervals and regions are given using the threshold values of 1.92 and 3.00 log-likelihood units below the maximum for univariate and bivariate profiles, respectively [413, 417, 418]. It is convenient to work with a normalised profile log-likelihood

$$
\hat{\ell}_p(\psi; \mathrm{Y}^L, L) = \sup_{\boldsymbol{\lambda}} \ell(\psi, \boldsymbol{\lambda}; \mathrm{Y}^L, L) - \ell(\hat{\boldsymbol{\theta}}^L; \mathrm{Y}^L, L), \quad \hat{\ell}_p \le 0.
\tag{6.19}
$$

Here, a 95% confidence interval is given where $\hat{\ell}_p(\psi; Y^L, L) \geq -1.92$, for example [413].

To compute numerical approximations to each profile log-likelihood, we employ the local optimisation routine BOBYQA [416]. The log-likelihood is profiled along a regular spaced grid, $(\psi_1, \psi_2, ..., \psi_M)$, in series, starting at the grid point closest to the MLE, using the MLE as the initial guess [401]. Subsequent grid points use the output from the previous grid points as an initial guess. Again, we look for a maximum with absolute threshold of 10^{-4}.

6.3 Results and Discussion

We interpret spatially-detailed, temporal, pore bridging data from a range of pore sizes using a relatively simple reaction-diffusion model. Our analysis considers data relating to the spatial characteristics of tissue growth—specifically, the tissue coverage and circularity of the tissue void—in addition to typical measurements, such as cell density. We aim to quantitatively determine whether there are fundamental differences in cell behaviour and tissue growth between different pore sizes, and verify the appropriateness of the reaction-diffusion model in explaining pore bridging, by comparing results across a series of experiments with various pore size. In particular, applications of reaction-diffusion models to describe tissue growth are typically limited to one-dimensional or unbounded geometries [36–38, 66, 246, 402–404, 406]; there is comparatively little guidance on applying these models to describe the geometrically constrained phenomena we study.

In figure 6.1e–h, we show a subset of the experimental images obtained for each pore size over the duration of the experiment, and in figure 6.4 we show the summary statistics collected from the processed experimental data for each pore size. As each scaffold is fixed prior to staining and imaging, we note that data collected from successive time points are statistically independent. We work with average cell density (figure 6.4a–d) instead of cell count to allow direct comparison between pore sizes. It is not until after day four that cells migrate from the fibres into the pore void, so we exclude data collected at earlier time points from the analysis, and calibrate the model with observations taken after day four. Observations from day four itself are excluded as cells primarily occupy the fibres, which the model does not consider (figure 6.1e–f). By the end of the experiment (28 days), the majority of the 300 and 400 μm pores are bridged (85% and 100%, respectively), and the cell density appears very close to a steady-state (the average cell densities are 102% and 93% of the edge density, respectively). In comparison, several of the 500 and 600 μm pores do not bridge at the conclusion of the experiment (70% and 60% bridged at day 28, respectively), and, for these pore sizes, cell growth is more evident between days 18 and 28 (cell densities are 78% and 69% of edge density, respectively).

Using information about the cell density and tissue coverage, we calibrate the Porous-Fisher model to obtain a maximum likelihood estimate (MLE), θ^L, individually for each pore size (Table 6.1). We show the solution of the model at the MLE, which we refer to as the best fit, along with the predicted tissue boundary in figure 6.2. Qualitatively, the behaviour predicted by the model matches that seen in figure 6.1 for the experimental data. First, the Porous-Fisher model predicts sharp-fronted migration, where regions ahead of the tissue boundary are devoid of cells. Second, we see cell migration drive tissue growth that bridges each pore. Pore bridging appears to occur at a slower rate for the larger pores, consistent with experimental observations. A counter-intuitive result that highlights the variability in pore bridging we see in the experimental data is that the 600 μm pores are predicted to bridge faster than the 500 μm pores: this is also seen in the experimental

data, where at day 18 tissue coverage is greater in the 600 μm than the 500 μm pores (figure 6.1).

In figure 6.4, we overlay a time-series of the best fit for each summary statistic with the experimental data, and in figure 6.5 we compare relationships between summary statistics predicted by the model to the experimental data. In all cases, we interpret realisations of the deterministic process model as the expected behaviour. To determine the distinct value of collecting information relating to the cell density and coverage, we also calculate the MLE in the case where we calibrate the model using (i) the cell density alone, and (ii) the coverage alongside day 28 observations of the cell density. Finally, to determine if the model can simultaneously match data across all pore sizes, we calculate the MLE using both cell density and tissue coverage information from all pore sizes (in this case, the initial density is allowed to vary between pore sizes). We show the best fits in these three additional scenarios in figure 6.4 and figure 6.5.
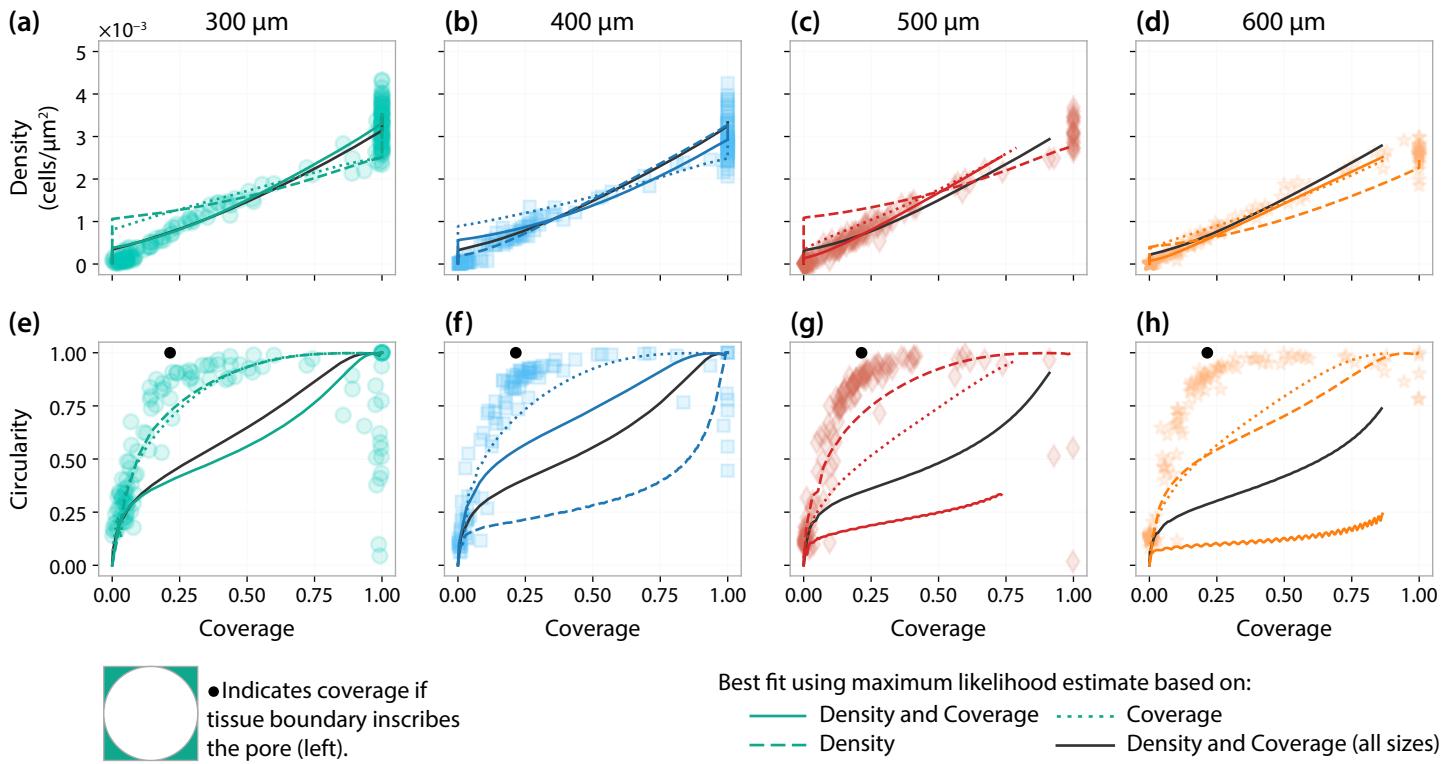
Results in figure 6.4a–h show a clear value in considering information relating to tissue coverage. We see an excellent match with experimental observations of cell density for all pore sizes (figure 6.4a–d), even for cases where only tissue coverage and day 28 cell density observations are used for model calibration. Overall, we also see an agreement with experimental observations of tissue coverage; however, when the model is calibrated using cell density information alone, the best fit does not appear to capture early time tissue formation correctly (figure 6.4e–h). The model also provides an excellent match to experimental cell density and coverage observations when calibrated to all pore sizes simultaneously. These results are important as the model does not explicitly incorporate geometric behaviour (aside from the initial and boundary conditions) yet is still able to capture features relating to tissue coverage and cell density in the experimental data. This agreement between the model and experimental data is not only the case when parameter estimates are allowed to vary between pore sizes, but also when a single set of parameters is used to describe data across all pore sizes.

Comparison between model fits and experimental observations in figure 6.4a–h highlight how variable experimental observations are despite a large sample size of $n = 618$ pores: the average density and coverage for the 400 μm pores, for example, decreases by 12% from day 7 to 10 (the model monotonically increases), and observations at day 14 of the same pore size encompass observations at nearly every other time, (figure 6.4b,f). For this reason, we have excluded day 7 observations of 400 μm pores from results in the main text. In the supplementary material, we demonstrate that including these observations leads to results inconsistent with the other pore sizes. We address possible reasons for high levels of variability data later in the discussion.

A critical area in which the model fails to capture the experimental observations is in its ability to match the circularity of the tissue boundary in the larger 500 and 600 μm pores (figure 6.4k,l and figure 6.2c,d). We

	D (μm^2 d^{-1})		λ (d^{-1})		K (cells/μm^2)	
300 μm	397	(290,653)	0.561	(0.353,0.858)	0.00352	(0.00338,0.00361)
400 μm	1030	(525,1690)	0.35	(0.191,0.694)	0.0033	(0.00314,0.00343)
500 μm	117	(40.6,269)	0.497	(0.238,1.21)	0.00361	(0.00322,0.00401)
600 μm	99.9	(54.7,240)	1.41	(0.621,2.0)	0.00294	(0.00271,0.00319)
All	426	(364,552)	0.339	(0.261,0.388)	0.00345	(0.00336,0.00355)

▲ Table 6.1. Maximum likelihood estimates obtained by calibrating the Porous-Fisher equation to information relating to the cell density and tissue coverage. Asymptotic 95% confidence intervals, approximated using the profile likelihoods (figure 6.6) are given in parentheses. All values are stated to three significant figures.

▲ Figure 6.5. Relationships between experimental and simulated summary statistics. Experimental data and model fit showing the relationship between (a–d) tissue coverage and cell density, and (e–h) coverage and circularity. In each case, a model prediction is shown based on the maximum likelihood estimate based on the cell density (dashed colour); cell density and tissue coverage (solid colour); tissue coverage with day 28 density measurement (dotted colour); and cell density and tissue coverage from all pore sizes (solid grey).
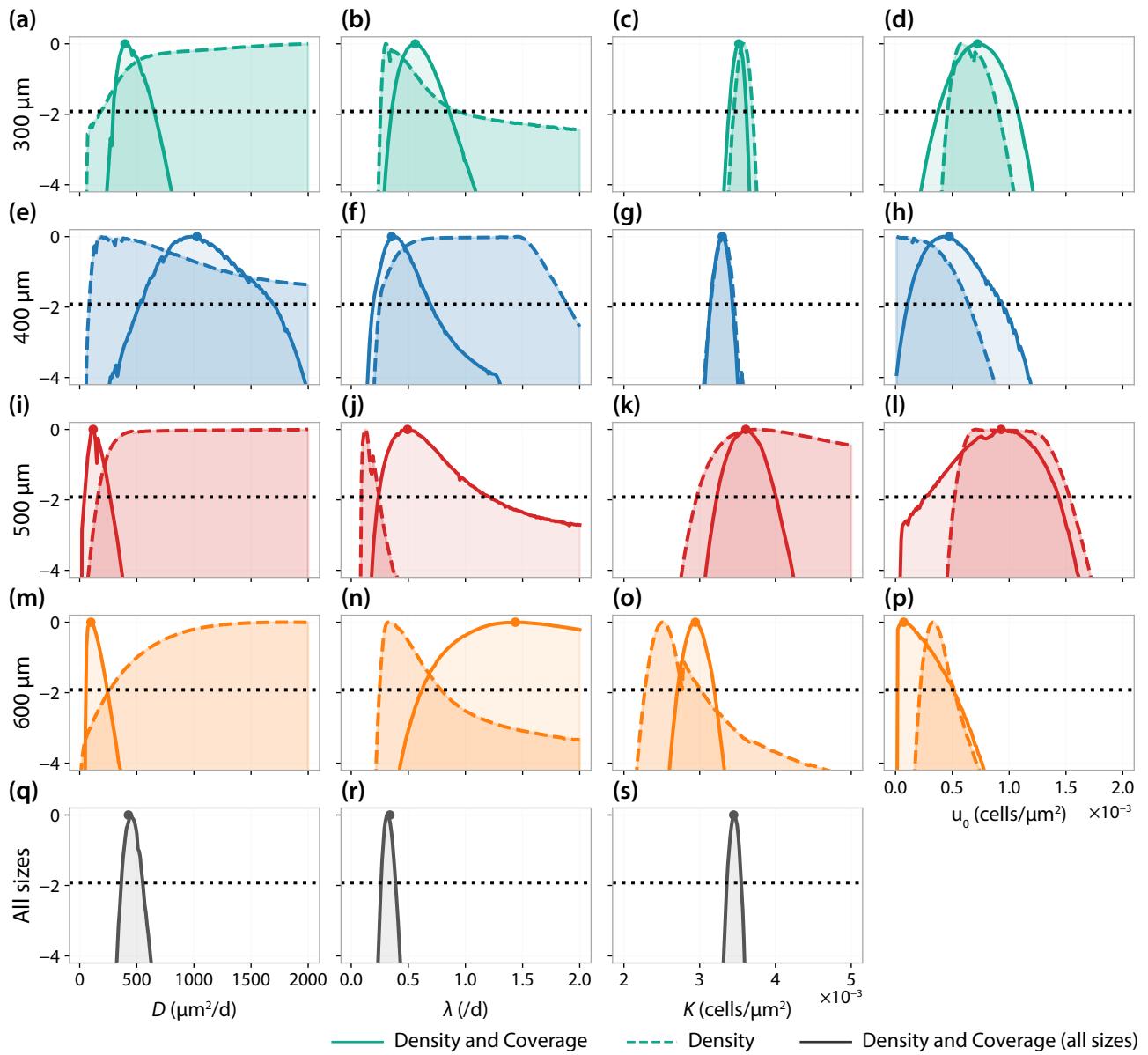
further verify this by calibrating the model to information that includes circularity, finding that the model best fit in this case does not match the circularity measurements seen in the experimental data (supplementary material). In figure 6.5e–h we explore the relationship between tissue coverage and circularity, which we note are both non-dimensional quantities and, therefore, can be directly compared between pore sizes. The relationships observed in the experimental data are remarkably consistent both between and within pore sizes, contrasting significantly to results in figure 6.4 that show highly variable observations. Comparing the tissue boundaries predicted by the model (figure 6.2) to the experimental data (figure 6.1) reveals why this may be the case. The model predicts initial tissue growth in both the corners and edge of the pore. In comparison, tissue growth in the experiments appears to occur initially only in the corners: it is not until the tissue boundary becomes almost completely circular, with a diameter equal to the pore size, that tissue growth occurs on the pore edge. We confirm this in figure 6.5e–h by calculating the coverage for a hypothetical, idealised, pore that forms a circular tissue void inscribed in the pore, equal to $100(1 - \pi/4)\%$ for all pore sizes (indicated in black). We note that since the manufacturing process never leads to pores that are perfectly square, we do not expect to see a perfectly circular pore with coverage $100(1 - \pi/4)\%$ in the experimental data. This corner corner bridging stage of tissue growth is not included in or captured by the Porous-Fisher model (figure 6.2). To develop a better understanding of corner bridging, we suggest future mathematical and experimental work focussed on corner bridging using scaffolds with pores large enough that tissues in adjacent corner tissues do not interact and start pore bridging [391, 396].

Point or maximum likelihood estimates for each parameter vary across pore sizes (Table 6.1), yet the model is able to match experimental observations of cell density and tissue coverage across all pore sizes with a
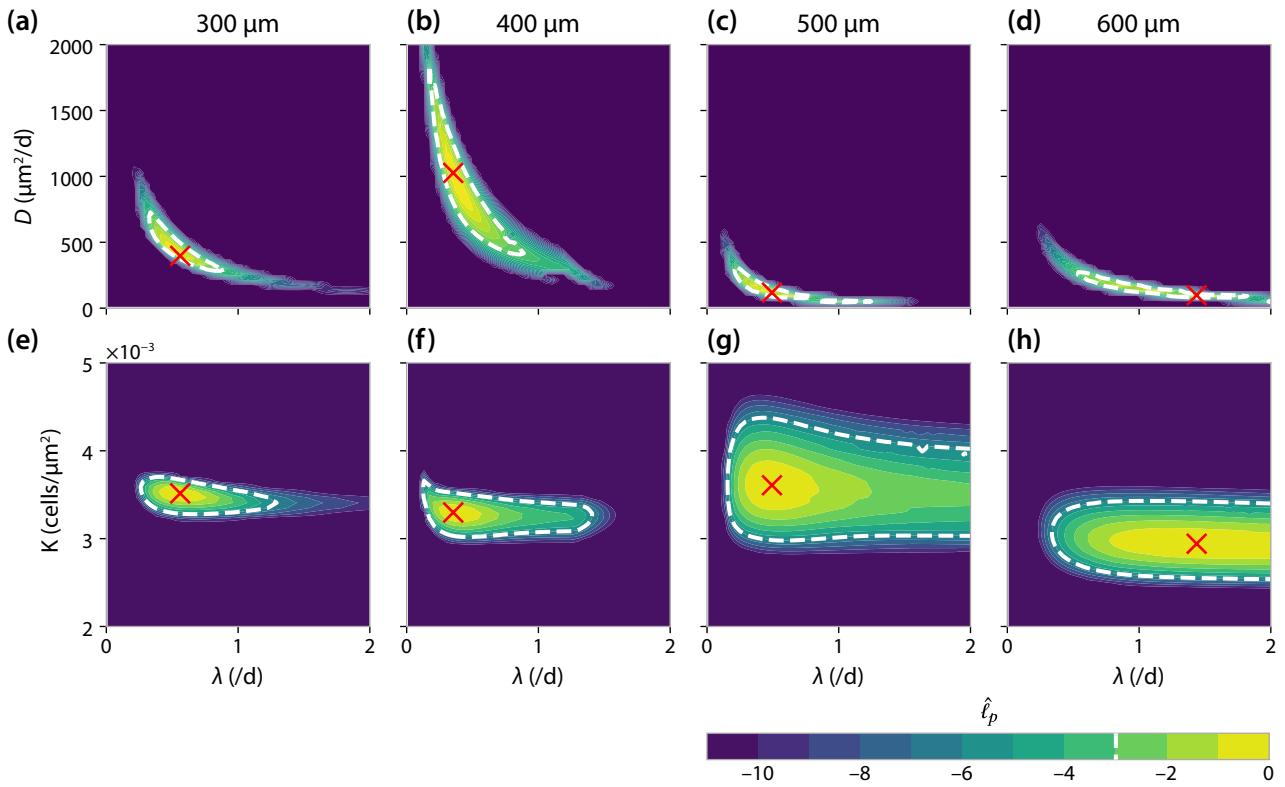
single set of parameters (figure 6.4). To allow for parameter uncertainty when comparing parameters across pore sizes, we compute profile likelihoods (figure 6.6) and approximate confidence intervals (Table 6.1) for each parameter. Although profile likelihoods must be interpreted with care given that they depend not only on the process model but also the noise model, they provide valuable information about the sensitivity of the likelihood estimates we obtain. This is important as the point estimates provided by the maximum likelihood estimate give no information about parameter uncertainty, sensitivity or identifiability [24]. Although point estimates for each parameter appear to vary between pore sizes (Table 6.1), confidence intervals (Table 6.1) and likelihood profiles (figure 6.6) largely overlap, providing no evidence that these parameters vary across pore sizes. MLEs obtained for the diffusivity, D, from cell density information alone are much larger than we might expect, but examination of the profile likelihoods, which provide a lower, but no upper, bounded confidence interval, indicates that D is cannot be established unless information relating to tissue coverage is included. The largest discrepancy between pore sizes is seen in the diffusivity: estimates range from 400 to 1000 $\mu m^2 \, d^{-1}$, for the 300 and 400 μm pores, respectively (the larger variability and inconsistencies in data for the 400 μm pores leads to a much wider confidence interval than for the 300 μm pores), to below 200 $\mu m^2 \, d^{-1}$ for the 500 and 600 μm pores. This variability is consistent with estimates for cell diffusivities in two-dimensional culture, which often vary over several magnitudes across experimental conditions [65, 297].

Estimates, profiles and confidence intervals for the proliferation rate, λ, are remarkably consistent between pore sizes. While the model does not capture the shape of the tissue boundary, it does capture both the cell density and tissue coverage, suggesting that the crowding effects which lead to logistic growth in the experiments are also captured. In particular, our results in figure 6.6 suggest proliferation of MC3T3-E1 is similar between scaffolds of different sizes and is lower than a rate of $\lambda \approx 1.1 \, d^{-1}$ observed in two-dimensional culture [409] (this is also seen in figure 6.7). Another interesting result is the consistency in estimates of carrying capacity, K, of approximately 0.00345 (95% combined CI: (0.00336,0.00355)), which corresponds to an average packing density where cells occupy the same amount of space as a disc with diameter of 19 μm, slightly lower than, and, therefore, consistent with a tissue comprising a monolayer of cells [419]. An exception is for the largest 600 μm pore, which produces an estimate much lower than the other pores (95% CI: (0.0027,0.0032)). While this lower estimate may be consistent with average cell density observations (figure 6.4d), the higher estimate from the combined MLE is more consistent with behaviour at the edge of the pore (figure 6.4p). In some cases, the assumption of a constant carrying capacity across the entire pore may not be appropriate. It is not clear from the data alone whether this observation is due to actual variation in carrying capacity within a pore, or because net cell growth in the centre of the pore has not yet plateaued due to crowding effects. To answer this question, data must be collected over a longer experimental duration for these larger pores.

In figure 6.7 we compute bivariate profiles to assess potential relationships between parameter estimates. First, examining the bivariate profiles between the proliferation rate, λ, and diffusivity, D, in figure 6.7a–d, reveals a hyperbolic relationship. This result is consistent with previous studies that establish only the product $D\lambda$ using information about the position of the tissue interface [13, 38], but that cannot establish individual values for these parameters. In our work, by using information relating to both cell density and tissue coverage, we are able to establish the individual values of D and λ within a region of compact support (a 95% confidence region is shown in figure 6.7a–d). Second, examining the bivariate profiles between the proliferation rate, λ, and carrying capacity, K, highlights the information obtainable from the 28 day

▲ Figure 6.6. Profile likelihoods for calibrated model parameters. Profile likelihoods for each inferred parameter and pore size where only the density is used (dashed) and where both the density and coverage are used (solid). Dotted horizontal black line indicates the −1.92 contour which corresponds to an asymptotic 95% confidence interval for each parameter. Also shown are profiles for D, λ and K where information relating to cell density and tissue coverage information from all pores is included simultaneously.

▲ Figure 6.7. **Bivariate profile likelihoods for calibrated model parameters.** Bivariate profile likelihoods showing the relationship between estimates for (a–d) D and λ, and (e–h) λ and K. Dotted white lines indicates the -3.00 contours which corresponds to an approximate asymptotic 95% confidence region for each parameter combination.

experiment for each pore size. On average, the larger 500 μm and 600 μm pores do not bridge by the conclusion of the experiment, and we see comparatively large uncertainties in both the estimated proliferation rate and estimated carrying capacity (figure 6.7g,h). In contrast, results for the smaller 300 μm and 400 μm pores—the majority of which bridge by day 18—show that we are able to establish these parameters with a relatively small region (figure 6.7e,f). Although point estimates for the proliferation rate vary across pore sizes (Table 6.1), the bivariate profiles show a significant overlap in possible parameter values, indicating that these parameters are similar between pore sizes.

When the data are analysed as summary statistics that depend upon time, as in figure 6.4, we see a large amount of variability that cannot be fully captured by the observation noise process we define in Section 6.2.2. However, when relationships between summary statistics of each pore are analysed with respect to each other, independent of time, we see notably less variability (figure 6.5). These results suggest that time alone is a poor predictor of each summary statistic. In contrast, the dependence between summary statistics in figure 6.5 suggests that summary statistics have well defined relationships with relatively little variability. In the deterministic process model, the initial condition (which describes the expected value of each summary statistic on day 4) is taken to be a fixed density of cells distributed around the fibres. The majority of the variability in the temporal pore bridging data may be due to variability in the initial condition, which affects initial pore formation. For example, we expect tissue growth to be slower, or stationary, for pores that initially have a smaller density of cells distributed around the pore boundary at day 4. One way around the limitation of providing a homogeneous initial distribution of the cells in the experiments is to collect time-series data, where the same pore is imaged at multiple time points. There are two ways this information

could be incorporated into the mathematical model. First, by including a time delay parameter for each data point that describes the delay until tissue formation inside the pore begins, that can be profiled out as a nuisance parameter in the analysis. Second, by capturing the variability directly by describing pore bridging as a differential equation where the initial density at the pore boundaries is a random variable.

Our results do not suggest significant differences in cell behaviour between pore sizes. Despite the Porous-Fisher model not explicitly incorporating geometric behaviour (aside from the initial and boundary conditions), we can capture information relating to both tissue coverage and cell density even when calibrated simultaneously to data from all pore sizes. By accounting for tissue coverage, we quantify a similar proliferation rate for all sizes based on a logistic growth assumption. The relationship between tissue boundary circularity and coverage is similar between all pore sizes. In all pores we see two stages of bridging: first, the corners bridge—this takes longer in the larger pores—and form an approximately circular tissue boundary; second, the pore closes and remains approximately circular in shape. These observations have also been made for triangular and hexagonally shaped pores [394], and convex pores [386]. Further experimental and modelling work is needed to disentangle the effect of each of these stages on overall pore bridging and tissue growth. For example, we suggest experimental work that investigates corner bridging and tissue establishment using non-constrained or "open" geometries [391, 396], rather than the current geometry where tissue growth eventually closes a pore of finite size. To reduce overall variability in the data, variability in the initial condition should be accounted for through time-series imaging, where information about each pore is available at multiple time-points, and throughout each distinct stage.

In the present work we assume that motility is described by the non-linear diffusion term $\mathscr{D}(u) = Du/K$, which ceases migration in regions where $u = 0$ and increases to D as $u \to K$. Identifying the dependence of motility on cell density will allow for calibration of more general models of the form

$$\frac{\partial u}{\partial t} = \nabla \cdot [\mathscr{D}(u)\nabla u] + \lambda u \left(1 - \frac{u}{K}\right), \quad \mathrm{x} \in \Omega. \tag{6.20}$$

For example, setting

$$\mathscr{D}(u) = Du \left(1 - \frac{u}{K}\right), \tag{6.21}$$

captures inhibited migration as the tissue growth ceases where $u \to K$, so that migration is maximised at intermediate densities. However, we note that it is often the case that, as $u \to K$, $\nabla u \approx 0$ so we expect our characterisation of diffusivity $\mathscr{D}(u) = Du/K$ to be sufficient in high-density regions.

With the information available, our work was not able to identify the mechanisms that lead to corner bridging and the circular shape of the tissue boundary. The migration of cells from the fibres to the pore interior may be captured more accurately through the generalised boundary conditions

$$a(u) + b(u)\nabla u \cdot \hat{\mathrm{n}} = 0, \tag{6.22}$$

where $\hat{\mathrm{n}}$ is a unit vector normal to the fibre, and $a(u)$ and $b(u)$ are functions that describe the behaviour of cells at the pore boundaries. However, we expect parameter identifiability issues to become more pronounced by the introduction of the functions $a(u)$ and $b(u)$ that depend on u and, therefore, implicitly on both location and time.

Our thin three-dimensional experimental framework, and two-dimensional depth-averaged mathematical

modelling framework, carry several advantages over more complex alternatives. In addition to information relating to tissue coverage, we are able to access detailed information about cell density, which we interpret with a mathematical model that quantifies cell behaviour with biophysical parameters such as proliferation and migration rates. This allows for comparison of cell behaviour between cell lines, allowing tissue growth optimisation with respect to cell line in addition to scaffold geometry. Our modelling framework is also extensible to co-culture systems that include multiple cell lines, which are more representative of *in vivo* tissue growth, through a coupled system of partial differential equations. Additionally, working with a thin three-dimensional experimental geometry reduces the need to account for additional extraneous factors on cell growth, such as nutrient availability. For example, typical *in vitro* three-dimensional tissue culture lack the vascular system that ensures homogeneous nutrient availability for *in vivo* tissue growth [12]. In comparison, our geometry results in a monolayer of cells that are all in direct contact with growth medium.

We suggest, in future, a hybrid modelling framework to describe each stage of pore bridging, rather than a single model that captures all stages of growth. While our analysis does not preclude generalisations of the Porous-Fisher model from capturing geometric features like circularity, reaction-diffusion models alone cannot account for both the corner bridging and pore closing stages of growth we see in the experimental data. Models based on continuum mechanics or curvature control have been successful in recapturing the initial stages of bridging seen experimental data [386, 394, 397], but typically neglect information relating to cell density. Once a circular tissue boundary is established, tissue growth may be quantified using density-dependent models such as those based on the Porous-Fisher equation, or agent based models [30, 420].

6.4 Conclusion and Outlook

We analyse experimental data from a series of pore bridging experiments using a relatively simple reaction-diffusion model based on the Porous-Fisher equation. In addition to commonly reported tissue coverage information, our model allows for the interpretation of information relating to cell density, and we see a clear value in considering both measurements. For example, the cell migration rate is often unidentifiable from information relating to cell density alone but becomes identifiable when information relating to tissue coverage is included. Compared to existing models of tissue growth that are largely phenomenological [391, 392], our framework characterises cell behaviour with parameters that carry a biologically meaningful interpretation, such as cell proliferation and migration rates. We find no evidence to suggest that cell behaviour is dependent upon pore size. The cell proliferation rates, which are lower than that observed for two-dimensional culture, and carrying capacities are found to be remarkably similar across different pore sizes. This outcome suggests that our experimental protocols lead to consistent, reproducible tissue growth. This conclusion is not apparent without interpretation of the experimental data with a mechanistic mathematical model.

Our analysis identifies two distinct stages of pore bridging that are consistent between pore sizes: an initial corner bridging stage, and a latter hole closing stage. The Porous-Fisher model does not describe the initial corner bridging stage and, therefore, does not reproduce the shape of the tissue boundary. However, the model does match features relating to cell population and tissue coverage, thus capturing crowding effects and providing confidence in the estimated cell proliferation rates. We suggest that a better understanding of pore bridging can be formed through distinct theoretical models and experimental analysis that individually

capture both the corner bridging and hole closing stages.

The experimental data used for model calibration suggests, at first, that pore bridging is a highly variable process. However, analysis of the relationships between summary statistics reveals this may not be the case. Rather, variability in both the initial distribution of cells on the scaffolds and corner bridging leads to a time-delay that cannot be accounted for with the information available from our data-collection method. These results highlight a potential value in designing an experiment to collect time-series observations, which will provide information about cell density and tissue coverage of each pore at multiple time points. This more detailed information will allow for the inclusion of more complicated mechanisms, such as directed migration through chemotaxis [18, 374], mechanical effects at the tissue boundary [86, 421], or the depletion of nutrients available to the cell population. At present, we find the complexity of the mathematical model is well suited to the level of information available in the experimental data, and we expect identifiability issues to arise if we were to interpret the current data with a more complex model.

Many of our conclusions could not have been made without considering data from multiple experimental geometries. The smaller pores, for example, give the impression that the model captures geometric features of pore closing; the inability of the model to capture these features is only evident when we analyse data for the larger pores. Comparing parameter estimates and profile likelihoods across experimental conditions is essential for constructing and verifying theoretical descriptions of pore bridging. Typical applications of mechanistic mathematical models to understand tissue formation usually involve working with a single experimental geometry, most often in a one-dimensional setting. These approaches cannot provide insight into the effect of high-dimensional geometric phenomena, such as corners, which we explore in our work.

In conclusion, our Porous-Fisher model successfully captures many of the key features of the experiments, providing a straightforward means of interpreting experimental observations in terms of the underlying cell proliferation and migration mechanisms that drive tissue growth. To the best of our knowledge, these mechanisms have never before been explicitly characterised for tissue growth in 3D-printed scaffolds.

Chapter 7

Conclusion

Mathematical models are routinely applied in biology and bioengineering to design experiments, interpret experimental data, and achieve control over biological systems. Tailoring model complexity to the availability of experimental data and the biological questions of interest is critical to avoid issues relating to parameter identifiability.

In this thesis, we apply ordinary, partial and stochastic differential equations, along with individual-based models and optimal control, to describe collective cell behaviour. The question of which model to select is not—and likely never will be—clear. However, our choices are always pragmatic and motivated directly by the availability of current or future experimental data. This approach avoids issues relating to parameter identifiability, ensuring the validity of model predictions and enabling future experimental work.

In this chapter, we outline the primary outcomes of this thesis and how they contribute to the literature. These outcomes relate to model choice and construction, to statistical techniques for application to experimental data, and to experimental insights that guide experimental design and work.

7.1 Summary and Contribution

In Chapter 2, motivated by the availability of highly detailed cell-positional information, we develop and calibrate an individual-based model. Our goals are to identify the nature of cell-to-cell interactions that give rise to the density-dependent collective behaviour seen in the experiments. As our questions relate to the net effect of interactions—whether they increase or decrease net motility or proliferation—we form a simplified model where interactions are described using kernel functions. Calibrating our model to experimental data allows us to identify both the strength and sign of these kernels, indicating with a single parameter whether proliferation, motility or a directional movement bias increase or decrease with density. We apply model selection to identify the simplest model that simultaneously matches several data sets, each initiated at a different cell density. In opposition to many typical models of cell behaviour, our analysis suggests that proliferation is relatively unaffected by cell density, whereas density-dependent motility is essential.

Overall, our analysis in Chapter 2 indicates that fundamental modelling assumptions relating to cell motility must be revisited. Whereas it is routine for linear-diffusion to be applied to describe cell motility, we suggest

a more appropriate model of cell motility is of the form

$$\frac{\partial u}{\partial t} = \underbrace{\frac{\partial}{\partial x}\left(\mathcal{D}(u)\frac{\partial u}{\partial x}\right)}_{\text{Contact stimulated motility}}, \quad \mathcal{D}'(u) > 0. \tag{7.1}$$

where $\mathcal{D}(u)$ is an increasing function of cell density, u. These findings strongly influence our model of cell motility in Chapter 6, where we set $\mathcal{D}(u) = Du$ (i.e., the Porous-Fisher equation) to describe contact-stimulated motility. Another significant result in Chapter 2 is that even from the highly detailed experimental data we consider, the effect of cell-to-cell interactions on proliferation is non-identifiable: models with and without density-dependent proliferation match the data. While we cannot conclude from our analysis what is required to identify the effect of interactions on proliferation, other studies suggest that experiment duration significantly affects identifiability of growth parameters [71].

Typical experimental protocols do not vary the initial number of cells, so models that are calibrated to experimental data are typically only done so with a single initial condition, leaving questions of over-fitting and parameter identifiability unanswered. For example, Jin *et al.* [40] find that calibrating one of the most common mean-field models of cell migration—the Fisher-Kolmogorov equation—to experimental data across a range of densities revealed discrepancies in estimates of cell motility, motivating our research in Chapter 2 and Chapter 6. Mathematically, it is typical for the initial condition to be modelled separately from the transient dynamics. This approach implicitly assumes that the transient dynamics, and hence the model parameters, are independent of the initial condition in the experiments. Therefore, it is important for model validation to compare parameter estimates to experimental data initiated across a range of densities. Consistency in parameter estimates indicates consistency in cell behaviour between initial densities and provides evidence of model appropriateness or, in the case of the discrepancy noted by Jin *et al.*, indicates areas for model refinement.

The success of applying kernel functions to describe cell-to-cell interactions in Chapter 2 motivates us to develop similar techniques to describe cellular decision making in a tractable modelling framework. In Chapter 3, we apply optimal control theory to model cellular decision making by bacteria in volatile environments, introducing our idea of *cellular hedging* to the literature. This approach invokes the assumption that the strategic response of cells to volatility optimises their productivity. While this assumption has been invoked in the literature before [22], our approach allows the characterisation of a genuinely adaptive strategy that depends explicitly on the external environment. Moreover, the application of the control alleviates the need to model the complex molecular-level mechanisms that give rise to decision-making strategies, ensuring that the model remains identifiable while still providing biological insight.

A fundamental result of our analysis in Chapter 3 is to show that bacterial persistence is only advantageous in the presence of environmental volatility. Yet, existing models of environmental variability in the literature are deterministic or a random sequence of discrete shocks [22, 103, 123, 124]. In Chapter 3 and the associated Appendix, we develop five unique models of volatility described using stochastic differential equations. We show that continuously fluctuating environments produce qualitatively different optimal responses than a sequence of discrete shocks. These observations draw parallels to so-called *evolutionarily-informed* therapies such as drug sequencing [146] that exploit the natural adaption of cells to environmental conditions to

produce a more favourable clinical response to therapy. Given that bacterial persistence is thought to be responsible for the formation and incurability of many infections [110, 111], a better understanding of how these strategies arise is vital to future clinical research.

A prerequisite for accurate parameter estimation is that model parameters be identifiable. Despite the ubiquity of stochasticity in biology—including in the study of bacterial persistence in Chapter 3—there are no commonly adopted methodologies to diagnose parameter identifiability in SDE models. Consequently, it is rarely apparent at the outset of analysis that adding to complexity by modelling noise will be more informative than a deterministic approach. In Chapter 4, we demonstrate how existing techniques for the analysis of ODE models can be applied to SDE models in systems biology. This analysis is critical to model and experimental design. For example, identifiability analysis can rapidly discern the quality and quantity of experimental data required to adequately estimate model parameters. When model parameters are structurally non-identifiable, it is often suggested in the literature that the model should be re-parameterised or reduced in complexity [273, 342]. Our analysis in Chapter 4 explicitly demonstrates that, in some instances, non-identifiability can be rectified by leveraging information in the noise and moving from an ODE to an SDE description. These insights are crucial for tailoring the mathematical model to the available data.

In Chapter 5 and Chapter 6, we apply mathematical models to two applications in bioengineering. In this context, detailed quantification of cell behaviour through predictive models is paramount to allow for exploitation and accurate control.

Our first application, in Chapter 5, is to introduce a theoretical biological subsystem that induces energy into a mechanical metamaterial. This chapter addresses a significant limitation of current signal transmitting metamaterials: limited by an initial energy state, these passive materials are single-use. In contrast, our metamaterial allows for reversible mechanical signal transmission. Through model simplification and travelling wave analysis, we identify the region of the parameter space that allows for signal transmission, placing constraints on the biological subsystem. Aside from theoretical manufacturing outcomes, this work yields mathematical analysis applicable more broadly to the mathematical biology community. For example, our research of this material reveals parallels between other fast-slow systems, such as the FitzHugh-Nagumo model for neuron excitation and relaxation [367–369, 372], meaning our analysis is broadly applicable to these systems. While work in Chapter 5 is purely theoretical, results demonstrate how a new class of biologically inspired metamaterials can produce useful new functionalities.

Our second application, in Chapter 6, is to quantify collective cell behaviour in 3D printed scaffolds. Such quantification is vital to answer key engineering questions to control and optimise tissue growth, allowing for future clinical applications that replace synthetic grafts with artificially regenerated tissues. Motivated by the findings in Chapter 2, we describe cell motility using the Porous-Fisher equation and validate the model by comparing parameter estimates across experimental conditions. These results suggest that cell proliferation and the maximal cell packing density are relatively insensitive to physical changes in scaffold geometry. Analogous to observations that motivate our study in Chapter 2 [40], we find and diagnose evidence of model misspecification, observing that the Porous-Fisher model is unable to capture the shape of the growing tissue. Our work, therefore, provides vital modelling guidance not only for the application to tissue growth in these scaffolds but more generally for mathematical models of collective migration in high-dimensional geometries.

Work in this thesis develops data-driven models in mathematical biology, with applications in biology

and bioengineering. We demonstrate how models can identify cell-to-cell interactions and are the first to bring stochastic optimal control theory to study decision making in biology. To guide model and experimental design, we elucidate identifiability analysis techniques for the stochastic models that are fast becoming commonplace through mathematical biology. Through two novel applications to bioengineering, we demonstrate how collective cell behaviour can be exploited—guided by mathematical modelling—to create functional new materials and synthetic skin grafts. Overall, we illustrate how mathematical models can be applied to deal with complexity in biology and bioengineering.

7.2 Future work

The developments in this thesis enable several areas for new experimental and theoretical models in cell biology that we now outline.

- **Develop identifiability analysis tools for partial differential equation models.**
  Two of the major themes in this thesis are spatial and stochastic models, yet, aside from our work in Chapter 4, there is very little guidance on identifiability analysis for these two common classes of model. Recent theoretical work extends the differential algebra approach for structural identifiability analysis of ODE models to age-structured PDE models [422]. Applying this work to reaction-diffusion-type PDE models, such as that considered in Chapter 6, will allow for optimal experiment design and model construction, with application in and beyond the mathematical biology modelling community. Since structural identifiability is a prerequisite for practical identifiability [23], development of tools to rapidly analyse identifiability of PDE models under various measurements will alleviate some of the computational challenges in diagnosing these issues using inference techniques, as encountered in Chapter 6.

  Any development of identifiability analysis for PDE models has implicit outcomes for SDE models, the transient behaviour of which can be described using the Fokker-Plank equation, a time-dependent PDE. We expect, however, that dimensionality will remain a major challenge. Whereas structural identifiability analysis tools for ODEs readily extend to large systems of equations, an n-dimensional SDE can only be formulated into an n-dimensional PDE. Therefore, we expect the MCMC and moment-dynamics approach to identifiability we explore in Chapter 4 to remain viable options.

- **Exploit adaption of bacteria to environmental volatility to design antibiotic alternatives.**
  Optimal control is often applied from a clinician's perspective to design therapies that combat malignant growth [4]. These ideas show promise to build treatments that are highly effective, personalised, and robust against the unexpected [5]. One point of criticism for the application of optimal control in this context is model misspecification. For example, existing optimal control studies for the disinfection of persister-producing bacteria colonies assume simple dynamics where the bacteria do not adapt to the antibiotic treatment [149]. Analysis in Chapter 3 avoids this by describing the optimal response of bacteria to treatment, i.e., a clinically *worst-case* model.

  The understanding of how bacteria might respond to antibiotic treatment we provide in Chapter 3 forms the basis for further work that exploits the natural adaption of bacteria to create new types of treatment. Drug sequencing therapies [152] apply multiple drugs in sequence to maintain antibiotic

susceptibility in the colony. This parallels observations by Gefen *et al.* [153], who find that exposing a colony to fresh media reduces persistence. Applying optimal control onto the dynamical model we develop in Chapter 3 may reveal non-intuitive therapies that account for the adaptation of bacteria to disinfect a colony through a combination therapy that controls both the growth rate and drug input.

- Model cell-to-cell interactions using optimal control.

  In Chapter 2, we apply a deterministic kernel function to describe cell-to-cell interactions that affect proliferation or motility, and in Chapter 3, we apply a control that determines the reaction of cells to their external environment. In both cases, we develop an identifiable model by summarising complex, molecule-level mechanisms using a function that describes the outcome. These studies raise the possibility of modelling cell-to-cell interactions as a control to study the nature of mechanisms that maximise or achieve an objective. For example, cancer cells may develop cell-to-cell interactions that maximise the expected growth rate, whereas epithelial cells might interact in such a way as to optimise the mechanical structure of the resultant tissue. So-called *inverse optimal control* can be applied to learn the objective function from a model and experimental data [423], potentially providing additional insight that can characterise why a particular cell population exhibits the behaviour that it does.

- Incorporate spatial effects in optimal control model of bacterial persistence.
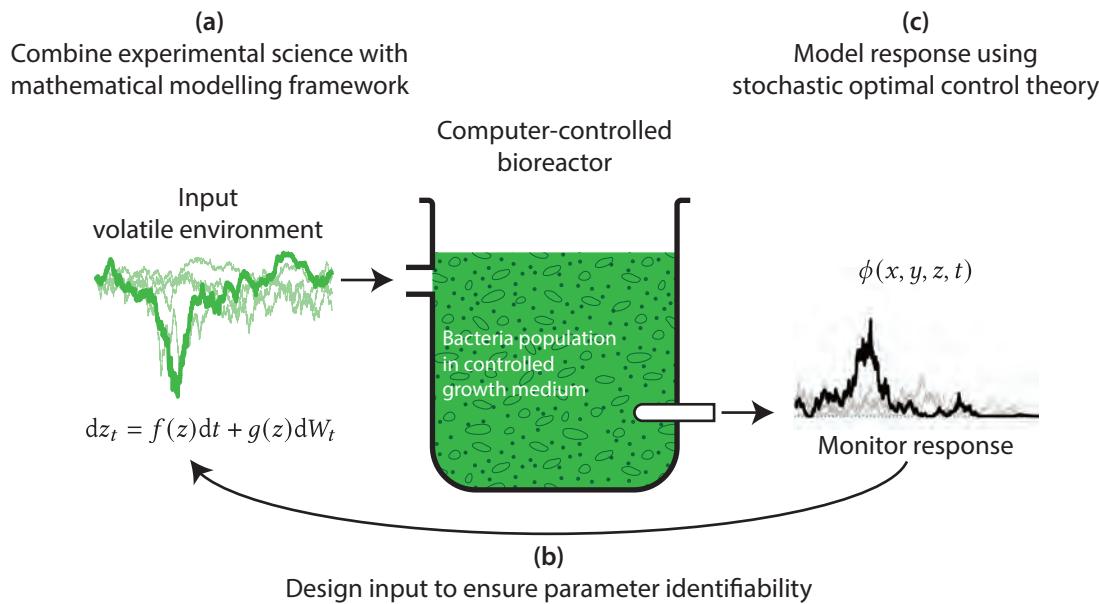
  In Chapter 3 we develop a space-averaged model to describe bacterial persistence using techniques from control theory. Given the computational complexity and nature of typical experimental data, in this initial piece of work we neglect spatial effects. In reality, spatial effects often play a crucial role in mechanisms such as quorum sensing, which enables bacteria to communicate and collectively respond to stimulus [154].

  One approach is to describe the behaviour of bacteria in a spatially-fluctuating environment using a stochastic PDE model, where chemical signals that regulate quorum sending and, hence, control signal input, are explicitly modelled [424]. However, applying optimal control to stochastic PDE models is not straightforward, and will come with a significant computational cost. An alternative approach is to apply the identified population-level control to an individual-based model, such as that explored in Chapter 2. The cell-to-cell interactions developed through kernel functions in Chapter 2 can be applied to describe the local prevalence of persisters to inform the control. In this approach, a diffusible chemical species, such as a species used for quorum sending, can be concurrently modelled [425].

- Apply optimal control theory to metamaterial model.

  Crucial to any potential implementation of the active metamaterial introduced in Chapter 5 is the ability to *control* the behaviour. In our work, we demonstrate that the material can selectively transmit and retransmit mechanical signals. However, it remains unclear whether concurrent or oscillatory signals are transmissible, or even compatible with the energy reset provided by the biological subsystem. Optimal control techniques (Chapter 3) can determine the required inputs to achieve, if possible, the aforementioned behaviours. As such techniques are typically based around an objective, they can also be applied to optimise energy expenditure and tune the material.

  Aside from theoretical outcomes relating to the material itself, optimal control analysis of the system studied in Chapter 5 has broader outcomes. For example, similar fast-slow systems are seen in neural

(a) Combine experimental science with mathematical modelling framework

(c) Model response using stochastic optimal control theory

Computer-controlled bioreactor

Input volatile environment

$\phi(x, y, z, t)$

Bacteria population in controlled growth medium

Monitor response

$dz_t = f(z)dt + g(z)dW_t$

(b) Design input to ensure parameter identifiability

▲ Figure 7.1. Combination of theoretical and experimental work to understand the adaption of biological systems to uncertainty. A volatile environment described by a known stochastic process, z, is provided as the input to a bioreactor. The response of the biological system is monitored and used to calibrate the mathematical model.

and cardiac excitation. In these systems, small perturbations and stochastic effects may play essential roles that must be accounted for to enable robust control. Here, interdisciplinary ideas developed in engineering [426] can be applied to study control in these biological systems.

- Design and conduct experimental study of bacterial persistence.

Recent experimental technologies, such as computer-controlled bioreactors [427], allow precise control of the nutrient and drug inflow in experiments. Mathematical work by Treloar et al. [428], for example, achieves control over bacteria growth in a bioreactor through reinforcement learning. These technologies open interdisciplinary opportunities to study cellular decision making under true uncertainty, quantified by the class of mathematical models developed in Chapter 3.

Bacteria are highly proliferative and rapidly adapt to environmental changes, including the introduction of antibiotics. Van den Bergh et al. [109] demonstrate that daily exposure to antibiotics increased colony survival by a factor of up to 300 after just three treatments. Exposing bacteria to a truly volatile environment, but an environment with known mathematical properties, will reveal how biological systems adapt to uncertainty. Work in Chapter 3 enables such studies, by providing mathematical tools to generate stochastic environments and the theory to predict an optimal evolutionary response (figure 7.1). The advantage of applying this cellular hedging approach over models that describe the molecular mechanisms that give rise to this behaviour [130–132] is model complexity and identifiability: the optimal control framework adds only a single tuning parameter to the already routine models of persistence [16]. Given that bacterial persistence is thought to be responsible for the formation and incurability of many common infections, a comprehensive understanding of how these cellular hedging strategies form will be of significant clinical importance.

- Incorporate stochastic effects in model of tissue growth.

A major conclusion of Chapter 6 is that much of the variability in pore bridging is due to the initial

condition. This mirrors results from Chapter 2, where experiments seeded at the same seeding density differ significantly in the initial number of cells. As the analysis in Chapter 2 concludes, this variability in the initial condition can significantly affect the behaviour and should be accounted for. Provided that time-lapse experimental data were available, there are two approaches to account for the variability in the initial condition in the model. First, to model the initial condition as a random variable in the existing, deterministic, PDE model. This approach neglects intrinsic variability in tissue growth and, therefore, may only be appropriate for larger pores or experiments where the average seeding density is higher. A second approach is to describe pore bridging using an individual-based model. Here, the initial locations of cells in the experiments can specify the initial condition in the model. A secondary advantage of this approach is to allow for the inclusion of more complex mechanisms, particularly at the domain boundaries where cells first migrate off fibres into the pore.

7.3 Final remarks

Alan Turing is famously quoted "The model will be a simplification and an idealization, and consequentially a falsification" [429]. The purpose of mathematical modelling is not to describe everything but to describe enough given the questions posed. This thesis develops new mathematical models and statistical techniques that enable future research and take full advantage of experimental data in biology and bioengineering.

Appendices

Appendix A

Supplementary material for Chapter 2, "Identifying density-dependent interactions in collective cell behaviour"
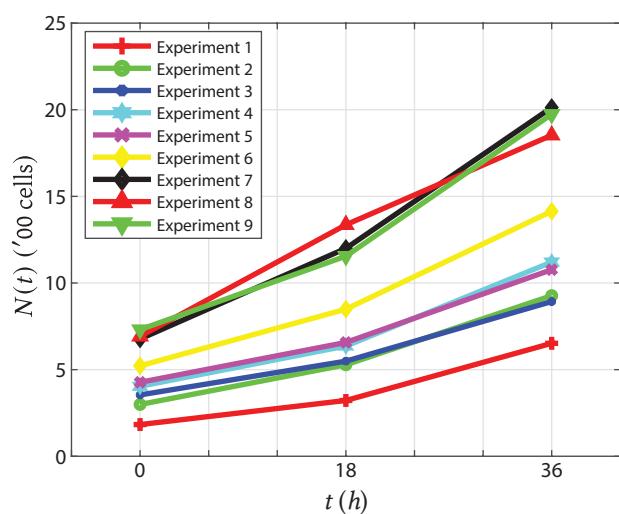
A.1 Code availability

Code used in this work available at `https://github.com/ap-browning/scratchIBM`.

A.2 Experimental data

Experiment	$N(0)$	$N(18)$	$N(36)$	Fold change
1	183	322	652	3.56
2	299	528	927	3.10
3	354	549	893	2.52
4	404	636	1121	2.77
5	427	657	1077	2.52
6	522	849	1414	2.71
7	677	1200	2013	2.97
8	692	1285	1853	2.67
9	731	1155	1974	2.70

◄ Table A.1. Summary of experimental data showing the cell count, $N(t)$, and fold change, $N(36)/N(0)$, for all nine experiments.

◄ Figure A.1. Summary of experimental data showing the cell count, $N(t)$, for all nine experiments.

A.3 ABC algorithms

Here, we present the ABC rejection algorithm (algorithm A.1) and ABC SMC algorithm (algorithm A.2).

A.3.1 ABC rejection algorithm

▼ **Algorithm A.1.** ABC rejection sampling algorithm.

1: Draw parameter samples from the joint prior $\boldsymbol{\theta}_j \sim \pi(\boldsymbol{\theta})$.

2: Set discrepancy of jth sample $\kappa_j = 0$, experiment index $i = 1$.

   2.1: Set agent locations, $\{x_n\}_{n=1}^{N(0)}$, to match experimental data $X_{\text{obs}}^{(i)}$ at $t = 0$.

   2.2: Simulate model with parameters $\boldsymbol{\theta}_j$ for $t \leq 36$, storing the agent locations at $t = 18$ h and $t = 36$ h, denoted $X_{\text{sim}}^{(i)}$.

   2.3: Update the discrepancy $\kappa_j \leftarrow \kappa_j + d(X_{\text{obs}}^{(i)}, X_{\text{sim}}^{(i)})$, where $d(\cdot, \cdot)$ is the discrepancy function.

   2.4: Move to the next replicate by setting $i = i + 1$ and repeat steps 2.1–2.4 until $i = 9$.

3: Repeat steps 1–2 until 10^5 samples $\{\boldsymbol{\theta}_j, \varepsilon_j\}_{j=1}^{10^5}$ are simulated.

4: Order $\{\boldsymbol{\theta}_j, \kappa_j\}_{j=1}^{10^5}$ by κ_j such that $\kappa_j < \kappa_{j+1}$.

5: Retain the first 1% ($\alpha = 0.01$) of prior samples $\boldsymbol{\theta}_j$, as posterior samples, $\{\boldsymbol{\theta}_j\}_{j=1}^{10^5 \alpha}$.

A.3.2 SMC algorithm

We apply the SMC model selection algorithm of Toni *et al.* [70], given in algorithm A.2. We choose the perturbation kernel to be a multivariate Gaussian with independent components and variances approximately equal to the ABC rejection posterior variances [81].
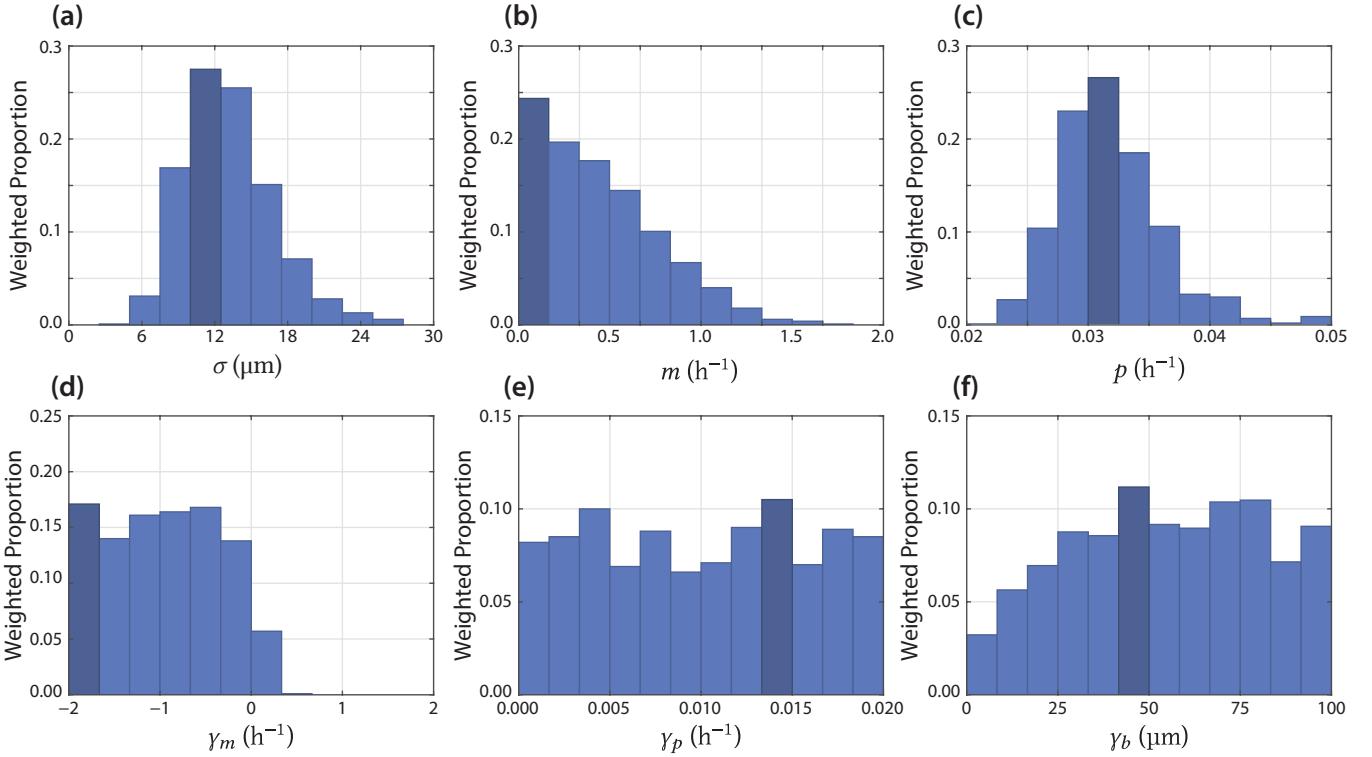
▼ **Algorithm A.2.** ABC SMC sampling algorithm for model selection with uniform priors [70].

1: Choose $\varepsilon_1, ..., \varepsilon_T$ such that $\varepsilon_k > \varepsilon_{k+1}$ and the desired total number of particles, N_{samples}. Set the population indicator $k = 1$.

2: Set the particle indicator $j = 1$.

    2.1: Sample model indicator, $M_a^* \sim \pi(M_a)$.

    2.2: If $k = 1$, sample proposal $\boldsymbol{\theta}^{**} \sim \pi_a(\boldsymbol{\theta})$ where $\pi_a(\boldsymbol{\theta})$ is the prior given model M_a^*. Go to step 2.4.

    2.3: If $k > 1$, sample $\boldsymbol{\theta}^*$ from the subset of the previous population of particles for M_a, $\Theta^{(a)}(k-1)$. If population is empty, return to 2.1. Perturb $\boldsymbol{\theta}^{**} \sim K(\boldsymbol{\theta}|\boldsymbol{\theta}^*)$, where $K(\cdot, \cdot)$ is a symmetric perturbation kernel. If $\pi_a(\boldsymbol{\theta}^{**}) = 0$, return to step 2.1.

    2.4: Set discrepancy $\kappa = 0$, experiment index $i = 1$.

        2.4.1: Set agent locations, $\{x_n\}_{n=1}^{N(0)}$, to match experimental data $X_{\text{obs}}^{(i)}$ at $t = 0$.

        2.4.2: Simulate model M_a with parameters $\boldsymbol{\theta}^{**}$ for $t \leq 36$, storing the agent locations at $t = 18$ h and $t = 36$ h, denoted $\mathcal{X}_{\text{sim}}^{(i)}$.

        2.4.3: Update the discrepancy $\kappa = \kappa + d(X_{\text{obs}}^{(i)}, X_{\text{sim}}^{(i)})$, where $d(\cdot, \cdot)$ is the discrepancy function.

        2.4.4: If $\kappa > \varepsilon_k$, reject particle and go back to 2.1. Else, move to next replicate by setting $i = i + 1$ and repeat steps 2.4.1–2.4.4 until $i = 9$.

    2.5: Add $\boldsymbol{\theta}^{**}$ to the population of particles $\Theta^{(a)}(k) = \{\boldsymbol{\theta}_j(k)\}_{j=1}^{N_{\text{samples}}^{(a)}}$, and calculate its weight as

$$
w_j^{(a)}(k) = \begin{cases} 1, & k = 1, \\ \left(\sum_{j=1}^{N_{\text{samples}}^{(a)}} w_j^{(a)}(k-1) K(\boldsymbol{\theta}^{**}|\boldsymbol{\theta}_j(k-1))\right)^{-1}, & k > 1. \end{cases}
$$

    2.6: Set $j = j + 1$ and repeat steps 2.1–2.6 until $j = N_{\text{samples}}$.

3. Set $k = k + 1$ and normalise the weights within each model. Repeat step 2 until $k = T$.

A.4 Pilot ABC results

### A.4.1 Pilot ABC to determine σ
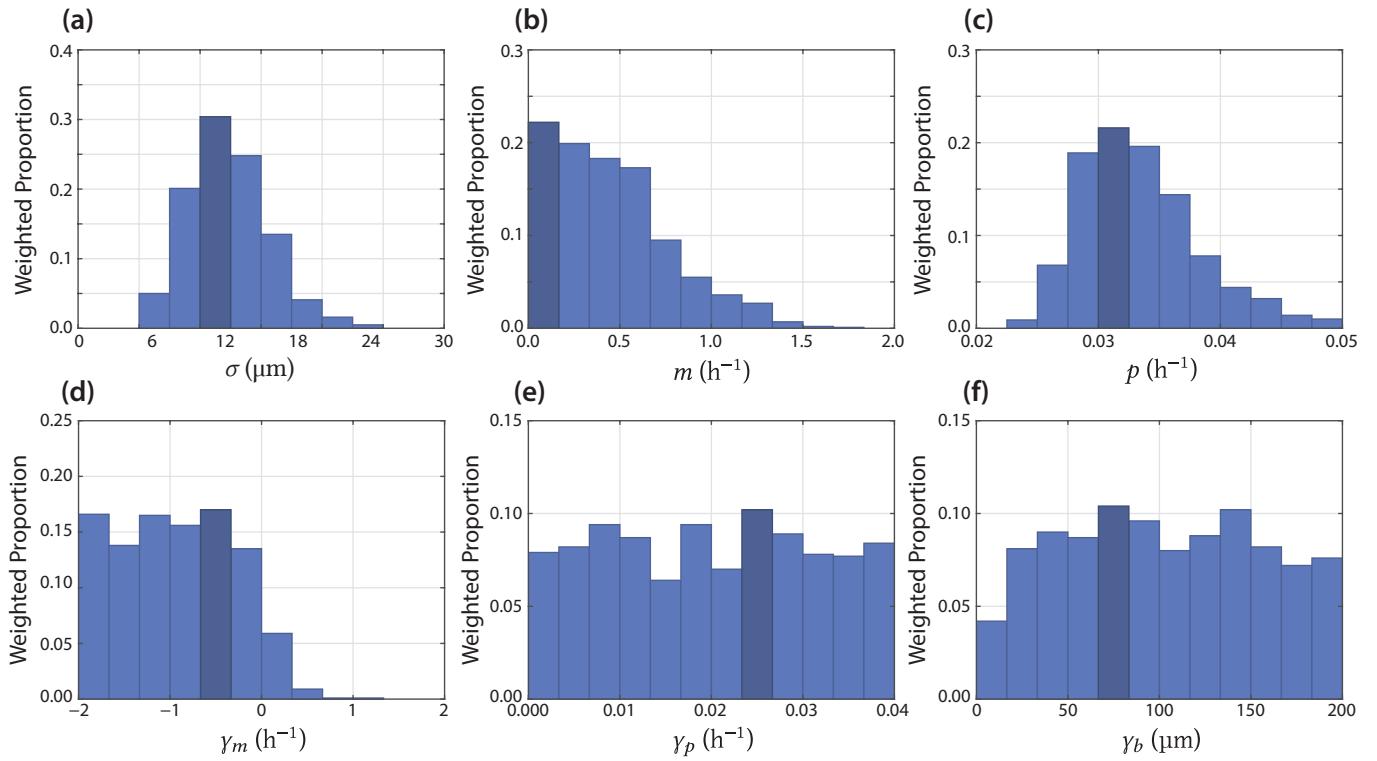
In order to reduce the number of unknown parameters, we estimate and fix the kernel width parameter, σ. To do this, we perform a pilot run using ABC rejection, where $\pi(\sigma) = U(2, 30)$ using algorithm A.1, the results of which are shown in figure A.2. These results show a posterior mode of approximately $\sigma \approx 12\,\mu m$, which we treat as constant for the rest of this study. In this appendix, we also produce some results for $\sigma = 24\,\mu m$ to explore the sensitivity of this choice (figures S4, S6, S9 and S10).

▲ Figure A.2. ABC rejection with σ as an unknown with $\pi(\sigma) = U(2, 30)$.

A.4.2 Wider priors

We notice in figure A.2e,f that the posterior support appears to cover the prior support. To investigate this, we widen the corresponding priors for γ_p and γ_b by a factor of two and perform ABC rejection. The results are shown in figure A.3. These results suggest that γ_p and γ_b are non-identifiable since the re-calculated posteriors are relatively flat and cover the same interval as the widened prior.
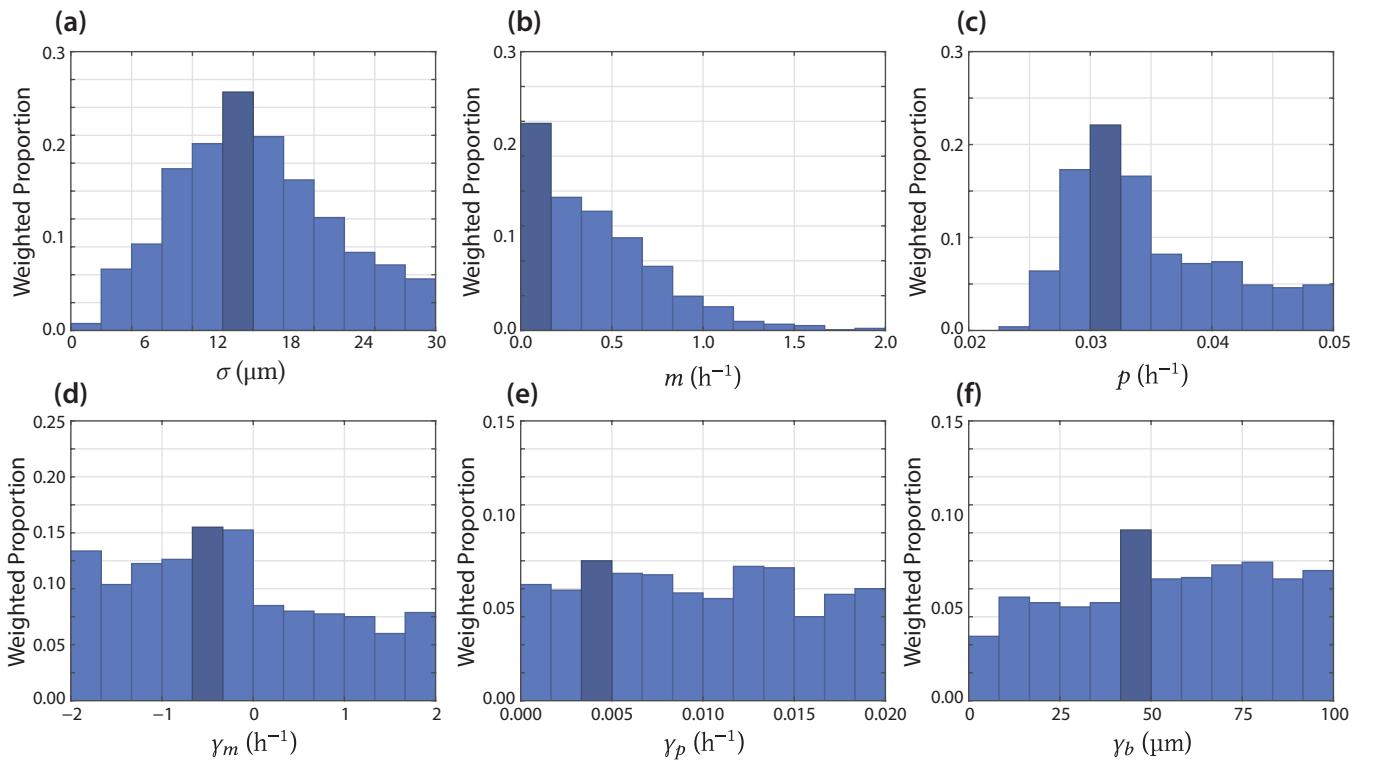
▲ Figure A.3. Figure A.2 with wider priors for γ_p and γ_b.

A.4.3 Excluding $\mathcal{P}$ from inference

To investigate the amount of information contained in the pair correlation function, $\mathcal{P}$, we perform ABC rejection in the case $\mathcal{P}$ is removed from the distance metric so that

$$d(X_{obs}, X_{sim}) = \sum_{t \in \{18,30\}} \left(\frac{[N_{sim}(t) - N_{obs}(t)]^2}{N_{obs}(t)^2} \right. \tag{A.1}$$

$$\left. + \frac{\sum_{j=I_{mid}-20}^{I_{mid}+20} [\mathcal{D}_{sim}(j,t) - \mathcal{D}_{obs}(j,t)]^2}{\sum_{j=I_{mid}-20}^{I_{mid}+20} \mathcal{D}_{obs}(j,t)^2} \right).$$
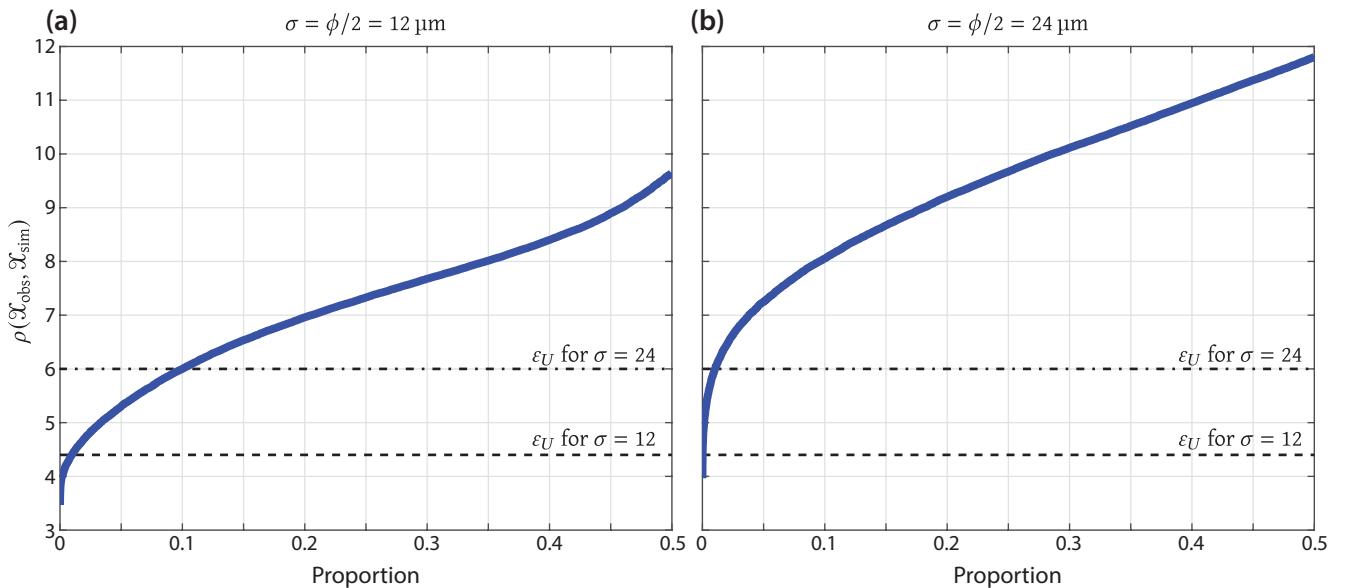
The results are shown in figure A.4. We see a large reduction in information in the posteriors in figure A.4 compared to figure A.2. In particularly, the sign of γ_m is less clear in the case the pair correlation is excluded.

▲ Figure A.4. Figure A.2 where summary statistics relating the the pair correlation, $\mathcal{P}$ are excluded.

A.5 Quantile plots

To determine the appropriate ABC SMC sequence of thresholds, we produce a quantile plot of the distance metric obtained from 100,000 prior samples where $\sigma = 12\,\mu m$ (figure A.5a) and $\sigma = 24\,\mu m$ (figure A.5b). In each case, we choose the final discrepancy, ε_U, to correspond to an ABC rejection rate of approximately 1%. We choose the sequence base upon acceptance probabilities of approximately 50%, 25%, 12.5%, 6.25%, 3.125%, 1.5625% and 1% [81]. The sequence of thresholds for results in the main chapter where $\sigma = 12\,\mu m$ is $\{9.6, 7.3, 6.3, 5.5, 4.9, 4.6, 4.4\}$; and the sequence of thresholds for results in this appendix where $\sigma = 24\,\mu m$ is $\{11.8, 9.7, 8.4, 7.5, 6.8, 6.3, 6.0\}$.
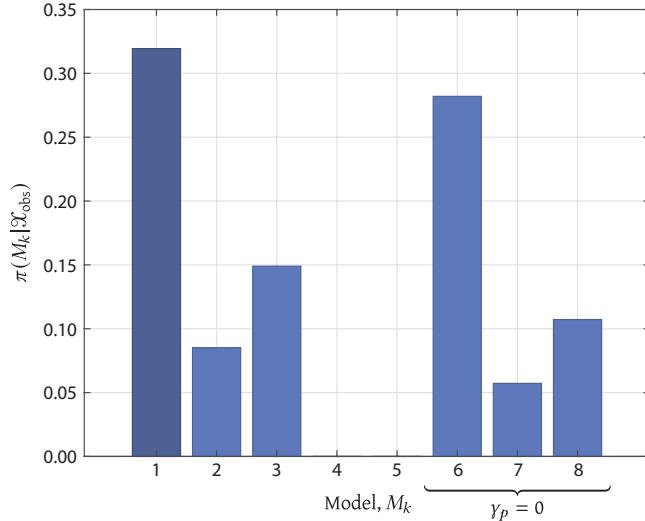
▲ **Figure A.5.** Quantile plot of the distance metric from 100,000 prior samples with (a) $\sigma = 12\,\mu m$; and, (b) $\sigma = 12\,\mu m$. The final discrepancy, ε_U, is indicated by a horizontal line in each case.

A.6 Exponential growth

To explore the possibility that $\gamma_p = 0$, which corresponds to exponential growth, we consider three additional models: Models 6, 7 and 8, that correspond to Models 1, 2 and 3, with $\gamma_p = 0$.

	$\boldsymbol{\theta}_k$	Density Dependence
Model 1	$(m, p, \gamma_m, \gamma_p, \gamma_b)$	Proliferation, Motility and Direction
Model 2	$(m, p, \gamma_p, \gamma_b)$	Proliferation and Direction
Model 3	$(m, p, \gamma_m, \gamma_p)$	Proliferation and Motility
Model 4	(m, p, γ_p)	Proliferation only (Fisher-Kolmogorov [61, 62])
Model 5	(m, p)	None (Skellam [80])
Model 6	$(m, p, \gamma_m, \gamma_b)$	Motility and Direction
Model 7	(m, p, γ_b)	Direction only
Model 8	(m, p, γ_m)	Motility only

▲ **Table A.2.** Here we consider three additional models: Models 6, 7 and 8 correspond to Models 1, 2 and 3, where $\gamma_p = 0$.

▲ **Figure A.6.** Figure 4 of the main chapter where we consider three additional models: Models 6, 7 and 8 correspond to Models 1, 2 and 3 where we set $\gamma_p = 0$ to remove the proliferation interaction from the model.
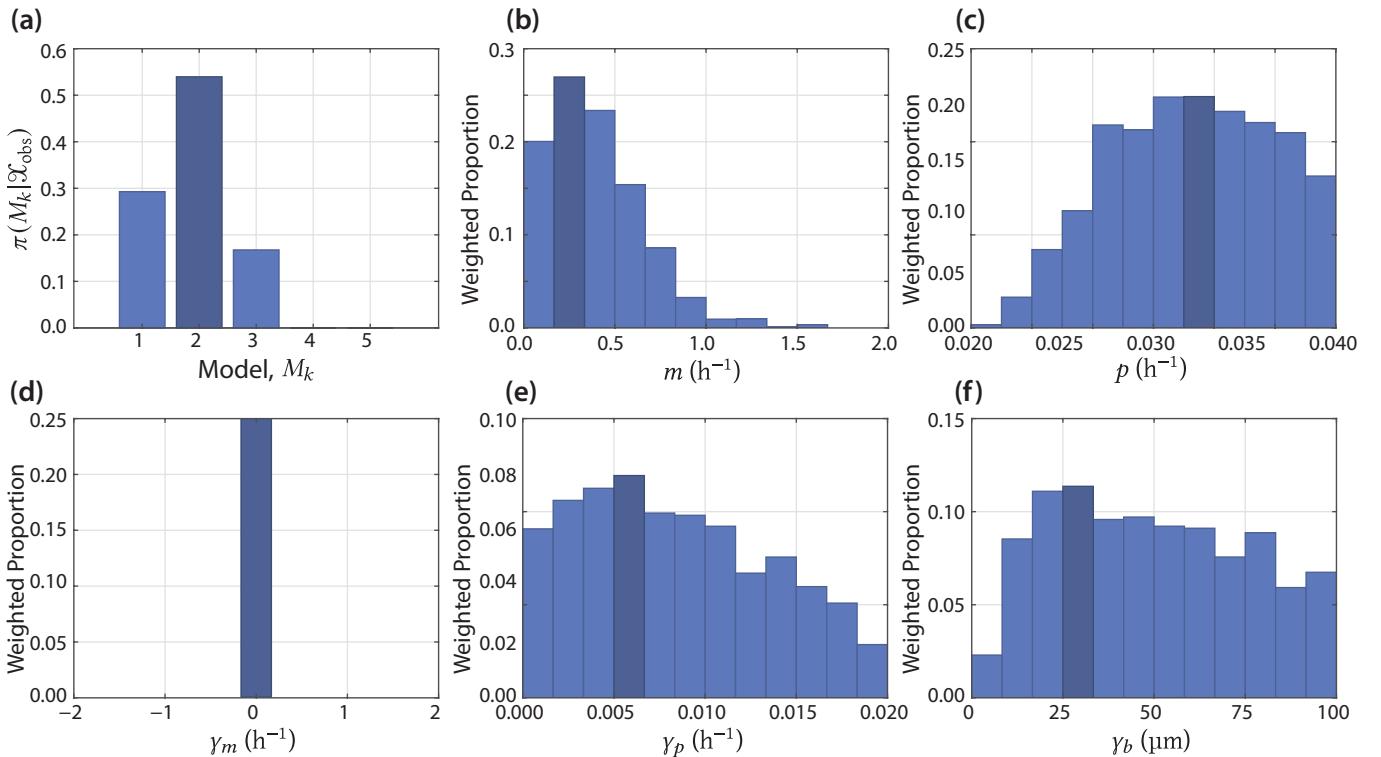
	Bayes factor, $\mathscr{B}_k$	Evidence in favour of Model 1
Model 1	1.000	—
Model 2	0.267	Positive
Model 3	0.467	Weak
Model 4	0.000	Very Strong
Model 5	0.000	Very Strong
Model 6	0.883	Weak
Model 7	0.179	Positive
Model 8	0.336	Weak

▲ Table A.3. Bayes factor for each model, which describes the evidence in favour of Model 1 over Model k. A Bayes factor close to 1 indicates limited evidence in favour of Model 1 over Model k, and a Bayes factor close to 0 indicates very strong evidence in favour of Model 1 over Model k [70]. Here, Models 6, 7 and 8 correspond to Models 1, 2 and 3, where $\gamma_p = 0$.

A.7 Model selection for $\sigma = 24\,\mu m$

Here, we reproduce results from figure 5 of the main chapter, in the case we fix $\sigma = 24\,\mu m$.

Results in figure A.7a differ from those in the main chapter, in that Model 2 now has the highest posterior density. We show the marginal distributions for each parameter in Model 2 in figure A.7b–f. These results are consistent with the main chapter in showing that models without a density dependent motility mechanism (Models 4 and 5) are unable to simultaneously match data from all nine experiments. Examining results in figure A.11 shows that $\sigma = 24\,\mu m$ is not able to match the spatial structure in the experimental data as closely as $\sigma = 12\,\mu m$, results for which are shown in figure A.9.
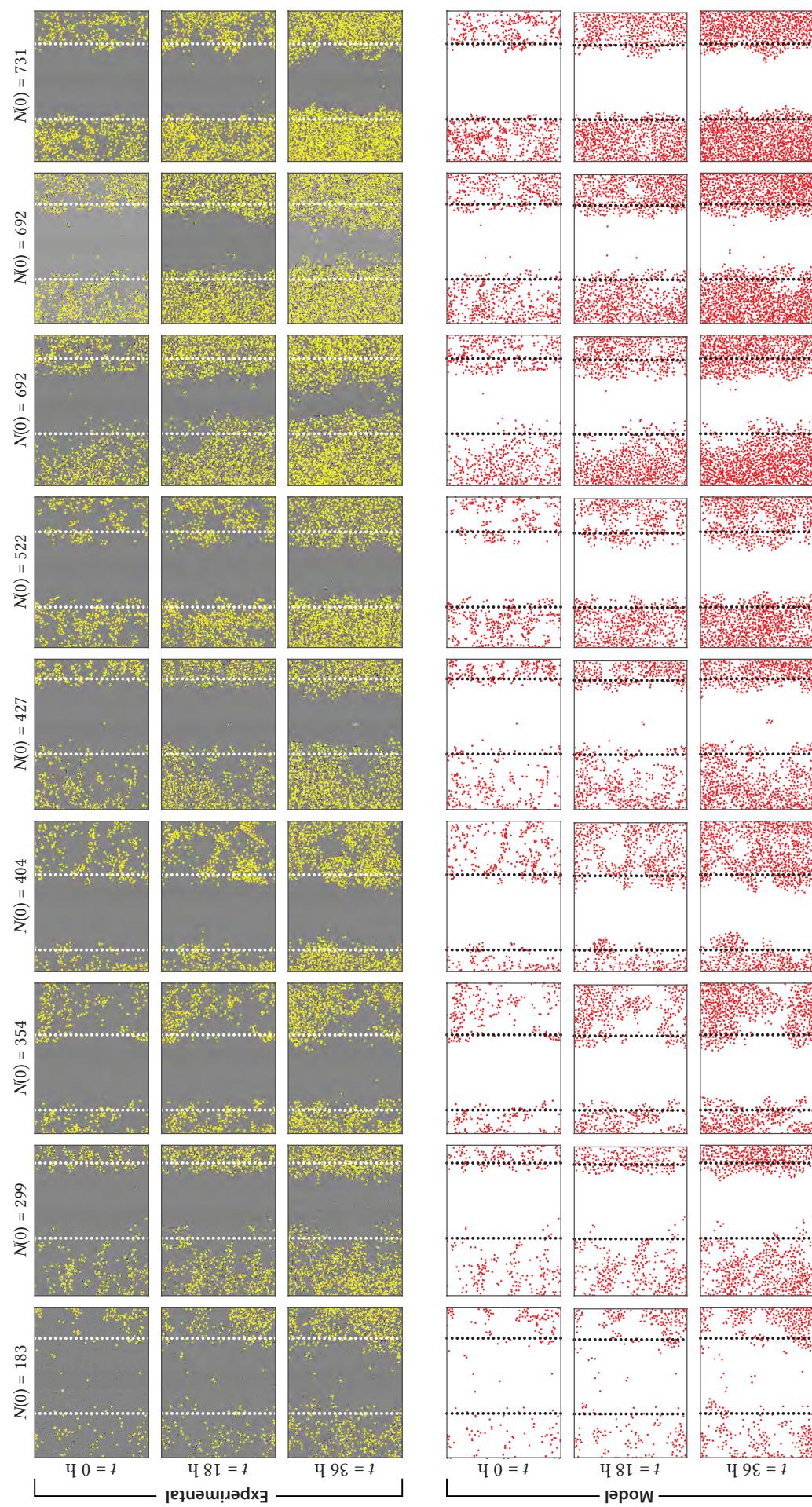
▲ Figure A.7. Reproduction of results in figure 6 of the main chapter, with $\sigma = \varphi = 24\,\mu m$. (a) Posterior for the model index, $\pi(M_k|\mathcal{X}_{\mathrm{obs}})$, showing that Model 2 (density-independent motility) as the posterior mode. (b)–(f) Marginal posterior distributions for each parameter in Model 2, shown as weighted histograms.
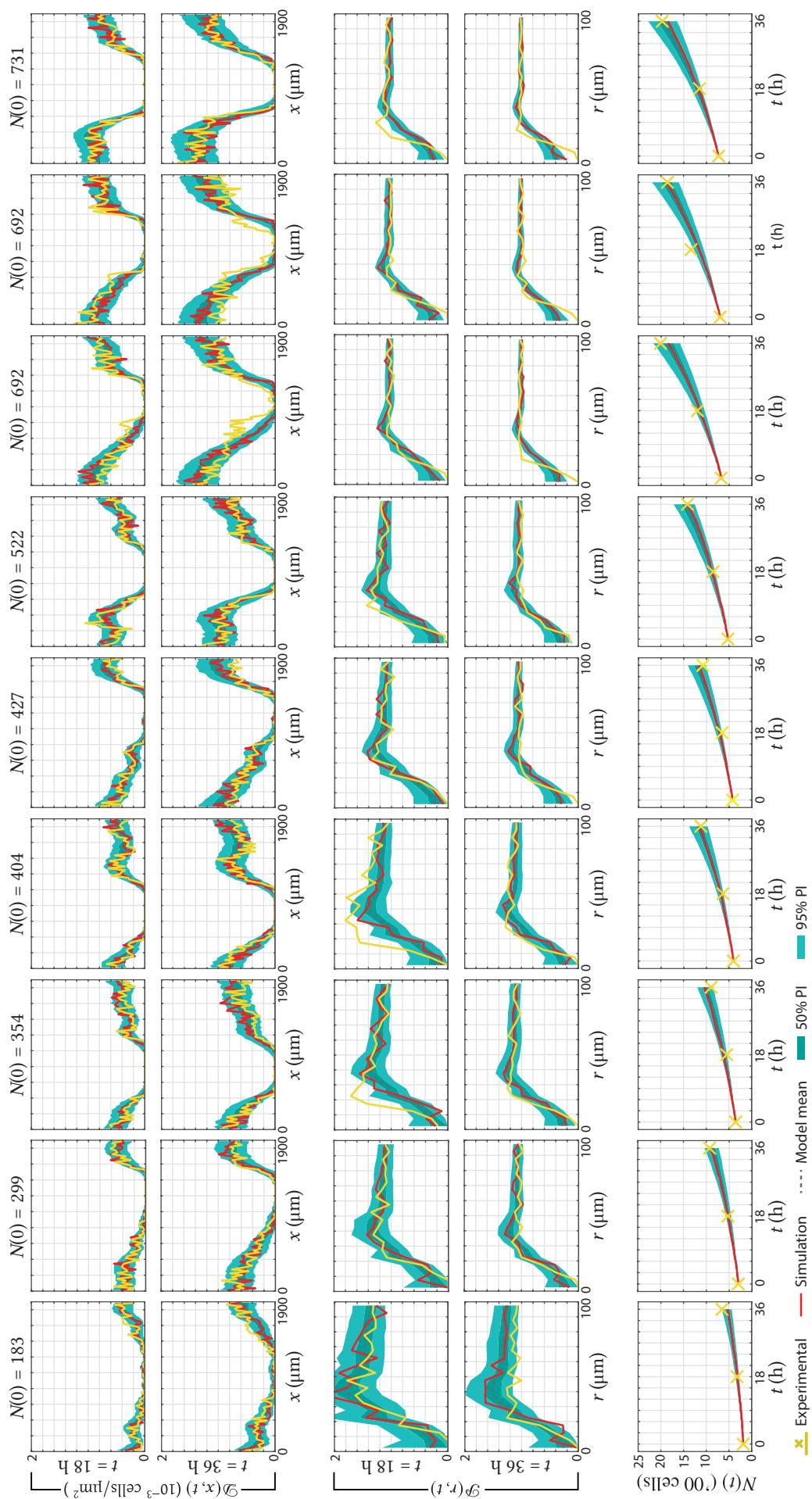
A.8 Results for all nine experiments

In the main chapter, we show results for experimental replicates 1, 3, 6 and 9 at $t = 36\,h$ in figure 6. Here, we reproduce figure 6 and show results for all nine experimental replicates at both $t = 18\,h$ and $t = 36\,h$.
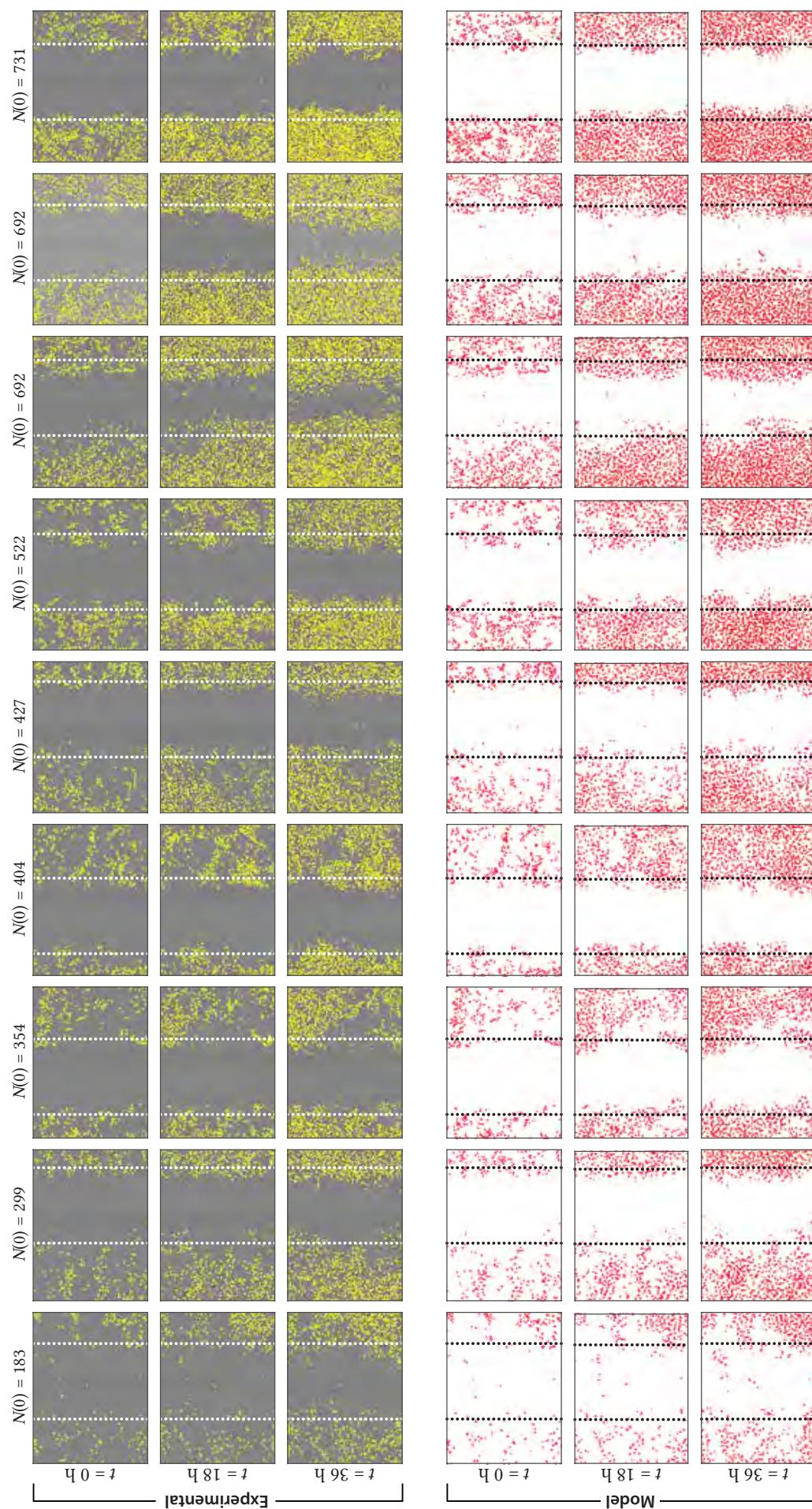
In figures A.8 and A.9 we show these results for $\sigma = 12\,\mu m$, the value from the main chapter. In figures A.10 and A.11 we show these results for $\sigma = 24\,\mu m$.

▲ Figure A.8. Reproduction of results in figure 6a–l of the main chapter, for all replicates, for $\sigma = 12$ µm.
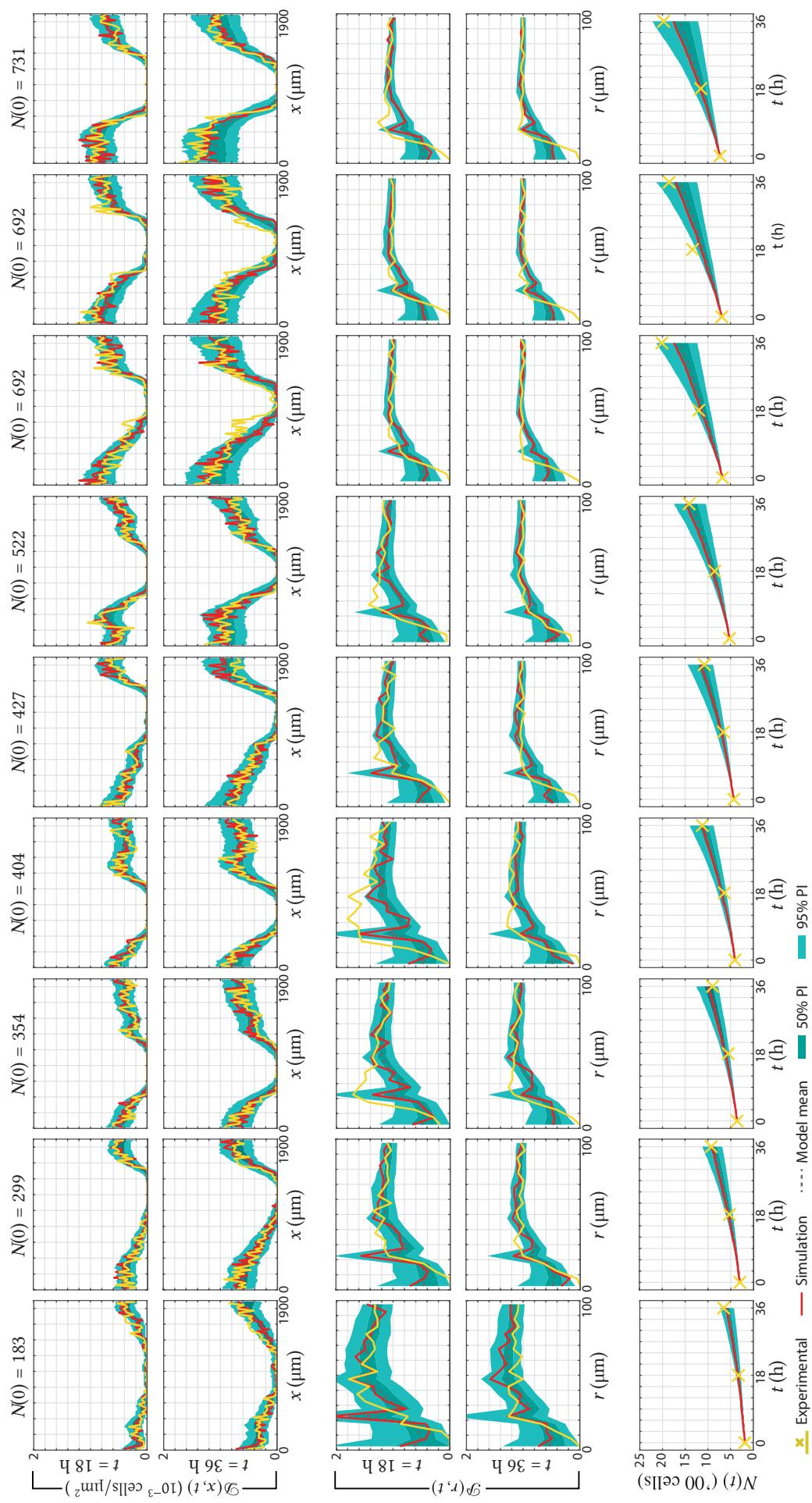
▲ Figure A.9. Reproduction of results in figure 6m–x of the main chapter, for all replicates, for $\sigma = 12$ μm.

139

▲ Figure A.10. Reproduction of results in figure 6*a–l* of the main chapter, for all replicates, for $\sigma = 24\ \mu m$.

▲ Figure A.11. Reproduction of results in figure 6m–x of the main chapter, for all replicates, for σ = 24 μm.

Appendix B

Supplementary material for Chapter 3, "Persistence as an optimal hedging strategy"

B.1   Code availability

Code used in this work is available at `https://github.com/ap-browning/persisters`.

B.2   Additional environments

In Chapter 3, we present results for three stochastic environments (Constant, Monod and Poisson). In this Appendix, we present and produce results for two additional environments. These environments are shown in figure B.1a–e. In figure B.1 we reproduce results in figure 3.2including these two additional environments.

4. *Ornstein-Uhlenbeck.* (figure B.1d) By considering that the environment moves continuously around $\mu_t = \mu_G$ and always reverts to this value, we can model ζ_t as an Ornstein-Uhlenbeck process [118] with $m(\zeta) = \zeta$. In this case we have

$$\mathrm{d}\zeta_t = \gamma(\mu_G - \zeta_t)\,\mathrm{d}t + \kappa\,\mathrm{d}W_t^{(3)}, \tag{B.1}$$

where γ and κ represent the strength of the mean-reversion process and variability, respectively; and $\mathrm{d}W_t^{(3)}$ is a Wiener process independent to both $\mathrm{d}W_t^{(1)}$ and $\mathrm{d}W_t^{(2)}$ that satisfies $W_{t+h}^{(i)} - W_t^{(i)} \sim \mathcal{N}(0, h)$ where $\mathcal{N}(0, h)$ denotes a normal distribution with mean zero and variance h. For numerical results, we choose $\mu_G = 2\,\mathrm{h}^{-1}$, $\gamma = 1$ and $\kappa = 0.1$.

5. *Duffing.* (figure B.1e) We can mimic the effects of the Poisson switching process using a Duffing oscillator. We consider that the environment, ζ_t, switches randomly between two stable states at $\zeta_t = 0, 1$. These states are separated by an asymmetric instability, such that the state at $\zeta_t = 0$ is rarer than that at $\zeta_t = 1$. We describe this with the Duffing oscillator

$$\mathrm{d}\zeta_t = \gamma\zeta_t(1 - \zeta_t)(\zeta_t - b)\,\mathrm{d}t + \kappa\,\mathrm{d}W_t^{(3)}, \tag{B.2}$$

where γ is the strength of the oscillator; b is the location of the instability, κ represents the strength of the fluctuations of the oscillator; and $\mathrm{d}W_t^{(3)}$ is a Wiener process independent to both $\mathrm{d}W_t^{(1)}$ and $\mathrm{d}W_t^{(2)}$ that satisfies $W_{t+h}^{(i)} - W_t^{(i)} \sim \mathcal{N}(0, h)$ where $\mathcal{N}(0, h)$ denotes a normal distribution with mean zero and variance h.

We couple the dependence of the growth rate to the environment by setting $m(\zeta_t) = \mu_G + (\mu_G - \mu_S)\zeta_t$ so that for $\zeta_t \approx 1$, $m(\zeta_t) \approx \mu_G$ and for $\zeta_t = 0$, $\mu \approx \mu_S$. We choose $\gamma = 20$, $b = 0.4$, $\kappa = 0.5$ and $\mu_S = -2\,\mathrm{h}^{-1}$.

In figure B.2 we examine how a persister strategy that is optimal under one type of environment behaves under another, unfamiliar environment. We assume that cells monitor their growth rate, μ_t, and couple the growth rate to the optimal control by considering $\phi_t^* = \phi^*(\theta_t, \zeta_t)$ where $\zeta_t = m^{-1}(\cdot)$ denotes the inverse of the growth rate coupling function $m(\zeta)$. In other words, the cells measure the current environment state using the growth rate. In figure B.2e–h we show how a persister strategy that is optimal under the Monod environment behaves when exposed to a growth rate from: the Monod environment (figure B.2e); the Poisson environment (figure B.2f); the Ornstein-Uhlenbeck environment (figure B.2g); and, the Duffing environment (figure B.2h). In subsequent rows we repeat this for persister strategies optimal under: the Poisson environment (figure B.2i–l); the Ornstein-Uhlenbeck environment (figure B.2m–p); and, the Duffing environment (figure B.2q–t).
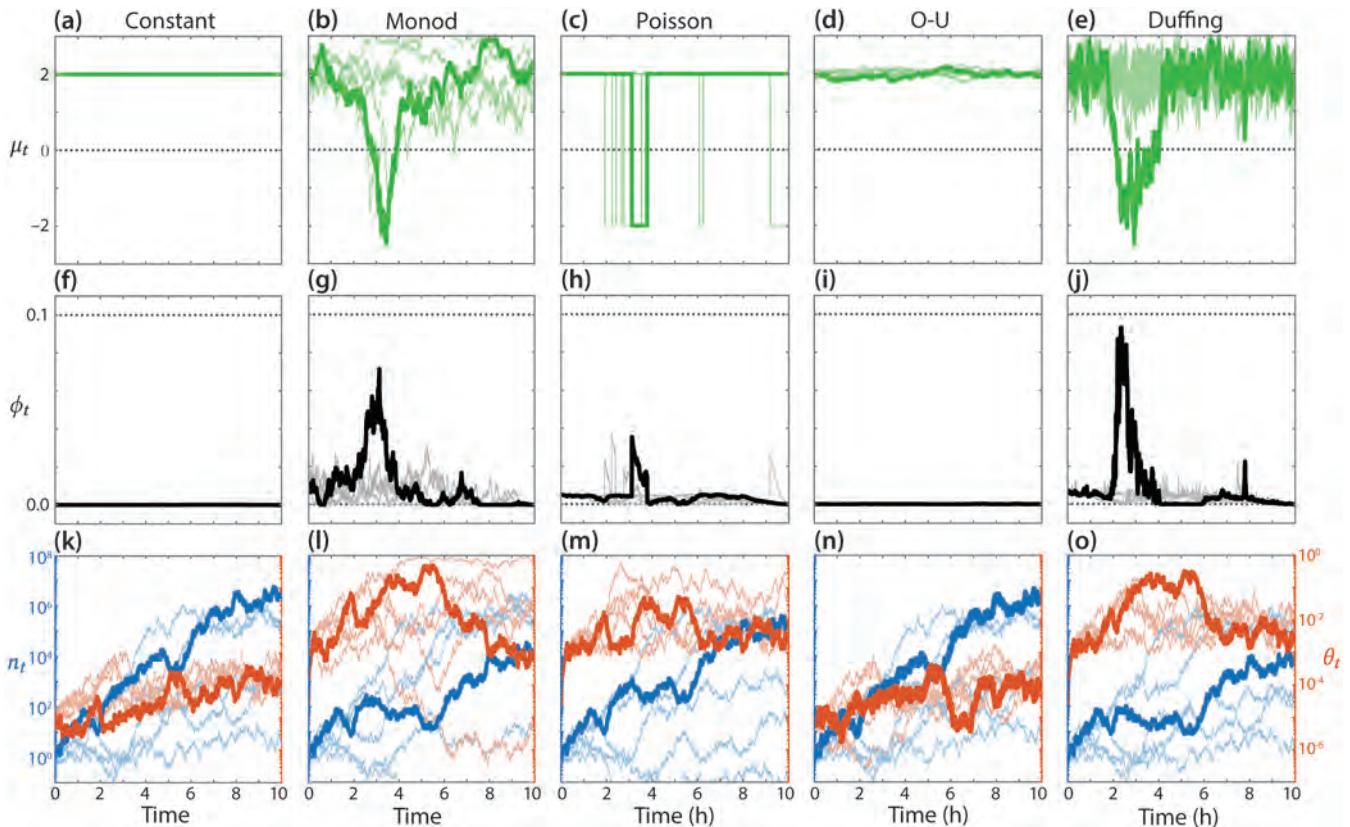
B.3 Constant persister production

Here, we assume that the cell population can use the switching rates, u and v, to steer the steady state proportion, $\hat{\theta}_t$, based on an optimality assumption. We only consider this type of persistence in the presence of a constant environment in growth, so that $\mu_t = \mu_G$. Allowing $\hat{\theta}_t \in [0, 1]$ to be a control reduces the system to a one-dimensional control problem with state equation given by the Itô form
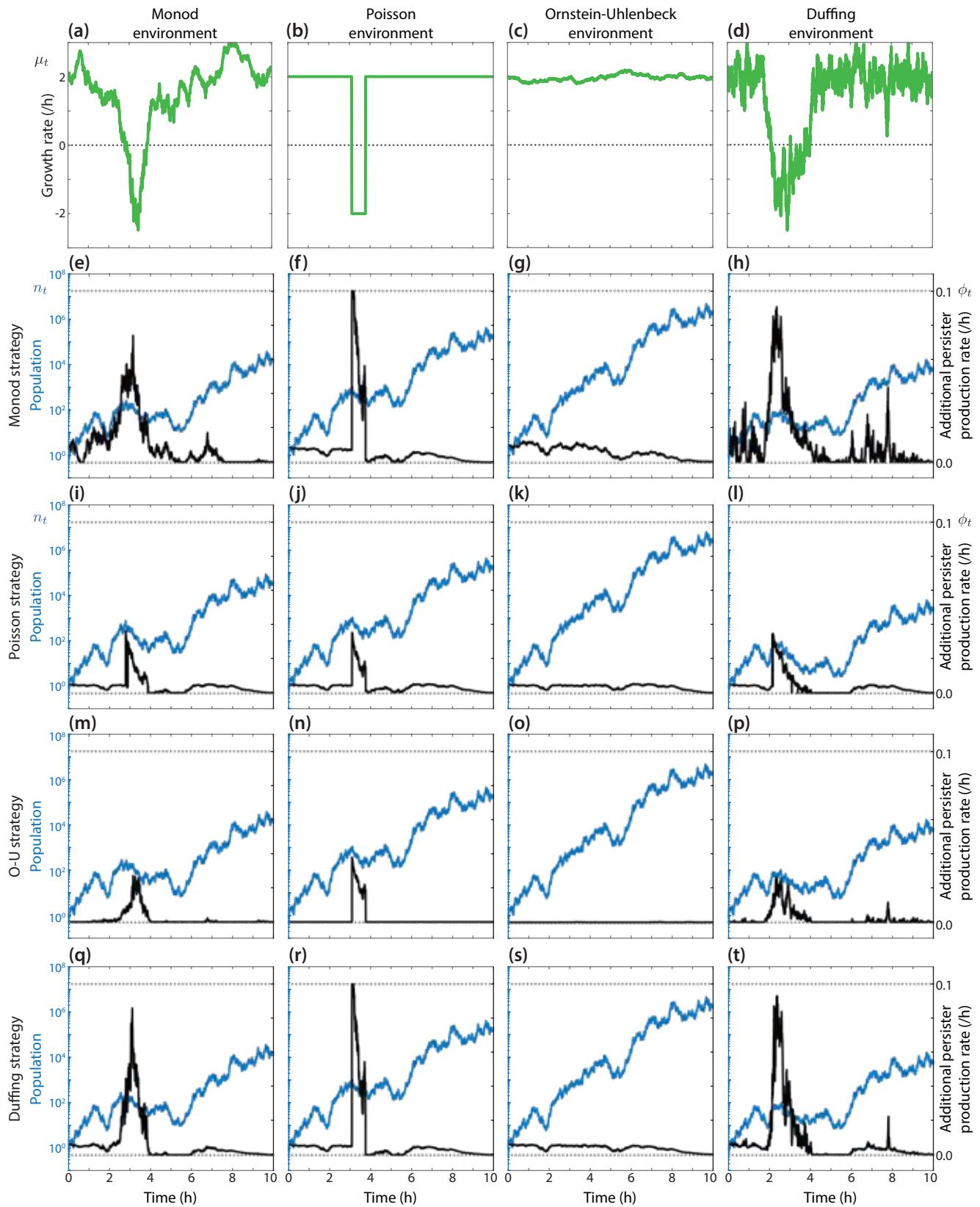
$$\frac{\mathrm{d}n_t}{n_t} = (1 - \hat{\theta}_t(1 - \varepsilon))\mu_G\,\mathrm{d}t + \sqrt{(1 - \hat{\theta}_t)^2\sigma^2 + \hat{\theta}_t^2\eta^2}\,\mathrm{d}W_t, \tag{B.3}$$

where $\hat{\theta}_t$ is now a control. Setting $\mathbb{E}(\mathrm{d}\theta_t) = 0$ in equation (3.4b), reveals the relationship between u, v and $\hat{\theta}_t$:

$$0 = (1 - \theta_t)u - \theta_t v - (1 - \theta_t)\theta_t\big((1 - \varepsilon)\mu_t - \sigma^2(1 - \theta_t) + \eta^2\theta_t\big),$$

$$\Rightarrow u = \hat{\theta}_t\left(\frac{v}{1 - \hat{\theta}_t} + \eta^2\hat{\theta}_t - (1 - \hat{\theta}_t)\sigma^2 + (1 - \varepsilon)\mu\right).$$

▲ Figure B.1. Optimal persister production in cell populations under various types of environment showing: (a)-(e) the growth rates produced by each environment; (f)-(j) the persister production; and, (k)-(o) the population (blue, left scale) and persister proportion (red, right scale). The seeds used to generate the Wiener process were fixed for all environments. Also shown are five additional, independent, realisations (semi-transparent).

▲ **Figure B.2.** Behaviour under each cell strategy in response to an unfamiliar environment. The rows correspond to each optimal cell strategy, the columns to each environment which the cells respond to. Note that (e), (j), (o) and (t) correspond to (g)–(j) of figure B.1.

As $\phi_t = 0$, the fitness measure (equation (3.7)) reduces to

$$J_{\hat{\theta}} = \mathbb{E}\left[\log_e(n_T)\right],$$

where n_t is described by equation (B.3), and n_T represents n_t at terminal time, $t = T$.

We solve this optimal control problem using Hamilton-Jacobi-Bellman (HJB) optimal control theory [118, 125]. We define the value function

$$V(x, s) = \max_{\hat{\theta}_t} \mathbb{E}\left[\log_e(n_T)\right], \quad \text{where } n_s = x, \tag{B.4}$$

and n_t is described by equation (B.3). Therefore, V describes the optimal payoff obtainable, starting at state x at time s. HJB describes V as the solution of a partial differential equation (PDE) [118, 125], given by

$$0 = \max_{\hat{\theta}_t} \left(\frac{\partial V}{\partial s} + (1 - \hat{\theta}_s(1 - \varepsilon))\mu_G x \frac{\partial V}{\partial x} + \frac{1}{2}x^2 \left((1 - \hat{\theta}_t)^2 \sigma^2 + \hat{\theta}_s^2 \eta^2\right) \frac{\partial^2 V}{\partial x^2} \right), \quad V(x, T) = \log_e(x). \tag{B.5}$$

Here, the terminal condition of the PDE is obtained by substituting $s = T$ in equation (B.4).

As equation (B.5) is quadratic in $\hat{\theta}_t$, we can carry out the maximization by setting the derivative to zero to see that the optimal control, $\hat{\theta}_t^*$, is given by

$$\hat{\theta}_s^* = \frac{(1 - \varepsilon)\mu_G \dfrac{\partial V}{\partial x} + x\sigma^2 \dfrac{\partial^2 V}{\partial x^2}}{x\left(\sigma^2 + \eta^2\right) \dfrac{\partial^2 V}{\partial x^2}}. \tag{B.6}$$

Substituting equation (B.6) into equation (B.5) gives a nonlinear PDE that can be solved to give $V(x, s)$ and hence $\hat{\theta}_t^*$. Noting the form of the terminal condition, the ansatz

$$V(x, s) = h(s) + \log_e(x), \quad h(T) = 0, \tag{B.7}$$

is consistent with the PDE, with

$$h(s) = \exp\left(\frac{(T - s)\left\{\mu_G\left[(1 - \varepsilon)^2 \mu_G + 2\varepsilon\sigma^2\right] + \eta^2(2\mu_G - \sigma^2)\right\}}{2(\sigma^2 + \eta^2)} \right).$$

To obtain an algebraic expression for the optimal control strategy, we can substitute the analytical solution for $V(x, s)$ (equation (B.7)) into equation (B.6) to obtain an expression for $\hat{\theta}_t^*$. Noting that $\hat{\theta}_t^* \in [0, 1]$, we can write this as

$$\hat{\theta}_t^* = \max\left(0, \min\left(1, \frac{\sigma^2 - (1 - \varepsilon)\mu_G}{\eta^2 + \sigma^2}\right)\right). \tag{B.8}$$

As equation (B.5) is quadratic in the control, we can verify the optimal solution is, indeed, a maximum by examining the sign of the coefficient of $\hat{\theta}_t^2$. We see that

$$\text{sign}\left(\frac{1}{2}x^2\eta^2 \frac{\partial^2 V}{\partial x^2}\right) = -1,$$

so we always obtain a maximum.

B.4 Variable persister production

We first apply Itô's lemma[1] [118] to the change of variables,

$$\tilde{n}_t = \log(r_t + p_t), \quad \theta_t = \frac{p_t}{n_t},$$

giving

$$d \begin{bmatrix} \tilde{n}_t \\ \theta_t \end{bmatrix} = \begin{bmatrix} f^{(\tilde{n})}(\theta_t, \zeta_t) \\ f^{(\theta)}(\theta_t, \zeta_t, \phi_t) \end{bmatrix} dt + \Sigma(\theta_t)\, dW_t, \tag{B.9}$$

where $dW_t = [dW_t^{(1)}, dW_t^{(2)}]^{\text{tr}}$ denotes a two-dimensional Wiener process with independent components, where superscript $^{\text{tr}}$ the matrix transpose, and

$$\Sigma(\theta_t)\Sigma(\theta_t)^{\text{tr}} = \begin{bmatrix} \Sigma^{(\tilde{n}\tilde{n})}(\theta_t) & \Sigma^{(\tilde{n}\theta)}(\theta_t) \\ \Sigma^{(\tilde{n}\theta)}(\theta_t) & \Sigma^{(\theta\theta),}(\theta_t) \end{bmatrix}, \tag{B.10a}$$

$$f^{(\tilde{n})}(\theta_t, \zeta_t) = (1 - (1-\varepsilon)\theta_t)m(\zeta_t) - \frac{\sigma^2(1-\theta_t)^2}{2} - \frac{\eta^2\theta_t^2}{2}, \tag{B.10b}$$

$$f^{(\theta)}(\theta_t, \zeta_t, \phi_t) = (u + \phi_t)(1 - \theta_t) - v\theta_t - (1-\theta_t)\theta_t(\eta^2\theta_t + m(\zeta_t)(1-\varepsilon) - (1-\theta_t)\sigma^2) \tag{B.10c}$$

$$\Sigma^{(\tilde{n}\tilde{n})}(\theta_t) = \sigma^2(1-\theta_t)^2 + \eta^2\theta_t^2, \tag{B.10d}$$

$$\Sigma^{(\theta\theta)}(\theta_t) = (\sigma^2 + \eta^2)(1-\theta_t)^2\theta_t^2, \tag{B.10e}$$

$$\Sigma^{(\tilde{n}\theta)}(\theta_t) = (1-\theta_t)\theta_t[\eta^2\theta_t - \sigma^2(1-\theta_t)]. \tag{B.10f}$$

Here, ζ_t is described by a state equation that differs between environments and we express $\Sigma = \Sigma(\theta_t)$ to denote the dependence of, for example, Σ on only θ_t.

B.4.1 Continuous environments

For each continuous environment, we can express the state equation for ζ_t as

$$d\zeta_t = f^{(\zeta)}(\zeta_t)\, ds + \sqrt{\Sigma^{(\zeta\zeta)}(\zeta_t)}\, dW_t^{(3)}. \tag{B.11}$$

Here, $dW_t^{(3)}$ is a Wiener process scaled by an intensity $\sqrt{\Sigma^{(\zeta\zeta)}(\zeta_t)}$, and is independent of both $dW_t^{(1)}$ and $dW_t^{(2)}$

To find the optimal control, we define the value function, $V(x, y, z, s)$, such that

$$V(x, y, z, s) = \max_{\phi_t \in \mathcal{U}} \mathbb{E}\left[\int_s^T \alpha\phi_t^2\, ds + x_T\right], \quad (\tilde{n}_s, \theta_s, \zeta_s) = (x, y, z), \tag{B.12}$$

---

[1] **Itô's lemma.** For the stochastic process given by $dX_t = M(X_t)\, dt + S(X_t)\, dW_t$ — where $X_t \in \mathbb{R}^2$, $M : \mathbb{R}^2 \to \mathbb{R}^2$, $S : \mathbb{R}^2 \to \mathbb{R}^{2\times 2}$ and W_t is a two-dimensional Wiener process — then Itô's lemma states that

$$df(X_t) = \left[(\nabla_X f)^{\text{tr}} M(X_t) + \frac{1}{2}\text{Trace}(S(X_t)^{\text{tr}}(H_X f)S(X_t))\right] dt + (\nabla_X f)^{\text{tr}} S(X_t)\, dW_t.$$

Here, $H_X f$ denotes the Hessian matrix for the function f; superscript $^{\text{tr}}$ the matrix transpose; and Trace$(\cdot)$ the matrix trace. In our study, the untransformed variables are denoted $X = [r_t, p_t]^{\text{tr}}$ and the transformed system denoted $Y = [f(r_t, p_t), g(r_t, p_t)]^{\text{tr}} = [\tilde{n}_t, \theta_t]^{\text{tr}}$.

where ϕ_t is the control.

HJB describes V as the solution of PDE [118, 143], given by

$$0 = \max_{\phi_s \in \mathcal{U}} \left(\mathscr{A}V + \alpha \phi_t^2 \right) = \max_{\phi_t \in \mathcal{U}} \mathscr{H}$$

where $\mathscr{A}V$ is the infinitesimal generator of V with respect to the stochastic process given by corresponding state equations, and we have defined $\mathscr{H} = \mathscr{A}V + \alpha \phi_t^2$ for notational clarity. For the the continuous environment, the state equations are given by equations (B.9) and (B.11) so V is described by the PDE

$$0 = \max_{\phi_t \in \mathcal{U}} \left\{ \frac{\partial V}{\partial s} + f^{(\tilde{n})} \frac{\partial V}{\partial x} + f^{(\theta)} \frac{\partial V}{\partial y} + f^{(\zeta)} \frac{\partial V}{\partial z} + \frac{\Sigma^{(\tilde{n}\tilde{n})}}{2} \frac{\partial^2 V}{\partial x^2} + \frac{\Sigma^{(\theta\theta)}}{2} \frac{\partial^2 V}{\partial y^2} + \frac{\Sigma^{(\zeta\zeta)}}{2} \frac{\partial^2 V}{\partial z^2} \right.$$
$$\left. + \Sigma^{(\tilde{n}\theta)} \frac{\partial^2 V}{\partial x \partial y} + \alpha \phi_t^2 \right\},$$
$$V(x, y, z, T) = x.$$

As all of the coefficients, $f^{(\cdot)}$ and $\Sigma^{(\cdots)}$, do not depend on x, we introduce the ansatz $V(x, y, z, s) = \Psi(y, z, s) + x$, leading to a two-dimensional PDE

$$0 = \max_{\phi_t \in \mathcal{U}} \left\{ \frac{\partial \Psi}{\partial s} + f^{(\tilde{n})} + f^{(\theta)} \frac{\partial \Psi}{\partial y} + f^{(\zeta)} \frac{\partial \Psi}{\partial z} + \frac{\Sigma^{(\theta\theta)}}{2} \frac{\partial^2 \Psi}{\partial y^2} + \frac{\Sigma^{(\zeta\zeta)}}{2} \frac{\partial^2 \Psi}{\partial z^2} + \alpha \phi_t^2 \right\}, \quad \Psi(y, z, T) = 0. \qquad \text{(B.13)}$$

Note that in the main chapter, we do not apply the transformation $x \to \log_e(x)$, so the ansatz is equivalent to $V(x, y, z, s) = \Psi(y, z, s) + \log_e(x)$.

For brevity, we express $\mathscr{H}$ as

$$\mathscr{H} = \mathscr{H}_\phi(y, \phi_t, s) + \mathscr{H}_\star(y, z, s),$$

so that

$$\underset{\phi_t \in \mathcal{U}}{\operatorname{argmax}} \, \mathscr{H} = \underset{\phi_t \in \mathcal{U}}{\operatorname{argmax}} \, \mathscr{H}_\phi = \phi_t^*,$$

and

$$\mathscr{H}_\phi = \alpha \phi_t^2 + \frac{\partial V}{\partial y} \, (1 - y) \phi_t,$$

for all continuous environments. This gives the optimal control $\phi_t^* \in [0, 0.1]$,

$$\phi_t^* = \max \left[0, \min \left[0.1, -\frac{1}{2\alpha} \frac{\partial V}{\partial y} \, (1 - y) \right] \right]. \qquad \text{(B.14)}$$

It is evident that $\partial^2 \mathscr{H}_\phi / \partial \phi^2 = 2\alpha < 0$ (since $\alpha < 0$), so we always obtain a maximum.

B.4.2 Poisson environment

We can replace the variable rate Poisson process, $\mathrm{d}P_t$ (described by equation (3.6) in Chapter 3) with

$$\mathrm{d}\zeta_t = \rho(\zeta_t) \, \mathrm{d}P_t = \rho_1(\zeta_t) \, \mathrm{d}P_t^{(1)} + \rho_2(\zeta_t) \, \mathrm{d}P_t^{(2)}, \qquad \text{(B.15)}$$

where $dP_t^{(1)}$ and $dP_t^{(2)}$ have constant rates $\lambda_1 = 1/\tau_G$ and $\lambda_2 = 1/\tau_S$, respectively, and jump distances

$$\rho_1(\zeta_t) = -(\zeta_t + 1) = \begin{cases} -2, & \zeta_t = +1, \\ 0, & \zeta_t = -1, \end{cases} \quad \rho_2(\zeta_t) = -(\zeta_t - 1) = \begin{cases} 0, & \zeta_t = +1, \\ +2, & \zeta_t = -1. \end{cases}$$

For the the Poisson environment, the state equations are given by equations (B.9) and (B.15) so V is described by the PDE

$$0 = \max_{\phi_t \in \mathcal{U}} \left\{ \frac{\partial V}{\partial s} + f^{(\tilde{n})} \frac{\partial V}{\partial x} + f^{(\theta)} \frac{\partial V}{\partial y} + \frac{\Sigma^{(\tilde{n}\tilde{n})}}{2} \frac{\partial^2 V}{\partial x^2} + \frac{\Sigma^{(\theta\theta)}}{2} \frac{\partial^2 V}{\partial y^2} + \lambda_1 \left[V|_{z=-1} - V + (z+1)\frac{\partial V}{\partial z} \right] \right.$$
$$\left. + \lambda_2 \left[V|_{z=+1} - V + (z-1)\frac{\partial V}{\partial z} \right] + \alpha\phi_t^2 \right\}, \quad V(x, y, z, T) = x.$$

Introducing the ansatz $V(x, y, z, t) = \Psi(y, z, t) + x$; and, since $m(\zeta_t) = \text{sign}(\zeta_t)$ so that $\partial V/\partial z = 0$, we have that

$$0 = \max_{\phi_t \in \mathcal{U}} \left\{ \frac{\partial \Psi}{\partial s} + f^{(\tilde{n})} + f^{(\theta)} \frac{\partial \Psi}{\partial y} + \frac{\Sigma^{(\theta\theta)}}{2} \frac{\partial^2 \Psi}{\partial y^2} + \lambda_1 \left[\Psi|_{z=-1} - \Psi \right] + \lambda_2 \left[\Psi|_{z=+1} - \Psi \right] + \alpha\phi_t^2 \right\}, \quad \Psi(y, z, T) = 0.$$

$$\text{(B.16)}$$

As equation (B.16) and the state equations only depend on z spatially at $z = \pm 1$, we write equation (B.16) as a coupled system of PDEs where Ψ^+ denotes Ψ at $z = 1$; and Ψ^- denotes Ψ at $z = -1$,

$$0 = \frac{\partial \Psi^+}{\partial s} + f^{(\tilde{n})}(y, +1) + f^{(\theta)}(y, +1, \phi_t^*)\frac{\partial \Psi^+}{\partial y} + \frac{\Sigma^{(\theta\theta)}(y)}{2} \frac{\partial^2 \Psi^+}{\partial y^2} + \lambda_1 \left[\Psi^- - \Psi^+ \right] + (\phi_t^*)^2\alpha, \quad \text{(B.17a)}$$

$$0 = \frac{\partial \Psi^-}{\partial s} + f^{(\tilde{n})}(y, -1) + f^{(\theta)}(y, -1, \phi_t^*)\frac{\partial \Psi^-}{\partial y} + \frac{\Sigma^{(\theta\theta)}(y)}{2} \frac{\partial^2 \Psi^-}{\partial y^2} + \lambda_2 \left[\Psi^+ - \Psi^- \right] + (\phi_t^*)^2\alpha. \quad \text{(B.17b)}$$

Here, the optimal control, ϕ_t^* is given by equation (B.14).

B.5 Numerical techniques

To solve the HJB equation (equation (B.13)) we employ a numerical scheme with a logarithmically spaced grid in y near $y = 0$ and $y = 1$ (figure B.3a), and constant spacing in z. We use up-winding to approximate derivatives in the y direction (figure B.3b). At the boundaries, we note that $\Sigma^{(\theta\theta)}(0, z) = \Sigma^{(\theta\theta)}(1, z) = 0$, so we only need to approximate the first derivative on the boundaries in y. For all other boundaries, we linearly extrapolate the two nearest nodes to estimate the derivatives on the boundary (figure B.3c). We use the explicit Euler method with $\Delta t = 1/30000$ to integrate backwards time from the terminal condition at $t = T$ to $t = 0$. To verify the accuracy of our solution, we compare the integral of the solution with both $\Delta t = 1/30000$ and $\Delta t = 1/60000$, and find a negligible difference. We show the numerical solutions at $t = 0$ (the PDE is solved backwards in time from $t = T$), for each environment, in figure B.4.

We use the Euler-Maruyama algorithm to simulate the SDEs forward. In solving the HJB equation, we store the discretized solution for the control, and interpolate using `interpn` in MATLAB [430] to get the control. In figure B.5 we verify the accuracy of our numerical methods. The numerical solution to the HJB equation at

$t = 0$ gives, by definition (equation (B.12)),

$$V(x, y, z, 0) = \mathbb{E}\left[\int_0^T \alpha \phi_t^2 \, dt + \log_e(n_T)\right],$$

(B.18)

which we can compare to the ensemble average of 5000 realizations of the SDE. In figure B.5 we show both our numerical solution and the ensemble average at $\mu_0 = 2$ and $n_0 = 1$.

In the following subsections we provide full details of our numerical solutions. Code used to produce the numerical results is available on GitHub at `github.com/ap-browning/persisters`.

B.5.1 Spatial discretization

We note that $y \in [0, 1]$ and $z \in \mathbb{R}$, however typically we find that $z \sim \mathcal{O}(1)$ in realizations of the SDE for all the environments considered. We denote the discretized domain as y_i and z_j, so that our finite difference approximation of the solution at each grid point, $\Psi(y_i, z_j, s)$, is denoted by

$$\Psi(y_i, z_j, s) \approx \overline{\Psi}_{i,j}(s), \ i \in \mathcal{I} = \{1, 2, ..., I\}, \ j \in \mathcal{J} = \{1, 2, ..., J\}.$$

Discretization in y. We apply apply variable grid spacing in y, demonstrated in figure B.3a,b. We set $y_1 = 0$ and $y_I = 1$. For $10^{-5} \leq y \leq 0.1$ and $0.9 \leq y \leq 1 - 10^{-5}$ we apply logarithmic grid spacing, such that $y_{50} = 0.1$ and $y_{81} = 0.9$. For $0.1 < y < 0.9$, we apply constant grid spacing. In this work, we find $I = 130$ is sufficient under visual inspection.

Discretization in z. We apply constant grid spacing in z such that $z_j - z_{j-1} = \Delta z \ \forall \ j$. We solve over a sufficiently large domain in the z direction to contain all realizations of corresponding SDE. In this work, we find that $J = 100$ is sufficient under visual inspection, except the Ornstein-Uhlenbeck environment where we set $J = 200$. The discretization in z is shown in figure B.3b.

B.5.2 Derivatives
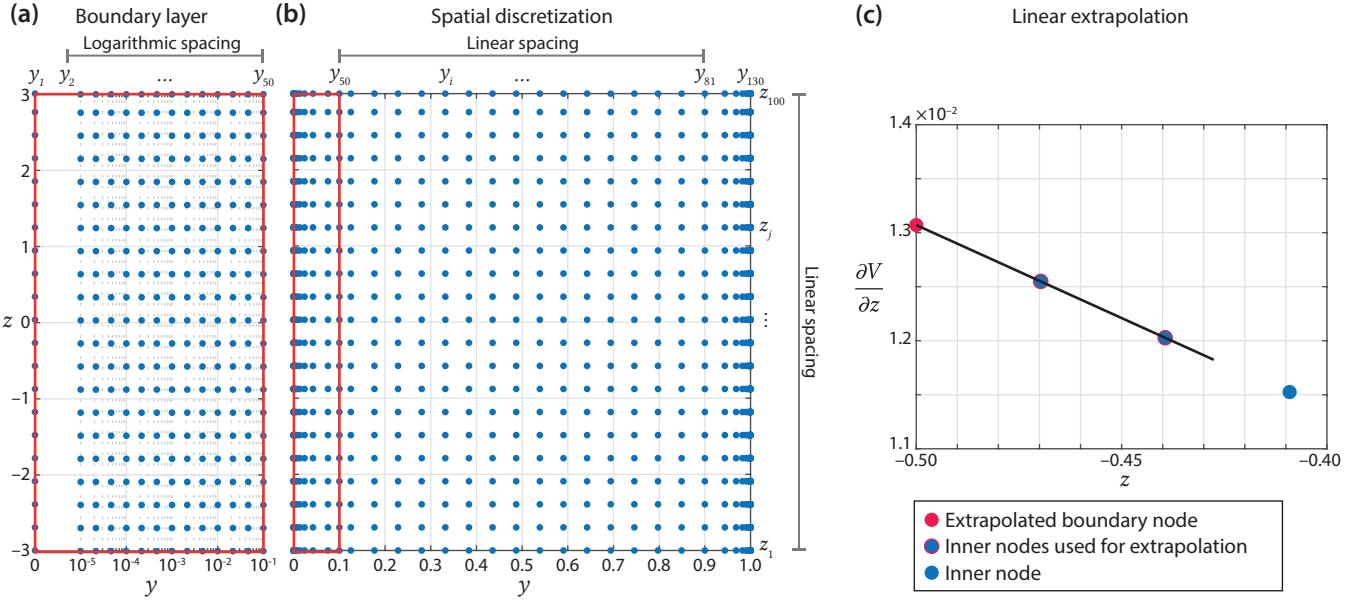
We approximate derivatives in for central nodes using standard techniques, that we detail here. For the boundaries, we linearly extrapolate using the two nearest nodes, as shown in figure B.3c.

Derivatives in y. We apply an up-winding scheme for the first derivative in y. Accounting for the variable mesh in the y direction, on the interior nodes we define

$$\frac{\partial \overline{\Psi}_{i,j}}{\partial y}^{\text{forw}} = \frac{\overline{\Psi}_{i+1,j}(s) - \overline{\Psi}_{i,j}(s)}{y_{i+1} - y_i},$$

$$\frac{\partial \overline{\Psi}_{i,j}}{\partial y}^{\text{back}} = \frac{\overline{\Psi}_{i,j}(s) - \overline{\Psi}_{i-1,j}(s)}{y_i - y_{i-1}},$$

so that our finite difference approximation for the first derivative in y is given by

$$\frac{\partial \Psi(y_i, z_i, s)}{\partial y} \approx \frac{\partial \overline{\Psi}_{i,j}}{\partial y} = \mathbf{1}_{\mathbb{R}^+}\left(f^{(\theta)}(y_i, z_j, (\overline{\phi}_t^*)_{i,j})\right) \frac{\partial \overline{\Psi}_{i,j}}{\partial y}^{\text{forw}} + \mathbf{1}_{\mathbb{R}^-}\left(f^{(\theta)}(y_i, z_j, (\overline{\phi}_t^*)_{i,j})\right) \frac{\partial \overline{\Psi}_{i,j}}{\partial y}^{\text{back}}.$$

(B.19)

(a) Boundary layer · Logarithmic spacing
(b) Spatial discretization · Linear spacing
(c) Linear extrapolation

▲ **Figure B.3.** Numerical scheme details. (a) and (b) demonstrate the spatial discretization. (a) is an inset of (b) (indicated in red) showing the logarithmic spacing for $y \leq 0.1$. (c) Linear extrapolation used to approximate the derivatives on the boundaries. A linear function is fit to the nearest two inner nodes (blue circles outlined in red) to extrapolate the value of the derivative on the boundary (red circle).

Here, $\mathbf{1}_{\mathbb{R}^+}(x)$ is an indicator function that takes the value 1 if $x \in \mathbb{R}^+$, and 0 otherwise. equation (B.19) applies for $i \in \mathcal{I} \setminus \{1, I\}$ and $j \in \mathcal{J}$. On the boundaries, where $i = 1$ and $i = I$, we linearly extrapolate using the nearest two nodes (figure B.3c).

We apply a central difference for the second derivative in y, such that

$$\frac{\partial^2 \Psi(y_i, z_i, s)}{\partial y^2} \approx \frac{\partial^2 \overline{\Psi}_{i,j}}{\partial y^2} = \frac{2(y_i - y_{i-1})(\overline{\Psi}_{i+1,j}(s) - \overline{\Psi}_{i,j}(s)) - (y_{i+1} - y_i)(\overline{\Psi}_{i,j}(s) - \overline{\Psi}_{i-1,j}(s))}{(y_{i+1} - y_{i-1})(y_i - y_{i-1})(y_{i+1} - y_i)},$$

(B.20)

equation (B.20) applies for $i \in \mathcal{I} \setminus \{1, I\}$ and $j \in \mathcal{J}$. For $i = 1$ and $i = I$, the the coefficient of $\partial^2 \Psi / \partial y^2$ in the PDE, $\Sigma^{(\theta\theta)}(y_0) = \Sigma^{(\theta\theta)}(y_I) = 0$, so we do not need to extrapolate the second derivative in y.

Derivatives in z. On the interior nodes, we use central differences to approximate the first and second derivative in z, such that
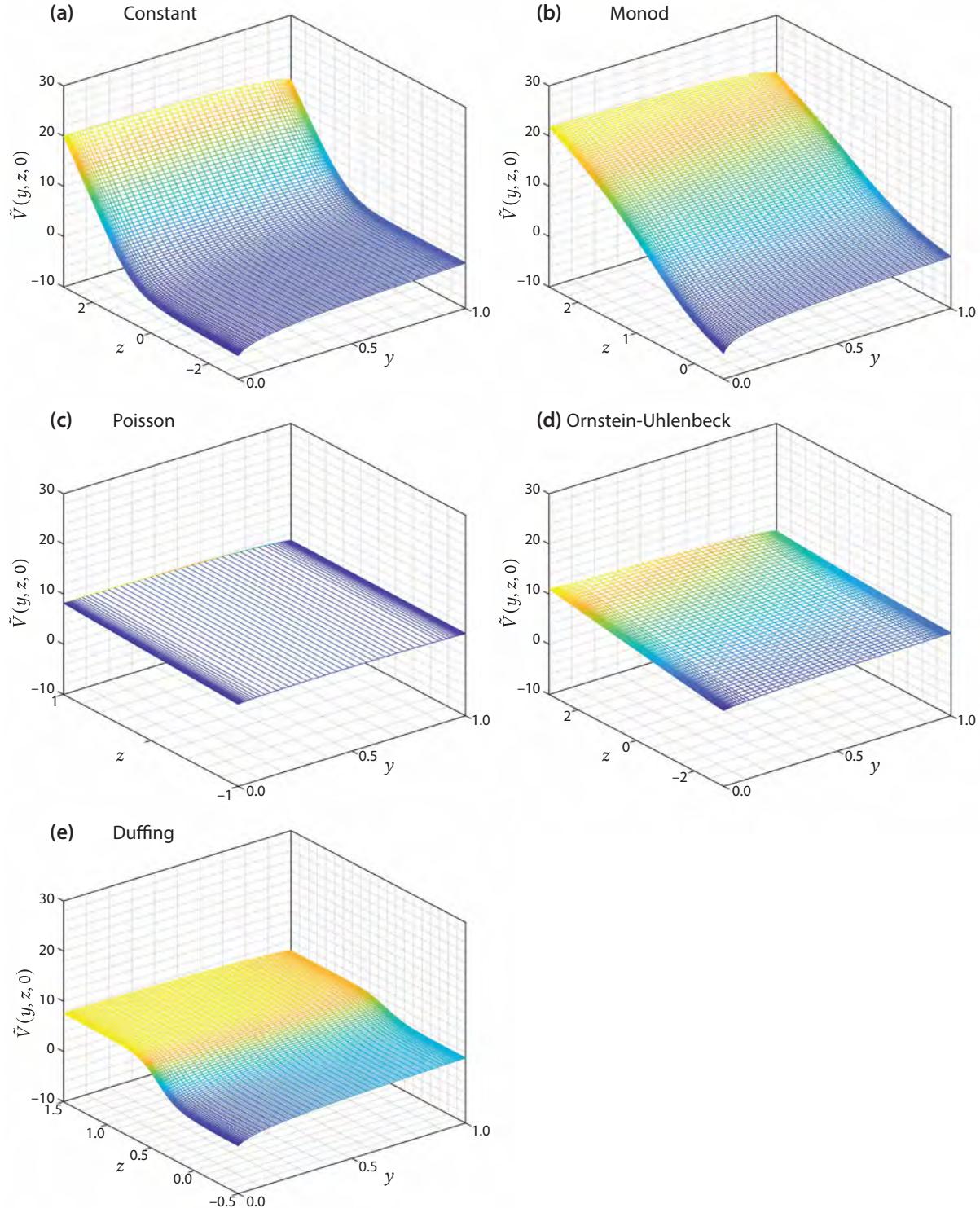
$$\frac{\partial \Psi(y_i, z_i, s)}{\partial z} \approx \frac{\partial \overline{\Psi}_{i,j}}{\partial z} = \frac{\overline{\Psi}_{i,j+1}(s) - \overline{\Psi}_{i,j-1}(s)}{2\Delta z},$$

(B.21)

$$\frac{\partial^2 \Psi(y_i, z_i, s)}{\partial z^2} \approx \frac{\partial^2 \overline{\Psi}_{i,j}}{\partial z^2} = \frac{\overline{\Psi}_{i,j+1}(s) - 2\overline{\Psi}_{i,j}(s) + \Psi_{i,j-1}}{\Delta z^2}.$$

(B.22)
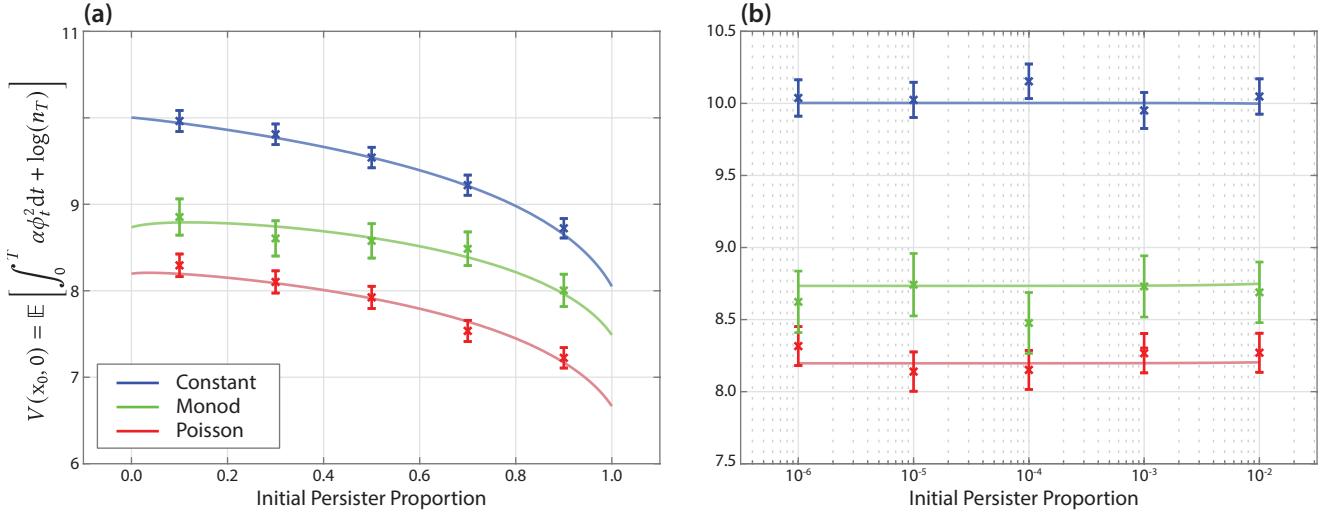
equations (B.21) and (B.22) apply for $i \in \mathcal{I}$ and $j \in \mathcal{J} \setminus \{1, J\}$. For $j = 1$ and $j = J$, we linearly extrapolate using the nearest two nodes (figure B.3c).

B.5.3 Time integration

We apply the first-order explicit Euler method to integrate in time. Whereas the continuous environments involve solving a PDE for $\Psi(y, z, s)$, the Poisson environment involves solving the system of PDEs given by

▲ **Figure B.4.** Solution to the HJB equations, for each environment, at $t = 0\,$h (the HJB equations are solved backwards in time from $T = 10\,$h), given by $\Psi(y, z, 0)$. In each case, $y = \theta_t$ and $z = \zeta_t$. The mesh for the Poisson environment only has two grid points in the z direction, as the system corresponds to a system of two PDEs for $z \geq 0$ and $z < 0$.

▲ **Figure B.5.** Verification of numerical solution to the HJB equations. For each environment in the main chapter, we show the expected payoff, $V(x_0, 0)$ (solid curve) as a function of the initial persister proportion, for $\mu_0 = 2\,\mathrm{h}^{-1}$ and $n_0 = 1$. In (a), we show the initial persister proportion in the linear scale; and in (b), we show the initial persister proportion in the log scale. We verify this numerical solution by showing the mean payoff from 5000 realisations of the SDE (crosses) along with a 95% confidence interval, assuming the payoff is normally distributed.

$\{\Psi^+(y, s), \Psi^-(y, s)\}$. Here, we describe our time-integration implementation separately for each type of PDE.

Continuous environments. Following the substitution of the optimal control (equation (B.14)), we may express the PDE (equation (B.13)) as

$$\frac{\partial \Psi}{\partial s} = F\left(s, y, z, \phi_t^*, \frac{\partial \Psi}{\partial y}, \frac{\partial \Psi}{\partial z}, \frac{\partial^2 \Psi}{\partial y^2}, \frac{\partial^2 \Psi}{\partial z^2}\right), \tag{B.23}$$

where we define

$$F := -f^{(\tilde{n})}(y, z) - f^{(\theta)}(y, z, \phi_t^*)\frac{\partial \Psi}{\partial y} - f^{(\zeta)}(z)\frac{\partial \Psi}{\partial z} - \frac{\Sigma^{(\theta\theta)}(y)}{2}\frac{\partial^2 \Psi}{\partial y^2} - \frac{\Sigma^{(\zeta\zeta)}(z)}{2}\frac{\partial^2 \Psi}{\partial z^2} - \alpha\phi_t^2.$$

Applying the spatial discretization, we may write equation (B.23) as the system of ordinary differential equations (ODEs),

$$\frac{\mathrm{d}}{\mathrm{d}s}\overline{\Psi}_{i,j}(s) = \overline{F}_{i,j}(s) := F\left(s, y_i, z_i, (\overline{\phi_t^*})_{i,j}, \frac{\partial \overline{\Psi}_{i,j}}{\partial y}, \frac{\partial \overline{\Psi}_{i,j}}{\partial z}, \frac{\partial^2 \overline{\Psi}_{i,j}}{\partial y^2}, \frac{\partial^2 \overline{\Psi}_{i,j}}{\partial z^2}\right) \tag{B.24}$$

where

$$(\overline{\phi_t^*})_{i,j} = \max\left(0, -\frac{1}{2\alpha}\frac{\partial \overline{\Psi}_{i,j}}{\partial y}\right),$$

which corresponds to the discretization of equation (B.14).

The ODE problem (equation (B.24)) is coupled to a terminal condition at $s = T$. Therefore, we must integrate backwards in time to $s = 0$. To integrate from s to $s - \Delta s$, we apply the explicit Euler scheme,

so that

$$\overline{\Psi}_{i,j}(s - \delta s) = \overline{\Psi}_{i,j}(s) + \int_s^{s - \Delta s} \overline{F}_{i,j}(\tau) \, \mathrm{d}\tau,$$

$$\approx \overline{\Psi}_{i,j}(s) - \Delta s \, \overline{F}_{i,j}(s).$$

Poisson environment. The system of PDEs for the Poisson environment (equations (B.17)) do not depend on z, and so we only discretize the domain in the y direction. We define $\mathbf{\Psi} = [\Psi^+, \Psi^-]^{\mathrm{tr}}$ for brevity. Following the substitution of the optimal control (equation (B.14)), we may express the system (equations (B.17)) as

$$\frac{\partial \mathbf{\Psi}}{\partial s} = \mathrm{F}\left(s, y, \phi_t^*, \Psi^+, \Psi^-, \frac{\partial \Psi^+}{\partial y}, \frac{\partial \Psi^-}{\partial y}, \frac{\partial^2 \Psi^+}{\partial y^2}, \frac{\partial^2 \Psi^-}{\partial y^2}\right), \tag{B.25}$$

where we define

$$\mathrm{F} := \begin{bmatrix} -f^{(\tilde{n})}(y, +1) - f^{(\theta)}(y, +1, \phi_t^*)\dfrac{\partial \Psi^+}{\partial y} - \dfrac{\Sigma^{(\theta\theta)}(y)}{2}\dfrac{\partial^2 \Psi^+}{\partial y^2} - \lambda_1\left[\Psi^- - \Psi^+\right] - (\phi_t^*)^2\alpha, \\[2em] -f^{(\tilde{n})}(y, -1) - f^{(\theta)}(y, -1, \phi_t^*)\dfrac{\partial \Psi^-}{\partial y} - \dfrac{\Sigma^{(\theta\theta)}(y)}{2}\dfrac{\partial^2 \Psi^-}{\partial y^2} - \lambda_2\left[\Psi^+ - \Psi^-\right] - (\phi_t^*)^2\alpha. \end{bmatrix} \tag{B.26}$$

Applying the spatial discretization, we may write equation (B.5.3) as the system of ordinary differential equations (ODEs),

$$\frac{\mathrm{d}}{\mathrm{d}s}\overline{\mathbf{\Psi}}_i(s) = \overline{\mathrm{F}}_i(s) := \mathrm{F}\left(s, y_i, (\overline{\phi_t^*})_i, \overline{\Psi^+}_i, \overline{\Psi^-}_i, \frac{\partial\overline{\Psi^+}_i}{\partial y}, \frac{\partial\overline{\Psi^-}_i}{\partial y}, \frac{\partial^2\overline{\Psi^+}_i}{\partial y^2}, \frac{\partial^2\overline{\Psi^-}_i}{\partial y^2}\right) \tag{B.27}$$

where

$$(\overline{\phi_t^*})_i = \max\left(0, -\frac{1}{2\alpha}\frac{\partial\overline{\Psi}_i}{\partial y}\right),$$

which corresponds to the discretization of equation (B.14).

The ODE problem (equation (B.24)) is coupled to a terminal condition at $s = T$. Therefore, we must integrate backwards in time to $s = 0$. To integrate from s to $s - \Delta s$, we apply the explicit Euler scheme, so that

$$\overline{\mathbf{\Psi}}_i(s - \delta s) = \overline{\mathbf{\Psi}}_i(s) + \int_s^{s - \Delta s} \overline{\mathrm{F}}_i(\tau) \, \mathrm{d}\tau,$$

$$\approx \overline{\mathbf{\Psi}}_i(s) - \Delta s \, \overline{\mathrm{F}}_i(s).$$

B.5.4 Solving the controlled SDE

We apply the Euler-Maruyama algorithm [144] to solve the SDE with the optimal control. The numerical algorithm described above provides a discretized optimal control,

$$(\overline{\phi_t^*})_{i,j} = \overline{\phi_t^*}(y_i, z_i) \approx \phi_t^*(y_i, z_i).$$

To approximate $\phi_t^*(y, z)$ we interpolate the discretized function using `interpn` in MATLAB [430].

For all environments, the state equations for $\tilde{n}_t$ and θ_t are given by

$$
d \begin{bmatrix} \tilde{n}_t \\ \theta_t \end{bmatrix} = f\big(\theta_t, \zeta_t, \phi_t^*(\theta_t, \zeta_t)\big)\, dt + \Sigma(\theta_t)\, dW_t \,,
$$

where

$$
f := \begin{bmatrix} f^{(\tilde{n})}(\theta_t, \zeta_t) \\ f^{(\theta)}\big(\theta_t, \zeta_t, \phi_t^*(\theta_t, \zeta_t)\big) \end{bmatrix}
$$

and $f^{(\tilde{n})}$, $f^{(\tilde{\theta})}$ and $\Sigma(\theta_t)$ are given in equation (B.10).

For all environments, we initialize the numerical results by setting $\tilde{n}_0 = 0$, $\theta_0 = 1.19 \times 10^{-5}$ and choose ζ_0 such that $\mu_t = m(\zeta_0) = 2$. To integrate $\tilde{n}_t$ and θ_t from t to $t + \Delta t$, we use the Euler-Maruyama algorithm, such that

$$
\begin{bmatrix} \tilde{n}_{t+\Delta t} \\ \theta_{t+\Delta t} \end{bmatrix} = f\big(\theta_t, \zeta_t, \phi_t^*(\theta_t, \zeta_t)\big)\Delta t + \Sigma(\theta_t)\sqrt{\Delta t} \begin{bmatrix} \Delta W_t^{(1)} \\ \Delta W_t^{(2)} \end{bmatrix}.
$$

Here, $\Delta W_t^{(i)} \sim \mathcal{N}(0, 1)$ for $i = 1, 2$. In our study, we fix $\Delta t = 0.002$.

Continuous environments. To integrate ζ_t from t to $t + \Delta t$, we use the Euler-Maruyama algorithm, such that

$$
\zeta_{t+\Delta t} = f^{(\zeta)}(\zeta_t)\Delta t + \sqrt{\Sigma^{(\zeta\zeta)}(\zeta_t)\Delta t}\Delta W_t^{(3)}.
$$

Here, $\Delta W_t^{(3)} \sim \mathcal{N}(0, 1)$.
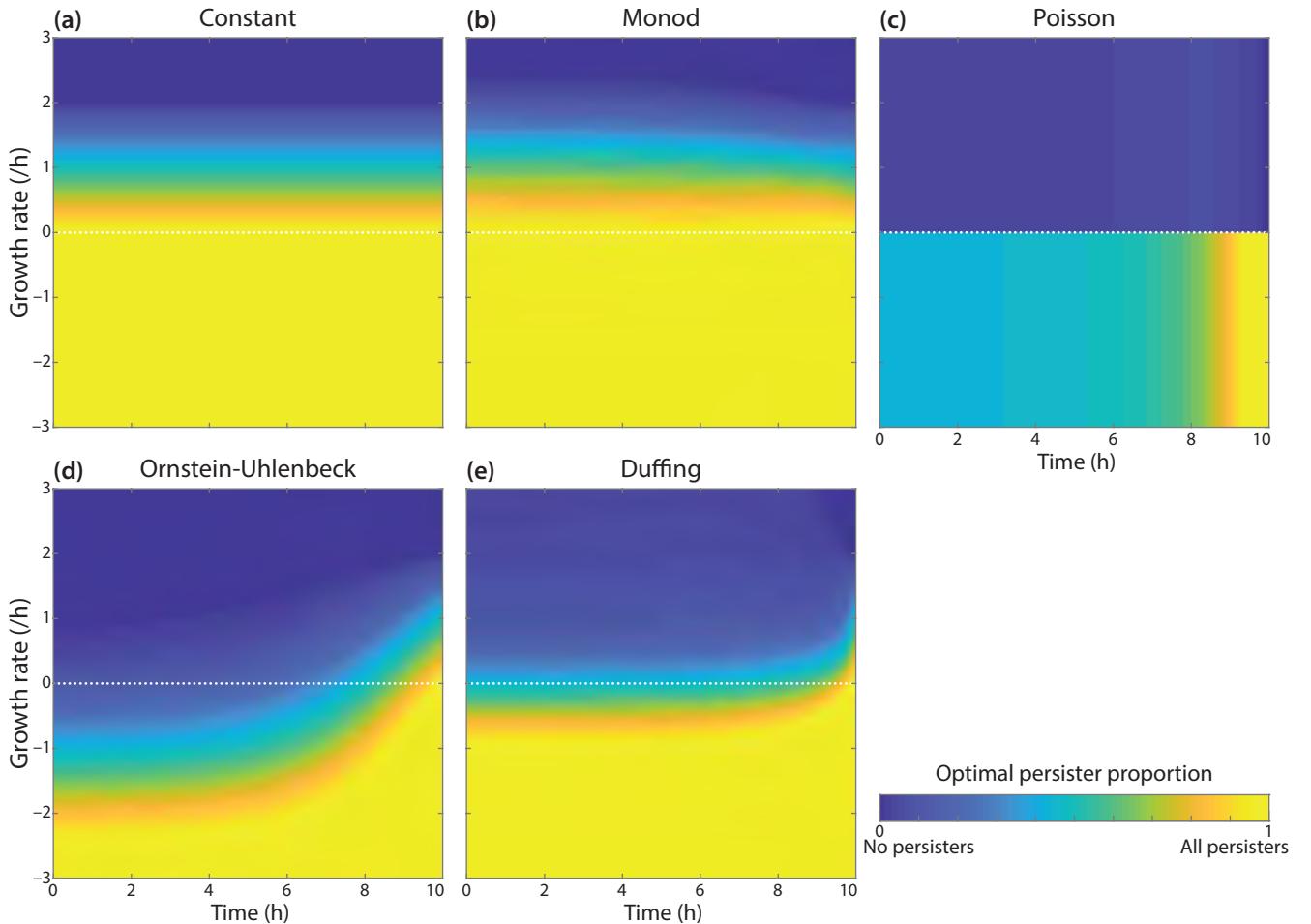
Poisson environment. Here, we integrate ζ_t from t to $t + \Delta t$ by considering

$$
\zeta_{t+\Delta t} = \rho(\zeta_t)\Delta P_t,
$$

where $\Delta P_t \sim \text{Po}(\Delta t \lambda(\zeta_t))$. Here, $\rho(\zeta_t)$ and $\lambda(\zeta_t)$ are given in equation (3.6).

B.6 Optimal persister proportion

The analytic expression for the optimal variable persister production (equation (B.14)) suggests that the population uses ϕ_t^* to steer the population toward the optimal persister proportion, where $\partial V/\partial y = 0$. In figure B.6 we approximate this optimal persister proportion using the numerical solution. To do this, we find the persister proportion, y that minimises the absolute value of the finite difference approximation to $\partial V/\partial y = \partial \Psi/\partial y$, given by equation (B.19). In figure B.6 we show this as a function of the time, s, and the growth rate, $\mu_t = m(z)$.

▲ Figure B.6. Optimal persister proportion. The optimal persister proportion, as a function of time, t (horizontal axis), and growth rate, μ_t, for each environment. For (a), the optimal persister proportion is given by the analytical expression calculated in the main chapter. For (b)–(e), we consider the optimal persister proportion, θ_t^* to the value of y that minimizes $\partial \Psi(y, z, t)/\partial y$, which we approximate using the numerical solution. In (c), the environment can only take two values $\zeta \in \{-1, 1\}$, which we assume corresponds to $\mu_t \geq 0$ and $\mu_t < 0$.

Appendix C

Supplementary material for Chapter 4, "Identifiability analysis for stochastic differential equation models in systems biology"

C.1   Code availability

This appendix is supplementary to, and refers to, code available on Github at github.com/ap-browning/SDE-Identifiability.

C.2 Stochastic simulation algorithm

We simulate the first three models as bio-chemical reaction networks with an event-driven stochastic simulation algorithm (SSA), developed by Gillespie [279]. Each model comprises Q reactions, with propensities and stoichiometries denoted

$$\{(a_k(X_t; \boldsymbol{\theta}), \boldsymbol{\nu}_k)\}_{k=1}^Q, \tag{C.1}$$

respectively. We simulate from the initial condition X_0 at $t = 0$ to the final state X_{t_f} at $t = t_f$. While simulating from the SSA, $X_t \in \mathbb{Z}^N$. We summarise the SSA in algorithm C.3 and provide our implementation in `Module/Models/SimulateSSA.jl`.

Algorithm C.3 Stochastic simulation algorithm

1: Initialise by setting $t = 0$ and $X_t = X_0$.

2: Compute total event rate, $a(X_t) = \sum_{k=1}^Q a_k(X_t; \boldsymbol{\theta})$.

3: Sample time-step $\Delta t \sim \text{Exp}(a(X_t))$. If $t + \Delta t > t_f$ stop the algorithm. Else, set $t \leftarrow t + \Delta t$.

4: Sample event, k, such that $\mathbb{P}(k) = a_k(X_t; \boldsymbol{\theta})/a(X_t)$.

5: Update state, $X_t \leftarrow X_t + \boldsymbol{\nu}_k$.

6: Repeat steps 2 – 6.

C.3 Euler-Maruyama algorithm

We simulate Itô SDEs using the Euler-Maruyama algorithm [285] with reflecting boundaries at $X_{i,t} = 0$ to ensure positivity [226]. Each SDE is defined by

$$dX_t = \boldsymbol{\alpha}(X_t, t; \boldsymbol{\theta}) \, dt + \boldsymbol{\sigma}(X_t, t; \boldsymbol{\theta}) \, dW_t. \tag{C.2}$$

Here X_t is an N-dimensional vector; $\boldsymbol{\alpha}(\cdot)$ maps to an N-dimensional vector; $\boldsymbol{\sigma}(\cdot)$ maps to an $N \times Q$ matrix; and W_t is an Q-dimensional Wiener process with independent components. We simulate from the initial condition X_{t_0} at $t = t_0$ to the final state X_{t_f} at $t = t_f$. We summarise the Euler-Maruyama algorithm in algorithm C.4 and provide our implementation in `Module/Models/EulerMaruyama.jl`.

Algorithm C.4 Euler-Maruyama algorithm

1: Initialise by setting $t = t_0$ and $X_t = X_{t_0}$. Choose time-step, Δt.

2: Sample Wiener increment $\Delta W = (\Delta W_1, \Delta W_2, \ldots, \Delta W_Q)^T$ where $\Delta W_k \sim \mathcal{N}(0, \Delta t)$.

3: Update state $X_t \leftarrow X_t + \boldsymbol{\alpha}(X_t, t; \boldsymbol{\theta})\Delta t + \boldsymbol{\sigma}(X_t, t; \boldsymbol{\theta})\Delta W$.

4: Set $X_{i,t} = |X_{i,t}|$ where $X_t = (X_{1,t}, X_{2,t}, \ldots, X_{N,t})^T$.

5: Update time, $t \leftarrow t + \Delta t$.

6: Repeat steps 2 – 5 until $t \geq t_f$.

C.4 MCMC algorithms

C.4.1 Metropolis-Hastings algorithm

We simulate from the posterior distribution

$$p(\boldsymbol{\theta}|\mathcal{D}) \propto \mathcal{L}(\mathcal{D}|\boldsymbol{\theta})p(\boldsymbol{\theta}), \tag{C.3}$$

using the Metropolis-Hastings algorithm [291, 292] and Markov-chain Monte-Carlo (MCMC). In this work, we employ a multivariate normal proposal kernel with covariance Σ such that

$$q(\boldsymbol{\theta}|\boldsymbol{\theta}_{s-1}) = \text{MVN}(\boldsymbol{\theta}_{s-1}, \Sigma). \tag{C.4}$$

We summarise our implementation of the Metropolis-Hastings MCMC algorithm in algorithm C.5 and provide our implementation in `Module/Inference/MetropolisHastings.jl`.

Algorithm C.5 Metropolis-Hastings algorithm with a symmetric proposal kernel.

1: Sample $\boldsymbol{\theta}_0 \sim p(\boldsymbol{\theta})$ to initialise chain at $s = 0$. Choose desired number of iterations, S.

2: Update iteration $s \leftarrow s + 1$.

3: Propose transition $\boldsymbol{\theta}^* \sim q(\boldsymbol{\theta}|\boldsymbol{\theta}_{s-1})$, where $q(\boldsymbol{\theta}|\boldsymbol{\theta}_{s-1})$ is a multivariate Gaussian.

4: Calculate acceptance probability

$$\alpha_{\text{MH}}(\boldsymbol{\theta}^*|\boldsymbol{\theta}_{s-1}) = \min\left(1, \frac{p(\boldsymbol{\theta}^*)\mathcal{L}(\mathcal{D}|\boldsymbol{\theta}^*)}{p(\boldsymbol{\theta}_{s-1})\mathcal{L}(\mathcal{D}|\boldsymbol{\theta}_{s-1})}\right).$$

5: Accept proposal, $\boldsymbol{\theta}_s \leftarrow \boldsymbol{\theta}^*$ with probability $\alpha_{\text{MH}}(\boldsymbol{\theta}^*|\boldsymbol{\theta}_{s-1})$; else, reject and set $\boldsymbol{\theta}_s \leftarrow \boldsymbol{\theta}_{s-1}$.

6: Repeat steps 2 – 5 until $s = S$.

C.4.2 Particle MCMC algorithm

For most SDE models, the likelihood function, $\mathcal{L}(\mathcal{D}|\boldsymbol{\theta})$, is intractable. We approximate the likelihood using a bootstrap particle filter in a particle MCMC algorithm [257], replacing $\mathcal{L} \leftarrow \hat{\mathcal{L}}$ in algorithm C.5.

The data are defined by $\mathcal{D} = \left\{\left\{t_{n,i}, Y_{\text{obs}}^{n,i}\right\}_{n=1}^{N_E}\right\}_{i=1}^{E}$, where

$$Y_{\text{obs}}^{n,i} \sim g(Y|X_{t_n}^i, t_n; \boldsymbol{\theta}). \tag{C.5}$$

Here, $\left\{X_{t_n}^i\right\}_{n=1}^{N_E}$ are observations from experiment $i = 1, 2, \ldots, N_E$, modelled by a single trace of the SDE (equation (C.2)) initiated at $X_0 = h_i(\boldsymbol{\theta})$. The vector-valued function $h_i(\boldsymbol{\theta})$ captures the fact that the initial condition may depend on unknown parameters, $\boldsymbol{\theta}$. This occurs in the epidemic model where only partial observations of the state are made so that the full initial condition is unknown.

We summarise our implementation of bootstrap particle filter algorithm in algorithm C.6 and provide our implementation in `Module/Inference/PMLogLikeParticleFilter.jl`.

Algorithm C.6 Bootstrap particle filter

1: Initialise likelihood estimate, $\hat{\mathscr{L}} = 1$, and experiment index $i = 0$.

2: Update experiment index $i \leftarrow i + 1$ and initialise R particles for the ith experiment, $\{X_{t_0}^r\}_{r=1}^R$, where $X_0^r = h_i(\boldsymbol{\theta})$ and $t_0 = 0$. Set observation index, $n = 0$.

3: Update observation index, $n \leftarrow n + 1$, and simulate all particles forward from $t = t_{n-1}$ to t_n using the Euler-Maruyama algorithm (algorithm C.4) to obtain $\{X_{t_n}^r\}_{r=1}^R$.

4: Compute weight of each particle, $\{W_n^r\}$, such that $W_n^r = g(Y_{\text{obs}}^{n,i}|X_{t_n}^r)$.

5: Update likelihood estimate,

$$\hat{\mathscr{L}} \leftarrow \hat{\mathscr{L}} \times \frac{1}{R}\sum_{r=1}^R W_n^r.$$

6: Resample R particles from $\{X_{t_n}^r\}_{r=1}^R$ with replacement according to weights $\{W_n^r\}$.

7: Repeat steps 3 – 6 until $n = N_E$.

8: Repeat steps 2 – 7 until $i = E$.

C.4.3 Likelihood for ODE models

When describing data using an ODE model, we make the typical assumption that residuals are independent. The likelihood function is then tractable and given by

$$\mathscr{L}(\mathscr{D}|\boldsymbol{\theta}) = \sum_{i=1}^E \sum_{n=1}^{N_E} g\big(Y_{\text{obs}}^{n,i}|X_{t_n}^i, t_n; \boldsymbol{\theta}\big). \tag{C.6}$$

Here, $\{X_{t_n}^i\}_{n=1}^{N_E}$ are observations from experiment i ($i = 1, 2, \ldots, N_E$), modelled by the ODE model initiated at $X_0 = h_i(\boldsymbol{\theta})$.

C.5 Moment equations

In the main text, we define the raw moments,

$$m_{i_1 i_2 \ldots i_N}(t) = \left\langle \prod_{j=1}^{N} X_{j,t}^{i_j} \right\rangle, \tag{C.7}$$

of the random variable X_t, which is itself described by the Itô SDE (equation (C.2)). Here, we derive the equation describing the time-evolution of the moments.

First, we define $\phi_i(X_t, t) = \prod_{j=1}^{N} X_{j,t}^{i_j}$, where $i = (i_1, i_2, \ldots, i_N)$. By Itô's lemma,

$$d\phi_i(X, t) = f_\phi(X_t, t; \boldsymbol{\theta}) dt + \sum_{i=1}^{N} g_\phi(X_t, t; \boldsymbol{\theta}) dW_t, \tag{C.8}$$

where

$$f_\phi(X_t, t; \boldsymbol{\theta}) = \boldsymbol{\alpha}(X_t, t; \boldsymbol{\theta}) \cdot \boldsymbol{\nabla} \left(\prod_{j=1}^{N} X_{j,t}^{i_j} \right) + \frac{1}{2} \mathrm{Tr} \left(\boldsymbol{\sigma}^T(X_t, t; \boldsymbol{\theta}) \mathrm{H} \left(\prod_{j=1}^{N} X_j^{i_j} \right) \boldsymbol{\sigma}(X_t, t; \boldsymbol{\theta}) \right), \tag{C.9}$$

$$g_\phi(X_t, t; \boldsymbol{\theta}) = \boldsymbol{\nabla} \left(\prod_{j=1}^{N} X_{j,t}^{i_j} \right)^T \boldsymbol{\sigma}(X_t, t; \boldsymbol{\theta}). \tag{C.10}$$

Our aim is to find an ODE describing the time evolution of $\langle \phi_i(X) \rangle$, where $\langle \cdot \rangle$ is the expectation taken with respect to the probability measure of the random variable X_t. Full details are available in [286].

Consider the expectation of equation (C.8) in integral form:

$$\langle \phi_i(X, t) \rangle = \langle \phi_i(X, 0) \rangle + \underbrace{\left\langle \int_0^t f_\phi(X_u, u; \boldsymbol{\theta}) du \right\rangle}_{(*)} + \underbrace{\sum_{n=1}^{N} \left\langle \int_0^t g_\phi^{(n)}(X_u, u; \boldsymbol{\theta}) dW_{n,u} \right\rangle}_{(**)}. \tag{C.11}$$

Here, we denote $W = (W_{1,t}, W_{2,t} \ldots, W_{N,t})^T$ and $g_\phi(\cdot) = \left(g_\phi^{(1)}(\cdot), g_\phi^{(2)}(\cdot), \ldots, g_\phi^{(N)}(\cdot) \right)^T$. Under certain conditions, which are satisfied when all elements of g_ϕ are polynomial, the stochastic integrals in equation (C.11)(**) will vanish [431] due to the Itô formulation. This is the case for the SDEs we consider, which are derived through the chemical Langevin equation with polynomial propensity functions, so that all components of g_ϕ are also polynomials. Therefore, after interchanging the order of time-integration to that of the expectation in (*), we obtain

$$\langle \phi_i(X, t) \rangle = \langle \phi_i(X, 0) \rangle + \int_0^t \left\langle f_\phi(X_u, u; \boldsymbol{\theta}) \right\rangle du. \tag{C.12}$$

We note that, by definition, $m_{i_1 i_2 \ldots i_N}(t) = \langle \phi_i(X, t) \rangle$, and so

$$m_{i_1 i_2 \ldots i_N}(t) = m_{i_1 i_2 \ldots i_N}(0) + \int_0^t \left\langle f_\phi(X_u, u; \boldsymbol{\theta}) \right\rangle du, \tag{C.13}$$

which corresponds to the ODE

$$\frac{\mathrm{d}m_{i_1 i_2 \ldots i_N}}{\mathrm{d}t} = \left\langle f_\phi(\mathrm{X}_u, u; \boldsymbol{\theta}) \right\rangle. \tag{C.14}$$

We provide `Mathematica` scripts used to derive the moment equations for each model in the `Mathematica` folder on Github.

C.5.1 Epidemic model

The CLE for the epidemic model is

$$
d\mathrm{X}_t = \begin{pmatrix} -\theta_1 X_{1,t} X_{3,t} \\ \theta_1 X_{1,t} X_{3,t} - \theta_2 X_{2,t} \\ \theta_2 X_{2,t} - \theta_3 X_{3,t} \\ \theta_3 X_{3,t} \end{pmatrix} dt + \begin{pmatrix} -\sqrt{\theta_1 X_{1,t} X_{3,t}} & 0 & 0 \\ \sqrt{\theta_1 X_{1,t} X_{3,t}} & -\sqrt{\theta_2 X_{2,t}} & 0 \\ 0 & \sqrt{\theta_2 X_{2,t}} & -\sqrt{\theta_3 X_{3,t}} \\ 0 & 0 & \sqrt{\theta_3 X_{3,t}} \end{pmatrix} d\mathrm{W}_t. \tag{C.15}
$$

We denote

$$
m_{i_1 i_2 i_3 i_4}(t) = \left\langle X_{1,t}^{i_1} X_{2,t}^{i_2} X_{3,t}^{i_3} X_{4,t}^{i_4} \right\rangle. \tag{C.16}
$$

To second order, the unclosed moment equations are

$$
\begin{aligned}
\frac{dm_{1000}}{dt} &= -\theta_1 m_{1010}, \\
\frac{dm_{0100}}{dt} &= \theta_1 m_{1010} - \theta_2 m_{0100}, \\
\frac{dm_{0010}}{dt} &= \theta_2 m_{0100} - \theta_3 m_{0010}, \\
\frac{dm_{0001}}{dt} &= \theta_3 m_{0010}, \\
\frac{dm_{2000}}{dt} &= -2\theta_1 m_{2010} + \theta_1 m_{1010}, \\
\frac{dm_{0200}}{dt} &= 2\theta_1 m_{1110} + \theta_1 m_{1010} + \theta_2 m_{0100} - 2\theta_2 m_{0200}, \\
\frac{dm_{0020}}{dt} &= \theta_2 m_{0100} + 2\theta_2 m_{0110} + \theta_3 m_{0010} - 2\theta_3 m_{0020}, \\
\frac{dm_{0002}}{dt} &= \theta_3 m_{0010} + 2\theta_3 m_{0011}, \\
\frac{dm_{1100}}{dt} &= -\theta_1 m_{1110} + \theta_1 m_{2010} - \theta_1 m_{1010} - \theta_2 m_{1100}, \\
\frac{dm_{1010}}{dt} &= -\theta_1 m_{1020} + \theta_2 m_{1100} - \theta_3 m_{1010}, \\
\frac{dm_{1001}}{dt} &= -\theta_1 m_{1011} + \theta_3 m_{1010}, \\
\frac{dm_{0110}}{dt} &= \theta_1 m_{1020} - \theta_2 m_{0100} - \theta_2 m_{0110} + \theta_2 m_{0200} - \theta_3 m_{0110}, \\
\frac{dm_{0101}}{dt} &= \theta_1 m_{1011} - \theta_2 m_{0101} + \theta_3 m_{0110}, \\
\frac{dm_{0011}}{dt} &= \theta_2 m_{0101} - \theta_3 m_{0010} - \theta_3 m_{0011} + \theta_3 m_{0020}.
\end{aligned} \tag{C.17}
$$

Here, we have indicated moments of order three and above in red font. To close the system, we approximate the higher order moments using the following closures.

Mean-field. The mean-field closure approximates third order moments with the appropriate product of first order moments. This results in the following approximations for each of the third-order raw moments in

equation (C.17):

$$m_{2010} \approx m_{1000}^2 m_{0010},$$ (C.18a)

$$m_{1110} \approx m_{1000} m_{0100} m_{0010},$$ (C.18b)

$$m_{1020} \approx m_{1000} m_{0020}^2,$$ (C.18c)

$$m_{1011} \approx m_{1000} m_{0010} m_{0001}.$$ (C.18d)

Pair-wise. The pair-wise closure approximates third order moments with a product and quotient of first and second order moments. This results in the following approximations for each of the third-order raw moments in equation (C.17):

$$m_{2010} \approx \frac{m_{2000} m_{1010}}{m_{1000}},$$ (C.19a)

$$m_{1110} \approx \frac{m_{1100} m_{0110}}{m_{0100}},$$ (C.19b)

$$m_{1020} \approx \frac{m_{1010} m_{0020}}{m_{0010}},$$ (C.19c)

$$m_{1011} \approx \frac{m_{1010} m_{0011}}{m_{0010}}.$$ (C.19d)

Gaussian. The Gaussian closure sets the third order *central* moments to zero. That is,

$$0 = \hat{m}_{i_1 i_2 i_3 i_4}(t) = \left\langle \prod_{j=1}^{4} \left(X_{i,t} - \langle X_{i,t} \rangle \right)^{i_j} \right\rangle.$$

This results in the following approximations for each of the third-order raw moments in equation (C.17):

$$m_{2010} \approx -2 m_{0010} m_{1000}^2 + 2 m_{1000} m_{1010} + m_{0010} m_{2000},$$ (C.20a)
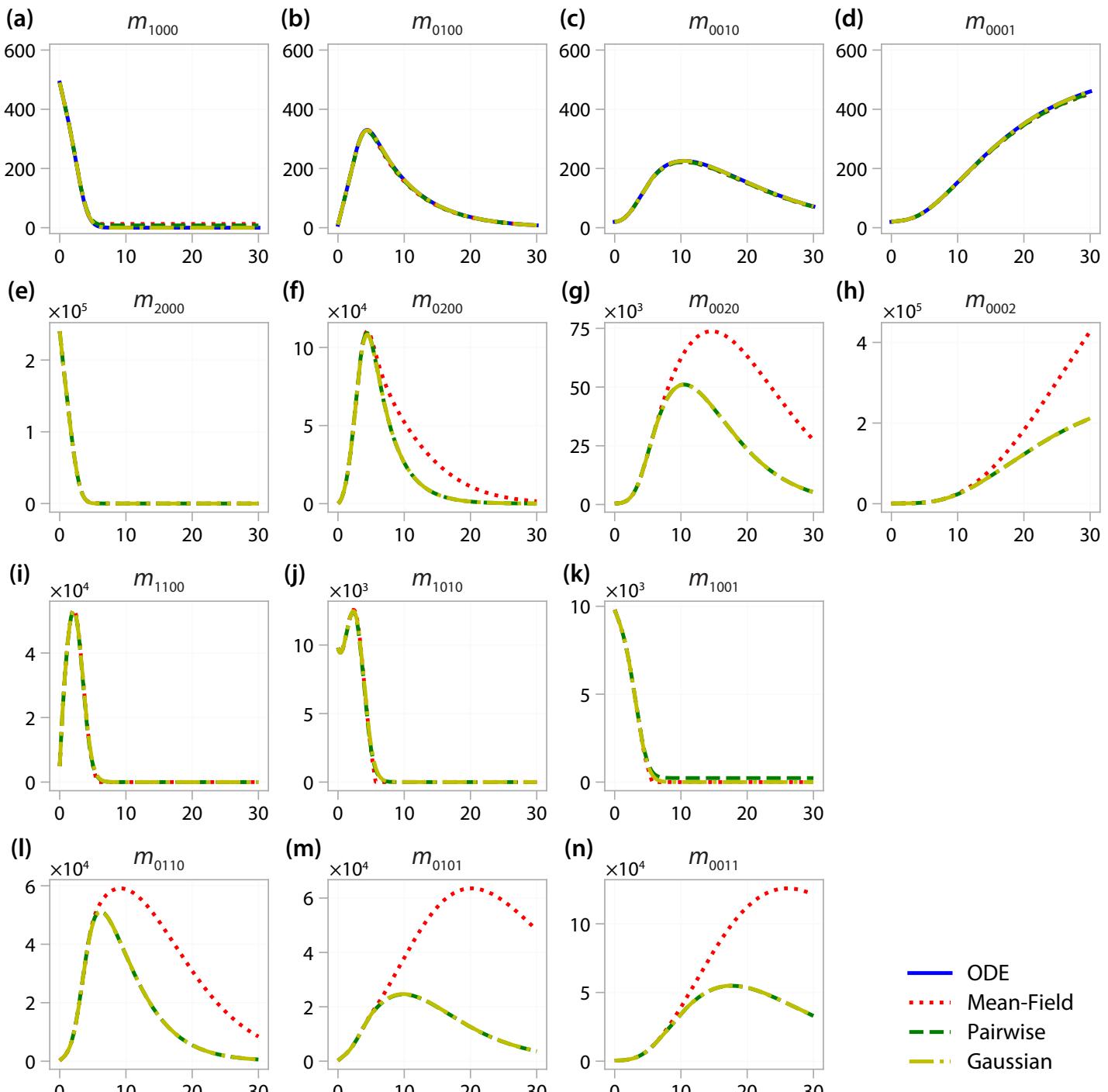
$$m_{1110} \approx -2 m_{0010} m_{0100} m_{1000} + m_{0110} m_{1000} + m_{0100} m_{1010} + m_{0010} m_{1100},$$ (C.20b)

$$m_{1020} \approx -2 m_{0010} m_{1000}^2 + 2 m_{0010} m_{1010} + m_{0020} m_{1000},$$ (C.20c)

$$m_{1011} \approx -2 m_{0001} m_{0010} m_{1000} + m_{0011} m_{1000} + m_{0010} m_{1001} + m_{0001} m_{1010}.$$ (C.20d)

Comparison of closures

In figure C.1 we compare the ODE model to the mean-field closure, pair-wise closure and Gaussian closure for the epidemic model. The mean-field model was solved using the positivity-preserving Patankar-type method [303] (of order one) with time-step $\Delta t = 0.001$, the other closures and the ODE model using the Tsit5 routine (a 4th and 5th order Runge-Kutte regime) in Julia [302].

▲ Figure C.1. Comparison of closure methods for the epidemic model.

Appendix D

Supplementary material for Chapter 5, "Reversible signal transmission in an active mechanical metamaterial"

Code used in this work is available at `https://github.com/ap-browning/rspa-2019`.

D.1 Numerical solution of the continuous model

In this section, we explain the numerical scheme used to solve the continuous model in the main chapter. The non-dimensional continuous model is given by

$$
\begin{aligned}
0 &= \frac{\partial^2 \hat{u}}{\partial \hat{t}^2} - \frac{\partial^2 \hat{u}}{\partial \hat{x}^2} + \frac{\partial \hat{u}}{\partial \hat{t}} + v(\hat{u}^2 - 1)\left(\frac{\hat{u}}{\eta} - \hat{a}\right), \\
0 &= \frac{\partial \hat{a}}{\partial \hat{t}} - \kappa\left(\frac{\hat{u}}{\eta} - \hat{a}\right),
\end{aligned}
\tag{D.1}
$$

where $\hat{x} \in (0, L)$, with no-flux boundary conditions on $\hat{x} \in \{0, \hat{L}\}$, and initial conditions given by

$$
\begin{aligned}
\hat{u}(\hat{x}, 0) &= \begin{cases} 1, & 0 \le \hat{x} < Q, \\ -1, & Q \le \hat{x} \le \hat{L}, \end{cases} \\
\hat{a}(\hat{x}, 0) &= \begin{cases} 1/\eta, & 0 \le \hat{x} < Q, \\ -1/\eta, & Q \le \hat{x} \le \hat{L}. \end{cases}
\end{aligned}
\tag{D.2}
$$

To solve the system given by equation (D.1), we employ a finite difference technique. We first introduce

$$\hat{w} = \frac{\partial \hat{u}}{\partial \hat{t}}, \tag{D.3}$$

so that the system in equation (D.1) can be written as the first order system of partial differential equations (PDEs),

$$
\begin{aligned}
\frac{\partial \hat{u}}{\partial t} &= \hat{w}, \\
\frac{\partial \hat{w}}{\partial t} &= \frac{\partial^2 \hat{u}}{\partial \hat{x}^2} - \hat{w} - \nu(\hat{u}^2 - 1)(\hat{u} - \hat{a}), \\
\frac{\partial \hat{a}}{\partial \hat{t}} &= \kappa\left(\frac{\hat{u}}{\eta} - \hat{a}\right).
\end{aligned}
\tag{D.4}
$$

Initially, the system is at rest, so that

$$\hat{w}(\hat{x}, 0) = 0. \tag{D.5}$$

Next, we divide the domain, $\hat{x} \in [0, \hat{L}]$ into S equally spaced subintervals, each of width $\delta\hat{x} = \hat{L}/S$. We index the boundary of each subinterval with $n \in \{0, 1, ..., S\}$, such that $\hat{u}_n(t) \approx \hat{u}(n\delta\hat{x}, \hat{t})$. Doing this, we obtain the following system of first order, non-linear, ordinary differential equations,

$$
\begin{aligned}
\frac{d\hat{u}_n}{dt} &= \hat{w}_n, \ \forall n, \\
\frac{d\hat{w}_0}{dt} &= \frac{\hat{u}_1 - \hat{u}_0}{\Delta x^2} - \hat{w}_1 - \nu(\hat{u}_0^2 - 1)(\hat{u}_0 - \hat{a}_0), \\
\frac{d\hat{w}_n}{dt} &= \frac{\hat{u}_{n+1} - 2\hat{u}_n + \hat{u}_{n-1}}{\Delta x^2} - \hat{w}_n - \nu(\hat{u}_n^2 - 1)(\hat{u}_n - \hat{a}_n), \ 1 \le n \le S-1, \\
\frac{d\hat{w}_S}{dt} &= \frac{\hat{u}_S - \hat{u}_{S-1}}{\Delta x^2} - \hat{w}_S - \nu(\hat{u}_S^2 - 1)(\hat{u}_S - \hat{a}_S), \\
\frac{da_i}{dt} &= \kappa\left(\frac{\hat{u}_n}{\eta} - \hat{a}_n\right) \ \forall n.
\end{aligned}
\tag{D.6}
$$

Next, we let $\mathbf{X}(t) = \langle \hat{u}_0, ..., \hat{u}_S, \hat{w}_0, ..., \hat{w}_S, \hat{a}_0, ..., \hat{a}_S \rangle$ such that equation (D.6) can be written as $\mathbf{X}' = \mathbf{f}(\mathbf{X})$. We can then integrate equation (D.6) using a second order trapezoidal rule,

$$\frac{\mathbf{X}(t + \Delta t) - \mathbf{X}(t)}{\Delta t} \approx \frac{1}{2}\Big(\mathbf{f}[\mathbf{X}(t + \Delta t)] + \mathbf{f}[\mathbf{X}(t)]\Big). \tag{D.7}$$

We solve equation (D.7) using Picard iteration [432], where $\mathbf{X}(t)$, which is always known, is taken to be the initial guess at each timestep. The initial conditions are chosen according to equations (D.2) and (D.5).

Unless specified otherwise, results in the main chapter are produced using a discretisation where $\delta t = 0.01$ and $S = 10000$.

D.2 Signal initiation

In the main text, we focus on the materials ability to transmit signals that are initiated at the left boundary. In addition to this, results shown in the main text allow sufficient time between signal reception at the right boundary and retransmission. In figure D.1 we demonstrate how a second signal initiated from the right boundary at $t_2 = 344$ is unable to propagate, whereas a second signal initiated from the right boundary at

$t_2 = 345$ is able. In figure D.2 we demonstrate that a second signal initiated from the left boundary at $t_2 = 80$ is unable to propagate, whereas those initiated at $t_2 = 90$ and $t_2 = 100$ are able, albeit with an initially slower transmission speed. Finally, in figure D.3 we show how two signals initiated simultaneously from both the left and right boundaries collide.

D.3 Wavespeed approximation technique

In this section, we detail the technique we use to approximate the wavespeed, c, from the solution of the continuous model. The technique used to solve the continuous model numerically is described in section D.1. The numerical parameters used to solve the continuous model: for the purpose of estimating the wavespeed in table D.3, are shown in table D.1; and, for the purpose of estimating the wavespeed in figure 5.7 in the main chapter, are shown in table D.2.

We first convert space-time data to wave-location time series data by defining the location of the front, $x_0(t)$, where

$$\hat{x}_0(t) = \hat{x} \ : \ \hat{u}(\hat{x}, \hat{t}) = 0. \tag{D.8}$$
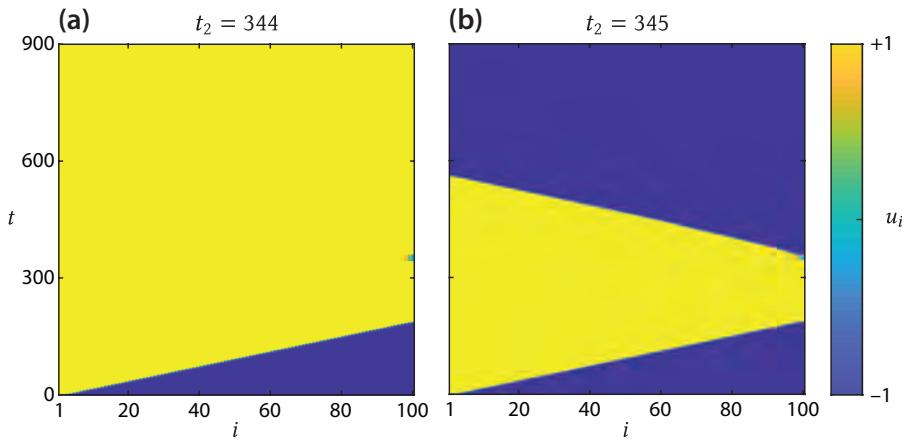
As our data is typically discrete, we find $x_0(t)$ using a linear interpolation. We now have discrete data for $\hat{x}_0(t)$ where $0 \leq \hat{t} \leq \hat{t}_{max}$. An example of this data is shown in figure D.2a.

Next, we determine a section of the wavefront curve that is at "late time" and sufficiently far from the right boundary. In this study, we found an appropriate region, to be automatically determined as
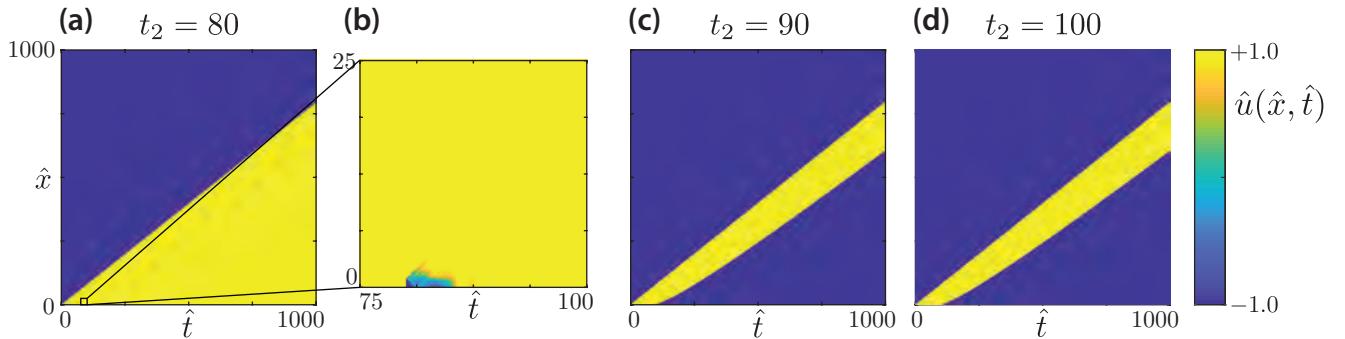
$$\hat{t} \in (\hat{t}_{lower}, \hat{t}_{upper}) = (0.8\hat{t}_{max}, 0.9\hat{t}_{max}). \tag{D.9}$$

This process gives us (approximately) linear data of the form $\{t^{(i)}, x_0^{(i)}\}$, where the slope is the estimate of the wavespeed, c. An example of this (approximately) linear data is shown in figure D.2b.
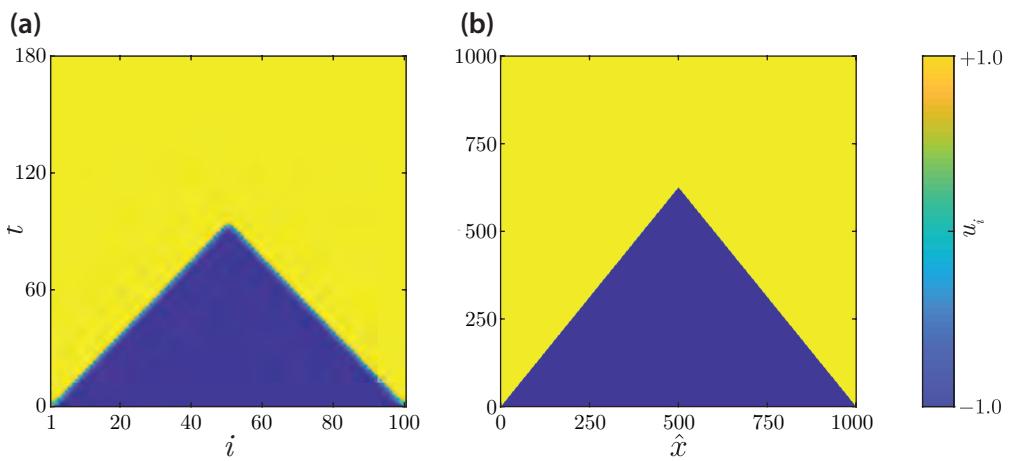
To approximate c, we linearly interpolate the data $\{t^{(i)}, x_0^{(i)}\}$, and choose c to be the gradient of the
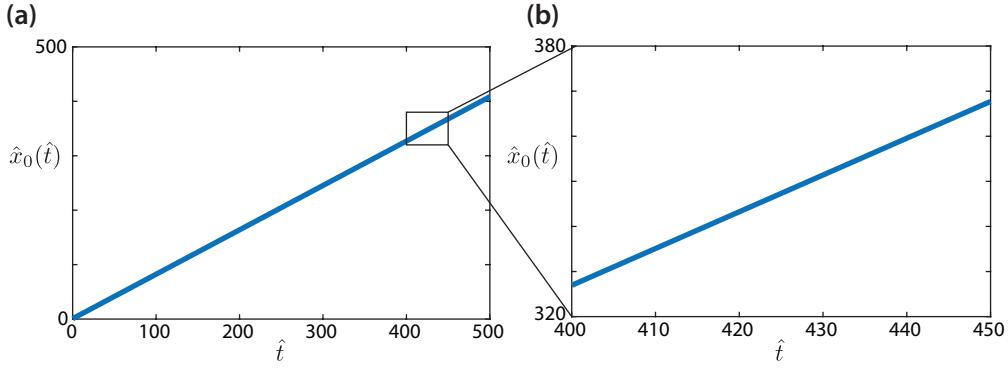

▲ Figure D.1. Signal propagation through the material described by the discrete model showing $u_i(t)$, the displacement, where i is the mass index. The first signal was initiated by moving the first element from a displacement of $-\delta$ to δ at $t = 0$, and was retransmitted by moving the last element from a displacement of δ to $-\delta$ at $t = t_2$. In (a) the second signal was not able to propagate for this value of ϵ, in (b) the signal is able to propagate. Parameters used are $m = 1$ g, $k = 1$ g m/s^2, $\gamma = 1$ g/s, $\Delta = 0.002$ m, $\delta = 1$ m, $\epsilon = 0.01$ /s, $\eta = 2$, $\nu = 1$ g/(m^{2}s^2) and $N = 101$ masses.

▲ Figure D.2. The solution to the continuous model showing $\hat{u}(\hat{x}, \hat{t})$. The first signal is initiated by setting $\hat{u}(\hat{x}, \hat{t}) = 1$ for $\hat{x} < 1$ at $t = 0$. The second signal was initiated by setting $\hat{u}(\hat{x}, \hat{t}) = -1$ for $\hat{x} < 1$ at $t = t_2$. In (a) the second signal is not able to propagate (signal initiation is shown in the inset (b)). In (c) and (d), the second signal is able to propagate.

▲ Figure D.3. (a) The solution to the discrete model showing showing $u_i(t)$, the displacement, where i is the mass index. A signal is initiated simultaneously at each end of the domain by moving the first and last element to $u_i = 1$ at $t = 0$. (b) The solution to the continuous model showing the displacement field, $\hat{u}(\hat{x}, \hat{t})$. A signal is initiated simultaneously at each end of the domain by setting $\hat{u}(\hat{x}, \hat{t}) = -1$ for $\hat{x} \leq 1$ and $\hat{x} \geq 499$ at $t = 0$. The parameters used in the discrete model are $m = 1$ g, $k = 1$ g m/s^2, $\gamma = 1$ g/s, $\Delta = 0.002$ m, $\delta = 1$ m, $\epsilon = 0.01$ /s, $\eta = 2$, $v = 1$ g/(m^{2}s^2) and $N = 101$ masses. The parameters used in the continuous model are $v = 4$, $\eta = 2$ and $\kappa = 0.01$.

(a)

(b)

▲ Figure D.4. Example wave front time series data, produced using a solution to the continuous model for $v = 4$, $\eta = 2$ and $\kappa = 0$. Numerical parameters are detailed in table D.1. (a) Shows the front data for $\hat{t} \in [0, \hat{t}_{\max}]$. (b) Shows the inset of (a) where $\hat{t} \in (\hat{t}_{\text{lower}}, \hat{t}_{\text{upper}})$, which corresponds to data from which the wavespeed is approximated from.

regression line.

η	v	$\hat{t}_{\max}$	$\hat{x}_{\max}$	$\delta\hat{t}$	S
1.5	1	500	500	0.01	10000
1.5	2	500	500	0.01	10000
1.5	4	500	500	0.01	10000
2	1	500	500	0.01	10000
2	2	500	500	0.01	10000
2	4	500	500	0.01	10000
3	1	500	500	0.01	10000
3	2	500	500	0.01	10000
3	4	500	500	0.01	10000
4	1	500	2000	0.01	20000
4	2	500	2000	0.01	20000
4	4	500	2000	0.01	20000

◀ Table D.1. Numerical parameters used to solve the continuous model when estimating the wavespeed for $\kappa \ll 1$ in table D.3 of this appendix.

D.4 Perturbation solution

In this section, we present our method for finding a singular perturbation expansion to the travelling wave model, given by the system

$$0 = \frac{\mathrm{d}^2 f}{\mathrm{d}z^2} + \frac{c}{1 - c^2}\frac{\mathrm{d}f}{\mathrm{d}z} - \frac{v}{1 - c^2}(f^2 - 1)(f - h), \quad f(\pm\infty) = \mp 1, \tag{D.10}$$

$$0 = \frac{\mathrm{d}h}{\mathrm{d}z} + \frac{\kappa}{c}\left(\frac{f}{\eta} - h\right), \quad h(\infty) = -\frac{1}{\eta}, \tag{D.11}$$

We pose a perturbation expansion for the wavespeed,

$$c = c_0 + c_1\kappa + \mathcal{O}(\kappa^2), \tag{D.12}$$

η	ν	$\hat{t}_{\max}$	$\hat{x}_{\max}$	$\delta\hat{t}$	S
1.5	1	500	10000	0.01	10000
1.5	2	500	1000	0.01	10000
1.5	4	500	1000	0.01	10000
2	1	500	10000	0.01	10000
2	2	500	1000	0.01	10000
2	4	500	1000	0.01	10000
3	1	500	10000	0.01	10000
3	2	500	1000	0.01	10000
3	4	500	1000	0.01	10000
4	1	500	10000	0.01	20000
4	2	500	10000	0.01	20000
4	4	500	10000	0.01	20000

◄ Table D.2. Numerical parameters used to solve the continuous model when estimating the wavespeed for $\kappa > 0$ in figure 5.7 in the main chapter.

where c_0 is known. We take time to note that, for this expansion, we also have

$$\frac{1}{c} = \frac{1}{c_0} - \frac{c_1}{c_0^2}\kappa + \mathcal{O}(\kappa^2).$$

$$\frac{c}{1-c^2} = \frac{c_0}{1-c_0^2} + \left(\frac{c_1}{1-c_0^2} + \frac{2c_0^2 c_1}{(1-c_0^2)^2}\right)\kappa + \mathcal{O}(\kappa^2),$$

$$\frac{\nu}{1-c^2} = \nu\left(\frac{1}{1-c_0^2} + \frac{2c_0 c_1}{(1-c_0^2)^2}\kappa + \mathcal{O}(\kappa^2)\right).$$

We now proceed to find perturbation solutions in the inner and outer regions. We denote solutions in the inner region $z \sim \mathcal{O}(\mu_0^{-1})$ using lowercase variables, $f(z)$ and $h(z)$. We denote solutions in the outer region $Z = \kappa z \sim \mathcal{O}(\kappa^{-1})$ using uppercase variables, $F^{(L)}(Z)$, $F^{(R)}(Z)$, $H^{(L)}(Z)$ and $H^{(R)}(Z)$, where a superscript, (L) and (R), denotes solutions in left, $Z < 0$ and right, $Z > 0$ regions of the outer region, respectively.

Outer Region

Substituting $Z = \kappa z$ into equations (D.10) and (D.11) leads to a system described by

$$0 = \kappa^2 \frac{d^2 F}{dZ^2} + \frac{c\kappa}{1-c^2}\frac{dF}{dZ} - \frac{\nu}{1-c^2}(F^2 - 1)(F - H),$$

$$0 = \frac{dH}{dZ} + \frac{1}{c}\left(\frac{F}{\eta} - H\right), \tag{D.13}$$

The left and right solutions in the outer region will match the corresponding left and right boundary conditions from equations (D.10) and (D.11), such that

$$F^{(L)}(-\infty) = 1, \qquad\qquad H^{(L)}(-\infty) = \frac{1}{\eta}, \tag{D.14}$$

$$F^{(R)}(\infty) = -1, \qquad\qquad H^{(R)}(\infty) = -\frac{1}{\eta}. \tag{D.15}$$

Provided $\kappa \ll \mu_0$, for $Z \sim \mathbb{O}(1)$ the solution to the system for $\kappa = 0$ is given by

$$f\left(\frac{Z}{\kappa}\right) = \tanh\left(\frac{\mu_0}{\kappa}Z\right) \sim -\text{sign}(Z),$$

since $\mu_0/\kappa \ll 1$. It is then trivial to see that the expansion for $F(Z)$ contains only the leading order term, so that

$$F(Z) = -\text{sign}(Z) \implies \begin{array}{ll} F^{(L)}(Z) = 1, & Z < 0, \\ F^{(R)}(Z) = -1, & Z > 0. \end{array} \tag{D.16}$$

Therefore, the solution for $H(Z)$ where $Z > 0$, $H^{(R)}(Z)$, is given by

$$H^{(R)}(Z) = -\frac{1}{\eta}. \tag{D.17}$$

To find the solution for $H(Z)$ where $Z < 0$, $H^{(L)}(Z)$, we pose a perturbation solution of the form

$$H^{(L)}(Z) = H_0 + H_1\kappa + \mathbb{O}(\kappa^2). \tag{D.18}$$

Substituting this perturbation expansion into equation (D.13) gives

$$0 = H_0' + \frac{1}{c_0}\left(\frac{1}{\eta} - H_0\right) + \left[H_1' - \frac{1}{c_0}H_1 + \frac{c_1}{c_0^2}\left(\frac{1}{\eta} - H_0\right)\right]\kappa + \mathbb{O}(\kappa^2), \tag{D.19}$$

where the boundary conditions are given by

$$\lim_{Z\to-\infty} = \begin{cases} -\dfrac{1}{\eta}, & i = 0, \\[2mm] 0, & i > 1. \end{cases} \tag{D.20}$$

The $\mathbb{O}(1)$ term in equation (D.19) gives the separable ordinary differential equation

$$0 = H_0' + \frac{1}{c_0}\left(\frac{1}{\eta} - H_0\right), \ Z < 0, \ \lim_{Z\to-\infty}H_0(Z) = \frac{1}{\eta}, \tag{D.21}$$

which has the solution

$$H_0(Z) = \alpha_1 \exp\left(\frac{Z}{c_0}\right) + \frac{1}{\eta},$$

where α_1 is to be determined by matching solutions in the inner and outer regions to $\mathbb{O}(1)$.

The $\mathbb{O}(\kappa)$ term in equation (D.19) gives the ordinary differential equation

$$0 = H_1' - \frac{1}{c_0}H_1 + \frac{c_1}{c_0^2}\alpha_1 \exp\left(\frac{Z}{c_0}\right), \ Z < 0, \ \lim_{Z\to-\infty}H_1(Z) = 0, \tag{D.22}$$

which has the solution

$$H_1(Z) = \left(\alpha_2 - \frac{c_1}{c_0^2}\alpha_1 Z\right)\exp\left(\frac{Z}{c_0}\right), \tag{D.23}$$

where α_2 is to be determined by matching solutions in the inner and outer regions to $\mathbb{O}(\kappa)$.

Inner Region

In the inner region, we pose a perturbation solution of the form

$$f(z) = f_0(z) + f_1(z)\kappa + \mathcal{O}(\kappa^2), \tag{D.24}$$

$$h(z) = h_0(z) + h_1(z)\kappa + \mathcal{O}(\kappa^2), \tag{D.25}$$

where f_0 and h_0 are the solutions for $\kappa = 0$,

$$f_0(z) = -\tanh(\mu_0 z), \tag{D.26}$$

$$h_0(z) = -\frac{1}{\eta}. \tag{D.27}$$

Substituting equations (D.24) and (D.25) into equations (D.10) and (D.11) gives

$$
\begin{aligned}
0 = &\frac{d^2 f_0}{dz^2} + \frac{1}{c_0}\frac{df_0}{dz} - \frac{\nu}{1 - c_0^2}(f_0^2 - 1)(f_0 - h_0) \\
&+ \left\{ \frac{d^2 f_1}{dz^2} + \frac{1}{c_0}\frac{df_1}{dz} - \frac{\nu}{1 - c_0^2}\left(2f_0(f_0 - h_0) + (f_0^2 - 1) \right)f_1 \right. \\
&+ \left(\frac{c_1}{1 - c_0^2} + \frac{2c_0^2 c_1}{(1 - c_0^2)^2} \right)\frac{df_0}{dz} - \frac{2\nu c_0 c_1}{(1 - c_0^2)^2}(f_0^2 - 1)(f_0 - h_0) \\
&\left. + \frac{\nu}{1 - c_0^2}(f_0^2 - 1)h_1 \right\}\kappa + \mathcal{O}(\kappa^2),
\end{aligned}
\tag{D.28}
$$

$$0 = \frac{dh_0}{dz} + \left\{ \frac{dh_1}{dz} + \frac{1}{c_0}\left(\frac{f_0}{\eta} - h_0 \right) \right\}\kappa + \mathcal{O}(\kappa^2). \tag{D.29}$$

The $\mathcal{O}(\kappa)$ term in equation (D.29) is directly integrable. Consider

$$0 = \frac{dh_1}{dz} + \frac{1}{c_0}\left(\frac{f_0}{\eta} - h_0 \right), \tag{D.30}$$

which has solution

$$h_1(z) = \frac{1}{c_0\mu_0\eta}\left(\log(\cosh(\mu_0 z)) - \mu_0 z + \alpha_3 \right), \tag{D.31}$$

where α_3 is to be determined by matching solutions in the inner and outer regions to $\mathcal{O}(\kappa)$.

To solve for $f_1(z)$, we consider that the $\mathcal{O}(\kappa)$ term in equation (D.28) can be written as

$$0 = \frac{d^2 f_1}{dz^2} + p(z)\frac{df_1}{dz} + q(z)f_1 + c_1 r(z) + s(z), \quad \lim_{z\to\pm\infty} f_1(z) = 0, \tag{D.32}$$

where c_1 is a yet undetermined correction term to the wave-speed, and

$$p(z) = \frac{1}{c_0},$$

$$q(z) = -\frac{v}{1 - c_0^2}\left(3\tanh^2(\mu_0 z) - \frac{2}{\eta}\tanh(\mu_0 z) - 1\right),$$

$$r(z) = -\frac{\text{sech}^2(\mu_0 z)}{1 - c_0^2}\left[\mu_0\left(1 + \frac{2c_0^2}{1 - c_0^2}\right) + \frac{2c_0 v}{1 - c_0^2}\left(\tanh(\mu_0 z) - \frac{1}{\eta}\right)\right],$$

$$s(z) = -\frac{v\,\text{sech}^2(\mu_0 z)}{(1 - c_0^2)c_0\mu_0\eta}\left(\log(\cosh(\mu_0 z)) - \mu_0 z + \log(2)\right).$$

(D.33)

The boundary conditions are chosen so that $f_1(z)$ is able to match the correction term in the outer solution, $F_1(Z) \equiv 0$. We describe the numerical technique we use to solve this boundary value problem in section D.6 of this appendix.

Matching

Here, we apply van Dyke's matching rule [377] to determine the integration constants that have appeared in each solution.

First, we match the fast solution to the slow right solution to $\mathcal{O}(\kappa)$,

$$\lim_{\kappa \to 0} H^{(R)}(\kappa z) = \lim_{\kappa \to \infty}\left[h_0\left(\frac{Z}{\kappa}\right) + h_1\left(\frac{Z}{\kappa}\right)\kappa\right],$$

$$-\frac{1}{\eta} = -\frac{1}{\eta} + \lim_{\kappa \to \infty}\frac{1}{c_0\mu\eta}\left(\log\left(\cosh\left(\frac{\mu}{\kappa}Z\right)\right) - \frac{\mu}{\kappa}Z + \alpha_3\right),$$

$$\Rightarrow \alpha_3 = \log(2).$$

Next, we match the fast solution to the slow left solution to $\mathcal{O}(1)$,

$$\lim_{\kappa \to 0} H_0(\kappa z) = \lim_{\kappa \to \infty} h_0\left(\frac{Z}{\kappa}\right),$$

$$\lim_{\kappa \to 0}\left(\alpha_1\exp\left(\frac{\bar{c}_0\kappa\zeta}{k}\right) + \frac{1}{\eta}\right) = \lim_{\kappa \to \infty} -\frac{1}{\eta},$$

$$\Rightarrow \alpha_1 = -\frac{2}{\eta},$$

and to $\mathcal{O}(\kappa)$,

$$\lim_{\kappa \to 0} H_1(\kappa z)\kappa = \lim_{\kappa \to \infty} h_1\left(\frac{Z}{\kappa}\right)\kappa,$$

$$\lim_{\kappa \to 0}\left(\alpha_2 - \frac{2c_1}{\eta c_0^2}\kappa z\right)\exp\left(\frac{z}{\kappa c_0}\right)\kappa = \lim_{\kappa \to \infty}\frac{1}{c_0\mu\eta}\left(\log\left(\cosh\left(\frac{\mu}{\kappa}Z\right)\right) - \frac{\mu}{\kappa}Z + \log(2)\right)\kappa,$$

$$\Rightarrow \alpha_2 = 0.$$

In summary, we have that

$$F(Z) = -\text{sign}(Z),$$

$$f(z) = -\tanh(\mu_0 z) + f_1(z)\kappa + \mathcal{O}(\kappa^2),$$

$$H^{(L)}(Z) = \frac{1}{\eta} - \frac{2}{\eta}\exp\left(\frac{Z}{c_0}\right) - \frac{2c_1}{\eta c_0^2}Z\exp\left(\frac{Z}{c_0}\right)\kappa + \mathcal{O}(\kappa^2),$$

$$H^{(R)}(Z) = -\frac{1}{\eta},$$

$$h(z) = -\frac{1}{\eta} + \frac{1}{c_0\mu_0\eta}\left(\log(\cosh(\mu_0 z)) - \mu_0 z + \log(2)\right)\kappa + \mathcal{O}(\kappa^2).$$

D.5 Numerical solution of the boundary value problem

In section D.4 of this appendix, we find that $f_1(z)$ and c_1 are described by the boundary value problem

$$0 = \frac{d^2 f_1}{dz^2} + p(z)\frac{df_1}{dz} + q(z)f_1 + c_1 r(z) + s(z), \quad \lim_{z\to\pm\infty} f_1(z) = 0, \tag{D.34}$$

where c_1 is a yet undetermined correction term to the wavespeed, and

$$p(z) = \frac{1}{c_0},$$

$$q(z) = -\frac{\nu}{1-c_0^2}\left(3\tanh^2(\mu_0 z) - \frac{2}{\eta}\tanh(\mu_0 z) - 1\right),$$

$$r(z) = -\frac{\text{sech}^2(\mu_0 z)}{1-c_0^2}\left[\mu_0\left(1 + \frac{2c_0^2}{1-c_0^2}\right) + \frac{2c_0\nu}{1-c_0^2}\left(\tanh(\mu_0 z) - \frac{1}{\eta}\right)\right], \tag{D.35}$$

$$s(z) = -\frac{\nu\,\text{sech}^2(\mu_0 z)}{(1-c_0^2)c_0\mu_0\eta}\left(\log(\cosh(\mu_0 z)) - \mu_0 z + \log(2)\right).$$

We approximate the solution to the infinite-domain boundary value problem given by equation (D.34) as the solution to the finite domain problem,

$$0 = \frac{d^2 Y}{dz^2} + p(z)\frac{dY}{dz} + q(z)Y + c_1 r(z) + s(z), \quad Y(\pm L) = 0, \tag{D.36}$$

where $f_1(z) = \lim_{L\to\infty} Y(z)$. We find that $L = 10$ is sufficient to obtain accurate estimates of $f_1(z)$ and c_1.

Next, we divide the domain $z \in [-L, L]$ into S equally space subintervals, each of with δz. We index the boundary of each subinterval with $n \in \{0, 1, ..., S\}$, such that $Y_n(z) \approx Y(n\delta z)$. Doing this we obtain the following system of linear equations in terms of the unknowns $Y_1, ... Y_{S-1}, c_1$,

$$-s_n = \left(\frac{1}{\delta z^2} - \frac{p_n}{2\delta z}\right)Y_{n-1} + \left(-\frac{2}{\delta z^2} + q_n\right)Y_n + \left(\frac{1}{\delta z^2} + \frac{p_n}{2\delta z}\right)Y_{n+1} + r_n c_1, \tag{D.37}$$

$$1 \leq n \leq S - 1,$$

where $Y_0 = Y_S = 0$ enforces the boundary conditions, $s_n = s(n\delta z)$, and similar for $p(z), q(z)$ and $r(z)$.

Denoting the unknowns $\mathbf{X} = [Y_1, ..., Y_{S-1}, c_1]^T \in \mathbb{R}^S$, we may write the system as the under-determined

matrix system

$$\mathbf{AX} = \mathbf{B}, \tag{D.38}$$

where $\mathbf{B} = [s_2, ..., s_S]^T \in \mathbb{R}^S$ and $\mathbf{A} \in \mathbb{R}^{S\times(S-1)}$. We obtain a solution $\mathbf{X}$ from the underdetermined system using the Moore-Penrose pseudoinverse [433, 434] with the `pinv` command in MATLAB [434], which gives the solution with the smallest norm of all possible solutions [433, 434]. We find that the component of the normalised null-space relating to the free parameter, c_1, is $\mathcal{O}(10^{-5})$, so solutions to the system of equations that have $Y(Z)$ small have similar values of the free parameter, c_1. Because of this, we find that the solution obtained from the pseudoinverse matches numerical results, obtained from the solution to the continuous model, to two significant figures, providing confidence in the approximate solution to the travelling wave obtained from the continuous model.

D.6 Comparison of wavespeed calculations

η	ν	c_0		c_1		
		PDE	Eq (3.5)	PDE	BVP	Eq (3.26)
1.5	1	0.69	0.69	−0.26	−0.26	−0.30
1.5	2	0.80	0.80	−0.10	−0.10	−0.12
1.5	4	0.88	0.88	−0.03	−0.03	−0.04
2	1	0.58	0.58	−0.43	−0.43	−0.48
2	2	0.71	0.71	−0.20	−0.20	−0.22
2	4	0.82	0.82	−0.07	−0.08	−0.09
3	1	0.43	0.43	−0.67	−0.67	−0.73
3	2	0.55	0.55	−0.37	−0.37	−0.40
3	4	0.69	0.69	−0.17	−0.17	−0.19
4	1	0.33	0.33	−0.81	−0.81	−0.86
4	2	0.45	0.45	−0.49	−0.49	−0.52
4	4	0.58	0.58	−0.26	−0.27	−0.28

▲ Table D.3. Comparison of calculations of the wavespeed, c_0, and the first correction term in a perturbation solution in κ, about $\kappa = 0$, c_1, shown to two significant figures. Approximations to c_1 using the PDE are performed by fitting a fifth-order polynomial through estimates of c obtained for $\kappa \in \{0, 0.01, ..., 0.05\}$. The numerical scheme, including numerical parameters chosen, are outlined in this appendix. Equations refer to the main chapter.

Appendix E

Supplementary material for Chapter 6, "Model-based data analysis of tissue growth in thin 3D printed scaffolds"

E.1 Code availability

This appendix is supplementary to, and refers to, code available on GitHub at github.com/ap-browning/Pore-Bridging.

E.2 Numerical solution of the PDE

We write the partial differential equation (PDE) model as

$$
\begin{aligned}
\frac{\partial u}{\partial t} &= \frac{\partial}{\partial x}\left(\mathcal{D}(u)\frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial y}\left(\mathcal{D}(u)\frac{\partial y}{\partial x}\right) + f(u), & t &> t_0,\ \mathrm{x} \in \Omega, \\
\frac{\mathrm{d}u}{\mathrm{d}t} &= f(u), & t &> t_0,\ \mathrm{x} \in \partial\Omega, \\
u(\mathrm{x}, t) &= 0 & t &= t_0,\ \mathrm{x} \in \Omega, \\
u(\mathrm{x}, t) &= u_0 & t &= t_0,\ \mathrm{x} \in \partial\Omega.
\end{aligned}
\tag{E.1}
$$

Here, Ω represents the set of points inside the pore (such that $0 < x < L$ and $0 < y < L$) and $\partial\Omega$ represents the set of points on the pore boundary (such that $x \in \{0, L\}$ or $y \in \{0, L\}$). Due to the symmetry of the problem, we solve the PDE on a quarter of the full domain (figure E.1c) and apply reflecting (symmetric) boundary conditions on the interior boundary.

We used a method of lines based finite difference approach to approximate the solution of the PDE model, $u(x, t)$, with a system of ordinary differential equations (ODEs), $\{u_{i,j}(t)\}_{i,j}$. Our implementation is available in `Module/Functions/SolvePDE.jl` on GitHub. We discretise the domain into a square mesh with N^2 nodes, denoting $\mathrm{x}_{i,j} = (x_i, y_j) = ((i-1)\Delta, (j-1)\Delta)$ for $i, j = 1, 2, ..., N+1$ the location of points that surround each element (figure E.1d). This yields

$$
\begin{aligned}
\frac{\mathrm{d}u_{i,j}}{\mathrm{d}t} &= \frac{1}{2\Delta^2}\Big[\big(\mathcal{D}(u_{i,j}) + \mathcal{D}(u_{i+1,j})\big)(u_{i+1,j} - u_{i,j}) - \big(\mathcal{D}(u_{i,j}) + \mathcal{D}(u_{i-1,j})\big)(u_{i,j} - u_{i-1,j})\Big] \\
&+ \frac{1}{2\Delta^2}\Big[\big(\mathcal{D}(u_{i,j}) + \mathcal{D}(u_{i,j+1})\big)(u_{i,j+1} - u_{i,j}) - \big(\mathcal{D}(u_{i,j}) + \mathcal{D}(u_{i,j-1})\big)(u_{i,j} - u_{i,j-1})\Big] + f(u_{i,j}),
\end{aligned}
\tag{E.2}
$$
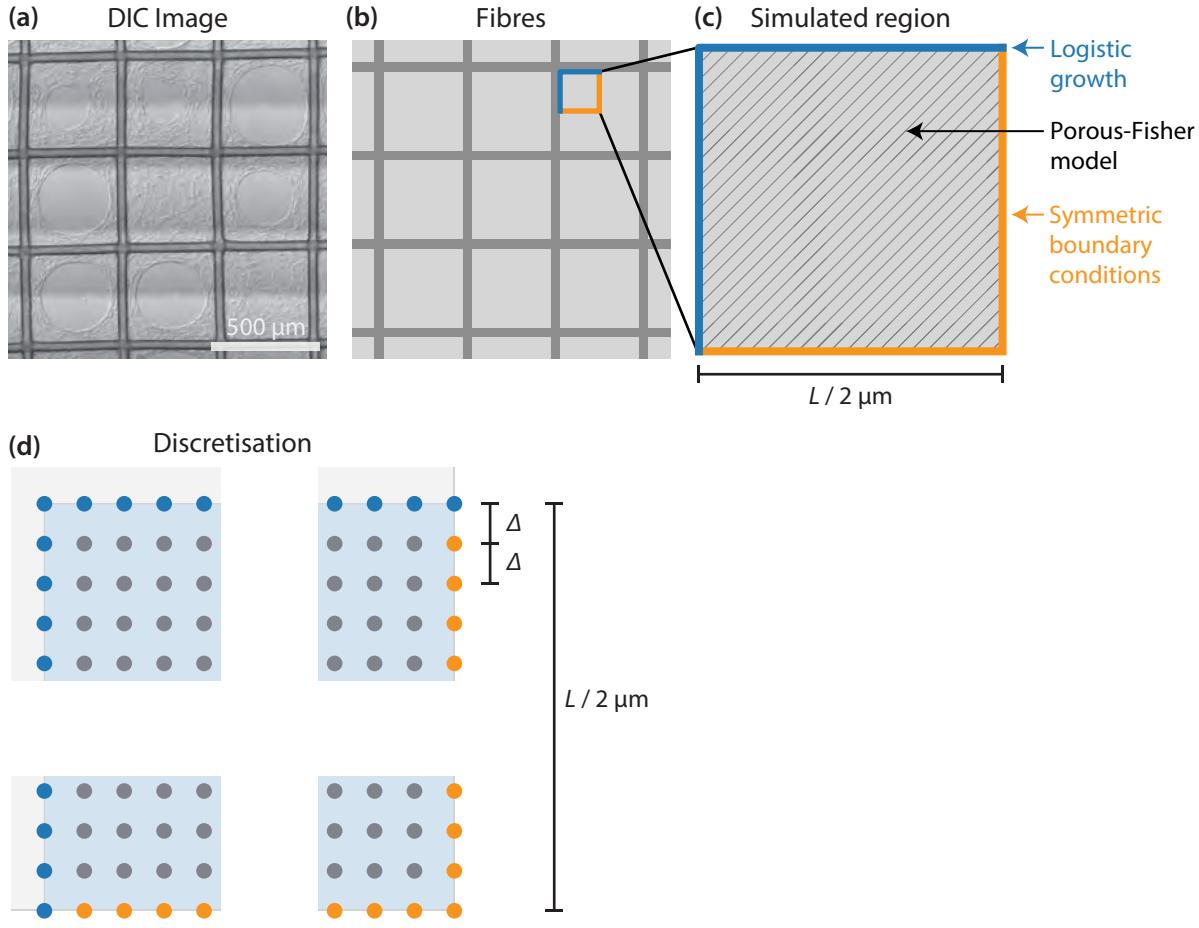
which applies on the interior nodes, $i, j = 2, 3, ..., N$ (figure E.1d, grey markers). On the interior boundary where $i = N+1$ or $j = N+1$ (figure E.1d, orange markers) we apply equation (E.2) and replace $u_{i+1,j} = u_{N-1,j}$ at $i = N+1$, $u_{i,j+1} = u_{i,N-1}$ at $j = N+1$, and $u_{i+1,j+1} = u_{N-1,N-1}$ at $i = N+1$, $j = N+1$. On pore boundary nodes (figure E.1d, blue markers) $i = 1$ or $j = 1$, the boundary condition can be applied exactly

$$
\frac{\mathrm{d}u_{i,j}}{\mathrm{d}t} = f(u_{i,j}).
\tag{E.3}
$$

We apply the initial condition by setting

$$
u_{i,j}(t_0) = \begin{cases} u_0, & i = 1 \text{ or } j = 1, \\ 0, & \text{otherwise.} \end{cases}
\tag{E.4}
$$

We integrate the ODE system using the `Tsit5` (a 4th and 5th order Runge-Kutta ODE solver with variable timestep) in `Julia` [275, 302].

(a) DIC Image

(b) Fibres

(c) Simulated region

← Logistic growth

Porous-Fisher model

← Symmetric boundary conditions

$L/2$ µm

(d) Discretisation

Δ

Δ

$L/2$ µm

▲ Figure E.1. Computational domain and finite difference discretisation.

E.3 Interpolating tissue void boundary from PDE solution
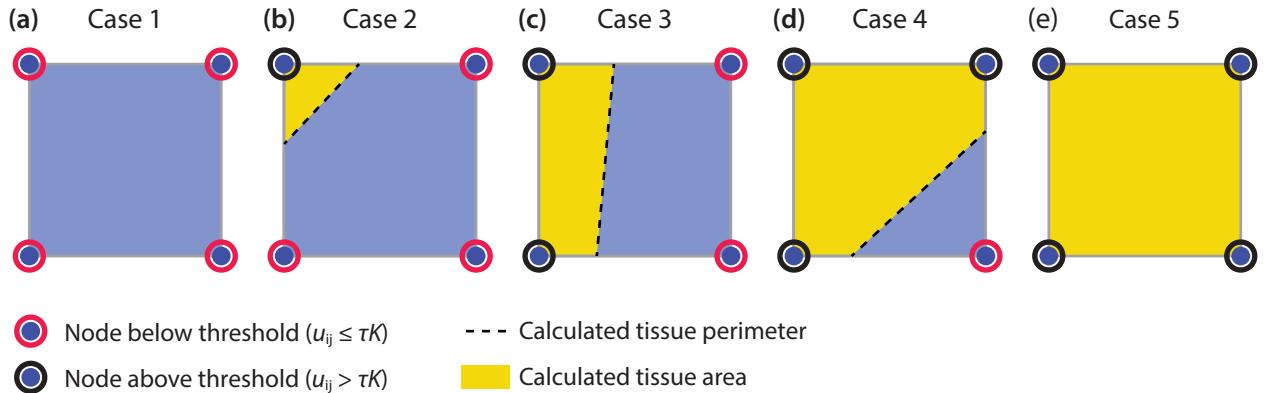
It is desirable, to ensure the optimisation routines used to calculate the maximum likelihood estimate converge to a maxima, that all summary statistics are continuous functions of the parameter space, $\boldsymbol{\theta}$. While the numerical solution of the PDE, $\{u_{ij}(t)\}_{i,j=1}^{N}$ varies continuously in $\boldsymbol{\theta}$, the naive approximation of tissue coverage calculated using nodes above the tissue boundary, $\{u_{ij}(t) > \tau K\}_{i,j=1}^{N}$, is not continuous. Therefore, we apply a linear interpolation procedure to estimate the tissue area and perimeter contribution from each finite element surrounded by nodes with tissue densities $\{u_{i,j}, u_{i+1,j}, u_{i,j+1}, u_{i+1,j+1}\}$, which we denote $A(u_1, u_2, u_3, u_4)$ and $P(u_1, u_2, u_3, u_4)$. The area and perimeter of the tissue void, used in calculation of the coverage and circularity summary statistics in the main chapter, are then given by

$$A_{\text{void}} = 4 \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} A(u_{i,j}, u_{i+1,j}, u_{i,j+1}, u_{i+1,j+1}), \tag{E.5}$$

$$P_{\text{void}} = 4 \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} P(u_{i,j}, u_{i+1,j}, u_{i,j+1}, u_{i+1,j+1}). \tag{E.6}$$

Note the factor of four as the numerical solution comprises a quarter of each pore.

The interpolation procedure is illustrated in figure E.2 and is implemented in `Module/Functions/` `SummaryStatistics.jl`, available on Github. In summary, the number of corners above the threshold,

(a) Case 1   **(b)** Case 2   **(c)** Case 3   **(d)** Case 4   **(e)** Case 5

◉ Node below threshold ($u_{ij} \leq \tau K$)   --- Calculated tissue perimeter

◉ Node above threshold ($u_{ij} > \tau K$)   ▮ Calculated tissue area

▲ Figure E.2. Interpolation procedure to calculate the area covered by tissue and the perimeter of the tissue void.
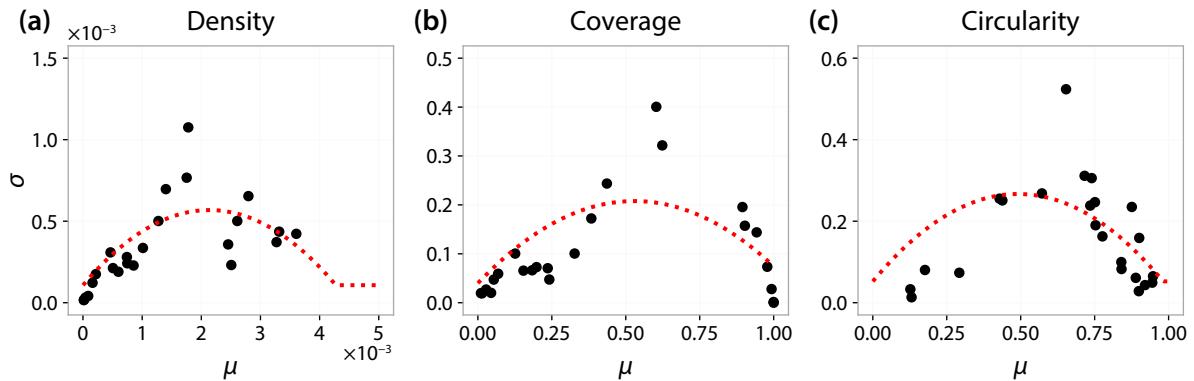
c_{above}, is first determined. The finite element in question is, therefore, represented by Case ($c_{\text{above}} + 1$) in figure E.2. In our implementation, we rotate and/or reflect the element to match the configuration of the corresponding case in figure E.2 exactly. We then linearly interpolate along the edges of the element to approximate the contour $u(t) = 0$. The region containing tissue is then approximated as a polygon (yellow region) formed by joining the interpolated points on the edges with a chord (black-dashed line). The area of this polygon, and the length of the chord, approximate the area and perimeter contribution from this element, giving $A(\cdot)$ and $P(\cdot)$.

E.4 Noise model

We pre-estimate a noise model [411] of the form

$$\sigma_i(\mu_i) = \max(0, a_i\mu_i^2 + b_i\mu_i + c_i), \tag{E.7}$$
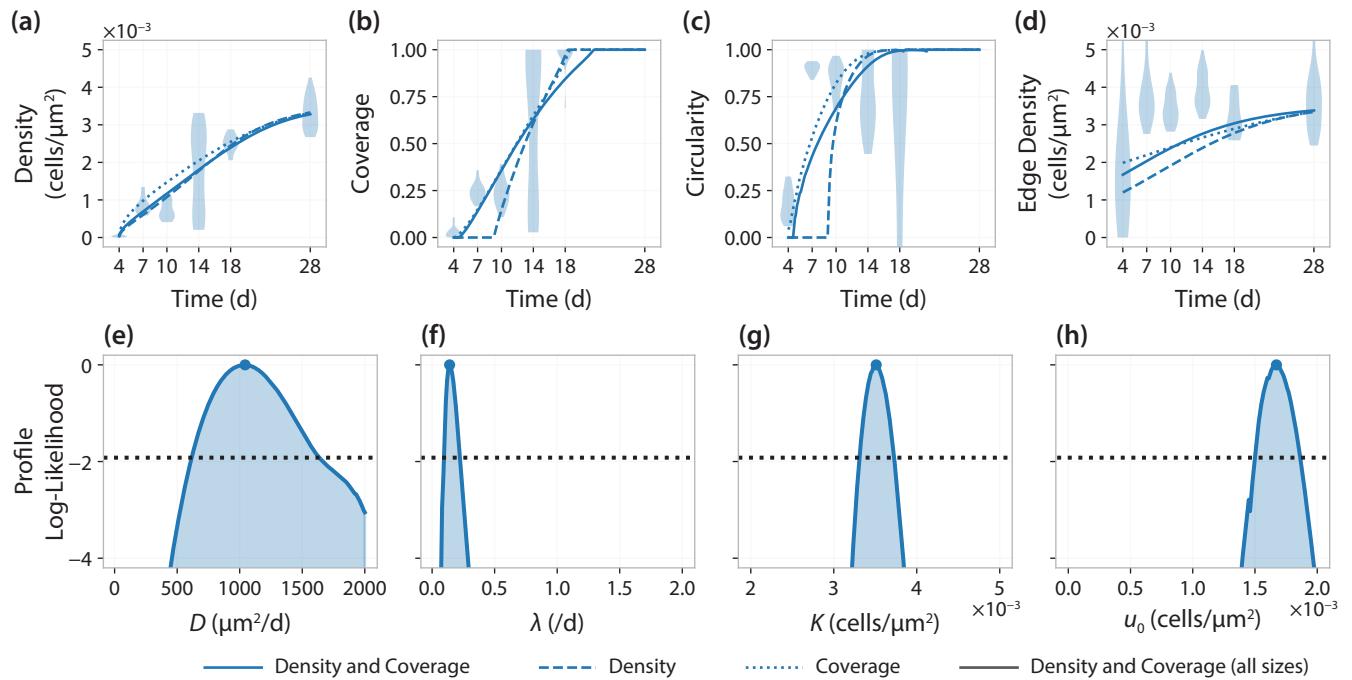
using the observed group standard deviations in the experimental data, pooled for all pore sizes figure E.3. In equation (E.7), $\sigma_i(\mu_i)$ is the standard deviation for the ith summary statistic, where μ_i is the mean (equivalent to the model prediction). In all cases, we fix the intercept c_i equal to 10% of the maximum standard deviation observed in the experimental data, and estimate a_i and b_i by minimising the sum of squares of the residual.

▲ Figure E.3. Pre-estmated noise model. Observed group standard deviations (black markers) and fitted noise model (red dashed).

E.5 Including all $L = 400\,\mu m$ data
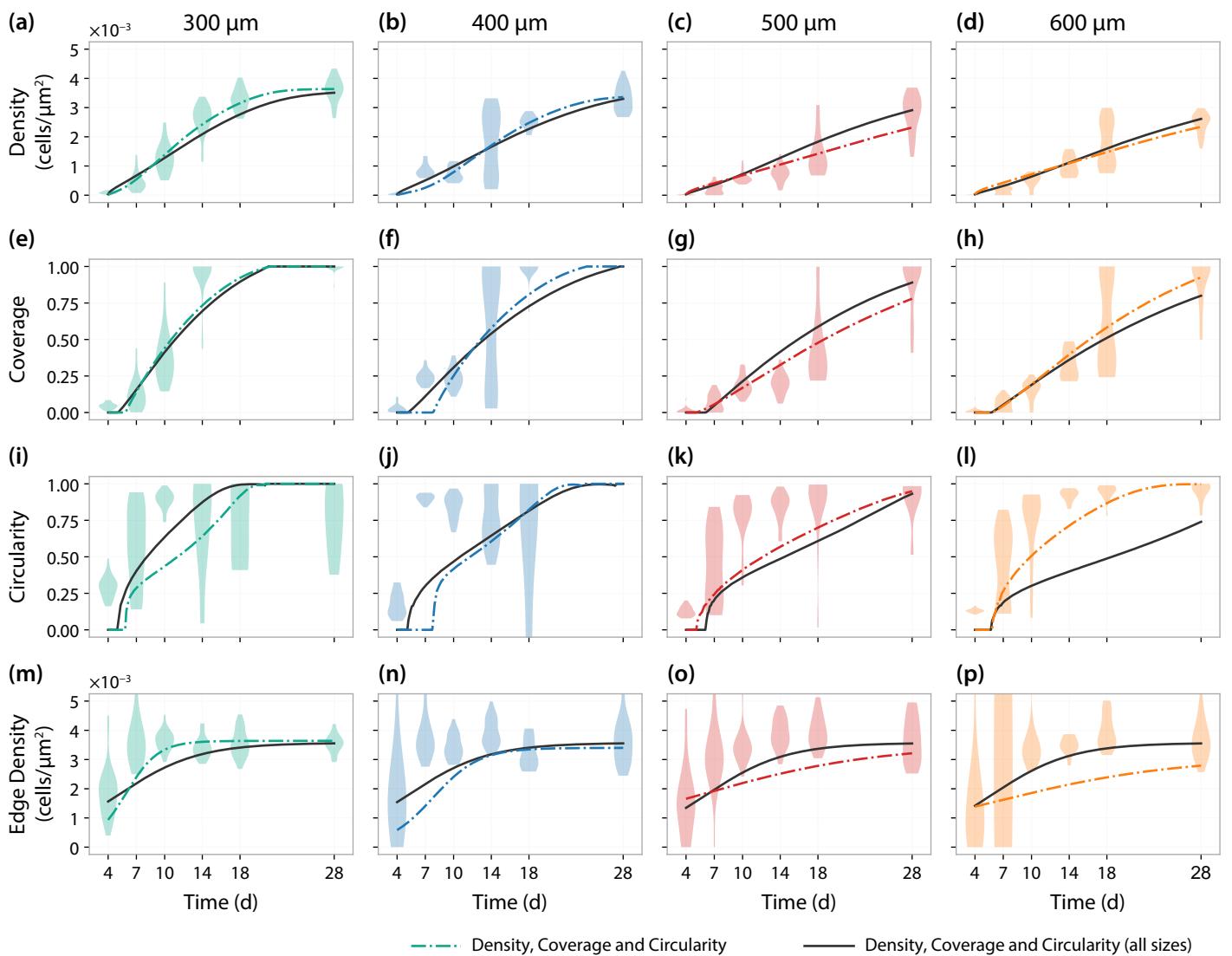
In figure E.4 we repeat results from the main chapter for the 400 μm pores in the case that all 400 μm data is included. We see results inconsistent with the inferred results for the other pore sizes, seeing very low estimates for the proliferation rate, λ. When we exclude data from $t = 7\,d$ for the main results, we see estimates for the proliferation rate that are similar to the other pore sizes, suggesting that these experimental replicates may be outliers. As scaffolds are fixed prior to staining and imaging, data from these scaffolds at earlier times is not available, and it is, therefore, difficult to ascertain why these scaffolds may exhibit behaviour different to the others.
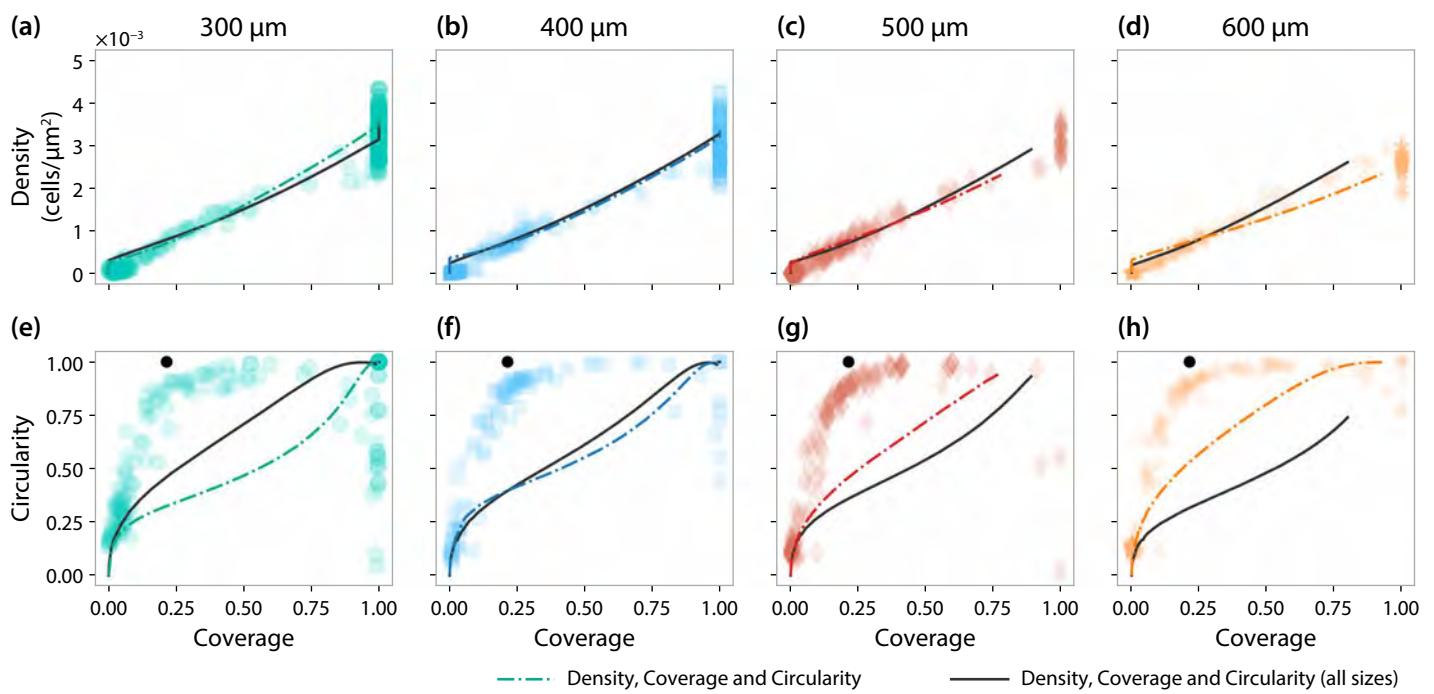

▲ Figure E.4. Results including all $L = 400\,\mu m$ data. (a–d) Violin plots show the experimental data. In each case, a model prediction is shown based on the maximum likelihood estimate where the cell density (dashed colour); cell density and tissue coverage (solid colour); tissue coverage with day 28 density measurement (dotted colour); and cell density and tissue coverage from all pore sizes (solid grey). (e–h) Profile likelihoods for each inferred parameter where both the density and coverage are used. Dotted horizontal black line indicates the -1.92 contour which corresponds to an asymptotic 95% confidence interval for each parameter.

E.6 Inference using circularity

In figure E.5 and figure E.6 we repeat results from the main chapter where the circularity is included alongside cell density and tissue coverage information. We see that, even when we calibrate the model to circularity data, we still do not match the circularity measurements seen in the experimental data, in particular for the case where inference is performed on data from all pore sizes simultaneously.

▲ **Figure E.5. Processed experimental data and model fits where the circularity is also used for inference.** Experimental data and model fit showing (a–d) the density (cells/μm²), (e–h) the coverage, (i–l) the circularity, and (m–p) the edge density. Violin plots show the experimental data. In each case, a model prediction is shown based on a maximum likelihood estimate that includes the cell density, tissue coverage, and tissue void circularity from (coloured) each individual pore size; and, (grey) simultaneously from all pore sizes.

▲ Figure E.6. Correlations between experimental and simulated summary statistics where the circularity is also used for inference. Experimental data and model fit showing the relationship between (a–d) tissue coverage and cell density, and (e–h) coverage and circularity. In each case, a model prediction is shown based on a maximum likelihood estimate that includes the cell density, tissue coverage, and tissue void circularity from (coloured) each individual pore size; and, (grey) simultaneously from all pore sizes.
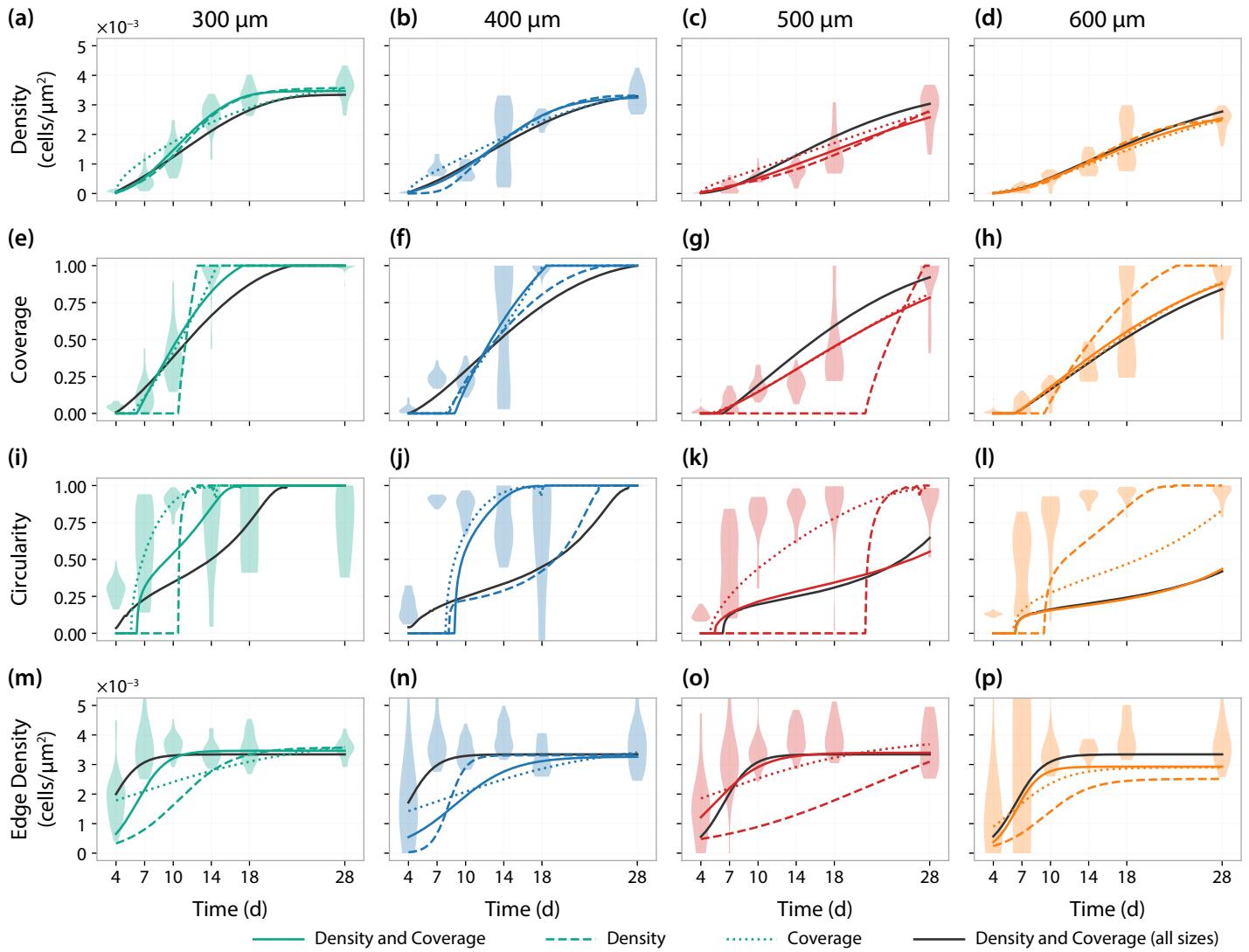
E.7 Fisher-Kolmogorov model

In the main chapter, we note that cells cannot move freely on the experimental substrate. Rather, cells work together to bridge the pore through interconnecting material such as extracellular matrix and intracellular actin filaments. This assumption leads to a governing equation of the form
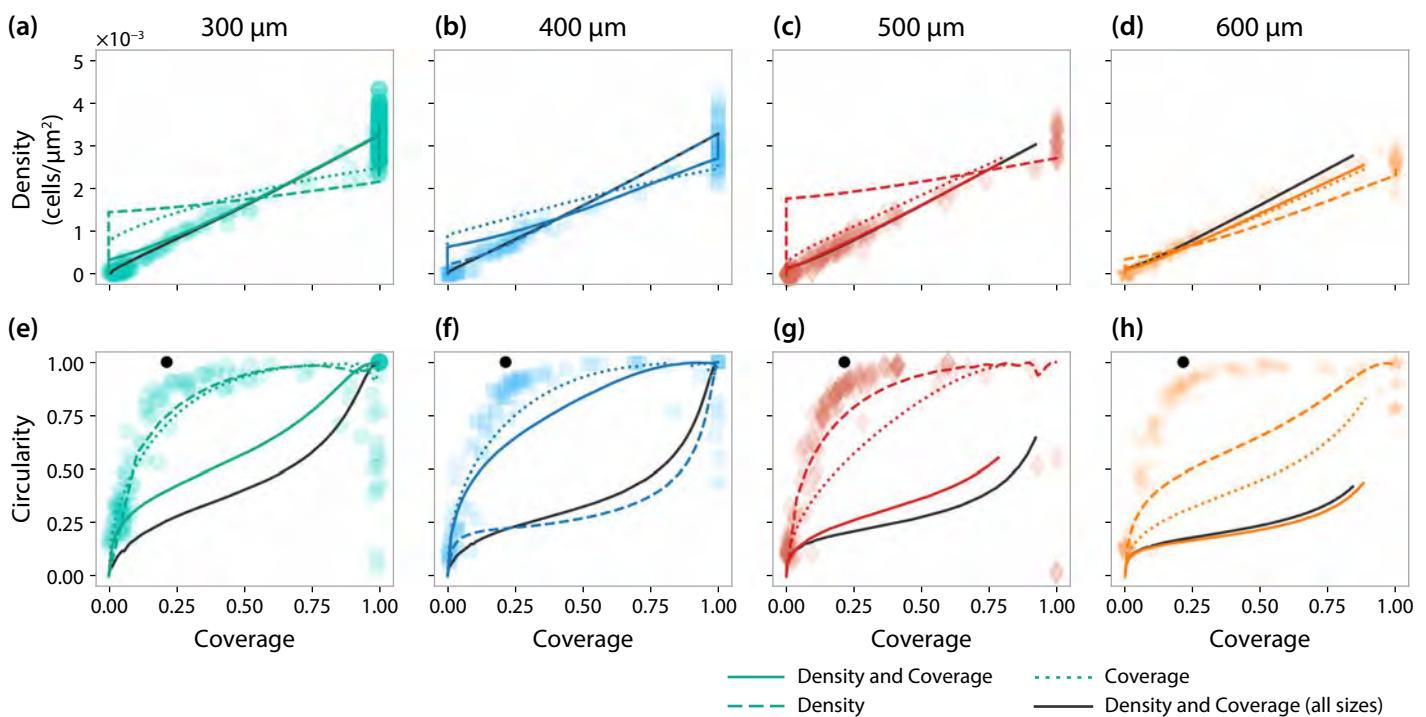
$$\frac{\partial u}{\partial t} = D\nabla \cdot \left[\left(\frac{u}{K} \right)^{\alpha} \nabla u \right] + \lambda u \left(1 - \frac{u}{K} \right), \quad \mathbf{x} \in \Omega, \tag{E.8}$$

where $\alpha = 1$, often referred to as the *Porous-Fisher equation.* Here, we relax the assumption of collaborative cell migration and reproduce the main results in the case where $\alpha = 0$, i.e., cell migration is a linear diffusion process. This choice recovers what is commonly referred to as the *Fisher-Kolmogorov model* [37], however does not align with the collaborative migration process that gives rise to tissue growth in the experiments, and will not yield a sharp-fronted tissue interface.
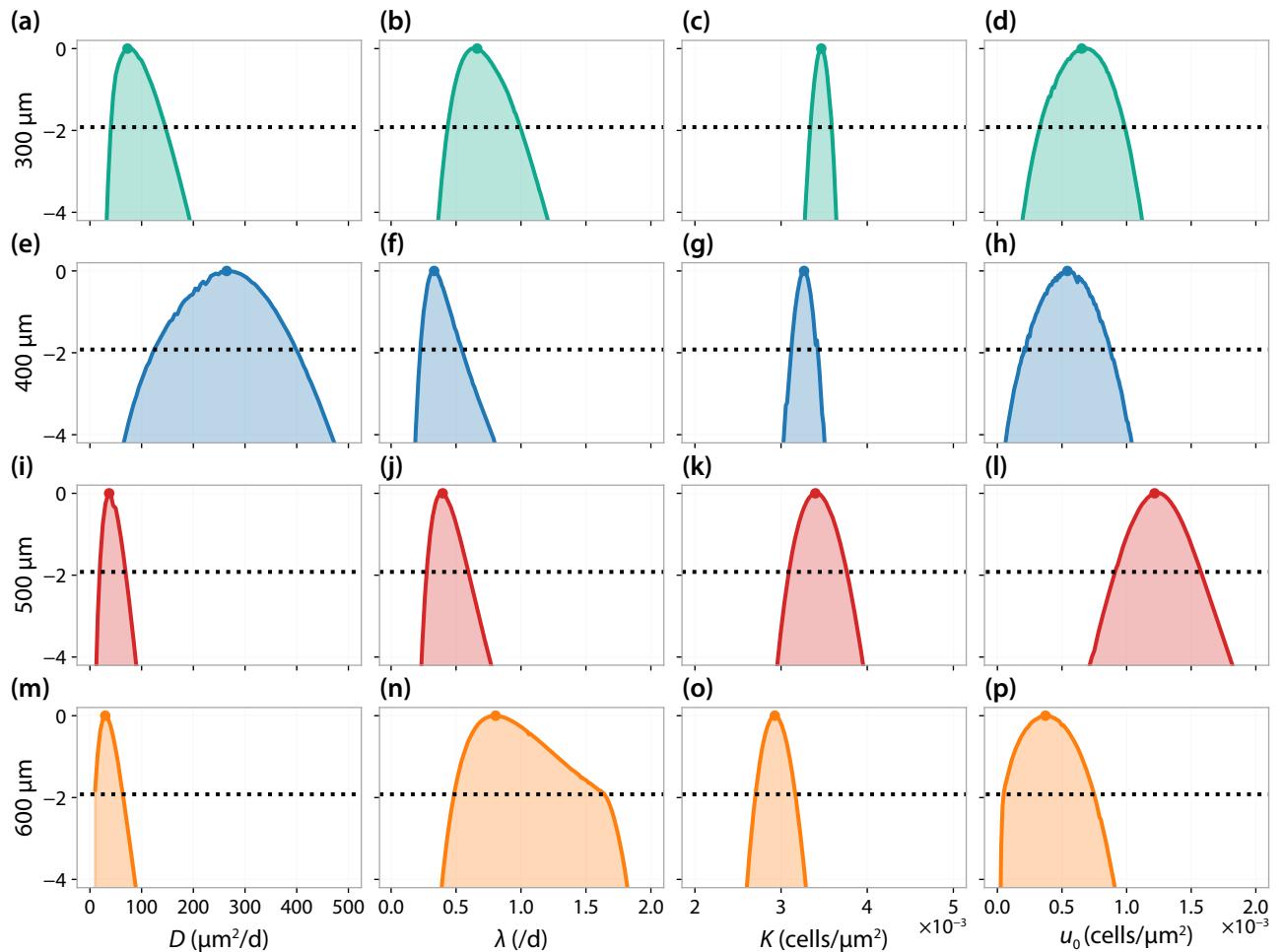
In figure E.7 and figure E.8, we compare experimental data with model best fits from the Fisher-Kolmogorov model. In figure E.9 we show profile likelihoods for this model. As with results in the main chapter, we find we are unable to match features in the experimental data relating to circularity. However, as we are able to account for both cell density and tissue coverage, we are able to quantify cell proliferation rates for each parameter, which we find to be similar to results in the main chapter.

▲ Figure E.7. Processed experimental data and model fits using the Fisher-Kolmogorov model. Experimental data and model fit showing (a–d) the density (cells/μm^2), (e–h) the coverage, (i–l) the circularity, and (m–p) the edge density. Violin plots show the experimental data. In each case, a model prediction is shown based on the maximum likelihood estimate where the cell density (dashed colour); cell density and tissue coverage (solid colour); tissue coverage with day 28 density measurement (dotted colour); and cell density and tissue coverage from all pore sizes (solid grey).

▲ Figure E.8. Correlations between experimental and simulated summary statistics using the Fisher-Kolmogorov model. Experimental data and model fit showing the relationship between (a–d) tissue coverage and cell density, and (e–h) coverage and circularity. In each case, a model prediction is shown based on the maximum likelihood estimate where the cell density (dashed colour); cell density and tissue coverage (solid colour); tissue coverage with day 28 density measurement (dotted colour); and cell density and tissue coverage from all pore sizes (solid grey).
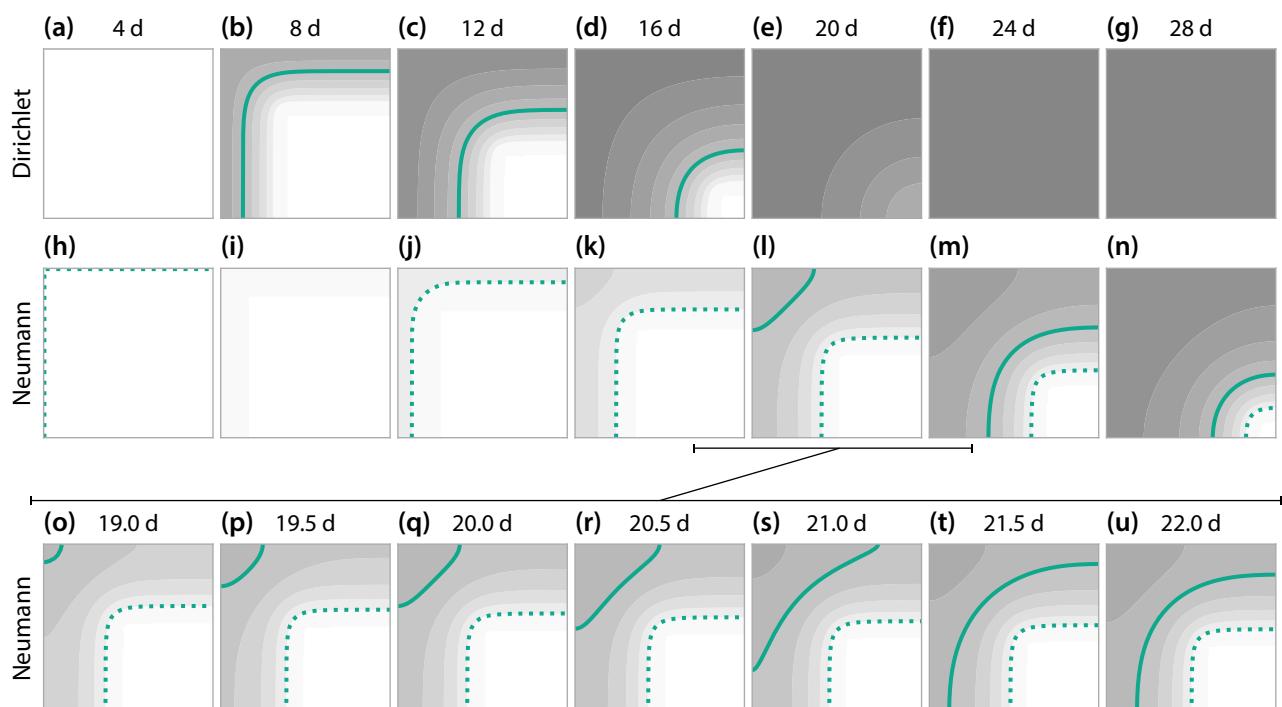
▲ Figure E.9. **Profile likelihoods for calibrated model parameters.** Profile likelihoods for each inferred parameter and pore size where only the density is used (dashed) and where both the density and coverage are used (solid). Dotted horizontal black line indicates the -1.92 contour which corresponds to an asymptotic 95% confidence interval for each parameter. Also shown are profiles for D, λ and K where information relating to cell density and tissue coverage information from all pores is included simultaneously.

E.8 Boundary conditions

In the main chapter, we apply time-dependent Dirichlet boundary conditions to the edge of each pore. This modelling assumption captures a source of cells from the fibres to the pore interior, where we assume that growth on the fibres is logistic.

In figure E.10a–g we show the solution of the Porous-Fisher equation, using parameters from the combined MLE (given in table 1 of the main chapter) in the 300 µm pores. In figure E.10h–n we show the solution to the model with homogeneous Neumann (zero normal derivative) boundary conditions applied using the same parameter combination.

Compared to the time-dependent Dirichlet boundary condition, the homogeneous Neumann boundary condition does not, initially, capture the behaviour seen in the experiments. In figure E.10o–u we show the short period of time in which the tissue becomes visible and moves off the pore edge. Cell density aggregates in the pore corners: we see a much nigher density of cells in the pore corner, always, compare to the rest of the tissue (this is not seen in the experiments). Further, tissue first appears to be convex, not smooth, in the corners. These observations, combined with the source of cells from the fibres, lead us to assume time-dependent Dirichlet boundary conditions for our analysis.

▲ Figure E.10. Comparison between time-dependent Dirichlet and homogeneous Neumann boundary conditions. Solution of the Porous-Fisher model in a 300 µm pore using the combined MLE (main chapter, table 1) using (a–g) Time-dependent Dirichlet boundary conditions; and, (h–u) Homogeneous Neumann boundary conditions. In (h–u), we show the predicted tissue boundary using the reduced threshold of $\tau = 0.1$ (dashed).
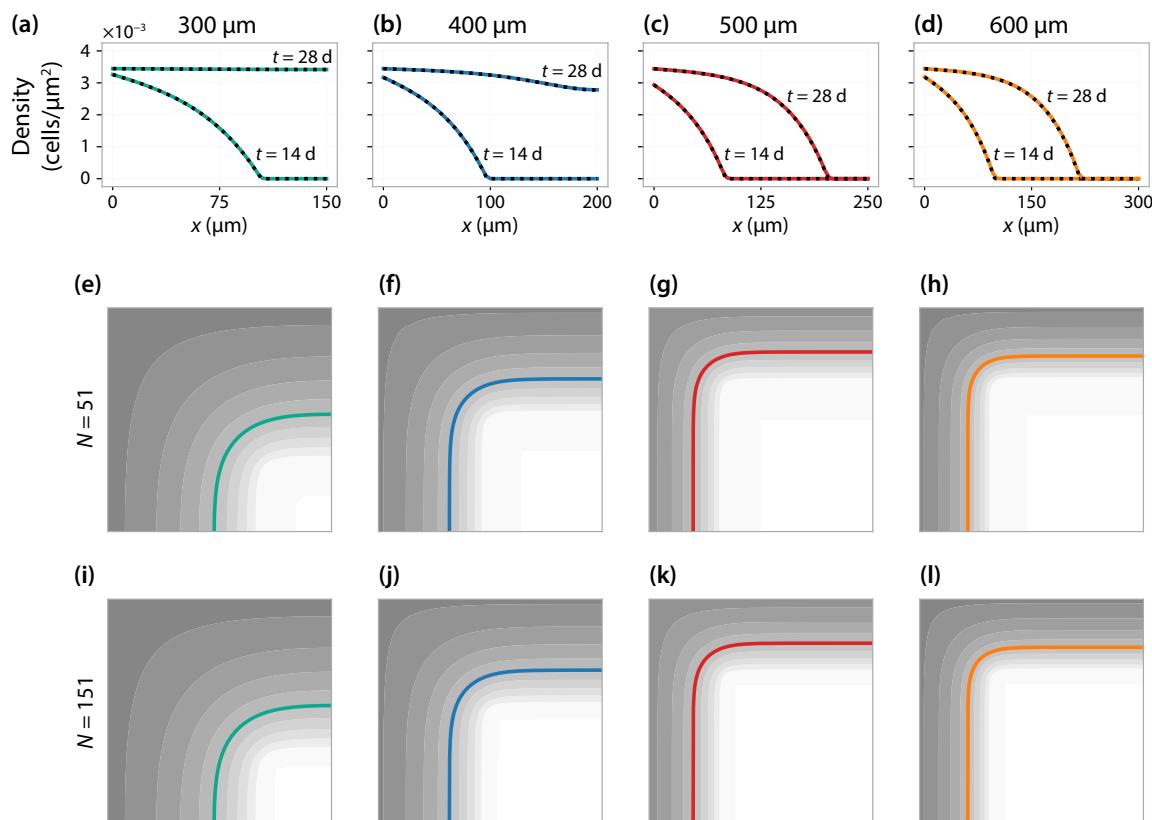
E.9 Numerical sensitivity

We solve the Porous-Fisher model numerically using a method of lines based finite difference approach (Section E.2). To keep the computational cost of inference manageable, we choose a spatial discretisation where $N = 51 \times 51$ nodes are placed in the quarter domain, as shown in figure E.1. Note that we apply an ODE solver with adaptive time-stepping to solve the resulting system of ordinary differential equations.
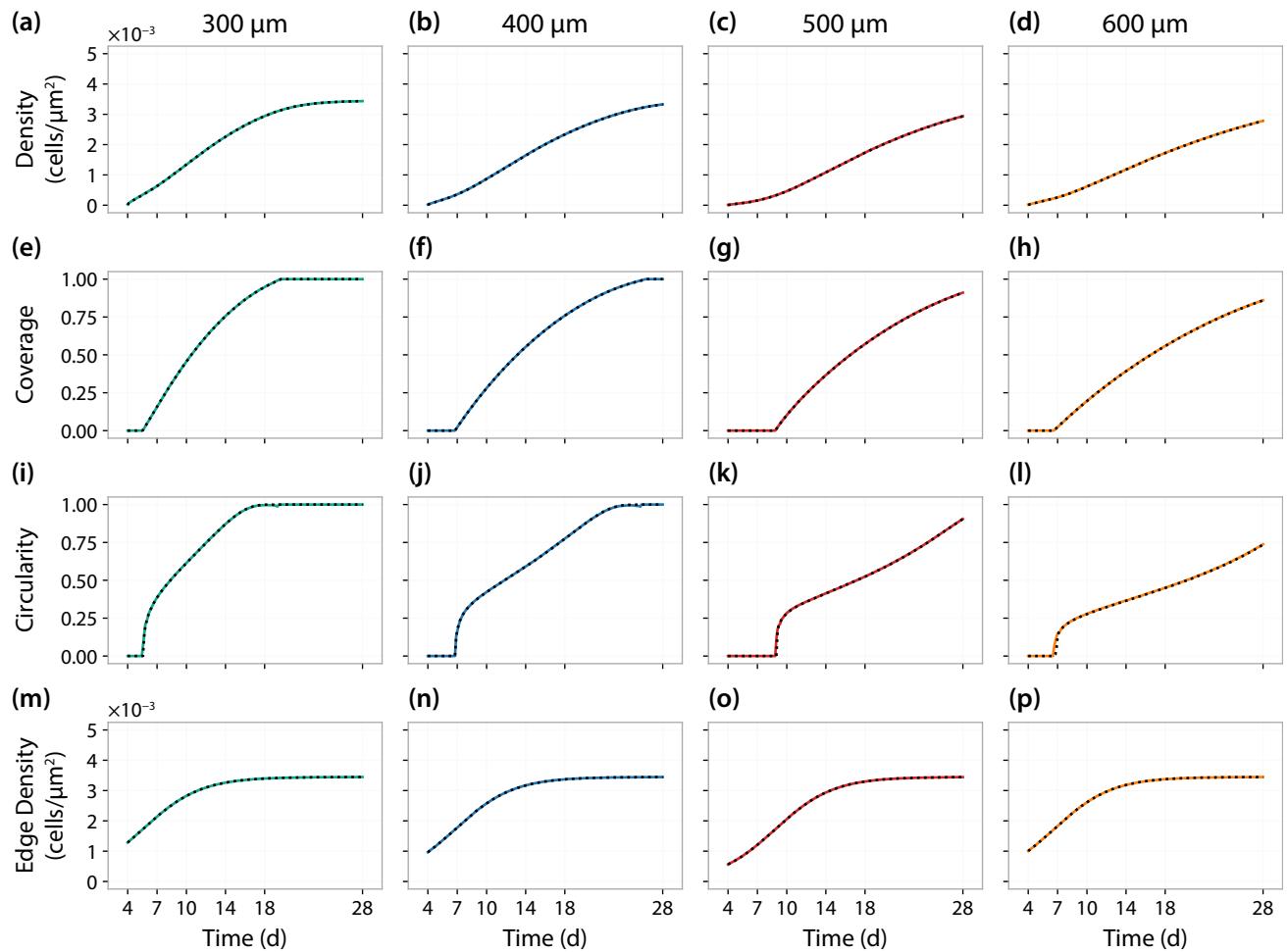
In figures E.11 and E.12 we compare the numerical solution of the model using $N = 51 \times 51$ nodes to a finer scheme where $N = 151 \times 151$ nodes are used, using parameters from the combined MLE (given in table 1 of the main chapter) for each pore size. While these results do not preclude a larger discrepancy between discretisation in other regions of the parameter space, they confirm our choice of $N = 51 \times 51$ nodes provides an acceptable level of accuracy, particularly compared to the variability in the data, without a larger computational cost.

In figure E.11a–d we show density profiles at $t = 14$ d and $t = 28$ d across the centre of each pore (the edge of the quarter-domain). While we don't expect these curves to be identical, we don't see any visual difference. In figure E.11e–h and figure E.11i–l we show the tissue boundary at $t = 14$ d for using $N = 51 \times 51$ and $N = 151 \times 151$ nodes, respectively. Visually, the choice of discretisation has minimal, if any, impact on the shape of the tissue boundary.

In figure E.12 we compare the time-evolution of the summary statistics, for each pore size, for each discretisation. In particular, results in figure E.12i–l confirm that the discretisation has minimal impact on the shape (circularity) of the pore. The exception to this is a small, noticeable artefact in circularity measurement right before pore closure.

▲ Figure E.11. Sensitivity of numerical solution to spatial discretisation I. (a–d) Cell density profile through the centre of each pore at $t = 14\,$d and $t = 28\,$d using $N = 51 \times 51$ nodes (coloured) and $N = 151 \times 151$ nodes (black dashed). (e–h) Tissue boundary at $t = 14\,$d using $N = 51$ nodes. (i–l) Tissue boundary at $t = 14\,$d using $N = 151$ nodes. All simulations used the combined MLE (given in table 1 of the main chapter).

▲ Figure E.12. Sensitivity of numerical solution to spatial discretisation II. Time-evolution of each summary statistic using $N = 51 \times 51$ nodes (coloured) and $N = 151 \times 151$ nodes (black dashed). All simulations used the combined MLE (given in table 1 of the main chapter).

Bibliography

[1] Melnikov SV, Stevens DL, Fu X, Kwok HS, Zhang JT, Shen Y, Sabina J, Lee K, Lee H, Söll D. 2020 Exploiting evolutionary trade-offs for posttreatment management of drug-resistant populations. *Proceedings of the National Academy of Sciences* **117**, 202003132.

[2] Janzén DLI, Bergenholm L, Jirstrand M, Parkinson J, Yates J, Evans ND, Chappell MJ. 2016 Parameter identifiability of fundamental pharmacodynamic models. *Frontiers in Physiology* **7**, 590.

[3] Karolak A, Markov DA, McCawley LJ, Rejniak KA. 2018 Towards personalized computational oncology: from spatial models of tumour spheroids, to organoids, to tissues. *Journal of The Royal Society Interface* **15**, 20170703.

[4] Sharp JA, Browning AP, Mapder T, Burrage K, Simpson MJ. 2019 Optimal control of acute myeloid leukaemia. *Journal of Theoretical Biology* **470**, 30–42.

[5] Gluzman M, Scott JG, Vladimirsky A. 2020 Optimizing adaptive cancer therapy: dynamic programming and evolutionary game theory. *Proceedings of the Royal Society B* **287**, 20192454.

[6] Enderling H, Wolkenhauer O. 2020 Are all models wrong? *Computational and Systems Oncology* **1**, e1008.

[7] Raney JR, Nadkarni N, Daraio C, Kochmann DM, Lewis JA, Bertoldi K. 2016 Saztable propagation of mechanical signals in soft media using stored elastic energy. *Proceedings of the National Academy of Sciences* **113**, 9722–9727.

[8] Hwang M, Arrieta AF. 2018 Input-independent energy harvesting in bistable lattices from transition waves. *Scientific Reports* **8**, 3630.

[9] Browning AP, Woodhouse FG, Simpson MJ. 2019 Reversible signal transmission in an active mechanical metamaterial. *Proceedings of the Royal Society A* **475**, 20190146.

[10] Gábor A, Banga JR. 2015 Robust and efficient parameter estimation in dynamic models of biological systems. *BMC Systems Biology* **9**, 74.

[11] Barthes J, Özçelik H, Hindié M, Ndreu-Halili A, Hasan A, Vrana NE. 2014 Cell microenvironment engineering and monitoring for tissue engineering and regenerative medicine: the recent advances. *BioMed Research International* **2014**, 1–18.

[12] Ambrosi D, Amar MB, Cyron CJ, DeSimone A, Goriely A, Humphrey JD, Kuhl E. 2019 Growth and remodelling of living tissues: perspectives, challenges and opportunities. *Journal of The Royal Society Interface* **16**, 20190233.

[13] Buenzli PR, Lanaro M, Wong CS, McLaughlin MP, Allenby MC, Woodruff MA, Simpson MJ. 2020 Cell proliferation and migration explain pore bridging dynamics in 3D printed scaffolds of different pore size. *Acta Biomaterialia* **114**, 285–295.

[14] Cabral D, Wurster J, Belenky P. 2018-03 antibiotic persistence as a metabolic adaptation: stress, metabolism, the host, and new directions. *Pharmaceuticals* **11**, 14–19.

[15] Browning AP, Haridas P, Simpson MJ. 2019 A Bayesian sequential learning framework to parameterise continuum models of melanoma invasion into human skin. *Bulletin of Mathematical Biology* **81**, 676—698.

[16] Balaban NQ, Merrin J, Chait R, Kowalik L, Leibler S. 2004 Bacterial persistence as a phenotypic switch. *Science* **305**, 1622–1625.

[17] Merton RC. 1971 Optimum Consumption and Portfolio Rules in a Continuous-Time Model. *Journal of Economic Theory* **3**, 373–413.

[18] Buttenschön A, Edelstein-Keshet L. 2020 Bridging from single to collective cell migration: A review of models and links to experiments. *PLOS Computational Biology* **16**, e1008411.

[19] Karr J, Sanghvi J, Macklin D, Gutschow M, Jacobs J, Bolival B, Assad-Garcia N, Glass J, Covert M. 2012 A whole-cell computational model predicts phenotype from genotype. *Cell* **150**, 389–401.

Bibliography

[20] Babtie AC, Stumpf MPH. 2017 How to deal with parameters for whole-cell modelling. *Journal of The Royal Society Interface* **14**, 20170237.

[21] Feng J, Kessler DA, Ben-Jacob E, Levine H. 2014 Growth feedback as a basis for persister bistability. *Proceedings of the National Academy of Sciences* **111**, 544–549.

[22] Kussell E, Koshony R, Balaban NQ, Leibler S. 2005 Bacterial Persistence: A Model of Survival in Changing Environments. *Genetics* **169**, 1807–1814.

[23] Browning AP, Warne DJ, Burrage K, Baker RE, Simpson MJ. 2020 Identifiability analysis for stochastic differential equation models in systems biology. *Journal of The Royal Society Interface* **17**, 20200652.

[24] Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U, Timmer J. 2009 Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* **25**, 1923–1929.

[25] Chiş O, Banga JR, Balsa-Canto E. 2011 GenSSI: a software toolbox for structural identifiability analysis of biological models. *Bioinformatics* **27**, 2610–2611.

[26] Villaverde AF, Banga JR. 2014 Reverse engineering and identification in systems biology: strategies, perspectives and challenges. *Journal of The Royal Society Interface* **11**, 20130505.

[27] Simpson MJ, Baker RE, Vittadello SV, Maclaren OJ. 2020 Parameter identifiability analysis for spatiotemporal models of cell invasion. *Journal of the Royal Society Interface* **17**, 20200055.

[28] Erban R, Chapman SJ. 2009 Stochastic modelling of reaction–diffusion processes: algorithms for bimolecular reactions. *Physical Biology* **6**, 046001.

[29] Kaighn ME, Narayan KS, Ohnuki Y, Lechner JF, Jones LW. 1979 Establishment and characterization of a human prostatic carcinoma cell line (PC-3). *Investigative Urology* **17**, 16–23.

[30] Browning AP, Jin W, Plank MJ, Simpson MJ. 2020 Identifying density-dependent interactions in collective cell behaviour. *Journal of The Royal Society Interface* **17**, 20200143.

[31] Liang CC, Park AY, Guan JL. 2007 *In vitro* scratch assay: a convenient and inexpensive method for analysis of cell migration *in vitro*. *Nat Protoc* **2**, 329–333.

[32] Beaumont KA, Mohana-Kumaran N, Haass NK. 2013 Modeling melanoma *in vitro* and *in vivo*. *Healthcare* **2**, 27–46.

[33] Johnston ST, McElwain DLS, Binder BJ, Ross JV, Simpson MJ. 2014 Interpreting scratch assays using pair density dynamics and approximate Bayesian computation. *Open Biology* **4**, 140097.

[34] Browning AP, McCue SW, Binny RN, Plank MJ, Shah ET, Simpson MJ. 2018 Inferring parameters for a lattice-free model of cell migration and proliferation using experimental data. *Journal of Theoretical Biology* **437**, 251–260.

[35] Bobadilla AVP, Arévalo J, Sarró E, Byrne HM, Maini PK, Carraro T, Balocco S, Meseguer A, Alarcón T. 2019 *In vitro* cell migration quantification method for scratch assays. *Journal of the Royal Society Interface* **16**, 20180709.

[36] Murray JD. 2002 *Mathematical Biology*. Berlin: Springer-Verlag 3 edition.

[37] Sherratt JA, Murray JD. 1990 Models of epidermal wound healing. *Proceedings of the Royal Society B* **241**, 29–36.

[38] Maini PK, McElwain DLS, Leavesley DI. 2004 Traveling wave model to interpret a wound-healing cell migration assay for human peritoneal mesothelial cells. *Tissue Engineering* **10**, 475–482.

[39] Gnerucci A, Faraoni P, Sereni E, Ranaldi F. 2020 Scratch assay microscopy: A reaction–diffusion equation approach for common instruments and data. *Mathematical Biosciences* **330**, 108482.

[40] Jin W, Shah ET, Penington CJ, McCue SW, Chopin LK, Simpson MJ. 2016 Reproducibility of scratch assays is affected by the initial degree of confluence: Experiments, modelling and model selection. *Journal of Theoretical Biology* **390**, 136–145.

[41] Allen LJS. 2011 *An Introduction to Stochastic Processes with Applications to Biology*. Boca Raton, Florida: Chapman & Hall/CRC Press.

[42] Gillespie DT. 1977 Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry A* **81**, 2340–2361.

[43] Plank MJ, Simpson MJ. 2012 Models of collective cell behaviour with crowding effects: comparing lattice-based and lattice-free approaches. *Journal of the Royal Society Interface* **9**, 2983–2996.

[44] Hollister SJ. 2005 Porous scaffold design for tissue engineering. *Nature Materials* **4**, 518–524.

[45] Mao AS, Mooney DJ. 2015 Regenerative medicine: Current therapies and future directions. *Proceedings of the National Academy of Sciences* **112**, 14452–14459.

[46] Li JJ, Ebied M, Xu J, Zreiqat H. 2018 Current approaches to bone tissue engineering: the interface between biology and engineering. *Advanced Healthcare Materials* **7**, 1701061.

[47] Bertoldi K, Vitelli V, Christensen J, van Hecke M. 2017 Flexible mechanical metamaterials. *Nature Reviews Materials* **2**, 17066.

[48] Fang H, Wang KW, Li S. 2017 Asymmetric energy barrier and mechanical diode effect from folding multi-stable stacked-origami. *Extreme Mechanics Letters* **17**, 7–15.

[49] Deng B, Raney JR, Tournat V, Bertoldi K. 2017 Elastic vector solitons in soft architected materials. *Physical Review Letters* **118**, 204102.

[50] Ion A, Wall L, Kovacs R, Baudisch P. 2017 Digital mechanical metamaterials. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17* pp. 977–988 Denver.

[51] Rouvala M, Amaratunga G, Wei D, Colli A. 2010 *Nanotechnologies for future mobile devices*. Cambridge: Cambridge University Press.

[52] Burgos-Morales O, Gueye M, Lacombe L, Nowak C, Schmachtenberg R, Hörner M, Jerez-Longres C, Mohsenin H, Wagner HJ, Weber W. 2021 Synthetic biology as driver for the biologization of materials sciences. *Materials Today Bio* **11**, 100115.

[53] Browning AP, Sharp JA, Mapder T, Baker CM, Burrage K, Simpson MJ. 2021a Persistence as an optimal hedging strategy. *Biophysical Journal* **120**, 133–142.

[54] Browning AP, Maclaren OJ, Buenzli PR, Lanaro M, Allenby MC, Woodruff MA, Simpson MJ. 2021b Model-based data analysis of tissue growth in thin 3D printed scaffolds. *Journal of Theoretical Biology* **528**, 110852.

[55] Swanson KR, Bridge C, Murray J, Alvord EC. 2003 Virtual and real brain tumors: using mathematical modeling to quantify glioma growth and invasion. *Journal of the Neurological Sciences* **216**, 1–10.

[56] Jackson PR, Juliano J, Hawkins-Daarud A, Rockne RC, Swanson KR. 2015 Patient-specific mathematical neuro-oncology: using a simple proliferation and invasion tumor model to inform clinical practice. *Bulletin of Mathematical Biology* **77**, 846–856.

[57] Rockne RC, Trister AD, Jacobs J, Hawkins-Daarud AJ, Neal ML, Hendrickson K, Mrugala MM, Rockhill JK, Kinahan P, Krohn KA, Swanson KR. 2015 A patient-specific computational model of hypoxia-modulated radiation resistance in glioblastoma using ^{18}F-FMISO-PET. *Journal of the Royal Society Interface* **12**, 20141174.

[58] Ward JP, King JR. 1997 Mathematical modelling of avascular-tumour growth. *Mathematical Medicine and Biology* **14**, 39–69.

[59] Lambert B, MacLean AL, Fletcher AG, Combes AN, Little MH, Byrne HM. 2018 Bayesian inference of agent-based models: a tool for studying kidney branching morphogenesis. *Journal of Mathematical Biology* **76**, 1673–1697.

[60] Nardini JT, Chapnick DA, Liu X, Bortz DM. 2016 Modeling keratinocyte wound healing dynamics: Cell–cell adhesion promotes sustained collective migration. *Journal of Theoretical Biology* **400**, 103–117.

[61] Fisher RA. 1937 The wave of advance of advantageous genes. *Annals of Eugenics* **7**, 355–369.

[62] Kolmogorov A, Petrovsky I, Piscounoff N. 1937 Étude de l'équation de la diffusion avec croissance de la quantité de matière et son application á un problème biologique. *Moscow University Mathematics Bulletin* **1**, 1–25.

[63] Fadai NT, Baker RE, Simpson MJ. 2019 Accurate and efficient discretizations for stochastic models providing near agent-based spatial resolution at low computational cost. *Journal of the Royal Society Interface* **16**, 20190421.

[64] Cai AQ, Landman KA, Hughes BD. 2006 Modelling directional guidance and motility regulation in cell migration. *Bulletin of Mathematical Biology* **68**, 25–52.

[65] Cai AQ, Landman KA, Hughes BD. 2007 Multi-scale modeling of a wound-healing cell migration assay. *Journal of Theoretical Biology* **245**, 576–594.

[66] Sengers BG, Please CP, Oreffo RO. 2007 Experimental characterization and computational modelling of two-dimensional cell spreading for skeletal regeneration. *Journal of The Royal Society Interface* **4**, 1107–1117.

[67] Simpson MJ, Baker RE, McCue SW. 2011 Models of collective cell spreading with variable cell aspect ratio: A motivation for degenerate diffusion models. *Physical Review E* **83**, 021901.

[68] McCue SW, Jin W, Moroney TJ, Lo KY, Chou SE, Simpson MJ. 2019 Hole-closing model reveals exponents for nonlinear degenerate diffusivity functions in cell biology. *Physica D* **398**, 130–140.

Bibliography

[69] Binny RN, James A, Plank MJ. 2016 Collective cell behaviour with neighbour-dependent proliferation, death and directional bias. *Bulletin of Mathematical Biology* **78**, 2277–2301.

[70] Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MPH. 2009 Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of The Royal Society Interface* **6**, 187–202.

[71] Browning AP, McCue SW, Simpson MJ. 2017 A Bayesian computational approach to explore the optimal duration of a cell proliferation assay. *Bulletin of Mathematical Biology* **10**, 1888–1906.

[72] Warne DJ, Baker RE, Simpson MJ. 2019 Simulation and inference algorithms for stochastic biochemical reaction networks: from basic concepts to state-of-the-art. *Journal of the Royal Society Interface* **16**, 20180943.

[73] Schneider CA, Rasband WS, Eliceiri KW. 2012 NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**, 671–675.

[74] Jin W, Shah ET, Penington CJ, McCue SW, Maini PK, Simpson MJ. 2017 Logistic proliferation of cells in scratch assays is delayed. *Bulletin of Mathematical Biology* **79**, 1028–1050.

[75] Plank MJ, Simpson MJ, Binny RN. 2020 Small-scale spatial structure influences large-scale invasion rates. *Theoretical Ecology* **13**, 277–288.

[76] Law R, Murrell DJ, Dieckmann U. 2003 Population growth in space and time: spatial logistic equations. *Ecology* **84**, 252–262.

[77] Forbes C, Evans M, Hastings N, Peacock B. 2011 *Statistical Distributions, Third Edition*. New Jersey: Wiley.

[78] Binder BJ, Simpson MJ. 2013 Quantifying spatial structure in experimental observations and agent-based simulations using pair-correlation functions. *Physical Review E* **88**, 022705.

[79] Binder BJ, Simpson MJ. 2015 Spectral analysis of pair-correlation bandwidth: application to cell biology images. *Royal Society Open Science* **2**, 140494.

[80] Skellam JG. 1951 Random dispersal in theoretical populations. *Biometrika* **38**, 196–218.

[81] Filippi S, Barnes CP, Cornebise J, Stumpf MP. 2013 On optimality of kernels for approximate Bayesian computation using sequential Monte Carlo. *Statistical Applications in Genetics and Molecular Biology* **12**, 87–107.

[82] Prangle D. 2016 Lazy ABC. *Statistics and Computing* **26**, 171–185.

[83] Kass RE, Raftery AE. 1995 Bayes factors. *Journal of the American Statistical Association* **90**, 773–795.

[84] Warne DJ, Baker RE, Simpson MJ. 2019 Using experimental data and information criteria to guide model selection for reaction-diffusion problems in mathematical biology. *Bulletin of Mathematical Biology* **19**, 1760–1804.

[85] Tremel A, Cai A, Tirtaatmadja N, Hughes B, Stevens G, Landman K, O'Connor A. 2009 Cell migration and proliferation during monolayer formation and wound healing. *Chemical Engineering Science* **64**, 247–253.

[86] Murphy RJ, Buenzli PR, Baker RE, Simpson MJ. 2019 A one-dimensional individual-based mechanical model of cell movement in heterogeneous tissues and its coarse-grained approximation. *Proceedings of the Royal Society A* **475**, 20180838.

[87] Stichel D, Middleton AM, Müller BF, Depner S, Klingmüller U, Breuhahn K, Matthäus F. 2017 An individual-based model for collective cancer cell migration explains speed dynamics and phenotype variability in response to growth factors. *NPJ Syst Biol Appl* **3**, 5.

[88] Nan P, Walsh DM, Landman KA, Hughes BD. 2018 Distinguishing cell shoving mechanisms. *PLoS One* **13**, e0193975.

[89] Binny RN, Law R, Plank MJ. 2020 Living in groups: Spatial-moment dynamics with neighbour-biased movements. *Ecol Model* **415**, 108825.

[90] Warne DJ, Baker RE, Simpson MJ. 2017 Optimal quantification of contact inhibition in cell populations. *Biophysical Journal* **113**, 1920–1924.

[91] Vittadello ST, McCue SW, Gunasingh G, Haass NK, Simpson MJ. 2019 Mathematical models incorporating a multi-stage cell cycle replicate normally-hidden inherent synchronization in cell proliferation. *Journal of the Royal Society Interface* **16**, 20190382.

[92] Hanahan D, Weinberg RA. 2011 Hallmarks of cancer: the next generation. *Cell* **144**, 646–674.

[93] Sarapata EA, de Pillis LG. 2014 A comparison and catalog of intrinsic tumor growth models. *Bulletin of Mathematical Biology* **76**, 2010–2024.

[94] Bobadilla AVP, Carraro T, Byrne HM, Maini PK, Alarcón T. 2019 Age structure can account for delayed logistic proliferation of scratch assays. *Bulletin of Mathematical Biology* **81**, 2706–2724.

[95] Codling EA, Plank MJ, Benhamou S. 2008 Random walk models in biology. *Journal of the Royal Society Interface* **5**, 813–834.

[96] Ozcelikkale A, Dutton JC, Grinnell F, Han B. 2017 Effects of dynamic matrix remodelling on *en masse* migration of fibroblasts on collagen matrices. *Journal of the Royal Society Interface* **14**, 20170287.

[97] Bigger JW. 1944 Treatment of staphylococcal infections with penicillin by intermittent sterilisation. *The Lancet* **244**, 497–500.

[98] Veening JW, Smits WK, Kuipers OP. 2008 Bistability, epigenetics, and bet-hedging in bacteria. *Annual Review of Microbiology* **62**, 193–210.

[99] Wolf DM, Vazirani VV, Arkin AP. 2005 Diversity in times of adversity: probabilistic strategies in microbial survival games. *Journal of Theoretical Biology* **234**, 227–253.

[100] Williams PD, Hastings A. 2011 Paradoxical persistence through mixed-system dynamics: towards a unified perspective of reversal behaviours in evolutionary ecology. *Proceedings of the Royal Society B* **278**, 1281–1290.

[101] Jia C, Qian M, Kang Y, Jiang D. 2014 Modeling stochastic phenotype switching and bet-hedging in bacteria: stochastic nonlinear dynamics and critical state identification. *Quantitative Biology* **2**, 110–125.

[102] Nichol D, Robertson-Tessi M, Jeavons P, Anderson ARA. 2016 Stochasticity in the genotype-phenotype map: implications for the robustness and persistence of bet-hedging. *Genetics* **204**, 1523–1539.

[103] Ardaševa A, Gatenby RA, Anderson ARA, Byrne HM, Maini PK, Lorenzi T. 2020 Evolutionary dynamics of competing phenotype-structured populations in periodically fluctuating environments. *Journal of Mathematical Biology* **80**, 775–807.

[104] Martín PV, Muñoz MA, Pigolotti S. 2019 Bet-hedging strategies in expanding populations. *PLOS Computational Biology* **15**, e1006529.

[105] Wood TK, Knabel SJ, Kwan BW. 2013 Bacterial persister cell formation and dormancy. *Applied and Environmental Microbiology* **79**, 7116–7121.

[106] Brauner A, Shoresh N, Fridman O, Balaban NQ. 2017 An Experimental Framework for Quantifying Bacterial Tolerance. *Biophysical Journal* **112**, 2664–2671.

[107] Şimşek E, Kim M. 2019 Power-law tail in lag time distribution underlies bacterial persistence. *Proceedings of the National Academy of Sciences* **116**, 17635–17640.

[108] Barrett TC, Mok WWK, Murawski AM, Brynildsen MP. 2019 Enhanced antibiotic resistance development from fluoroquinolone persisters after a single exposure to antibiotic. *Nature Communications* **10**, 1–11.

[109] Van den Bergh B, Michiels JE, Wenseleers T, Windels EM, Boer PV, Kestemont D, Meester LD, Verstrepen KJ, Verstraeten N, Fauvart M, Michiels J. 2016 Frequency of antibiotic application drives rapid evolutionary adaptation of *Escherichia coli* persistence. *Nature Microbiology* **1**, 16020.

[110] Bartell JA, Sommer LM, Haagensen JAJ, Loch A, Espinosa R, Molin S, Johansen HK. 2019 Evolutionary highways to persistent bacterial infection. *Nature Communications* **10**, 629.

[111] Fauvart M, Groote VND, Michiels J. 2011 Role of persister cells in chronic infections: clinical relevance and perspectives on anti-persister therapies. *Journal of Medical Microbiology* **60**, 699–709.

[112] Windels EM, Michiels JE, Fauvart M, Wenseleers T, van den Bergh B, Michiels J. 2019 Bacterial persistence promotes the evolution of antibiotic resistance by increasing survival and mutation rates. *The ISME Journal* **13**, 1239–1251.

[113] Amato SM, Orman MA, Brynildsen MP. 2013 Metabolic control of persister formation in *Escherichia coli*. *Molecular Cell* **50**, 475–487.

[114] Harms A, Maisonneuve E, Gerdes K. 2016 Mechanisms of bacterial persistence during stress and antibiotic exposure. *Science* **354**, aaf4268.

[115] Moyed HS, Bertrand KP. 1983 *hipA*, a newly recognized gene of *Escherichia coli* K-12 that affects frequency of persistence after inhibition of murein synthesis. *Journal of Bacteriology* **155**, 768–775.

[116] Alonso AA, Molina I, Theodoropoulos C. 2014 Modeling bacterial population growth from stochastic single-cell dynamics. *Applied and Environmental Microbiology* **80**, 5241–5253.

[117] Rivoire O, Leibler S. 2011 The value of information for populations in varying environments. *Journal of Statistical Physics* **142**, 1124–1166.

Bibliography

[118]   Øksendal B. 1998 *Stochastic Differential Equations*. Berlin, Heidelberg: Springer.

[119]   Turelli M. 1977 Random environments and stochastic calculus. *Theoretical Population Biology* **12**, 140–178.

[120]   Merton RC. 1976 Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics* **3**, 125 144.

[121]   Merton RC. 1973 Theory of rational option pricing. *The Bell Journal of EconomicsEcon.* **4**, 141.

[122]   Black F, Scholes M. 1973 The pricing of options and corporate liabilities. *Journal of Political Economy* **81**, 637–654.

[123]   Gaál B, Pitchford JW, Wood AJ. 2010 Exact results for the evolution of stochastic switching in variable asymmetric environments. *Genetics* **184**, 1113–1119.

[124]   Müller J, Hense BA, Fuchs TM, Utz M, Pötzsche C. 2013 Bet-hedging in stochastically switching environments. *Journal of Theoretical Biology* **336**, 144–157.

[125]   Hanson FB. 2007 *Applied stochastic processes and control for jump diffusions*. SIAM.

[126]   Hidalgo J, Pigolotti S, Muñoz MA. 2015 Stochasticity enhances the gaining of bet-hedging strategies in contact-process-like dynamics. *Physical Review E* **91**, 032114.

[127]   Garcia-Bernardo J, Dunlop M. 2015 Noise and low-level dynamics can coordinate multicomponent bet hedging mechanisms. *Biophysical Journal* **108**, 184–193.

[128]   Varughese MM, Fatti LP. 2008 Incorporating environmental stochasticity within a biological population model. *Theoretical Population Biology* **74**, 115–129.

[129]   Wong E, Zakai M. 1965 On the convergence of ordinary integrals to stochastic integrals. *The Annals of Mathematical Statistics* **36**, 1560–1564.

[130]   Thattai M, van Oudenaarden A. 2004 Stochastic gene expression in fluctuating environments. *Genetics* **167**, 523–530.

[131]   Acar M, Mettetal JT, van Oudenaarden A. 2008 Stochastic switching as a survival strategy in fluctuating environments. *Nature Genetics* **40**, 471–475.

[132]   Maisonneuve E, Gerdes K. 2014 Molecular mechanisms underlying bacterial persisters. *Cell* **157**, 539–548.

[133]   Pai A, You L. 2009 Optimal tuning of bacterial sensing potential. *Molecular Systems Biology* **5**, 1–11.

[134]   Ghosh A, Baltekin Ö, Wäneskog M, Elkhalifa D, Hammarlöf DL, Elf J, Koskiniemi S. 2018 Contact-dependent growth inhibition induces high levels of antibiotic-tolerant persister cells in clonal bacterial populations. *The EMBO Journal* **37**, e98026.

[135]   Xue B, Leibler S. 2017 Bet hedging against demographic fluctuations. *Physical Review Letters* **119**, 108103.

[136]   Roberts ME, Stewart PS. 2005 Modelling protection from antimicrobial agents in biofilms through the formation of persister cells. *Microbiology* **151**, 75–80.

[137]   Cogan NG. 2006 Effects of persister formation on bacterial response to dosing. *Journal of Theoretical Biology* **238**, 694–703.

[138]   Carvalho G, Balestrino D, Forestier C, Mathias JD. 2018 How do environment-dependent switching rates between susceptible and persister cells affect the dynamics of biofilms faced with antibiotics? *npj Biofilms and Microbiomes* **4**, 1–8.

[139]   Jablonka E, Oborny B, Molnár I, Kisdi E, Hofbauer J, Czárán T. 1995 The adaptive advantage of phenotypic memory in changing environments. *Philosophical Transactions of the Royal Society B* **350**, 133–141.

[140]   Tourigny DS. 2020 Dynamic metabolic resource allocation based on the maximum entropy principle. *Journal of Mathematical Biology* **80**, 2395–2430.

[141]   Bellman R. 1954 Dynamic programming and a new formalism in the calculus of variations. *Proceedings of the National Academy of Sciences* **40**, 231–235.

[142]   Kirk D. 2004 *Optimal Control Theory: An Introduction*. Dover Books on Electrical Engineering Series. Dover Publications.

[143]   Kushner HJ, Dupuis PG. 2001 *Numerical Methods for Stochastic Control Problems in Continuous Time*. New York: Springer Verlag 2nd edition.

[144]   Higham DJ. 2001 An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Rev.* **43**, 525–546.

[145]   Maltas J, Wood KB. 2019 Pervasive and diverse collateral sensitivity profiles inform optimal strategies to limit antibiotic resistance. *PLoS Biol.* **17**, e3000515.

[146] Nichol D, Rutter J, Bryant C, Hujer AM, Lek S, Adams MD, Jeavons P, Anderson ARA, Bonomo RA, Scott JG. 2019 Antibiotic collateral sensitivity is contingent on the repeatability of evolution. *Nature Communications* **10**, 334.

[147] de Jong IG, Haccou P, Kuipers OP. 2011 Bet hedging or not? A guide to proper classification of microbial survival strategies. *BioEssays* **33**, 215–223.

[148] Rotem E, Loinger A, Ronin I, Levin-Reisman I, Gabay C, Shoresh N, Biham O, Balaban NQ. 2010 Regulation of phenotypic variability by a threshold-based mechanism underlies bacterial persistence. *Proceedings of the National Academy of Sciences* **107**, 12541–12546.

[149] Cogan NG, Brown J, Darres K, Petty K. 2012 Optimal control strategies for disinfection of bacterial populations with persister and susceptible dynamics. *Antimicrobial Agents and Chemotherapy* **56**, 4816–4826.

[150] Sharp JA, Browning AP, Mapder T, Burrage K, Simpson MJ. 2020 Designing combination therapies using multiple optimal controls. *Journal of Theoretical Biology* **497**, 110277.

[151] Iram S, Dolson E, Chiel J, Pelesko J, Krishnan N, Güngör O, Kuznets-Speck B, Deffner S, Ilker E, Scott JG, Hinczewski M. 2020 Controlling the speed and trajectory of evolution with counterdiabatic driving. *Nature Physics* **17**, 135–142.

[152] Nichol D, Jeavons P, Fletcher AG, Bonomo RA, Maini PK, Paul JL, Gatenby RA, Anderson ARA, Scott JG. 2015 Steering evolution with sequential therapy to prevent the emergence of bacterial antibiotic resistance. *PLOS Computational Biology* **11**, e1004493.

[153] Gefen O, Gabay C, Mumcuoglu M, Engel G, Balaban NQ. 2008 Single-cell protein induction dynamics reveals a period of vulnerability to antibiotics in persister bacteria.. *Proceedings of the National Academy of Sciences* **105**, 6145–6149.

[154] Emerenini BO, Hense BA, Kuttler C, Eberl HJ. 2015 A mathematical model of quorum sensing induced biofilm detachment. *PLOS One* **10**, e0132385.

[155] Perkins TJ, Swain PS. 2009 Strategies for cellular decision-making. *Molecular Systems Biology* **5**, 326.

[156] Beaumont HJE, Gallie J, Kost C, Ferguson GC, Rainey PB. 2009 Experimental evolution of bet hedging. *Nature* **461**, 90–93.

[157] Rodriguez-Beltran J, Hernandez-Beltran JCR, DelaFuente J, Escudero JA, Fuentes-Hernandez A, MacLean RC, Peña-Miller R, Millan AS. 2018 Multicopy plasmids allow bacteria to escape from fitness trade-offs during evolutionary innovation. *Nature Ecology & Evolution* **2**, 873—881.

[158] Boettiger C, Bode M, Sanchirico JN, Lariviere J, Hastings A, Armsworth PR. 2016 Optimal management of a stochastically varying population when policy adjustment is costly. *Ecological Applications* **26**, 808–817.

[159] Haass NK, Beaumont KA, Hill DS, Anfosso A, Mrass P, Munoz MA, Kinjyo I, Weninger W. 2014 Real-time cell cycle imaging during melanoma growth, invasion, and drug response. *Pigment Cell Melanoma Research* **27**, 764–776.

[160] Kaznatcheev A, Peacock J, Basanta D, Marusyk A, Scott JG. 2019 Fibroblasts and alectinib switch the evolutionary games played by non-small cell lung cancer. *Nature Ecology & Evolution* **3**, 450–456.

[161] Stumpf MPH, Laidlaw Z, Jansen VAA. 2002 Herpes viruses hedge their bets. *Proceedings of the National Academy of Sciences* **99**, 15234–15237.

[162] Rouzine IM, Weinberger AD, Weinberger LS. 2015 An evolutionary role for HIV latency in enhancing viral transmission. *Cell* **160**, 1002–1012.

[163] Lu M, Jolly MK, Levine H, Onuchic JN, Ben-Jacob E. 2013 MicroRNA-based regulation of epithelial-hybrid-mesenchymal fate determination. *Proceedings of the National Academy of Sciences* **110**, 18144–18149.

[164] Abkowitz JL, Catlin SN, Guttorp P. 1996 Evidence that hematopoiesis may be a stochastic process *in vivo*. *Nature Medicine* **2**, 190–197.

[165] Elowitz MB, Leibler S. 2000 A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338.

[166] Wilkinson DJ. 2009 Stochastic modelling for quantitative description of heterogeneous biological systems. *Nature Reviews Genetics* **10**, 122–133.

[167] Neuert G, Munsky B, Tan RZ, Teytelman L, Khammash M, van Oudenaarden A. 2013 Systematic identification of signal-activated stochastic gene regulation. *Science* **339**, 584–587.

[168] Székely T, Burrage K. 2014 Stochastic simulation in systems biology. *Computational and Structural Biotechnology Journal* **12**, 14–25.

Bibliography

[169] Kang HW, KhudaBukhsh WR, Koeppl H, Rempała GA. 2019 Quasi-steady-state approximations derived from the stochastic model of enzyme kinetics. *Bulletin of Mathematical Biology* **81**, 1303–1336.

[170] Xu B, Kang HW, Jilkine A. 2019 Comparison of deterministic and stochastic regime in a model for Cdc42 oscillations in fission yeast. *Bulletin of Mathematical Biology* **81**, 1268–1302.

[171] Swameye I, Muller TG, Timmer J, Sandra O, Klingmuller U. 2003 Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling. *Proceedings of the National Academy of Sciences* **100**, 1028–1033.

[172] Heron EA, Finkenstädt B, Rand DA. 2007 Bayesian inference for dynamic transcriptional regulation; the Hes1 system as a case study. *Bioinformatics* **23**, 2596–2603.

[173] Locke JCW, Elowitz MB. 2009 Using movies to analyse gene circuit dynamics in single cells. *Nature Reviews Microbiology* **7**, 383–392.

[174] Young JW, Locke JCW, Altinok A, Rosenfeld N, Bacarian T, Swain PS, Mjolsness E, Elowitz MB. 2011 Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy. *Nature Protocols* **7**, 80–88.

[175] Cho H, Rockne RC. 2019 Mathematical modeling with single-cell sequencing data. *bioRxiv* p. 710640.

[176] Ritchie K, Shan XY, Kondo J, Iwasawa K, Fujiwara T, Kusumi A. 2005 Detection of non-Brownian diffusion in the cell membrane in single molecule tracking. *Biophysical Journal* **88**, 2266–2277.

[177] Isaacson SA. 2008 Relationship between the reaction–diffusion master equation and particle tracking models. *Journal of Physics A: Mathematical and Theoretical* **41**, 065003.

[178] Rienzo CD, Piazza V, Gratton E, Beltram F, Cardarelli F. 2014 Probing short-range protein Brownian motion in the cytoplasm of living cells. *Nature Communications* **5**, 5891.

[179] Schnoerr D, Grima R, Sanguinetti G. 2016 Cox process representation and inference for stochastic reaction–diffusion processes. *Nature Communications* **7**, 11729.

[180] Brückner DB, Ronceray P, Broedersz CP. 2020 Inferring the dynamics of underdamped stochastic systems. *Physical Review Letters* **125**, 058103.

[181] Golightly A, Wilkinson DJ. 2006 Bayesian sequential inference for stochastic kinetic biochemical network models. *Journal of Computational Biology* **13**, 838–851.

[182] Liepe J, Kirk P, Filippi S, Toni T, Barnes CP, Stumpf MPH. 2014 A framework for parameter estimation and model selection from experimental data in systems biology using approximate Bayesian computation. *Nature Protocols* **9**, 439–456.

[183] Schnoerr D, Sanguinetti G, Grima R. 2017 Approximation and inference methods for stochastic biochemical kinetics—a tutorial review. *Journal of Physics A: Mathematical and Theoretical* **50**, 093001.

[184] Bressloff PC. 2017 Stochastic switching in biology: from genotype to phenotype. *Journal of Physics A: Mathematical and Theoretical* **50**, 133001.

[185] Bosco DB, Kenworthy R, Zorio DAR, Sang QXA. 2015 Human mesenchymal stem cells are resistant to paclitaxel by adopting a non-proliferative fibroblastic state. *PLOS One* **10**, e0128511.

[186] Wilkinson DJ. 2012 *Stochastic Modelling for Systems Biology*. Boca Raton, Florida: CRC Press 2nd edition.

[187] Liao S, Vejchodský T, Erban R. 2015 Tensor methods for parameter estimation and bifurcation analysis of stochastic reaction networks. *Journal of The Royal Society Interface* **12**, 20150233.

[188] Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP. 2007 Universally sloppy parameter sensitivities in systems biology models. *PLOS Computational Biology* **3**, e189.

[189] Hines KE, Middendorf TR, Aldrich RW. 2014 Determination of parameter identifiability in nonlinear biophysical models: A Bayesian approach. *The Journal of General Physiology* **143**, 401–16.

[190] Roosa K, Chowell G. 2019 Assessing parameter identifiability in compartmental dynamic models using a computational approach: application to infectious disease transmission models. *Theoretical Biology and Medical Modelling* **16**, 1.

[191] Kreutz C, Raue A, Timmer J. 2012 Likelihood based observability analysis and confidence intervals for predictions of dynamic models. *BMC Systems Biology* **6**, 120.

[192] Cedersund G. 2016 Prediction uncertainty estimation despite unidentifiability: an overview of recent developments. In *Uncertainty in Biology* pp. 449–466. Cham: Springer International Publishing.

[193] Villaverde AF, Raimúndez E, Hasenauer J, Banga JR. 2019 A comparison of methods for quantifying prediction uncertainty in systems biology. *IFAC-PapersOnLine* **52**, 45–51.

[194] Bellman R, Åström K. 1970 On structural identifiability. *Mathematical Biosciences* **7**, 329–339.

[195] Cobelli C, DiStefano JJ. 1980 Parameter and structural identifiability concepts and ambiguities: a critical review and analysis. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* **239**, 7–24.

[196] Walter E. 1987 *Identifiability of Parametric Models.* London, United Kingdom: Elsevier Science & Technology.

[197] Jaqaman K, Danuser G. 2006 Linking data to models: data regression. *Nature Reviews Molecular Cell Biology* **7**, 813–819.

[198] Bellu G, Saccomani MP, Audoly S, D'Angiò L. 2007 DAISY: A new software tool to test global identifiability of biological and physiological systems. *Computer Methods and Programs in Biomedicine* **88**, 52–61.

[199] Miao H, Xia X, Perelson AS, Wu H. 2011 On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM Review* **53**, 3–39.

[200] Eisenberg MC, Hayashi MA. 2014 Determining identifiable parameter combinations using subset profiling. *Mathematical Biosciences* **256**, 116–126.

[201] Daly AC, Gavaghan D, Cooper J, Tavener S. 2018 Inference-based assessment of parameter identifiability in nonlinear biological models. *Journal of The Royal Society Interface* **15**, 20180318.

[202] Raj A, van Oudenaarden A. 2008 Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135**, 216–226.

[203] Balázsi G, van Oudenaarden A, Collins J. 2011 Cellular decision making and biological noise: from microbes to mammals. *Cell* **144**, 910–925.

[204] Bar-Joseph Z, Gitter A, Simon I. 2012 Studying and modelling dynamic biological processes using time-series gene expression data. *Nature Reviews Genetics* **13**, 552–564.

[205] Ruess J, Lygeros J. 2015 Moment-based methods for parameter inference and experiment design for stochastic biochemical reaction networks. *ACM Transactions on Modeling and Computer Simulation* **25**, 8.

[206] Soltani M, Vargas-Garcia CA, Antunes D, Singh A. 2016 Intercellular variability in protein levels from stochastic expression and noisy cell cycle processes. *PLOS Computational Biology* **12**, e1004972.

[207] Smith S, Grima R. 2018 Single-cell variability in multicellular organisms. *Nature Communications* **9**, 345.

[208] Hovorka R, Canonico V, Chassin LJ, Haueter U, Massi-Benedetti M, Federici MO, Pieber TR, Schaller HC, Schaupp L, Vering T, Wilinska ME. 2004 Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes. *Physiological Measurement* **25**, 905—920.

[209] Facchinetti A. 2016 Continuous glucose monitoring sensors: past, present and future algorithmic challenges. *Sensors* **16**, 1—12.

[210] Siekmann I, Sneyd J, Crampin E. 2012 MCMC can detect nonidentifiable models. *Biophysical Journal* **103**, 2275–2286.

[211] Choi B, Rempala GA, Kim JK. 2017 Beyond the Michaelis-Menten equation: accurate and efficient estimation of enzyme kinetic parameters. *Scientific Reports* **7**, 17018.

[212] Turner TE, Schnell S, Burrage K. 2004 Stochastic approaches for modelling *in vivo* reactions. *Computational Biology and Chemistry* **28**, 165–178.

[213] Ruess J, Milias-Argeitis A, Lygeros J. 2013 Designing experiments to understand the variability in biochemical reaction networks. *Journal of The Royal Society Interface* **10**, 20130588.

[214] Parsons TL, Lambert A, Day T, Gandon S. 2018 Pathogen evolution in finite populations: slow and steady spreads the best. *Journal of The Royal Society Interface* **15**, 20180135.

[215] Gillespie DT. 2000 The chemical Langevin equation. *The Journal of Chemical Physics* **113**, 297–306.

[216] Mummert A, Otunuga OM. 2019 Parameter identification for a stochastic SEIRS epidemic model: case study influenza. *Journal of Mathematical Biology* **79**, 705–729.

[217] Kristensen NR, Madsen H, Jørgensen SB. 2004 Parameter estimation in stochastic grey-box models. *Automatica* **40**, 225–237.

Bibliography

[218] Duun-Henriksen AK, Schmidt S, Røge RM, Møller JB, Nørgaard K, Jørgensen JB, Madsen H. 2013 Model identification using stochastic differential equation grey-box models in diabetes. *Journal of Diabetes Science and Technology* **7**, 431–440.

[219] Munsky B, Trinh B, Khammash M. 2009 Listening to the noise: random fluctuations reveal gene network parameters. *Molecular Systems Biology* **5**, 318.

[220] Leander J, Lundh T, Jirstrand M. 2014 Stochastic differential equations as a tool to regularize the parameter estimation problem for continuous time dynamical systems given discrete time measurements. *Mathematical Biosciences* **251**, 54–62.

[221] Enciso G, Kim J. 2019 Embracing noise in chemical reaction networks. *Bulletin of Mathematical Biology* **81**, 1261–1267.

[222] Enciso G, Erban R, Kim J. 2020 Identifiability of stochastically modelled reaction networks. *European Journal of Applied Mathematics* pp. 1–23.

[223] Boys RJ, Wilkinson DJ, Kirkwood TBL. 2008 Bayesian inference for a discretely observed stochastic kinetic model. *Statistics and Computing* **18**, 125–135.

[224] Ruttor A, Opper M. 2009 Efficient statistical inference for stochastic reaction processes. *Physical Review Letters* **103**, 230601.

[225] Simpson MJ, Browning AP, Drovandi C, Carr EJ, Maclaren OJ, Baker RE. 2020 Profile likelihood analysis for a stochastic model of diffusion in heterogeneous media. *Proceedings of the Royal Society A* **477**, 20210214.

[226] Dangerfield CE, Kay D, Burrage K. 2012 Modeling ion channel dynamics through reflected stochastic differential equations. *Physical Review E* **85**, 051907.

[227] Dangerfield CE, Kay D, Burrage K. 2011 Comparison of continuous and discrete stochastic ion channel models. *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* **2011**, 704–707.

[228] University, Center for Systems Science and Engineering (CSSE) at Johns Hopkins. 2020 COVID-19 data repository. `https://github.com/CSSEGISandData/COVID-19`. Accessed: 7th July 2020.

[229] Vigers T, Chan CL, Snell-Bergeon J, Bjornstad P, Zeitler PS, Forlenza G, Pyle L. 2019 cgmanalysis: an R package for descriptive analysis of continuous glucose monitor data. *PLOS One* **14**, e0216851.

[230] Hengl S, Kreutz C, Timmer J, Maiwald T. 2007 Data-based identifiability analysis of non-linear dynamical models. *Bioinformatics* **23**, 2612–2618.

[231] Chis OT, Banga JR, Balsa-Canto E. 2011 Structural identifiability of systems biology models: a critical comparison of methods. *PLOS One* **6**, e27755.

[232] Reiersøl O. 1950 Identifiability of a linear relation between variables which are subject to error. *Econometrica* **18**, 375.

[233] Pohjanpalo H. 1978 System identifiability based on the power series expansion of the solution. *Mathematical Biosciences* **41**, 21–33.

[234] White LJ, Evans ND, Lam TJGM, Schukken YH, Medley GF, Godfrey KR, Chappell MJ. 2001 The structural identifiability and parameter estimation of a multispecies model for the transmission of mastitis in dairy cows. *Mathematical Biosciences* **174**, 77–90.

[235] Maclaren OJ, Nicholson R. 2019 What can be estimated? Identifiability, estimability, causal inference and ill-posed inverse problems. *arXiv*. `https://arxiv.org/abs/1904.02826`.

[236] Margaria G, Riccomagno E, Chappell MJ, Wynn HP. 2001 Differential algebra methods for the study of the structural identifiability of rational function state-space models in the biosciences. *Mathematical Biosciences* **174**, 1–26.

[237] Saccomani MP, Audoly S, D'Angiò L. 2003 Parameter identifiability of nonlinear systems: the role of initial conditions. *Automatica* **39**, 619–632.

[238] Brouwer AF, Eisenberg MC. 2018 The underlying connections between identifiability, active subspaces, and parameter space dimension reduction. *arXiv*. `https://arxiv.org/abs/1802.05641`.

[239] Jacquez JA, Greif P. 1985 Numerical parameter identifiability and estimability: Integrating identifiability, estimability, and optimal sampling design. *Mathematical Biosciences* **77**, 201–227.

[240] Saccomani MP, Thomaseth K. 2018 The union between structural and practical identifiability makes strength in reducing oncological model complexity: a case study. *Complexity* **2018**, 1–10.

[241] Raue A, Karlsson J, Saccomani MP, Jirstrand M, Timmer J. 2014 Comparison of approaches for parameter identifiability analysis of biological systems. *Bioinformatics* **30**, 1440–1448.

[242] Meshkat N, Kuo CEz, DiStefano J. 2014 On finding and using identifiable parameter combinations in nonlinear dynamic systems biology models and COMBOS: a novel web implementation. *PLOS One* **9**, e110261.

[243] Villaverde AF, Barreiro A, Papachristodoulou A. 2016 Structural identifiability of dynamic systems biology models. *PLOS Computational Biology* **12**, e1005153.

[244] Hong H, Ovchinnikov A, Pogudin G, Yap C. 2019 SIAN: software for structural identifiability analysis of ODE models. *Bioinformatics* **35**, 2873–2874.

[245] Brouwer AF, Meza R, Eisenberg MC. 2017 Parameter estimation for multistage clonal expansion models from cancer incidence data: A practical identifiability analysis. *PLOS Computational Biology* **13**, e1005431.

[246] Johnston ST, Ross JV, Binder BJ, McElwain DLS, Haridas P, Simpson MJ. 2016 Quantifying the effect of experimental design choices for *in vitro* scratch assays. *Journal of Theoretical Biology* **400**, 19–31.

[247] Lehmann EL, Fienberg S, Casella G. 1998 *Theory of Point Estimation*. Secaucus: Springer 2nd edition.

[248] Murphy SA, van der Vaart AW. 2000 On profile likelihood. *Journal of the American Statistical Association* **95**, 449–465.

[249] Komorowski M, Costa MJ, Rand DA, Stumpf MPH. 2011 Sensitivity, robustness, and identifiability in stochastic chemical kinetics models. *Proceedings of the National Academy of Sciences* **108**, 8645–8650.

[250] Gunawan R, Cao Y, Petzold L, Doyle FJ. 2005 Sensitivity analysis of discrete stochastic systems. *Biophysical Journal* **88**, 2530–2540.

[251] Tavaré S, Balding DJ, Griffiths RC, Donnelly P. 1997 Inferring coalescence times from DNA sequence data. *Genetics* **145**, 505–518. 9071603.

[252] Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW. 1999 Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution* **16**, 1791–1798.

[253] Beaumont MA, Zhang W, Balding DJ. 2002 Approximate Bayesian computation in population genetics. *Genetics* **162**, 2025–2035. 12524368.

[254] Sunnåker M, Busetto AG, Numminen E, Corander J, Foll M, Dessimoz C. 2013 Approximate Bayesian computation. *PLOS Computational Biology* **9**, e1002803.

[255] Wilkinson RD. 2013 Approximate Bayesian computation (ABC) gives exact results under the assumption of model error. *Statistical Applications in Genetics and Molecular Biology* **12**, 129–141.

[256] Beaumont MA. 2003 Estimation of population growth or decline in genetically monitored populations. *Genetics* **164**, 1139–1160. 12871921.

[257] Andrieu C, Roberts GO. 2009 The pseudo-marginal approach for efficient Monte Carlo computations. *The Annals of Statistics* **37**, 697–725.

[258] Golightly A, Wilkinson DJ. 2008 Bayesian inference for nonlinear multivariate diffusion models observed with error. *Computational Statistics & Data Analysis* **52**, 1674–1693.

[259] Andrieu C, Doucet A, Holenstein R. 2010 Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **72**, 269–342.

[260] Golightly A, Wilkinson DJ. 2011 Bayesian parameter inference for stochastic biochemical network models using particle Markov chain Monte Carlo. *Interface Focus* **1**, 807–820.

[261] Warne DJ, Baker RE, Simpson MJ. 2020 A practical guide to pseudo-marginal methods for computational inference in systems biology. *Journal of Theoretical Biology* **496**, 110255.

[262] Nestel PJ, Whyte HM, Goodman DS. 1969 Distribution and turnover of cholesterol in humans. *Journal of Clinical Investigation* **48**, 982–991.

[263] Kermack WO, McKendrick AG. 1927 A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London Series A* **115**, 700–721.

[264] Tuncer N, Le TT. 2018 Structural and practical identifiability analysis of outbreak models. *Mathematical Biosciences* **299**, 1–18.

[265] Alahmadi A, Belet S, Black A, Cromer D, Flegg JA, House T, Jayasundara P, Keith JM, McCaw JM, Moss R, Ross JV, Shearer FM, Tun STT, Walker J, White L, Whyte JM, Ada WC, Zarebski AE. 2020 Influencing public health policy with data-informed mathematical models of infectious diseases: Recent developments and new challenges. *Epidemics* **32**, 100393.

Bibliography

[266] Topp B, Promislow K, Devries G, Miura RM, Finegood DT. 2000 A model of β-cell mass, insulin, and glucose kinetics: pathways to diabetes. *Journal of Theoretical Biology* **206**, 605–619.

[267] Karin O, Swisa A, Glaser B, Dor Y, Alon U. 2016 Dynamical compensation in physiological circuits. *Molecular Systems Biology* **12**, 886.

[268] Villaverde AF, Tsiantis N, Banga JR. 2019 Full observability and estimation of unknown inputs, states and parameters of nonlinear biological models. *Journal of The Royal Society Interface* **16**, 20190043.

[269] Engblom S. 2006 Computing the moments of high dimensional solutions of the master equation. *Applied Mathematics and Computation* **180**, 498–515.

[270] Lakatos E, Ale A, Kirk PDW, Stumpf MPH. 2015 Multivariate moment closure techniques for stochastic kinetic models. *The Journal of Chemical Physics* **143**, 094107.

[271] Kuehn C. 2016 Moment closure—a brief review. In *Control of Self-Organizing Nonlinear Systems* pp. 253–271.

[272] Fan S, Geissmann Q, Lakatos E, Lukauskas S, Ale A, Babtie AC, Kirk PDW, Stumpf MPH. 2016 MEANS: python package for Moment Expansion Approximation, iNference and Simulation. *Bioinformatics* **32**, 2863–2865.

[273] Brouwer AF, Meza R, Eisenberg MC. 2017 A systematic approach to determining the identifiability of multistage carcinogenesis models. *Risk Analysis* **37**, 1375–1387.

[274] Ligon TS, Fröhlich F, Chiş OT, Banga JR, Balsa-Canto E, Hasenauer J. 2017 GenSSI 2.0: multi-experiment structural identifiability analysis of SBML models. *Bioinformatics* **34**, 1421–1423.

[275] Bezanson J, Edelman A, Karpinski S, Shah VB. 2017 Julia: a fresh approach to numerical computing. *SIAM Review* **59**, 65–98.

[276] Gillespie DT. 2001 Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics* **115**, 1716–1733.

[277] Schnoerr D, Sanguinetti G, Grima R. 2014 The complex chemical Langevin equation. *The Journal of Chemical Physics* **141**, 024103.

[278] Higham DJ. 2008 Modeling and simulating chemical reactions. *SIAM Review* **50**, 347–368.

[279] Gillespie DT. 1977 Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry* **81**, 2340–2361.

[280] Kurtz TG. 1972 The relationship between stochastic and deterministic models for chemical reactions. *The Journal of Chemical Physics* **57**, 2976–2978.

[281] Gibson MA, Bruck J. 2000 Efficient exact stochastic simulation of chemical systems with many species and many channels. *The Journal of Physical Chemistry A* **104**, 1876–1889.

[282] Rao CV, Wolf DM, Arkin AP. 2002 Control, exploitation and tolerance of intracellular noise. *Nature* **420**, 231–237.

[283] Samad HE, Khammash M, Petzold L, Gillespie D. 2005 Stochastic modelling of gene regulatory networks. *International Journal of Robust and Nonlinear Control* **15**, 691–711.

[284] Golightly A, Wilkinson DJ. 2005 Bayesian inference for stochastic kinetic models using a diffusion approximation. *Biometrics* **61**, 781–788.

[285] Maruyama G. 1955 Continuous Markov processes and stochastic equations. *Rendiconti del Circolo Matematico di Palermo* **4**, 48.

[286] Socha L. 2008 *Linearization Methods for Stochastic Dynamic Systems*. Lecture Notes in Physics. Berlin Heidelberg: Springer-Verlag.

[287] Hausken K, Moxnes JF. 2010 A closure approximation technique for epidemic models. *Mathematical and Computer Modelling of Dynamical Systems* **16**, 555–574.

[288] Isserlis L. 1918 On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika* **12**, 134.

[289] Singh A, Hespanha JP. 2007 A derivative matching approach to moment closure for the stochastic logistic model. *Bulletin of Mathematical Biology* **69**, 1909–1925.

[290] Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. 2014 *Bayesian Data Analysis*. CRC Press 3 edition.

[291] Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. 1953 Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* **21**, 1087–1092.

[292] Hastings WK. 1970 Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.

[293] Geyer CJ. 1992 Practical Markov chain Monte Carlo. *Statistical Science* **7**, 473–483.

[294] Roberts GO, Rosenthal JS. 2001 Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science* **16**, 351–367.

[295] Gelman A, Rubin DB. 1992 Inference from iterative simulation using multiple sequences. *Statistical Science* **7**, 457–472.

[296] Brooks SP, Gelman A. 1998 General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* **7**, 434.

[297] Johnston ST, Shah ET, Chopin LK, McElwain DLS, Simpson MJ. 2015 Estimating cell diffusivity and cell proliferation rate by interpreting IncuCyte ZOOM^{TM} assay data using the Fisher-Kolmogorov model. *BMC Systems Biology* **9**, 38.

[298] Matsiaka OM, Baker RE, Shah ET, Simpson MJ. 2019 Mechanistic and experimental models of cell migration reveal the importance of cell-to-cell pushing in cell invasion. *Biomedical Physics & Engineering Express* **5**, 045009.

[299] Poovathingal SK, Gunawan R. 2010 Global parameter estimation methods for stochastic biochemical systems. *BMC Bioinformatics* **11**, 414–414.

[300] Dargatz C. 2008 A diffusion approximation for an epidemic model. *Sonderforschungsbereich* **386**, 517.

[301] Warne DJ, Ebert A, Drovandi C, Mira A, Mengersen K. 2020 Hindsight is 2020 vision: Characterisation of the global response to the COVID-19 pandemic. *BMC Public Health* **20**, 1868.

[302] Rackauckas C, Nie Q. 2016 DifferentialEquations.jl – A performant and feature-rich ecosystem for solving differential equations in Julia. *Journal of Open Research Software* **5**.

[303] Burchard H, Deleersnijder E, Meister A. 2003 A high-order conservative Patankar-type discretisation for stiff systems of production–destruction equations. *Applied Numerical Mathematics* **47**, 1–30.

[304] Villaverde AF. 2019 Observability and structural identifiability of nonlinear biological systems. *Complexity* **2019**, 1–12.

[305] Brown KS, Sethna JP. 2003 Statistical mechanical approaches to models with many poorly known parameters. *Physical Review E* **68**, 021904.

[306] Erguler K, Stumpf MPH. 2011 Practical limits for reverse engineering of dynamical systems: a statistical analysis of sensitivity and parameter inferability in systems biology models. *Molecular BioSystems* **7**, 1593–1602.

[307] Transtrum MK, Machta BB, Brown KS, Daniels BC, Myers CR, Sethna JP. 2015 Perspective: Sloppiness and emergent theories in physics, biology, and beyond. *The Journal of Chemical Physics* **143**, 010901.

[308] Dufresne E, Harrington HA, Raman DV. 2018 The geometry of sloppiness. *Journal of Algebraic Statistics* **9**, 30–68.

[309] Smadbeck P, Kaznessis YN. 2013 A closure scheme for chemical master equations. *Proceedings of the National Academy of Sciences* **110**, 14261–14265.

[310] Zechner C, Ruess J, Krenn P, Pelet S, Peter M, Lygeros J, Koeppl H. 2012 Moment-based inference predicts bimodality in transient gene expression. *Proceedings of the National Academy of Sciences* **109**, 8340–8345.

[311] Roberts GO, Rosenthal JS. 2009 Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics* **18**, 349–367.

[312] Moral PD, Doucet A, Jasra A. 2006 Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B* **68**, 411–436.

[313] Giles MB. 2008 Multilevel Monte Carlo path simulation. *Operations Research* **56**, 607–617.

[314] Jasra A, Kamatani K, Law KJH, Zhou Y. 2017 Multilevel particle filters. *SIAM Journal on Numerical Analysis* **55**, 3068–3096.

[315] Warne DJ, Baker RE, Simpson MJ. 2018 Multilevel rejection sampling for approximate Bayesian computation. *Computational Statistics & Data Analysis* **124**, 71–86.

[316] Quiroz M, Kohn R, Villani M, Tran MN. 2019 Speeding up MCMC by efficient data subsampling. *Journal of the American Statistical Association* **114**, 831–843.

[317] Pooley CM, Bishop SC, Marion G. 2015 Using model-based proposals for fast parameter inference on discrete state space, continuous-time Markov processes. *Journal of The Royal Society Interface* **12**, 20150225.

[318] Burrage K, Burrage P. 1996 High strong order explicit Runge-Kutta methods for stochastic ordinary differential equations. *Applied Numerical Mathematics* **22**, 81–101.

[319] Mingas G, Bottolo L, Bouganis CS. 2017 Particle MCMC algorithms and architectures for accelerating inference in state-space models. *International Journal of Approximate Reasoning* **83**, 413–433.

Bibliography

[320] Lee YS, Liu OZ, Hwang HS, Knollmann BC, Sobie EA. 2013 Parameter sensitivity analysis of stochastic models provides insights into cardiac calcium sparks. *Biophysical Journal* **104**, 1142–1150.

[321] Botha I, Kohn R, Drovandi C. 2020 Particle methods for stochastic differential equation mixed effects models. *Bayesian Analysis.*

[322] Picchini U. 2012 Inference for SDE models via approximate Bayesian computation. *Journal of Computational and Graphical Statistics* **23**, 1080–1100.

[323] Buckwar E, Tamborrino M, Tubikanec I. 2020 Spectral density-based and measure-preserving ABC for partially observed diffusion processes. An illustration on Hamiltonian SDEs. *Statistics and Computing* **30**, 627–648.

[324] Liepe J, Filippi S, Komorowski M, Stumpf MPH. 2013 Maximizing the information content of experiments in systems biology. *PLOS Computational Biology* **9**, e1002888.

[325] Evans ND, White LJ, Chapman MJ, Godfrey KR, Chappell MJ. 2005 The structural identifiability of the susceptible infected recovered model with seasonal forcing. *Mathematical Biosciences* **194**, 175–197.

[326] Chapman JD, Evans ND. 2008 The structural identifiability of SIR type epidemic models with incomplete immunity and birth targeted vaccination. *IFAC Proceedings Volumes* **41**, 9075–9080.

[327] Beskos A, Papaspiliopoulos O, Roberts GO. 2006 Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli* **12**, 1077–1098.

[328] Fröhlich F, Thomas P, Kazeroonian A, Theis FJ, Grima R, Hasenauer J. 2016 Inference for stochastic chemical kinetics using moment equations and system size expansion. *PLOS Computational Biology* **12**, e1005030.

[329] Elf J, Ehrenberg M. 2003 Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. *Genome Research* **13**, 2475–2484.

[330] Fearnhead P, Giagos V, Sherlock C. 2014 Inference for reaction networks using the linear noise approximation. *Biometrics* **70**, 457–466.

[331] Plotnikov A, Zehorai E, Procaccia S, Seger R. 2011 The MAPK cascades: Signaling components, nuclear roles and mechanisms of nuclear translocation. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1813**, 1619–1633.

[332] Xiu D, Karniadakis GE. 2002 The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing* **24**, 619–644.

[333] Archambeau C, Cornford D, Opper M, Shawe-Taylor JS. 2007 Gaussian process approximations of stochastic differential equations. *Proceedings of Machine Learning Research* **1**, 1–16.

[334] Mirams GR, Pathmanathan P, Gray RA, Challenor P, Clayton RH. 2016 Uncertainty and variability in computational and mathematical models of cardiac physiology. *The Journal of Physiology* **594**, 6833–6847.

[335] Kaintura A, Dhaene T, Spina D. 2018 Review of polynomial chaos-based methods for uncertainty quantification in modern integrated circuits. *Electronics* **7**, 30.

[336] Ran ZY, Hu BG. 2017 Parameter identifiability in statistical machine learning: a review. *Neural Computation* **29**, 1151–1203.

[337] Lill D, Timmer J, Kaschek D. 2019 Local Riemannian geometry of model manifolds and its implications for practical parameter identifiability. *PLOS One* **14**, e0217837.

[338] Livingstone S, Girolami M. 2014 Information-geometric Markov chain Monte Carlo methods using diffusions. *Entropy* **16**, 3074–3102.

[339] Archambeau C, Opper M, Shen Y, Cornford D, Shawe-Taylor JS. 2008 Variational inference for diffusion processes. In Platt JC, Koller D, Singer Y, Roweis ST, editors, *Advances in Neural Information Processing Systems* vol. 20 pp. 17–24. Curran Associates, Inc.

[340] Raue A, Kreutz C, Theis FJ, Timmer J. 2013 Joining forces of Bayesian and frequentist methodology: a study for inference in the presence of non-identifiability. *Philosophical Transactions of the Royal Society A* **371**, 20110544.

[341] Faller D, Klingmüller U, Timmer J. 2003 Simulation methods for optimal experimental design in systems biology. *SIMULATION* **79**, 717–725.

[342] Walter E, Lecourtier Y. 1981 Unidentifiable compartmental models: what to do? *Mathematical Biosciences* **56**, 1–25.

[343] Paulose J, Meeussen AS, Vitelli V. 2015 Selective buckling via states of self-stress in topological metamaterials. *Proceedings of the National Academy of Sciences* **112**, 7639–7644.

[344] Hwang M, Arrieta AF. 2018 Solitary waves in bistable lattices with stiffness grading: Augmenting propagation control. *Physical Review E* **98**, 042205.

[345] Nadkarni N, Daraio C, Abeyaratne R, Kochmann DM. 2016 Universal energy transport law for dissipative and diffusive phase transitions. *Physical Review B* **93**, 1340–1347.

[346] Deng B, Wang P, He Q, Tournat V, Bertoldi K. 2018 Metamaterials with amplitude gaps for elastic solitons. *Nature Communications* **9**, 3410.

[347] Kochmann DM, Bertoldi K. 2017 Exploiting microstructural instabilities in solids and structures: from metamaterials to structural transitions. *Applied Mechanics Reviews* **69**, 050801.

[348] Dudek KK, Gatt R, Dudek MR, Grima JN. 2018 Negative and positive stiffness in auxetic magneto-mechanical metamaterials. *Proceedings of the Royal Society A* **474**, 20180003.

[349] Serra-Garcia M, Molerón M, Daraio C. 2018 Tunable, synchronized frequency down-conversion in magnetic lattices with defects. *Philosophical Transactions of the Royal Society A* **376**, 20170137.

[350] Silverberg JL, Evans AA, McLeod L, Hayward RC, Hull T, Santangelo CD, Cohen I. 2014 Using origami design principles to fold reprogrammable mechanical metamaterials. *Science* **345**, 647–650.

[351] Paulose J, Chen B, Vitelli V. 2015 Topological modes bound to dislocations in mechanical metamaterials. *Nature Physics* **11**, 153–156.

[352] Turco E, Giorgio I, Misra A, dell'Isola F. 2017 King post truss as a motif for internal structure of (meta)material with controlled elastic properties. *Royal Society Open Science* **4**, 171153.

[353] Matlack KH, Bauhofer A, Krödel S, Palermo A, Daraio C. 2016 Composite 3D-printed metastructures for low-frequency and broadband vibration absorption. *Proceedings of the National Academy of Sciences* **113**, 8386–8390.

[354] Chen T, Bilal OR, Shea K, Daraio C. 2018 Harnessing bistability for directional propulsion of soft, untethered robots. *Proceedings of the National Academy of Sciences* **115**, 5698–5702.

[355] Chen T, Mueller J, Shea K. 2017 Integrated design and simulation of tunable, multi-state structures fabricated monolithically with multi-material 3D printing. *Scientific Reports* **7**, 45671.

[356] Baardink G, Souslov A, Paulos J, Vitelli V. 2018 Localizing softness and stress along loops in 3D topological metamaterials. *Proceedings of the National Academy of Sciences* **115**, 489–494.

[357] Ding Z, Yuan C, Peng X, Wang T, Qi HJ, Dunn ML. 2017 Direct 4D printing via active composite materials. *Science Advances* **3**, e1602890.

[358] Wehner M, Tolley MT, Mengüç Y, Park YL, Mozeika A, Ding Y, Onal C, Shepherd RF, Whitesides GM, Wood RJ. 2014 Pneumatic energy sources for autonomous and wearable soft robotics. *Soft Robotics* **1**, 263–274.

[359] Wiktor V, Jonkers HM. 2011 Quantification of crack-healing in novel bacteria-based self-healing concrete. *Cement and Concrete Composites* **33**, 763–770.

[360] Miniaci M, Krushynska A, Movchan AB, Bosia F, Pugno NM. 2016 Spider web-inspired acoustic metamaterials. *Applied Physics Letters* **109**, 071905.

[361] Jiang Y, Korpas LM, Raney JR. 2019 Bifurcation-based embodied logic and autonomous actuation. *Nature Communications* **10**, 128.

[362] Woodhouse FG, Ronellenfitsch H, Dunkel J. 2018 Autonomous actuation of zero modes in mechanical networks far from equilibrium. *Physical Review Letters* **121**, 178001.

[363] Pollard TD, Cooper JA. 1986 Actin and actin-binding proteins. A critical evaluation of mechanisms and functions. *Annual Review of Biochemistry* **55**, 987–1035.

[364] Blanchoin L, Boujemaa-Paterski R, Sykes C, Plastino J. 2014 Actin dynamics, architecture, and mechanics in cell motility. *Physiological Reviews* **94**, 235–263.

[365] Kumar S, Mansson A. 2017 Covalent and non-covalent chemical engineering of actin for biotechnological applications. *Biotechnology Advances* **35**, 867–888.

Bibliography

[366] Nicolau DV Jr, Lard M, Korten T, van Delft FCMJM, Persson M, Bengtsson E, Mansson A, Diez S, Linke H, Nicolau DV. 2016 Parallel computation with molecular-motor-propelled agents in nanofabricated networks. *Proceedings of the National Academy of Sciences* **113**, 2591–2596.

[367] FitzHugh R. 1955 Mathematical models of threshold phenomena in the nerve membrane. *Bulletin of Mathematical Biophysics* **17**, 257–278.

[368] Rubin J, Wechselberger M. 2008 The selection of mixed-mode oscillations in a Hodgkin-Huxley model with multiple timescales. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **18**, 015105.

[369] Beck M, Jones CKRT, Schaeffer D, Wechselberger M. 2008 Electrical waves in a one-dimensional model of cardiac tissue. *SIAM Journal on Applied Dynamical Systems* **7**, 1558–1581.

[370] Baker RE, Parker A, Simpson MJ. 2019 A free boundary model of epithelial dynamics. *Journal of Theoretical Biology* **481**, 61–74.

[371] Billingham J, Needham DJ. 1991 The development of travelling waves in quadratic and cubic autocatalysis with unequal diffusion rates. I. Permanent form travelling waves. *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences* **334**, 1–24.

[372] Keener J, Sneyd J. 2009 *Mathematical Physiology.* New York: Springer.

[373] Landman KA, Simpson MJ, Slater JL, Newgreen DF. 2005 Diffusive and chemotactic cellular migration: smooth and discontinuous traveling wave solutions. *SIAM Journal on Applied Mathematics* **65**, 1420–1442.

[374] Simpson MJ, Landman KA, Hughes BD, Newgreen DF. 2006 Looking inside an invasion wave of cells using continuum models: Proliferation is the key. *Journal of Theoretical Biology* **243**, 343 360.

[375] Nadkarni N, Arrieta AF, Chong C, Kochmann DM, Daraio C. 2016 Unidirectional transition waves in bistable lattices. *Physical Review Letters* **116**, 244501.

[376] Nizovtseva IG, Galenko PK. 2018 Travelling-wave amplitudes as solutions of the phase-field crystal equation. *Philosophical Transactions of the Royal Society A* **376**, 20170202.

[377] Hinch EJ. 1991 *Perturbation Methods.* Cambridge University Press.

[378] Snita D, Sevcikova H, Marek M, Merkin JH. 1997 Travelling waves in an ionic autocatalytic chemical system with an imposed electric field. *Proceedings of the Royal Society A* **453**, 2325–2351.

[379] Simpson MJ, Landman KA, Bhaganagarapu K. 2007 Coalescence of interacting cell populations. *Journal of Theoretical Biology* **247**, 525–543.

[380] Roseti L, Parisi V, Petretta M, Cavallo C, Desando G, Bartolotti I, Grigolo B. 2017 Scaffolds for bone tissue engineering: state of the art and new perspectives. *Materials Science and Engineering: C* **78**, 1246–1262.

[381] Groll J, Boland T, Blunk T, Burdick JA, Cho DW, Dalton PD, Derby B, Forgacs G, Li Q, Mironov VA, Moroni L, Nakamura M, Shu W, Takeuchi S, Vozzi G, Woodfield TBF, Xu T, Yoo JJ, Malda J. 2016 Biofabrication: reappraising the definition of an evolving field. *Biofabrication* **8**, 013001.

[382] Forrestal DP, Klein TJ, Woodruff MA. 2017 Challenges in engineering large customized bone constructs. *Biotechnology and Bioengineering* **114**, 1129–1139.

[383] Paxton NC, Ren J, Ainsworth MJ, Solanki AK, Jones JR, Allenby MC, Stevens MM, Woodruff MA. 2019a Rheological characterization of biomaterials directs additive manufacturing of strontium-substituted bioactive glass/polycaprolactone microfibers. *Macromolecular Rapid Communications* **40**, 1900019.

[384] Paxton NC, Allenby MC, Lewis PM, Woodruff MA. 2019b Biomedical applications of polyethylene. *European Polymer Journal* **118**, 412–428.

[385] Bidan CM, Kommareddy KP, Rumpler M, Kollmannsberger P, Bréchet YJM, Fratzl P, Dunlop JWC. 2012 How linear tension converts to curvature: geometric control of bone tissue growth. *PLOS One* **7**, e36336.

[386] Bidan CM, Kommareddy KP, Rumpler M, Kollmannsberger P, Fratzl P, Dunlop JWC. 2013 Geometry as a factor for tissue growth: towards shape optimization of tissue engineering scaffolds. *Advanced Healthcare Materials* **2**, 186–194.

[387] Alias MA, Buenzli PR. 2018 Osteoblasts infill irregular pores under curvature and porosity controls: a hypothesis-testing analysis of cell behaviours. *Biomechanics and Modeling in Mechanobiology* **17**, 1357–1371.

212

[388] Callens SJ, Uyttendaele RJ, Fratila-Apachitei LE, Zadpoor AA. 2020 Substrate curvature as a cue to guide spatiotemporal cell and tissue organization. *Biomaterials* **232**, 119739.

[389] Simpson MJ. 2009 Depth-averaging errors in reactive transport modeling. *Water Resources Research* **45**.

[390] Treloar KK, Simpson MJ, McElwain DLS, Baker RE. 2014 Are *in vitro* estimates of cell diffusivity and cell proliferation rate sensitive to assay geometry? *Journal of Theoretical Biology* **356**, 71–84.

[391] Bidan CM, Wang FM, Dunlop JW. 2013 A three-dimensional model for tissue deposition on complex surfaces. *Computer Methods in Biomechanics and Biomedical Engineering* **16**, 1–15.

[392] Gamsjäger E, Bidan C, Fischer F, Fratzl P, Dunlop J. 2013 Modelling the role of surface stress on the kinetics of tissue growth in confined geometries. *Acta Biomaterialia* **9**, 5531–5543.

[393] Salbreux G, Jülicher F. 2017 Mechanics of active surfaces. *Physical Review E* **96**, 032404.

[394] Rumpler M, Woesz A, Dunlop JW, van Dongen JT, Fratzl P. 2008 The effect of geometry on three-dimensional tissue growth. *Journal of the Royal Society Interface* **5**, 1173–1180.

[395] Guyot Y, Papantoniou I, Chai YC, Bael SV, Schrooten J, Geris L. 2014 A computational model for cell/ECM growth on 3D surfaces using the level set method: a bone tissue engineering case study. *Biomechanics and Modeling in Mechanobiology* **13**, 1361–1371.

[396] Ehrig S, Schamberger B, Bidan CM, West A, Jacobi C, Lam K, Kollmannsberger P, Petersen A, Tomancak P, Kommareddy K, Fischer FD, Fratzl P, Dunlop JWC. 2019 Surface tension determines tissue shape and growth kinetics. *Science Advances* **5**, eaav9394.

[397] Lee HG, Park J, Yoon S, Lee C, Kim J. 2019 Mathematical model and numerical simulation for tissue growth on bioscaffolds. *Applied Sciences* **9**, 4058.

[398] Alias MA, Buenzli PR. 2020 A level-set method for the evolution of cells and tissue during curvature-controlled growth. *International Journal for Numerical Methods in Biomedical Engineering* **36**, e3279.

[399] Campbell DA, Chkrebtii O. 2013 Maximum profile likelihood estimation of differential equation parameters through model based smoothing state estimates. *Mathematical Biosciences* **246**, 283–292.

[400] Wieland FG, Hauber AL, Rosenblatt M, Tönsing C, Timmer J. 2021 On structural and practical identifiability. *Current Opinion in Systems Biology* **25**, 60–69.

[401] Boiger R, Hasenauer J, Hroß S, Kaltenbacher B. 2016 Integration based profile likelihood calculation for PDE constrained parameter estimation problems. *Inverse Problems* **32**, 125009.

[402] Simpson MJ, Treloar KK, Binder BJ, Haridas P, Manton KJ, Leavesley DI, McElwain DLS, Baker RE. 2013 Quantifying the roles of cell motility and cell proliferation in a circular barrier assay. *Journal of The Royal Society Interface* **10**, 20130007.

[403] Jin W, Lo KY, Chou S, McCue SW, Simpson MJ. 2018 The role of initial geometry in experimental models of wound closing. *Chemical Engineering Science* **179**, 221–226.

[404] Lagergren JH, Nardini JT, Baker RE, Simpson MJ, Flores KB. 2020 Biologically-informed neural networks guide mechanistic modeling from sparse experimental data. *PLOS Computational Biology* **16**, e1008462.

[405] Geris L, Lambrechts T, Carlier A, Papantoniou I. 2018 The future is digital: *in silico* tissue engineering. *Current Opinion in Biomedical Engineering* **6**, 92–98.

[406] Treloar KK, Simpson MJ. 2013 Sensitivity of edge detection methods for quantifying cell migration assays. *PLOS One* **8**, e67389.

[407] Baldock AL, Ahn S, Rockne R, Johnston S, Neal M, Corwin D, Clark-Swanson K, Sterin G, Trister AD, Malone H, Ebiana V, Sonabend AM, Mrugala M, Rockhill JK, Silbergeld DL, Lai A, Cloughesy T, McKhann GM, Bruce JN, Rostomily RC, Canoll P, Swanson KR. 2014 Patient-specific metrics of invasiveness reveal significant prognostic benefit of resection in a predictable subset of gliomas. *PLOS One* **9**, e99057.

[408] Gonçalves IG, Garcia-Aznar JM. 2021 Extracellular matrix density regulates the formation of tumour spheroids through cell migration. *PLOS Computational Biology* **17**, e1008764.

[409] Yan XZ, Yang W, Yang F, Kersten-Niessen M, Jansen JA, Both SK. 2014 Effects of continuous passaging on mineralization of MC3T3-E1 cells with improved osteogenic culture protocol. *Tissue Engineering Part C: Methods* **20**, 198–204.

[410] Mathworks. 2021 Image Processing Toolbox: Regionprops. .

[411] Carroll RJ, Ruppert D. 1988 *Transformation and Weighting in Regression*. Boca Raton: Chapman and Hall 1st edition.

Bibliography

[412] Collis J, Connor AJ, Paczkowski M, Kannan P, Pitt-Francis J, Byrne HM, Hubbard ME. 2017 Bayesian calibration, validation and uncertainty quantification for predictive modelling of tumour growth: a tutorial. *Bulletin of Mathematical Biology* **79**, 939–974.

[413] Pawitan Y. 2013 *In all likelihood: statistical modelling and inference using likelihood.* Oxford: Oxford University Press.

[414] Johnson SG. 2021 The NLopt module for Julia. `https://github.com/JuliaOpt/NLopt.jl`.

[415] Jones DR, Perttunen CD, Stuckman BE. 1993 Lipschitzian optimization without the Lipschitz constant. *Journal of Optimization Theory and Applications* **79**, 157–181.

[416] Powell MJD. 2009 The BOBYQA algorithm for bound constrained optimization without derivatives. Technical report Department of Applied Mathematics and Theoretical Physics Cambridge, England.

[417] Royston P. 2007 Profile likelihood for estimation and confidence intervals. *The Stata Journal* **7**, 376–387.

[418] Venzon DJ, Moolgavkar SH. 1988 A method for computing profile-likelihood-based confidence intervals. *Applied Statistics* **37**, 87.

[419] Gibon E, Batke B, Jawad MU, Fritton K, Rao A, Yao Z, Biswal S, Gambhir SS, Goodman SB. 2012 MC3T3-E1 osteoprogenitor cells systemically migrate to a bone defect and enhance bone healing. *Tissue Engineering Part A* **18**, 968–973.

[420] Tarle V, Gauquelin E, Vedula SRK, D'Alessandro J, Lim CT, Ladoux B, Gov NS. 2017 Modeling collective cell migration in geometric confinement. *Physical Biology* **14**, 035001.

[421] Hegarty-Cremer SGD, Simpson MJ, Andersen TL, Buenzli PR. 2021 Modelling cell guidance and curvature control in evolving biological tissues. *Journal of Theoretical Biology* **520**, 110658.

[422] Renardy M, Kirschner D, Eisenberg M. 2021 Structural identifiability analysis of PDEs: A case study in continuous age-structured epidemic models. *arXiv*.

[423] Tsiantis N, Balsa-Canto E, Banga JR. 2018 Optimality and identification of dynamic models in systems biology: an inverse optimal control framework. *Bioinformatics* **34**, 2433–2440.

[424] Dalwadi MP, Pearce P. 2021 Emergent robustness of bacterial quorum sensing in fluid flow. *Proceedings of the National Academy of Sciences* **118**, e2022312118.

[425] Koshy-Chenthittayil S, Archambault L, Senthilkumar D, Laubenbacher R, Mendes P, Dongari-Bagtzoglou A. 2021 Agent based models of polymicrobial biofilms and the microbiome—a review. *Microorganisms* **9**, 417.

[426] Blankenship G. 1986 Stochastic optimal control of systems with fast and slow states. *American Control Conference* **1986**, 864–870.

[427] Hansen E, Karslake J, Woods RJ, Read AF, Wood KB. 2020 Antibiotics can be used to contain drug-resistant bacteria by maintaining sufficiently large sensitive populations. *PLOS Biology* **18**, e3000713.

[428] Treloar NJ, Fedorec AJH, Ingalls B, Barnes CP. 2020 Deep reinforcement learning for the control of microbial co-cultures in bioreactors. *PLOS Computational Biology* **16**, e1007783.

[429] Turing AM. 1952 The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **237**, 37–72.

[430] Mathworks interpn: Interpolation for 1-D, 2-D, 3-D, and N-D gridded data in ndgrid format. `https://au.mathworks.com/help/matlab/ref/interpn.html`. Accessed: June, 2020.

[431] Kloeden PE, Platen E. 1992 *Numerical Solution of Stochastic Differential Equations.* Berlin: Springer-Verlag.

[432] Eberhard Z. 1996 *Oxford users' guide to mathematics* vol. 1. Oxford: Oxford University Press.

[433] Campbell SL, Meyer CD. 1991 *Generalized inverses of linear transformations.* New York: Dover.

[434] Mathworks. 2019 Moore-Penrose pseudoinverse. `https://au.mathworks.com/help/matlab/ref/pinv.html`. Accessed: 6/3/19.