

Developing and applying modelling and Bayesian inference tools for developmental biology



Jonathan U. Harrison
St Hugh's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Michaelmas 2018

Abstract

Developmental biology allows us to answer crucial questions about how patterned, polarized cells can organize robustly and repeatably to form living tissues and organisms. Quantitative imaging data is now available that can provide deep insights into developmental processes, but to make best use of these data, we need mathematical and computational models. Linking complex mechanistic models to quantitative experimental data is challenging, but Bayesian inference provides a principled framework for this. Together, mechanistic models can guide hypothesis-driven experiments, while quantitative experimental data can inform modelling through a data-driven framework for inference and testable predictions. However, challenges remain in developing methodological tools to relate mechanistic models to data effectively. We address a number of these challenges in this thesis, and demonstrate the use of such tools in an integrative framework to address hypotheses about the robustness and points of control governing mRNA localization during development. We begin by studying the bottlenecks controlling the regulation of mRNA transport, using a fine-grained stochastic model of transport within a single cell and relate this model to experimental data. We then study this stochastic model in a more general setting and ask how experimental design, such as the sampling frequency of observations via imaging, affects parameter estimation. Summary statistics are frequently used to simplify inference for problems with high-dimensional data. It is not clear how to select and combine summary statistics for approximate Bayesian computation, although much previous work has been done in this area. We address this problem with an adaptive algorithm that aims to maximize the distance between prior and approximate posterior distributions. Finally, we return to our study of mRNA localization by developing a coarse-grained model to address questions about how robustness of mRNA localization is regulated during development. Using a simple differential equation model, we make testable predictions, which demonstrate the surprising robustness of mRNA localization, and subsequently reveal crowding as a mechanism for regulation of this process. Together, the contributions of this work are towards developing modelling and inference tools to enable quantitative hypothesis-driven study of developmental and cell biology.

Acknowledgements

The work in this thesis would not have been possible without the support and guidance of my fantastic supervisor, Prof. Ruth Baker. I am extremely grateful also for the collaboration with Prof. Ilan Davis and my co-supervisor, Dr. Richard Parton. Amongst other things, thank you for showing me how the data is collected rather than it simply being numbers on a screen, and for letting me drive a microscope!

I would like to acknowledge funding from the Engineering and Physical Sciences Research Council via the Systems Biology DTC, and from the Society for Mathematical Biology, Newton Institute, and MBI, Ohio for travel funding to attend various workshops and conferences

Thanks to all the members of the Wolfson Centre for Mathematical Biology, for the cake, the puns (Josh), and getting to hear about the amazing research going on in the rest of the group.

A big shout out has to go to Dr. Dan Humphries and (soon-to-be-Dr.) Eddie Rolls, with whom I productively shared an office throughout my DPhil. Final standings on the predictions are: J 87, D 77, E 16. Still possible for Eddie to catch up, but we might need to share an office again for a while longer.

The following groups have enriched my time in Oxford enormously: St Hugh's College students, tutors, and staff; SysBio DTC; Jesus College Boat Club; and the Cowley Road Condors. I am grateful to all my friends and family that have been there through all the highs and lows. A DPhil is a marathon not a sprint; although perhaps a marathon would have been over much sooner! Nonetheless I am eternally grateful to all those who have helped me along the way.

Finally, I'd like to thank my long-suffering wife, who has helped pick me up whenever something has knocked me down. Mimi, thank you. We started on the DPhil road together and you may have pipped me to the post, but I suppose that's ok!

Statement of Originality

This thesis is submitted to the University of Oxford in partial fulfillment of the requirements of the degree of Doctor of Philosophy. This thesis contains no material which has been previously submitted for a degree or diploma at this University or any other institution. All experimental data presented in this thesis were obtained by Dr. Richard M. Parton whilst working in the laboratory of Prof. Ilan Davis at the Department of Biochemistry, University of Oxford. Otherwise, this thesis represents my own original work towards this research degree.

The research in Chapter 4 has been published in the journal *PLoS Computational Biology* as Harrison and Baker [1]. The work in Chapters 5 and 6 is available as pre-print articles Harrison and Baker [2] and Harrison et al. [3], respectively.

Jonathan U. Harrison

University of Oxford

December 2018

Contents

1	Introduction	1
1.1	Grand challenges	4
1.1.1	Statistical challenges: summary statistics	4
1.1.2	Experimental design informed by computational modelling . .	5
1.1.3	Biological challenges: mRNA localization	6
1.2	Aims and outline	7
2	Bayesian and approximate Bayesian inference	9
2.1	Bayesian inference	10
2.1.1	Bayes' theorem	10
2.1.2	Expectation and Monte Carlo	11
2.1.3	Importance sampling	13
2.1.4	Markov chain Monte Carlo	14
2.1.5	Sequential Monte Carlo	19
2.1.6	Fitting differential equation models	21
2.1.7	Identifiability	22
2.2	Likelihood-free inference	23
2.2.1	ABC	24
2.2.2	Summary statistics	24
2.2.3	Improvements to ABC	26
2.2.4	ABC-SMC	28
2.2.5	Lazy and multifidelity approaches to ABC	31

2.2.6	Other likelihood-free methods	33
2.2.7	Particle MCMC	34
2.3	Model comparison	35
2.3.1	Information criteria	36
2.3.2	Model evidence and Bayes Factors	38
2.3.3	Alternatives	39
2.3.4	Prediction	40
2.3.5	Cross-validation	41
2.3.6	Pseudo Bayesian model averaging	41
2.3.7	Bayesian model averaging	42
2.3.8	Stacking	43
2.3.9	So, can I fit an elephant?	43

3 An interdisciplinary approach to elucidate the regulation of mRNA

	transport in <i>Drosophila</i> oogenesis	45
3.1	Regulation of RNA controls expression of protein in space and time .	46
3.2	<i>Drosophila</i> development	47
3.2.1	<i>Drosophila</i> oogenesis	49
3.2.2	Dynamic cytoskeleton forms part of the subcellular environment	50
3.2.3	mRNA localization in <i>Drosophila</i>	52
3.2.4	Advanced imaging tools shed light on mRNA localization . . .	55
3.3	The role of modelling	55
3.3.1	Previous modelling approaches to <i>Drosophila</i> development . .	56
3.3.2	Summary	58
3.4	Fine-grained stochastic model of mRNA transport within a nurse cell	58
3.4.1	Master equation	61
3.4.2	Geometry	62
3.4.3	Reorientation kernels based on cytoskeleton distribution . . .	63
3.4.4	Boundary conditions	65

3.5	Analysis of the microscopic model	65
3.5.1	Parameter estimates from direct measurements	65
3.5.2	Simulation of first passage times	66
3.5.3	Sensitivity analysis highlights the role of the cytoskeleton . . .	68
3.5.4	Spatial distribution as a summary statistic	72
3.6	Bayesian inference methods to link the microscopic model to data . .	73
3.6.1	Implementation of ABC	75
3.6.2	Inference on synthetic data	76
3.6.3	Combined summary statistics	79
3.6.4	Using ABC to fit to real data	81
3.7	Testable prediction of FPT for a Dynein mutant	83
3.8	Discussion	85
4	The impact of temporal sampling resolution on parameter inference for biological transport models	89
4.1	Introduction	91
4.1.1	Velocity jump process models	92
4.1.2	Inference for velocity jump process models via particle MCMC	95
4.2	Methods	96
4.2.1	Model without measurement noise	99
4.2.2	Model with measurement noise	102
4.2.3	Derivation of emission probabilities	103
4.2.4	Comparison of expression for emission probabilities with simulations	106
4.2.5	Particle MCMC algorithm	108
4.2.6	Implementation of MCMC methods	108
4.2.7	Approximate Bayesian computation	110
4.3	Results	112
4.3.1	Increasing Δt results in an abrupt breakdown in the posterior	113

4.3.2	Increasing σ increases the variance of the posterior for λ . . .	113
4.3.3	Probability of hidden state sequences	116
4.3.4	More data provides better estimates	116
4.3.5	Experimental constraints	117
4.3.6	Comparison to ABC	118
4.3.7	The computational cost of pMCMC and ABC per effective sample size.	122
4.3.8	Model misspecification	123
4.3.9	Application to RNA transport dataset	127
4.4	Discussion	131
5	Adaptive weighting of summary statistics for ABC	135
5.1	Role of summary statistics in ABC	137
5.1.1	Previous literature on selecting summary statistics in ABC . .	138
5.1.2	Outline	140
5.2	An algorithm for automatic weighting of summary statistics in ABC .	140
5.2.1	Adaption of weights	141
5.2.2	Distance between distributions	142
5.3	Theoretical justification	145
5.3.1	Convergence of the nearest neighbour estimator	145
5.3.2	Convergence of adaptive ABC-SMC	147
5.4	Examples	148
5.4.1	Uniform toy model	149
5.4.2	Bimodal toy model	149
5.4.3	Gaussian toy model	151
5.4.4	Death process	153
5.4.5	Dimerization system	155
5.4.6	Simple spatial model	158
5.4.7	Computational overhead	158

5.4.8	Consistent weights	160
5.4.9	Comparison to dimensionality reduction methods	161
5.5	Discussion	162

6 Experimental verification of a coarse-grained model of mRNA

localization reveals robustness regulated by crowding at ring canals164

6.1	Introduction	165
6.2	Modelling approaches to RNA regulation in oogenesis	167
6.3	A coarse-grained model for <i>gurken</i> mRNA localization	168
6.4	Bayesian inference framework	170
6.4.1	Choice of prior distributions	173
6.4.2	Implementation of Monte Carlo methods	173
6.4.3	Image analysis	174
6.4.4	Identification of developmental time via an exponential growth model	176
6.5	Quantification of higher-order assembly of <i>gurken</i> mRNA complexes in the oocyte	177
6.6	Biased transport through ring canals revealed by parameter estima- tion at quasi-steady state	179
6.6.1	Analytic solution to the coarse grained ODE model	179
6.6.2	Parameter estimation at quasi-steady state	181
6.7	Estimates of production and transport rates show production and transport are balanced	183
6.7.1	Sensitivity to rate parameters	183
6.7.2	Balance between parameter estimates may be due to uniden- tifiability	186
6.8	Perturbing mRNA production yields results inconsistent with model predictions	187
6.8.1	Estimates of γ	187

6.8.2	Prediction of the distribution of mRNA complexes for an over-expression mutant	189
6.9	Localization of mRNA in the oocyte of the over-expression mutant reveals robustness	191
6.10	Biological mechanisms to explain the over-expression data can be represented as alternative models	192
6.11	Model comparison	194
6.12	Model comparison reveals blocking of ring canals as the most plausible mechanism	194
6.13	Reexamining microscopy data provides evidence of queuing at ring canals	195
6.14	Model generalizations support the same conclusions about the blocking mechanism	196
6.14.1	Extension to model with mRNA decay	198
6.14.2	Analysis of inhomogeneous variance model	198
6.15	Discussion	200
7	Discussion	204
7.1	Discussion	205
7.1.1	Chapter 3	205
7.1.2	Chapter 4	206
7.1.3	Chapter 5	206
7.1.4	Chapter 6	207
7.2	Future work	208
7.2.1	Further work to extend inference methods	208
7.2.2	Further work to test current biological hypotheses	208
7.2.3	Further work to extend modelling of mRNA localization	209
7.2.4	Further work on software and reproducibility	211
7.2.5	Details on computational methods used	211

7.2.6	Conclusions	212
A	Supplementary Material for Chapter 3	213
A.1	Experimental methods and materials	214
A.1.1	Fly strains and tissue preparation	214
A.1.2	Live-Cell imaging	214
B	Supplementary Material for Chapter 5	215
B.1	Theoretical justification	216
C	Supplementary Material for Chapter 6	223
C.1	Experimental methods and materials	224
C.1.1	Fly strains and tissue preparation	224
C.1.2	Single molecule FISH	224
C.1.3	Fixed-cell imaging	224
C.2	Quantifying production via nascent transcription	225
	Bibliography	227

List of Figures

1.1	The predict-test-refine loop connecting mathematical modelling with the underlying biological process via testable experimental predictions	3
2.1	Illustration of regression adjustment for a rejection ABC algorithm	27
2.2	Illustration of the ABC-SMC algorithm	29
2.3	Illustration of the relationships between collections of particles at successive generations of a SMC algorithm	30
3.1	Distribution of <i>gurken</i> mRNA in <i>Drosophila</i> egg chambers	48

3.2	Ring canals form cytoplasmic bridges between nurse cells and the oocyte	49
3.3	Stage 6 <i>Drosophila</i> egg chamber highlighting the connections between cells due to ring canals	51
3.4	The distribution of microtubules in the nurse cells of a stage 5 <i>Drosophila</i> egg chamber	52
3.5	Typical behaviour of VJP model	59
3.6	States and transitions between the states for mRNA complexes mov- ing according to the fine-grained VJP model	60
3.7	The 3D ellipsoid geometry used to represent a nurse cell in the fine- grained VJP model	63
3.8	Representation of the biased microtubule distribution in the fine- grained model	64
3.9	The FPT distribution as simulated via the fine-grained model	68
3.10	The evolution of a typical path from the fine-grained model until absorption at the posterior ring canal	69
3.11	Sensitivity analysis of the FPT as simulated with the fine-grained model	71
3.12	Sensitivity analysis of the spatial distribution as simulated from the fine-grained model	74
3.13	Marginal plots of the posterior distribution using individual summary statistics	78
3.14	Marginal plots of the posterior distribution using population sum- mary statistics	79
3.15	Marginal plots of the posterior distribution using combined summary statistics	80
3.16	<i>gurken</i> mRNA complexes moving in the nurse cells tracked over a timescale of 30s	81
3.17	Marginal plots of the posterior distribution using observed real data .	82
3.18	Prediction for the FPT of a Dynein mutant	84

3.19	Photo-activation of <i>grk</i> *MCP-Dendra	87
4.1	Four example VJP trajectories	93
4.2	Measurement of angle changes at discrete times	97
4.3	Hidden and observed states in a partially observed Markov process model	98
4.4	Example paths for each pattern of hidden states	100
4.5	True and observed angle changes based on discrete time observations	103
4.6	Comparison between simulated results and theoretical predictions for the observed angle change	107
4.7	Traceplots and autocorrelation functions for MCMC chains	109
4.8	Dependence of the variance of estimates of the log-likelihood on the number of particles used in the particle filter	111
4.9	The data used in the pMCMC inference discretized at differing reso- lutions	114
4.10	Results of parameter estimation via pMCMC varying experimental design parameters	115
4.11	The probability of sequences of hidden states as Δt varies	116
4.12	Results of parameter estimation via pMCMC varying the size of the dataset	117
4.13	Results of parameter estimation via pMCMC jointly varying experi- mental design parameters	119
4.14	Parameter estimation via ABC rejection sampling	121
4.15	The computational cost of parameter estimation with pMCMC and ABC per effective sample size depends on Δt	124
4.16	Comparison between assumed distribution of observed angle changes and simulated distributions under model misspecification	126
4.17	Estimates of the log likelihood via a particle filter	128
4.18	The posterior distribution obtained under model misspecification . . .	129

4.19	Posterior distribution for model of RNA transport in a <i>Drosophila</i> oocyte	131
5.1	Posterior for parameter θ of the uniform toy model for different weights in the ABC distance function	150
5.2	Posterior for parameters θ of the bimodal toy model for different weights in the ABC distance function	152
5.3	Posterior for parameters θ of the Gaussian toy model for different weights in the ABC distance function	153
5.4	Posteriors for parameters k_1 and σ in the death process test problem for different weights in the ABC distance function	155
5.5	Posteriors for parameters θ of the dimerization system for different weights in the ABC distance function	157
5.6	Posteriors for parameter θ in the simple diffusion model for different weights in the ABC distance function	159
5.7	The optimal distance weights found from the search procedure	161
6.1	Behaviour of the ODE model	171
6.2	Typical behaviour of the coarse-grained ODE model across each of the cells in the egg chamber	172
6.3	Prior predictive distribution	174
6.4	Traceplots indicate convergence of MCMC chains	175
6.5	Exponential growth model to give timescale based on area measurements	177
6.6	Assembly of higher order <i>gurken</i> complexes in the oocyte	178
6.7	Estimation of the bias in transport through ring canals	182
6.8	Posterior predictive distribution for the wild type	184
6.9	Pairwise posterior plot for the full dynamic coarse-grained ODE model	185
6.10	Sensitivity analysis for the ODE model	186

6.11	Comparison of phenotypes between wild type and over-expression mutant	188
6.12	Posterior distribution for the ratio between production in wild type and the over-expression mutant	189
6.13	Posterior predictive distribution for the over-expression mutant	190
6.14	Rate of mRNA increase is robust between wild type and mutant . . .	191
6.15	Nascent transcription measured in the nuclei of nurse cells	192
6.16	Estimated inhomogeneous production values across cells	193
6.17	Model comparison based on predictions for the over-expression mutant	196
6.18	Blocking of ring canals results in significant accumulation of mRNA far from the oocyte	197
6.19	Posterior predictive distribution for the model with mRNA decay . .	199
6.20	Posterior predictive distribution for the overdispersion model	200
6.21	Model comparison weights between homogeneous and inhomogeneous overdispersion models	201
7.1	Positioning of ring canals in two different nurse cells	210
B.1	The estimator $L(\mathbf{w}) = \hat{D}_h \left(X_{1:N} Y_{1:N}^{(\mathbf{w})} \right)$ is piecewise constant	219

List of Abbreviations

2D	Two-dimensional
ABC	Approximate Bayesian computation
AIC	Akaike information criterion
ATP	Adenosine triphosphate
<i>bcd</i>	<i>bicoid</i>
BGD	Background
BIC	Bayesian information criterion
BMA	Bayesian model averaging
CPU	Central processing unit
DIC	Deviance information criterion
ESS	Effective sample size
FC	Follicle cells
GFP	Green fluorescent protein
<i>grk</i>	<i>gurken</i>
GUI	Graphical user interface
HMC	Hamiltonian Monte Carlo
HMM	Hidden Markov model
KL divergence	Kullback-Leibler divergence

MAD		Median absolute deviation
MCMC		Markov chain Monte Carlo
mRNA		Messenger ribonucleic acid
<i>nos</i>		<i>nanos</i>
NC		Nurse cell
NPC		Nuclear pore complex
ODE		Ordinary differential equation
<i>osk</i>		<i>oskar</i>
OO		Oocyte
PDE		Partial differential equation
pdf		Probability density function
pMCMC		Particle Markov Chain Monte Carlo
RNA		Ribonucleic acid
RNP		Ribonucleoprotein
SDE		Stochastic differential equation
SMC		Sequential Monte Carlo
smFISH		Single molecule <i>in situ</i> hybridization
SIS		Sequential Importance Sampling
VJP		Velocity Jump Process
WAIC		Widely applicable information criterion

Chapter 1

Introduction

In biological systems, where non-linear behaviour is common, intuition can be a misleading guide. Mathematical and computational models can aid researchers in interpreting data by formalizing hypotheses about nonlinear and indirect interactions between components of a model [4]. Interfacing mathematical modelling and quantitative experimental data is key to unlocking the full potential of both of these techniques to investigate biological hypotheses, but remains a major challenge. Due to advances in experimental technologies, there is an abundance of quantitative imaging data available, particularly in the study of developmental biology and cancer. This mountain of data represents a key challenge to biology and medicine; the development of appropriate tools to link mathematical models with quantitative data to enable predictions and testing of biological hypotheses is therefore crucial. The aim of this thesis is the development and application of such tools in the context of developmental biology.

Quantitative modelling enables us to formalise hypotheses about how biological systems operate [5, 6]. Through modelling, we can gain an understanding of the underlying biological mechanisms controlling a certain process. In addition, a fully quantitative framework, comprising a mathematical or computational model together with quantitative data and a parameterization to fit the model to the data, enables us to make predictions about behaviour under perturbations to the biological system, such as genetic mutations. Experimentally testable predictions enable us to close the predict-test-refine loop (see Figure 1.1), improving our biological understanding in the process, and subsequently also refining our computational model. Reliable model predictions have the potential to enable the use of quantitative models as a screening tool to explore behaviour *in silico* that may be more costly to investigate *in vivo* [7, 8]. Adopting a quantitative framework also allows appropriate propagation of uncertainty through a model [9].

Throughout this work, we take a Bayesian perspective on statistical inference in order to link models with quantitative data. This perspective lets us account for

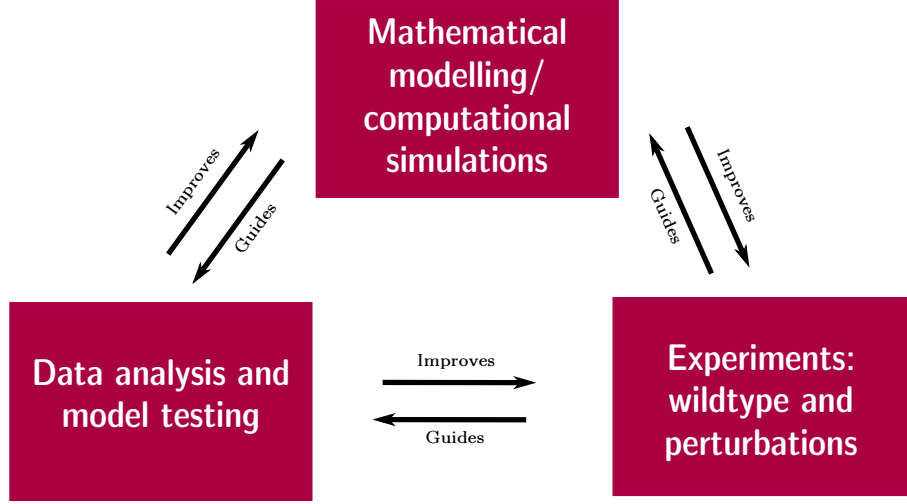


Figure 1.1. The predict-test-refine loop connecting mathematical modelling with the underlying biological process via testable experimental predictions. This approach enables an improved understanding of relevant biological mechanisms, and improved experimental predictions.

measurement error in the data collection process, in addition to underlying stochastic variability in the biological process itself. When we make testable predictions about the behaviour of a system under a biological perturbation, it is important to account for the uncertainty in this prediction, which is not available from a point estimate.

In both Bayesian and frequentist settings, the likelihood is a central quantity. Given observed data, $\mathbf{y}$, and model parameters to infer, θ , then Bayes' theorem,

$$p(\theta|\mathbf{y}) \propto p(\mathbf{y}|\theta)p(\theta), \quad (1.1)$$

relates the posterior distribution, $p(\theta|\mathbf{y})$, which provides information about the parameters based on the observed data, to our prior beliefs encoded via $p(\theta)$ and the likelihood, $p(\mathbf{y}|\theta)$, which describes the probability of observing data, $\mathbf{y}$, given model parameters, θ . Closed form solutions and methods to sample from the posterior distribution rely on the likelihood, $p(\mathbf{y}|\theta)$. For complex biological models, the likelihood function can be intractable in many cases, due to the complexity of the model or due to computational cost. Likelihood-free methods [10–13] such as approximate Bayesian computation (ABC) [14–19] have been developed to facilitate inference

for models with intractable likelihoods.

1.1 Grand challenges

The problem of parameterizing a model becomes even more critical, but also more challenging, as biological models grow in complexity; consider, for example, whole cell models [20–22], multiscale models [23–26], and hybrid models [27–30]. These modelling frameworks may offer opportunities to study emergent phenomena across scales, or to perform realistic patient-specific simulations and analysis. Patient-specific modelling requires accurate parameterization of computational models in addition to detailed understanding of the underlying biology [31, 32], but can enable design of patient-specific treatment protocols [33, 34]. Progress towards parameterizing these complex models requires the development of new methodological techniques due to the computational challenges of evaluating these expensive models, and the statistical challenges of incorporating information from diverse sources of data.

1.1.1 Statistical challenges: summary statistics

ABC avoids evaluation of the likelihood by comparing simulated data with observed data. In practice, this comparison can be challenging and inefficient for high-dimensional data [35]. Summary statistics are used to reduce the dimensionality of both observed and simulated data, thus aiding in comparison. If sufficient statistics are available then no information is lost in comparing summary statistics rather than the data directly. In practice, sufficient statistics are not usually available and the choice of summary statistics is far from obvious, even with guidance from biological domain knowledge. Using insufficient statistics results in an extra layer of approximation during inference with ABC and a poor choice of summary statistics can be prohibitive to inference. Open questions remain around the role

of summary statistics in ABC [2, 36, 37] relating to selection of summary statistics and how best to combine different sources of data.

1.1.2 Experimental design informed by computational modelling

Computational approaches can offer key insights into experimental design [1, 38] which, in turn, can enable more effective study of biological phenomena. Care must be taken since experimental design can influence the biological conclusions drawn when comparing models [39]. Choice of summary statistics for inference relates directly to experimental design, in the sense that, from one point of view, experiments involve collecting summary statistics of an underlying (biological) data-generating process. Interdisciplinary approaches are needed to bring together modelling, experiment, and inference in ways that get the best out of each of these separate parts, such that testable predictions can be made to probe hypotheses.

When drawing conclusions from a computational study, it is important that the results of this work do not depend on the computational implementation of the study. Robustness to the exact methodology used is desirable if the conclusions made are to relate to the biological mechanism rather than simply the representation of this mechanism in the computational model. This applies both to the implementation of a complex computational model, such as a cell-based model [40] or agent-based model [41], as well as to the statistical inference techniques used [42]. Concerns that the implementation of computational studies may influence results can govern experimental design choices [41] to ensure that, when a model is combined with data, the conclusions drawn are reliable.

1.1.3 Biological challenges: mRNA localization

Proteins provide the machinery for cells to perform a variety of different functions. In order for the proteins to perform their functions, they need to be in the right place at the right time. It is easier to regulate and move mRNA and then translate the mRNA into protein in position, as opposed to moving large proteins directly, particularly for large polarized cells such as oocytes and neurons. A single mRNA can be localized in a dormant state to give rise to many proteins locally once activated; it can then be degraded, which allows precise spatial and temporal control of gene expression. mRNA localization is important for setting up the body axes during development [43, 44] and for synaptic plasticity in the brain [45]. Studying mRNA localization offers both an opportunity to gain insight into important biological phenomena and also a good case study for the application of quantitative modelling and inference methods.

In development, precise coordination of the expression of mRNA and protein must be robust to changes, so that development can occur in the same way across many individuals. It remains unclear exactly how this robustness is controlled to ensure that development progresses consistently despite genetic and environmental variation [46].

Insight into how a biological process is regulated can be gained by studying bottlenecks of the process, since such bottlenecks often act as points of control. Such understanding can guide further experiments or even synthetic control of the biological process. The process of mRNA localization involves several different steps, including transcription, nuclear export, active transport on the cytoskeleton, and transit between cells via ring canals. A key question is how the timescales of these processes relate to one another, and whether there are any obvious bottlenecks in the mRNA localization process. Any such bottlenecks would be important targets for regulation by the cell.

1.2 Aims and outline

The overarching aim of this thesis is to develop and apply Bayesian inference tools for the study of key phenomena in cell and developmental biology. In Chapter 2, we review necessary background on Bayesian inference tools from the point of view of a mathematical or computational biologist looking to fit models to biological data and make predictions with these models.

Following this, Chapter 3 outlines the main biological application considered in this thesis: mRNA localization in the development of the fruit fly *Drosophila melanogaster*. We review experimental and modelling work on regulation of mRNA in *Drosophila* development, focusing on transport and localization of mRNA. By developing a fine-grained stochastic model, we examine how transport of mRNA is regulated within a single nurse cell. Using ABC methods, we fit this microscopic model to experimental data and assess the effect of different data types. Parameterizing the model in this way allows us to make experimentally testable predictions.

Motivated by the consideration of the experimental design needed to gather data for the fine grained model, we present in Chapter 4 a computational study on the effects of experimental design on parameter inference in biological transport models. We formulate our biological transport model within a hidden Markov model framework to enable inference via particle Markov chain Monte Carlo methods. This enables us to consider the effect of varying experimental design parameters, such as how often we observe the system, or how much experimental noise is present, and make decisions about trade-offs between these.

Fitting the microscopic model to data via ABC in Chapter 3 provides further motivation for a theoretical study in Chapter 5 that aims to provide a method to automatically select summary statistics relevant to parameter inference using ABC. In practice, it is far from obvious how to decide on what summary statistics to use when implementing ABC. We outline a method that provides weights for a

collection of summary statistics within the ABC distance function and demonstrate the method on some test problems.

Revisiting the mRNA localization process considered previously in Chapter 3, we consider the same process on a larger spatial scale in Chapter 6, wherein we develop a coarse-grained differential equation model of transport between nurse cells and the oocyte. By making testable predictions from this coarse-grained model for a genetic mutant, we highlight the tight regulation of mRNA localization within the *Drosophila* egg chamber, and show that crowding and blocking of ring canals can be an additional mechanism that contributes to the robustness of mRNA localization.

Finally, in Chapter 7 we discuss the implications of this work, both biologically and for the wider modelling community. We make recommendations about sensible approaches to interdisciplinary science in the form of interactions between modelling and experiment, as well as highlighting useful frameworks for modelling and inference.

All experimental results shown in this thesis are due to a collaboration with Dr. Richard Parton and Prof. Ilan Davis. Experiments and data collection were performed by Dr. Richard Parton. My contribution lies in the data analysis, modelling, inference, and development of methodology.

Chapter 2

Bayesian and approximate

Bayesian inference

2.1 Bayesian inference

Decision making under uncertainty is an important part of science. Is the growth rate of a certain cell type greater before or after treatment with a drug? How many people in the UK will become infected with flu in the next week? Bayesian inference offers a principled framework for answering questions about the world around us, and allows decisions to be made based on a combination of prior beliefs about the world and observations. In the following chapter, we will provide an overview of Bayesian approaches and methods for parameter estimation and prediction, focusing on methods appropriate for stochastic and differential equation models.

2.1.1 Bayes' theorem

The foundation of Bayesian inference is in conditional probability and Bayes' theorem. Suppose we have observed data, $\mathbf{y}$, and a generative model, $f(\mathbf{x}|\theta)$, dependent on parameters, θ . Bayes' Theorem allows us to relate the posterior distribution for parameters θ , that is the conditional distribution of the model parameters given the observed data, $\mathbf{y}$, to the likelihood, $p(\mathbf{y}|\theta)$, and our prior information, $p(\theta)$, about the parameters, θ :

$$p(\theta|\mathbf{y}) = \frac{p(\mathbf{y}|\theta)p(\theta)}{p(\mathbf{y})}. \quad (2.1)$$

Bayes' theorem (2.1) can be interpreted as an updating of beliefs. The prior distribution, $p(\theta)$, encodes current scientific knowledge about a biological system of interest due to previous studies and other external data sources. After observing new experimental data, $\mathbf{y}$, we update our beliefs about the parameters, θ , in light of how consistent they are with the new data, $\mathbf{y}$, which we quantify via the likelihood. The likelihood, $p(\mathbf{y}|\theta)$, gives the probability of observing data $\mathbf{y}$ given parameters θ . The denominator $p(\mathbf{y})$, which is known as the evidence or marginal likelihood, provides a normalizing constant to ensure that the left hand side of Equation (2.1)

is a probability distribution.

In practice, even for simple models, it is not analytically tractable to compute the posterior distribution in any closed form way. Conjugate models can permit this but are not widely used in computational biology; see Hines [47] for details. We resort instead to simulation-based approaches.

2.1.2 Expectation and Monte Carlo

In order to extract the rich information contained in the posterior distribution, $p(\theta|\mathbf{y})$, we need to compute expectations with respect to this distribution. We consider expectations of the form

$$\mathbb{E}[h] = \int h(q)\pi(q)dq, \quad (2.2)$$

where h is the function of interest, and π is the density with respect to which we are computing the integral (often $\pi(\theta) = p(\theta|\mathbf{y})$ in a Bayesian inference setting). For example, taking $h(q) = q$ allows one to compute the mean or $h(q) = \mathbb{1}_{q>10}$ can help to compute the probability that the model parameters are greater than 10.

Similarly, in a Bayesian setting, we may need to make predictive inferences about unknown observable variables. Before considering data, $\mathbf{y}$, the distribution of the unknown observable, $\tilde{\mathbf{y}}$, is given by the prior predictive distribution obtained by integrating over θ in the joint distribution of $\tilde{\mathbf{y}}$ and θ .

$$p(\tilde{\mathbf{y}}) = \int p(\tilde{\mathbf{y}}, \theta)d\theta = \int p(\theta)p(\tilde{\mathbf{y}}|\theta)d\theta. \quad (2.3)$$

After observing data, $\mathbf{y}$, we update our beliefs via Bayes' Theorem (2.1) to obtain a posterior distribution, and can make predictions by the same process. This forms the posterior predictive distribution $p(\tilde{\mathbf{y}}|\mathbf{y})$ via

$$p(\tilde{\mathbf{y}}|\mathbf{y}) = \int p(\tilde{\mathbf{y}}|\theta)p(\theta|\mathbf{y})d\theta. \quad (2.4)$$

In this way, the problem of extracting information from data has been reduced to computing integrals of functions with respect to the posterior density.

Quadrature approaches to compute integrals scale very badly with the dimension of the space over which the integration is performed. To form a grid of n points in each of d dimensions will require n^d function evaluations. Instead, for high-dimensional integrals, Monte Carlo methods are popular, as these have convergence rates independent of the dimension, d . A natural estimator to take for an integral formulated as an expectation is to average the function, h , evaluated at samples drawn from the distribution of interest, as follows

$$\hat{h}_N = \frac{1}{N} \sum_{i=1}^N h(q_i), \quad (2.5)$$

where $q_i \sim \pi(q)$ for $i = 1, \dots, N$. By the Strong Law of Large Numbers, $\hat{h}_N \rightarrow \mathbb{E}[h]$ as $N \rightarrow \infty$. When the expectation of h^2 under $\pi(q)$ is also finite, then the speed of convergence of the estimate $\hat{h}$ can also be assessed [48], and is proportional to $1/\sqrt{N}$. For sufficiently large N ,

$$\hat{h}_N = \frac{1}{N} \sum_{i=1}^N h(q_i) \quad (2.6)$$

$$\sim \mathcal{N}\left(\mathbb{E}[h], \frac{\text{Var}[h]}{N}\right). \quad (2.7)$$

In practice, this convergence rate can be prohibitively slow. Techniques have been developed for improving the Monte Carlo convergence rate, many of which focus on variance reduction techniques to reduce $\text{Var}[h]$. Recent work in this area has focused on Multilevel Monte Carlo methods [49–52], which use a telescoping sum to incorporate coupled simulations from low accuracy methods along with fewer simulations from high accuracy methods, and Multifidelity Monte Carlo methods [53], which aim to expand these ideas to a more general framework for use of arbitrary surrogate models.

Further to this work, Quasi Monte Carlo methods [54–56] have proved successful in improving the overall scaling of the convergence rate for Monte Carlo methods from $\mathcal{O}(N^{-1/2})$ to up to $\mathcal{O}((\log N)^d N^{-1})$. These methods rely on low discrepancy sequences of numbers rather than (pseudo) random numbers. For some stochastic biological models, however, the discrete nature of the problems can be prohibitive in achieving these optimal rates [57].

2.1.3 Importance sampling

Often we may be unable to draw samples directly from the target distribution of interest, $\pi(\theta)$. Instead, importance sampling provides an alternative whereby we draw samples from another distribution, $g(\theta)$, which we consider as an approximation to the target distribution, $\pi(\theta)$, and weight the terms in the Monte Carlo sum appropriately. We rewrite the expectation with respect to $\pi(\theta)$ as an expectation with respect to $g(\theta)$, assuming that $\pi(\theta) > 0$ implies $g(\theta) > 0$:

$$\mathbb{E}[h(\theta)] = \frac{\int h(\theta)\pi(\theta)d\theta}{\int \pi(\theta)d\theta} \quad (2.8)$$

$$= \frac{\int [h(\theta)\pi(\theta)/g(\theta)] g(\theta)d\theta}{\int [\pi(\theta)/g(\theta)] g(\theta)d\theta}. \quad (2.9)$$

This can be viewed as a change of measure from $\pi(\theta)d\theta$ to $g(\theta)d\theta$. Taking samples $\theta_1, \dots, \theta_N$ from $g(\theta)$ allows the estimation of $\mathbb{E}[h]$ via the following Monte Carlo sum:

$$\frac{\frac{1}{N} \sum_{i=1}^N h(\theta_i) w(\theta_i)}{\frac{1}{N} \sum_{i=1}^N w(\theta_i)}, \quad (2.10)$$

where

$$w(\theta_i) = \frac{\pi(\theta_i)}{g(\theta_i)} \quad (2.11)$$

are importance weights. If the proposal density $g(\theta)$ can be chosen such that $w(\theta)h(\theta)$ is approximately unity, then precise estimates of the integral are possible [58]. We have assumed that, although we are unable to draw samples directly

from the target, $\pi(\theta)$, we are nonetheless able to evaluate the value of $\pi(\theta_i)$ at a given point θ_i . It is possible to relax this requirement slightly and formulate importance sampling for unnormalized densities where we are unable to evaluate the normalizing constant [58].

2.1.4 Markov chain Monte Carlo

Markov chain Monte Carlo (MCMC) methods are a broad class of algorithms to construct a Markov chain of samples $\theta_1, \theta_2, \dots$, such that these samples are distributed according to the target distribution $\pi(\theta)$. A Markov chain is a sequence of random variables $\theta_1, \dots, \theta_i, \dots$ such that for any $i \geq 2$, the next random variable only depends on the most recent variable. That is $\theta_i | (\theta_{i-1}, \dots, \theta_1)$ has the same distribution as $\theta_i | \theta_{i-1}$. By constructing these methods in the right way, we can ensure that the Markov chain has an equilibrium or stationary distribution equal to the target distribution of interest, $\pi(\theta)$. Provided the Markov chain we construct is ergodic, then this guarantees convergence of the empirical average of the samples to the desired expectations that we wish to compute [48]:

$$\frac{1}{N} \sum_{i=1}^N f(\theta_i) \rightarrow \mathbb{E}_{\pi}[h(\theta)] \text{ as } N \rightarrow \infty. \quad (2.12)$$

A number of questions remain, including how should we determine the transition between states θ_{i-1} and θ_i of the Markov chain, and how do we know whether the Markov chain has converged to the target distribution when in practice we must stop any computations after a finite number of steps? In a Bayesian context, if it was so hard to compute the the posterior distribution in closed form, and we also cannot sample from it directly, how is it going to be any easier to construct an ergodic Markov chain with the posterior as a stationary distribution? We now outline brief sketches of some answers to these questions.

Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm is a general class of MCMC methods based on an adapted random walk with an acceptance/rejection rule to ensure convergence to the correct target distribution, $\pi(\theta)$. It is based on the original work of Metropolis et al. [59] in the context of interacting particles in statistical mechanics.

Suppose that we have a proposal distribution $K(\tilde{\theta}|\theta)$, which will dictate how jumps to new states of the Markov chain are proposed. Given an initial state θ_0 , which can be drawn at random or chosen deterministically:

1. Propose a new state $\tilde{\theta} \sim K(\cdot|\theta_{t-1})$.

2. Take

$$\theta_t = \begin{cases} \tilde{\theta} & \text{with probability } \alpha(\theta_{t-1}, \tilde{\theta}), \\ \theta_{t-1} & \text{with probability } 1 - \alpha(\theta_{t-1}, \tilde{\theta}), \end{cases} \quad (2.13)$$

where the acceptance rate is

$$\alpha(x, y) = \min \left(\frac{\pi(y)K(x|y)}{\pi(x)K(y|x)}, 1 \right).$$

In the case where the proposal distribution, $K(\tilde{\theta}|\theta)$, is symmetric, then the acceptance rate reduces to

$$\alpha(x, y) = \min \left(\frac{\pi(y)}{\pi(x)}, 1 \right).$$

Often a multivariate normal distribution is used for the proposal distribution, $K(\tilde{\theta}|\theta)$. This case is known as the random walk Metropolis-Hastings algorithm. A key implementation hyperparameter is the covariance matrix of the multivariate normal distribution. There is a trade-off in the choice of how large to make the jumps in exploring parameter space. With very large jumps, then with high probability the proposed state will be far from the target distribution, particularly in high dimensions, which is likely to be rejected by the accept/reject step, resulting in a

low acceptance probability. When very small jumps are used, the proposed state will remain close to the current position, moving across the target distribution by diffusion, which can be slow to explore the full distribution. As such, both very large and very small jumps result in inefficient implementations of the Metropolis-Hastings algorithm.

Suppose we take a proposal distribution of the form

$$K(\tilde{\theta}|\theta) = \mathcal{N}(\theta, \mathbf{I}_d \sigma_d^2),$$

where $\mathbf{I}_d$ is the identity matrix in dimension d , and

$$\sigma_d^2 = \frac{l^2}{d},$$

for scaling factor l . Theoretical results from Roberts et al. [60] demonstrate that there is an optimal scaling l^* giving the optimal acceptance rate of $\alpha^* = 0.234$. This allows the most efficient exploration of the target distribution. In practice, to tune MCMC algorithms it is recommended to use trial runs of the MCMC to tune the acceptance rate first [61]. We note also that these theoretical results are specific to the Metropolis-Hastings class of algorithms.

Hamiltonian Monte Carlo

Even with optimized implementations to tune the proposal distribution, Metropolis-Hastings MCMC algorithms can be prohibitively inefficient in high dimensions. In high-dimensional parameter spaces, the region that contributes the vast majority of the value of an integral (known as the ‘typical set’ [62]) can be vanishingly small. To ensure that a Markov chain stays in this region when generating new samples, gradient information can be very helpful. Hamiltonian Monte Carlo (HMC) provides a class of MCMC algorithms that make use of geometric information about the target distribution in the form of gradient information [63, 64]. This requires stronger

information about the target distribution, $\pi(\theta)$, than usual MCMC algorithms in the form of the gradient of the log-density of the target distribution, $\nabla \log \pi(\theta)$. The position variable in state space is augmented with an auxiliary momentum variable, which together are used to define kinetic and potential energies. Hamiltonian dynamics based on these energies provide the governing equations of motion in augmented state space [63]. Avoiding the random walk behaviour of Metropolis-Hastings style algorithms is one of the major benefits of Hamiltonian Monte Carlo.

In modern statistical computing, the availability of general purpose probabilistic programming languages, such as Stan [65] and Edward [66], allows for easy specification of flexible generative models. In particular, Stan uses HMC in the form of an implementation of the No-U-Turn sampler [67]. To obtain the gradient information required for HMC, Stan employs reverse-mode automatic differentiation [68]. Software of this kind is important in enabling modellers who are not experts in the implementation of Bayesian statistical methods to perform robust and efficient inference with mechanistic generative models.

MCMC diagnostics

Geometric ergodicity, which bounds how close a Markov chain is to its stationary distribution after a certain number of steps, is sufficient to ensure a Central Limit Theorem for Markov chain Monte Carlo estimators [69]. In turn, the Central Limit Theorem ensures that these estimators converge in a sensible manner as expected. However, demonstrating such properties for even the simplest model is not practical. Instead, MCMC practitioners use diagnostic tests to check whether MCMC chains have desired properties. For example, the split $\hat{R}$ statistic [58] relies on running an ensemble of several Markov chains simultaneously and comparing the results of these to quantify the variation between chains. Inconsistencies between chains are suggestive of pathological behaviour.

Since MCMC chains generate correlated samples, using 1000 draws from a distribution via MCMC will not tell you as much about a distribution as 1000 independent draws. The impact of this correlation can be assessed via the effective sample size (ESS), n_{eff} . In the importance sampling case of Section 2.1.3, the effective sample size can be calculated as

$$n_{\text{eff}} = \frac{\left(\sum_{i=1}^N w_i\right)^2}{\sum_{i=1}^N w_i^2}. \quad (2.14)$$

The maximum value that the effective sample size can take is N , which is equivalent to an independent sample.

MCMC methods aim to sample from a target distribution, which is the stationary distribution of the MCMC chain. Often we do not know in advance where the mass of the target distribution is. An initial condition for the MCMC chain must be chosen. If the initial condition does not lie close to the target distribution, a chain will not provide samples from its stationary distribution initially. After running a chain for long enough, it should approach the mass of the target distribution, but estimators that include the initial part of the chain may still retain some bias. By ignoring the first B steps of the chain (known as using a burn-in or warm-up of length B), the bias in estimators based on MCMC samples can be reduced [58]. Often the first half of a chain is discarded. Use of burn-in is not universally recommended, with some statisticians arguing that due to the Markov property of a Markov chain, all the burn-in does is help to pick a sensible initial condition [70]. Using half the available computational effort to pick the initial condition may seem wasteful, and other suggestions include using an optimizer to help find an initial condition.

Simulation-based checks, such as visualizing the prior predictive distribution and posterior predictive distribution, can be highly informative [71]. Particularly for complex models, such visualizations are strongly recommended [72]. These offer diagnostic information regarding the model and prior in addition to information about the MCMC methods themselves. For validating the outputs of Bayesian

computation algorithms, simulating rank statistics via simulation-based calibration methods can offer insight into how an algorithm has failed [73].

Non-reversible MCMC

The MCMC methods discussed thus far are discrete-time algorithms with reversibility properties. Alternative methods that expand the class of MCMC algorithms have recently been proposed based on non-reversible methods in continuous time [74–76]. These methods use dynamics similar to the velocity jump process to explore the parameter state space, rather than traditional MCMC methods based on variations of the Metropolis-Hastings algorithm, which can be more similar to random walks in state space. Gradient information about the target distribution is used to inform the changes in direction in state space.

2.1.5 Sequential Monte Carlo

Sequential Monte Carlo (SMC) methods encompasses a class of algorithms to sample from a distribution that function by building successive approximations to the distribution [77, 78]. At each generation of a SMC method, importance sampling is used to weight a collection of samples (or particles). To produce samples for the next generation, the existing samples are perturbed to new values and new weights are calculated. Details are provided in Algorithm 1 for a simple bootstrap particle filter [79], also known as a Sequential Importance Sampling (SIS) scheme, since the samples form a sequence of importance sampling approximations to the target distribution.

Issues can arise with degeneracy of particles in cases where only a small subset of the samples at a generation, t , have a non-negligible weight. If the majority of samples are in a region with very low density, their corresponding weights will be close to zero. Upon resampling at the next generation, the population of samples can become dominated by a single sample. The effect of this is inefficient explo-

Algorithm 1 Bootstrap particle filter

- 1: Sample a collection of particles $\{x_1^1, \dots, x_1^M\}$ from an initial density $p(x_1)$.
- 2: **for** $i = 1$ to M **do**
- 3: Compute the weights for each particle, i , via $w_1^i = p(y_1|x_1^i, \theta)$.
- 4: Find the normalized weights

$$\tilde{w}_1^i = \frac{w_1^i}{\sum_{j=1}^M w_1^j}.$$

- 5: **end for**
- 6: Resample M times with replacement from the collection of particles $\{x_1^1, \dots, x_1^M\}$ with probabilities given by the normalized weights $\{\tilde{w}_1^1, \dots, \tilde{w}_1^M\}$.
- 7: **for** $t = 1$ to $(T - 1)$ **do**
- 8: **for** $i = 1$ to M **do**
- 9: Evolve the current collection of particles according to the forward model, by drawing $x_{t+1}^i \sim p(x_{t+1}|x_t^i, \theta)$.
- 10: Compute the weights for each particle, i , via $w_{t+1}^i = p(y_{t+1}|x_{t+1}^i, \theta)$.
- 11: Find the normalized weights

$$\tilde{w}_{t+1}^i = \frac{w_{t+1}^i}{\sum_{j=1}^M w_{t+1}^j}.$$

- 12: **end for**
- 13: Resample M times with replacement from the collection of particles $\{x_{t+1}^1, \dots, x_{t+1}^M\}$ with probabilities given by the normalized weights $\{\tilde{w}_{t+1}^1, \dots, \tilde{w}_{t+1}^M\}$.
- 14: **end for**
- 15: Obtain an estimate of the marginal likelihood using the unnormalized weights, via

$$\hat{p}(\mathbf{y}|\theta) = \prod_{t=1}^T \frac{1}{M} \sum_{i=1}^M w_t^i.$$

- 16: **return** $\hat{p}(\mathbf{y}|\theta)$
-

ration of the target distribution. To avoid these problems with particle degeneracy, the particles can be resampled according to their weights, either when the ESS in Equation (2.14) falls below a certain value, or at every generation. Multinomial resampling is a commonly used method to perform the resampling. Many other resampling schemes, and many other sequential Monte Carlo algorithms are available; see Del Moral et al. [78] for details.

2.1.6 Fitting differential equation models

Differential equations are a powerful tool used throughout applied mathematics to understand processes involving changes in space and time. Although simple linear differential equations can be solved analytically, many biological differential equation models must be solved numerically in practice [80].

Suppose we have a simple ordinary differential equation (ODE) model of the form

$$\frac{d\mathbf{x}}{dt} = F(\mathbf{x}, t; \theta), \quad (2.15)$$

where $\mathbf{x}$ is an observable variable and F describes some mechanistic model of its behaviour dependent on parameters of interest θ . To perform inference for this model, an observation model for how we measure $\mathbf{x}$ should be specified, which will impose a likelihood function, $L(\theta; \mathbf{y}) = p(\mathbf{y}|\theta)$, where $\mathbf{y}$ is an observation of the variable $\mathbf{x}$. If we assume normally distributed noise with variance σ^2 , then the log-likelihood function $l(\theta; \mathbf{y}) = \log(L(\theta; \mathbf{y}))$ is given by

$$l(\theta; \mathbf{y}) = \frac{1}{2} \log(2\pi T \sigma^2) + \sum_{k=1}^T -\frac{(\mathbf{x}(t_k) - \mathbf{y}(t_k))^2}{2\sigma^2},$$

where the t_k are the T observation times, and $\mathbf{x}(t_k)$ is obtained by solving the differential equation (2.15). Since evaluating the likelihood involves solving the differential equation in this case, it can be computationally expensive. MCMC methods presented thus far in this work rely on evaluating the likelihood many times to generate samples from the posterior distribution, meaning that performing Bayesian inference for differential equation models can be extremely challenging.

For these reasons, frequentist methods for inference have remained popular for many mathematical and computational biologists using differential equation models [81], or the focus has remained on showing possible qualitative behaviours of models for certain parameter regimes [82]. These methods remain valuable approaches, able to offer important insights into the mechanisms underlying biological systems,

and were key in many early studies in the field of mathematical biology [83, 84]. Nevertheless, to harness the full potential for quantitative prediction and uncertainty quantification through Bayesian inference, further improvement of existing computational methods is required.

2.1.7 Identifiability

For some models, each different value of model parameters will result in a different model output. For other models, several different parameter values can result in the same model output. Bellman and Åström [85] highlighted how this can cause issues in identifying model parameters. If the same data can be explained by multiple parameter values (or a manifold of parameter values) then it is not possible to find parameters uniquely corresponding to the observed data.

Issues of identifiability particularly affect differential equation models [86]. A model is structurally identifiable if it is possible to determine estimates of the parameters based on observed outputs from the model [87]. It is practically identifiable if parameters can be determined based on realistic noisy measurements of the system. If parameters of interest only occur in the model in combination with other parameters, they are susceptible to problems with (structural) identifiability. Suppose for example, that we have a population undergoing birth and death proportional to the current population at birth and death rates b and d , respectively, such that

$$\frac{dx}{dt} = (b - d)x. \quad (2.16)$$

Since the birth rate, b , only occurs in combination with the death rate, d , these parameters will not be uniquely identifiable in this case, based on observations of the population, x . Rescaling or reparameterizing models can be effective in avoiding some identifiability problems. For example, the birth-death model in Equation (2.16) can be reparameterized via a net birth rate, $r = b - d$, which would be identifiable.

Various methods have been developed to assess identifiability of models. If observables can be expressed as a linear system of model parameters, then considering the rank of the design matrix provides a test for structural identifiability [88]. Local sensitivity analysis approaches can provide insight into identifiability, for example via the singular value decomposition based method of Sher et al. [89]. Raman et al. [90] argue that local analysis is insufficient for assessing the reliability of parameter estimation and suggest alternative methods based on multiscale sloppiness [91]. Tests for identifiability of parameters based on data cloning have also been suggested [92], where data cloning involves taking many copies or clones of a dataset [93]. Bayesian approaches with MCMC can avoid some of the issues with identifiability by integrating out over parameters [88], even if the mass of the posterior is spread over a manifold (although this may make the posterior distribution harder to sample from). Alternatively, MCMC and ABC methods can be used to diagnose identifiability for unidentifiable biophysical models [94, 95].

2.2 Likelihood-free inference

There are a wide variety of situations that can lead to the likelihood function being intractable. Computational agent-based and stochastic models are powerful biological modelling tools for realistic modelling of complex biological systems. Multiscale models provide tools to examine behaviour across spatial and temporal scales of biological processes. For many models of this form, it is relatively easy to forward simulate samples from the model, but hard or computationally very expensive to calculate the likelihood function. Similar issues with differential equation models have been outlined in Section 2.1.6. To solve this problem of intractable likelihood functions, likelihood-free methods for Bayesian inference have been developed. We will outline a selection of these methods below.

2.2.1 ABC

Suppose we wish to infer a posterior distribution over parameters θ of a generative model such that we can simulate from $\mathbf{x} \sim f(\mathbf{x}|\theta)$. In ABC, parameters θ are drawn from a prior, $\pi(\theta)$, and data, $\mathbf{x}^*$, is simulated from the generative model using those parameters, such that $\mathbf{x}^* \sim f(\mathbf{x}|\theta)$. The distance between the simulated dataset, $\mathbf{x}^*$, and the real data, $\mathbf{y}$, is calculated using a distance function $d(\mathbf{x}^*, \mathbf{y})$. In cases where the data are discrete, we can consider whether simulated data, $\mathbf{x}$, are exactly equal to observed data, $\mathbf{y}$, and accept parameters correspondingly, which gives exact samples from the posterior ($\epsilon = 0$ case). Often, though, mechanistic models provide continuous data and the acceptance rate from an exact comparison is prohibitively small. Instead, if the distance is below a certain non-zero tolerance, ϵ , then the parameters θ can be accepted into the approximate posterior sample. This introduces an approximation that becomes exact in the limit $\epsilon \rightarrow 0$. This method is known as ABC rejection sampling and is summarized in Algorithm 2. The resulting ABC approximation to the posterior distribution is given by:

$$p_{\text{ABC}}(\theta|\mathbf{y}) \propto \int p(\mathbf{y}|\theta)p(\theta)\mathbb{1}[d(\mathbf{x}, \mathbf{y}) \leq \epsilon] \mathrm{d}\mathbf{x}.$$

Choice of the tolerance ϵ can be avoided, to some extent, by simulating a large number, N , of parameter samples and datasets, calculating the corresponding distances for these and accepting the proportion α that lie closest to the real data [95, 96].

2.2.2 Summary statistics

In many practical circumstances, the data (either collected experimentally or simulated from the *in silico* model) will be very high dimensional. High-dimensional data pose difficulties within the ABC framework, as it is difficult to sensibly estimate when the output of a particular simulation is ‘close’ to the data. Even taking ac-

Algorithm 2 ABC rejection sampling

```
1: Set number of accepted particles  $i = 0$ .
2: while  $i < N$  do
3:   Sample parameter value  $\theta$  from the prior  $p(\theta)$ .
4:   Generate synthetic dataset from the model  $\mathbf{x} \sim f(\mathbf{x}|\theta)$ .
5:   if  $d(s(\mathbf{x}), s(\mathbf{y})) < \epsilon$ , for a distance function,  $d(\cdot, \cdot)$ , summary statistics,  $s(\cdot)$ ,
     and tolerance  $\epsilon$ , then
6:     Accept sample  $\theta$  and set  $\theta_i = \theta$ . Set  $i = i + 1$ .
7:   end if
8: end while
9: return  $\{\theta_i\}_{i=1}^N$ 
```

count of domain expertise, it can be hard to determine which features of the data are important. This issue of comparing high-dimensional data is further compounded using stochastic models where there is noise in the process model in addition to measurement noise. Repeatedly drawing from a stochastic model with the same parameter values can give vastly different outputs.

As such, it is often necessary to work with a lower-dimensional vector of summary statistics, $s(\mathbf{x})$, of the data, such that we require the distance between summary statistics to be less than the tolerance, $d(s(\mathbf{x}^*), s(\mathbf{y})) < \epsilon$. Using (insufficient) summary statistics introduces another approximation into the method. Examples of summary statistics may be data points within a time series, an average transition time between different states of a system, or the moments of a certain species within a model.

Selecting a good choice of summary statistics can have a large effect on the efficiency of an ABC algorithm. Expert knowledge can help guide choice of summary statistics. Nonetheless, designing appropriate summary statistics has proved to be a challenging problem for ABC, and various researchers have sought to address these issues [36, 37, 97]. We review many of these approaches further in Chapter 5 and also introduce an adaptive method to combine summary statistics by automatically selecting weights within the ABC distance function.

2.2.3 Improvements to ABC

A number of additional improvements to the simple ABC rejection sampling procedure of Algorithm 2 have been proposed. We will outline these in the following sections.

Regression adjustments

Post-processing methods to aid with some of the approximations of ABC have been advocated since the earliest literature on ABC [15]. Regression adjustments can correct for the relationship between parameters, θ , and summary statistics, $s(\mathbf{x})$, as shown in Figure 2.1, in the sense of accounting for the imperfect match between observed and simulated data [17]. Further extensions include non-linear, heteroscedastic regression adjustment [98] and ridge regression approaches [35]. Theoretical results on the convergence of regression-adjusted ABC [99] indicate that regression adjustment can be beneficial in the sense that in the limit of infinite data it provides the correct point estimate and uncertainty for the posterior with an acceptance probability that tends to unity for a well-designed algorithm. However, under the assumption of model misspecification, Frazier et al. [100] show that regression adjustments concentrate posterior mass on a different pseudo-true parameter value than for rejection ABC. Recently, a recalibration post-processing adjustment method has been developed [101] based on a diagnostic coverage property for ABC [102].

ABC-MCMC

Rejection sampling algorithms can be very inefficient with a low acceptance rate when the proposal distribution is very different to the target distribution, but designing good proposal distributions can be difficult. This motivated the development of likelihood-free MCMC methods [103, 104], also known as ABC-MCMC methods, to leverage some of the benefits of generating a correlated Markov chain of samples as outlined in Section 2.1.4.

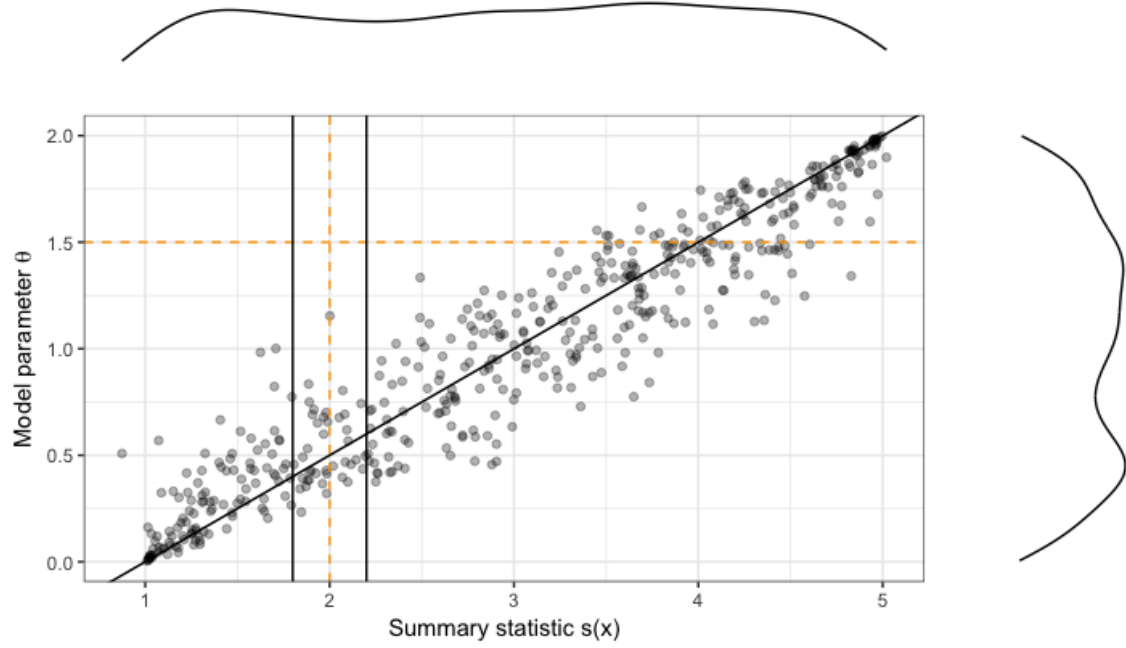


Figure 2.1. Illustration of regression adjustment for a rejection ABC algorithm. Points shown are sampled from the joint distribution, $(\theta, s(\mathbf{x}))$. The prior distribution, $p(\theta)$, is shown as the marginal on the right hand side. The marginal likelihood, $p(s(\mathbf{x}))$, is shown above the plot. ABC aims to approximate the posterior distribution $p(\theta|s(\mathbf{y}))$ based on observed data $\mathbf{y}$, where $s(\mathbf{y})$ is illustrated by the vertical orange dashed line. The horizontal orange dashed line shows a slice through the joint distribution corresponding to the likelihood $p(s(\mathbf{x})|\theta)$ for given θ . An acceptance region is used shown by vertical black lines, such that points between these lines are accepted for the approximate posterior, and points outside are rejected. Conditional density estimation via regression adjustment can aid with the accuracy of the approximation to the posterior distribution. The sloping black line illustrates the regression line used for this adjustment (line is based on local regression about $s(\mathbf{y})$ and so is only fitted to points inside the accepted region).

Model error

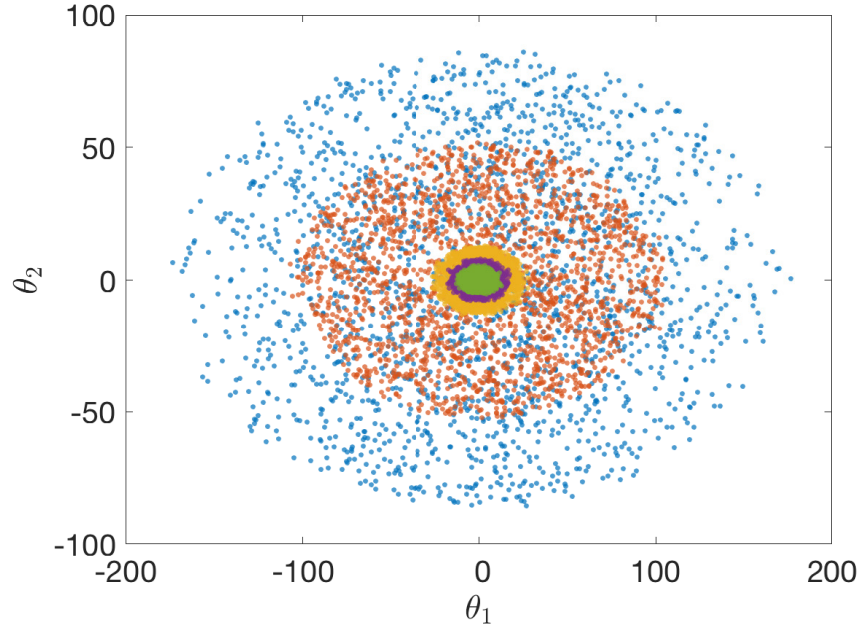
It has been shown that under the assumption of model error, inference via ABC is exact [105], where the acceptance kernel and tolerance, ϵ , govern the distribution of the model error. In practice, when modelling a real biological system, models will be misspecified due to a lack of full understanding of the biological processes.

2.2.4 ABC-SMC

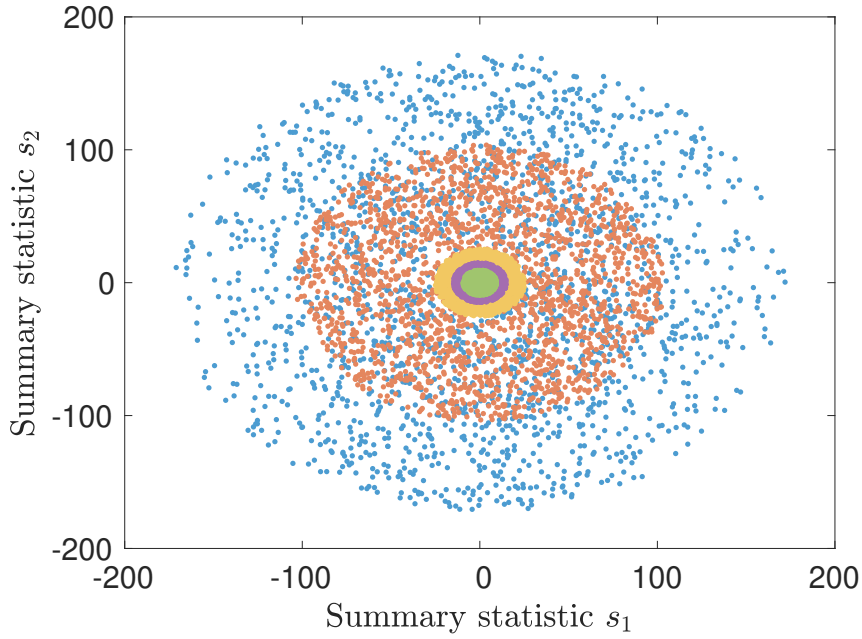
In cases where the prior and posterior distributions are very different, the rejection sampling version of ABC described above can have very low acceptance rates. Algorithm 3 summarizes how samples from an approximate posterior can be generated via a more efficient version of ABC using sequential Monte Carlo techniques, known as ABC-SMC [16, 78, 106]. Importance sampling is used iteratively so that instead of sampling repeatedly from the prior, parameters are sampled from an approximate posterior at each generation of the algorithm. A weight must be given to each sample to correct for the fact that it is not drawn from the prior. In Figure 2.2, the successive approximations to the posterior distribution are demonstrated for a simple toy model. This example also illustrates the acceptance region in summary statistic space, which shrinks here as the tolerances are decreased. An advantage of ABC-SMC methods over ABC-MCMC (see Section 2.2.3) is that, while ABC-MCMC methods can become stuck in regions of low probability, ABC-SMC can avoid getting stuck in such regions if sufficiently many particles are used in the SMC algorithm. Figure 2.3 shows the genealogy of particles in ABC-SMC, illustrating how samples with a high weight are likely to be sampled again at the next generation, whereas samples with a low weight in the tails of the posterior distribution are unlikely to have descendants in the next generation of ABC-SMC.

Improvements to ABC-SMC

In addition to the choice of summary statistics, there are several aspects of the ABC-SMC algorithm that require user specification, and a natural question is how to choose these. Primarily these aspects are the proposal kernel, $K_t(\theta|\theta^*)$, the tolerance thresholds, ϵ_t , and the distance function, $d(s(\mathbf{y}), s(\mathbf{x}))$. Optimal proposals for ABC-SMC have been studied [107] for Gaussian kernels via the Kullback–Leibler (KL)



(a)



(b)

Figure 2.2. Illustration of the ABC-SMC algorithm (see Algorithm 3). (a) shows the accepted samples in the ABC approximation to the joint posterior distribution at successive generations. (b) shows the acceptance region in the space of summary statistics for successive generations of ABC-SMC. Successive generations are shown in different colours, with blue as the first generation. The model used is as follows: $s_1 \sim \mathcal{N}(\theta_1, 2^2)$, and $s_2 \sim \mathcal{N}(2\theta_2, 1^2)$ independently, with prior distribution $\mathcal{N}(0, 100^2 I_2)$, Euclidean distance function, and observed data $(s_1, s_2) = (0, 0)$.

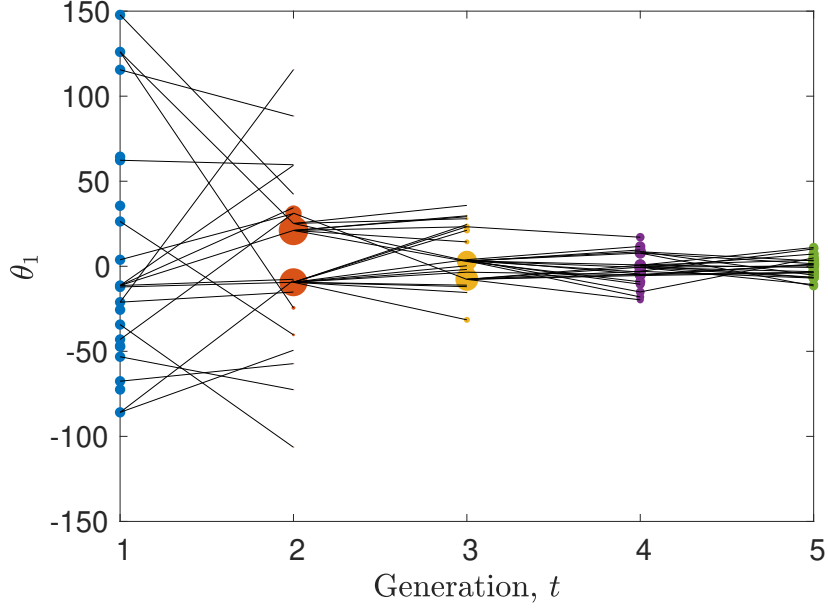


Figure 2.3. Illustration of the relationships between collections of particles at successive generations of a Sequential Monte Carlo algorithm, due to resampling. Points represent a collection of $N = 20$ particles generated via ABC-SMC for the same toy model used in Figure 2.2. The size of the particles correspond to the weights of the particles. At each successive generation, particles are resampled according to their weights. For particles with small weights, there is a high probability that they will not be sampled in the next generation.

divergence¹ between proposal and target, minus the log acceptance ratio. Filippi et al. [108] extended these results and derived optimal kernels of different forms, including local perturbation kernels. Instead, Alsing et al. [109] maximize the effective number of samples per parameter proposal. Optimal threshold design schemes have been proposed [110]. In practice, tolerances are often chosen by taking a quantile of the distribution of ABC distances at each generation of ABC-SMC [95, 96]. Prangle [111] has studied choice of weighted Euclidean distance functions, and developed an adaptive method to choose distance weights within an ABC-SMC framework.

¹The Kullback-leibler divergence is a distance between distributions and is defined as

$$D_{KL}(F, G) = \int F(\mathbf{x}) \log \left(\frac{F(\mathbf{x})}{G(\mathbf{x})} \right) d\mathbf{x},$$

for continuous distributions $F(\mathbf{x})$ and $G(\mathbf{x})$.

Algorithm 3 ABC-SMC

```
1: for  $t = 1$  to  $T$  do
2:   Set  $i = 1$ .
3:   while  $i \leq N$  do
4:     Sample  $\theta^*$  from previous population  $\{\theta_{t-1}^i\}_{i=1}^N$  with weights  $\{v_{t-1}^i\}_{i=1}^N$ .
     Peturb  $\theta^*$  to give  $\theta^{**} \sim K_t(\theta|\theta^*)$ . If  $\pi(\theta^{**}) = 0$ , return to step 5.
5:     Simulate dataset  $\mathbf{x}^{*i} \sim f(\mathbf{x}|\theta^{**})$  and calculate distance  $d(s(\mathbf{x}^{*i}), s(\mathbf{y}))$ .
6:     if  $d(s(\mathbf{x}^{*i}), s(\mathbf{y})) < \epsilon_t$  then
7:       Accept the parameters.
       Set  $\theta_t^i = \theta^{**}$ .
       Calculate the weight  $v_t^i$  for particle  $\theta_t^i$  via

$$v_t^i = \frac{\pi(\theta_t^i)}{\sum_{j=1}^M v_{t-1}^j K_t(\theta_{t-1}^j, \theta_t^i)}.$$

8:       Set  $i = i + 1$ .
9:     end if
10:   end while
11:   Normalize the particle weights  $\{v_t^i\}_{i=1}^N$ .
12: end for
13: return  $\{\theta_T^i\}_{i=1}^N, \{v_T^i\}_{i=1}^N$ 
```

2.2.5 Lazy and multifidelity approaches to ABC

For certain practical problems, it can become clear before a simulation is complete that it is unlikely that the pseudo-data generated by that simulation will be close to the observed data. Lazy ABC [112] offers a method to take advantage of such situations. Suppose that for a proposed parameter value, θ^* , generating simulated data via $\mathbf{x}^* \sim f(\mathbf{x}|\theta^*)$ results in $d(\mathbf{x}^*, \mathbf{y}) > \epsilon$. Additionally, by stopping the simulation early in some way, we could determine that the simulated data, $\mathbf{x}$, will be far from the observed data, $\mathbf{y}$, and save computational cost. In performing the early stopping, we do not want the resulting Monte Carlo estimators to be biased. To ensure that the parameter inference process still targets the correct distribution, not every unlikely, expensive computation is stopped, but instead the simulation is continued with some probability, $a(\mathbf{y})$. This allows an importance weight for continued simulations to be calculated. The benefit of this procedure is that it can save computational cost on simulations that would have been rejected by the ABC

algorithm anyway, at the price of some increased variance of the Monte Carlo estimators. Depending on the problem of interest, the magnitude of the computational savings can easily outweigh any increase in variance.

To take advantage of lazy ABC, there must be a stopping rule available to enable us to indicate the success of a simulation before the simulation is complete. For example, for an ODE (as in Equation (2.15)) or stochastic differential equation (SDE) model with observations in the time interval $[0, T]$, a stopping rule could be determined by integrating numerically up to time $T/2$ and comparing to observed data over $[0, T/2]$. A variety of stopping rules are possible and can be based on analytic results or approximate models in addition to early stopping of simulations.

Recently, these ideas have been further extended to include early acceptance of some parameter values, using ideas from machine learning to guide where it is best to use an approximate model and where to spend computational effort on the full model. Wiqvist et al. [113] use Gaussian process emulators as an approximate model on which to base early accept-reject decisions. Prescott and Baker [114] provide a general framework for use of multifidelity approximate models within ABC, and demonstrate how to make best use of a limited computational budget.

From a more philosophical perspective, lazy ABC methods can be viewed as a way of encoding prior information about a problem (depending on the choice of stopping rule). When performing statistical inference to gain knowledge about a scientific problem, it is common to take broad uninformative priors about parameters of interest, and aim to learn from data without assuming anything about the problem in advance. It can be more effective to attempt to encode what we know about a problem within the prior distribution. An alternative approach to lazy ABC in certain subsets of cases could be a more careful design of prior distributions to capture our scientific or computational knowledge about a problem.

2.2.6 Other likelihood-free methods

Variational methods rely on approximating a distribution of interest by a simpler, more tractable distribution for which it is possible to perform inference via simpler methods [115]. The aim is to find the closest distribution to the exact target within the approximating family of distributions. A popular variational approximation is the mean field approximation, which, given a distribution of interest jointly over many parameters, assumes that this distribution factorizes as a product of distributions over these parameters independently. Variational methods typically seek to minimize the KL divergence between the approximating distribution and the exact distribution.

Indirect inference [116] bases inference on two steps, the first of which is based on an incorrect model, which can be corrected at the second step. Synthetic likelihood methods [11, 13] approximate an intractable likelihood by a normal distribution. The approximation is at the level of the likelihood function, and MCMC methods can be used as usual with this approximate likelihood. This strong normality assumption has been relaxed via a semi-parametric approach to the Bayesian synthetic likelihood method [117].

Similarly, Gaussian process emulator methods have also been proposed [118], which assume continuity when approximating the log-likelihood, an assumption that can help save computational effort. Other machine learning methods such as neural networks have been used for likelihood-free inference based around density estimation for the posterior [119–121]. For example, Papamakarios and Murray [119] assume a parametric form for a posterior approximation as a mixture of Gaussians, and aim to estimate parameters for this mixture using neural networks, showing that many fewer simulator evaluations are required compared to ABC methods.

2.2.7 Particle MCMC

Particle MCMC (pMCMC) methods provide a Bayesian framework for parameter estimation by allowing samples to be drawn from the posterior distribution of the model parameters, given observed data, without needing to evaluate the likelihood function directly. Instead of evaluating the likelihood directly, we can use (unbiased) estimates of the likelihood within an MCMC algorithm. Estimating the likelihood of the observed data given certain parameters can be achieved with a particle filter [79] for a fixed, finite, number of particles (see Algorithm 1). The results of Andrieu et al. [122] demonstrate that, even when a finite number of particles are used in the filter to estimate the likelihood, the MCMC algorithm will still target the correct posterior distribution. These methods have been applied in the context of modelling epidemics [123] and biochemical reaction networks [124, 125].

Suppose we have an observed dataset $\mathbf{y} = \{y_k \mid k = 1, 2, \dots, T\}$. The Metropolis-Hastings MCMC algorithm [59] described in Section 2.1.4 involves an acceptance rate for new parameters

$$\alpha(x, y) = \min \left(\frac{\pi(y)K(x|y)}{\pi(x)K(y|x)}, 1 \right),$$

where $K(y|x)$ is the proposal distribution, and $\pi(x)$ is the target distribution which in this case will be the posterior distribution, $p(\theta|\mathbf{y})$. Evaluation of the likelihood, $p(\mathbf{y}|\theta)$, is required in calculating the acceptance probability for the proposed move, via Equation (2.1): $p(\theta|\mathbf{y}) \propto p(\theta)p(\mathbf{y}|\theta)$. We replace the likelihood with an unbiased estimate of the likelihood, $\hat{p}(\mathbf{y}|\theta)$, obtained from a particle filter.

Details of the bootstrap particle filter algorithm [79] used within the pMCMC algorithm (Algorithm 4) are given in Algorithm 1. A particle filter represents the state of the system via a population of weighted particles [77]. We obtain an estimate of the likelihood by successively updating the hidden state of the system (represented via the particles) and comparing this hidden state with the observed data at each

Algorithm 4 Particle MCMC

- 1: Initialize parameters, θ_0 .
- 2: Run a particle filter (see Algorithm 1) to compute an estimate of the marginalized likelihood $\hat{p}(\mathbf{y}|\theta_0)$, where $\mathbf{y}$ is the observed data.
- 3: **for** $j = 1$ to N **do**
- 4: Draw parameters θ^* from a proposal distribution $K(\cdot|\theta_{j-1})$.
- 5: Run a particle filter to compute an estimate of the marginalized likelihood $\hat{p}(\mathbf{y}|\theta^*)$.
- 6: Accept the proposed move with probability α where

$$\alpha = \min \left(1, \frac{p(\theta^*)K(\theta_{j-1}|\theta^*)\hat{p}(\mathbf{y}|\theta^*)}{p(\theta_{j-1})K(\theta^*|\theta_{j-1})\hat{p}(\mathbf{y}|\theta_{j-1})} \right).$$

 If the move is accepted, set $\theta_j = \theta^*$, otherwise set $\theta_j = \theta_{j-1}$.

7: **end for**

observed time point. This procedure provides new weights for the particles according to how well they match the observed data. To prevent a degenerate situation where the state of the system is represented solely by a single particle, it is necessary to resample from the population of particles according to their weights.

2.3 Model comparison

A large part of the motivation for using modelling relates to representing hypotheses as models. Valuable biological insight can be gained from comparing hypotheses representing competing perspectives on a biological system in the form of models. Methods for model comparison allow us to answer questions about which hypothesis is best supported by the data. However, there can be criticism as to whether we should base decisions on a single criterion or single number from a model comparison method [126]. As scientists, we want to know in a straightforward way which hypothesis is ‘right’ and the temptation is to use model comparison techniques as a black box to make a decision between competing hypotheses. However, as statisticians, we should recognize that this decision-making process is complex and to summarize it in a single number can be misleading under certain circumstances.

To assess a model, we often may not have access to any additional data beyond

the data, $\mathbf{y}$, used to infer parameters of the model. Evaluating a model based on the same data used to fit the parameters can lead to biased results [127]. To correct for this bias due to using the same data, $\mathbf{y}$, twice, various methods incorporating different choices of bias correction to prevent overfitting have been proposed.

In the following section, we will outline methods for comparison between models and discuss their strengths and weaknesses. We focus primarily on Bayesian approaches for mechanistic biological models, although it should be highlighted that some of the approaches described here are not limited to the Bayesian paradigm.

2.3.1 Information criteria

Information theory based methods to assess the ‘best’ model based on data have been in use for several decades. The Akaike information criterion (AIC) [128, 129] is a widely used method for model selection, based on foundations of information theory. Akaike derived an estimator of the (relative) expectation of the KL divergence based on Fisher’s maximized log-likelihood. AIC is based on this estimator and includes a bias correction that accounts for the number of parameters, k , in the model. The criterion takes the following form:

$$\text{AIC} = -2(\log(\hat{L}(\theta; \mathbf{y})) - k), \quad (2.17)$$

where

$$\hat{L}(\theta; \mathbf{y}) = \max_{\theta \in \Theta} p(\theta | \mathbf{y}) \quad (2.18)$$

is the maximum of the log-likelihood for a given model. Akaike recognized the log-likelihood as a measure of model fit [128] and penalized this based on the number of parameters in the model to prevent overfitting, thus accounting for assessing the model with the same data as used to fit it.

An alternative criterion is the Bayesian information criterion (BIC) proposed by Schwarz [130], who views model comparison as a problem of estimating the correct

dimensionality of model to use. The BIC is related to the AIC, but is not an estimate of KL divergence in the same way [131]. The BIC is defined as

$$\text{BIC} = -2(\log(\hat{L}(\theta; \mathbf{y})) - \frac{1}{2}k \log n), \quad (2.19)$$

where $\hat{L}(\theta; y)$ is as defined in Equation (2.18), k is the number of model parameters and n is the dimensionality of the data $\mathbf{y}$. An approximation to the Bayes factor (see Section 2.3.2) comparing two models, M_1 and M_2 , can be obtained via $\exp(-\Delta\text{BIC}/2)$, where ΔBIC is the difference in the BIC values for models M_1 and M_2 [132]. This approximation is motivated by performing an expansion of the marginal likelihood via Laplace's method. The BIC favours lower dimensional models compared to the AIC when $n \geq 8$ [130]. Warne et al. [133] have shown Bayesian approaches such as BIC have benefits over other information criteria when selecting between continuum models of cell spreading and proliferation.

The deviance information criterion (DIC) [134] makes use of Monte Carlo estimation methods to estimate generalization performance of a model, M . The DIC is defined as:

$$\text{DIC} = \frac{1}{n} \sum_{i=1}^n \log p(y_i | \bar{\theta}, M) - \frac{p_{\text{eff}}}{n}, \quad (2.20)$$

where $\bar{\theta}$ is the posterior mean estimate of the parameters θ , and p_{eff} is an estimate of the effective number of parameters given by

$$p_{\text{eff}} = 2 \sum_{i=1}^n (\log p(y_i | \bar{\theta}, M) - \mathbb{E}[\log p(y_i | \theta, M)]) ,$$

with expectation taken with respect to the posterior distribution. As in the AIC and the BIC, the DIC still uses a point estimate of the parameters, θ , rather than taking a full Bayesian approach by integrating out over the parameters [127]. Although the DIC has been widely used due to computational convenience, concerns have also been raised that the DIC is inconsistent with the results of other model comparison

methods such as Bayes Factors, and the DIC can fail to choose the true model in the case that the true model is available within the set of models to be compared [135].

A successor to these information criteria [135] is the Widely Applicable Information Criterion (WAIC) [136]. The WAIC can be computed as follows:

$$\text{WAIC} = \frac{1}{n} \sum_{i=1}^n \log p(y_i | \mathbf{y}, M) - \frac{V}{n}, \quad (2.21)$$

where V is the functional variance given by

$$V = \sum_{i=1}^n \left(\mathbb{E}[(\log p(y_i | \theta, M))^2] - \mathbb{E}[\log p(y_i | \theta, M)]^2 \right), \quad (2.22)$$

with expectations taken with respect to the posterior distribution. The WAIC is asymptotically equal to using Bayesian leave-one-out cross-validation [136] (see Section 2.3.5). Compared to the DIC, WAIC involves estimation of predictive densities as opposed to log densities.

2.3.2 Model evidence and Bayes Factors

The model evidence, or marginal likelihood for a model, $p(\mathbf{y} | M)$, offers an assessment of evidence in support of a given model M . In order to compare competing models, Bayes factors provide a method to quantify the support for one model, M_1 , over an alternative model, M_2 . We can express Bayes' Theorem (compare with Equation (2.1)) as follows:

$$p(M_i | \mathbf{y}) = \frac{p(\mathbf{y} | M_i) p(M_i)}{p(\mathbf{y})}, \quad (2.23)$$

for $i = 1, 2$. The Bayes Factor, $B_{1,2}$, comparing models M_1 and M_2 is given by

$$B_{1,2} = \frac{p(\mathbf{y} | M_1)}{p(\mathbf{y} | M_2)} = \frac{p(M_1 | \mathbf{y}) p(M_2)}{p(M_2 | \mathbf{y}) p(M_1)}. \quad (2.24)$$

In the case of equal priors for M_1 and M_2 , the Bayes factor reduces to a likelihood ratio. Guides on how to interpret the values of Bayes factors have been provided [132, 137]. Nonetheless, some statisticians are uncomfortable with the dependence of Bayes factors on the priors for each model, $p(M_1)$ and $p(M_2)$, which can result in arbitrary increases/decreases to a Bayes factor when using vague priors [138].

Methods to compute Bayes factors via MCMC can be challenging. Thermodynamic integration has been proposed as an effective method for estimating Bayes factors in the context of phylogenetic models [139], but the authors warn that this method is both theoretically involved and requires 10 times more CPU time than simple posterior sampling for the more expensive of the two models being compared. SMC methods (see Section 2.1.5) can provide estimates of the model evidence as a convenient by-product for model comparison [16, 140], and population-based MCMC methods can also prove effective [141]. In the context of approximate Bayesian inference though, there can be further issues with model comparison via Bayes factors. Due to the approximations involved with insufficient summary statistics in ABC, model comparison based on approximate ABC posteriors can result in misleading conclusions [142]. This need not be a concern if the full dataset is used, rather than insufficient summary statistics, or if the posterior distribution is defined via the desired system outputs [143].

2.3.3 Alternatives

With such a zoo of methods for model comparison, each with strengths and weaknesses, it is hard to know which method to use. Further alternatives breaking with the commonly used paradigms for model comparison have been suggested. For example, many of the methods described, including Bayes factors, are based on the logarithmic score function. The Hyvärinen score has been proposed [138] leading to a method robust to vague priors, in contrast to Bayes factors. Some statisticians suggest moving away from decisions based on a single criterion resulting in binary

0-1 decisions [144, 145], proposing instead taking a broader picture of model comparison. Kamary et al. [144] suggest testing hypotheses represented as models M_1 and M_2 via inference for the mixture model:

$$p(\mathbf{y}|\theta) = \mu p(\mathbf{y}|\theta_1, M_1) + (1 - \mu)p(\mathbf{y}|\theta_2, M_2), \quad (2.25)$$

where μ is a mixture weight, and θ_1 and θ_2 are the parameters for the respective models. A mixture weight close to either 0 or 1 can provide evidence in favour of a particular model. Mixture models can be challenging to fit robustly in practice due to identifiability problems (see Section 2.1.7). This may, however, provide a good diagnostic that models M_1 and M_2 are hard to distinguish.

2.3.4 Prediction

Models can be evaluated based on the quality of their predictions about future behaviour, based on unknown future observations, $\tilde{\mathbf{y}}$ [146]. The posterior predictive distribution for a model, M , based on observed data, $\mathbf{y}$, can be written as:

$$p(\tilde{\mathbf{y}}|\mathbf{y}, M) = \int p(\tilde{\mathbf{y}}|\theta, M)p(\theta|\mathbf{y}, M) d\theta, \quad (2.26)$$

by marginalizing out the model parameters, θ . Evaluation of the quality of predictions requires a utility function, $u(M, \tilde{\mathbf{y}})$ [127]. A common choice for this utility function is the logarithmic score

$$u(M, \tilde{\mathbf{y}}) = \log p(\tilde{\mathbf{y}}|\mathbf{y}, M), \quad (2.27)$$

which possesses desirable properties, such as invariance to reparameterization [147], and a connection to the KL divergence between the true distribution of $\tilde{\mathbf{y}}$ and the distribution based on predictions from model M [127].

Consideration of the log score as a utility function leads to the following Monte

Carlo estimator for the expected log predictive distribution (elpd) [148]:

$$\mathbb{E} \left[\log \left(\int p(\tilde{\mathbf{y}}|\theta)p(\theta|\mathbf{y}) d\theta \right) \right] \approx \frac{1}{n} \sum_{i=1}^n \log \left(\int p(\tilde{y}_i|\theta)p(\theta|\mathbf{y}) d\theta \right), \quad (2.28)$$

where unknown observations $\tilde{\mathbf{y}} = (\tilde{y}_1, \dots, \tilde{y}_n)$. We note that in some cases where data are limited, additional observations $\tilde{\mathbf{y}}$ may be unavailable. The following Section 2.3.5 describes strategies appropriate for reuse of a single dataset, $\mathbf{y}$, to assess prediction, in place of additional data, $\tilde{\mathbf{y}}$.

2.3.5 Cross-validation

Cross validation allows the predictive performance of a model to be assessed by efficient use of a single dataset, rather than requiring collection of a full new independent dataset on which to assess predictive performance. A dataset is divided into equally sized subsets, one of which is retained to evaluate the performance of a model fitted to the remaining data. For example, from a dataset of n points of observed data $\mathbf{y}$, we can produce n subsets of the data $\mathbf{y}_{-i}$ for which the single point y_i has been removed. Using these subsets to evaluate predictive performance of a model results in leave-one-out cross validation [149, 150].

2.3.6 Pseudo Bayesian model averaging

Weights for a collection of models under comparison can be obtained via the elpd defined in Equation (2.28). Using a Bayesian leave-one-out estimator of the elpd [151] gives

$$\text{elpd}_{\text{loo}} = \sum_{i=1}^n \log(p(y_i|\mathbf{y}_{-i})) \quad (2.29)$$

where $p(y_i|\mathbf{y}_{-i}) = \int p(y_i|\theta)p(\theta|\mathbf{y}_{-i}) d\theta$ is the leave-one-out predictive density omitting the i th data point. Suppose $\text{elpd}_{\text{loo}}^j$ is the estimated expected log predictive density for the model M_j , then the corresponding weight w_j for model M_j can be

computed as

$$w_j = \frac{\exp(\text{elpd}_{\text{loo}}^j)}{\sum_{j=1}^J \exp(\text{elpd}_{\text{loo}}^j)}. \quad (2.30)$$

These weights are known as Pseudo-BMA (Bayesian model averaging) weights. The Pseudo-BMA weights can provide an assessment of which model is most predictive from a set of models. Bayesian bootstrap methods can compute uncertainties for the leave-one-out estimates based on the n independent components from different data points [148], which leads to adjusted weights known as Pseudo-BMA+ weights.

2.3.7 Bayesian model averaging

Rather than drawing conclusions based on what has been assessed to be the ‘best’ model, more robust inferences can be drawn by considering several models or all of a set of models. Better predictive ability can be provided by averaging over *all* models in a set of models than for any single individual model from that set [152]. The resulting posterior distribution can be computed for a set of models $\mathcal{M} = \{M_1, \dots, M_J\}$ in the form

$$p(\theta|\mathbf{y}) = \sum_{j=1}^J p(\theta|\mathbf{y}, M_j) p(M_j|\mathbf{y}). \quad (2.31)$$

This is a weighted average of posterior distributions conditional on each of the models weighted by the posterior model probability. Model averaging in this way can help to find the optimal model configuration in the space spanned by all the individual models [148]. Bayesian model averaging provides a convenient way of accounting for uncertainty in the structure of a model, particularly when the aim is prediction.

In the setting where the true model is not contained in the set of models being compared, Bayesian model averaging will select the single model closest in KL divergence to the true model from the set of available models. This may be undesirable in the sense that the model misspecification has been concealed and may not be

immediately evident. We note also that in Equation (2.31), there is a dependence on $p(M_j|\mathbf{y}) \propto p(\mathbf{y}|M_j)$, where $p(\mathbf{y}|M_j)$ is the evidence or marginal likelihood for model M_j . The evidence is sensitive to the prior distribution, $p(\theta|M_j)$, for model M_j in undesirable ways, similar to the discussion for Bayes factors in Section 2.3.2.

2.3.8 Stacking

Recent work has proposed using stacking of predictive distributions to combine predictive distributions from different models together and aid in model comparison [148]. This approach is based on previous work on averaging of point estimates for multiple models [153, 154]. This technique can help avoid some of the criticisms of Bayesian model averaging and other model comparison methods.

In the stacking method of Yao et al. [148], the log score of the combined predictive distribution is maximized subject to the weights combining predictive distributions for each model being positive, normalized weights:

$$\max_{\mathbf{w}} \frac{1}{n} \sum_{i=1}^n \log \left(\sum_{j=1}^J w_j p(y_i|\mathbf{y}, M_j) \right) \quad \text{subject to } w_j \geq 0, \sum_{j=1}^J w_j = 1, \quad (2.32)$$

where w_k are the stacking weights for each model. Asymptotically, rather than concentrating on a single model closest to the true model on KL divergence, the stacking weights method will find the optimal predictive distribution within the convex set formed from the span of models. This may be more informative in the case where the true model is not contained in the set of models considered.

2.3.9 So, can I fit an elephant?

Models should have the ‘right’ amount of complexity, often measured by the number of free parameters, to fit to data, and no more. This concept of the simplicity of models is often related to Occam’s razor, but is highlighted well by the following quotation attributed to Von Neumann: “With four parameters I can fit an elephant,

and with five I can make him wiggle his trunk” [155]. The question of how many parameters are needed to fit an elephant has been studied [156], with recent results showing that a single parameter is enough using properties of the logistic map [157], thus suggesting that simply counting parameters is not a complete picture of model complexity. A variety of complex behaviours can be exhibited by a model with a small number of parameters, while for instance it may be possible to reduce a differential equation model with many parameters to a simpler model with fewer parameters since most of the behaviour of the original model lies only on a manifold. Model comparison comes in as a way of selecting a model that fits the observed data well and will also generalize well to new observations.

Selecting a model to meet these criteria can fail in one of two main ways. Overfitting of a model occurs when a flexible model is fitted to data without sufficient constraints. Constraints can be the complexity of the model itself, or some form of regularization. In a Bayesian context, the prior distribution can provide regularization through restricting to biologically plausible parameter regimes. In contrast, underfitting occurs when a model is not sufficiently flexible to explain the observed data due to model misspecification of some form.

This perspective on failures of overfitting or underfitting is usually associated with statistical machine learning models rather than mechanistic models. However, there is much that can be gained through sharing a synergy of ideas between these fields [158]. Mechanistic models hold the advantage that they constrain the class of models under consideration (which helps to avoid overfitting) and gain interpretability into biological mechanisms.

Chapter 3

An interdisciplinary approach to
elucidate the regulation of mRNA
transport in *Drosophila* oogenesis

3.1 Regulation of RNA controls expression of protein in space and time

During development, it is crucial for a wide variety of cells to direct proteins to particular locations so that they can carry out their functions [159]. Since it is challenging to move large proteins around the crowded environment of the cell, a better option can be to localize the mRNA and enable protein synthesis in place. Transporting mRNA rather than the proteins themselves can be more efficient and allow cells to respond quickly to stimuli, since a single mRNA can give rise to many proteins locally once activated, which enables precise spatial and temporal control. Intracellular mRNA localization is a commonly used mechanism across a range of cell types, particularly for large cells such as neurons and oocytes [160–163], although this ubiquitous process is not limited to large cells as it has also been observed in chicken fibroblasts, yeast and bacteria [164–166]. Controlling spatial and temporal expression of protein is important for a wide variety of processes in developing cells [167]. It remains unclear however, how mRNA localization is controlled and how the timescales involved in mRNA localization are regulated.

Identifying bottlenecks in a cellular process indicates key targets for regulation by the cell. However, in many cases, these rate-limiting steps are neither identified nor well understood. In many cases, mRNA localization by molecular-motor-driven transport is crucial for cell polarization, but the rate-limiting processes underlying the localization processes are not fully understood. We begin to address this question in this chapter.

To make progress, we employ an interdisciplinary approach to examine the rate-limiting steps in the localization of *gurken/TGF- α* mRNA in *Drosophila* egg chambers. First, we construct a fine-grained stochastic model of the transport of mRNA within a single nurse cell based on the current understanding of the mechanisms of transport. This computational model allows us to examine how the

timescale for transport of mRNA within a single cell depends on biophysical parameters by performing a sensitivity analysis. Using approximate Bayesian inference, we relate the fine-grained model to quantitative imaging data obtained specifically for this study. By considering different summary statistics extracted from the data, we gain differing amounts of information about different parameters. Informed by the fine-grained model and inference on synthetic data, we suggest future experiments to make best use of combined information from different sources of data. Finally, we make a testable prediction from the fine-grained model about the timescale for mRNA transport in a Dynein mutant with a reduced speed of molecular motor.

The research in this chapter motivates further investigation in three separate areas: the velocity jump process modelling framework for the fine-grained model, the Bayesian inference framework for fitting the model to data, and an interdisciplinary experimental collaboration to investigate hypotheses about the regulation of mRNA localization with a coarse-grained model. Subsequent chapters of this thesis are devoted to these investigations. In this chapter, we will begin by providing background material on development and localization of mRNA in *Drosophila*, before describing the fine-grained stochastic model, and how we relate this to experimental data to investigate the rate-limiting steps and processes controlling mRNA localization.

3.2 *Drosophila* development

Drosophila melanogaster is a commonly used and extremely widely studied model organism [168] since it allows easy use of genetic manipulations via technologies such as gene editing with CRISPR [169] and subcellular imaging [170]. In early *Drosophila* development, mRNA localization plays a key role in determination of the body axes. During oogenesis, one of the early stages of *Drosophila* development, maternal mRNA is transported from the nurse cells into the oocyte itself, where it is localized to specific locations (as illustrated in Figure 3.1) in order to target

proteins to regions where they can function. Polarity of the body axes is established by localizing *gurken* [171], *bicoid* [172], *oskar* [173] and *nanos* [174] mRNA within the developing oocyte. *Drosophila* rely heavily on the asymmetric localization of mRNAs to coordinate early development processes both spatially and temporally.

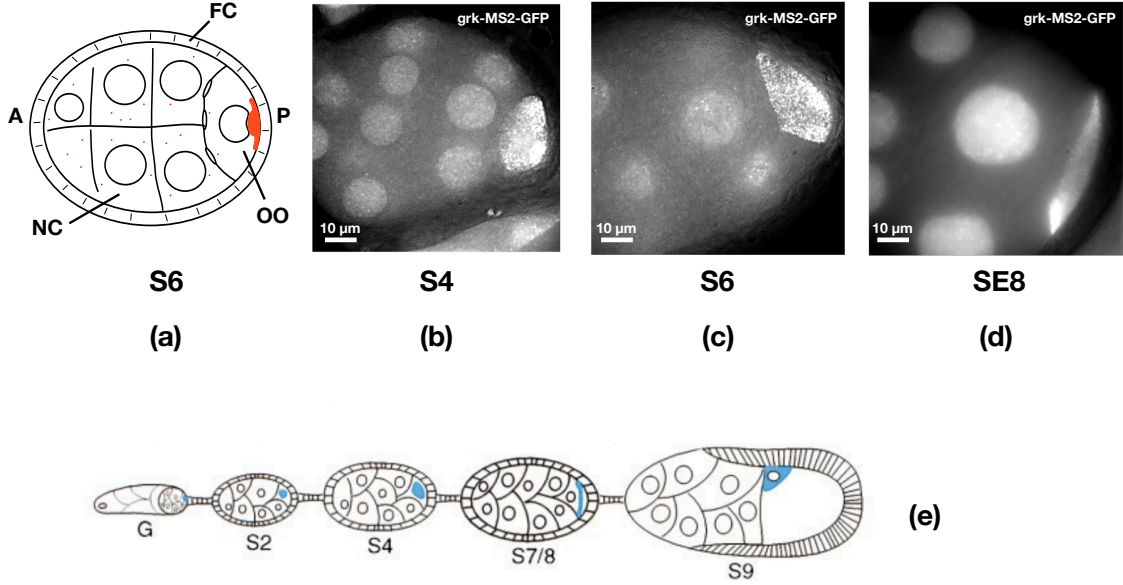


Figure 3.1. Distribution of *gurken* mRNA in *Drosophila* egg chambers. A schematic of a *Drosophila* egg chamber at stage 6 of development is shown in (a). The oocyte, marked as OO, is surrounded by 15 nurse cells, indicated as NC. The tissue is enclosed by an epithelial layer of follicle cells labelled as FC. *gurken* mRNA is localized to the posterior (P) pole of the oocyte, with the anterior of the egg chamber labelled as A. *gurken* complexes in transit are shown as small red dots; localized *gurken* is visible anchored at the posterior pole of the oocyte as a larger red region. In (b), (c) and (d), the distribution of *gurken* mRNA is shown at different stages of development. The nurse cells grow in size between these stages and the distribution of localized *gurken* mRNA also varies at different stages of development. In (e), several egg chambers at different stages of development are shown together, organized linearly in an ovariole, which acts as an assembly line of developing egg chambers [175]. The differences in *gurken* localization at different stages of development are further displayed in (e). (e) is adapted from Lasko [176]. Note that S is an abbreviation for stage; G indicates the germarium, which consists of germ line cells.

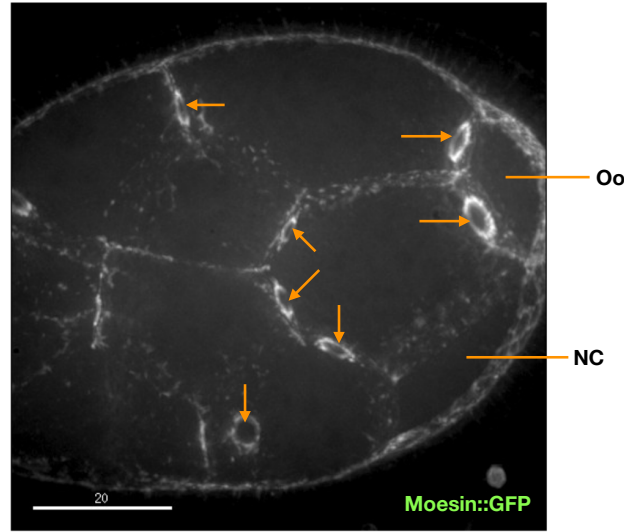


Figure 3.2. Ring canals form cytoplasmic bridges between nurse cells and the oocyte. These contractile rings between cells can be clearly seen here as small ellipses. An egg chamber at stage 5 of oogenesis is shown with F-actin marked with Moesin-GFP. To highlight the ring canals, the image has been projected in the z direction over a partial volume in z of depth $5\mu\text{m}$. The scale bar is $20\mu\text{m}$. Ring canals are highlighted by orange arrowheads. The oocyte is labelled as Oo and NC indicates a nurse cell.

3.2.1 *Drosophila* oogenesis

Oogenesis is the process of growth and development of the primary egg cell (or oocyte) and supporting tissue. In *Drosophila*, this begins with the asymmetric division of a germline stem cell into another germline stem cell and a cytoblast [177]. This is followed by further rounds of division until, after four rounds of division, 16 cells have formed into a cluster of interconnected germline cells known as a cyst, shown in Figure 3.1a). One of these cells will become the oocyte, while the other 15 maternal germline cells are known as nurse cells [178] whose main role is to provide nutrients for the oocyte [179]. Maternal mRNA is produced in the 15 nurse cells neighbouring the oocyte [180]; it is then packaged into complexes with various proteins [181] and transported into the oocyte [182].

Cytoplasmic bridges remain connecting the nurse cells and oocyte due to incomplete cellular divisions [183, 184]. The connections between cells are known as ring

canals, and these are made of cytoskeleton proteins such as actin forming structures of 1 to 2 μm in diameter [185]. The ring canals between nurse cells and the oocyte can be clearly seen in Figure 3.2 as small ellipses in the cell membranes. Figure 3.2 shows a maximum projection of several z slices through a *Drosophila* egg chamber with Moesin GFP marking F-actin on the cell membranes. The ring canals result in the 15 nurse cells and the oocyte being connected in a characteristic pattern [186]. A schematic of the early *Drosophila* egg chamber is shown in Figure 3.3, together with a microscopy image showing the ring canals (highlighted with an actin (phalloidin) marker), and a diagram illustrating the characteristic connections between nurse cells.

The egg chamber undergoes development over a timescale of hours to days, which is divided into 14 morphologically distinct stages [43, 178]. Transport of RNA and other cellular components must be coordinated within and between nurse cells through the ring canals to localize in the oocyte at the right time. In this thesis, we focus on early to mid stages of development, up to stage 8, and note that *gurken* mRNA is expressed and localized at these stages. After late stage 10, the remaining contents of the nurse cells are violently dumped into the oocyte [43], involving distinctly different dynamics of transport.

3.2.2 Dynamic cytoskeleton forms part of the subcellular environment

The environment within a cell is crowded and filled with a variety of cellular components. The cytoskeleton, consisting of tubular polymers known as microtubules and actin filaments, is a structure that provides shape and organization to the cell. The microtubule network has been shown to have a key role in the localization of mRNAs such as *gurken* [159, 187]. Figure 3.4 shows the microtubule network in the nurse cells of a stage 5 *Drosophila* egg chamber. Although the microtubule cytoskeleton was initially thought to be a stable static network, due to how it was analysed in

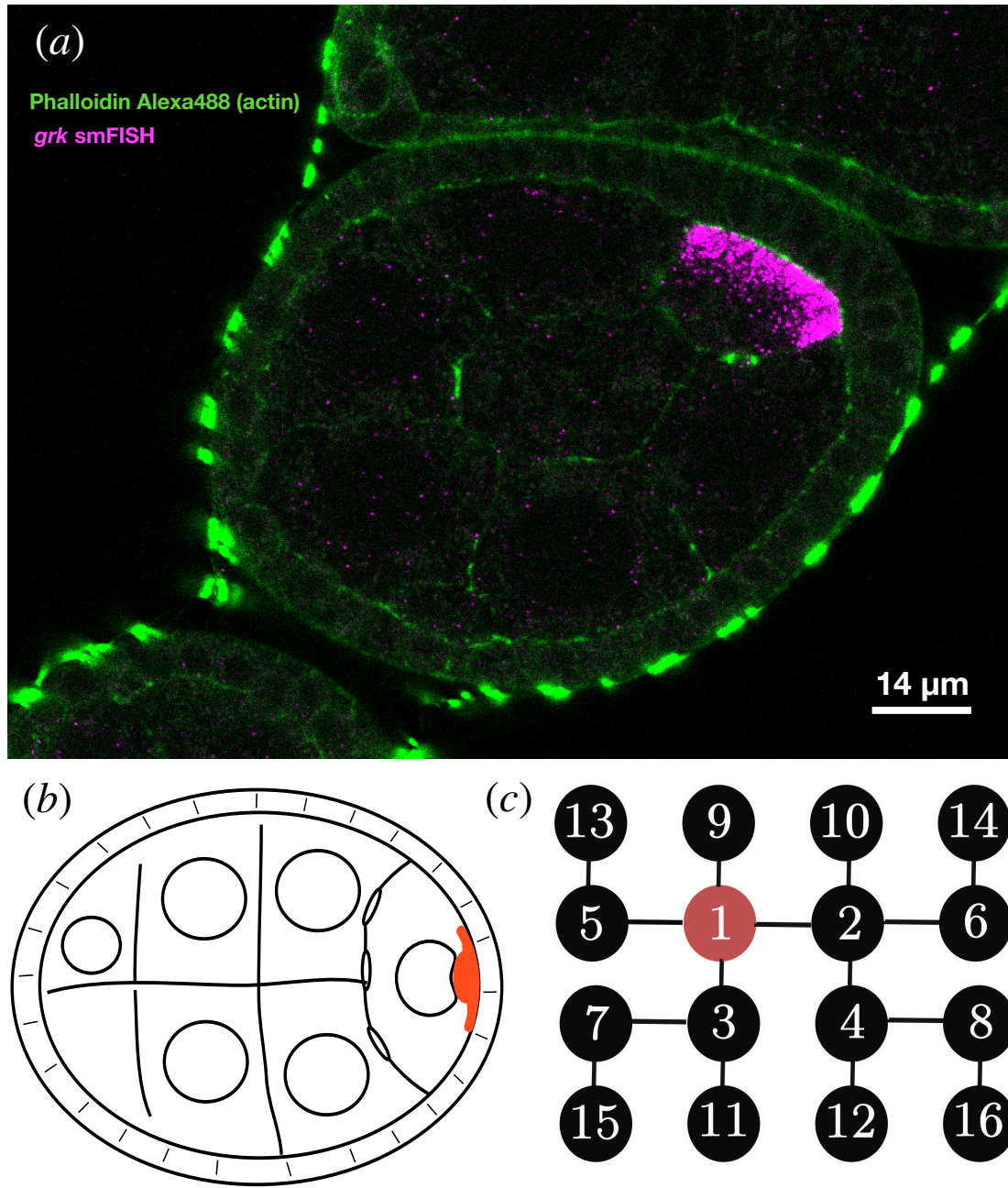


Figure 3.3. Stage 6 *Drosophila* egg chamber. (a) A microscopy image of a fixed egg chamber labelled in green with phalloidin-alexa488 (actin) and in magenta with *grk* smFISH probes (stellaris, Cal Fluor 590). Localization of *gurken* RNA is observed at the posterior of the oocyte. Ring canals providing cytoplasmic bridges connecting between nurse cells and to the oocyte can be seen highlighted by the actin marker. (b) Schematic of the early egg chamber. (c) Diagram of the connections between nurse cells and the oocyte via ring canals. The oocyte (red) is labeled 1 and the nurse cells are labeled 2 to 16.

fixed material, it has been revealed to be a dynamic network with a biased random orientation of microtubules [188]. This underlying structure had been suggested by the work of Zimyanin et al. [189] who observed complexes containing *osk* mRNA moving in a biased random walk.

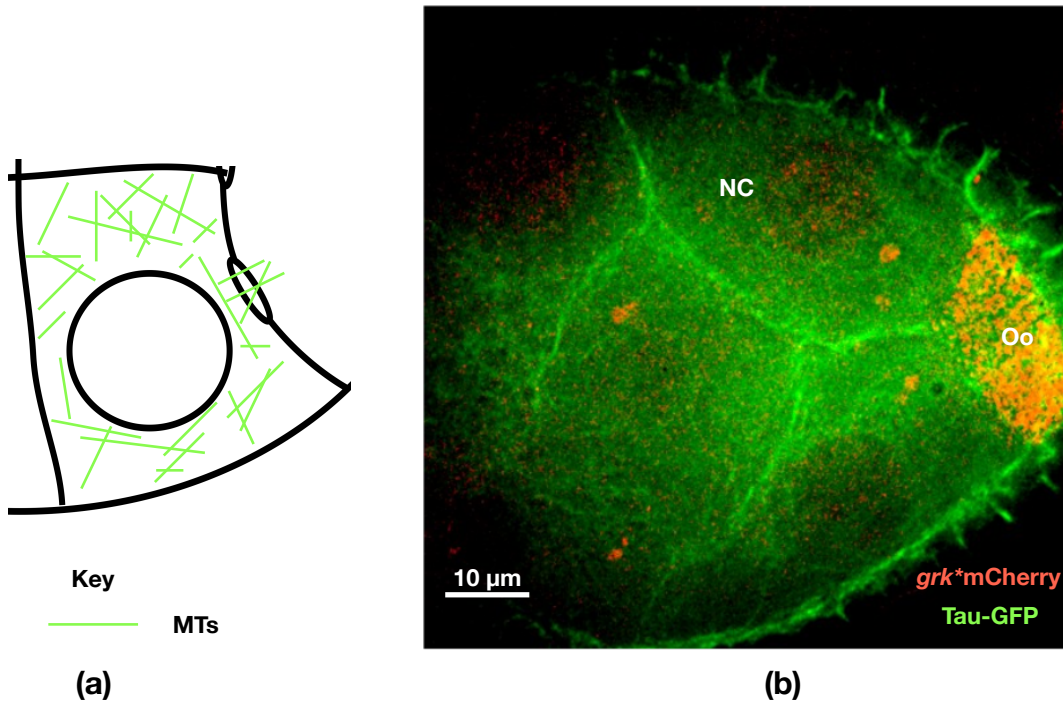


Figure 3.4. The distribution of microtubules in the nurse cells of a stage 5 *Drosophila* egg chamber is shown with a schematic diagram in (a) and via Tau-GFP in (b). In (a) microtubules are shown as short green line segments. A population of microtubules passes directly through ring canals at early stages of oogenesis, but the cytoskeleton organization is rearranged during mid-oogenesis. In (b), an image of the dynamic distribution of microtubules is shown at a single point in time. Microtubules are tagged via Tau-GFP and are visible as short green line segments. The label Oo highlights the oocyte and NC indicates a nurse cell.

3.2.3 mRNA localization in *Drosophila*

Once mRNA has been transcribed from DNA in the nucleus of the nurse cells, the process by which it reaches its final site of localization, where it is then translated into protein, can be divided into four main stages [44]: particle assembly; nuclear export; transport; and anchoring.

Assembly of particles

mRNA does not exist in isolation inside the cell. Instead the mRNA molecules bind to proteins to form particle complexes, which are also known as ribonucleoproteins (RNPs). These proteins perform a range of different functions, including translational regulation, preventing translation while the mRNA is in transit, and may determine the final destination of the particle [190]. There is also evidence that RNPs are dynamically remodelled during the transport process, to support translational regulation [191].

Nuclear export of particles

After the mRNA has assembled into RNP complexes, the RNP complexes must diffuse through crowded interchromatin spaces in the nucleus until they reach a nuclear pore complex (NPC) [192]. The RNP is then exported out of the NPC into the cytoplasm via interactions with co-factors. Certain NPCs are more active than others at different times [191], but the overall direction of export out of the nucleus is unbiased [162].

Transport of particles

The transport stage of mRNA localization occurs by different mechanisms for different mRNAs. The most common method observed is active transport via molecular motors moving along the dynamic cytoskeleton (either microtubules or actin filaments). However, *nanos* mRNA localizes using a diffusion and trapping technique [193] rather than by active transport. Active transport results in fast, directed motion with velocities of the order of $1\mu\text{ms}^{-1}$ [189, 194], which is an order of magnitude faster than movement by free diffusion in the cell.

Individual types of molecular motors are thought to move generally unidirectionally, with Dynein directed towards the minus end of microtubules and Kinesin directed towards the plus end [195]. Dynein-mediated transport on microtubules is a

common mechanism for mRNA localization [162]. Bidirectionality was observed by Vendra et al. [196] in the motion of mRNA complexes on microtubules in *Drosophila* blastoderm embryos. The number of motors required for a given mRNA complex can vary and may be governed by cis-acting localization elements in the mRNA [197]. Possible explanations for the bidirectionality include: the bidirectional disorganized distribution of the microtubule network; a tug of war between different molecular motors moving in opposing directions on different microtubules; a tug of war between different motor species all bound to the same mRNA complex; and reversal of a molecular motor moving on a single microtubule, possibly due to regulation by microtubule associated proteins (MAPs) [45]. The molecular motors must be correctly joined to their cargo and the motor cargo complex secured to the cytoskeleton. This is ensured by the linkers *Bicaudal D* (*BicD*) and *Egalitarian* (*Egl*) [44].

Anchoring of particles

Once the mRNA complex has reached its destination, these complexes must be maintained in position by an anchoring mechanism to keep the mRNA localized within the cell, as for the population of mRNA localized at the posterior of the oocyte in Figure 3.1. *grk* mRNA is anchored by the molecular motor Dynein [198] in the *Drosophila* oocyte and *nos* is trapped by actin [193]. Dynein was observed by Delanoue and Davis [198] to act as a static motor without requiring further energy in the form of ATP to function. An alternative hypothesis is that continuous active transport is required to ensure the localization of mRNA, since Weil et al. [194] found that Dynein motor activity is required to ensure *bcd* localization in the anterior cortex of the *Drosophila* embryo. The transport process may ensure that mRNA complexes are kept near their targeted region of localization by directing them back if they move away.

3.2.4 Advanced imaging tools shed light on mRNA localization

Recent advances in imaging techniques and image analysis technologies [170, 199, 200] have allowed progress in understanding of the mechanisms of mRNA localization [193, 201, 202]. The use of single molecule fluorescence *in situ* hybridization (smFISH) [203, 204] enables single molecules of mRNA to be labelled with high levels of sensitivity and specificity. The development of the MS2-MCP system, which consists of a MS2 bacteriophage RNA stem loop bound by MS2 coat protein fusion to a fluorescent protein [44], has provided great benefits for the imaging of live cells *in vivo*. The MS2-MCP system has been used successfully in the visualization of *nos*, *grk*, *bcd* and *osk* mRNAs in *Drosophila* [189, 193, 194, 205]. Although there are limitations to these technologies dependent on cell type, they have permitted an improvement of our understanding of intracellular motility and structure. Both these imaging tools, MS2-MCP and smFISH, are used to enable detailed analysis and quantification of the localization of mRNA in *Drosophila* for the research presented in this thesis. What has become clear is that a variety of different mechanisms are used to ensure localization of mRNAs [44].

3.3 The role of modelling

While we have provided an overview of the current understanding of the processes involved in localizing mRNA in *Drosophila* egg chambers, many of the mechanisms of transport and localization, including the cues and signals governing each stage of development, remain unclear. Computational and mathematical modelling can assist in elucidating aspects of these mechanisms. This can be achieved through formalizing a descriptive ‘word model’ form of a hypothesis into a quantitative framework able to make testable predictions. Modelling further allows more rapid prototyping and investigation of hypotheses at a cheaper cost compared to *in vivo* experiments

which may, for example, require lengthy genetic manipulation to generate mutant phenotypes in order to test a hypothesis.

More specifically, for this model system, modelling aids our understanding by simulating processes that are hard to observe directly. In practice, automated tracking of RNA-protein complexes is extremely challenging due to the spatial and temporal scales involved. Complexes frequently move out of the field of vision during imaging making both automated and manual tracking challenging, although some attempts at this hard problem have been made [206] and identifying particles can be simpler in fixed material [207]. Rather than directly observing full trajectories of complexes, constructing models of this behaviour based on our current understanding can reveal new aspects of the mechanisms involved.

A modelling description of the full *Drosophila* egg chamber capturing all the microscopic detail of mRNA transport and localization may seem initially to be desirable for querying biological hypotheses. Developing such a detailed model of multiple connected cells is challenging computationally and from a model development point of view. Few previous modelling studies in this area have considered multiple cells, although detailed models of mRNA localization in the oocyte itself have been developed [208–210]. To simplify to a tractable problem, we focus here on a fine-grained model of transport in a single nurse cell. This framework can be extended in future to encompass the full tissue of multiple cells and enable us to probe questions on larger spatial scales once we have a solid understanding of the regulation and rate-limiting processes of mRNA transport and localization within a single nurse cell.

3.3.1 Previous modelling approaches to *Drosophila* development

Although mathematical and computational modelling approaches have become increasingly widespread in biology [211, 212], such modelling approaches are more

mature in certain application areas than others. Much work has focused on morphogen gradient formation during *Drosophila* embryogenesis to quantify the spatio-temporal dynamics [213, 214]. For example, Liu and Niranjana [215] investigated competing hypotheses about *bicoid* mRNA stability regulation via differential equation and stochastic models, concluding their models support regulated stability of mRNA. Steps have been made towards quantitative analysis of genetics experiments through modelling [216] and through improved imaging. Modelling approaches in oogenesis have received rather less attention.

At mid to late stages of oogenesis, Kinesin-dependent cytoplasmic streaming becomes important in driving transport dynamics in the oocyte. By using an advection-diffusion model with noise in the local mean velocity of complexes represented via a correlated random field, Ganguly et al. [208] show that these cytoplasmic streaming dynamics reflect the microtubule cytoskeleton architecture and may aid in localization of complexes. Szymańska et al. [217] construct a generic model of intracellular mRNA and protein dynamics, including active transport on microtubules, showing that oscillations in protein concentrations are possible since the regulation of protein synthesis acts as a negative feedback. Trong et al. [209] have demonstrated how localization of mRNA can be achieved in a *Drosophila* oocyte using a partial differential equation (PDE) description of mRNA dynamics and a stochastic model of cytoskeleton dynamics. Ciocanel et al. [210] used a PDE model of RNA localization in *Xenopus* oocytes to quantify aspects of the transport process using FRAP (fluorescence recovery after photobleaching) data, and others have considered models of RNA transport to examine the behaviour of individual molecular motors [195, 218], or the cytoskeleton structure [219]. Here, we will describe the transport of mRNA in the nurse cells surrounding the oocyte using an individual-based stochastic description of mRNA complexes. This microscopic, individual-based description is more appropriate than continuous densities of mRNA at low copy numbers.

3.3.2 Summary

In this chapter, we will address the question of how mRNA localization is controlled within a single nurse cell and what the rate-limiting steps in this aspect of the mRNA localization process are. Specifically, we ask here whether transport of mRNA is a rate-limiting bottleneck during mRNA localization. Any rate-limiting bottleneck of a process will play a key role in its regulation since a small change at this point can have a large downstream effect. For this reason, study of such bottlenecks can provide a focus for future studies of the biological mechanisms governing localization of mRNA and highlight potential targets for synthetic control of this process. We construct a fine-grained model of *gurken* RNA transport and localization within a single nurse cell. We build a stochastic model for movement of individual mRNA-protein complexes from export at a NPC to the ring canal. To relate this model to experimental data, we use approximate Bayesian methods which allow us to identify the parameters of the model. The parameterized model enables us to make predictions about summary statistics, such as the first passage time for mRNA complexes to move from nucleus to ring canal, which agree with other previously published data on the timescales of RNA transport. Additionally, we are able to make a testable prediction about the behaviour of a genetic mutant with a reduced speed Dynein motor.

3.4 Fine-grained stochastic model of mRNA transport within a nurse cell

To investigate the regulation of mRNA localization and transport, we develop a fine-grained stochastic model of the transport of mRNA complexes. In particular, this model allows us to investigate the transport of *gurken* mRNA within and between nurse cells. We focus on a single nurse cell and study the regulation of transport of mRNA from export at the NPC to transit through a ring canal into the oocyte.

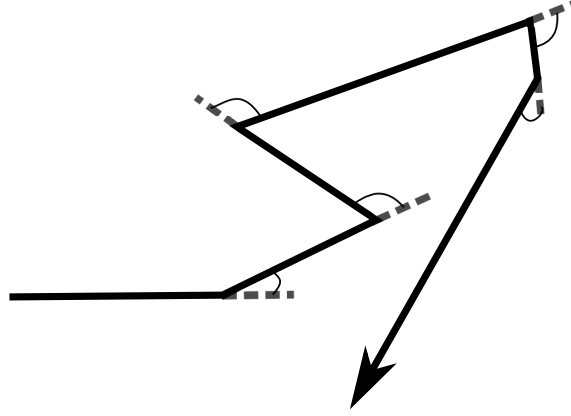


Figure 3.5. Typical behaviour of VJP model showing reorientations and a corresponding change of direction. Biologically these reorientations would correspond to mRNA complexes falling off microtubules, reattaching, and moving off in a different direction.

By using the model, we can study the role of molecular motors versus that of the biased cytoskeleton network in controlling the timescale for transport, which is hard to tease apart experimentally. We model a nurse cell neighbouring the oocyte with a single ring canal into the oocyte, and will later discuss generalizations of this setup. The timescale for transport from the nucleus to the ring canal is important as this characterizes transport between cells.

The movement of mRNA complexes is often described as a biased random walk [189]. Rather than modelling this transport process as a simple random walk in position, we let the velocity of motion vary. This type of model is known as a velocity jump process (VJP) and has been successfully applied to directed migration of animals and cells [220–222].

The VJP model involves fast directed motion in a given direction followed by switching events, at which an mRNA complex will reorientate and subsequently move in a different direction (see Figure 3.5). This process continues until the mRNA complex reaches a small target, the ring canal. The VJP is a good choice of model here because it represents well the mechanistic movement process via active transport on microtubules. Trajectories from a VJP can be highly correlated and persistent, similar to observed trajectories of mRNA complexes.

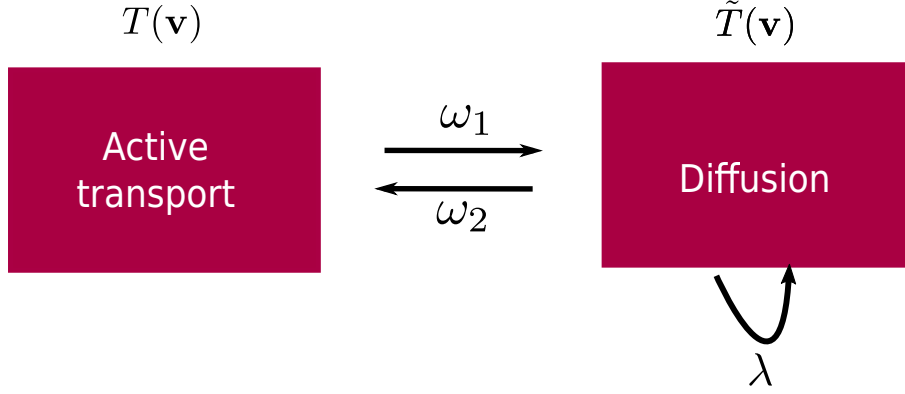


Figure 3.6. States and transitions between the states for mRNA complexes moving according to the fine-grained VJP model. The speeds in active transport and diffusion are given by ν_1 and ν_2 , respectively. Transition rates between states are given by ω_1 , ω_2 and λ as shown.

There are two phases of motion included in the model: an active transport phase when the mRNA complex is attached to a molecular motor and is moving on the microtubules with constant speed ν_1 ; and a slower diffusive phase with constant speed ν_2 . In the diffusive phase, the complex is detached from the microtubules and moves by diffusion in the cytoplasm until it reaches another microtubule. Switching occurs between these two phases of motion with exponential waiting times¹ between switching events. Biologically, this switching corresponds to the molecular motor falling off and reattaching to the microtubule network. As illustrated in Figure 3.6, switching occurs from the active transport state to diffusion with rate ω_1 . From the diffusive phase, switching events within the same phase occur with rate λ , providing a representation of diffusion between microtubules via frequent movements and reorientations, and switching to the active transport phase takes place at rate ω_2 .

After each switching event from diffusion to the active transport state, the mRNA complex changes direction (see Figure 3.5) by picking a new velocity, $\nu_1 \mathbf{v}$, where the direction $\mathbf{v} \in S^2 = \{x \in \mathbb{R}^3 : \|x\| = 1\}$ is drawn from a reorientation kernel $T(\mathbf{v}, \mathbf{v}')$, which in general can depend on the previous direction $\mathbf{v}'$. Following a transition to

¹Exponential distributions for the waiting times are the simplest possible choice for waiting time distributions, which helps keep the model tractable, and ensure the memorylessness property.

the diffusive state, or a reorientation within the diffusive state, the particle travels with velocity $\nu_2 \mathbf{v}$ where the direction $\mathbf{v}$ is chosen via a reorientation kernel $\tilde{T}(\mathbf{v}, \mathbf{v}')$. In this work, we will assume that new directions are chosen independent of the previous direction of motion, $\mathbf{v}'$, since reorientations correspond to mRNA complexes detaching from microtubules or moving by diffusion. Therefore we will instead write $T(\mathbf{v})$ in place of $T(\mathbf{v}, \mathbf{v}')$.

3.4.1 Master equation

We can describe this model via a master equation of the following form [220]:

$$\frac{\partial p_1}{\partial t} + \nu_1 \mathbf{v} \cdot \nabla p_1 = -\omega_1 p_1 + \omega_2 \int_{S^2} T(\mathbf{v}) p_2(x, \mathbf{v}', t) d\mathbf{v}', \quad (3.1)$$

$$\begin{aligned} \frac{\partial p_2}{\partial t} + \nu_2 \mathbf{v} \cdot \nabla p_2 = & -\lambda p_2 + \lambda \int_{S^2} \tilde{T}(\mathbf{v}) p_2(x, \mathbf{v}', t) d\mathbf{v}' \\ & - \omega_2 p_2 + \omega_1 \int_{S^2} \tilde{T}(\mathbf{v}) p_1(x, \mathbf{v}', t) d\mathbf{v}', \end{aligned} \quad (3.2)$$

where $p_j(\mathbf{x}, \mathbf{v}, t)$, $j = 1, 2$, is the probability density of a complex being at position $\mathbf{x} \in \mathbb{R}^3$ moving in a direction $\mathbf{v} \in S^2$ at time t in the active transport state, or in the diffusive state, respectively; $\nu_j > 0$, $j = 1, 2$, are the speeds in active transport and diffusion; $\lambda > 0$ and $\omega_j > 0$, $j = 1, 2$, are transition rates between states, as shown in Figure 3.6; $T(\mathbf{v})$ and $\tilde{T}(\mathbf{v})$ are reorientation kernels. The left hand side of Equation (3.1) represents an advection term in $(\mathbf{x}, \mathbf{v})$. The integral terms on the right hand side are due to mRNA complexes arriving in a given state of motion due to reorientations. The negative terms on the right hand side are due to complexes leaving a given state to transition to a new state of motion.

Due to the integral terms, as well as the challenging geometry and boundary conditions, it is difficult to gain insight from this master equation by attempting to solve directly or integrate numerically. Nonetheless, we can simulate trajectories from the VJP model and use Monte Carlo methods to gain insight into the variability of the process and any bottlenecks affecting the regulation of mRNA localization.

3.4.2 Geometry

We use a realistic three-dimensional (3D) geometry for the fine-grained stochastic model to enable us to study the regulation of mRNA transport within a single nurse cell. We model the nurse cell as an ellipsoid

$$\Omega = \left\{ \mathbf{x} = (x, y, z) : \frac{x^2}{L_x^2} + \frac{y^2}{L_y^2} + \frac{z^2}{L_z^2} = 1 \right\},$$

with the x axis aligned along the anterior-posterior axis of the oocyte, where L_x, L_y, L_z describe typical length scales in each dimension. The ellipsoid geometry of the system is shown in Figure 3.7. Within the nurse cell, the cell nucleus is modelled as a spherical region of given radius, R , placed centrally in the cell at the origin. The nucleus is excluding to the mRNA complexes in the model. The remainder of the nurse cell is accessible to the complexes, which seems reasonable since experimentally mRNA complexes are observed in all parts of the nurse cell. Initially, mRNA complexes are placed randomly at time $t = 0$ on the circumference of the nucleus to represent their export from NPCs. Complexes move randomly, as described by Equation (3.1), until eventually they hit a small target at the posterior end of the nurse cell representing the ring canal joining to the oocyte, where they are absorbed and removed from the system. mRNA complexes are assumed to be non-interacting; since the volume of a complex is much smaller than the volume of the cell it is reasonable to neglect volume exclusion of complexes in this case. Throughout all simulations in this work, we fix $L_x = 26\mu\text{m}$, $L_y = L_z = 18.5\mu\text{m}$, $R = 10\mu\text{m}$ and the diameter of the ring canal as $2\mu\text{m}$ based on measurements performed on nurse cells in egg chambers at stage 5 of oogenesis using FIJI (ImageJ V1.51d; *fiji.sc*, [223]).

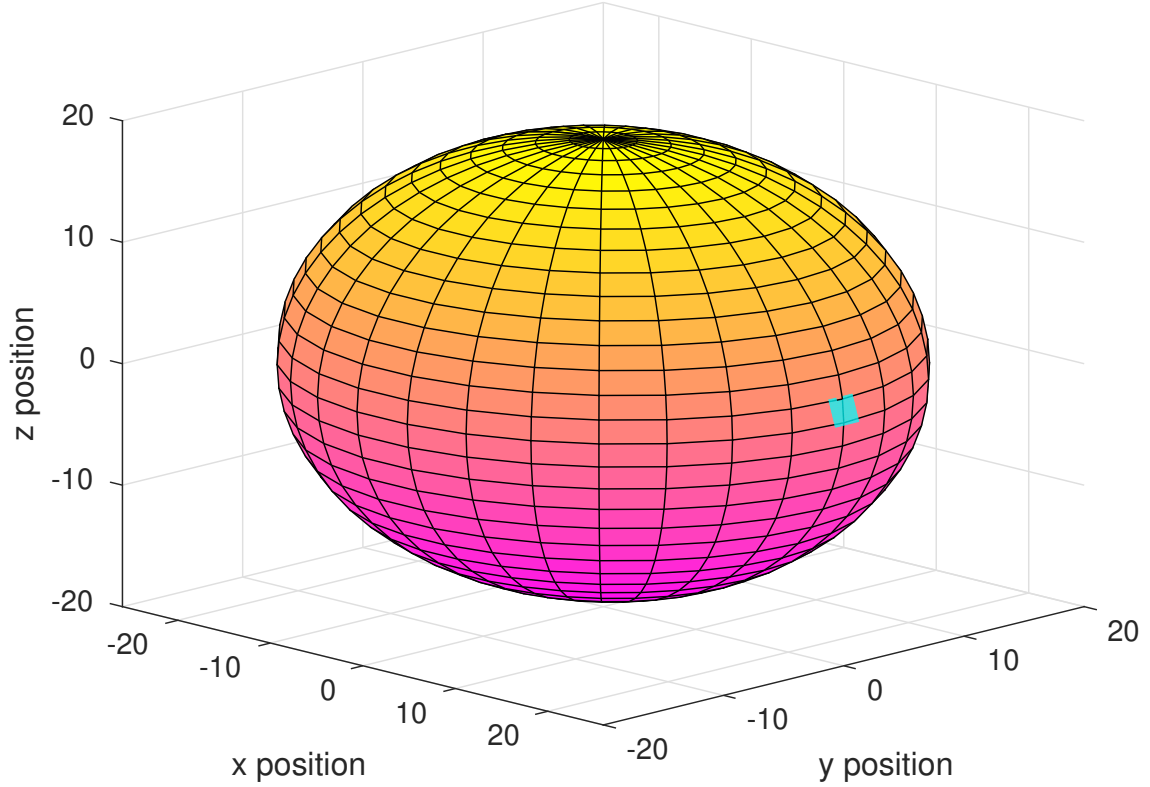


Figure 3.7. The 3D ellipsoid geometry used to represent a nurse cell in the fine-grained VJP model. The ring canal is visible at the posterior pole of the nurse cell shown as a cyan region. The length scales used for the nurse cell geometry are obtained from measurements of experimental data and are as follows: $L_x = 26\mu\text{m}$, $L_y = L_z = 18.5\mu\text{m}$.

3.4.3 Reorientation kernels based on cytoskeleton distribution

Here we choose to define the reorientation kernel, $T(\mathbf{v})$, empirically based on biological data reflecting the observed biased distribution of microtubules [188], as described in Section 3.2.2, since mRNA complexes move in active transport on microtubules. We define a proportion of microtubules, $\phi \in [0, 1]$, aligned in the posterior direction (parallel to the x axis in our 3D model, see Section 3.4.2) and thus take the following for $T(\mathbf{v})$ using spherical coordinates (θ, ψ) :

$$T(\mathbf{v}) = T(\theta, \psi) = \begin{cases} \frac{\phi}{4\pi} & \theta \in [0, \pi] \\ \frac{1-\phi}{4\pi} & \theta \in [\pi, 2\pi), \end{cases} \quad (3.3)$$

where $\theta \in [0, 2\pi]$ is the azimuthal angle and $\psi \in [0, \pi]$ is the polar angle. The resulting distribution of reorientation angles from this reorientation kernel is shown in Figure 3.8. For the reorientation kernel for transitions into the diffusive state,

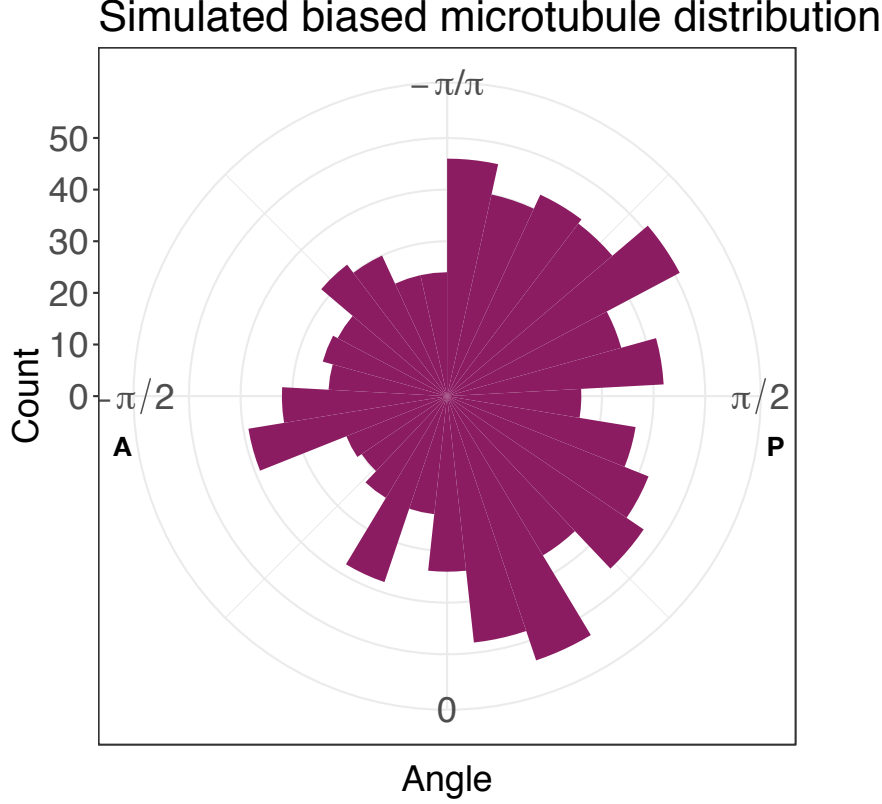


Figure 3.8. The distribution of microtubules in the nurse cells of a stage 5 *Drosophila* egg chamber as represented in the model. A circular histogram of 1000 reorientation angles simulated from the reorientation kernel given in Equation 3.3 with $\phi = 0.58$ is shown, corresponding to the biased microtubule distribution observed in Parton et al. [188]. (Compare to Figure 3.4 which shows an image of the microtubule network at this stage.) Note that A indicates the anterior direction in the egg chamber and P indicates the posterior direction along the anterior-posterior axis.

$\tilde{T}(\mathbf{v})$, we take a uniform distribution on the unit sphere:

$$\tilde{T}(\mathbf{v}) = \frac{1}{4\pi} \text{ for } \mathbf{v} \in S^2. \quad (3.4)$$

3.4.4 Boundary conditions

In addition to specifying the geometry of a nurse cell, an important aspect of the dynamics is how mRNA complexes interact with the cell membrane. Based on observations from live imaging, when a complex approaches the membrane it pauses and reorientates. To represent this in the computational model, we implement a boundary condition such that if a complex has hit the boundary, it stops on the boundary and reorientates so that its next direction of movement is inside the cell. This is a non-standard boundary condition, but best captures the observed biological behaviour. In addition, we have an absorbing condition at the ring canal such that mRNA complexes are removed from simulations upon hitting the ring canal.

3.5 Analysis of the microscopic model

We proceed by analyzing behaviour of the microscopic model. Knowledge of properties of the model can help guide us in relating this model to experimental data and gaining insight from it.

3.5.1 Parameter estimates from direct measurements

The fine-grained VJP model described in Section 3.4 is mechanistic and several parameter values correspond to biophysical quantities that are directly measurable, such as the speed of particles moving in active transport on microtubules. By analysing data from manually tracked *gurken* mRNA complexes in the nurse cells, we are able to obtain direct biophysical measurements of some model parameters and estimate the rate parameters indirectly.

Based on measurements from manually tracked *grk*-GFP mRNA complexes in *Drosophila* nurse cells, we are able to make preliminary estimates of the parameters of our model. All data were obtained from Davidson [224]. Particles are classed as static, paused or active. Paused and active particles will correspond to those in

the diffusive and active transport phases of our model, respectively. Static particles will be assumed to be either anchored or undergoing anomalous behaviour. Using a two-dimensional (2D) imaging slice in a single z plane, the number of total particles counted was $N = 340$ in a $40 \times 40 \mu\text{m}$ area of a nurse cell. The proportions of particles in each of the classes are 25% active, 50% paused and 21% static. Over the timescale of the experiment, particles were not observed to transition between states. Particle speeds were assessed by observing a $10 \times 10 \mu\text{m}$ area for 50 timepoints, with images taken in a single z slice at three frames per second. The average speed in the active transport state ν_1 was $1.16 \pm 0.08 \mu\text{ms}^{-1}$ from $n = 33$ trajectories. The average run length was $2.79 \pm 0.66 \mu\text{m}$. In the paused phase, an average speed, ν_2 , of $0.8 \pm 0.6 \mu\text{ms}^{-1}$ was observed from 58 particles with an average run length of $0.84 \pm 0.06 \mu\text{m}$.

Considering the process as a continuous time Markov chain, with active transport and diffusive states as shown in Figure 3.6, we obtain that the steady state of the continuous time Markov chain is $[\omega_2/(\omega_1 + \omega_2), \omega_1/(\omega_1 + \omega_2)]$. Then, assuming ergodicity and neglecting static particles, this approach gives us $\omega_2/(\omega_1 + \omega_2) = 2/3$ and $\omega_1/(\omega_1 + \omega_2) = 1/3$ by comparing to the proportions of particles observed in each state.

Assuming there are no internal transitions back to the active transport state, then we have $\omega_1 = 0.42\text{s}^{-1}$ and hence $\omega_2 = 0.84\text{s}^{-1}$. We can also deduce that the turning rate $\lambda = 0.95 - 0.84 = 0.11\text{s}^{-1}$. In addition, we assume the distribution of microtubule orientations in the nurse cells is similar to that in the oocyte, since there is data available for the oocyte [188] but not for the nurse cells. Therefore we take the microtubule bias parameter as $\phi = 0.58$. Parameter estimates are summarized in Table 3.1.

3.5.2 Simulation of first passage times

Parameter	Value	Units	Description
ν_1	1.16	μms^{-1}	Speed in active transport
ν_2	0.80	μms^{-1}	Speed in diffusion
ω_1	0.42	s^{-1}	Transition rate from active transport to diffusion
ω_2	0.84	s^{-1}	Transition rate from diffusion to active transport
λ	0.11	s^{-1}	Transition rate within diffusive state
ϕ	0.58	-	Proportion of microtubule bias

Table 3.1. Parameter estimates for the fine-grained VJP model based on direct measurements

The timescale for transport between nurse cells (or from a nurse cell to the oocyte) can be characterized by the timescale for transport from the nucleus to the ring canal of a nurse cell. We estimate this timescale via Monte Carlo simulation methods and demonstrate dependence on biophysical model parameters. Based on simulations from the fine-grained VJP model as described in Section 3.4, we obtain a distribution of first passage times (FPTs), computed across 1024 particles, with a median of 25 mins and a 95% interval of [5, 109] mins. The distribution of FPTs is shown in Figure 3.9. These results agree with experimental observations of Clark et al. [182], who report that 15-20 mins after injection of *gurken* RNA accumulation of RNA was observed around ring canals in 88% of egg chambers ($n = 8$). We plot the evolution of a typical path from the fine-grained VJP model in Figure 3.10. This highlights how movement of mRNA complexes by the mechanisms assumed in the model results in indirect paths that do not take a direct route to the ring canal, as observed experimentally [189].

Given that stage 5 in *Drosophila* oocytes lasts five hours, the timescale for the FPT appears to be an order of magnitude shorter than the timescales over which *grk* mRNA transport is observed in oogenesis. This suggests that transport of a single complex within a single nurse cell from the nucleus to the ring canal is fast, and so not the limiting step in the transport and localization of the mRNAs. Other

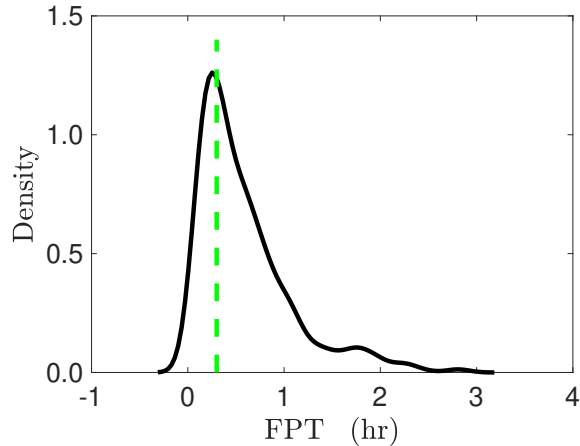


Figure 3.9. The FPT distribution for a mRNA complex to travel from the nucleus of a nurse cell to a ring canal simulated as simulated via the fine-grained VJP model for parameters in Table 3.1 obtained by direct measurements as described in Section 3.5.1. Simulations were performed with 1024 mRNA complexes. The green line shows an experimentally observed timescale of 0.3 hrs at which accumulation near ring canals of RNA was noticeable.

candidates for any bottleneck in mRNA transport and localization (if such a bottleneck is present) include transcription and export of mRNA complexes, availability of molecular motors to transport multiple complexes simultaneously, and transit through ring canals. Experimental observations from Clark et al. [182] show slower speeds of mRNA complexes in ring canals than moving on cytoskeleton in the cytoplasm, for example. These aspects are not captured well by the current model so are hard to assess in this context. Nonetheless, we have been able to rule out transport within a single nurse cell as a bottleneck within the mRNA localization process by demonstrating that this intracellular transport is fast relative to the timescales of oogenesis and localization of *gurken* during oogenesis.

3.5.3 Sensitivity analysis highlights the role of the cytoskeleton

We explore the dependence of the fine-grained VJP model on model parameters by performing a sensitivity analysis, varying each parameter to the extent of biologically plausible parameter regimes. This allows us to study how the transport of

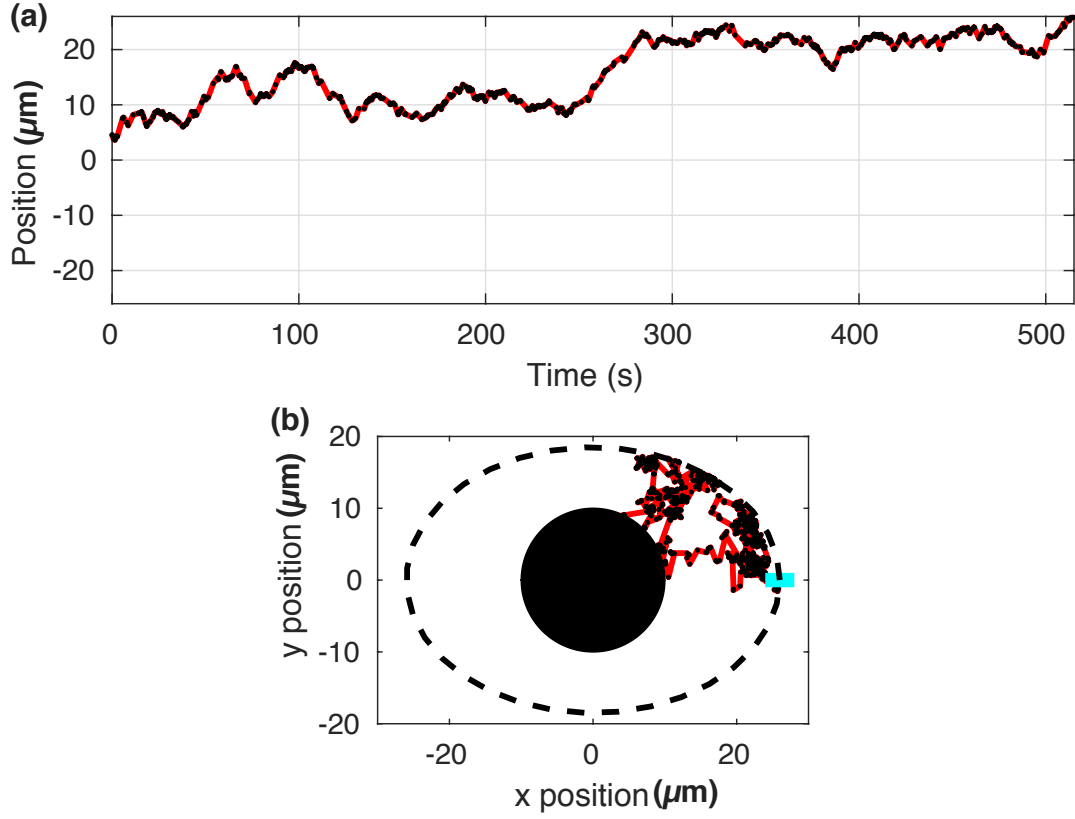


Figure 3.10. The evolution of a typical path from the fine-grained model until absorption at the posterior ring canal, showing also the geometry of the model projected into two dimensions. In (a), the position of an mRNA complex is shown along the x axis (parallel to the anterior-posterior axis) over time. The complex reaches the ring canal after just over 500s in this case. In (b), the path of the mRNA complex is shown in the (x, y) plane until the complex reaches the ring canal. Note that (a) and (b) are two different representations of the same data. Small black circles on the path show switching events between the active transport and diffusion states. The dashed ellipse represents the boundary of the nurse cell, the large filled black circle shows the nurse cell nucleus, and the thick cyan line at the posterior pole represents the ring canal into the oocyte. Model parameters used are: $\nu_1 = 1.16\mu\text{ms}^{-1}$, $\nu_2 = 0.80\mu\text{ms}^{-1}$, $\omega_1 = 0.42\text{s}^{-1}$, $\omega_2 = 0.84\text{s}^{-1}$, $\lambda = 0.11\text{s}^{-1}$, $\phi = 0.58$.

mRNA complexes is regulated. The fine-grained transport model is dependent on six parameters: $\nu_1, \nu_2, \omega_1, \omega_2, \lambda, \phi$. We show in Figure 3.11 the effect of varying key parameters on the FPT from the nucleus to the ring canal into the oocyte. The results of this sensitivity analysis demonstrate that decreasing the speed parameters, ν_1 and ν_2 , or the microtubule bias, ϕ , leads to a sharp rise in the FPT (see Figure

3.11b) and Figure 3.11e)). Travelling slower on microtubules and in the diffusive phase increases the time to reach the ring canal, but moving much faster has a more limited effect on the FPT distribution due to effects of the geometry of the nurse cell.

The transition rates ω_1 and ω_2 have opposite effects with an increase in ω_1 increasing the FPT, and a decrease in ω_2 increasing the FPT (see Figure 3.11c) and Figure 3.11d)). Biologically, ω_1 is the rate of falling off a microtubule to move out of the active transport state, and ω_2 is the rate of attaching onto a microtubule to enter the active transport state. This agrees qualitatively with our intuition in that spending less time in the active transport state should increase the FPT. The turning rate within the diffusive state, λ , has no noticeable effect on the FPT (see Figure 3.11a)), since it only effects movement within the diffusive phase.

The strongest effects on the FPT are due to variation in the parameter ϕ describing the reorientation bias due to the distribution of the cytoskeleton. These effects can be seen in the shift in the distribution of the FPT for values of ϕ close to 0.5 and the elongation of the tail of the distribution. This dependence on ϕ emphasises the role of the cytoskeleton in transport of mRNA complexes within nurse cells; cytoskeleton morphology has been similarly highlighted by other studies [225] focusing on mRNA localization in different biological model organisms. Interestingly, the microtubule bias, ϕ , observed by Parton et al. [188] in the oocyte corresponds to a value $\phi = 0.58$ which is approximately the minimum bias value consistent with the observed FPT data. Together these results offer quantitative insight into how altering the mechanisms of transport would effect the timescales for transport of mRNA within a nurse cell.

We note that although varying many of the parameters results in a stretching of the distribution of FPTs possible from the model, only reducing the speed and increasing the switching rate ω_1 result in a marked shift in the support of the distribution of FPTs. For the remaining model parameters, a wide range of parameters

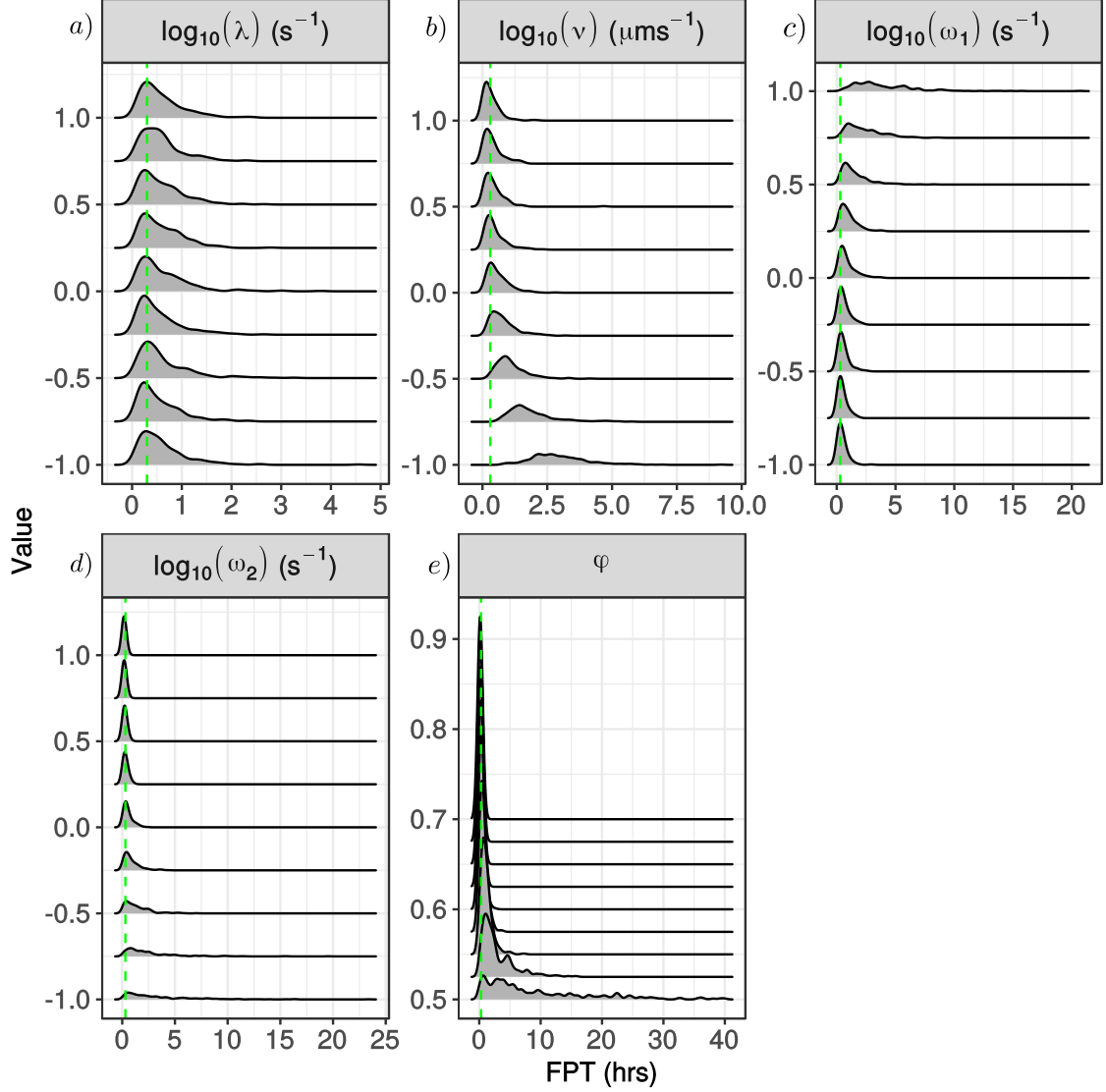


Figure 3.11. Sensitivity analysis of the FPT for mRNA-protein complexes to travel from the nucleus to the ring canal of a nurse cell, as simulated with the fine-grained VJP model. Results show the FPT times for 256 particles as individual parameters are varied to demonstrate the sensitivity to each parameter. Parameters used for these simulations are $\phi = 0.58$, $\nu_1 = 1.16\mu\text{ms}^{-1}$, $\nu_2 = 0.8\mu\text{ms}^{-1}$, $\omega_1 = 0.42\text{s}^{-1}$, $\omega_2 = 0.84\text{s}^{-1}$, $\lambda = 0.11\text{s}^{-1}$, as in Table 3.1. We vary the speed in the diffusion and active transport states together such that $\nu_2 = \nu_1/2$ while varying ν_1 , whereas for the remaining subplots we vary individual parameters $\omega_1, \omega_2, \lambda$, and ϕ . For comparison, the timescale (0.3 hrs) over which Clark et al. [182] report experimental observations of RNA build-up at ring canals is shown as the green dashed line for comparison. Note the different scales on the x axes of each subplot.

are capable of producing FPT data similar to experimentally observed timescales. This may indicate potential identifiability problems as outlined in Section 2.1.7 when using only the FPT as data, and motivates investigation of further summary statistics.

3.5.4 Spatial distribution as a summary statistic

Although studying FPTs provides valuable information about the timescales of mRNA transport, this process is inherently spatial and since spatial data is available via live imaging, it makes sense also to study summary statistics that can be informative about the spatial properties of mRNA transport. We investigate the sensitivity of certain spatial summary statistics to parameters of the fine-grained model. These spatial summary statistics will be informative when we consider inference for the fine-grained model.

In particular, we exhibit the dependence of the spatial distribution of mRNA complexes simulated from the fine-grained model on parameters of the model, as shown in Figure 3.12. The spatial distribution is studied parallel to the anterior-posterior axis (x axis in Figure 3.7), since this direction of transport is of greatest interest biologically. We simulate the movements of mRNA complexes with the fine-grained model up until a fixed time T , and assess the spatial distribution by dividing the domain into bins and recording the number of particles that fall in each bin. Any particles that have already been absorbed at the ring canal are recorded in the final compartment. Since the timescale used here, $T = 0.075 \text{ hr} = 4.5 \text{ min}$, is less than the biological timescale for the FPT described above in Section 3.5.3, we expect that not all the particles will be localized at this time.

For increased motor speeds, ν_1 and ν_2 , the mRNA complexes spread out faster from the nucleus and are directed towards the posterior of the nurse cell. By increasing the microtubule bias, ϕ , the motion of the mRNA complexes becomes much more directed towards the posterior with accumulation near the posterior and ab-

sorption upon reaching the ring canal there. For high values of the switching rate, ω_1 , the distribution of complexes is more spread out and has not reached the posterior, as little time is spent in the active transport state. For low values of ω_1 , the mRNA complexes stay in the active transport state for longer on average, so greater accumulation near the ring canal is seen for intermediate values of ω_1 . We note that, for large values of ω_2 , there is an accumulation of particles near the posterior, but that these are not quickly absorbed, since the mRNA complexes spend little time in the diffusive phase of motion. The diffusive state is not required for absorption in the model, but the size of the small target together with the geometry induces this effect. Changes in the turning rate, λ , have little effect on the spatial distribution of particles in the x direction.

3.6 Bayesian inference methods to link the microscopic model to data

In order to make testable predictions from our model, we aim to relate the fine-grained model directly to experimental data such that we can obtain estimates of uncertainty in our parameter estimates. Testable predictions allow us to verify the model and evaluate new biological hypotheses [5]. We make use of the framework of Bayesian inference described in Chapter 2 and consider fitting the fine-grained model to experimental data directly using ABC. A Bayesian inference framework allows us to make full use of the quantitative image data available to us, and to make quantitative predictions. However, since the fine-grained model relies on a complex geometry and boundary conditions, evaluating the likelihood is expensive, although we are able to perform forward simulation from the model. ABC allows us to draw from an approximate posterior distribution from our model based only on forward simulation, as described in Section 2.2.1. Challenges here in implementing ABC for the fine-grained model stem from the choice of summary statistics and the computa-

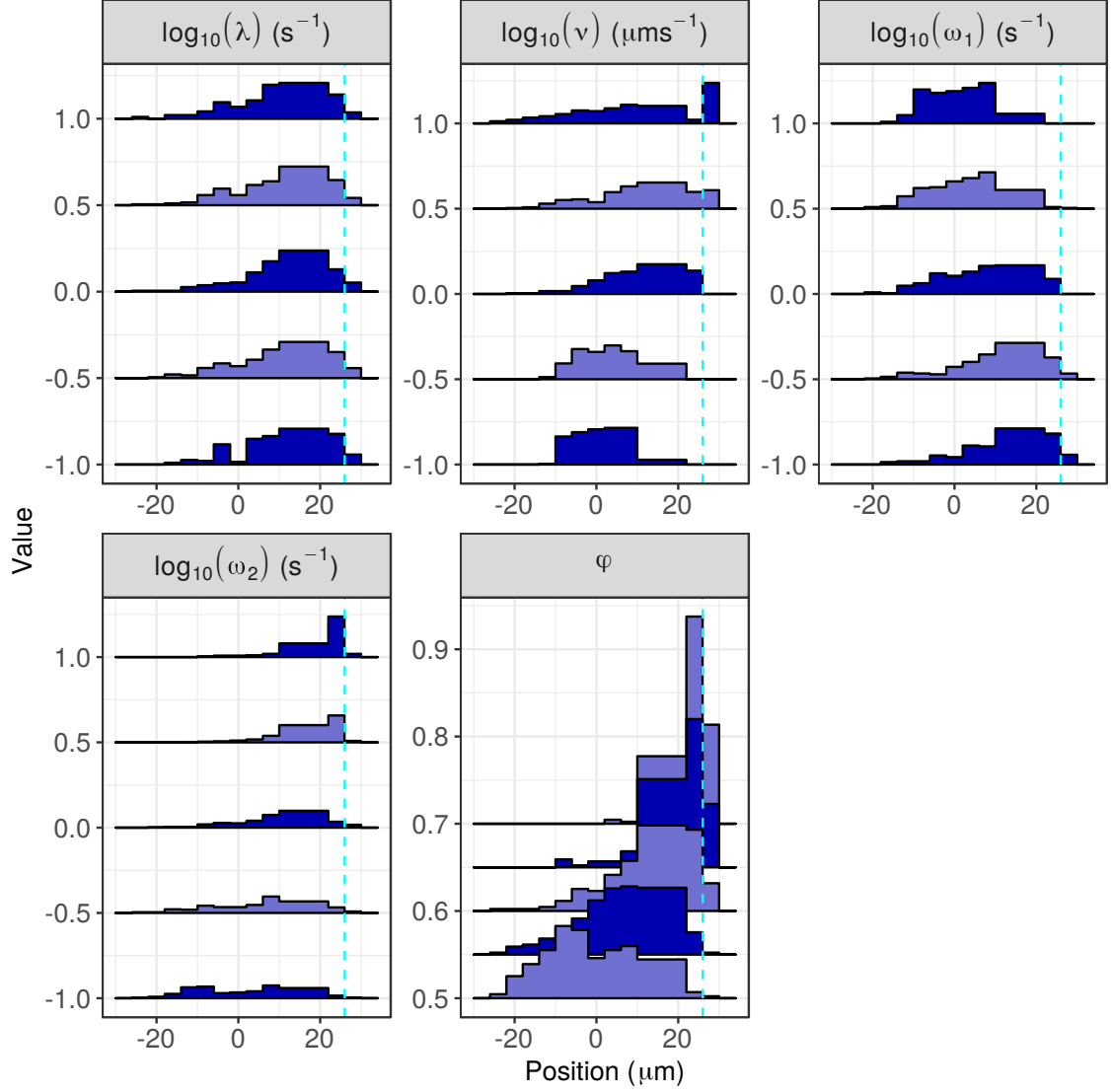


Figure 3.12. Sensitivity analysis of the spatial distribution in the x direction for mRNA complexes simulated with the fine-grained model for parameters as in Table 3.1. As previously, we vary the speed in diffusion and active transport together such that $\nu_2 = \nu_1/2$ while varying ν_1 , whereas for the remaining subplots we vary individual parameters $\omega_1, \omega_2, \lambda, \phi$, respectively. The spatial distributions shown are computed by dividing the domain into bins of a histogram and recording the number of particles in each bin, with 256 particles shown at time $T = 270\text{s} = 0.075\text{hr}$. The final spatial bin to the right of the cyan dashed line represents particles that have hit the ring canal and been absorbed.

tional cost of performing forward simulations from the model in three dimensions in this complex geometry. To overcome some of the computational cost of ABC with this model, we apply ABC-SMC (see Section 2.2.4) and use a prior distribution that excludes non-biological parameter values that would be expensive to simulate. The implementation of these methods is outlined in Section 3.6.1 below. We describe microscopic summary statistics based on properties from individual complexes, and macroscopic summary statistics focused on population level behaviour.

3.6.1 Implementation of ABC

Priors and distance function

We assume truncated normal priors for parameters ν_1 , ν_2 , λ , ω_1 and ω_2 with $\mathcal{N}(0, 1)$ priors truncated at zero. Based on biological knowledge and the previous estimate described in Section 3.5.1, we expect these parameters to be positive and on the unit scale. Values of $\phi < 0.5$ result in much larger FPTs, which can be very expensive to simulate. Experimental evidence suggests the microtubule distribution is biased towards the posterior direction of the oocyte and thus we constrain the prior for the microtubule bias, ϕ , to the interval $[0.5, 1]$, uniform on this interval.

Nonetheless, for certain choices of parameter values supported by this prior, simulations from the fine-grained model can be extremely expensive due to very frequent transitions between states, or long trajectories before the mRNA complex hits the ring canal. Stopping simulations after a shorter fixed time or using an analytic approximation can prevent long simulations that for parameters with low posterior density that are likely to be rejected. Methods to enable this via Lazy ABC are described in Section 2.2.5. Here instead we use a simple analytic quantity as a guide for ignoring nonbiological parameter values by excluding them from our prior distribution. We note that excluding regions of parameter space results in zero posterior density for these parameter values, whereas Lazy ABC uses importance sampling to reweight the resulting samples. Since the excluded parameter regions are

nonbiological, it is not an issue here to exclude such parameter regions completely.

The mean distance travelled in the active transport state in a single run is given by ν_1/ω_1 . It would be outside of the typical biological parameter regime to have very large runs, so we exclude parameters for which the following holds:

$$\frac{\nu_1}{\omega_1} > 20 \mu\text{m}.$$

This criterion can be implemented by rejection sampling or by a stretch transformation of parameters drawn over $\mathbb{R}^5 \times [0.5, 1]$. We find this exclusion of nonbiological parameter regions to be effective at reducing computational cost whilst retaining otherwise weakly informative priors.

A scaled Euclidean distance function is used to measure how close the simulated and observed data are:

$$d(\mathbf{x}, \mathbf{y})^2 = \sum_i \frac{(x_i - y_i)^2}{\sigma_i^2}.$$

The scaling factors, σ_i , are chosen adaptively according to the method of Prangle [111]. See Chapter 5 for further details on this method.

ABC-SMC

We use ABC-SMC (see Algorithm 3) with $N = 512$ particles over 5 generations. At each generation, t , we generate synthetic datasets until M samples satisfy the tolerances, ϵ_i for $i < t$; that is they satisfy the tolerances for all previous generations of ABC-SMC. The tolerance at generation t is chosen as the $\alpha = 0.5$ quantile of the distribution of distances between simulated and observed data. Taking $M = \lceil N/\alpha \rceil$ results in N samples.

3.6.2 Inference on synthetic data

It is advisable in many circumstances to begin with synthetic data generated with known parameters from a known model, and check if it is possible to recover the

same parameters. This avoids issues with misspecification of models and can help uncover identifiability problems at an initial stage. Here we begin by considering the performance of different summary statistics for parameter estimation via ABC on synthetic data, with a view to using these summary statistics for analysis on experimental data.

Microscopic summary statistics

Individual behaviour of mRNA complexes can be studied *in vivo* by tracking individual complexes. Properties such as the FPT offer insight into the timescales of mRNA transport within a nurse cell. We perform inference based on summary statistics that use properties of individual mRNA complex trajectories. These are the average speed of mRNA complexes moving in active transport on microtubules and the average number of switching events², both calculated over a short trajectory of 30s, along with the FPT. The median is used to compute these averages, and we further average over a small population of 12 trajectories. From these microscopic summary statistics, we obtain the following posterior as shown in Figure 3.13, based on synthetic data created with parameters as in Table 3.1.

Macroscopic summary statistics

Since the microscopic summary statistics rely on properties of the path, they are hard to collect when tracking of particles is difficult. This is the case for the live imaging data that we have access to due to mRNA complexes moving out of the imaging plane in z (thus resulting in very short tracks), low signal to noise ratio and rapid bleaching when taking a full z stack. Instead we investigate population-level summary statistics to inform us of the population-level behaviour of particles. We use estimates of the distribution of particles spatially at several time points as a summary statistic, as described in Section 3.5.4. We simulate trajectories for $T =$

²Switching events were detected by considering framewise angle changes above a tolerance value of 0.5.

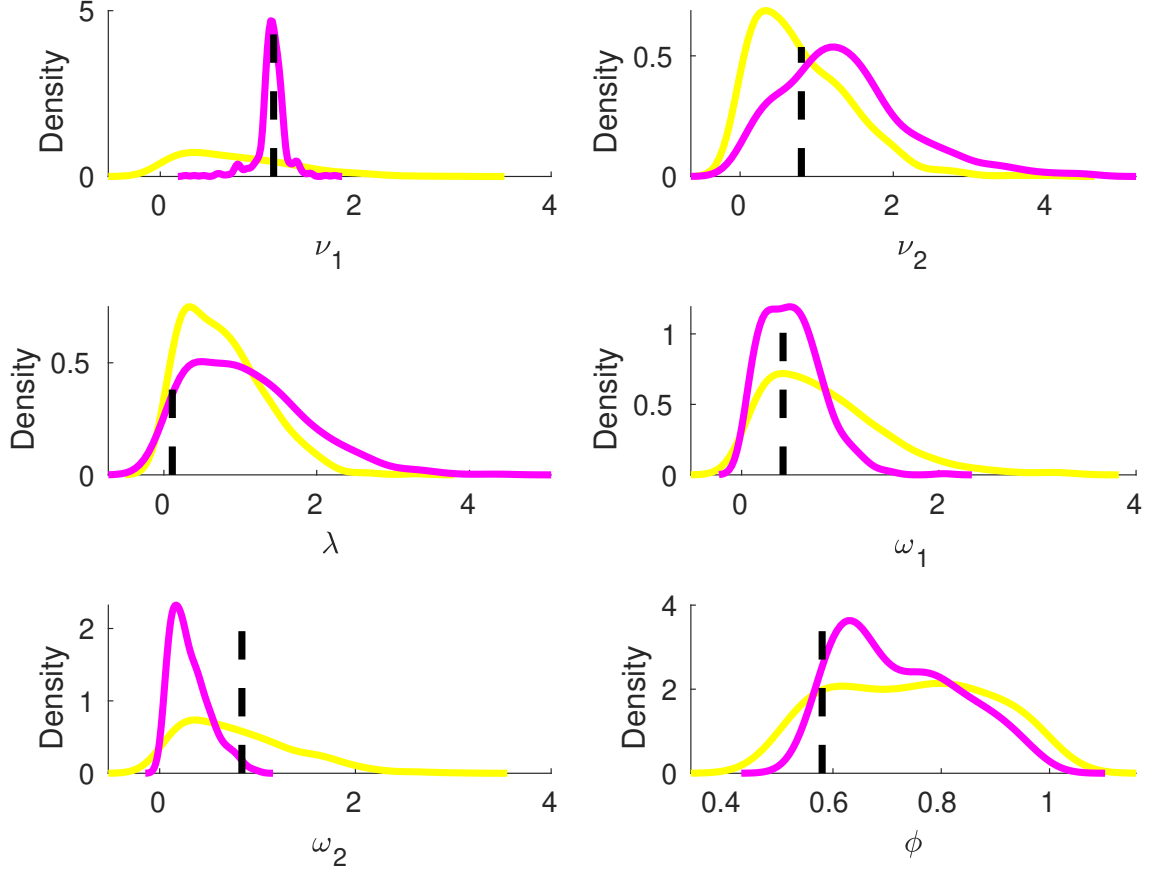


Figure 3.13. Marginal plots of the posterior distribution for the fine-grained VJP model with synthetic data using individual summary statistics. The dashed black lines show point values of the parameters in Table 3.1. The magenta curve shows the posterior distribution with the yellow curve showing the prior distribution for comparison. Parameters used for the VJP model are given in Table 3.1. The posterior distribution was sampled via ABC-SMC with $N = 512$ particles and $\alpha = 0.5$ over 10 generations as described in Section 3.6.1.

0.075 hrs = 4.5 mins and observe these every 1.5 mins, recording the quartiles of the spatial distribution. This results in four observations in time (including the initial condition) each consisting of three statistics summarizing the spatial distribution, giving a summary statistic of dimension 12. As it is expensive to simulate from the fine-grained VJP model, we construct this spatial distribution based on only a small population of complexes ($n = 12$). From these spatial summary statistics we obtain the following posterior shown in Figure 3.14, using synthetic data simulated with the same parameters in Section 3.6.2.

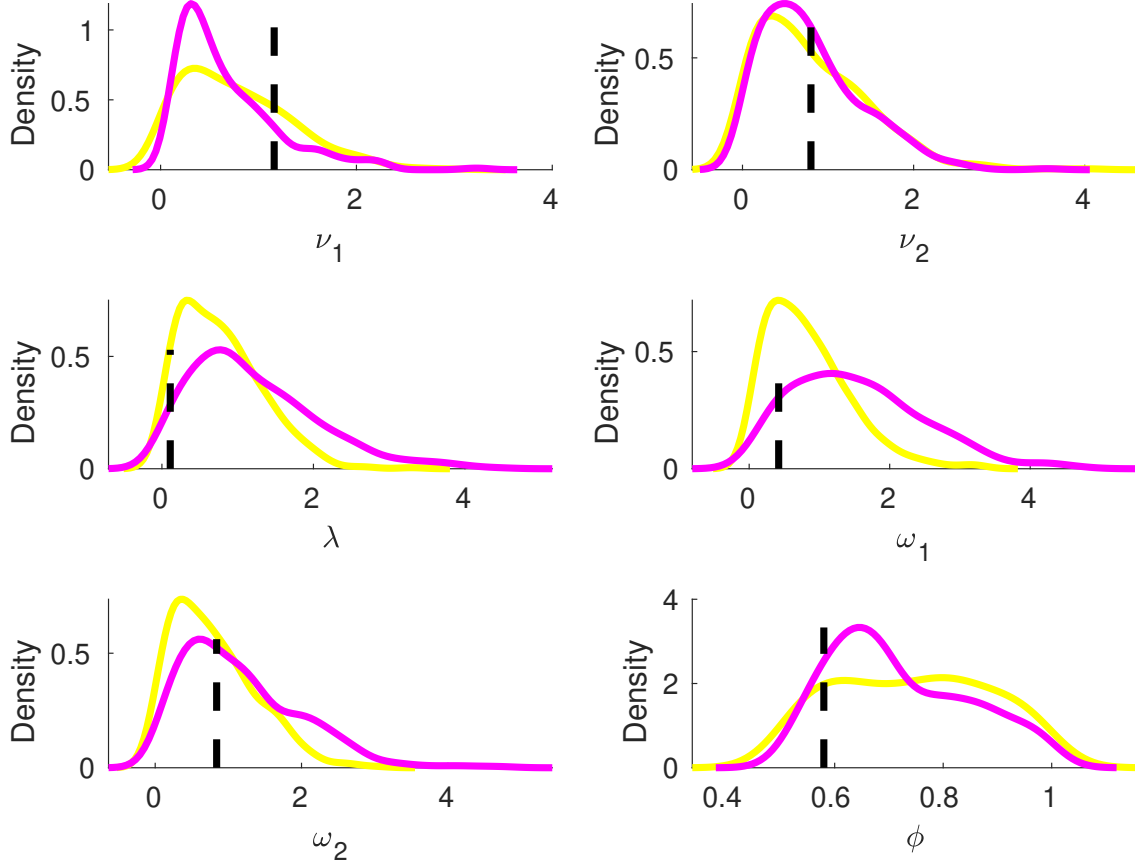


Figure 3.14. Marginal plots of the posterior distribution for the fine-grained VJP model with synthetic data using population summary statistics. The dashed black lines show point values of the parameters in Table 3.1. The magenta curve shows the posterior distribution with the yellow curve showing the prior distribution for comparison. Parameters used for the VJP model are given in Table 3.1. The posterior distribution was sampled via ABC-SMC with $N = 512$ particles and $\alpha = 0.5$ over 5 generations as described in Section 3.6.1.

3.6.3 Combined summary statistics

The microscopic and macroscopic summary statistics provide complementary contributions towards the information gained about different model parameters for our fine-grained VJP model. The microscopic summary statistics offer detailed information about local dynamics and are informative about the speed of travel, ν_1 , and turning rate, λ . In contrast, the macroscopic summary statistics describe high-level behaviour and are informative about the microtubule bias, ϕ .

Various approaches are possible to combine information from multiple different experimental sources in a Bayesian framework, such as performing multiple anal-

yses successively where the posterior from one analysis informs the prior for the next analysis [226]. Here, we provide the data from multiple sources together as one dataset. We concatenate together the microscopic and macroscopic summary statistics into a single vector of summary statistics. Performing inference via ABC for the combined summary statistics results in the posterior distribution presented in Figure 3.15, which is able to benefit from contributions from both summary statistics.

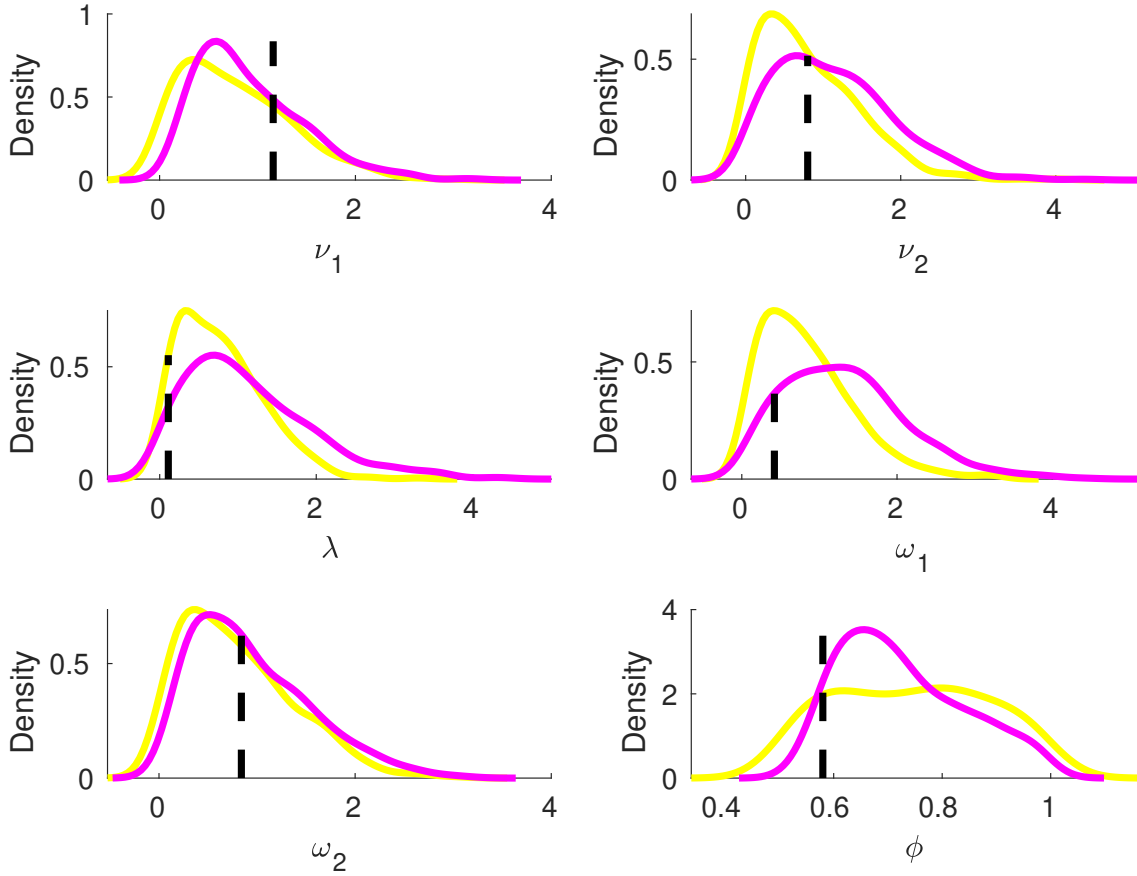


Figure 3.15. Marginal plots of the posterior distribution for the fine-grained VJP model with synthetic data using combined microscopic and macroscopic summary statistics. The dashed black lines show point values of the parameters in Table 3.1. The magenta curve shows the posterior distribution with the yellow curve showing the prior distribution for comparison. Parameters used for the VJP model are given in Table 3.1. The posterior distribution was sampled via ABC-SMC with $N = 512$ particles and $\alpha = 0.5$ over 5 generations as described in Section 3.6.1.

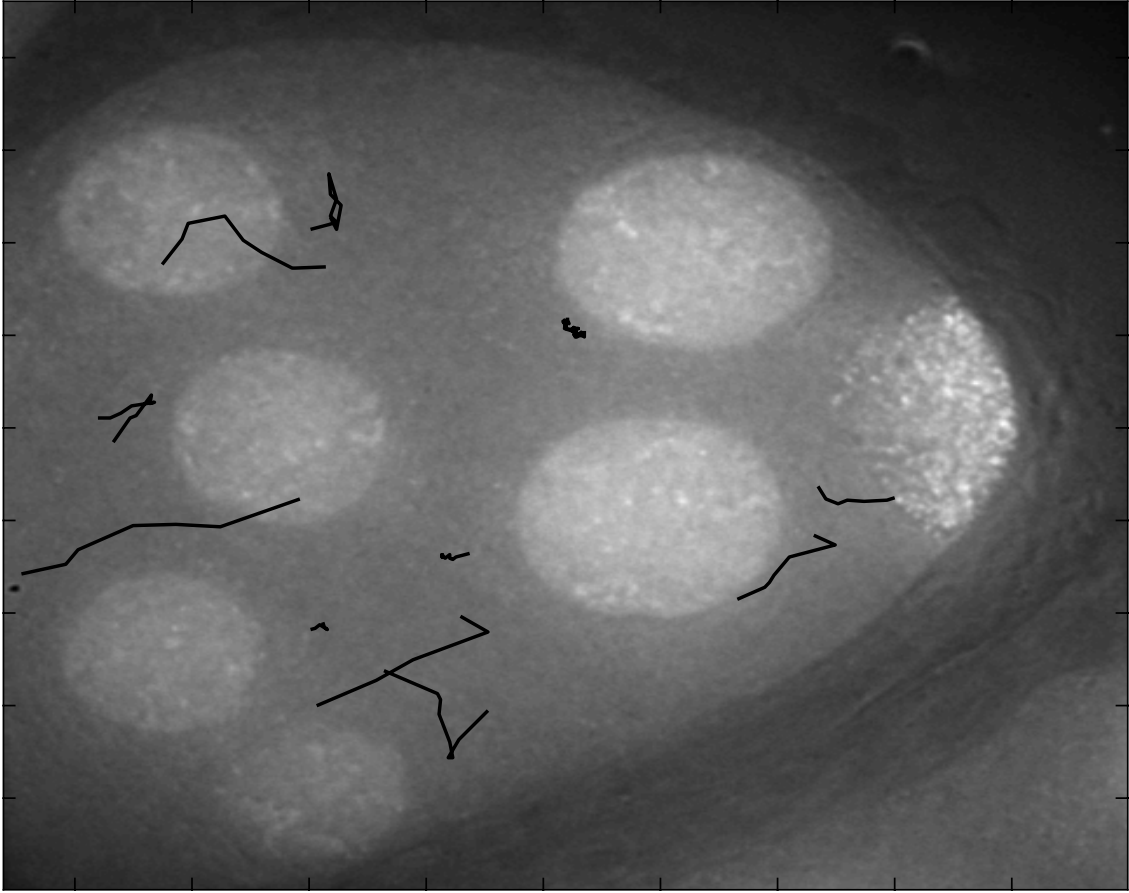


Figure 3.16. *gurken* mRNA complexes moving in the nurse cells tracked over a timescale of 30s. From a stage 5 egg chamber, 12 tracks were obtained and summary statistics computed based on these tracks. The tracks are shown overlaid over an original image obtained via live imaging of *grk*-MS2-GFP.

3.6.4 Using ABC to fit to real data

We perform parameter estimation via ABC for the fine-grained model using the FPT for transport to the ring canal of a nurse cell as a summary statistic, along with individual microscopic summary statistics (the median number of reorientations and median speed over a 30s trajectory) computed from tracks of mRNA complexes over short timescales. We obtain an experimental measurement of 18 mins for the FPT from Clark et al. [182] based on observations of accumulation of *grk* mRNA near ring canals after injection in a nurse cell. The individual summary statistics are computed based on 12 tracks from a stage 5 egg chamber, shown in Figure 3.16. Experimentally, measurements of the distribution of a population of mRNA com-

plexes proved challenging. For this reason we were restricted to summary statistics based on individual properties. This aspect is discussed further in Section 3.8.

Marginal posterior distributions for each parameter are shown in Figure 3.17 in order to display the results of inference for these data. We find similar results to those from using the microscopic summary statistics on simulated data in Figure 3.13 in that the shape of marginal posterior distributions are similar. The speed in

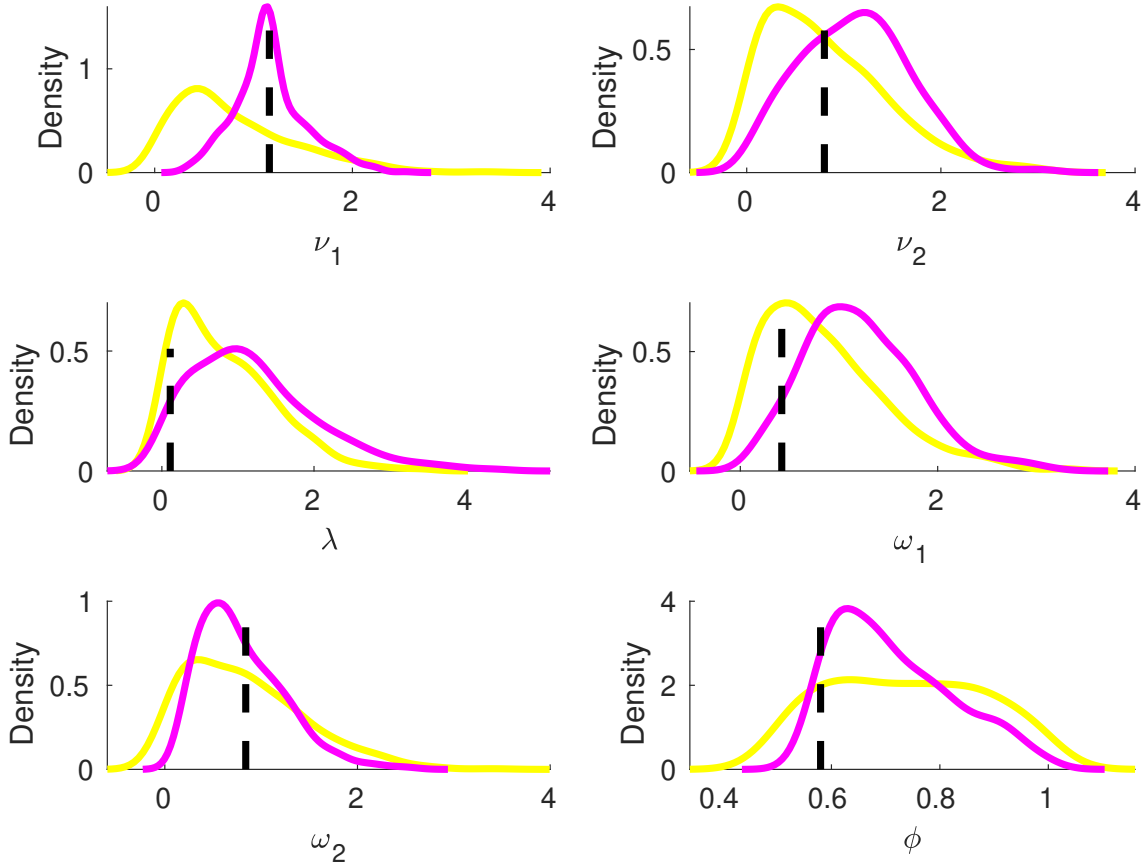


Figure 3.17. Posterior marginal distributions for the parameters of the VJP model with real observed data using microscopic summary statistics, sampled via AMC-SMC. The dashed black lines show point values of the parameters in Table 3.1. The magenta curve shows the posterior distribution with the yellow curve showing the prior distribution for comparison. The posterior distribution was sampled via ABC-SMC with $N = 512$ particles and $\alpha = 0.5$ over 10 generations as described in Section 3.6.1.

active transport, ν_1 , is identified clearly by the summary statistic relating directly to this parameter (median speed). The marginal posterior distribution for the microtubule bias parameter ϕ excludes any posterior density at very low bias values

suggesting these are inconsistent with observed data, and shows a shift from the uniform prior to greater posterior density at low to intermediate values of this bias parameter. However, posterior distributions for the remaining parameters in the model are very broad and comparable to the prior distributions. This suggests that the data used were not hugely informative about these parameters overall and that a range of combinations of parameter values could equally well explain the observed data. This behaviour is indicative of an unidentifiable model given the microscopic summary statistics from tracking data currently available. Although a lack of identifiability means we are unable to uniquely identify parameters for the model without further data, we can still make predictions from the model. Since the posterior predictive distribution $p(\tilde{\mathbf{y}}|\mathbf{y})$ involves marginalizing over the model parameters, we can integrate out the uncertainty over these parameters. Additionally, combined with the sensitivity analysis of Section 3.5.3, these results reveal significant insights into the key mechanisms and parameters affecting control over this mRNA localization process.

3.7 Testable prediction of FPT for a Dynein mutant

We conclude by making a testable prediction from the fine-grained model about the behaviour of a Dynein Heavy Chain hypomorphic mutant [162]. Dynein is the molecular motor responsible for active transport of *grk* mRNA on microtubules. Wilkie and Davis [162] report that, for this Dynein mutant, injected RNA particles move at 60 to 70% speed compared with wild type. We therefore assume that the speed of complexes in active transport for the mutant is $0.65 \times \nu_1$. We make a prediction for the FPT from nucleus to ring canal for the Dynein mutant, shown in Figure 3.18. The median FPT for the wild type based on the posterior predictive distribution is 19 mins with a 95% credible interval of [4, 69] mins for the FPT. In

contrast, for the Dynein mutant, the median FPT is 25 min with a 95% credible interval of [6, 133] mins. The difference between the phenotypes for the median values for this timescale is not large since we assumed the only parameter that changed is ν_1 . However, the difference is more noticeable if one considers the tails of the distributions. For example, it should be very unlikely for mRNA complexes in the wild type to take longer than 90 mins to be transported from the nucleus to ring canal, but should be observable for the mutant. The fine-grained model could also be used to make similar predictions for other summary statistics, or for other mutant phenotypes.

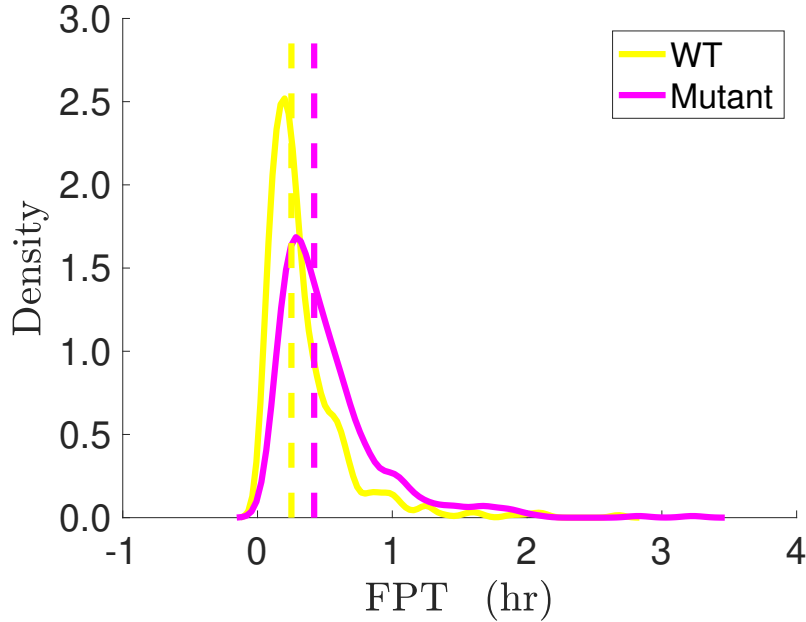


Figure 3.18. Prediction for the FPT of a Dynein Heavy Chain hypomorphic mutant affecting the Dynein molecular motor. We assume this phenotype affects the parameter ν_1 only, with a reduction to $0.65 \times \nu_1$. We simulate the posterior predictive distribution showing the FPT as a summary statistic. The yellow curve shows the predictive distribution for the wild type, and the magenta curve shows the predictive distribution for the Dynein mutant. The dashed vertical lines show the median of the FPT distribution for each phenotype to highlight the difference between the two distributions.

3.8 Discussion

In this chapter, we have developed a fine-grained stochastic model of mRNA transport within a single nurse cell. This allowed us to investigate hypotheses about the regulation and timescales of mRNA localization. We have demonstrated that mRNA transport within a single nurse cell operates on a fast timescale relative to the full mRNA localization process, which rules out transport within a cell as a regulatory bottleneck. Results of the sensitivity analysis in Section 3.5.3 indicate that reducing the motor speed by as much as an order of magnitude from observed values would be required for transport of individual mRNA complexes within a single nurse cell to be rate-limiting for the mRNA localization process. We propose possible alternative bottleneck mechanisms to be investigated such as availability of molecular motors, transcription and export through NPCs, and transit through ring canals.

With the fine-grained stochastic model, we showed a simulated FPT consistent with experimentally observed values, based on model parameters identified by biophysical measurements. A sensitivity analysis highlighted the role of cytoskeleton organization in controlling the mRNA transport process and we comment that the experimentally observed microtubule bias is the minimum bias value capable of producing the observed FPT data.

Approximate Bayesian inference methods enabled us to infer an approximate posterior distribution for model parameters and to propagate forward uncertainty to testable predictions. We performed inference here using individual summary statistics from tracking data and a FPT measurement, but have also shown via synthetic data that improved posterior distributions can be obtained using summary statistics that incorporate spatial information about the mRNA transport process. We have recommended alternative experiments to allow combined summary statistics employing spatial information to be used in future. Together, this has allowed us to make a testable model prediction about the FPT of a Dynein mutant to enable us to close the predict-test-refine loop (see Figure 1.1).

Based on currently available data in the form of microscopic summary statistics, parameters of the fine-grained VJP model are not all identifiable (see Section 2.1.7), as is common for computational and systems biology models [86]. A range of contrasting biological behaviours could equally well explain the observed data. This raises the possibility that a nurse cell could transition between different regulatory behaviours over time. For example, both a strongly polarized network with a slower motor speed, and a weakly polarized network with a higher motor speed could result in similar summary statistics. Some reorganization of the microtubule network is known to occur after stage 6 [224] and this could mark a transition between different regimes of regulation corresponding to different regions of parameter space. *Drosophila* oogenesis operates robustly under a range of different conditions, including a range of different temperatures which affect biophysical quantities such as motor speed. It is therefore important for the robustness of the mRNA localization process that a change in one parameter, such as motor speed, can be compensated for by varying other parameters.

Summary statistic choice related to experimental design

Computational studies allow us to explore experimental design setups before performing expensive or challenging experiments. In Section 3.6.3, we demonstrated the benefits of using both microscopic and macroscopic summary statistics combined together in inference with ABC on synthetic data. This result helps guide experimental design to parameterize the fine-grained model directly with experimental data. In particular, we suggest performing both short timescale imaging to enable tracking of mRNA complexes in three dimensions for microscopic summary statistics. This is nonetheless challenging since z slices must be imaged sequentially, and mRNA complexes move between imaging different slices. In addition, we recommend also performing injection or photo-activation experiments, such as with Dendra [227, 228], to study macroscopic behaviour of a population of mRNA

complexes on a timescale of minutes.

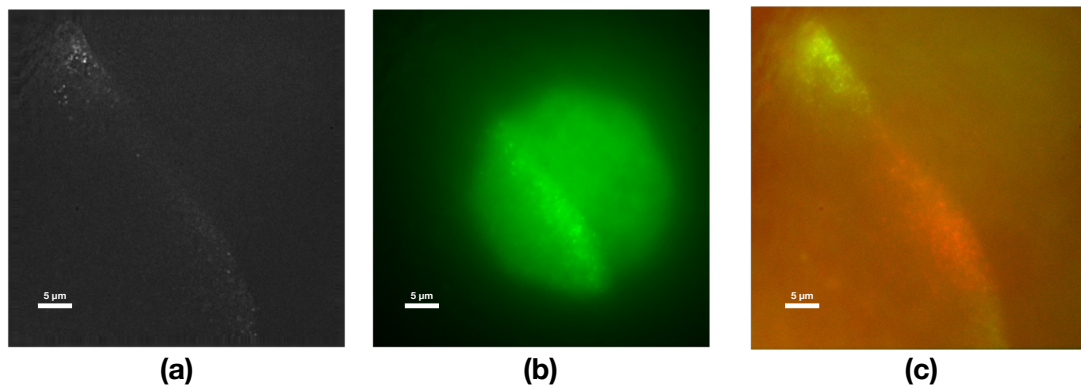


Figure 3.19. Photo-activation of *grk**MCP-Dendra can be used to study spatial distributions of mRNA complexes over time. Switching of Dendra in a stage 9 oocyte is illustrated. In (a), the distribution of *gurken* mRNA complexes is shown in the green channel prior to photo-activation. The activation of Dendra in a region of interest is displayed in (b). Activation occurs via excitation through the green channel. In (c), the oocyte is shown in two channels with red showing the photo-activated mRNA complexes, and green showing complexes that have not been activated. These data can then be used to study spatial distributions of mRNA complexes over time. As discussed in Section 3.8, technical details make collection and analysis of such data challenging.

However, although photo-activation experiments offer a wealth of possibilities for studying dynamic behaviour of mRNA regulation in space and time, they can be challenging to perform. We show typical data from a photo-activation experiment in the oocyte of a stage 9 egg chamber in Figure 3.19. Photo-activation involves using an intense irradiation at a specific wavelength to change the fluorescent state of a fluorescent protein [228]. The key issue is one of the signal to noise ratio compared to the signal to background ratio. Ideally, switching a subpopulation of tagged transcripts would allow that population to be studied easily. However this requires high signal to background ratio from a precisely defined point in time and space. The activation can be initiated in a local region of interest within the egg chamber, but the intense irradiation will not be entirely limited to this region of interest, meaning that some switching will also occur in other spatial regions. Since the switching is not instantaneous, some complexes will move in and out of the

switching region while the activation is taking place. In the oocyte with a higher concentration of complexes, it can be easier to define the switching region [229] but these experiments seem to be currently prohibitive in the nurse cells. Tracking of individual complexes for extended periods has only recently become possible in the nurse cells [224] using the MS2 system with enhanced GFP, which is brighter than Dendra2.

After the photo-activation, we would aim to observe activated and non-activated populations of complexes to study their dynamics. However, the Dendra protein is easily switched by further imaging, thus obscuring the dynamics of the activated mRNA complexes. Rapid bleaching of the sample is another obstacle to overcome. Currently, these issues are prohibitive to quantitative analysis of photo-activation data for the purposes of extracting spatial distributions of mRNA complexes for parameter estimation. We discuss this further in Section 7.2.2.

The fine-grained model motivates further investigation in three areas: experimental design within the VJP modelling framework used for the fine-grained model, the ABC inference framework for fitting the model to data, and a coarse-grained model to interrogate hypotheses about the regulation of mRNA localization on different scales.

Chapter 4

The impact of temporal sampling
resolution on parameter inference
for biological transport models

Imaging data has become an essential tool to explore key biological questions at various scales, for example the motile behaviour of bacteria or the transport of mRNA, and it has the potential to transform our understanding of important transport mechanisms. Often these imaging studies require us to compare biological species or mutants, and to do this we need to quantitatively characterize their behaviour. Mathematical models offer a quantitative description of a system that enables us to perform this comparison, but to relate mechanistic mathematical models to imaging data, we need to estimate their parameters. In this chapter, we study how collecting data at different temporal resolutions impacts our ability to infer parameters of biological transport models; performing exact inference for simple velocity jump process models in a Bayesian framework. The question of how best to choose the frequency with which data is collected is prominent in a host of studies because the majority of imaging technologies place constraints on the frequency with which images can be taken, and the discrete nature of observations can introduce errors into parameter estimates. In this chapter, we mitigate such errors by formulating the velocity jump process model within a hidden states framework. This allows us to obtain estimates of the reorientation rate and noise amplitude for noisy observations of a simple velocity jump process. We demonstrate the sensitivity of these estimates to temporal variations in the sampling resolution and extent of measurement noise. We use our methodology to provide experimental guidelines for researchers aiming to characterize motile behaviour that can be described by a velocity jump process. In particular, we consider how experimental constraints resulting in a trade-off between temporal sampling resolution and observation noise may affect parameter estimates. Finally, we demonstrate the robustness of our methodology to model misspecification, and then apply our inference framework to a dataset that was generated with the aim of understanding the localization of RNA-protein complexes. The research contained in this chapter has been published in the journal *PLoS Computational Biology* as Harrison and Baker [1].

4.1 Introduction

Biological transport processes occur on a wide range of spatial and temporal scales, and a common mechanism for transport involves two phases: fast active transport, and a quasi-stationary reorientation phase. This pattern of movements has been observed at a range of scales from the intracellular transport of cellular components such as mRNA particles moving on a microtubule network [44], to the run-and-tumble motion of bacteria such as *Escherichia coli* [220, 230, 231], and the flights of birds between nesting sites [222]. To capture appropriately these two phases of motion, a class of models known as VJP models [220–222, 232, 233] (also known as correlated random walks [234–238] or Levy Walks [239–242]) have been developed. The VJP model framework was utilized for the fine-grained stochastic model of Chapter 3 and will be explored in further detail here.

Estimating the parameters of these models can give us mechanistic information relating to the underlying biological process, such as the rate of reorientation. Being able to obtain accurate estimates, with appropriate uncertainty, for these parameters allows us to compare different biological species or mutants, and gain an understanding of the underlying mechanistic behaviour. Importantly, parameterizing models and quantifying the uncertainty in parameter estimates, as can be achieved via Bayesian inference, enables us to use models to make quantitative predictions of behaviour in new conditions with quantifiable uncertainty. By performing experiments to test model predictions, we can evaluate the areas in which a given model fails to describe experimental data, and so iteratively refine our understanding of a given system or phenomenon.

In this work, we consider the effects that experimental design can have on the information we can obtain from a data set, in terms of using that data to estimate parameters of a mechanistic model. In particular, for time series data describing a biological transport process, we vary the time between successive measurements. We demonstrate a framework for estimating the parameters of a VJP model for

data of this form in the presence of noise, and examine how the posterior estimates of the model parameters change for more coarsely sampled and noisier datasets. Our framework formulates the VJP model as a process with hidden states, as in a hidden Markov model (HMM), which allows us to use pMCMC methods to perform exact Bayesian inference. We use our framework to suggest sensible experimental design choices in the context of microscopy studies, where there may be a trade-off between how frequently it is possible to image, and the noise resulting from more or less frequent observations of the process. We present a comparison between the pMCMC framework described in this work and ABC for parameter inference in this context. We demonstrate robustness to model misspecification in a situation where we attempt inference on synthetic data generated with a different model to the assumed model. Finally, we apply our method to an imaging dataset of tracks of RNA-protein complexes allowing us to estimate motility parameters for a mechanistic model.

4.1.1 Velocity jump process models

VJP models have been developed to describe biological transport processes where there is persistent or biased random motion [220, 222, 243]. Correlated, persistent motion is observed experimentally in a variety of contexts [244, 245]. VJP models are most appropriate for motion consisting of multiple phases, such as a fast directed phase and a stationary or reorientation phase, and these models describe how an object moves in one direction before reorienting and moving in a new direction. This type of “run and reorientate” motion is exhibited, for example, by *Escherichia coli* [230, 231], fibroblasts [245], and RNA-protein complexes [44].

We present here a mathematical description of a VJP model. Suppose we have a running time distribution, with probability density function (pdf) f_τ , and a waiting time distribution, with pdf f_μ . Random variables drawn from these distributions dictate the lengths of time spent in the fast active transport, or running, phase of

the VJP and the reorientation phase, respectively. After the reorientation phase, a velocity for the new run is chosen according to a transition kernel, f_v . This transition kernel could incorporate a distribution of speeds as well as directions, or could rely on a fixed speed and specify only the directionality of the run. In the fixed speed case,

$$f_v(\mathbf{v}) = \delta(|\mathbf{v}| - c)f_\Phi(\phi), \quad (4.1)$$

where δ is the Dirac delta function, f_Φ is a reorientation kernel describing the angle change, and c is the constant running speed. For biological processes with a distinct separation of timescales between the running and reorientation phases, it is often possible to assume that reorientations between successive runs occur instantaneously and therefore neglect the reorientation phase in a model of the process [220, 222]. Repeated simulation from this model can be used to generate trajectories of individual particles (see Figure 4.1, for example).

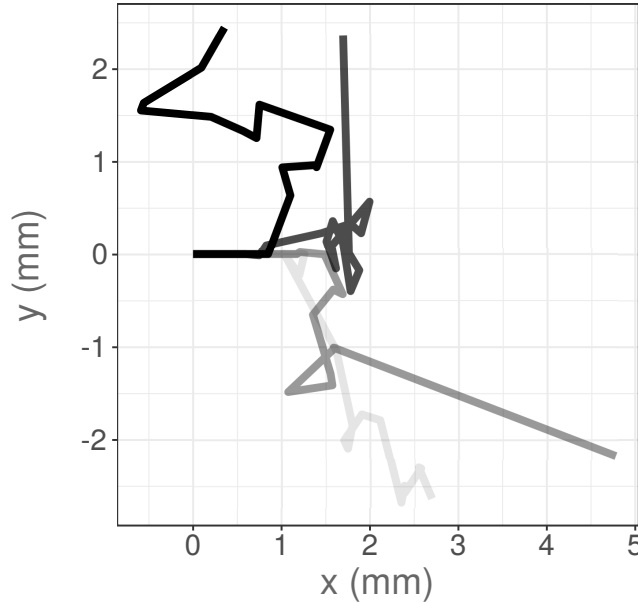


Figure 4.1. Four example VJP trajectories, simulated with a uniform reorientation kernel, running speed $c = 50 \mu\text{ms}^{-1}$, exponential running time distribution, $f_\tau(t) = \lambda e^{-\lambda t}$, with reorientation frequency $\lambda = 0.2 \text{ s}^{-1}$ and no waiting time between runs, for a duration of $T_{\text{final}} = 64 \text{ s}$. These trajectories start from the origin, and are orientated initially parallel to the positive x -axis.

In the simplest case, where we assume that the running time distribution is memoryless, that is, exponentially distributed, with $f_\tau(t) = \lambda \exp(-\lambda t)$, the long time behaviour of the mean squared displacement scales linearly with time, t [220]. Further moments of the motion of individuals displaying such VJP behaviour have been characterized by making certain closure assumptions [221]. For the case of a more general running time distribution, with finite mean and variance, in the large time limit the probability of the particle being at position x at time t follows a diffusion equation [222].

In practice, VJP models are often parameterized by obtaining measurements of the effective diffusion coefficient or the mean squared displacement, and using these data to estimate the parameters of a specific running distribution [220, 222]. Rosser et al. [231] parameterized a HMM via maximum likelihood estimation, whilst Nicosia et al. [246] fitted a hidden state random walk model to animal movement data using an expectation-maximization algorithm. These frequentist approaches can provide useful point estimates of the VJP parameters, and quantify the error in these estimates. However, Bayesian approaches can propagate uncertainty from both process noise and measurement noise, which can be crucial when dealing with noisy biological data (see Figure 4.1). In addition, they provide the added benefit that we can interpret the results of a Bayesian analysis as probabilistic statements about the model parameters. This quantification of uncertainty enables the generation of predictions of further biological behaviour using the model. In addition, we can consider the effects of noisy data measurements upon the accuracy of the inferred parameter distributions, something which has previously been difficult to deal with in practice, or has been neglected.

For simplicity, in this work we will assume that there is a separation of timescales between the lengths of the running and reorientation phases, such that reorientations can be considered instantaneous. We will restrict ourselves to considering a two dimensional (2D) model. For some of the relevant biological processes for VJP

models, such as bacterial motility, experiments are conducted in an environment that can be well approximated by a 2D plane [231]. The arguments presented in this chapter could be extended to a further spatial dimension for a relevant biological application.

In addition, we assume that the running time distribution, f_τ , is exponentially distributed, that particles run at a fixed speed, c , and the reorientation kernel is a uniform distribution on $[-\pi, \pi)$. That is $f_\tau(t) = \lambda \exp(-\lambda t)$ and $f_\Phi(\phi) = 1/2\pi \mathbb{1}_{[-\pi, \pi)}(\phi)$. We follow the trajectory of a single, motile individual, and take, as experimental measurements, the change in angle between successive observed positions (calculated relative to the previous observed position, see Figure 4.2), subject to measurement noise drawn from a wrapped Normal distribution, $\mathcal{N}(0, \sigma^2)$, where σ is the magnitude of the noise. Our Bayesian inference approach will target estimation of the reorientation rate, λ , and the magnitude of the noise, σ .

4.1.2 Inference for velocity jump process models via particle MCMC

Inference for partially observed Markov processes can be performed using particle Markov Chain Monte Carlo (pMCMC), as developed by Andrieu and Roberts [247] and Andrieu et al. [122]. pMCMC, as described in Section 2.2.7, provides a Bayesian framework for parameter estimation by allowing samples to be drawn from the posterior distribution of the model parameters, given observed data, without needing to evaluate the likelihood function directly. For partially observed Markov process models, the model structure makes directly evaluating the likelihood difficult or expensive. (We note that VJP models can be viewed in this form by introducing hidden states, as explained in Section 4.2.) Instead of evaluating the likelihood directly, we can use (unbiased) estimates of the likelihood within an MCMC algorithm. Estimating the likelihood of the observed data given certain parameters can be achieved with a particle filter [79] for a fixed, finite, number of particles. The

results of Andrieu et al. [122] demonstrate that, even when a finite number of particles are used in the filter to estimate the likelihood, the MCMC algorithm will still target the correct posterior distribution. However, pMCMC methods for parameter inference have not previously been applied to spatial agent-based models, such as the VJP model considered here. The details of the pMCMC algorithm used in this work are given in Section 4.2.

4.2 Methods

We will demonstrate how to exploit pMCMC methods to obtain posterior parameter estimates for the VJP outlined in Section 4.1 by formulating the model within an appropriate framework that incorporates hidden states. This formulation additionally allows us to incorporate a model for measurement noise in our observations, as well as explicitly accounting for the temporal discretization of the data. The hidden states (also known as latent variables) in our model will describe whether or not a reorientation event occurred between observations of the VJP. This is not a variable that we can observe directly, since we only measure the observed angle change. For example, in Figure 4.2 there was a reorientation event between observations at $t = k\Delta t$ and $t = (k + 1)\Delta t$, and an observed angle change of θ_1 . On the other hand, there was no reorientation event between observations at $t = (k + 1)\Delta t$ and $t = (k + 2)\Delta t$ but still a non-zero observed angle change of θ_2 . Note that the true reorientation angle for the reorientation event that takes place between observations at $t = k\Delta t$ and $t = (k + 1)\Delta t$ is ϕ .

We assume that we observe the system at times $\{k\Delta t : k \in \{0, 1, \dots, T\}\}$, up until a final time $T_{\text{final}} = T\Delta t$, by measuring the position of the individual of interest and recording the observed angle change. We define the hidden variable,

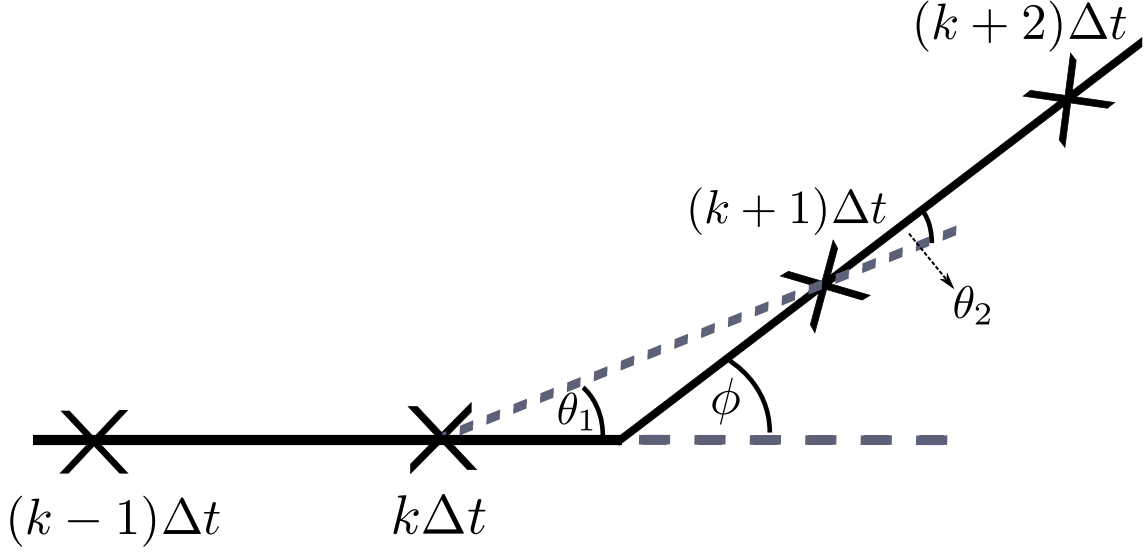


Figure 4.2. The motile individual is observed at discrete times $t = k\Delta t$ for $k \in \{0, 1, \dots, T\}$ (marked by crosses). In each time interval between observations, we measure the angle change between the observed direction of travel between successive observations. We assign a hidden state to each time interval according to whether a reorientation event occurred in that time interval or not. In this example, which excludes measurement noise, there is a reorientation, through angle ϕ , between observations at times $t = k\Delta t$ and $t = (k+1)\Delta t$; however, the measured angle change between these observations is θ_1 . Conversely, there is no reorientation event between observations at $t = (k+1)\Delta t$ and $t = (k+2)\Delta t$ and an observed angle change of θ_2 .

X_k , as follows:

$$X_k = \begin{cases} 1, & \text{if a reorientation event occurred during } [k\Delta t, (k+1)\Delta t), \\ 0, & \text{otherwise.} \end{cases} \quad (4.2)$$

The observed state is the observed angle change, θ_k , defined as the difference between the observed direction of travel during $[(k-1)\Delta t, k\Delta t)$ and $[k\Delta t, (k+1)\Delta t)$, $k \geq 1$, as illustrated in Figure 4.2. For example, if we directly observe as data the positions $\{(x_k, y_k) : k \in \{0, 1, \dots, T\}\}$, then we can calculate the observed angle change as

$$\left\{ \theta_k = \arctan \left(\frac{y_{k+1} - y_k}{x_{k+1} - x_k} \right) : k \in \{0, 1, \dots, T-1\} \right\}.$$

We illustrate the sequence of hidden and observed states in Figure 4.3, with dependence between these states given by the transition and emission probabilities, denoted β_i and p_{ij} , respectively. The hidden variable, X_k , evolves according to the VJP model. Here, since we have an exponential distribution for the running time, $P(\text{reorientation event in } [(k-1)\Delta t, k\Delta t]) = 1 - \exp(-\lambda\Delta t)$.

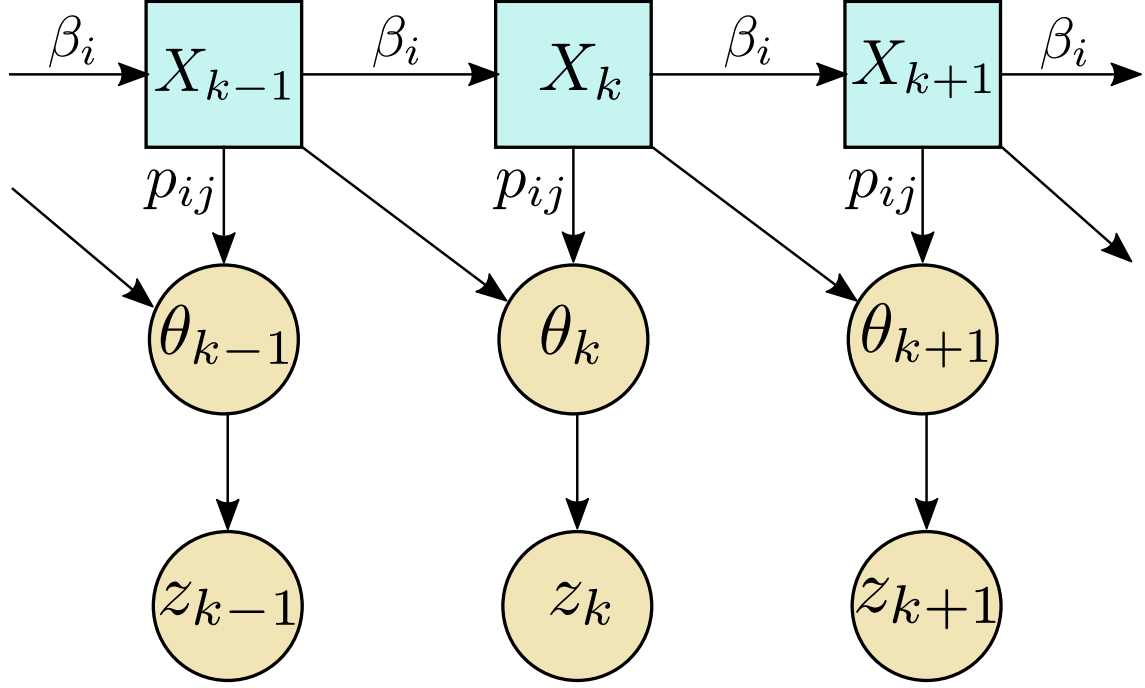


Figure 4.3. Hidden and observed states in a partially observed Markov process model. We observe an angle change, θ_k , which is dependent on a hidden state X_k defined in Equation (4.2). This dependency is shown by the arrows between states. Here, we require dependencies on the previous hidden state, X_{k-1} , since observation times will not coincide with reorientation events in general. Additionally, we introduce an extra layer of state dependencies to capture the measurement noise in this noisy biological system. We suppose that θ_k is the true observed angle change and our noisy observed version of this is z_k , with dependencies shown by the arrows.

We assume that reorientation events are rare relative to the sampling rate such that $\lambda\Delta t \ll 1$. This assumption allows us to neglect multiple reorientations in a single time interval, enabling us to greatly simplify the problem, so that it is tractable via the binary hidden variables X_k . We highlight that the model structure as shown in Figure 4.3 is an approximation. In reality, observed data will not be

a function of only the current and previous hidden states, X_k and X_{k-1} , but may depend on other previous states also. The assumption that $\lambda\Delta t \ll 1$ enables us to neglect dependence on earlier states.

4.2.1 Model without measurement noise

We initially make progress by simplifying the problem and its exposition via the assumption of zero measurement error. To relate the unobserved hidden state to the observed angle change, which we can measure, we derive probability distributions for the angle change given the hidden state. Note that there is additional dependence not only on the current hidden state, but also on the previous hidden states, as shown in Figure 4.3. We can obtain expressions for the emission probabilities by considering the path that is taken under different sequences of hidden states (Figure 4.4); we will outline simplifying assumptions as they are made.

Suppose we observe the system at times $(k-1)\Delta t, k\Delta t$, and $(k+1)\Delta t$ for some k , as shown in Figure 4.4. The sequence of hidden states i, j , corresponds to whether there were reorientation events in the time intervals $[(k-1)\Delta t, k\Delta t)$ and $[k\Delta t, (k+1)\Delta t)$, respectively. Let $p_{ij}(\theta)$ be the pdf of observing a reorientation angle of θ given the sequence of hidden states i, j . We assume angle change θ was observed over the time interval $[k\Delta t, (k+1)\Delta t)$ corresponding to the hidden state j . In the case where no reorientation occurs in either of the time intervals (corresponding to the situation in Figure 4.4a)), then, assuming no noise, we would observe zero angle change. That is, we have $p_{00}(\theta) = \delta(\theta)$. If a reorientation occurs in the time interval $[k\Delta t, (k+1)\Delta t)$, but not in the preceding time interval, $[(k-1)\Delta t, k\Delta t)$, (as shown in Figure 4.4b)), then the observed reorientation angle is θ_1 , as labelled in Figure 4.4b). This gives $p_{01}(\theta) := p_{\Theta_1}(\theta)$. The marginal distribution of Θ_1 is derived in Section 4.2.3, and depends on the running time distribution and the reorientation kernel.

If, immediately after a reorientation, we have no reorientation in the following

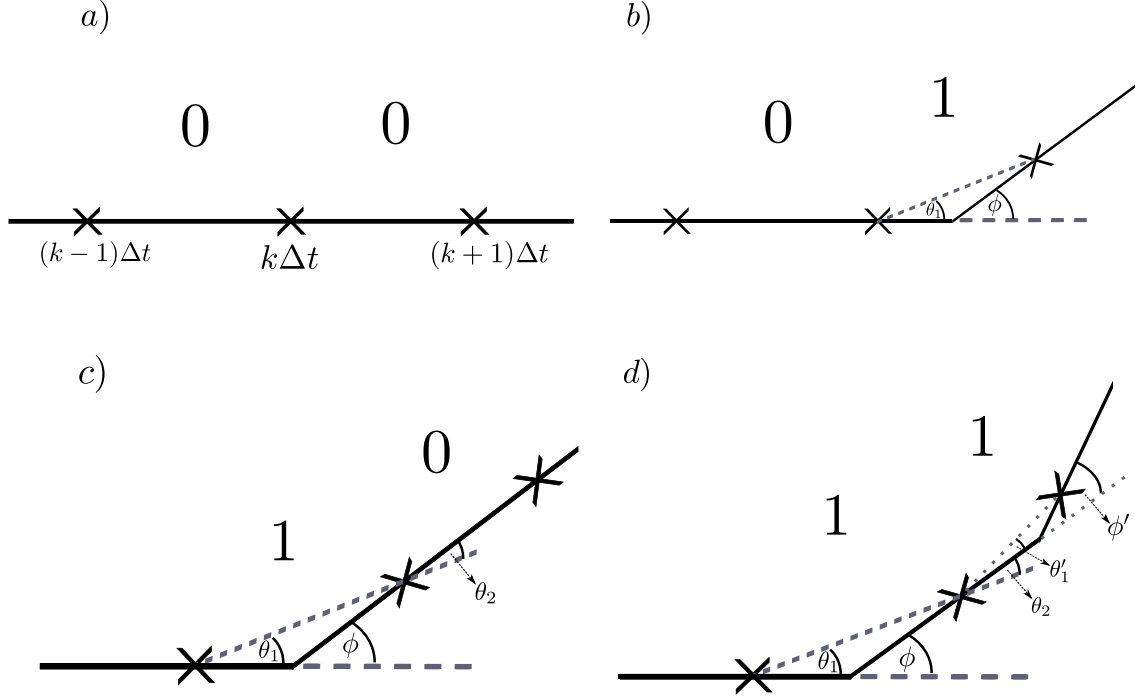


Figure 4.4. Example paths for each pattern of hidden states. The particle of interest moves from left to right, and is observed at times $(k-1)\Delta t$, $k\Delta t$, and $(k+1)\Delta t$. In a), there are no reorientation events, so the particle continues along a straight trajectory. In b), there is a single reorientation in the time interval $[k\Delta t, (k+1)\Delta t)$, where the particle turns through an angle ϕ , but the observed angle change is θ_1 due to the discrete nature of the observations. In c), there is a single reorientation in the time interval $[(k-1)\Delta t, k\Delta t)$. The particle turns through an angle ϕ and continues on its new trajectory. We observe an angle change θ_1 for the time interval $[(k-1)\Delta t, k\Delta t)$, and observe an angle change θ_2 for the next time interval $[k\Delta t, (k+1)\Delta t)$ even though there was no reorientation during this time interval. Similarly, in d), there was a reorientation of true angle change ϕ during $[(k-1)\Delta t, k\Delta t)$ followed by another reorientation of true angle change ϕ' in the time interval $[k\Delta t, (k+1)\Delta t)$. In this case, we observe angle change θ_1 for the time interval $[(k-1)\Delta t, k\Delta t)$, and observe an angle change of $\theta_2 + \theta'_1$ for the time interval $[k\Delta t, (k+1)\Delta t)$.

time interval, then we may still observe a nonzero angle change since our discrete observation times do not in general coincide with the reorientation events. In this case, we have a reorientation event during $[(k-1)\Delta t, k\Delta t)$, and no reorientation during $[k\Delta t, (k+1)\Delta t)$. Such a situation with a pattern of hidden states 1, 0 is shown in Figure 4.4c), and the observed angle change during the time interval $[k\Delta t, (k+1)\Delta t)$ corresponds to θ_2 in the diagram. We note that, by geometric arguments, we

have

$$\theta_1 + \theta_2 = \phi, \quad (4.3)$$

where ϕ is the true angle change. Therefore, given a pattern of hidden states $1, 0$ the angles θ_1 and θ_2 are not independent, and we have $p_{10}(\theta_1, \theta_2) = p_{\Theta_2|\Theta_1=\theta_1}(\theta_2)$. Note that throughout we will assume that we can observe the previous angle change directly, which is equivalent to assuming that the pattern of hidden states is in fact $0, 1, 0$.

The remaining cases involve reorientation events in successive time intervals. Suppose we have the case where we have two successive reorientation events, giving a pattern of hidden states $1, 1$, which is shown graphically in Figure 4.4d). This case is similar to the case with hidden states $1, 0$ (see Figure 4.4c)), in that we observe a contribution to the reorientation angle, θ_2 , from the correction for the previous reorientation event, and also a contribution from the new reorientation event, θ'_1 . As can be seen in Figure 4.4d), these contributions sum to give an observed angle change for the time interval $[k\Delta t, (k+1)\Delta t]$ of $\theta = \theta_2 + \theta'_1$. The probability density for sums of random variables can be expressed as a convolution [248]. Hence, $p_{11}(\theta_1, \theta) = (p_{\Theta_1} * p_{\Theta_2|\Theta_1=\theta_1})(\theta)$, where $*$ is the convolution operator. Any further cases involve multiple reorientation events within a single time interval which occurs rarely, with probability on the order of $\mathcal{O}((\lambda\Delta t)^2)$. We can safely neglect these provided $\lambda\Delta t \ll 1$.

To summarize therefore, the emission probabilities (that is the probability of observing a certain angle change given the sequence of hidden states) in the case

without noise can be given as follows:

$$p_{00}(\theta) = \delta(\theta); \quad (4.4)$$

$$p_{01}(\theta) = p_{\Theta_1}(\theta); \quad (4.5)$$

$$p_{10}(\theta) = p_{\Theta_2|\Theta_1}(\theta); \quad (4.6)$$

$$p_{11}(\theta) = (p_{\Theta_2|\Theta_1} * p_{\Theta'})(\theta). \quad (4.7)$$

4.2.2 Model with measurement noise

To account for noise, we introduce another layer of states in our diagram of state dependencies, as shown in Figure 4.3. Let z_k be the noisy observed angle change at time $k\Delta t$, which is a noisy observation of θ_k , such that for a noise model K we have $z_k \sim K(\cdot, \theta_k)$. Alternatively, we can write the noisy observed angle as

$$z_k = \theta_k + \epsilon_k, \quad (4.8)$$

where $\epsilon_k \sim K(\cdot, 0)$. Under the assumption of a wrapped normal noise model, $\epsilon_k \sim \mathcal{N}(0, \sigma^2)$, where σ is the magnitude of the measurement noise. Since Equation (4.8) represents z_k as a sum of random variables, it becomes clear that we can obtain the noisy emission probabilities, q_{ij} , corresponding to a pattern of hidden states i, j via the following convolution:

$$q_{ij}(\theta) = (p_{ij} * K)(\theta) \quad (4.9)$$

$$= \int_{\Theta} p_{ij}(\theta - x) K(x, 0) dx. \quad (4.10)$$

To extend our previous results to the noisy case, we therefore need to compute these convolutions, a task that must be carried out numerically.

4.2.3 Derivation of emission probabilities

In previous work [249], marginal distributions for the observed angle change, Θ_1 , were obtained and we summarize the arguments here. Figure 4.5 shows the true angle change, Φ , and the observed angle change, Θ_1 , based on discrete time observations of the VJP for the case of hidden states 0, 1.

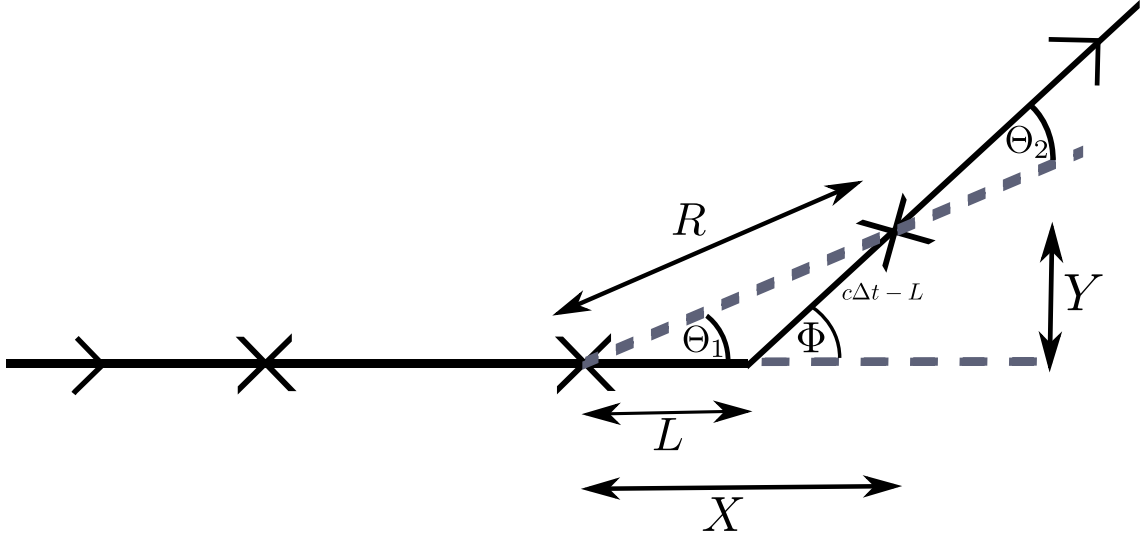


Figure 4.5. True and observed angle changes (Φ and Θ_1 , respectively) based on discrete time observations of a VJP without measurement error. Cartesian co-ordinates X and Y are shown in addition to polar co-ordinates R and Θ_1 . The particle is moving along the trajectory from left to right as shown by the arrows.

By changing co-ordinates from the displacement, L , (which is given by the running time distribution f_τ) and true angle change, Φ , to Cartesian co-ordinates X and Y , and then to polar co-ordinates, R and Θ_1 , we can obtain a joint distribution for R and Θ_1 of

$$f_{R,\Theta_1}(r, \theta_1) = \begin{cases} \frac{r}{c\Delta t(c\Delta t - r \cos \theta_1)} f_\Phi \left(\arccos \left[\frac{r^2 \sin^2 \theta_1 - (c\Delta t - r \cos \theta_1)^2}{r^2 \sin^2 \theta_1 + (c\Delta t - r \cos \theta_1)^2} \right] \right), & \text{if } (r, \theta_1) \in C, \\ 0, & \text{otherwise,} \end{cases} \quad (4.11)$$

where $C = \{r \leq c\Delta t, \theta \in [-\pi, \pi)\}$ is the set of permissible values for r and θ , Δt is

the discretization in time and c is the running speed. Integrating $f_{R,\Theta_1}(r, \theta_1)$ over r allows us to obtain the marginal distribution for θ_1 via

$$f_{\Theta_1}(\theta_1) = \int_0^\infty f_{R,\Theta_1}(r, \theta_1) dr. \quad (4.12)$$

In the case where we assume a uniform reorientation kernel, $f_\Phi(\phi) = 1/2\pi \mathbf{1}_{[-\pi, \pi)}(\phi)$, then this integral becomes

$$f_{\Theta_1}(\theta_1) = \int_0^{c\Delta t} \frac{r}{2\pi c\Delta t(c\Delta t - r \cos \theta_1)} dr \quad (4.13)$$

$$= \frac{-1}{2\pi c\Delta t \cos(\theta_1)} \int_0^{c\Delta t} \left(1 + \frac{\frac{c\Delta t}{\cos \theta_1}}{r - \frac{c\Delta t}{\cos \theta_1}} \right) dr \quad (4.14)$$

$$= \frac{-1}{2\pi c\Delta t \cos(\theta_1)} \left[r + \frac{c\Delta t}{\cos \theta_1} \log \left| r - \frac{c\Delta t}{\cos \theta_1} \right| \right]_0^{c\Delta t} \quad (4.15)$$

$$= \frac{-1}{2\pi \cos^2 \theta_1} (\cos \theta_1 + \log(1 - \cos \theta_1)). \quad (4.16)$$

We note that the marginal distribution for θ_1 does not depend on the speed, c , or the time discretization, Δt . This is intuitive because f_{Θ_1} is the pdf of a certain observed angle change, θ_1 , given there was a reorientation in that interval, irrespective of the length of that interval.

Derivation of the joint distribution of Θ_1 and Θ_2

By considering the displacement in the X and Y directions during a time step, we have

$$r \cos(\theta_1) = l + (c\Delta t - l) \cos(\phi), \quad (4.17)$$

$$r \sin(\theta_1) = (c\Delta t - l) \sin(\phi). \quad (4.18)$$

Dividing these expressions, we can relate θ_1 to l and ϕ by

$$\tan(\theta_1) = \frac{(c\Delta t - l) \sin(\phi)}{l + (c\Delta t - l) \cos(\phi)}, \quad (4.19)$$

and rearranging for l , we have

$$l = \frac{c\Delta t(\tan \phi - \tan \theta_1)}{\tan \phi + \tan \theta_1(\sec \phi - 1)}. \quad (4.20)$$

Differentiating with respect to θ_1 , we obtain

$$\frac{\partial l}{\partial \theta_1} = \frac{-c\Delta t \sec^2 \theta_1}{\tan \phi + \tan \theta_1(\sec \phi - 1)} - \frac{c\Delta t(\tan \phi - \tan \theta_1) \sec^2 \theta_1(\sec \phi - 1)}{(\tan \phi + \tan \theta_1(\sec \phi - 1))^2}, \quad (4.21)$$

which can be simplified to give

$$\frac{\partial l}{\partial \theta_1} = \frac{-c\Delta t \sin \phi}{(\sin \phi \cos \theta_1 + \sin \theta_1(1 - \cos \phi))^2}. \quad (4.22)$$

To transform from coordinates (L, Φ) to coordinates (Θ_1, Φ) , we can use the Jacobian $J_{L,\Phi}$, where

$$\det J_{L,\Phi} = \frac{\partial l}{\partial \theta_1} = \frac{-c\Delta t \sin \phi}{(\sin \phi \cos \theta_1 + \sin \theta_1(1 - \cos \phi))^2}. \quad (4.23)$$

Therefore the joint distribution of θ_1 and ϕ is

$$f_{\Theta_1, \Phi}(\theta_1, \phi) = \begin{cases} \frac{|c\Delta t \sin \phi|}{(\sin \phi \cos \theta_1 + \sin \theta_1(1 - \cos \phi))^2} f_\phi(\phi) \cdot \frac{1}{c\Delta t}, & \text{if } \theta_1, \phi \in C, \\ 0, & \text{otherwise,} \end{cases} \quad (4.24)$$

where $C = \{(\theta_1, \phi) : \phi \in [-\pi, \pi), \theta_1 \in (\min(0, \phi), \max(0, \phi))\}$. Under the assumption that ϕ is uniform on $[-\pi, \pi)$, as described in the Section 4.1.1, such that

$f_{\Phi}(\phi) = 1/2\pi \mathbb{1}_{[-\pi, \pi)}(\phi)$, we have

$$f_{\Theta_1, \Phi}(\theta_1, \phi) = \begin{cases} \frac{|\sin \phi|}{2\pi(\sin \phi \cos \theta_1 + \sin \theta_1(1 - \cos \phi))^2}, & \text{if } \theta_1, \phi \in C, \\ 0, & \text{otherwise.} \end{cases} \quad (4.25)$$

Changing variables again to (θ_1, θ_2) , via $\phi = \theta_1 + \theta_2$, we have

$$f_{\Theta_1, \Theta_2}(\theta_1, \theta_2) = \begin{cases} \frac{|\sin(\theta_1 + \theta_2)|}{2\pi(\sin(\theta_1 + \theta_2) \cos \theta_1 + \sin \theta_1(1 - \cos(\theta_1 + \theta_2)))^2}, & \text{if } \theta_1, \theta_2 \in C, \\ 0, & \text{otherwise,} \end{cases} \quad (4.26)$$

where $C = \{(\theta_1, \theta_2) : \theta_1 + \theta_2 \in [-\pi, \pi)\}$. From this joint distribution, we can then find the conditional distribution of θ_2 given θ_1 , which is required to give $p_{10}(\theta) = f_{\Theta_1, \Theta_2}(\theta_1, \theta)/f_{\Theta_1}(\theta_1)$.

4.2.4 Comparison of expression for emission probabilities with simulations

To verify our emission probabilities calculated in the Section 4.2.3, we can compare the distributions of observed angle changes from simulations with those predicted theoretically. We simulate from the VJP model for a known pattern of hidden states and assume no measurement noise is present. Performing these checks for hidden states of the form 0, 1, we obtain results as shown in Figure 4.6a); there is excellent agreement between simulation and theory which verifies our theoretical results. For hidden states of the form 1, 0, we must condition on the value of the previous angle change. We demonstrate agreement between the theory and simulations, for different values of the previous angle change, in Figure 4.6b). Similarly, we show corresponding results for hidden states of the form 1, 1 in Figure 4.6c).

Figure 4.6 shows a comparison between simulated results and theoretical predic-

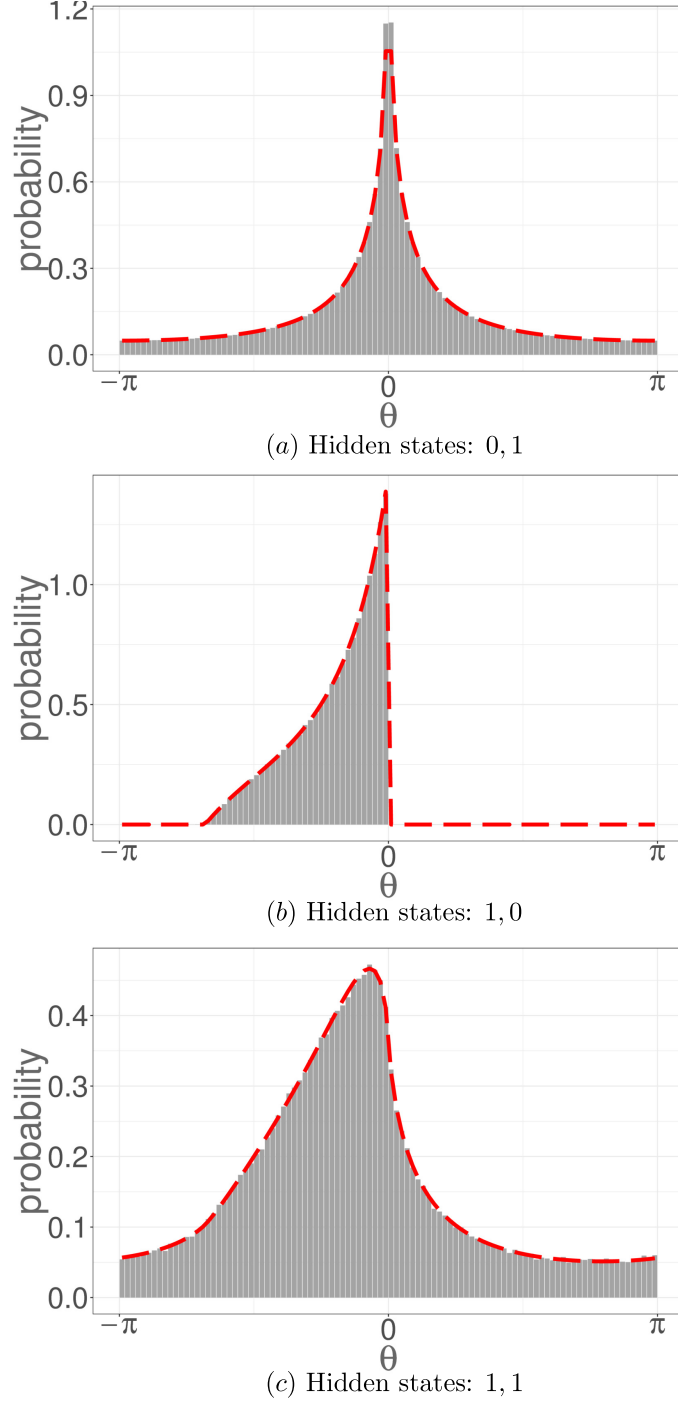


Figure 4.6. Comparison between simulated results and theoretical predictions of the observed angle change for hidden states of the form 0, 1 in a), 1, 0 in b), and 1, 1 in c). The simulated results are shown by the histogram and the theoretical prediction for the observed angle change distribution is shown as the red dashed line. For both b) and c), we have conditioned on an observed angle change in the previous time interval of 0.1 rad, and for c) we also conditioned on an observed angle change prior to that of -1.0 rad. To generate these results, we used $N = 10^7$ simulated trajectories with running speed $c = 50 \mu\text{ms}^{-1}$, uniform reorientation kernel, reorientation rate $\lambda = 0.2 \text{ s}^{-1}$ and time discretization $\Delta t = 1 \text{ s}$.

tions of the observed angle change for hidden states of the form 0, 1 in a), 1, 0 in b), and 1, 1 in c). The simulated results are shown by the histogram and the theoretical prediction for the observed angle change distribution is shown as the red dashed line. For both b) and c), we have conditioned on an observed angle change in the previous time interval of 0.1 rad, and for c) we also conditioned on an observed angle change prior to that of -1.0 rad. To generate these results, we used $N = 10^7$ simulated trajectories with running speed $c = 50 \mu\text{ms}^{-1}$, uniform reorientation kernel, reorientation rate $\lambda = 0.2 \text{ s}^{-1}$ and time discretization $\Delta t = 1 \text{ s}$.

4.2.5 Particle MCMC algorithm

The pMCMC algorithm used in this work is given in Algorithm 4 for an observed dataset $\mathbf{y} = \{y_k \mid k = 1, 2, \dots, T\}$ (see Section 2.2.7 for further detail). We use a Metropolis-Hastings MCMC algorithm [59], in which we need to evaluate the likelihood, $p(\mathbf{y}|\theta)$, in calculating the acceptance probability for the proposed move. We replace the likelihood with an unbiased estimate of the likelihood, $\hat{p}(\mathbf{y}|\theta)$, obtained from a particle filter. Details of the bootstrap particle filter algorithm [79] used within the pMCMC algorithm are given in Algorithm 1 (see Section 2.1.5 for further detail).

4.2.6 Implementation of MCMC methods

We use a Metropolis-Hastings algorithm to run the MCMC algorithm with a bootstrap particle filter [79] using 400 particles to provide an estimate of the likelihood. As a proposal distribution, we use the kernel $K(., \theta) \sim \mathcal{N}(\theta, \Sigma)$, where

$$\Sigma = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.05 \end{bmatrix}.$$

We run the Markov chain for $N = 50,000$ steps with thinning of $m = 2$ which gives a minimum effective sample size of $n_{\text{eff}} = 743$ for the slowest chain to converge, corresponding to the smallest value of Δt . We demonstrate the convergence of our Markov chains by showing traceplots and autocorrelation functions in Figure 4.7, for data generated with parameters $\lambda = 0.2 \text{ s}^{-1}$, $\Delta t = 0.25 \text{ s}$, $\sigma = 0.04 \text{ rad}$, as for the first posterior shown in Figure 4.10a). Similar results are seen in sampling for other posterior distributions shown.

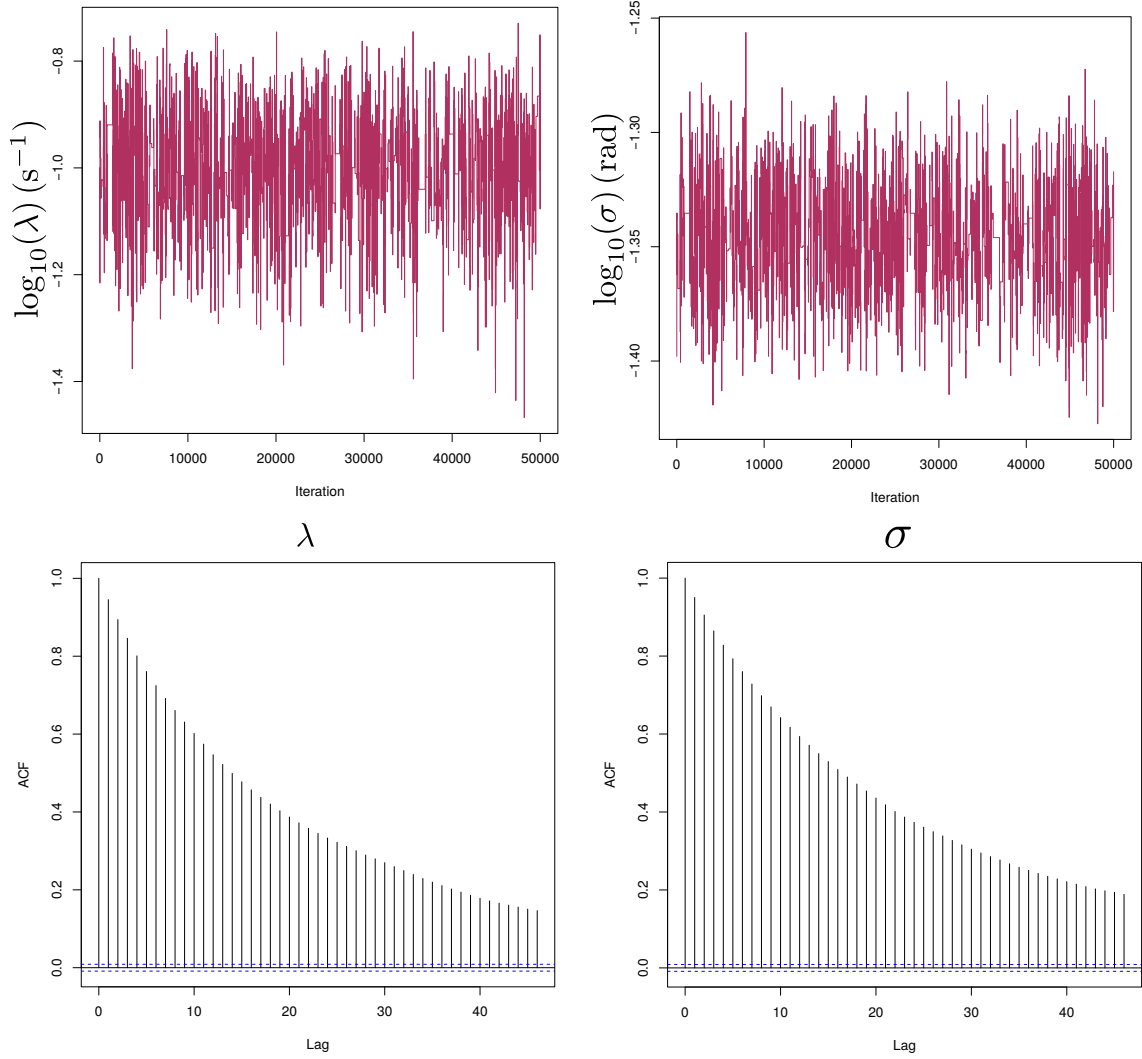


Figure 4.7. Traceplots and autocorrelation functions for MCMC chains to analyse convergence are shown for data generated with parameters $\lambda = 0.2 \text{ s}^{-1}$, $\Delta t = 0.25 \text{ s}$, $\sigma = 0.04 \text{ rad}$, as for the first posterior shown in Figure 4.10a). Similar results are seen in sampling for other posterior distributions shown.

We choose the number of particles in the filter by considering the variance in estimating the log likelihood using the particle filter at the true values of the parameters used to generate the synthetic data, and balancing this with the time needed to run the particle filter to obtain a single estimate. The number of particles used in the particle filter decreases the variance of estimates of the log-likelihood; however, a higher computational cost is required to use more particles, as shown in Figure 4.8. In a), we show how the distribution of the log-likelihood estimates varies (provided the filter does not become degenerate) as we change the number of particles. We use 1000 runs of the particle filter with the specified number of particles and estimate the log-likelihood at the true value of the parameters ($\lambda = 0.2$ s and $\sigma = 0.08$ rad) used to generate a synthetic dataset. In b), we illustrate the variance in the (non-degenerate) log-likelihood estimates, and the mean time to obtain a single estimate. A moderate increase in the compute time to run the particle filter offers substantial decrease in the variance of the log-likelihood estimates. To strike a reasonable balance, we use 400 particles to generate the results presented in this work.

To further tune the number of particles for optimal efficiency of the pMCMC algorithm, the recommendations in Sherlock et al. [250] and Pitt et al. [251] could be employed. The prior used in this work is uniform on the log of the parameters over the intervals $[-1.70, 1.30]$ and $[-5, 1]$ for λ and σ , respectively.

4.2.7 Approximate Bayesian computation

An alternative method commonly used for parameter estimation for models with intractable likelihoods is approximate Bayesian computation, or ABC [14, 15], as explained in detail in Section 2.2.1. The ABC method used in this chapter is the ABC rejection sampling algorithm given in Algorithm 2, and will show a comparison between application of pMCMC and ABC for parameter estimation of a VJP model based on time series data.

In practice, the choice of summary statistics can have a strong effect on the ef-

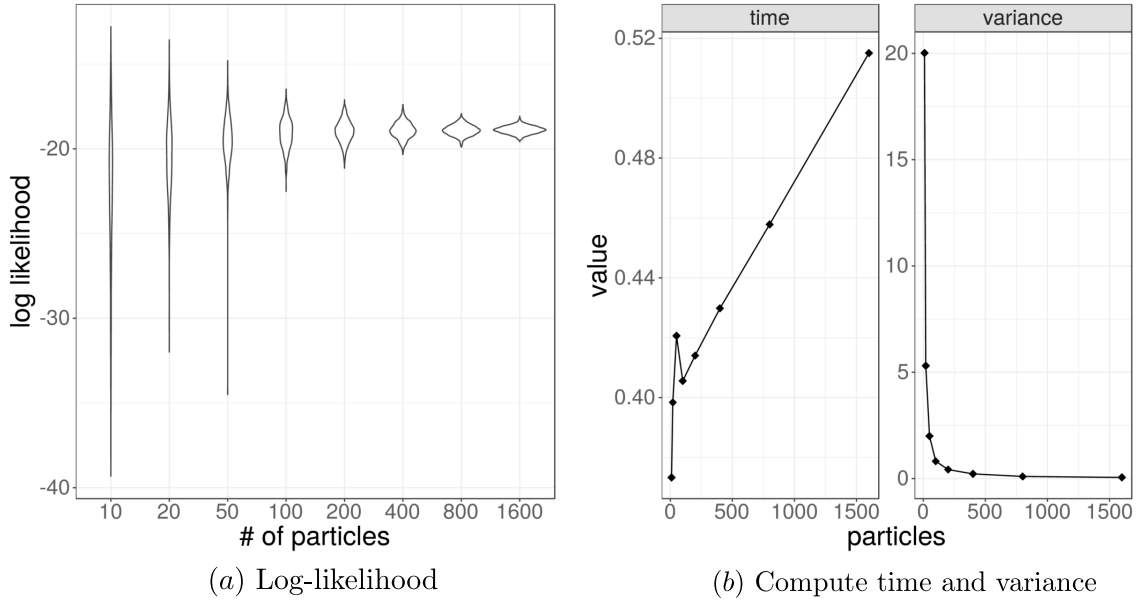


Figure 4.8. The number of particles used in the particle filter affects the variance of estimates of the log-likelihood. However, a higher computational cost is required to use more particles. In a), we show how the distribution of the log-likelihood estimates varies (provided the filter does not become degenerate) as we change the number of particles. We use 1000 runs of the particle filter with the specified number of particles and estimate the log-likelihood at the true value of the parameters ($\lambda = 0.2$ s and $\sigma = 0.08$ rad) used to generate a synthetic dataset. In b), we illustrate the variance in the (nondegenerate) log-likelihood estimates, and the mean time to obtain a single estimate. A moderate increase in the compute time to run the particle filter offers substantial decrease in the variance of the log-likelihood estimates. To strike a reasonable balance, we use 400 particles to generate the results presented in this work.

efficiency of the ABC algorithm. Here, we perform ABC with three different choices of summary statistic. First, we consider using the full data, a vector of the noisy observed angle changes, z_i , $i = 1, \dots, T$. Second, we use a simple threshold-based summary statistic which produces a count of the number of observed angle changes with magnitude above a certain threshold, h . That is $s(\mathbf{z}) = \sum_{i=1}^T \mathbb{1}_{|z_i| > h}$. This provides an intuitive one-dimensional summary statistic. Finally, we use a transition matrix as a summary statistic of the data, an approach described by Jones et al. [238]. A transition matrix allows us to summarize time series data via a two-dimensional binning of the observed angle changes, based on the current observed

angle change and the previous observed angle change. We bin the observed angle change time series data into an n by n matrix, where $n = 5$. This summary statistic provides a much lower dimensional summary of the data for small values of Δt , although for large values of Δt we may end up with higher dimensional (sparse) data compared to the full time series data, as in this case there will be fewer than n^2 observations. The distance function, $d(s(x), s(y))$, used also plays a role in the parameter estimation possible via ABC, and has been investigated in other work [2, 111, 238, 252].

In using Algorithm 2, we generate $N = 1,000,000$ synthetic datasets based on parameters sampled from the prior, calculate the Euclidean distance between these synthetic datasets and our observed data, and select the parameters corresponding to the 0.1% of the datasets closest to the observed data. As for the pMCMC approach, we take a duration of time series $T_{\text{final}} = 64\text{s}$, and a prior uniform on the log of the parameters on the intervals $[-1.70, 1.30]$ and $[-5, 1]$ for λ and σ , respectively.

4.3 Results

We are able to investigate the effects of experimental design such as the discretization, Δt , of a time series on the estimated biophysical parameters, using the framework for inferring the parameters of a VJP described in the Section 4.2. We demonstrate the effects of restrictions imposed by experimental constraints on a trade-off between discretization in time versus measurement noise. Our results can be used to provide guidance for experimental design choices. As we vary experimental design hyperparameters, such as Δt , we will produce a separate posterior distribution for each of the model parameters, λ and σ , for each value of the hyperparameter. This approach allows us to illustrate directly the effects of the experimental design hyperparameter on the inferred posterior distributions.

4.3.1 Increasing Δt results in an abrupt breakdown in the posterior

We first assume that the magnitude of the noise on the observed angle, σ , is fixed. We vary the sampling frequency used to collect the data (corresponding to using different values of Δt). We generate observed (*in silico*) data by simulating one single trajectory directly from the VJP model. We discretize this trajectory to generate observed angle changes with a temporal resolution of Δt , and add independent Gaussian noise with zero mean and variance σ^2 to these observations, to represent measurement noise. Using datasets generated from the same simulated path discretized at different temporal resolutions, Δt , as shown in Figure 4.9, we infer the posterior distribution for the reorientation rate, λ , of the VJP model and the magnitude of the measurement noise, σ , via pMCMC.

The results of using pMCMC to infer model parameters are shown in Figure 4.10 for $\Delta t = 1/8, 1/4, 1/2, 1, 2, 4, 8$ seconds, with noise of unknown magnitude σ , where $\sigma = 0.04$ rad. The estimated posterior distributions are very similar for small Δt , but we observe an abrupt breakdown in the quality of the posterior distributions obtained as Δt is increased. This breakdown in posterior quality arises for values of $\Delta t \geq 2$ s, as multiple reorientation events in a time interval start to become more common (see also Figure 4.11). Provided that Δt and λ are such that the probability of multiple reorientations in a time interval is small, we obtain accurate estimates of the joint posterior distribution for the model parameters.

4.3.2 Increasing σ increases the variance of the posterior for λ

To investigate sensitivity of the posterior to measurement noise, we now fix the discretization, Δt , and vary the measurement noise amplitude, σ , used to create each dataset. Applying the same analysis as in the previous subsection, for a fixed

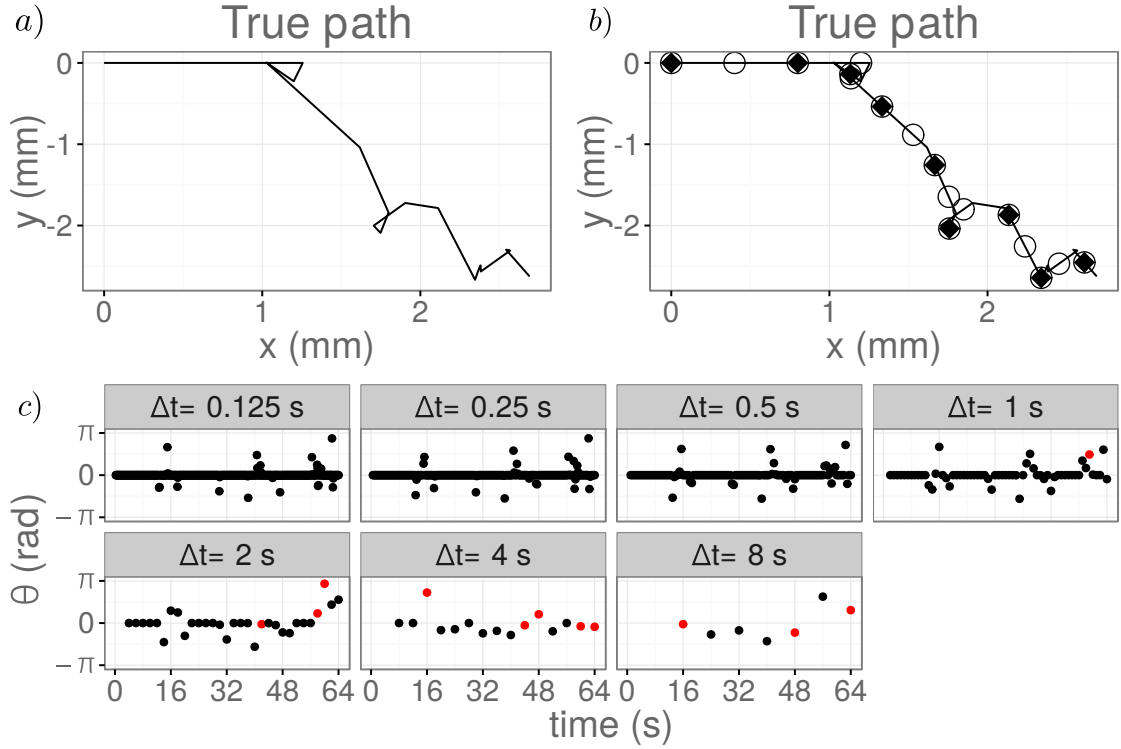


Figure 4.9. The data used in the pMCMC inference discretized at differing resolutions. The true trajectory is shown without discretization in a). The same path is shown in b) with circle markers to show the observed positions with $\Delta t = 4$ s and filled triangular markers to show observed positions with $\Delta t = 8$ s. The corresponding observed angle changes are shown in c) for different values of Δt . Observations corresponding to multiple reorientations in a single time interval are highlighted in red. Parameters used in generating these data were a reorientation rate of $\lambda = 0.2 \text{ s}^{-1}$, a run speed of $c = 50 \mu\text{ms}^{-1}$ and a total duration of observation of $T_{\text{final}} = 64$ s. A circular uniform distribution was used for the reorientation kernel. Observations are shown without measurement noise.

value of $\Delta t = 1$ s, we obtain posterior distributions for the reorientation rate, λ , and measurement noise, σ , as shown in Figure 4.10. We find that increasing the noise amplitude, σ , increases the variance in the posterior distribution obtained for the reorientation rate, λ . In addition, we are able to accurately estimate the value of σ used to generate the datasets.

We note that the presence of noise in the datasets results in a bias towards smaller estimates of the reorientation rate, λ . This can be explained intuitively by reasoning that some reorientations leading to very small observed angle changes are

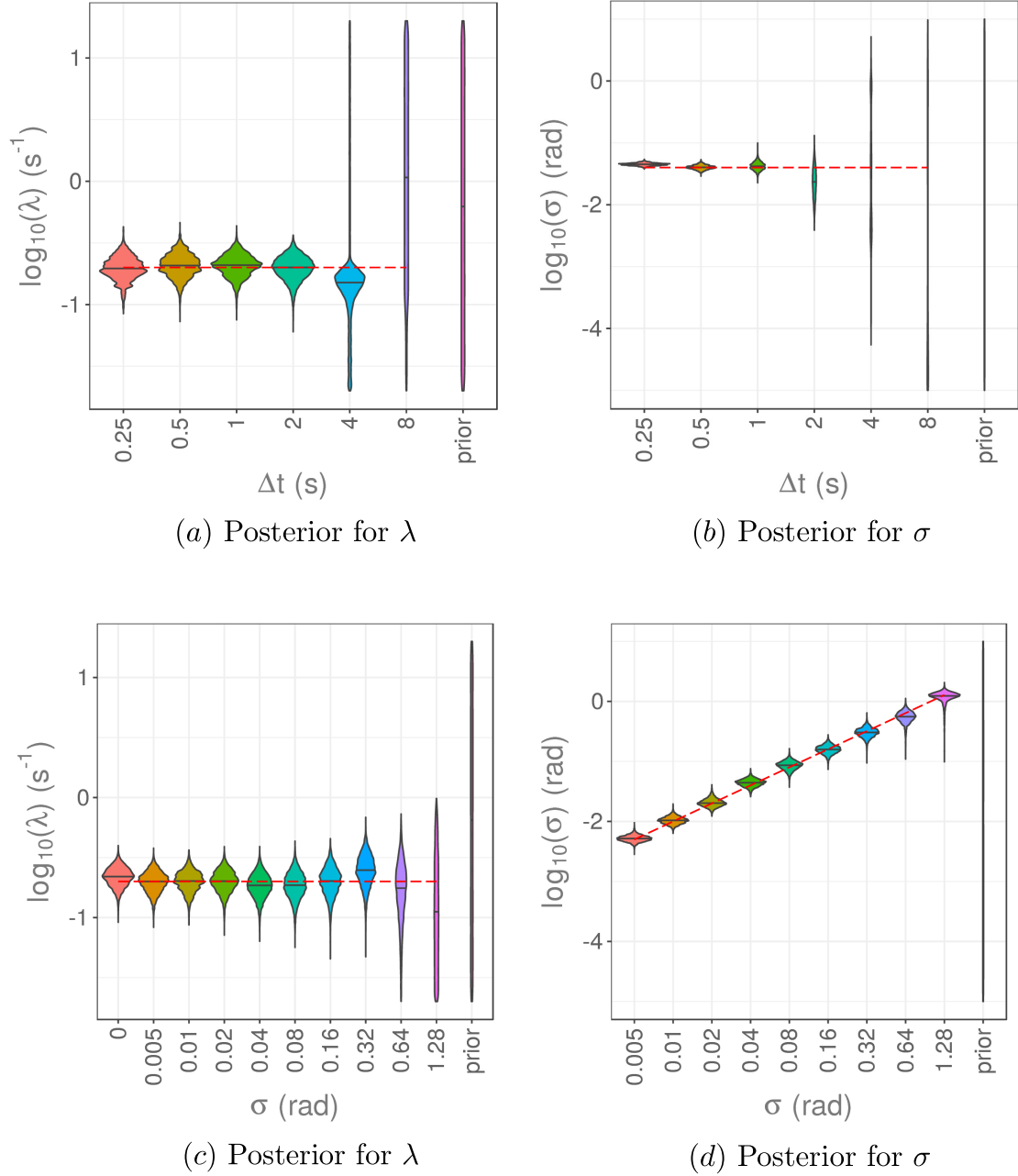


Figure 4.10. Results of parameter estimation via pMCMC for the reorientation rate, λ , and measurement noise, σ . Data collected with different values of Δt and fixed measurement noise, $\sigma = 0.04 \text{ rad}$, were used in a) and b). Data generated with a fixed value of $\Delta t = 1 \text{ s}$ and different values of the measurement noise, σ , were used in c) and d). The marginal posterior distributions for λ are given in a) and c), while the marginal posterior distributions for σ are shown in b) and d). The red dashed lines indicate the true values of parameters used in simulation of the datasets. A reorientation rate of $\lambda = 0.2 \text{ s}^{-1}$ was used throughout.

mistaken for noise as σ increases.

4.3.3 Probability of hidden state sequences

The probability of sequences of hidden states varies as Δt increases with reorientation rate fixed at $\lambda = 0.2 \text{ s}^{-1}$. For large values of Δt , the assumptions of the model start to break down as multiple reorientations appear within a single time interval, as shown in Figure 4.11.

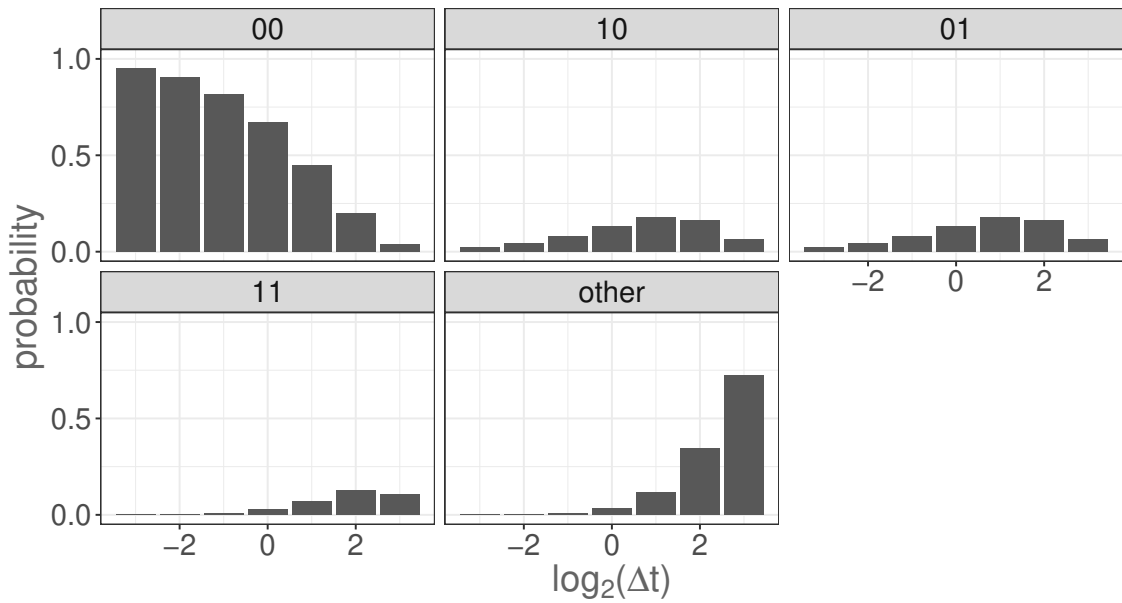


Figure 4.11. The probability of sequences of hidden states as Δt varies with reorientation rate $\lambda = 0.2 \text{ s}^{-1}$. For large values of Δt , the assumptions of the model start to break down as multiple reorientations appear within a single time interval.

4.3.4 More data provides better estimates

Let T_{final} be the total duration of our observations of the system. For a fixed value of the time discretization, Δt , varying T_{final} is equivalent to gathering a bigger dataset. We fix $\Delta t = 1 \text{ s}$ and allow T_{final} to vary so that we can consider the effects of collecting more data. In practice, collecting more data experimentally may come at a cost. Quantifying the benefits of collecting more data with an *in silico* model can

aid decisions about whether it is worthwhile to collect a larger dataset. Posteriors for the reorientation rate, λ , are shown in Figure 4.12.

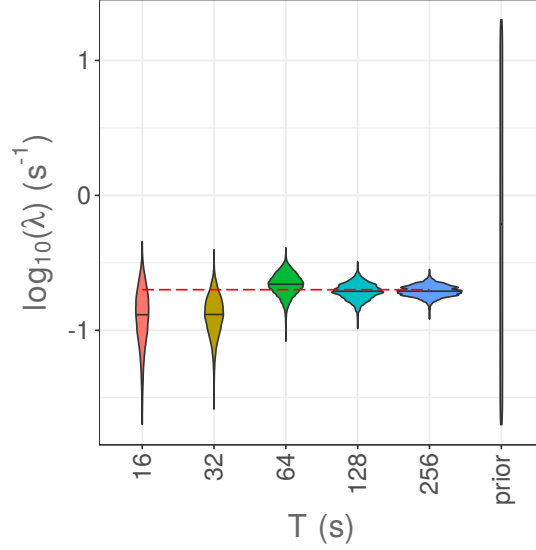


Figure 4.12. Results of parameter estimation via pMCMC for the reorientation rate, λ , with $\Delta t = 1$ s and $\sigma = 0$ rad, whilst varying T_{final} to give different sized datasets.

It is clear that larger datasets result in less bias and less variance for estimates of the reorientation rate, λ . However, we note that running the particle filter within the pMCMC algorithm becomes much more computationally intensive as the size of the dataset increases, scaling as $\mathcal{O}(T)$ as the size of the dataset increases.

4.3.5 Experimental constraints

We investigate the implications of these results for experimental design by considering an imaging experiment to observe the position of a particle of interest (for example, a bacterium) at regular time points. The behaviour of our system of interest happens over a certain timescale inherent to the biological process, so we fix the total duration for the imaging experiment, T_{final} , based on this timescale. We consider how best to choose the time between successive observations, Δt , given the restriction of a fixed photon budget. That is, we assume that the biological sample

can only be exposed to a fixed number of photons before phototoxicity or photo-bleaching significantly reduce the quality of, or destroy, any further potential data. We assume that the sample is only exposed to photons during imaging. The results of Zhao et al. [253] suggest that the signal to noise ratio (SNR) is proportional to the square root of the time between successive frames, Δt , giving a relationship of the form

$$\text{SNR} = \kappa \sqrt{\Delta t}, \quad (4.27)$$

where κ is a constant that depends on the imaging set up, but not other experimental design choices. Assuming that for our model the noise is of magnitude σ , we find an inverse square relationship between σ and Δt , such that

$$\sigma = \frac{K}{\sqrt{\Delta t}}, \quad (4.28)$$

for a new constant K which is κ times the average angle change.

To investigate how to choose Δt given a fixed photon budget, we set a value of the proportionality constant and vary σ and Δt according to this relationship. We take proportionality constants $K = 0.08$ and $K = 0.8$ in Figure 4.13. A larger value of the proportionality constant, K , corresponds to worse imaging conditions, in that for a fixed value of Δt , the noise in the images obtained will be greater. Therefore we expect our inference method to perform worse for a larger value of K . We then ask the question, for a given value of K , how should we choose Δt to improve the parameter estimation process? Our results in Figure 4.13 suggest that the value of Δt should be taken as small as possible, even if this increases the noise present in the data.

4.3.6 Comparison to ABC

A common approach for mathematical and computational models where the likelihood is intractable is to apply ABC for parameter inference [14, 15]. Although ABC

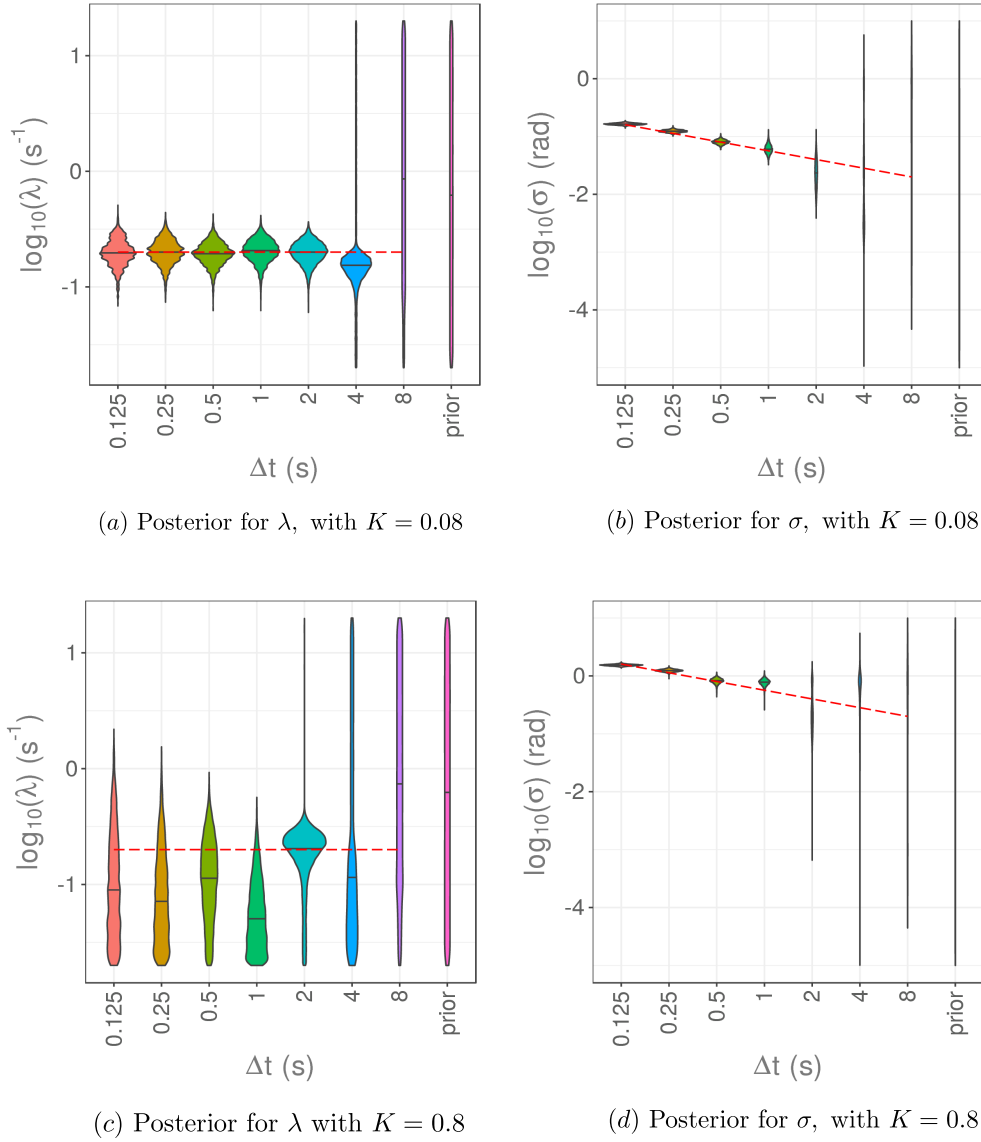


Figure 4.13. Results of parameter estimation via pMCMC for the reorientation rate, λ , and measurement noise, σ , using data generated with a fixed value of Δt and different values of the measurement noise, σ . We have varied both σ and Δt with an inverse square root relationship between these, as in Equation(4.28). The posterior distributions for λ are given in a) and c), while the posterior distributions for σ are shown in b) and d). The proportionality constant is $K = 0.08$ for a) and b), and $K = 0.8$ for c) and d).

produces samples from an approximate posterior, rather than the exact posterior distribution, it is intuitively simple to understand and implement. ABC methods have been applied to parameter estimation for biased, persistent random walk models very similar to our VJP model [237, 238, 254, 255]. For these methods, the choice

of which summary statistics to use has notable effects on the approximate posterior distributions obtained via ABC. We consider three different summary statistic choices: the full time series of observed angle changes, a threshold-based summary statistic, and a transition matrix summary statistic¹.

We compare the quality of resulting posterior distributions obtainable with ABC to those from pMCMC. Our results, shown in Figure 4.14, indicate that the choice of summary statistic has a substantial effect on the inferred posterior distribution. When the dimensionality of the data observed is high, ABC performs poorly at approximating the posterior for the parameters of our VJP, which we were able to sample exactly using pMCMC. For the full time series summary statistic, the dimensionality is high (up to 511 dimensions for $\Delta t = 0.125$ s), meaning that the approximation we obtain to the posterior is very poor (see Figure 4.14a) and b)). Since reorientation events are rare, particularly for small Δt , a greater fraction of the distance between simulated datasets and observed data is accounted for by the observation noise. ABC excludes part of the parameter space considered in the prior, but does poorly at identifying the reorientation rate, λ , since reorientations are rare events.

The threshold summary statistic is effective at identifying the number of reorientation events and provides a reasonable approximation of the marginal posterior for λ . However, it offers very little information about the observation noise, σ , except in relation to the threshold chosen. As a result the approximate marginal posteriors for σ are poor (see Figure 4.14c) and d)). Another summary statistic more informative about σ could be used in addition here to improve results.

When using the transition matrix summary statistic, the approximate posteriors for small Δt provide better estimates of the true reorientation rate and are much closer to the true posterior, as shown in Figure 4.14e) and f). The transition matrix

¹The transition matrix summary statistic consists of a two-dimensional discretization of the empirical frequency of observed angle changes, based on the current observed angle change and the previous observed angle change. Further details are described in Section 4.2.7 and in Jones et al. [238]

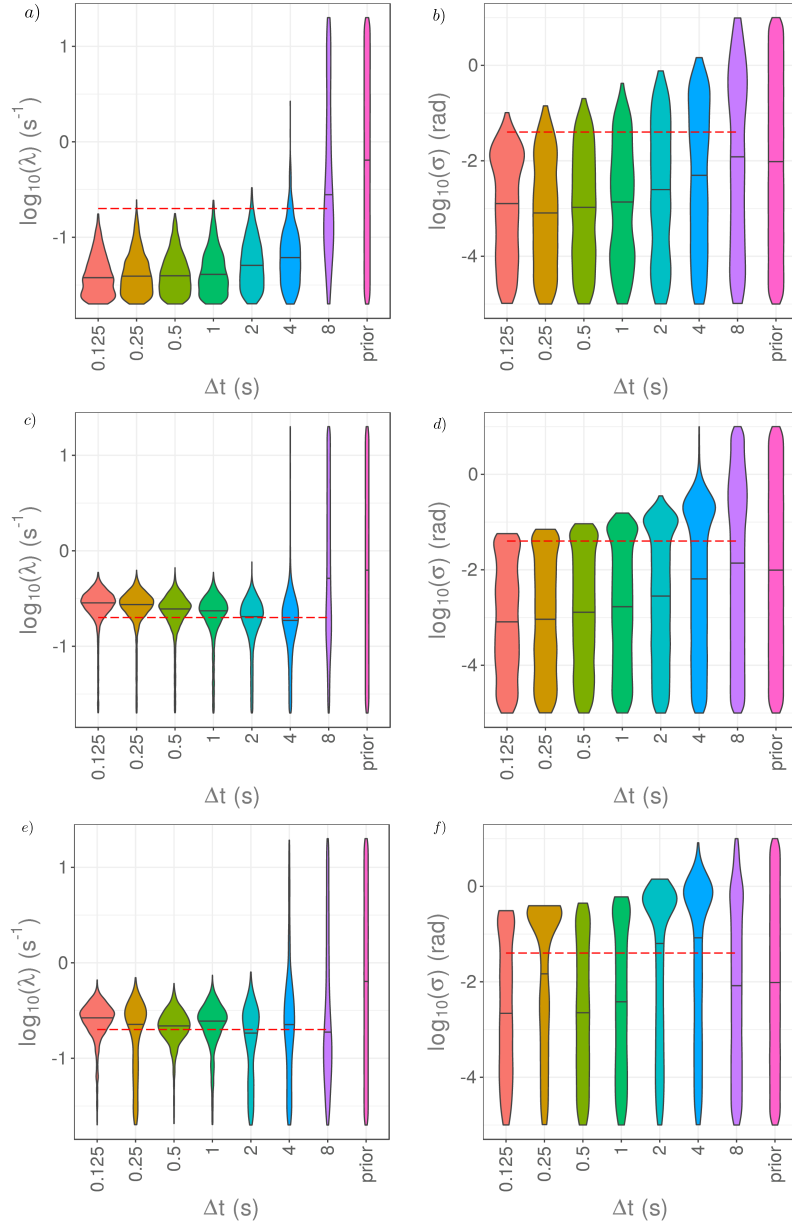


Figure 4.14. Parameter estimation via ABC rejection sampling for the reorientation rate, λ , in a), c), and e) and for the measurement noise, σ , in b), d), and f) using different summary statistics and data collected with noise of magnitude $\sigma = 0.04$ rad, for different values of Δt . $N = 1,000,000$ datasets were generated and parameter samples corresponding to the closest 0.1% of the datasets were retained to give the approximate posterior. As summary statistics, we take the full time series of observed angle changes in a) and b). In c) and d), we use a simple threshold summary statistic counting the number of observed angle changes with magnitude above a threshold, here taken as $h = 0.1$ rad. In d) and e), we use a transition matrix summary statistic using $n = 5$ bins to discretize the observed angle change.

is able to capture the distribution of the angle changes and dependence on recent history, whilst also reducing the dimensionality of the data. However, the transition matrix is unable to distinguish angle changes smaller than the resolution of the discretization, $2\pi/n$, and so offers limited information about the measurement noise, σ . This issue could be mitigated by increasing the number of bins used, n , to give a finer discretization of the observed angle changes (which would require sufficient data), or by targeting the noise with an additional summary statistic.

We still notice a deterioration in the quality of the approximate posterior distributions as Δt increases when using ABC for each choice of summary statistic even though there are no explicit assumptions made about multiple reorientations within a time interval, as was the case when using pMCMC. This suggests that the lack of information content about the parameters may be a property of the data, rather than the inference method. For large values of Δt , the data contains limited information about the model parameters, particularly the reorientation rate, λ .

To obtain potentially improved results for inference with ABC for this type of data (time series observations of angle changes), we could consider a more systematic choice of summary statistics [35–37] or a more efficient version of the ABC algorithm, such as ABC-SMC [16, 106], which applies sequential Monte Carlo methods to generate a sequence of approximations to the posterior distribution, using the previous posterior distribution to propose parameters for the next approximation. Other improvements could be possible by applying a regression adjustment, via linear regression [15] or using nonlinear regression techniques such as with a neural network [98].

4.3.7 The computational cost of pMCMC and ABC per effective sample size.

To further investigate a comparison between pMCMC and ABC in this context, we quantify the computational cost of each method. The computational cost of

parameter estimation with pMCMC and ABC per effective sample size depends on the sampling frequency of the data, Δt . We quantify here a comparison between the computational cost of the pMCMC and ABC methods for parameter estimation when varying Δt . Datasets were generated with $\sigma = 0.04$ rad as in Figure 4.10a). The cost to produce a sample via pMCMC increases as Δt decreases, as shown by the grey squares. The cost for ABC remains approximately constant with Δt as we simulate data from the model a fixed number of times, N . The acceptance rate, α , in ABC affects the sample size we produce for a fixed number of simulations, N . For the results in Figure 4.14, an acceptance rate of $\alpha = 0.1\%$ was used. This gives a cost per sample lower than for pMCMC, shown by the blue circles and triangles. For a smaller acceptance rate, $\alpha = 0.01\%$ (shown by red circles and triangles), the cost per sample is much higher and the plots of the posterior are unchanged compared to those for $\alpha = 0.1\%$. The results for pMCMC are given by squares, ABC with transition matrix summary statistics are shown as circles and ABC with time series summary statistics are shown as triangles. We note that the computational cost depends strongly on the problem considered and the implementation used.

4.3.8 Model misspecification

In general, in applying a model to a real world dataset, any model that we choose will be an approximation of the true data generating process. Before applying our inference methods to real world data, where we will fit to a model that is but an approximation to the true biological process, it is important to check the robustness of our method to fit a misspecified model. We investigate this robustness computationally by considering a misspecified model which should be no longer misspecified in an appropriate limit. It has previously been shown by Frazier et al. [100] that inference with ABC on a misspecified model can concentrate posterior mass around different pseudo-true parameter values depending on the version of ABC used, where the pseudo-true parameter values are those closest within the

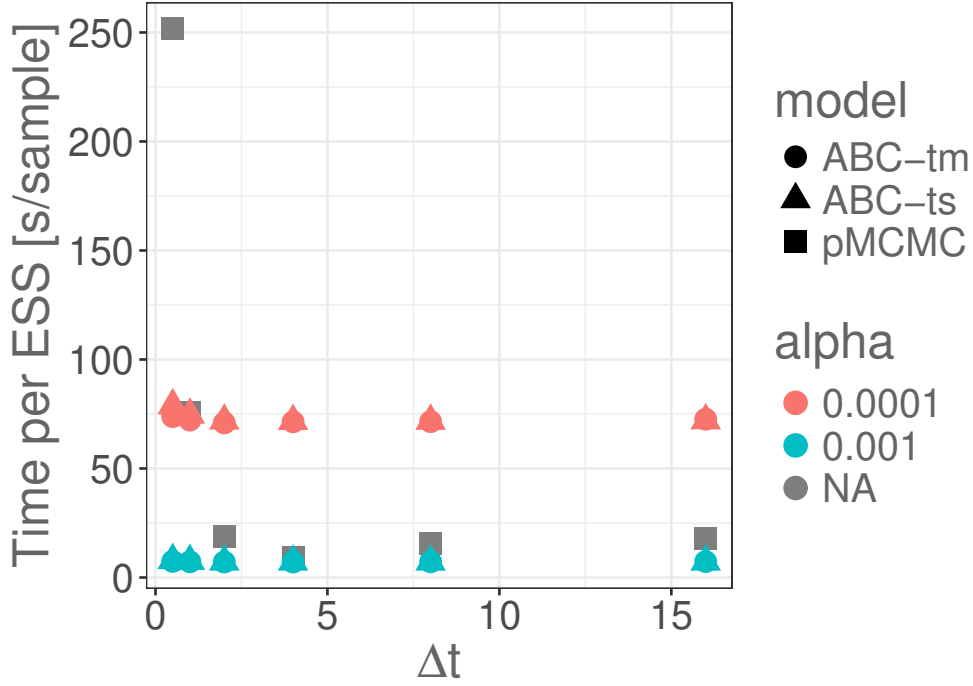


Figure 4.15. The computational cost of parameter estimation with pMCMC and ABC per effective sample size depends on Δt . We quantify here a comparison between the computational cost of the pMCMC and ABC methods for parameter estimation when varying Δt . Datasets were generated with $\sigma = 0.04$ rad as in Figure 4.10a). The cost to produce a sample via pMCMC increases as Δt decreases, as shown by the grey squares. The cost for ABC remains approximately constant with Δt as we simulate data from the model a fixed number of times, N . The acceptance rate, α , in ABC affects the sample size we produce for a fixed number of simulations, N . For the results in Figure 4.14, an acceptance rate of $\alpha = 0.1\%$ was used. This gives a cost per sample lower than for pMCMC, shown by the blue circles and triangles. For a smaller acceptance rate, $\alpha = 0.01\%$ (shown by the red circles and triangles), the cost per sample is much higher and the plots of the posterior are unchanged compared to those for $\alpha = 0.1\%$. The results for pMCMC are given by squares, ABC with transition matrix summary statistics are shown as circles and ABC with time series summary statistics are shown as triangles. We note that the computational cost depends strongly on the problem considered and the implementation used.

misspecified model ².

²A model is misspecified if

$$\inf_{\theta \in \Theta} d(\mathbf{y}, f(\theta)) > 0,$$

where $f(\theta)$ is a generative model dependent of parameters θ . In the case of a misspecified model, $f(\theta)$, with observed data $\mathbf{y}$, the pseudo-true parameter value, θ^* , is defined as

$$\theta^* = \arg \inf_{\theta \in \Theta} d(\mathbf{y}, f(\theta)).$$

Here, we demonstrate the robustness of our inference framework, using the pMCMC algorithm within a hidden states framework, to give relevant estimates for a misspecified model. We generate synthetic datasets using a wrapped normal reorientation kernel with dispersion parameter γ , but choose to estimate the model parameters with a model different to the data generating process by assuming a uniform reorientation kernel. As previously, we attempt to infer posterior distributions for the reorientation rate, λ , and the measurement noise, σ . The effect of misspecifying the model will be that some of the transmission probabilities used in the particle filter estimate of the likelihood of model parameters given the data will be wrong (see Figure 4.16).

In Figure 4.16, we show a comparison between the assumed distribution of observed angle changes and simulated distributions when using a misspecified model for the reorientation kernel. We simulate $N = 10^7$ angle changes using a wrapped normal reorientation kernel with dispersion parameter γ given a certain sequence of hidden states (0,1 in a), d), g); 1,0 in b), e), h); 1,1 in c), f), i)) and show a grey histogram of the simulated observed angle changes. To demonstrate the misspecification in the emission probabilities, we plot the assumed theoretical distribution of the observed angle changes as a red dashed line, based on assuming that the reorientation kernel is a uniform distribution. The model is more misspecified for a smaller value of the dispersion parameter γ . As in Figure 4.6, we have conditioned on an observed angle change in the previous time interval of 0.1 rad for b), c), e), f), h), and i). For c), f), and i), we also conditioned on an observed angle change prior to that of -1.0 rad. We used a dispersion parameter, γ , in the reorientation kernel of $\gamma = 0.4$ for a), b), and c), $\gamma = 1$ for d), e), and f), and $\gamma = 6.4$ for g), h), and i).

This error in the misspecified emission probabilities will give rise to some bias in our estimates of the likelihood within the particle filter (as demonstrated by Figure

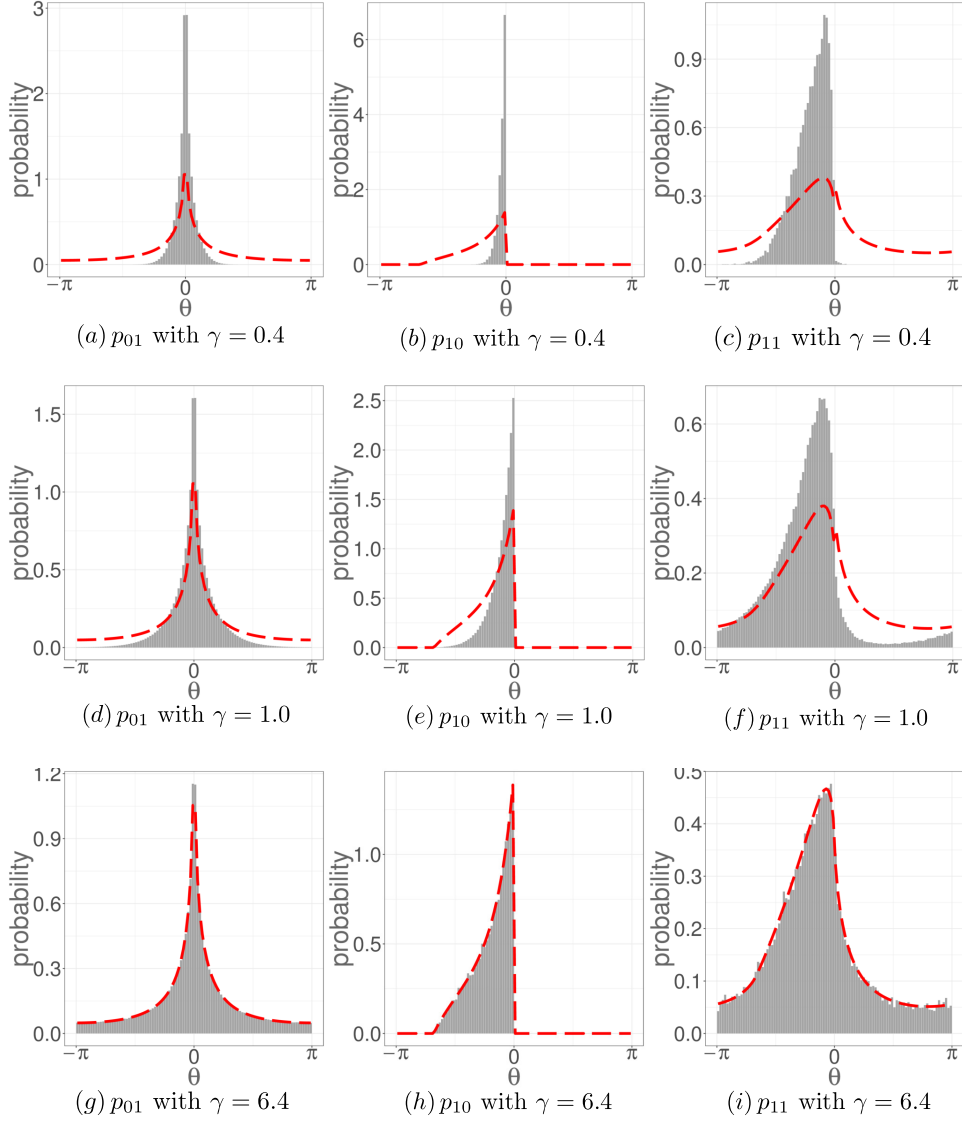


Figure 4.16. Comparison between assumed distribution of observed angle changes and simulated distributions when using a misspecified model for the reorientation kernel. We simulate $N = 10^7$ angle changes using a wrapped normal reorientation kernel with dispersion parameter γ given a certain sequence of hidden states (0,1 in a), d), g); 1,0 in b), e), h); 1,1 in c), f), i)) and show a grey histogram of the simulated observed angle changes. To demonstrate the misspecification in the emission probabilities, we plot the assumed theoretical distribution of the observed angle changes as a red dashed line, based on assuming that the reorientation kernel is a uniform distribution. The model is more misspecified for a smaller value of the dispersion parameter γ . As in Figure 4.6, we have conditioned on an observed angle change in the previous time interval of 0.1 rad for b), c), e), f), h), and i). For c), f), and i), we also conditioned on an observed angle change prior to that of -1.0 rad. We used a dispersion parameter, γ , in the reorientation kernel of $\gamma = 0.4$ for a), b), and c), $\gamma = 1$ for d), e), and f), and $\gamma = 6.4$ for g), h), and i).

4.17), and hence to some bias in our final approximation of the posterior. We will consider the effect of varying the dispersion parameter, γ , in the reorientation kernel used to generate the synthetic dataset, on the resulting posterior distribution for the model parameters. In the limit $\gamma \rightarrow \infty$, the wrapped normal distribution converges to the uniform distribution on $(-\pi, \pi]$, meaning that the model is no longer misspecified. The misspecification of the model is most pronounced for smaller values of γ .

We find, for a fixed value of the measurement noise, σ , that a reasonable approximation to the true posterior can be obtained for values of the dispersion, γ , larger than the measurement noise, σ . The approximate posterior distributions obtained are shown in Figure 4.18. When the dispersion, γ , is a similar magnitude to the measurement noise, which here is $\sigma = 0.04\text{rad}$, then reorientation events are frequently missed resulting in low estimates of the reorientation rate. For larger values of the dispersion parameter, we are able to robustly sample from the posterior distribution for λ and σ , despite notable differences in the misspecified reorientation kernel compared to the assumed uniform reorientation kernel. In this case, when a reorientation occurs, the observed angle changes are sufficiently different to the background measurement noise, that often it is interpreted as a reorientation event by the particle filter, despite error in the emission probabilities due to the model misspecification.

4.3.9 Application to RNA transport dataset

To demonstrate the effectiveness of our inference framework, we apply it here to a dataset of tracks obtained from imaging the transport of RNA-protein (RNP) complexes in a *Drosophila* oocyte. These complexes move on microtubules via molecular motors [44], as explained in Chapter 3. Occasionally, the complexes fall off a microtubule, and reattach on a different microtubule moving in a different direction. We neglect the diffusive stationary phase between falling off and reattaching on mi-

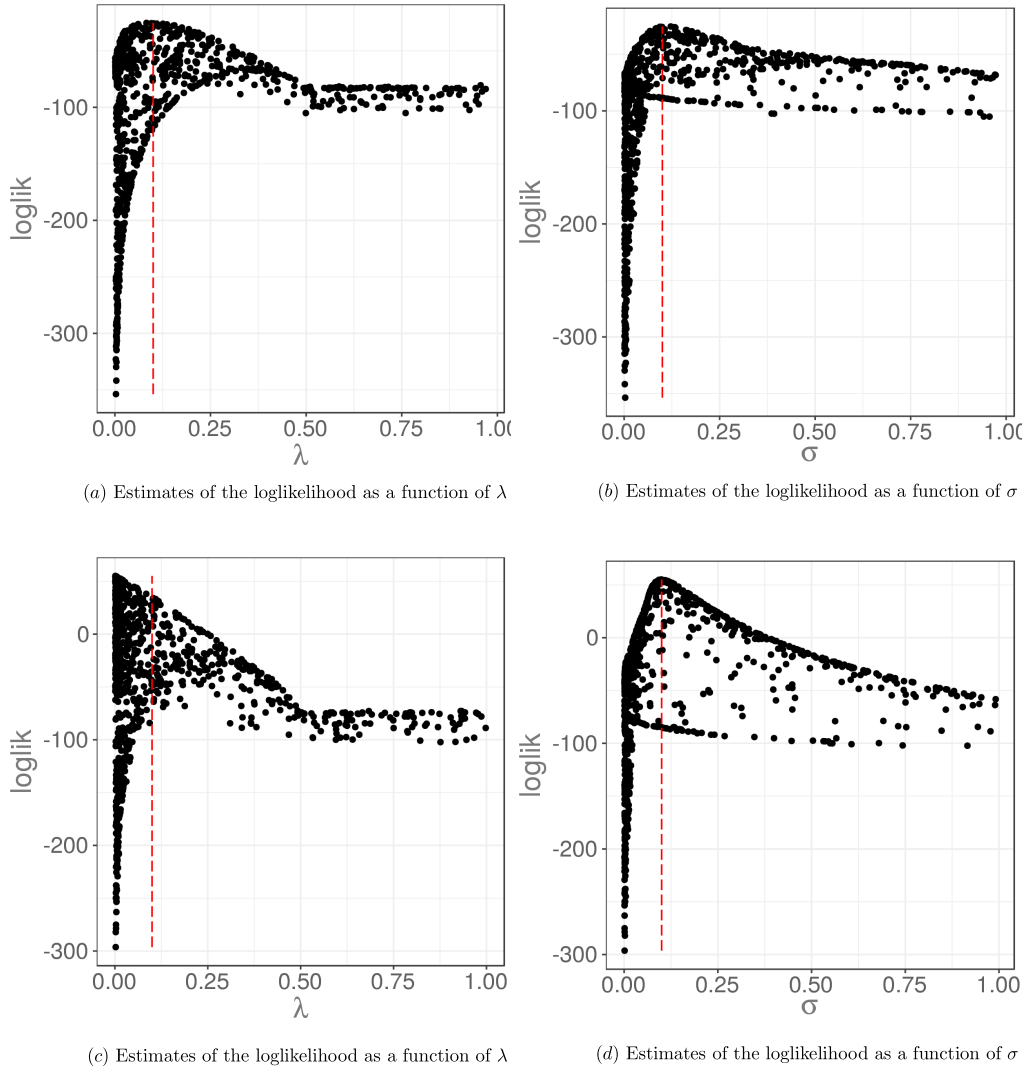


Figure 4.17. Estimates of the log likelihood via a particle filter for data generated with the true model and under a misspecified model. In a) and b), we show estimates of the log likelihood for different values of λ and σ based on a dataset simulated with a uniform reorientation kernel, and assuming this same reorientation kernel in estimating the log likelihood via the particle filter. The peak of the log likelihood coincides with the true values of the parameters shown by the red dashed line. In c) and d), we show estimates of the log likelihood for different values of λ and σ based on a dataset simulated with a wrapped normal reorientation kernel with dispersion $\gamma = 0.1$. However, we assume a uniform reorientation kernel in the particle filter. This model misspecification results in a slight shift in the peak of the log likelihood compared to the true parameter values used shown by the red dashed line. The log likelihood estimates will be biased in the case of the misspecified model.

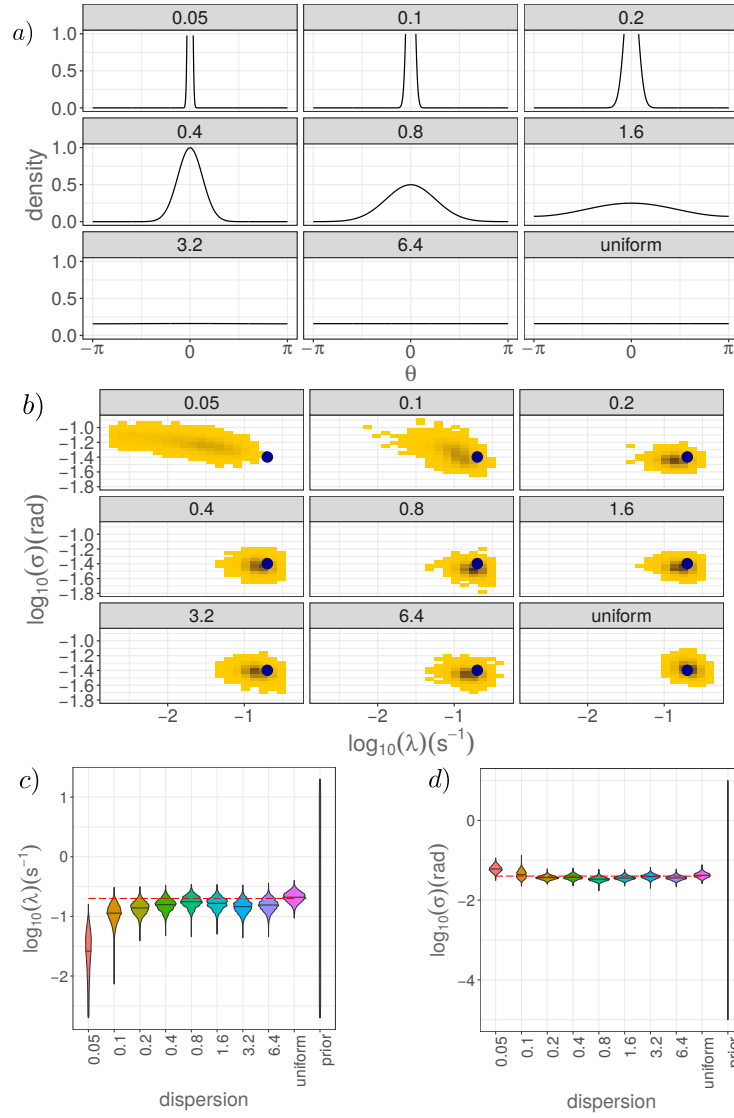


Figure 4.18. The posterior distribution obtained under model misspecification converges to the distribution obtained with the true model as the misspecified model approaches the true model. We assume a wrapped normal reorientation kernel with dispersion parameter γ and vary this dispersion parameter. For large γ , this tends in distribution towards a uniform reorientation kernel, as shown in a). The posterior distributions obtained from performing inference with data generated for different values of γ are shown in b), with the corresponding marginal distributions shown in c) and d). In b), the true parameter values used are shown by the dark blue circle. In c) and d), the true parameter values are given by the red dashed line. Parameters used to generate the synthetic datasets are $\lambda = 0.2 \text{ s}^{-1}$, $\sigma = 0.04 \text{ rad}$, $\Delta t = 0.25 \text{ s}$, $c = 50 \mu\text{ms}^{-1}$, and the details of the pMCMC are as in Figure 4.10.

crotochutes, meaning that we assume a running phase is followed immediately by another running phase. We apply our modelling framework, assuming an exponentially distributed running time distribution, $f_r(t) = \lambda \exp(-\lambda t)$, with constant running speed, c , and a uniform reorientation kernel, $f_\Phi(\phi) = 1/2\pi \mathbb{1}_{(-\pi, \pi]}(\phi)$.

We take 10 tracks of separate complexes moving in *Drosophila* nurse cells obtained from an *in vivo* imaging dataset of movements of staufen protein available in Zimyanin et al. [189]. Each track is short since the dataset is imaged in a single plane in the z direction, meaning that complexes move out of the frame of view frequently. The tracks used are shown in Figure 4.19a). We infer a subposterior distribution for each individual track and combine these together to produce a single posterior distribution combining data from different tracks. To combine the subposteriors, we use the consensus Monte Carlo algorithm of Scott et al. [256] developed for running MCMC on large datasets, via the implementation in the R [257] package `parallelMCMCcombine` [258]. The consensus Monte Carlo algorithm works by computing a weighted average of the Monte Carlo samples from each subposterior distribution, and is most appropriate when posterior distributions have a Gaussian-like shape. Application of the consensus Monte Carlo algorithm produces an approximate posterior distribution for the turning rate, λ , and measurement noise, σ , as shown in Figure 4.19b). We find 95% credible intervals for the turning rate λ of $[1.06, 1.58] \text{ s}^{-1}$ and for the measurement noise σ of $[0.27, 0.47] \text{ rad}$, respectively. Based on a VJP model corresponding to Brownian motion, the relation $\lambda = c^2/(2D)$ holds for a constant running speed c and diffusion constant D . The running speed in active transport for staufen RNPs is of the order $0.5 \mu\text{ms}^{-1}$ [189], and based on the size of the staufen protein, we can estimate its diffusion coefficient as $1 \mu\text{m}^2\text{s}^{-1}$ [259], which gives a similar order of magnitude for λ as the estimate obtained here. Evidently the dataset used here is very noisy, and a limitation of the dataset is that the transport process occurs in 3D, whereas the data and model are in 2D, so there may be limitations due to complexes moving out of plane, but nonetheless we are

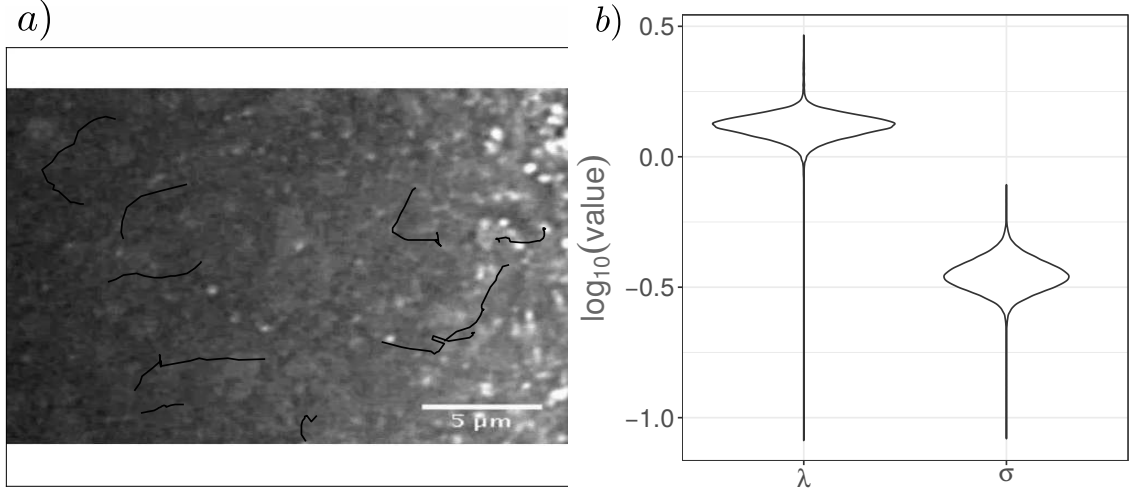


Figure 4.19. Posterior distribution for model parameters λ and σ for RNA transport in a *Drosophila* oocyte. The results are based on 10 tracks labelled manually and sampled at a resolution of $\Delta t = 0.478$ s, with subposteriors obtained for each track and combined to give a single posterior via Consensus Monte Carlo. The tracks are shown drawn on the first frame in a), although not all tracks start from the same time point. The marginal posterior distributions for the reorientation rate, λ , and the measurement noise, σ , are shown in b).

able to obtain estimates for the turning rate in a model of RNA transport.

4.4 Discussion

In this work, we have considered parameter estimation of a VJP model for biological transport and how insights from this parameter inference process can inform experimental design. We generated estimates of the posterior distribution for parameters of a VJP model based on noisy datasets collected at varying temporal resolutions. To perform this parameter inference, we used pMCMC and derived the appropriate emission probabilities. We observed an abrupt breakdown in the quality of the posteriors obtained when decreasing the temporal resolution. This transition corresponds to a breakdown in modelling assumptions underlying the derivation of the emission probabilities. For example, as Δt increases we see multiple reorientations in a time interval. These assumptions are necessary in the development of

our inference framework which relies on hidden states consisting of binary variables indicating whether or not a reorientation occurred in a time interval.

Increasing the magnitude of the noise in the data slowly decreased the quality of the median posterior estimates and increased the posterior variance. In general, better estimates were obtained when more data was available, either by decreasing Δt , or by increasing the total duration of the experimental observations, T_{final} . These results are intuitive, but the real benefit here is in quantifying the effects of choices of these experimental hyperparameters to enable researchers to make decisions about experimental design.

We also compared parameter estimation with pMCMC to that with ABC, and results suggest that different methods are appropriate in different situations, dependent on the data available. In particular, pMCMC performs well provided the assumptions made in deriving the emission probabilities hold, which is true for small Δt . For larger values of Δt , neither ABC nor pMCMC provide accurate samples from the full posterior distribution. The information about the model parameters present in the data for large Δt is limited. In terms of computational cost comparisons between ABC and pMCMC, an ABC rejection sampler is highly parallelizable over independent parameter samples [42] whereas pMCMC is less amenable to simple parallelization. Making use of this parallelization, our implementation of ABC rejection sampling was more computationally efficient than pMCMC. We show a quantitative comparison between the methods in Figure 4.15. We note that the computational cost is strongly problem and implementation dependent.

Through comparison between inference with pMCMC and ABC, our results highlight a weakness of inference with ABC, in that it performs poorly for high-dimensional data, and also a weakness of our application of pMCMC, which is that it relies upon assumptions about the number of reorientation events in a time step; when this fails our posterior estimates are no longer accurate. We note, additionally, that pMCMC allows us to sample from the exact posterior for parameters from a

model (given that model is appropriate to describe the data), whereas via ABC we obtain approximate posteriors for a fixed tolerance, ϵ , which only become exact in the limit as $\epsilon \rightarrow 0$.

In Section 4.2, we described a framework for inference using pMCMC, and made certain assumptions about the VJP model, such as a separation in timescales between runs and reorientations, and a memoryless exponential distribution for the running time. Although many of these are standard assumptions, it would be possible to perform the same analysis in a more general model. For instance, if a running time distribution was chosen that does not satisfy the memoryless property, we could introduce an extra hidden variable, s , for the time since the last reorientation. In addition, in this work we have described parameter estimation via a hidden states formulation of the VJP model using dependence on hidden states from two time intervals. This can be extended to hidden states from three time intervals, which can allow rare consecutive reorientation events to be handled more accurately, albeit at greater computational cost.

Given a fixed photon budget and a trade-off between temporal sampling frequency and measurement noise, our results in the Section 4.3.5 indicate that a small value of Δt should be used for the discretization in time i.e. that motile individuals should be imaged as frequently as possible. In practice, there may be disadvantages to this choice of Δt . Computationally, conducting parameter inference via pMCMC will be significantly more expensive, although the computational run time may still be small in comparison to the duration of an experimental protocol. Datasets with higher noise present may also be much harder to interpret. Here, we have assumed that the noise present in the data is applied to the observed angle change. In reality there is noise on each pixel, which may contribute to an uncertainty in identifying the observed position of the object of interest.

Additionally, we considered using our particle MCMC inference framework to fit data from a misspecified model; our results indicate that the framework is robust

to moderate amounts of model misspecification. This allowed us to be confident in applying our framework to a real dataset of RNA tracks to consider the motility of RNA-protein complexes moving on the cytoskeleton, demonstrating the ability of our framework to obtain parameter estimates in challenging conditions with very noisy data.

Chapter 5

Adaptive weighting of summary statistics for ABC

To infer the parameters of mechanistic models with intractable likelihoods, techniques such as approximate Bayesian computation (ABC) are increasingly being adopted. One of the main disadvantages of ABC in practical situations, however, is that parameter inference must generally rely on summary statistics of the data. This is particularly the case for problems involving high-dimensional data, such as biological imaging experiments. However, some summary statistics contain more information about parameters of interest than others, and it is not always clear how to weight their contributions within the ABC framework. In this chapter, we address this problem by developing an automatic, adaptive algorithm that chooses weights for each summary statistic.

When using quantitative models to explore biological or physical phenomena, as for example in the model of mRNA transport in Chapter 3, it is crucial to be able to estimate parameters of these models and account appropriately for uncertainty in both the parameters and model predictions. Bayesian statistics offers a wealth of tools in this regard [47, 260], as outlined in Chapter 2. However, much of the current theory surrounding the generation of posterior distributions for parameter inference relies on being able to evaluate the likelihood of the data given the parameters of a model. In practice, for a large class of mechanistic models the likelihood is not tractable, either due to computational or analytical complexity. Therefore, the use of likelihood-free methods for inference has become widespread [10] (see Section 2.2 for further details). In particular, ABC [14, 15, 17, 18, 261] has been widely adopted due to its ease of understanding and implementation. The focus on ABC in this chapter is consistent with the use of ABC-SMC for study of mRNA transport and localization in *Drosophila* described in Chapter 3.

Both ABC and ABC-SMC methods have been described in Sections 2.2.1 and 2.2.4, respectively. These methods rely on approximations at several different levels. Use of a mechanistic or statistical model results in an approximation since the model will not be an exact representation of reality. The tolerance ϵ introduces

an approximation leading to an approximate Bayesian method rather than an exact method for performing Bayesian computation. Monte Carlo error will be present due to approximating a distribution by samples. In addition to these errors which are always present in ABC, summary statistics can optionally be used instead of the full dataset, which introduces a further source of error when these summary statistics are not sufficient statistics. In the following, we explore motivations for using summary statistics, methods for selecting and weighting different summary statistics, and introduce an algorithm to adaptively weight summary statistics within the ABC distance function. The algorithm aims to maximize the distance between the prior and the approximate posterior by automatically adapting the weights within the ABC distance function. Computationally, we use a nearest neighbour estimator of the distance between distributions. We justify the algorithm theoretically based on properties of the nearest neighbour distance estimator, and show it targets the correct distribution. To demonstrate the effectiveness of the algorithm, we apply it to a variety of test problems, including several toy models, a stochastic model of a biochemical reaction network, and a spatial model of diffusion, and compare the results with existing algorithms.

5.1 Role of summary statistics in ABC

As outlined in Section 2.2.2, it is often necessary to work with a lower-dimensional vector of summary statistics, $s(\mathbf{x})$, of the data, so that we require the distance between summary statistics is less than the tolerance ϵ , such that $d(s(\mathbf{x}^*), s(\mathbf{y})) < \epsilon$. Examples of these summary statistics may be spatial moments of the position of a random walker, or the FPT from nucleus to ring canal studied in the model in Chapter 3. However, not all summary statistics are equally informative about the posterior. Common practice is to combine summary statistics based on some heuristic approach, such as the weighting the contribution of each summary statistic

according to its standard deviation. However, it is not clear whether these heuristic approaches result in optimal weighting of the various summary statistics available. As such, the aim of this chapter is to provide an automated and adaptive method for determining the weighting of available summary statistics in order to optimize the quality of the resulting posterior.

5.1.1 Previous literature on selecting summary statistics in ABC

Previous work has also considered how to weight or select summary statistics for ABC. Fearnhead and Prangle [37] developed a popular method to find informative linear combinations of summary statistics by fitting a regression for each model parameter. This approach uses a projection method to find informative linear combinations of statistics by fitting a regression for each parameter in the model. The result is a reduction from the original high-dimensional set of summary statistics to a new lower dimensional set of summary statistics with the same dimensionality as the parameter space. Improved results are seen by using a pilot run of ABC to choose a subset of parameter space as a training region for the regression. Further improvements are obtained by extending the vector of summary statistics by concatenating with a non-linear transformation of the same summary statistics, $s(\mathbf{x}) = (s, s^2, s^3, s^4)$, where s is a given vector of summary statistics and the superscripts indicating raising these to the given power. This method uses contributions from all of the summary statistics and should optimize the mean quadratic loss.

Other successful approaches rely on subset selection methods [36, 97, 262]. Joyce and Marjoram [97] use criteria for approximate sufficiency of a subset of statistics to test whether adding a new statistic results in a change in the posterior above a certain threshold. Nunes and Balding [262] minimize an information criterion based on knn-entropy over all subsets of summary statistics. Barnes et al. [36] use an approximate sufficiency criterion to select a subset of summary statistics

on which to base inference, by adding summary statistics until the KL divergence between the resulting posteriors is below a threshold in order to reduce any loss of information. All of these methods seek to choose a lower dimensional subset of a given list of summary statistics. Using this lower dimensional subset increases the acceptance rate for samples in ABC by avoiding the curse of dimensionality for the data. However, the results can depend on the order in which the summary statistics (or subsets) are analysed.

Further methods rely on assigning weights to each summary statistic within the ABC distance function and considering how best to choose these weights. Prangle [111] present a method for adaptively choosing summary statistic weights for ABC based on the scale of the summary statistics from the predictive distribution for summary statistics at each generation of an ABC-SMC algorithm. The median absolute deviation (MAD), a measure of spread of a statistic, is used for the scaling. All summary statistics are scaled by a measure of their spread, but if the summary statistics are not equally informative about the posterior distribution this may not be the best choice. A genetic algorithm has also been used to choose the weights of different summary statistics [263]. This algorithm attempts to optimize the mean squared error (MSE) of the posterior samples from the true parameter. Recently, Singh and Hellander [264] have proposed a multi-armed bandit approach to selecting summary statistics for ABC. Other work has avoided using summary statistics at all by considering the Wasserstein distance between full data sets [252] or by using classification techniques from machine learning to discriminate between datasets [265].

We approach the problem of combining summary statistics from the point of view of adapting the ABC distance function to information contained in the summary statistics, rather than selecting a certain subset of summary statistics. We provide an automatic algorithm that adaptively selects weights for each summary statistic within the ABC distance function.

5.1.2 Outline

The contribution of this chapter is to present a flexible, novel framework for improving inference with ABC by adapting the weights of different summary statistics to maximize the gain in posterior information from a dataset. This is helpful for avoiding bias and variance from redundant information in data (such as would be the case when including a summary statistic that is uncorrelated with the parameters of interest). A further advantage of this work is that it can alleviate the burden of designing and selecting summary statistics ‘by hand’, since a large collection of summaries can be used and weighted appropriately via the procedure presented in this chapter. It is also possible to combine this framework with existing dimensionality reduction techniques for summary statistics in ABC (see Section 5.4.9).

We outline in Section 5.2 an adaptive algorithm for combining summary statistics in an ABC framework. We provide theoretical justification for the algorithm in Section 5.3 and demonstrate that we obtain convergence to the posterior distribution. To demonstrate the utility of the algorithm, we apply it to several test problems based on biochemical reaction networks in Section 5.4. We compare results of parameter inference using the adaptive algorithm against benchmark results from applying ABC-SMC using other choices of weights for the summary statistics. Finally, in Section 5.5, we summarize the work presented in this article and compare the methodology for combining summary statistics with other techniques in the literature that are based on dimensionality reduction of a set of summary statistics.

5.2 An algorithm for automatic weighting of summary statistics in ABC

A Sequential Monte Carlo (SMC) algorithm produces a sequence of approximations that aid in moving between the prior and posterior distributions (see Section 2.1.5). The motivation for the method introduced here is to accelerate this movement from

the prior to the posterior by maximizing the distance between the prior and the current approximation to the posterior distribution, subject to the samples for this current approximation being generated from a valid ABC-SMC algorithm.

In order to use ABC-SMC (see Algorithm 3), we must specify a function to measure the distance between simulated and real datasets. Suppose we take a weighted Euclidean distance as the ABC distance function such that

$$d_{\mathbf{w}}(s(\mathbf{x}_1), s(\mathbf{x}_2)) = \sum_{i=1}^{\kappa} w_i (s_{1i} - s_{2i})^2,$$

where $s(\mathbf{x}) = (s_1, \dots, s_{\kappa}) \in \mathbb{R}^{\kappa}$ is a vector of summary statistics and the sum over i is taken over all the summary statistics considered. This distance function is a reasonable and flexible choice commonly used in the literature [266]. It is these distance weights, w_i , that control how the summary statistics are combined in this case. Given simulated pairs of parameter samples and datasets, we find weights, $\mathbf{w} = (w_1, \dots, w_{\kappa}) \in E \subset \mathbb{R}^{\kappa}$, that maximize a distance between the prior and the posterior that represents the maximum possible gain in information about the parameters from the given data. Constructing the weights in this way allows us to account for the scale of the summary statistics, as well as their relative contribution to a posterior.

5.2.1 Adaption of weights

We seek to optimize the weights, $\mathbf{w}$, so that we can place less emphasis on summary statistics that are not informative for the posterior, but also scale summary statistics appropriately so that we do not neglect to obtain information about certain parameters. We do this within the ABC-SMC framework [16, 78, 106] given in Algorithm 3 using the implementation outlined in Prangle [111]. We outline the proposed methodology in Algorithm 5.

At each generation, we search for the weights, $\mathbf{w}$, of the distance function that

maximize the distance between the prior and resulting posterior, given M ABC samples from the model for different θ values. This distance between prior and posterior gives a measure of the information gain in moving from the prior to the posterior.

5.2.2 Distance between distributions

We use the Hellinger distance to measure the discrepancy between the prior and approximate posterior, and so to quantify the information gained in moving towards the true posterior distribution. The Hellinger distance is defined, for distributions P and Q , with densities p and q , respectively, as

$$H^2(P, Q) = \frac{1}{2} \int \left(\sqrt{p(x)} - \sqrt{q(x)} \right)^2 dx \quad (5.1)$$

$$= 1 - \int \sqrt{p(x)q(x)} dx. \quad (5.2)$$

Alternative measures of distance between distributions such as the Euclidean distance or KL divergence can be used. The Hellinger distance performs better than alternatives, particularly for robustly identifying relatively small differences between posterior distributions, an observation that is supported by other work [238]. In addition, the Hellinger distance is finite when comparing distributions with different support (unlike the KL divergence). This property is desirable when comparing a broad prior with a posterior distribution where we have gained some knowledge of parameter space and can exclude certain regions.

Nearest neighbour distance estimator

To estimate the distance between two distributions, based on samples from these distributions, we use a k nearest neighbour estimator [267–269] developed to describe a family of distances between distributions known as α divergences, of which the Hellinger distance is a special case. Suppose we have two probability distributions,

P and Q with densities $p(x)$, and $q(x)$, and are interested in the distance between these. We suppose that we have samples, $X_{1:N}$ and $Y_{1:N}$ from p and q . If we define $D_\alpha(p||q) = \int p^\alpha(x)q^{1-\alpha}(x) dx$ for $\alpha \in \mathbb{R}$, then the Hellinger distance, which is related to the divergence D_α for $\alpha = 1/2$, is defined as

$$D_h(p||q) = 1 - D_{1/2}(p||q).$$

The k nearest neighbour estimator that we use depends only on distances between observations in a sample. Let $\rho_k(i)$ be the Euclidean distance from the sample X_i to its k th nearest neighbour in $X_{1:N}$. Similarly, let $\nu_k(i)$ be the distance from X_i to its k th nearest neighbour in $Y_{1:N}$. Then the estimator [267] is given by

$$\hat{D}_\alpha(X_{1:N}||Y_{1:N}) = \frac{1}{N} \sum_{i=1}^N \left(\frac{(N-1)\rho_k(i)}{N\nu_k(i)} \right)^{1-\alpha} B_{k,\alpha}, \quad (5.3)$$

where

$$B_{k,\alpha} = \frac{\Gamma(k)^2}{\Gamma(k-\alpha+1)\Gamma(k+\alpha-1)}.$$

At each generation of ABC-SMC we seek to find weights $\mathbf{w} \in E \subset \mathbb{R}^\kappa$ such that

$$\mathbf{w}^* = \operatorname{argmax}_{\mathbf{w} \in E} \left(1 - \hat{D}_\alpha(\{\xi_i\}_{i=1}^N || \{\theta_i\}_{i=1}^N) \right), \quad (5.4)$$

where $\{\xi_i\}_{i=1}^N$ are samples from the prior distribution, and $\{\theta_i\}_{i=1}^N$ are samples from the approximate posterior distribution, which depends on the summary statistic weights, $\mathbf{w}$.

To perform the optimization in weight space in the implementation of Algorithm 5, we use a constrained nonlinear optimizer, implemented via `fmincon` in MATLAB [270], with multiple initializations of the optimizer. This nearest neighbour estimator is computationally intensive in high dimensions, where the high dimensional setting corresponds to using many particles in the SMC part of the adaptive ABC-SMC algorithm. In the context of ABC, the computational cost is dominated nonetheless

by the expensive model simulations.

Algorithm 5 Adaption of distance weights for ABC

- 1: Set generation index $t = 1$. Let $M = \lceil N/\alpha \rceil$. Set $i = 1$.
- 2: **while** $i \leq M$ **do**
- 3: Sample from prior $\theta^{**} \sim \pi(\theta)$.
- 4: Simulate dataset $\mathbf{x}^{*i} \sim f(\mathbf{x}|\theta^{**})$.
- 5: Accept the parameters.
 Set $\theta_t^i = \theta^{**}$.
 Calculate the weight v_t^i for particle θ_t^i as $v_t^i = 1$.
- 6: Set $i = i + 1$.
- 7: **end while**
- 8: Let

$$L(\mathbf{w}) = 1 - \hat{D}_{1/2} \left(\{\xi^i\}_{i=1}^N \parallel \{\theta_t^i\}_{i=1}^N \right),$$

where $\{\theta_t^i\}_{i=1}^N$ are the closest N samples when ranked according to ABC distance from the observed dataset, $\mathbf{y}$, via $d_{\mathbf{w}}(s(\mathbf{x}^{*i}), s(\mathbf{y}))$.

- 9: Maximize $L(\mathbf{w})$ as a function of summary statistic weights, $\mathbf{w}$.
- 10: Keep the samples $\{\theta_t^i\}_{i=1}^N$ corresponding to the maximum of $L(\mathbf{w})$ (i.e. the maximum distance between prior and approximate posterior).
- 11: Normalize the particle weights $\{v_t^i\}_{i=1}^N$.
- 12: Calculate tolerance, ϵ_t , as the α quantile of distances. Store the optimum weights $\mathbf{w}_t^*$.
- 13: **for** $t = 2$ to T **do**
- 14: Set $i = 1$.
- 15: **while** $i \leq M$ **do**
- 16: Sample θ^* from previous population $\{\theta_{t-1}^i\}_{i=1}^N$ with weights $\{v_{t-1}^i\}_{i=1}^N$.
 Perturb θ^* to give $\theta^{**} \sim K_t(\theta|\theta^*)$. If $\pi(\theta^{**}) = 0$, return to step 15.
- 17: Simulate dataset $\mathbf{x}^{*i} \sim f(\mathbf{x}|\theta^{**})$.
- 18: **if** $d_{\mathbf{w}_j^*}(s(\mathbf{x}^{*i}), s(\mathbf{y})) < \epsilon_j$ for all $j < t$ **then**
- 19: Accept the parameters.
 Set $\theta_t^i = \theta^{**}$.
 Calculate the weight v_t^i for particle θ_t^i as

$$v_t^i = \frac{\pi(\theta_t^i)}{\sum_{j=1}^M v_{t-1}^j K_t(\theta_{t-1}^j, \theta_t^i)}. \quad (5.5)$$

- 20: **end if**
- 21: Set $i = i + 1$.
- 22: **end while**
- 23: Let

$$L(\mathbf{w}) = 1 - \hat{D}_{1/2} \left(\{\xi^i\}_{i=1}^N \parallel \{\theta_t^i\}_{i=1}^N \right),$$

where $\{\theta_t^i\}_{i=1}^N$ are the closest N samples when ranked according to ABC distance from the observed dataset, $\mathbf{y}$, via $d_{\mathbf{w}}(s(\mathbf{x}^{*i}), s(\mathbf{y}))$.

- 24: Maximize $L(\mathbf{w})$ as a function of summary statistic weights, $\mathbf{w}$.
 - 25: Keep the samples $\{\theta_t^i\}_{i=1}^N$ corresponding to the maximum of $L(\mathbf{w})$ (i.e. the maximum distance between prior and approximate posterior).
 - 26: Normalize the particle weights $\{v_t^i\}_{i=1}^N$.
 - 27: Calculate tolerance, ϵ_t , as the α quantile of distances. Store the optimum weights $\mathbf{w}_t^*$.
 - 28: **end for**
 - 29: **return** $\{\theta_T^i\}_{i=1}^N, \{v_T^i\}_{i=1}^N$
-

5.3 Theoretical justification

In the following, we demonstrate that by using the nearest neighbour estimator, we can obtain optimum weights that maximize the distance between the prior and approximate posterior. Furthermore, we show by using standard arguments that the adaptive ABC-SMC algorithm given in Algorithm 5 will converge to the correct ABC posterior distribution.

5.3.1 Convergence of the nearest neighbour estimator

The estimator we use is a k nearest neighbour estimator relying only on distances between observations in a sample, as described above in Section 5.2.2. With this nearest neighbour estimator (5.3), under certain conditions on p and q , we have L_2 convergence of the estimator [267]. This ensures that, in the limit of more samples, estimates of the distance between p and q will become more concentrated around the true distance, such that optimizing the estimate of the distance will give the true optimum, $\mathbf{w}^*$. We note that although the Hellinger distance, $D_h(p||q)$, is symmetric in p and q , the estimator above in Equation (5.3) is not. By using the estimator from Equation (5.3) and choosing q as the distribution that depends on the weights, $\mathbf{w}$, we are able to make strong assumptions about p independent of the weights, $\mathbf{w}$, and make weaker assumptions about q . In the context of the adaptive algorithm for ABC, this allows us to treat p as the prior and q as the approximate posterior distribution.

Theorem 1. *Suppose that $k \geq 2$ and that $\mathcal{M} = \text{supp}(p)$. Assume that*

- (a) q is bounded above;*
- (b) p is bounded away from zero;*
- (c) p is uniformly Lebesgue approximable;*
- (d) $\exists \delta_0$ such that $\forall \delta \in (0, \delta_0)$*

$$\int_{\mathcal{M}} H(x, p, \delta, 1/2) p(x) dx < \infty, \quad (5.6)$$

where $H(x, p, \delta, \psi) =$

$$\sum_{j=0}^{k-1} \left(\frac{1}{j!}\right)^\psi \Gamma\left(\frac{1}{2} + j\psi\right) \left(\frac{p(x) + \delta}{p(x) - \delta}\right)^{j\psi} (p(x) - \delta)^{-\frac{1}{2}} \\ ((1 - \delta)\psi)^{-\frac{1}{2} - j\psi};$$

(e)

$$\int_{\mathcal{M}} \|x - y\|^\gamma p(y) dy < \infty \text{ for almost all } x \in \mathcal{M}, \\ \int \int_{\mathcal{M}^2} \|x - y\|^\gamma p(y) p(x) dy dx < \infty;$$

(f) $E \subset \mathbb{R}^\kappa$ is compact;

(g) a unique $\mathbf{w}^* \in E$ maximizes $D_h(p||q^{(\mathbf{w})})$.

Then

$$\lim_{N \rightarrow \infty} \mathbb{P} \left(\operatorname{argmax}_{w \in E} \hat{D}_h \left(X_{1:N} || Y_{1:N}^{(\mathbf{w})} \right) = \mathbf{w}^* \right) = 1. \quad (5.7)$$

Proof. See Appendix B for details. $\square$

The proof of Theorem 1 relies on results from Póczos and Schneider [267] showing L_2 consistency of the nearest neighbour estimator, which ensures that the estimates of the distance between p and q become more concentrated around the true values as more samples are used. This requires the construction of an integrable function as a bound such that Lebesgue's dominated convergence theorem can be applied. Most of the conditions on p and q are to ensure that this all holds. We assume that the space of ABC distance weights, E , is compact. In practice this is not a problem, since we can work with a constrained optimization problem and assume that the summary statistic weights lie within a large but finite region. Although there are several conditions on the prior distribution, p , most reasonable choices of prior distribution will satisfy these, and only a single condition on the approximate posterior distribution, q , is assumed. In the limit of having more samples from the distributions p and $q^{(\mathbf{w})}$, selecting summary statistic weights, $\mathbf{w}$, based on optimizing the estimate from $\hat{D}_h$ will give the true optimum, $\mathbf{w}^*$, of this distance between distributions.

5.3.2 Convergence of adaptive ABC-SMC

We demonstrate here that Algorithm 5 converges to the correct target distribution by using arguments from Prangle [111].

Theorem 2. *Prangle [111]*

Suppose that the following conditions hold:

- (a) $\theta \in \mathbb{R}^\lambda$, $s(\mathbf{x}) \in E \subset \mathbb{R}^\kappa$ and these random variables have joint density $\pi(\theta, s(\mathbf{x}))$ with respect to the Lebesgue measure;*
- (b) the sets $A_t = \{s(\mathbf{x}) \mid d_{\mathbf{w}_t^*}(s(\mathbf{x}), s(\mathbf{y})) < \epsilon_t\}$ are Lebesgue measureable, where w^* are the optimal weights at generation t , and ϵ_t are the tolerances at each generation;*
- (c) $\pi(s(\mathbf{y})) > 0$;*
- (d) $\lim_{t \rightarrow \infty} m(A_t) = 0$, where $m(\cdot)$ represents Lebesgue measure;*
- (e) the sets A_t have bounded eccentricity. That is, for any A_t , $\exists$ a ball $B_t = \{s(\mathbf{x}) \mid \|s(\mathbf{x}) - s(\mathbf{y})\|_2 \leq r_t\}$ about $s(\mathbf{y})$ with radius r_t such that $A_t \subset B_t$ and $m(A_t) \geq c m(B_t)$, where $\|\cdot\|$ is the Euclidean or L_2 norm and $c > 0$ is a constant.*

Then the posterior approximations converge,

$$\lim_{t \rightarrow \infty} p_{ABC,t}(\theta | s(\mathbf{y})) = p(\theta | s(\mathbf{y})) \text{ for almost all } (\theta, s(\mathbf{y})),$$

where $p_{ABC,t}(\theta | s(\mathbf{y}))$ is the ABC posterior, defined as

$$p_{ABC,t}(\theta | s(\mathbf{y})) \propto \int p(s | \theta) \pi(\theta) \mathbb{1}[d_{\mathbf{w}_t^*}(s(\mathbf{x}), s(\mathbf{y})) \leq \epsilon_t] d\mathbf{x}.$$

Proof. See Appendix B for details. □

Similarly to Prangle [111], we can take $A_t = \{s(\mathbf{x}) \mid d_{\mathbf{w}_i^*}(s(\mathbf{x}), s(\mathbf{y})) < \epsilon_t \text{ for all } i \leq t\}$ as the acceptance region for Algorithm 5, which ensures that $m(A_t)$ is decreasing in t . This acceptance region is not guaranteed to converge to zero under all circumstances due to model misspecification or particles becoming trapped in the wrong mode of a multimodal posterior within the SMC algorithm. However, these are issues that equally affect the conventional ABC-SMC algorithm, as discussed in Prangle [111]. The bounded eccentricity condition is ensured for the scaled Euclidean distance in Algorithm 5 by using weights $\mathbf{w} \in E$ with compact support, since the weights will be bounded by a maximum

upper bound at each generation. Thus the conditions for convergence of Algorithm 5 are essentially the same as those for convergence of standard ABC-SMC algorithms, provided that the weights considered in the distance function have compact support.

5.4 Examples

We apply the algorithm of automatic, adaptive weighting of summary statistics to a variety of test problems, including toy models and problems based on different chemical reaction networks. The dynamics of these networks are simulated stochastically using Gillespie’s direct method [271], which allows us to sample trajectories directly from the model. The summary statistics collected for each of the chemical reaction network problems are in the form of a time series, to imitate data that could be collected from a biological experiment. For some of these models it is possible to solve for the likelihood analytically, and we show comparisons with exact posterior distributions sampled via MCMC. However, for some of the more complex, but more biologically realistic, test problems, the likelihood is very computationally expensive and the exact posterior distribution is not available.

To demonstrate the effectiveness of taking a flexible choice of distance weights, we make two comparisons. Firstly, we compare results obtained using Algorithm 5 to those generated using a uniform choice of weights, $w_i = 1 \ \forall i$. Secondly, we compare to results generated using weights that scale with each summary statistic, as in the method of Prangle [111]. For this method, we use $w_i = 1/\sigma_i \ \forall i$, where σ_i is the MAD of the given summary statistic based on simulations from a given generation of ABC-SMC. We show approximate posterior distributions produced by each method for choosing weights. For the tractable test problems considered, we compute the MSE, Hellinger distance between prior and approximate posterior, and Hellinger distance between approximate posterior and exact posterior sampled via MCMC, all averaged over 40 repeats of Algorithm 5.

We note that a table summarizing the parameters used in the implementation

of all the test problems can be found in Appendix B. For the nearest neighbour estimator, $k = 5$ is used throughout, as recommended by previous work [268].

5.4.1 Uniform toy model

We first consider a tractable toy problem with a sufficient statistic to illustrate the method of Algorithm 5. We observe

$$x_i \sim \text{Unif}([0, \theta]), \quad (5.8)$$

for $i = 1, \dots, r$. In this case, the maximum of the observed values is a sufficient statistic, $s(\mathbf{x}) = \max_i x_i$. Instead of using only this single statistic, we take the full sample vector sorted in ascending order, such that the summary statistics are the order statistics of the sample $s(\mathbf{x}) = (x_{(1)}, \dots, x_{(r)})$. We can sample the true posterior distribution $p(\theta|s(\mathbf{y}))$ directly via (likelihood-based) MCMC and compare to the approximations obtained via ABC.

In Figure 5.1 we show the marginal posterior distribution generated using uniform weights, weights scaled with the MAD of each summary statistic via the adaptive method of Prangle [111], and adaptively chosen weights via the method outlined in Algorithm 5. The results indicate that the method in Algorithm 5 is able to produce a higher quality approximation of the posterior for a given number of parameter samples compared to other methods of weighting the summary statistics. The true parameter is $\theta = 10$, a prior uniform on the logarithm of the parameters over the interval $[10^0, 10^2]$ was used, and $r = 10$ samples of the uniform model were used as the dataset.

5.4.2 Bimodal toy model

Although distances between distributions can be intuitive for unimodal Gaussian-like distributions, we want to ensure the method of Algorithm 5 is robust to multimodal distributions. The benefit of using a general distance measure like the Hellinger

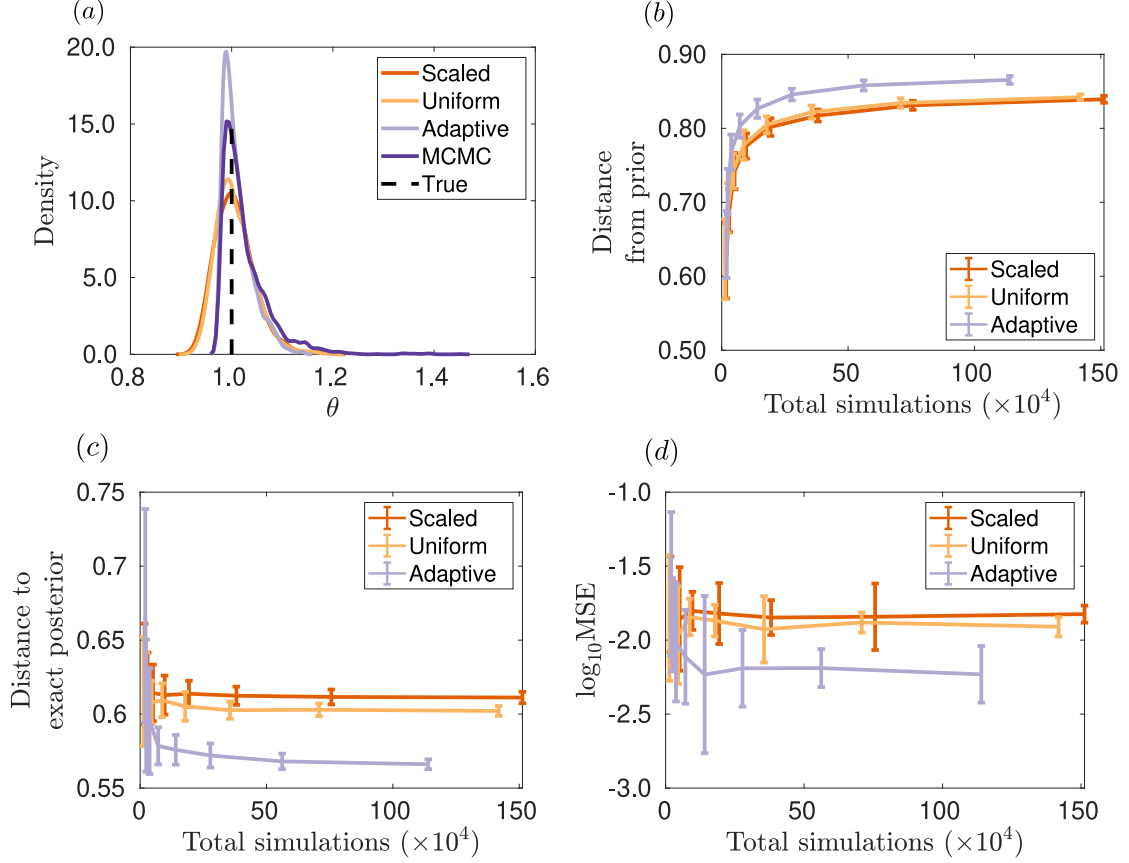


Figure 5.1. The posterior distribution for parameter θ of the uniform toy model for different weights in the ABC distance function is shown in (a). ABC-SMC was used to provide estimates of the posterior, with 10 generations and $N = 2,000$ particles at each generation with the posterior constructed from the closest 50% of the simulations ($\alpha = 0.5$). Metrics to evaluate the performance of Algorithm 5 are shown in (b), (c), and (d) as N varies resulting in different total numbers of simulations from the model. Results are averaged over 40 repeated runs. In (b), is shown the Hellinger distance from the prior to approximate posterior distribution, which is maximized directly within Algorithm 5. In (c), the Hellinger distance from the approximate posterior to the exact posterior distribution (as sampled via MCMC) is displayed. A lower value of this distance indicates a better approximation to the posterior. In (d), the MSE from the approximate posterior mean to the exact posterior mean is shown.

distance is that it is well suited to quantifying distances between such general distributions. We aim to infer $\theta = (\theta_1, \theta_2) \in \mathbb{R}^2$ based on observations

$$s(\mathbf{x}) \sim \mathcal{N} \left(\begin{pmatrix} \sin(\theta_1) \\ \sin(\theta_2) \end{pmatrix}, 0.1^2 I_2 \right).$$

We take a prior distribution for θ_1 and θ_2 uniform on $[0, 2\pi]$. We suppose that we observe $s(\mathbf{y}) = (\sqrt{2}/2, -\sqrt{2}/2)$, which leads to four distinct modes of a multimodal posterior distribution in two dimensions. As shown in Figure 5.2, the adaptive method performs robustly for multimodal posterior distributions and is competitive with the other methods considered. There are added complications in evaluating the quality of an approximation to a multimodal posterior distribution, since several distinct regions of parameter space can equally well explain the observed data. Using the symmetry of the posterior, we measure MSE by distance of the approximate posterior mean from $(\pi/2, \pi/2)$ at the centre of the four modes.

5.4.3 Gaussian toy model

Suppose we aim to infer a single parameter, θ . We observe two summary statistics $s(\mathbf{x}) = (x_1, x_2)$, where $x_1 \sim \mathcal{N}(\theta, 0.1^2)$ and $x_2 \sim \mathcal{N}(0, 1^2)$ such that the first summary statistic is informative and the second is uninformative. We take a broad prior distribution of $\mathcal{N}(0, 100^2)$ for θ . The effect of this is that by choosing weights for summary statistics based on the prior predictive distribution, the weight for the informative summary statistic, x_1 , is smaller than for the uninformative statistic, x_2 . This toy model was used previously by Prangle [111] to demonstrate their weight adaption method, by comparison with fixed weights based on summary statistics simulated from the prior predictive distribution. The method from Algorithm 2 improves on this method, as shown in Figure 5.3, by reaching faster adaption to the scales of the summary statistics, combined with how informative they are. In addition, in Figure 5.3 we show the scaling of the MSE as the number of simulations increases. Here there are two summary statistics on similar scales, so weighting these equally is a sensible choice. The adaptive method of Algorithm 5 matches the scaling of MSE of the uniform choice of weights, improving upon the scaled method.

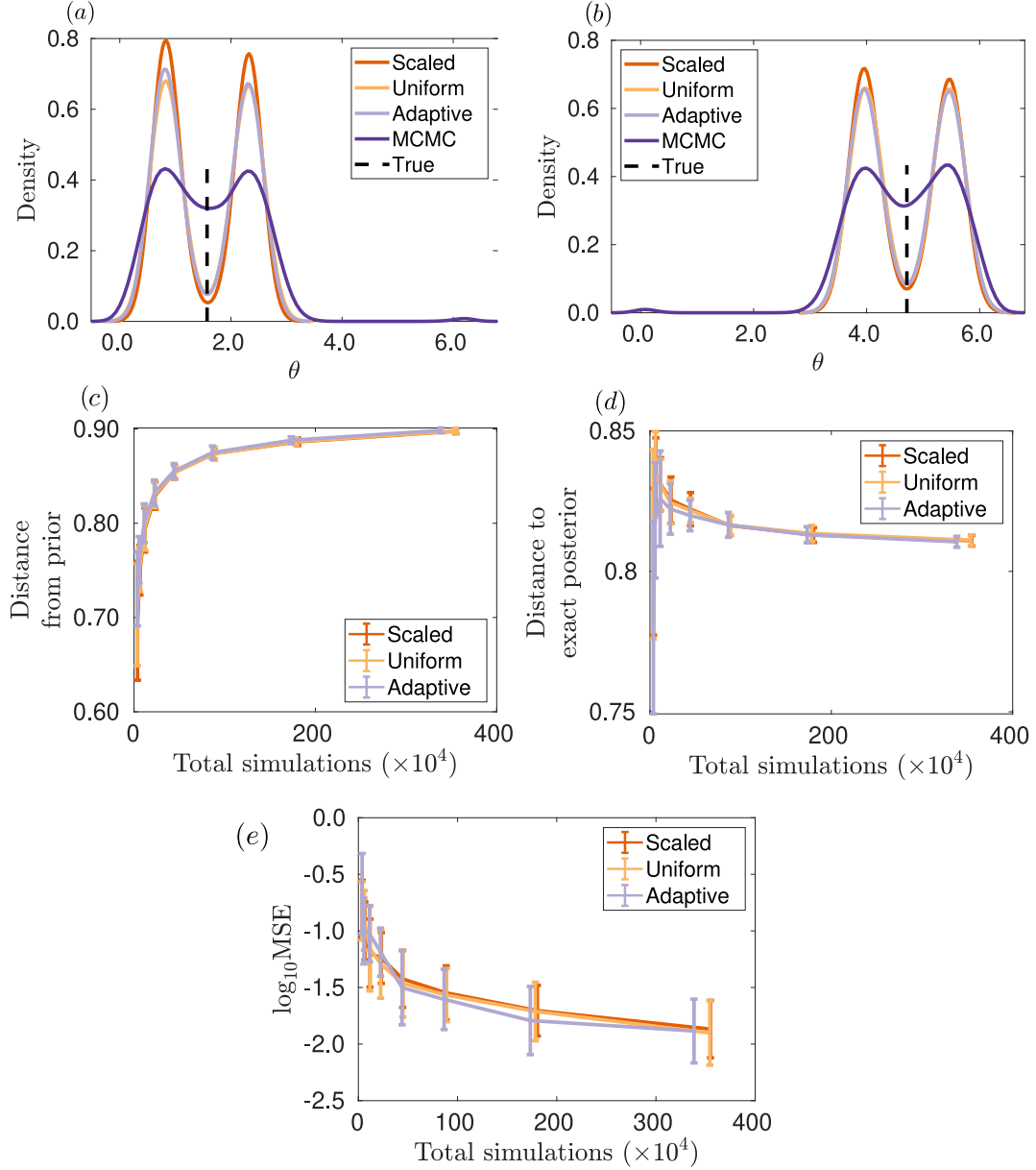


Figure 5.2. The posterior distribution for parameters $\theta = (\theta_1, \theta_2)$ of the bimodal toy model for different weights in the ABC distance function is shown in (a) and (b). ABC-SMC was used to provide estimates of the posterior, with 10 generations and $N = 2,000$ particles at each generation with the posterior constructed from the closest 50% of the simulations ($\alpha = 0.5$). Metrics to evaluate the performance of Algorithm 5 are shown in (c), (d), and (e) as N varies resulting in different total numbers of simulations from the model. Results are averaged over 40 repeated runs. In (c), is shown the Hellinger distance from the prior to approximate posterior distribution. In (d), the Hellinger distance from the approximate posterior to the exact posterior distribution is displayed. In (e), the MSE from the approximate posterior mean to the point $(\pi/2, 3\pi/2)$ is shown.

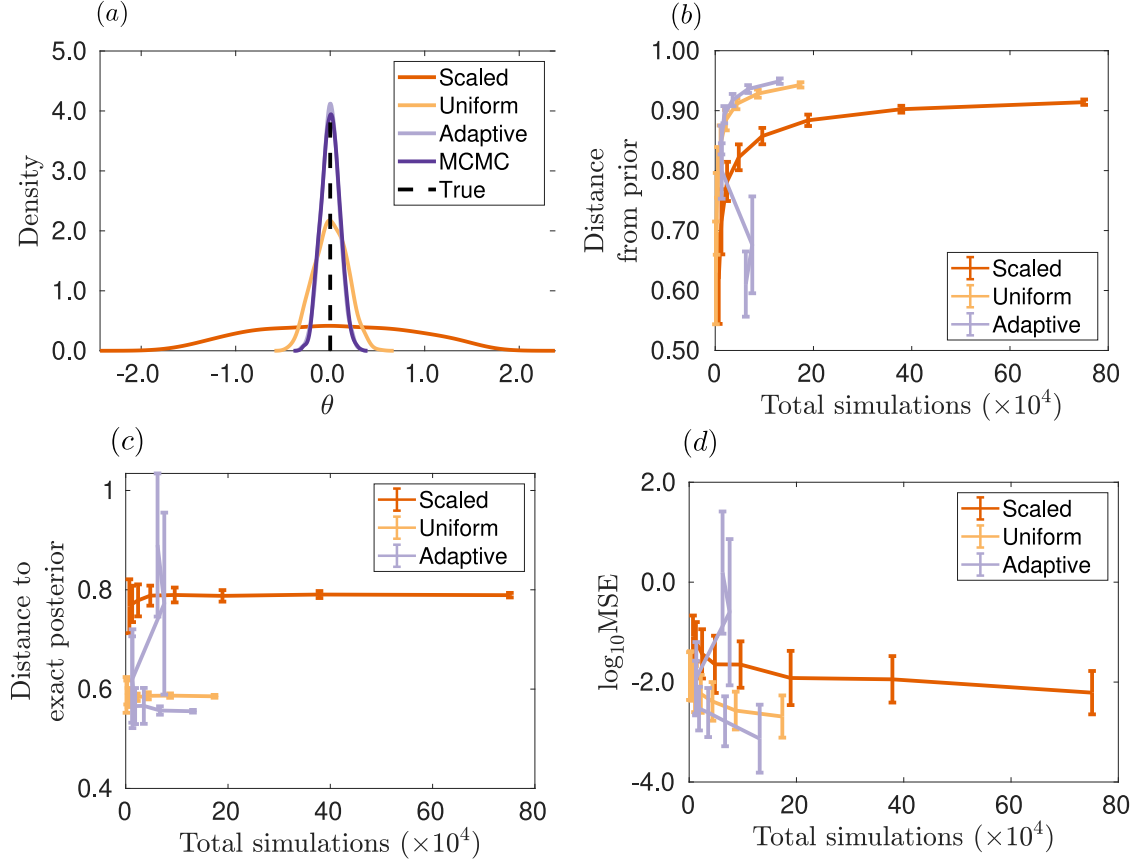


Figure 5.3. Posterior for parameters θ of the gaussian toy model for different weights in the ABC distance function. ABC-SMC was used to provide estimates of the posterior, with five generations and $N = 2,000$ simulations at each generation with the posterior constructed from the closest 50% of the simulations ($\alpha = 0.5$). Metrics to evaluate the performance of Algorithm 5 are shown in (b), (c), and (d) as N varies resulting in different total numbers of simulations from the model. Results are averaged over 40 repeated runs. In (b), is shown the Hellinger distance from the prior to approximate posterior distribution, which is maximized directly within Algorithm 5. In (c), the Hellinger distance from the approximate posterior to the exact posterior distribution (as sampled via MCMC) is shown. A lower value of this distance indicates a better approximation to the posterior. In (d), the MSE from the approximate posterior mean to the exact posterior mean is shown.

5.4.4 Death process

For our first test problem, we consider estimating the rate parameter for a single, first order degradation reaction:



We will consider for this, and subsequent, test problems that time has been non-dimensionalized. Initially, we assume there are $A(0) = 10$ particles in the system, which is observed over a (non-dimensional) time period $[0, 20]$. We assume it is possible to measure the state of the system (in this case the number of molecules of species A) without observation noise at given time points $t_0, t_1, \dots, t_n$. For this test problem, we assume that we measure at n equally spaced time intervals, where $n = 8$.

As our summary statistics, we take $s(\mathbf{x}) = [A(t_0), A(t_1), \dots, A(t_n), z]$ where z is an observation of a random variable $Z \sim \mathcal{N}(0, \sigma^2)$ that is uncorrelated with the death process. We suppose that the scale of the variance, σ , is different to the scale of the observations of the exponential decay process, giving a simple system with a two-dimensional parameter to infer: $\theta = (k_1, \sigma)$. Note that the scale of z is determined by the standard deviation, σ , but the scale of the death process is affected by the initial condition, $A(0) = 10$, resulting in two distinct scales in these summary statistics.

Results of parameter inference for this system using ABC-SMC are shown in Figure 5.4, where the true parameters used are $\theta = (0.1, 0.01)$ and a prior uniform on the logarithm of each of the parameters over the interval $[10^{-3}, 10^3]$ was used. We observe similar performance in identification of the decay parameter k_1 using uniform weights, scaled weights and the adaptive choice of weights. Scaling the summary statistics with the MAD results in problems here since some summary statistics may not vary much upon simulation from the prior predictive distribution initially, resulting in $\sigma_i = 0$ for the corresponding summary statistic. To solve this, we set the weight for a given summary statistic to a large fixed value if the MAD for that summary statistic was below a small threshold value. Note that only one summary statistic provides information about the random variable Z , whereas the other $n + 1$ summary statistics (which are observations of the decay process at time points $\{t_i\}_{i=0}^n$) provide information about the decay of species A . The summary

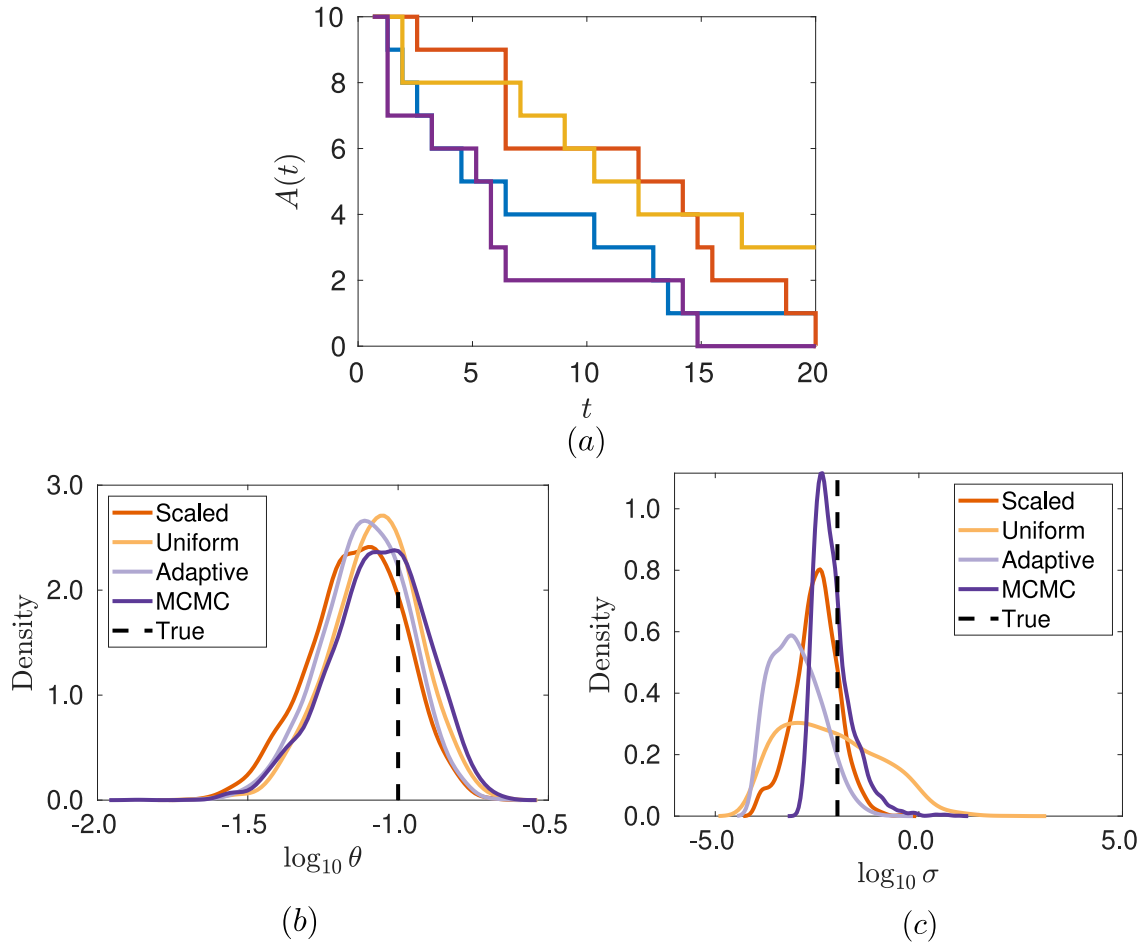


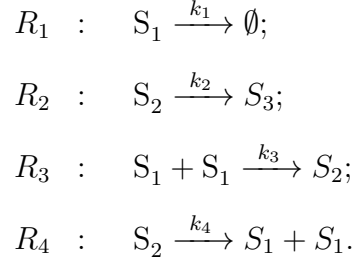
Figure 5.4. Posteriors for parameters k_1 and σ in the death process test problem for different weights in the ABC distance function. ABC-SMC was used to provide estimates of the posterior, with five generations and $N = 2000$ simulations at each generation with the posterior constructed from the closest 50% of the simulations ($\alpha = 0.5$). (a) shows typical output from the model for the true parameters. The posteriors for k_1 are given in (b) and for σ in (c).

statistic weights chosen via the search process outlined in Algorithm 5 give rise to a posterior that outperforms the posteriors generated using uniform weights and scaled weights for the second parameter σ , since only a single summary statistic provides relevant information for this parameter.

5.4.5 Dimerization system

To examine a system with multiple scales, we consider also a dimerization model, which undergoes a fast initial transient followed by slower subsequent dynamics [51].

This system has higher dimensional parameter and summary statistic spaces, and exhibits more complex behaviour typical of more realistic biological models. The dimerization system consists of the following reactions:



We take initial conditions $S_1(0) = 10^5$, $S_2(0) = 0$, $S_3(0) = 0$ and consider an observational time period of $[0, 100]$ with $n = 16$ geometrically spaced observations (to capture the multiple timescales present), without observational noise. For the dimerization system, we take the time series $s(\mathbf{x}) = [S_1(t_0), \dots, S_1(t_n), S_2(t_0), \dots, S_2(t_n), S_3(t_0), \dots, S_3(t_n)]$ as summary statistics and infer the four-dimensional parameters $\theta = (k_1, k_2, k_3, k_4)$. We note that for a choice of parameter $\theta^* = (1, 0.04, 0.002, 0.5)$, and the given initial conditions, we obtain a fast decay of species S_1 and accumulation of species S_2 , followed by a slower decay of S_2 and accumulation of S_3 (see Figure 5.5(a)).

The results of parameter inference for this system can be seen in Figure 5.5. The true parameters used are $\theta = (1, 0.04, 0.002, 0.5)$, and we apply a prior uniform on the logarithm of the parameters over the intervals $[10^{-2}, 10^2]$, $[10^{-3}, 10^1]$, $[10^{-5}, 10^{-1}]$, $[10^{-3}, 10^1]$, respectively, for each parameter. Parameters k_1 and k_2 are clearly identified by the adaptive choice of weights. The fast transient behaviour initially involves reactions at rate k_1 , while k_2 corresponds to the longer timescale accumulation of species S_3 . Parameters k_3 and k_4 are harder to identify with broader resulting posteriors, but again the adaptive algorithm does a better job at excluding regions of search space than a uniform choice of weights, or scaling with the MAD.

For this test problem (and also for the simple spatial model in Section 5.4.6), comparison to the true posterior distribution is not feasible. If an observation noise

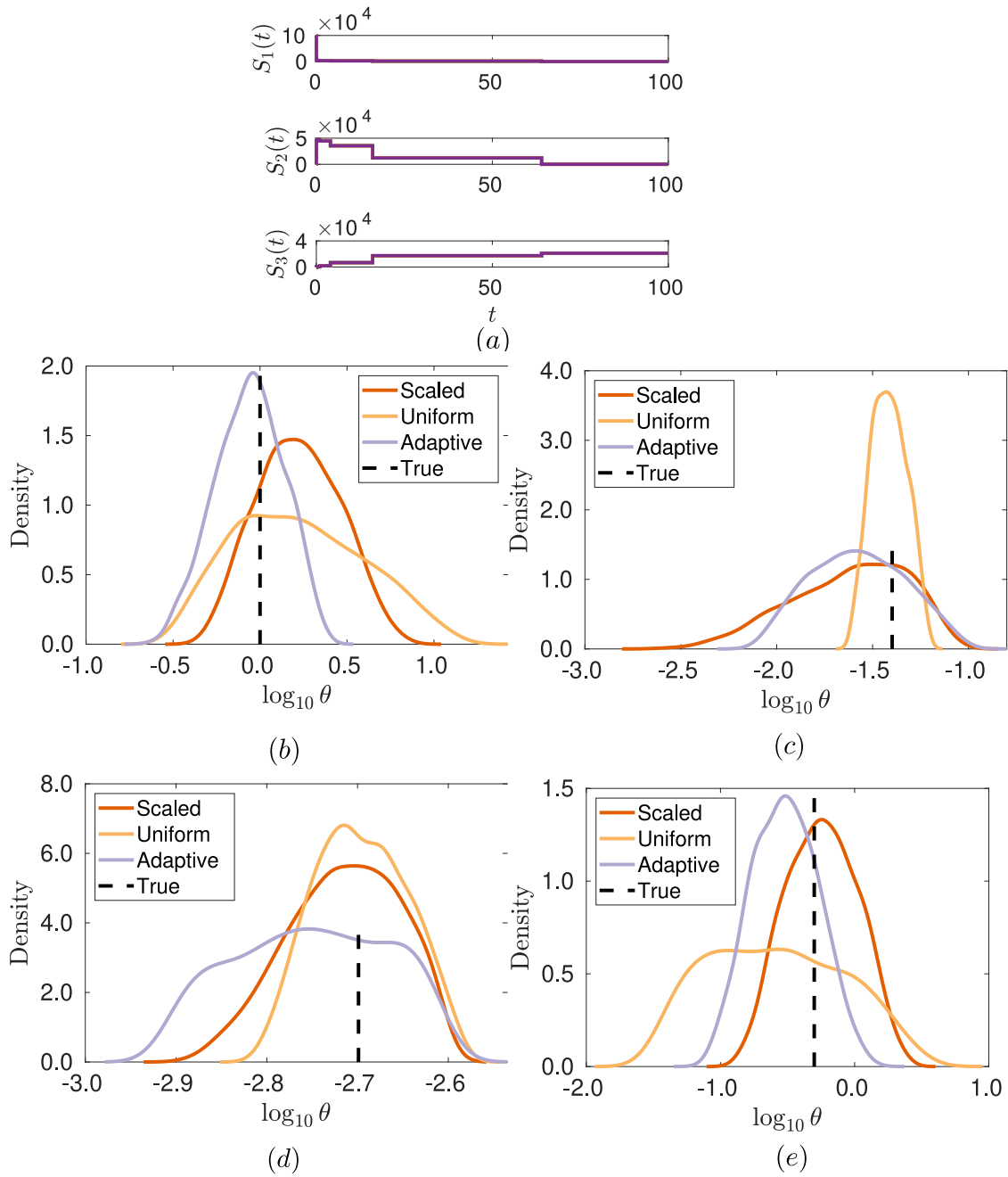


Figure 5.5. Posteriors for parameters $\theta = (k_1, k_2, k_3, k_4)$ in the dimerization system for different weights in the ABC distance function. ABC-SMC was used with 10 generations and $N = 2000$ particles at each generation with the posterior constructed from the closest 50% of the simulations ($\alpha = 0.5$). (a) shows typical output from the model for the true parameters, for each species, S_i . Posterior marginal distributions for parameters k_1, k_2, k_3, k_4 are shown in (b) to (e).

model were assumed, then sampling the true posterior distribution via particle MCMC would be possible, but the algorithm breaks down in the absence of an

observation noise model.

5.4.6 Simple spatial model

Spatial models produce very high dimensional data, containing information about dynamics in both space and time. Here, we consider a simple spatial model in one dimension to describe the spreading of particles by diffusion without volume exclusion. We divide our spatial domain $X \in [-1, 1]$ into m boxes or voxels, and label the numbers of particles in voxels $1, \dots, m$ as $S_1, \dots, S_m$, respectively. Particles can jump between neighbouring voxels at rate $\theta = D/h^2$, where D is the macroscopic diffusion constant and h is the width of the voxel. We assume zero flux conditions at $X = \pm 1$ and take $m = 8$, so that $h = 1/4$. As an initial condition, we place 10 particles in each of the $m/2$ voxels on the left-hand side of the domain where $x < 0$, and allow the system to evolve over the time interval $[0, 20]$. We observe the system at $n = 8$ equally spaced time points, and take as the summary statistic the time series for each voxel, $s(\mathbf{x}) = [S_1(t_0), \dots, S_1(t_n), S_2(t_0), \dots, S_2(t_n), \dots, S_m(t_0), \dots, S_m(t_n)]$, where $S_i(t_j)$ is the number of particles in voxel i at time point t_j . Using synthetic data simulated with $\theta = 0.1$, we attempt to recover the jump rate θ . The results of parameter inference for this problem are shown in Figure 5.6, where we have used a prior uniform on $\log_{10}(\theta)$ over the interval $[10^{-4}, 10^0]$. We successfully obtain an informative unbiased posterior for θ using the adaptive choice of weights, with a notable improvement in comparison to the other methods for selecting the weights.

5.4.7 Computational overhead

If the proposed approach of adapting the weights of each of the summary statistics is to be used in practice, we must ensure that the increases in the quality of the resulting posterior justify the computational overhead required for the search process. Otherwise, it would be preferable simply to generate the posterior using ABC-SMC

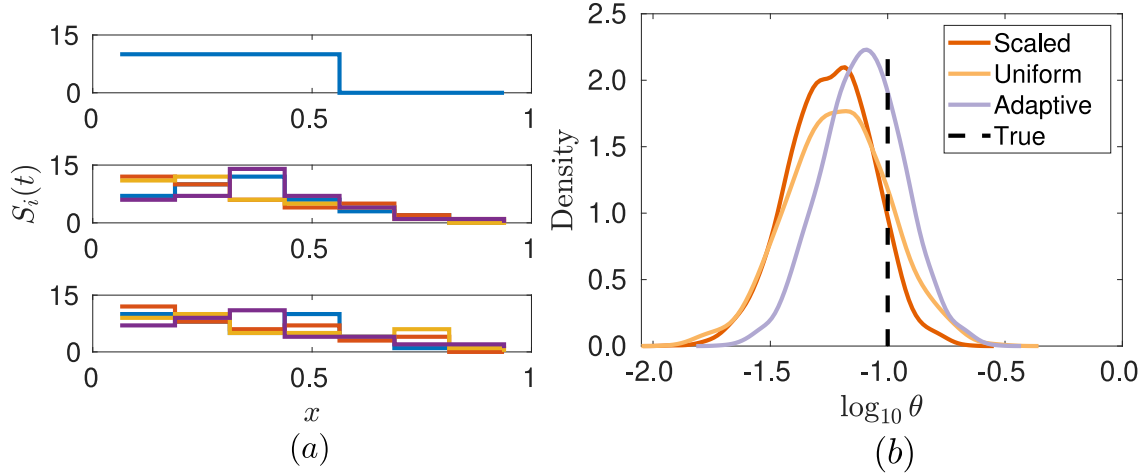


Figure 5.6. Posteriors for parameter θ in the simple diffusion model for different weights in the ABC distance function. ABC-SMC was used for the inference with 10 generations and $N = 2000$ particles at each generation with the posterior constructed from the closest 50% of the simulations ($\alpha = 0.5$). (a) shows the spatial profile at three different time points ($t = 0, 10, 20$) and demonstrates the variability in the output for this spatial process across four realizations with the same parameter, $\theta = 0.1$. In (b), we compare the posteriors obtained for θ with different choices of weights.

with more samples. Therefore we are interested in evaluating the computational overhead of the search process, and how to limit the cost of the search in higher dimensions.

Using the uniform toy problem, as described in Section 5.4.1, we ran Algorithm 5 with $N = 1,000$, $\alpha_1 = 0.5\%$ for one generation. To compare this to ABC-SMC with uniform weights, we performed parameter inference with uniform weights using both $N = 1,000$, $\alpha_1 = 0.5\%$ and $N = 1,000$, $\alpha_2 = 0.001\%$. The value of α_2 was chosen such that an equal length of computation time was spent in the search steps to find the summary statistic weights in Algorithm 5, as was spent in generating extra samples in ABC-SMC with uniform weights. The number of samples in the final posterior approximation was equivalent.

In this case, adaptively choosing weights using Algorithm 5 resulted in a significantly greater distance between the prior and posterior, and reduced the bias in the posterior compared to running ABC-SMC with more samples, as measured by the

Test problem	Hellinger distance between prior and posterior	Bias in posterior
Uniform toy model	0.800 / 0.809 / 0.828	0.184 / 0.141 / 0.134
Bimodal toy model	0.865 / 0.867 / 0.866	0.029 / 0.034 / 0.031
Dimerization	0.921 / 0.919 / 0.921	0.563 / 0.493 / 0.421
Diffusion	0.700 / 0.710 / 0.743	0.533 / 0.475 / 0.125

Table 5.1. Performance of Algorithm 5 compared with increasing the number of samples in ABC-SMC. Results are shown for each of the test problems in the form: ABC-SMC with N and α_1 / ABC-SMC with N and α_2 / Algorithm 5 with N and α_1 . Highlighted in bold is the method with best performance according to each metric.

distance between the maximum posterior estimate and the true parameters. These results, which represent improvements in the posterior for the same computational cost, are shown in Table 5.1 and the same procedure was used for the other test problems. Values shown have been averaged over 10 repeated runs for different datasets.

5.4.8 Consistent weights

Ideally, the search process should find the global optimum weight vector, so that if Algorithm 5 is run multiple times the same weight vector is obtained. In practice, for the examples we have explored, the function to be optimized (distance between prior and posterior as a function of the distance weights) is very flat with respect to some of the distance weights. This makes it hard to consistently identify a global maximum. In Figure 5.7, we explore how the chosen weights vary for the uniform toy model. We can interpret this as the algorithm identifying the informative summary statistics and appropriately using the information from these, while allowing weights for other summary statistics to take a range of values without much effect on the resulting posterior. The largest weight is given to the most informative summary statistic. In the case of the uniform toy model, this is $s(\mathbf{x}) = \max_i x_i$.

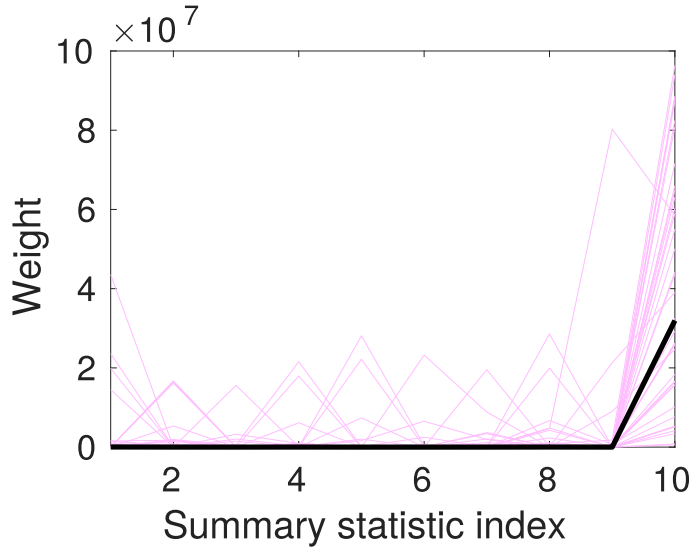


Figure 5.7. The optimal distance weights found from the search procedure after 40 successive runs of Algorithm 5 on the toy model test problem (described in Section 5.4.1 in (a) and on the death process test problem (described in Section 5.4.4) in (b). Parameters as for Figures 5.1 and 5.4 in each case. The faint purple lines show the resulting summary statistics weights from repeated runs of the ABC distance weight algorithm, while the black line shows the mean of the weights selected.

5.4.9 Comparison to dimensionality reduction methods

Adaptively choosing the summary statistic weights within the ABC distance function can be seen as achieving a similar goal to summary statistic dimension reduction techniques [35, 262]. These techniques either project high-dimensional summary statistics into a lower dimensional subspace, or select an optimal subset of summary statistics via some optimality criterion. In contrast, a similar effect is achieved here when the statistics are combined in the weighted Euclidean distance function, $d_{\mathbf{w}}(\mathbf{x}_1, \mathbf{x}_2)$, by weighting summary statistics to take account of both their inherent scale, and also their relative contribution towards the posterior distribution. Uninformative summary statistics are automatically assigned a lower weighting, while more informative summary statistics are given high weights relative to their scale.

We tested the adaptive weight selection algorithm against the semi-automatic ABC method [37] (using the implementation available via `abctools` [272]), and the

Test problem	Hellinger distance between prior and posterior	Bias in posterior
Uniform toy model	0.822 / 0.806 / 0.815	0.055 / 0.059 / 0.056
Bimodal toy model	0.873 / 0.872 / 0.855	0.025 / 0.024 / 0.028
Dimerization	0.930 / 0.914 / 0.925	1.109 / 1.274 / 0.783
Diffusion	0.723 / 0.680 / 0.709	0.358 / 0.399 / 0.373

Table 5.2. Comparison of the quality of the posteriors obtained using different methods to combine summary statistics. Results given as adaptive method/Barnes et al. [36]/Fearnhead and Prangle [37]. Bold text highlights the best performance on a metric for a test problem.

subset selection method of Barnes et al. [36] based on an approximate sufficiency criterion. In general, for the test problems considered, as described in Section 5.4, the method equals the performance of the competing methods, as shown by the metrics in Table 5.2. A larger value of the Hellinger distance indicates a greater distance between prior and posterior. The bias gives the distance between the posterior mean and the true parameter value. Values shown have been averaged over 10 successive runs of each method on different datasets. In implementing these methods, we have used only ABC rejection sampling, equivalent to a single generation of ABC-SMC, to compare the methods. This hinders the method of Algorithm 5 which makes use of adaptive selection of weights, but is evaluated only on a single generation. In practice, these results mean that the method outlined in Algorithm 5 for adaptively choosing the weights of summary statistics produces a more informative posterior than competing methods based on dimensionality reduction of summary statistics.

5.5 Discussion

In this work, we have presented a method for improving the quality of posteriors resulting from approximate inference using ABC-SMC by optimizing the weights of the ABC distance function, $d_{\mathbf{w}}(s(\mathbf{x}_1), s(\mathbf{x}_2))$. By applying the methodology to several test problems we have demonstrated that the novel, adaptive method allows effective combination of summary statistics. We see superior performance using

Algorithm 5 in comparison with naive choices of uniform weights or using the scale of the summary statistics based on the method of Prangle [111]. Further benefits of adapting the weights include removing the requirement for design and selection of summary statistics ‘by hand’.

In the future, our method for automatically adapting the weights of the ABC distance function could be combined with other methods for dimensionality reduction of summary statistics to further improve the quality of posteriors produced with ABC for given computational effort. A particular area to consider would be how best to combine optimization of the distance weights for ABC and dimensionality reduction of the summary statistics. These are related approaches that can work well together. One approach that could be explored, for example, is enforcing some sparsity of the weights during the search step of the weights optimization. By setting some weights to be explicitly zero, we exclude the corresponding summary statistics, effectively reducing the dimensionality of the summary statistics. Further investigations could explore how best to sample sparse subsets of weights in high dimensions. The adaptive framework for maximizing the distance between prior and posterior could also incorporate choice of other hyperparameters needed for implementation of ABC-SMC methods such as the tolerance ϵ , in addition to ABC distance weights.

In summary, we propose a computationally efficient search procedure to identify a set of optimum weights to allow us to combine summary statistics within the ABC distance function in such a way that the gain in information in the posterior over the prior is maximized.

Chapter 6

Experimental verification of a
coarse-grained model of mRNA
localization reveals robustness
regulated by crowding at ring
canals

6.1 Introduction

Robust control of gene expression in both space and time is of central importance in the regulation of cellular processes, and for multicellular development. However, the mechanisms by which robustness is achieved are generally not identified or well understood. For example, mRNA localization by molecular-motor-driven transport is crucial for cell polarization in numerous contexts, but the regulatory mechanisms that enable this process to take place in the face of noise or significant perturbations are not fully understood.

Biological systems are constantly subjected to noise from both internal and external sources, and also to random mutations. It is crucial that key driving mechanisms are able to buffer against such insults, thereby enabling organisms to display phenotypes that are robust to perturbation. Localization of mRNA is a fundamental, and crucially important, mechanism that enables the polarization of cells such as oocytes and early embryos [43–45] (see Chapter 3). However, despite the widespread nature of mRNA localization, there is still much that we do not understand, in particular what ensures robustness. To this end, we combine mathematical modelling with quantitative experimental data to investigate the regulatory mechanisms controlling mRNA localization, capturing, in particular, the roles of transport and production. We focus our study, as in Chapter 3, on early development of the model organism *Drosophila*, where maternal mRNA prepatterns the oocyte [43–45].

In this chapter, we use a combined experimental-theoretical approach to characterize the robustness of *gurken/TGF- α* mRNA localization in *Drosophila* egg chambers. We develop a coarse-grained ODE model of *gurken* (*grk*) transport through the *Drosophila* egg chamber, including all 15 nurse cells and the developing oocyte, with terms representing both the production and transport of mRNA¹. Using Bayesian inference methods, we fit this simple model to quantitative imag-

¹In referring to production, we include mRNA transcription, export through the nuclear pore complex, and assembly of RNA-protein complexes.

ing data obtained using smFISH (single molecule fluorescence *in situ* hybridization) [203, 204]. By analyzing both the steady state and dynamic behaviours of the model, we provide estimates for the rates of different steps of the localization process, as well as the extent of directional bias in transport through the ring canals. The model predicts that mRNA synthesis and transport must be balanced to maintain robustness, a prediction which we tested experimentally using an over-expression mutant. Surprisingly, the over-expression mutant fails to display the anticipated degree of overaccumulation of mRNA in the oocyte predicted by the model. Through careful model-based analysis of quantitative data from the over-expression mutant we conclude that saturation of transport of mRNA through ring canals engenders robustness of the localization process, in the face of significant variation in the levels of mRNA synthesis.

Combining quantitative modelling approaches with experimental data in this way allows us to explore underlying biological mechanisms through the generation of testable model predictions with appropriate quantification of uncertainties. In particular, it enables estimation of the mRNA production and transport rates in the model, and predicts that there is a balance between production and transport for the localization of *gurken* in the *Drosophila* egg chamber.

In the process of fitting the model to microscopy imaging data, we examine the formation of higher-order RNA-protein complexes in the oocyte. In the nurse cells, mRNA is assembled into complexes containing both mRNA and various proteins, and these RNA complexes are then transported into the oocyte. An important aspect of the transport process is that the RNA complexes are remodelled (so that more mRNA transcripts are contained in each complex) upon transit through the ring canals that connect the nurse cells to the oocyte [182, 273], leading to larger complexes in the oocyte [274]. We use quantitative data analysis to provide an estimate of the extent of this assembly of higher-order complexes for *gurken*.

In addition, we explore the question of whether transport of RNA complexes

occurs unidirectionally or bidirectionally through ring canals in the *Drosophila* egg chamber. Some evidence that transport through ring canals is unidirectional has previously been provided [273]. However, since ring canals are small relative to the nurse cells, the passage of complexes through them is relatively difficult to observe *in vivo*. Our study offers further evidence in support of the hypothesis that transport through ring canals is strongly biased towards the oocyte.

Finally, we use the coarse-grained ODE model to make testable predictions about the behaviour of an over-expression mutant with increased production of mRNA. We demonstrate strong agreement between the model predictions and observed data for nurse cells close to the oocyte, but find discrepancies for nurse cells far from the oocyte. Surprisingly, we find the numbers of complexes localized in the oocyte of the over-expression mutant are very similar to wild type, whereas the model predicts numbers should increase significantly. To probe the reasons for this disparity, we consider a suite of extended models incorporating inhomogeneous production, density dependent transport and crowding-induced blocking of transport through ring canals. We show, via statistical techniques that allow quantitative model comparison, that the crowding-induced blocking mechanism is best supported by the data, and that a model incorporating blocking is capable of producing distributions of RNA complexes similar to those observed experimentally. In strong support of this model-based prediction, we observe accumulation of complexes at ring canals in smFISH microscopy data for the over-expression mutant consistent with this crowding-induced blocking mechanism.

6.2 Modelling approaches to RNA regulation in oogenesis

Despite their increasing use across the biological sciences, mathematical and computational modelling approaches have not to-date been widely adopted in the study of mRNA localization. Trong et al. [209] have demonstrated how localization of mRNA

can be achieved in a *Drosophila* oocyte using a partial differential equation (PDE) description of mRNA dynamics and a stochastic model of cytoskeleton dynamics. Ciocanel et al. [210] used a PDE model of RNA localization in *Xenopus* oocytes to quantify aspects of the transport process using FRAP (fluorescence recovery after photobleaching) data, and others have considered models of RNA transport to examine the behaviour of individual molecular motors [195, 218], or the cytoskeleton structure [219]. However, these works have focused on dynamics within a single cell, the oocyte, as has other work in model systems such as mating budding yeast and the neuronal growth cone [275, 276]. Recent research by Alsous et al. [277] has quantified collective growth and size control dynamics in *Drosophila* nurse cells. Similarly, our work considers dynamic behaviours in multiple cells. However, our focus is to examine the regulation and robustness of mRNA localization by considering a coarse-grained ODE model of the entire *Drosophila* egg chamber, and using it to interrogate quantitative experimental data. Further, our work highlights that by combining an incredibly simple mathematical model with quantitative experimental data one can draw relevant conclusions and make testable predictions about key biological processes that would not, otherwise, have been possible.

6.3 A coarse-grained model for *gurken* mRNA localization

To investigate the mechanisms governing robustness of mRNA localization, we have developed a minimal, compartment-based ODE model of *gurken* localization in a *Drosophila* egg chamber. We view each cell as a separate compartment and assume that mRNA complexes are produced in each of the nurse cells [180]. We assume that there is transport of mRNA complexes between cells that are connected by ring canals (see Figure 3.3), and that there is negligible degradation (there is evidence that the majority of maternal mRNAs are stable during oogenesis over the timescales of interest [214, 278]; see also Section 6.14.1).

The model can be written as

$$\frac{d\mathbf{y}}{dt} = a\mathbf{v} + b\mathbf{B}(\nu)\mathbf{y}, \quad (6.1)$$

where the i^{th} entry of the vector $\mathbf{y}$ is the number of mRNA complexes in cell i (indexed as in Figure 3.3). mRNA complexes are produced at constant rate $a > 0$ in units of [particles hr⁻¹], and the vector $\mathbf{v} \in \mathbb{R}^{16}$ lists the cells that produce mRNA; for the *Drosophila* egg chamber, where mRNA is produced in all cells except the oocyte, $\mathbf{v} = (0, 1, \dots, 1)^\top$ since the oocyte is transcriptionally silent for most of oogenesis [279]. Transport of complexes between cells takes place at constant rate $b > 0$ in units of [hr⁻¹]. The matrix $\mathbf{B}$ describes the network of connected nurse cells (so that entry (i, j) is non-zero only when cells i and j are connected by a ring canal). Transport bias (towards the oocyte) is represented by parameter ν so that complexes in a given nurse cell move towards the oocyte at relative rate ν and away from it at relative rate $1 - \nu$ (see schematic in Figure 6.7). The precise form of $\mathbf{B}$ is provided in Equation (6.2). Note that a similar approach has been employed

$$\mathbf{B}(\nu) = \begin{pmatrix} -4+4\nu & \nu & \nu & 0 & \nu & 0 & 0 & 0 & \nu & 0 & 0 & 0 & 0 & 0 & 0 \\ 1-\nu & -3+2\nu & 0 & \nu & 0 & \nu & 0 & 0 & 0 & \nu & 0 & 0 & 0 & 0 & 0 \\ 1-\nu & 0 & -2+\nu & 0 & 0 & \nu & 0 & 0 & 0 & 0 & \nu & 0 & 0 & 0 & 0 \\ 0 & 1-\nu & 0 & -2+\nu & 0 & 0 & 0 & \nu & 0 & 0 & 0 & \nu & 0 & 0 & 0 \\ 1-\nu & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \nu & 0 & 0 \\ 0 & 1-\nu & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & \nu & 0 & 0 \\ 0 & 0 & 1-\nu & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & \nu & 0 \\ 0 & 0 & 0 & 1-\nu & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & \nu \\ 1-\nu & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\nu & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1-\nu & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\nu & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1-\nu & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\nu & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1-\nu & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\nu & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1-\nu & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\nu & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1-\nu & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\nu & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1-\nu & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\nu \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-\nu & 0 & 0 & 0 & 0 & 0 & 0 & -\nu \end{pmatrix} \quad (6.2)$$

by Alsous et al. [277] to show that differences in cell sizes in the *Drosophila* egg chamber result from the characteristic cell network generated through incomplete divisions, and the resulting ring canal connections. As an initial condition, we assume there are no mRNA complexes in any of the cells at time t_0 (where t_0 is determined from the linear model described in Section 6.4.4). The system of ODEs is solved numerically using the fourth order Runge-Kutta scheme implementation contained in the Boost C++ library [280].

Typical behaviour of the model is shown in Figure 6.1. Much of the biologically relevant behaviour occurs in the quasi-steady-state regime where we have linear growth in the number of mRNA complexes in the oocyte, and a characteristic distribution of mRNA complexes across the nurse cells. Typical dynamic behaviour for each nurse cell is shown in Figure 6.2.

6.4 Bayesian inference framework

We connect the model directly to quantitative experimental data so that it can provide relevant predictions and insight. We use a Bayesian inference framework to take account of both measurement and process uncertainty, incorporating the mechanistic model stated in Equation (6.1), a model of the measurement process, and prior knowledge of parameters.

The biological process model (Equation (6.1)) can be related to the observed data via a measurement model. The measurement model accounts for any errors in processing the data, in addition to raw experimental error. However, we first note that mRNA complexes consist of multiple individual mRNA transcripts, and recall that upon entry to the oocyte there is some higher-order assembly of these complexes [182, 273, 274]. Since we have quantitative data on the number of mRNA complexes in each cell, rather than the number of individual transcripts, we account for this higher-order assembly of complexes in the oocyte in the measurement model. We assume that in the oocyte we observe ϕy_1 complexes, where $\phi \in (0, 1]$ is a scaling parameter. We can interpret ϕ as a ratio between the number of transcripts per complex in the nurse cells as compared to the oocyte.

We use a negative binomial distribution² for the measurement model, so that

²We use a parameterization of the negative binomial distribution via a mean parameter, μ , and overdispersion parameter, ψ , such that

$$p(y|\mu, \psi) = \binom{y + \psi - 1}{y} \left(\frac{\mu}{\mu + \psi} \right)^y \left(\frac{\psi}{\mu + \psi} \right)^\psi.$$

With this parameterization, if $y \sim \text{NB}(\mu, \psi)$, then $\mathbb{E}[Y] = \mu$ and $\text{Var}[Y] = \mu + \mu^2/\psi$.

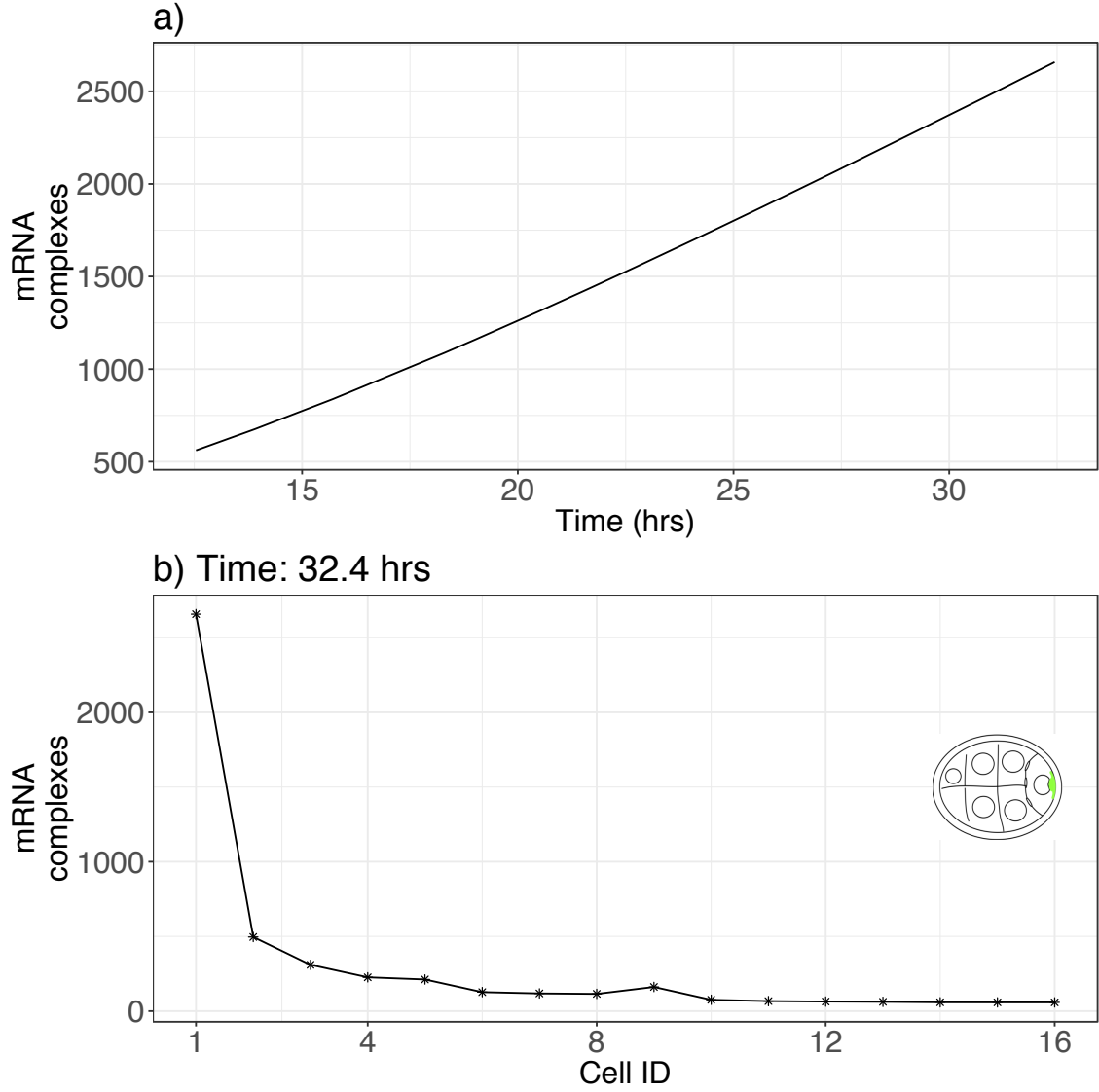


Figure 6.1. Predictions from the ODE model (6.1). (a) Dynamic behaviour of the number of complexes in the oocyte; the remaining nurse cells are shown in Figure 6.2. (b) The distribution of complexes across the cells in the egg chamber at $t = 32.4$ hrs demonstrates the accumulation of mRNA in the oocyte. The numbering for the Cell ID on the x axis corresponds to that given in Figure 3.3. Parameters are $a = 10$ particles hr^{-1} , $b = 0.20$ hr^{-1} and $\nu = 0.90$.

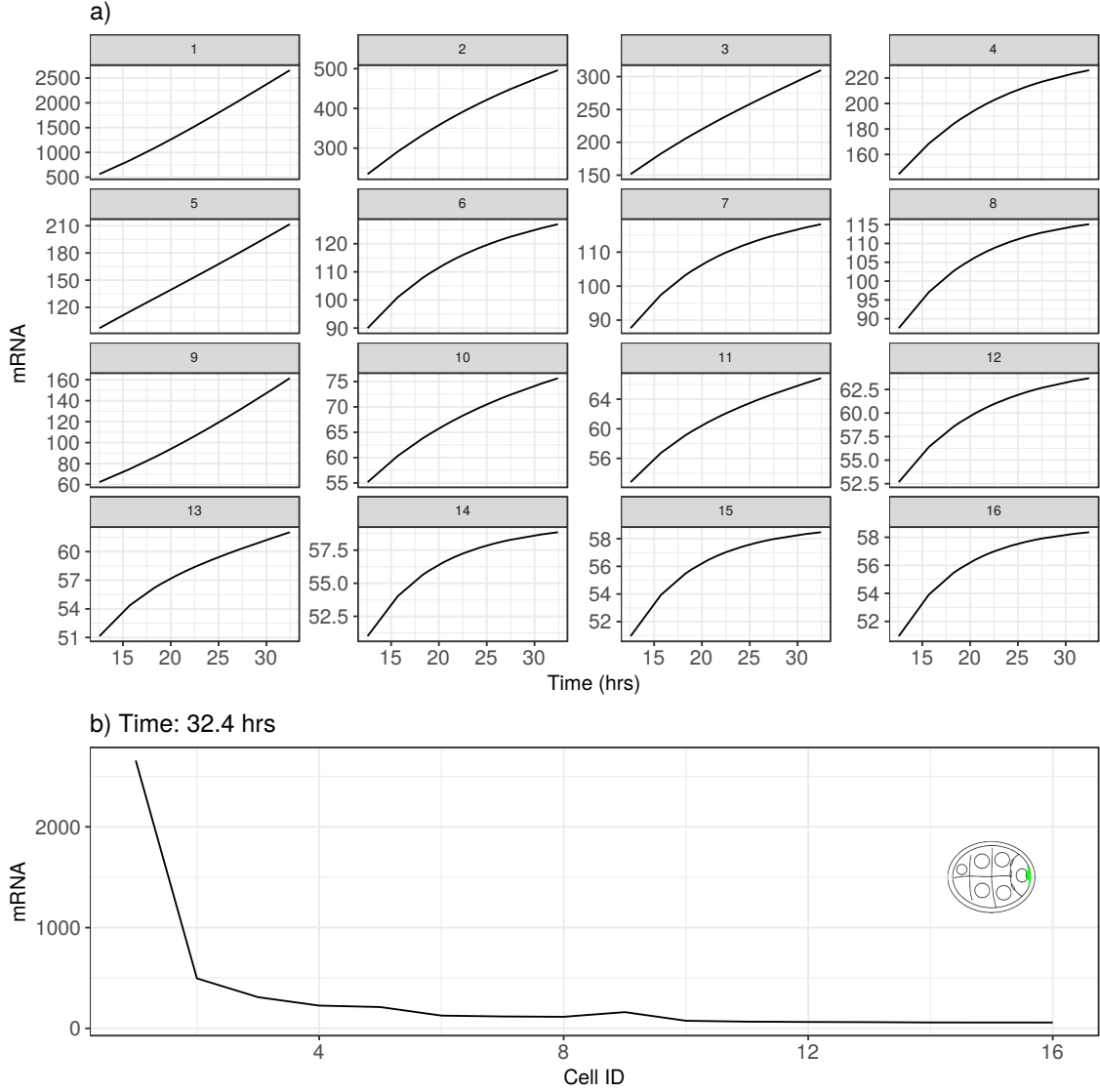


Figure 6.2. Typical behaviour of the coarse-grained ODE model across each of the cells in the egg chamber is shown in a). The distribution of mRNA across cells is shown at a single time point at 32.4 hrs. The numbering for the Cell ID on the x axis corresponds to that given in Figure 3.3. Parameters are $a = 10 \text{ particles hr}^{-1}$, $b = 0.20 \text{ hr}^{-1}$ and $\nu = 0.90$.

measured mRNA complex counts, $\mathbf{z}$, are given by

$$\mathbf{z} \sim \text{NB}(\Phi \mathbf{y}, \sigma), \quad (6.3)$$

where Φ is a diagonal matrix with entries $[\phi, 1, 1, \dots, 1]$, σ is a parameter controlling the magnitude of the measurement error, and $\mathbf{y}$ is the solution of the biological

process model (Equation (6.1)), giving the numbers of RNA complexes in each cell. We take the product of the likelihoods over each of the observed cells (including nurse cells and oocyte), with each biological sample corresponding to a unique time point, to give the full likelihood. The parameters to be inferred are a , b , ν , ϕ and σ .

6.4.1 Choice of prior distributions

The prior distributions are chosen to incorporate existing biological knowledge of feasible parameter ranges. For the rate parameters, a and b , we choose Truncated Normal³ priors with parameters $\mu = 0$, $\sigma = 10$ and support $(0, \infty)$ in units of [particles hr⁻¹] and [hr⁻¹], respectively. For the assembly parameter, ϕ , we use a Truncated Normal prior with $\mu = 0.30$, $\sigma = 0.036$ and support $(0, 1]$. This is a strong prior informed by experimental evidence (see Section 6.5). For the transport bias parameter, ν , we assume a uniform prior, $U(0.5, 1)$; it is clear that transport is biased towards the oocyte, but we avoid providing any further information about ν . For the measurement noise parameter, σ , we choose a broad Truncated Normal prior with $\mu = 0$, $\sigma = 10$ and support $(0, \infty)$. We check that these priors are compatible with the observed data by simulating from the prior predictive distribution, as suggested in the Bayesian workflow outlined by Gabry et al. [71] (see Figure 6.3).

6.4.2 Implementation of Monte Carlo methods

We use the open source software package Stan *mc* – *stan.org* (version 2.17.0) [65] to perform MCMC sampling from the posterior distribution. Stan uses Hamiltonian Monte Carlo techniques to navigate the geometry of the posterior distribution efficiently [64], along with automatic differentiation to compute gradients.

³ Note that we define a Truncated Normal distribution with parameters μ and σ and support on an interval $J \subset \mathbb{R}$ via the following rescaled probability density function

$$p(x; \mu, \sigma, J) = \frac{\mathcal{N}(x; \mu, \sigma^2)}{1 - \int_{\mathbb{R} \setminus J} \mathcal{N}(s; \mu, \sigma^2) ds},$$

where $\mathcal{N}(x; 0, 1^2)$ is the probability density function of a standard unit normal distribution.

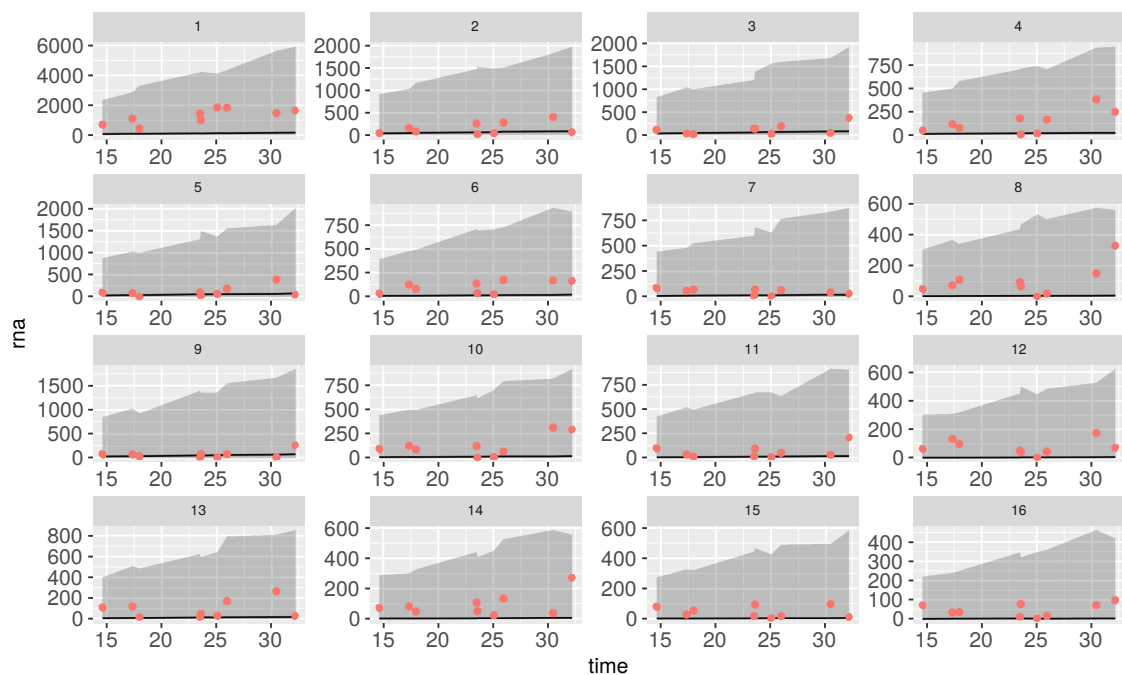
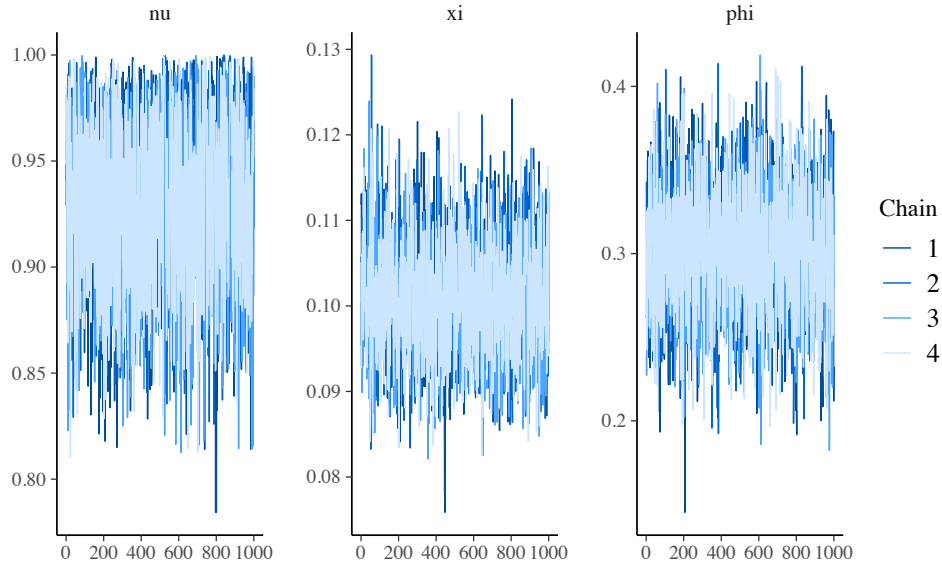


Figure 6.3. Predictions from the model given by Equation (6.1) before it has been fitted to the data, by drawing samples from the prior predictive distribution. Each separate facet plot gives predictions for a different nurse cell. The red dots show the observed wild type data. The shaded grey ribbon gives the 95% credible interval of predictions from the prior predictive distribution, while the black line gives the median prediction. The red dots clearly lie within the range of possible model predictions suggesting that the model and priors are flexible enough to generate the data.

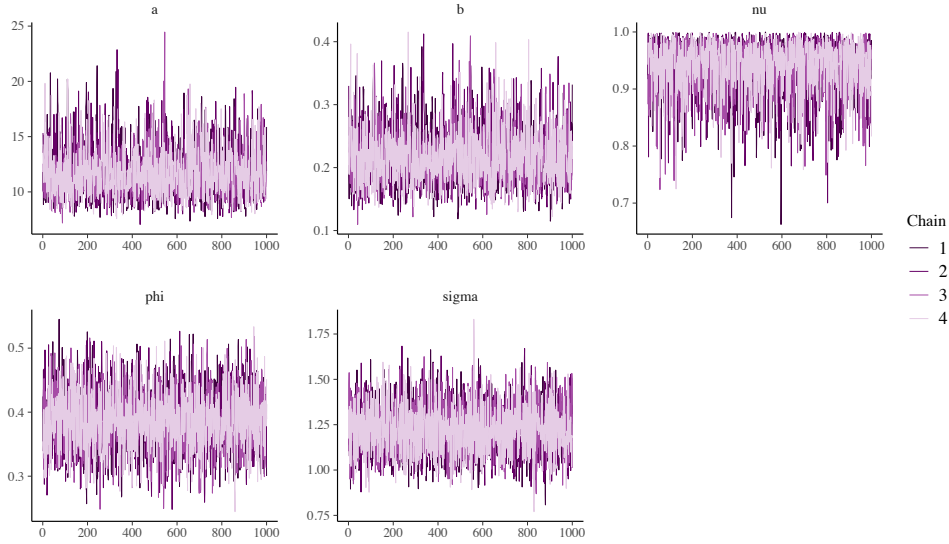
We run four chains in parallel for 2000 iterations with a burn in of 1000 iterations. Traceplots illustrating convergence of each of the chains are shown in Figure 6.4. Code to reproduce the analysis described in this work is available at https://github.com/shug3502/rstan_analysis.

6.4.3 Image analysis

Basic image handling and processing was carried out in FIJI (ImageJ V1.51d; *fiji.sc*, [223]). Image data was archived in OMERO (V5.3.5) [282]; image conversions were carried out using the BioFormats plugin in FIJI [283]. mRNA particles in the nurse cells and oocyte are identified and located using the FISH QUANT software [207]. We combine use of this software with manual segmentation of the



(a) Steady state



(b) Full dynamic model

Figure 6.4. Traceplots indicate convergence of MCMC chains for the model at steady state and the full model as described in Sections 6.6 and 6.7 respectively. Four Markov chains were run in parallel and sample from the same distribution, indicating convergence to the posterior distribution. HMC implemented via Stan [65] was used to perform this sampling, as described in Section 6.4.2. The potential scale reduction statistic [281], $\hat{R}$, is 1.0 for inference at steady state, and also 1.0 for the full problem indicating convergence in both cases. The minimum effective sample size, n_{eff} , across the model parameters is 1940 for inference at steady state, and 1630 for the full model.

nurse cells and oocyte performed in three dimensions using a GUI (graphical user interface) designed for the annotation of training and evaluation data for machine learning via software available from QBrain [284]. This enables quantification of the total numbers of complexes observed in each cell. Identification of cells within the egg chamber is performed by counting the number of ring canals connecting each cell, and which cells these connect. Data is processed from $n = 9$ egg chambers between stage 5 and stage 8 of oogenesis. Since the mRNA complexes are sparse in the nurse cells, using counts of particles is a valid method of quantifying the total *gurken* RNA expression [285].

6.4.4 Identification of developmental time via an exponential growth model

In order to relate the coarse-grained ODE model to data from a given egg chamber, we need to quantify how far through developmental time the egg chamber is, such that we can numerically integrate the ODE model up to this time. To identify a precise developmental time variable for each egg chamber, we apply an exponential growth model to measurements of the area of the median section. Effectively this allows us to use $\log(A)$, where A is the area of the median section, as a rescaled variable for the timescale of development, by fitting the following model by linear regression: $\log(A) = \log(A_0) + \tau t$, where the intercept $t_0 = \log(A_0)$ corresponds to the start of *gurken* mRNA production, and τ gives a rescaling of time. To fit this model, we manually identify the developmental stage of each egg chamber and take the time point corresponding to the midpoint of each stage as an estimate of its age (see Figure 6.5). This approach is supported by the results of Jia et al. [286], who use $\log(A)$ in a similar way to assist with automatic stage identification for *Drosophila* egg chambers.

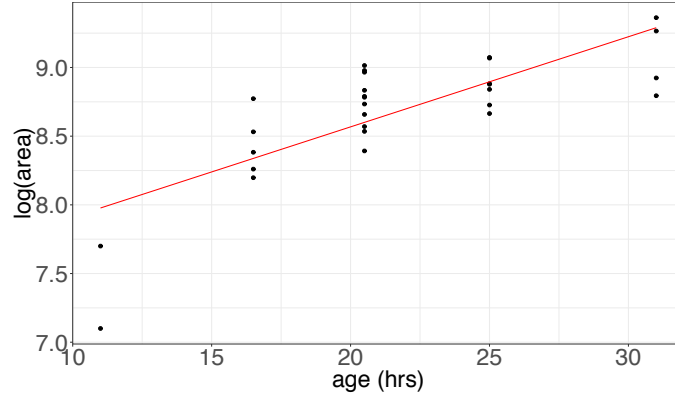


Figure 6.5. Fitting of exponential growth model to give timescale based on staging of egg chambers and measurements of area of the median section through the egg chamber. A linear model was used to fit the logarithm of the area as $\log(A) = \log(A_0) + \tau t$ via linear regression as described in Section 6.4.4. The values of $\log(A_0)$ and τ found were 7.57 and 0.0559 respectively.

6.5 Quantification of higher-order assembly of *gurken* mRNA complexes in the oocyte

Considering directly the total integrated intensity of mRNA complexes in the oocyte compared to those in the nurse cells allows us to directly estimate the relative numbers of mRNA transcripts in an mRNA complex in the oocyte compared to the nurse cells. We calculate the total integrated intensity for 448 foci in $n = 13$ example datasets for egg chambers from stage 5 to stage 8, including both the nurse cells and oocyte. By subtracting the background intensity, and dividing by the mean nurse cell intensity, we can obtain an estimate for ϕ , as shown in Figure 6.6: the median estimate gives $\phi = 0.345 \pm 0.048$.

We can now use this independent estimate of ϕ to specify a strong prior for ϕ when fitting the full Bayesian model. We obtain posterior estimates for the assembly parameter, ϕ , with 0.33 as a median value, and $[0.26, 0.39]$ as a 95% credible interval. This gives an estimate of between three and four nurse cell mRNA complexes assembled into one equivalent oocyte mRNA complex. The posterior distribution for ϕ is shown in pairwise plots of all the model parameters in Figure 6.9.

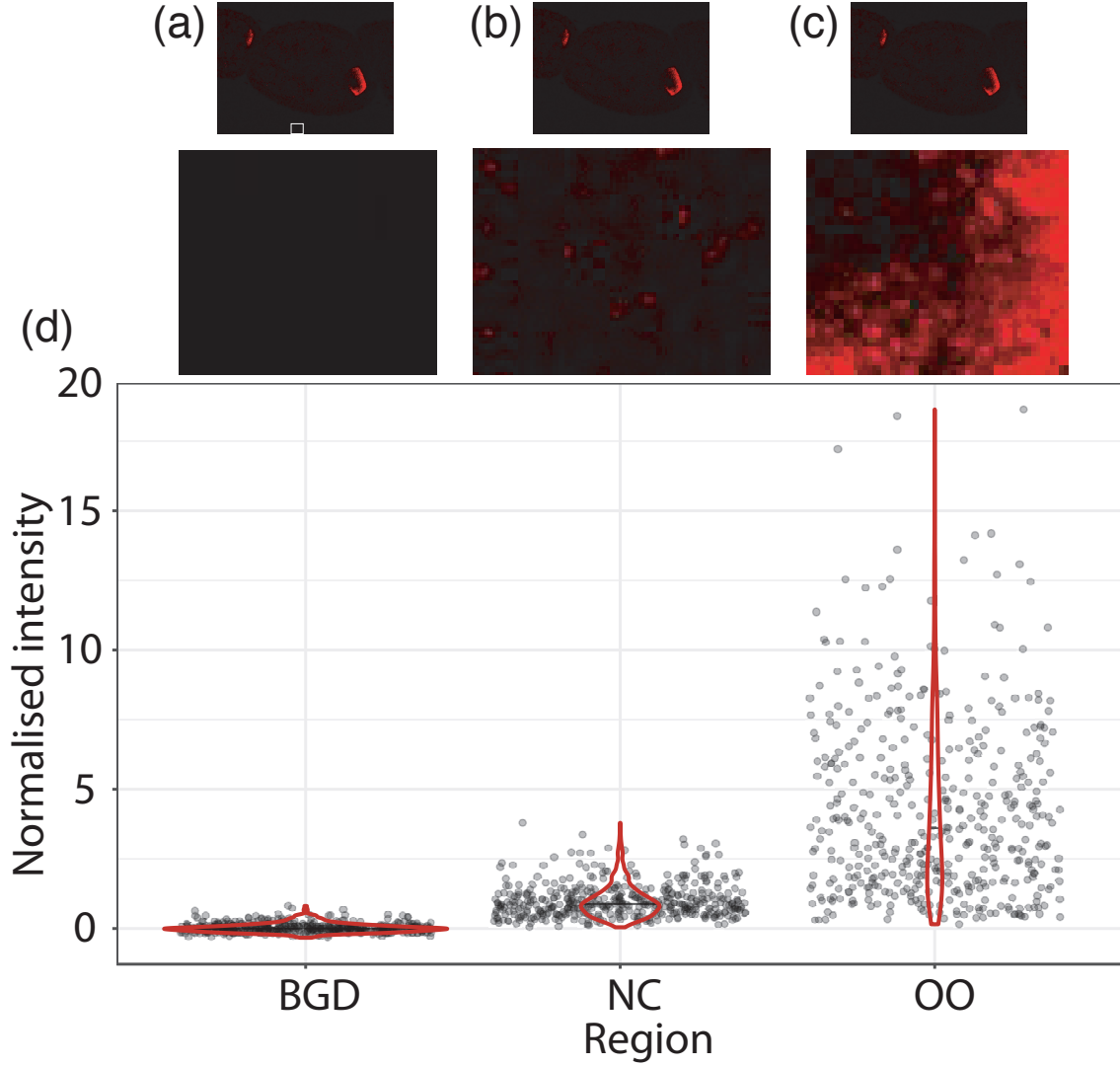


Figure 6.6. Assembly of higher order *gurken* complexes in the oocyte and quantification of the size of the assemblies in terms of equivalent complexes in the nurse cells. (a)-(c) typical complexes in background (BGD), nurse cells (NC) and the oocyte(OO). (d) By measuring total integrated intensity of these complexes in $n = 13$ egg chambers, we obtain distributions in each region, normalized by subtracting the average background value and scaling to a single nurse cell complex. We find that complexes in the oocyte are equivalent to a median of 3.33 times a single complex from the nurse cells.

6.6 Biased transport through ring canals revealed by parameter estimation at quasi-steady state

We now consider the question of whether transport of mRNA complexes through ring canals in *Drosophila* is unidirectional (towards the oocyte) or bidirectional (both towards and away from the oocyte). In addition, if transport is bidirectional, how biased is it towards the oocyte? Within the mathematical model (Section 6.3), the network structure (encoded in the matrix $\mathbf{B}$) and the bias parameter, ν , impose a distribution of mRNA complexes across the nurse cells. By inferring a posterior distribution for ν , through comparison of the model with quantitative data, we can estimate the extent to which transport through the ring canals is biased.

To simplify the model and ensuing Bayesian analysis, we consider the model behaviour in the large time, quasi-steady-state limit (which corresponds to the time scale on which observations are made). In this limit, we have $\mathbf{y} \approx 15a\mathbf{k}_2t$, where $\mathbf{k}_2$ is the quasi-steady-state distribution that can be uniquely calculated for a given matrix $\mathbf{B}$, as follows. (see Section 6.6.1).

6.6.1 Analytic solution to the coarse grained ODE model

The coarse-grained ODE model for the evolution of complex numbers in the egg chamber is given by

$$\frac{d\mathbf{y}}{dt} = a\mathbf{v} + b\mathbf{B}(\nu)\mathbf{y}, \quad (6.4)$$

where $\mathbf{y}$ gives the number of complexes in each of the cells of the egg chamber, $\mathbf{v}$ is a vector describing which cells produce mRNA, and $\mathbf{B}$ is a rate matrix where each row gives the relative rates at which RNA flows into and out of a cell. As an initial condition, we take $\mathbf{y}(t_0) = \mathbf{0}$.

Equivalently, we can write Equation (6.4) as

$$\mathbf{L}\mathbf{y} = f = a\mathbf{v},$$

where L is the linear operator

$$L\mathbf{y} := \frac{d\mathbf{y}}{dt} - b\mathbf{B}(\nu)\mathbf{y}.$$

We note that we can solve the homogeneous problem $L\mathbf{y} = 0$ by calculating the eigenvalues and eigenvectors of the matrix $b\mathbf{B}(\nu)$. The solution to the homogeneous problem is $\mathbf{y}_h = \mathbf{V}\mathbf{D}\mathbf{c}$, where $\mathbf{V}$ a matrix of eigenvectors of $b\mathbf{B}(\nu)$, D is a diagonal matrix with diagonal entries $e^{\lambda_i t}$, such that the λ_i are the corresponding eigenvalues, and $\mathbf{c}$ is a vector of constants determined by the initial conditions.

For the particular solution, we take $\mathbf{y}_p = \mathbf{k}_1 + \mathbf{k}_2 t$, for some vectors $\mathbf{k}_1$ and $\mathbf{k}_2$. Applying the linear operator to this particular solution gives

$$L\mathbf{y}_p = \mathbf{k}_2 - b\mathbf{B}\mathbf{k}_1 - b\mathbf{B}\mathbf{k}_2 t.$$

In addition, we note that the inhomogeneous problem, $L\mathbf{y} = a\mathbf{v}$, has a solution that has a constant term, but no term linear in t . Therefore, we must have $\mathbf{k}_2 \in \ker(\mathbf{B})$. Provided that $\nu \in [0, 1]$, this allows us to find $\mathbf{k}_2$ uniquely (up to a constant of proportionality), since the null space of $\mathbf{B}$ has dimension 1. That the null space has dimension 1 is due to the zero eigenvector of $\mathbf{B}$, which is present due to the fact that no complexes are lost from the system (we can also use this fact to determine the constant of proportionality).

Given we can determine $\mathbf{k}_2$ by finding the null space of $\mathbf{B}$, we can solve the following for $\mathbf{k}_1$:

$$-b\mathbf{B}\mathbf{k}_1 = a\mathbf{v} - \mathbf{k}_2. \tag{6.5}$$

Since $\mathbf{B}$ is not a full rank matrix, we cannot uniquely determine a solution to Equation (6.5); instead we obtain a line of solutions of the form $\mathbf{k}_1 = \mathbf{k}_1^* + \lambda\mathbf{k}_2$, for some parameter λ .

The constants in the vector $\mathbf{c}$ can be determined from the initial conditions, $\mathbf{y}_0$.

Initially at $t = 0$,

$$\begin{aligned} \mathbf{y}_0 &= \mathbf{V}\mathbf{c} + \mathbf{k}_1, \\ \implies \mathbf{c} &= \mathbf{V}^\top(\mathbf{y}_0 - \mathbf{k}_1^* - \lambda\mathbf{k}_2). \end{aligned}$$

Put together, we have a general solution for the ODE model of the form

$$\mathbf{y} = \mathbf{V}\mathbf{D}\mathbf{c} + \mathbf{k}_1 + 15a\mathbf{k}_2t. \quad (6.6)$$

6.6.2 Parameter estimation at quasi-steady state

Normalizing both the vector $\mathbf{k}_2$ (so that the normalized RNA complex count in the oocyte is unity), and the equivalent distribution measured from smFISH data, enables comparison of the distribution of complexes across the egg chamber under the assumption of a Gaussian measurement error model, $\mathcal{N}(0, \xi^2)$. We note that a different measurement error model is required here compared to Section 6.4, since the normalized mRNA distribution takes positive real values rather than integer counts. Priors for ν and ϕ are as stated above in Section 6.4.1. For the model at steady state, we additionally have a prior on the measurement noise, ξ , in the form of a Truncated Normal prior with $\mu = 0$, $\sigma = 0.1$ and support $(0, \infty)$.

In Figure 6.7, we show the marginal posterior distribution for the transport bias parameter, ν . The median of the marginal posterior distribution for ν is 0.95, with a 95% credible interval of $[0.81, 1.00]$. The posterior predictive distribution (Figure 6.7(c)) contains the observed data points within the 95% credible region for the complex distribution in each cell. Together these results strongly support the hypothesis that transport of complexes is significantly biased towards the oocyte.

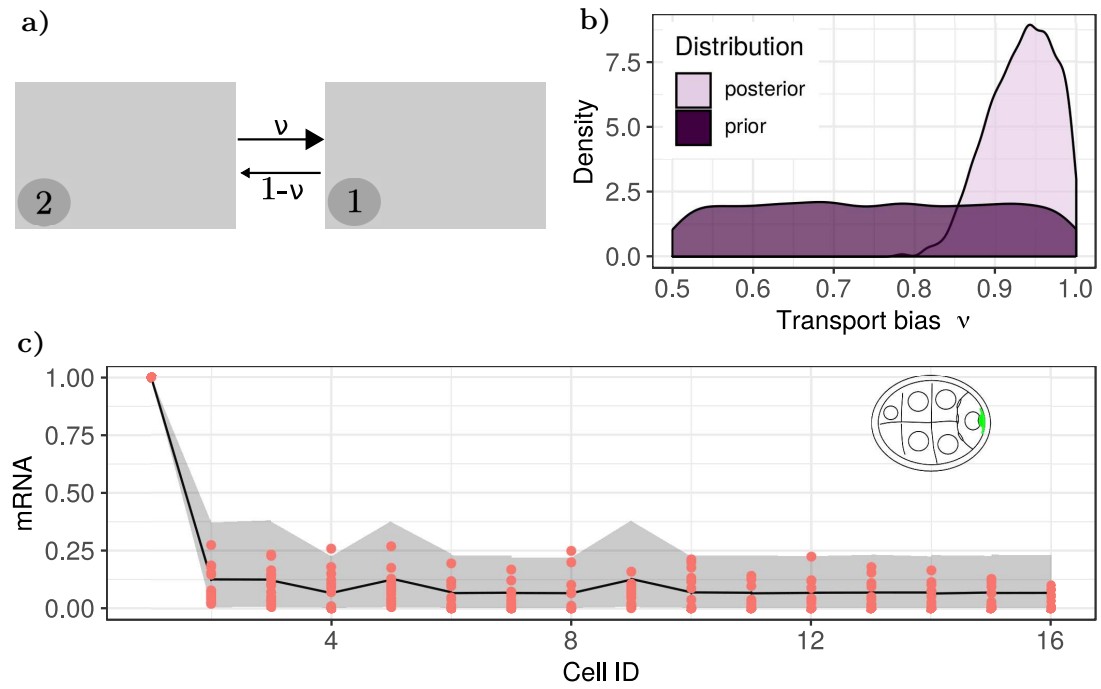


Figure 6.7. Estimation of the bias in transport through ring canals obtained via MCMC sampling for the coarse grained model at steady state, as described in Section 6.6. In a), we display a schematic diagram of biased transport between two compartments, with transport from cell 2 to cell 1 at rate ν , and transport from cell 1 to cell 2 at rate $1 - \nu$. In b), we display a density plot of the marginal posterior distribution for the transport bias parameter ν , compared to the prior for the same parameter which is uniform over $[0.5, 1]$. This shows evidence of strong bias in transport through ring canals. In c), we show the posterior predictive distribution with the raw data shown as red points. The shaded region shows a 95% credible interval of the distribution of predictions from the model. Here the distribution of mRNA across cells in the egg chamber is normalized such that the amount of mRNA in the oocyte is 1. The model provides a good fit to the data, since most of the data points lie within the grey envelope of the 95% credible interval.

6.7 Estimates of production and transport rates show production and transport are balanced

We obtain estimates for the rates of production, a , and transport, b , of mRNA complexes by applying the modelling approach outlined in Equation (6.1) and Section 6.4. We consider the full dynamic behaviour of the model, together with the smFISH dataset of $n = 9$ egg chambers (see Section C.1) and infer all model parameters (a , b , ν , ϕ and σ).

Upon sampling from the posterior for the full model, we find that the production rate, a , takes a median value of 12.9 and lies within a 95% credible interval of [9.5, 18.9] (in units of [particles hr⁻¹]). The transport rate, b , has a median value of 0.24 and lies within a 95% credible interval of [0.16, 0.35] (in units of [hr⁻¹]). The marginal posterior distributions for the rate parameters a and b are shown in Figure 6.8, and the posterior distribution for all the model parameters is shown pairwise in Figure 6.9.

Using these parameter estimates, we can compare the relative contributions of the production and transport terms in Equation (6.1) to the rate of change in the number of mRNA complexes. We have $12.9 \approx a \approx b\langle\tilde{\mathbf{y}}\rangle \approx 23.3$, where a is the complex production rate, b is the transport rate and $\langle\tilde{\mathbf{y}}\rangle = 97$ is the median number of *gurken* mRNA complexes at time t considered across all cells of the egg chamber. This result indicates that the mRNA complex production rate is of the same order of magnitude as the rate of complex transport, and therefore suggests that production and transport are balanced. We provide further evidence to support this hypothesis by evaluating the sensitivity of the model described in Section 6.3 to changes in the rate parameters a and b (see Section 6.7.1).

6.7.1 Sensitivity to rate parameters

We provide further evidence to support the hypothesis that production and transport are in tightly regulated balance by performing a sensitivity analysis of the model

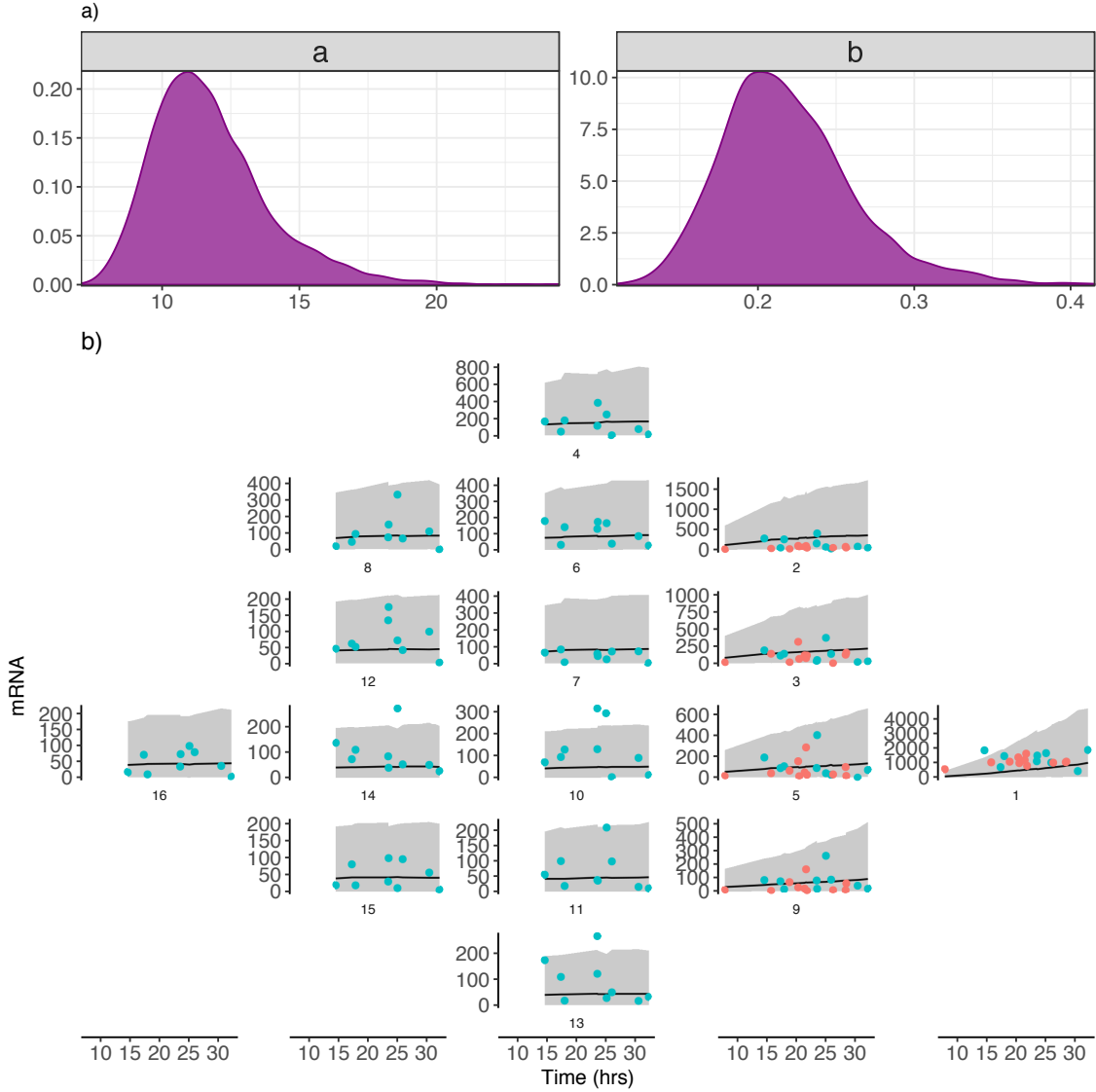


Figure 6.8. Estimation of all model parameters via MCMC sampling for the model described in Section 6.4 and predictions for numbers of mRNA complexes in each cell. In a), we display the posterior marginal distribution for the rate parameters a and b (other parameters can be found in Figure 6.9). In b), we show the posterior predictive distribution, which gives the predictions from the model with parameters fitted to the data. Each panel gives the prediction for each cell (numbered 1 for the oocyte through 2 to 16 for the nurse cells as in Figure 3.3) as a solid line, with measured experimental particle counts shown as points. A 95% credible interval for the predictions is shown as a shaded region. The blue points are observed data used to train the model, whilst the red points are observed data used to test the model (note that these red points are egg chambers for which only the oocyte and neighbouring nurse cells have been segmented). The subplots are arranged to highlight distance from the oocyte within the network of connections between nurse cells shown in Figure 3.3c).

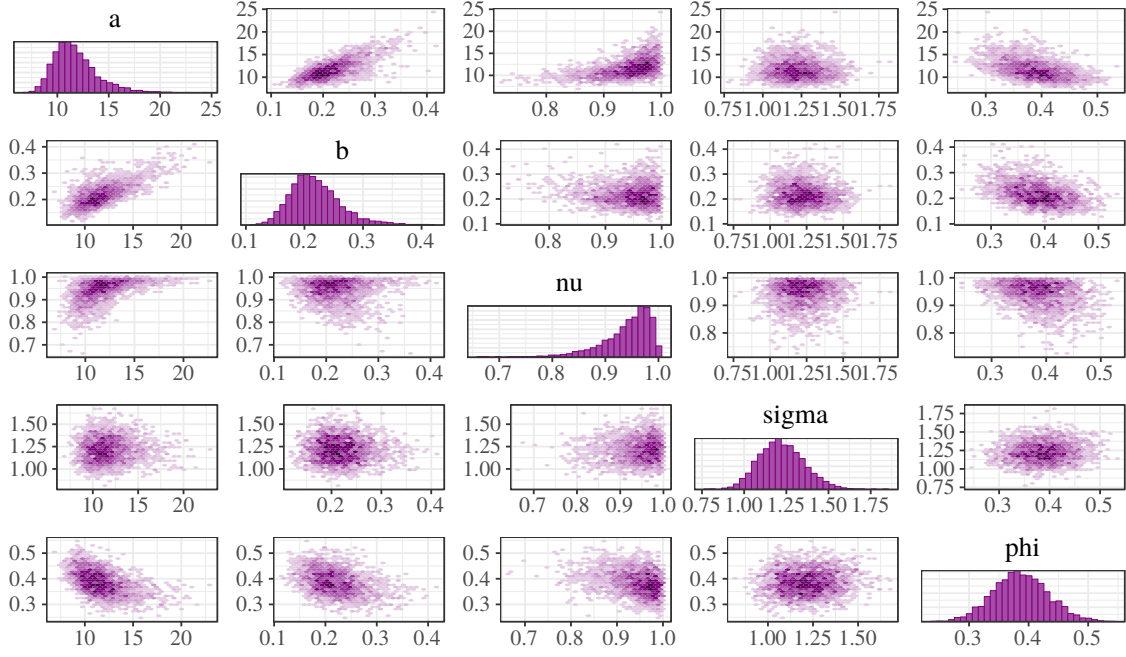


Figure 6.9. Pairwise posterior plot for the full dynamic coarse-grained ODE model (6.4). Marginal distributions for each parameter (a , b , ν , σ , ϕ) are shown on the diagonal. Pairwise plots are shown as hex plots with darker purple indicating higher density. We note the positive correlation between production rate a and transport rate b due to an unidentifiability in the model.

described in Section 6.6.1. We consider the effects of making a small perturbation to the production and transport rates a and b . Since Equation (6.1) is a linear system of ODEs, we can obtain an analytical solution for the evolution of $\mathbf{y}$, as given in Equation (6.6). By differentiating this solution with respect to model parameters a and b , we find an expression for how sensitive the number of RNA-protein complexes is to these parameters.

Using the general solution obtained in Equation (6.6), and differentiating with respect to the production rate, a , we have

$$\begin{aligned} \frac{\partial \mathbf{y}}{\partial a} &= \mathbf{V} \mathbf{D} \frac{\partial \mathbf{c}}{\partial a} + \frac{\partial \mathbf{k}_1}{\partial a} + \frac{\partial \mathbf{k}_2}{\partial a} t, \\ &= (\mathbf{I} - \mathbf{V} \mathbf{D} \mathbf{V}^{-1}) \frac{\partial \mathbf{k}_1}{\partial a} + \frac{\partial \mathbf{k}_2}{\partial a} t. \end{aligned}$$

We obtain a similar expression when differentiating with respect to the transport

rate, b , but without the term giving linear growth in time

$$\frac{\partial \mathbf{y}}{\partial b} = (I - \mathbf{V}\mathbf{D}\mathbf{V}^{-1}) \frac{\partial \mathbf{k}_1}{\partial b}.$$

We show in Figure 6.10 how sensitivity to the rate of production, a , and the rate of transport, b , vary across different parameter values, and highlight relevant biological parameters by the orange circle. At realistic parameter values: $(a, b) = (12.9, 0.24)$, then

$$\frac{\partial \mathbf{y}}{\partial a} \approx \frac{\partial \mathbf{y}}{\partial b}.$$

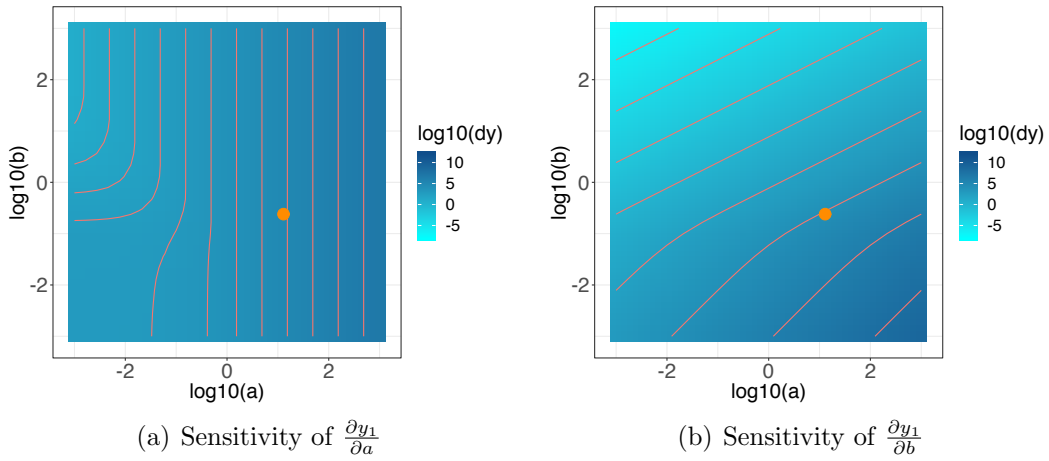


Figure 6.10. Sensitivity analysis for the ODE model given in (6.1). In (a), we show the sensitivity of the number of RNA particles in the oocyte to the rate of production, a , with realistic model parameters highlighted with an orange circle. Similarly in (b), we show sensitivity to the rate of transport, b . Results are presented at time $t = 30$ hrs.

6.7.2 Balance between parameter estimates may be due to unidentifiability

However, we note that the shape of the pairwise posterior distribution in Figure 6.9 between the production rate a and transport rate b suggests that there is an unidentifiability between these parameters given the data available to us since the posterior mass lies on a manifold relating these parameters. Similar results are also

seen when performing inference based on synthetic data simulated from the model. This indicates that this relationship between a and b is a property of the model. The same data can equally be explained by a high production rate combined with a high transport rate, or by a lower production rate with a lower transport rate, provided this is compensated for by the assembly parameter ϕ . Although we can observe that the production and transport rates, a and b , are balanced, we should be cautious in drawing further biological conclusions from the relationship between these parameters, since what appears to be a tight regulatory balance between production and transport could simply represent a consequence of the unidentifiability of these parameters.

6.8 Perturbing mRNA production yields results inconsistent with model predictions

To further validate our results, we explored whether the model can predict the response of the system to a perturbation in the mRNA complex production rate by considering an over-expression mutant with multiple copies of the *gurken* gene. The behaviour of the over-expression mutant phenotype is illustrated in Figure 6.11. We make the assumption that the mutant has mRNA complex production rate γa , where a is the production rate in wild type, and $\gamma > 1$ is a scale factor, but that all other model parameters are unchanged. The model specified in Equation (6.1) predicts that the total number of RNA complexes in an egg chamber should increase at rate $15a$ in wild type, and $15\gamma a$ in the over-expression mutant. Comparing this prediction with experimental data, we estimate a median value for γ as follows.

6.8.1 Estimates of γ

We assume that the over-expression mutant has RNA complex production rate γa , where a is the production rate in wild type, and $\gamma > 1$ is a scale factor. We determine γ by considering the total mRNA over time in all cells in the egg chamber (15 nurse

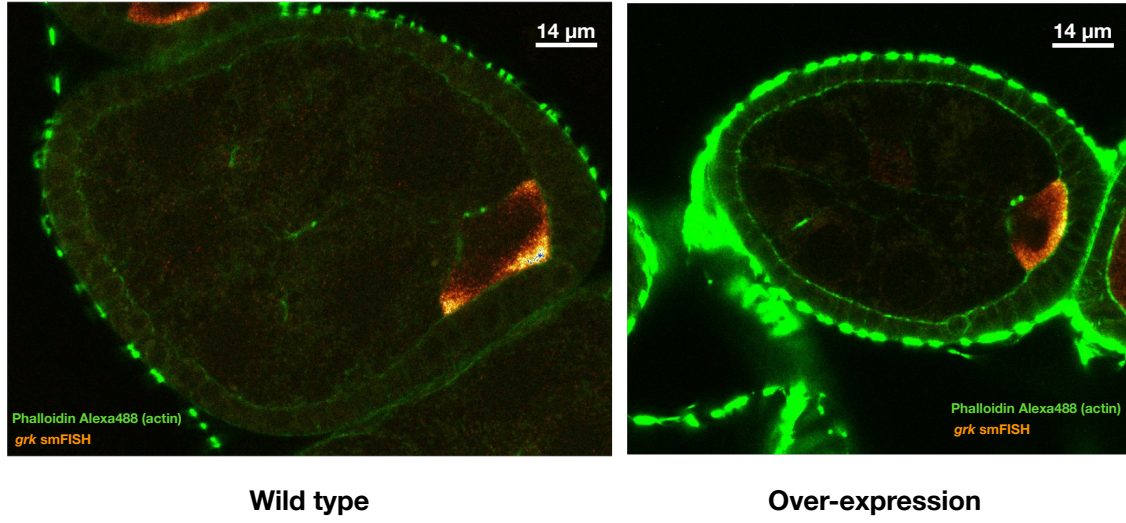


Figure 6.11. Comparison of phenotypes between wild type and over-expression mutant for egg chambers at similar stages of development and under identical imaging conditions. Similar amounts of *gurken* mRNA have been localized in each case. For the over-expression mutant, one of the nurse cells contains much more mRNA than the others, whereas the distribution across nurse cells for the wild type is more equal.

cells and the oocyte). Recall that mRNA is produced in the nurse cells, but not the oocyte which is transcriptionally silent, such that if we define the total mRNA across all cells as:

$$Y(t) = \sum_{i=1}^{16} y_i(t), \quad (6.7)$$

then by summing Equation (6.1), the transport terms cancel and we obtain

$$\frac{dY}{dt} = 15a. \quad (6.8)$$

Suppose that we observe $Z_{WT} \sim \mathcal{N}(Y_{WT}, \sigma_{WT}^2)$ and $Z_{OE} \sim \mathcal{N}(Y_{OE}, \sigma_{OE}^2)$ for the wild type and over-expression mutant, respectively, where we assume that each has production at rate a_{WT} and a_{OE} , with $a_{OE} = \gamma a_{WT}$.

We assume the following prior distributions: Truncated Normal distribution with $\mu = 0$, $\sigma = 10$ and support $(0, \infty)$ for the production rates a_{OE} and a_{WT} ; Half-Cauchy(0, 3) with support on $(0, \infty)$ for observation noise parameters σ_{OE} and

σ_{WT} . Results of performing MCMC for this model are shown in Figure 6.12, giving a median estimate for γ of 2.18, with a 95% credible interval of [1.08, 3.18]. We take the value $\gamma = 2$ for simplicity for the remainder of this chapter.

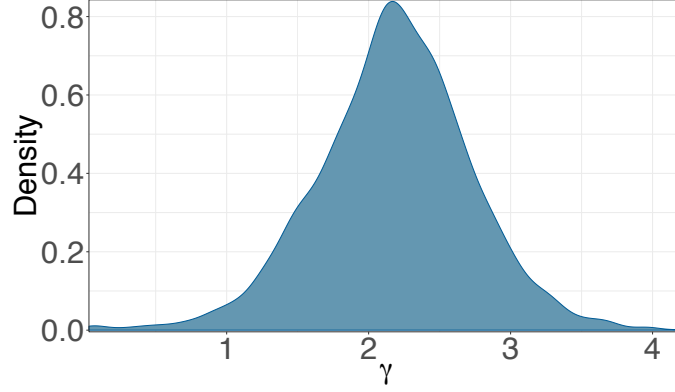


Figure 6.12. Posterior distribution for the ratio, γ , between production in wild type and the over-expression mutant.

6.8.2 Prediction of the distribution of mRNA complexes for an over-expression mutant

We now use the model to predict the distribution of mRNA complexes in each of the cells of the egg chamber by sampling from the posterior predictive distribution generated in Section 6.7 using wild type data only. In doing so, the implicit assumption is that, in the over-expression mutant, the only process significantly affected is the mRNA complex production rate. For the oocyte and neighbouring nurse cells, the model predictions are consistent with the observed data (Figure 6.13). However, for the nurse cells furthest away from the oocyte (such as 8,12,14,15,16 which are at least three ‘steps’ away from the oocyte) the observed experimental data contains nurse cells with much greater accumulation of mRNA complexes than predicted by the model (Figure 6.13). This result clearly indicates that the model does not capture key mechanistic details of the mRNA localization process that are relevant to the over-expression mutant. The aim of the ensuing analysis is to tease apart the nature of these missing mechanistic details.

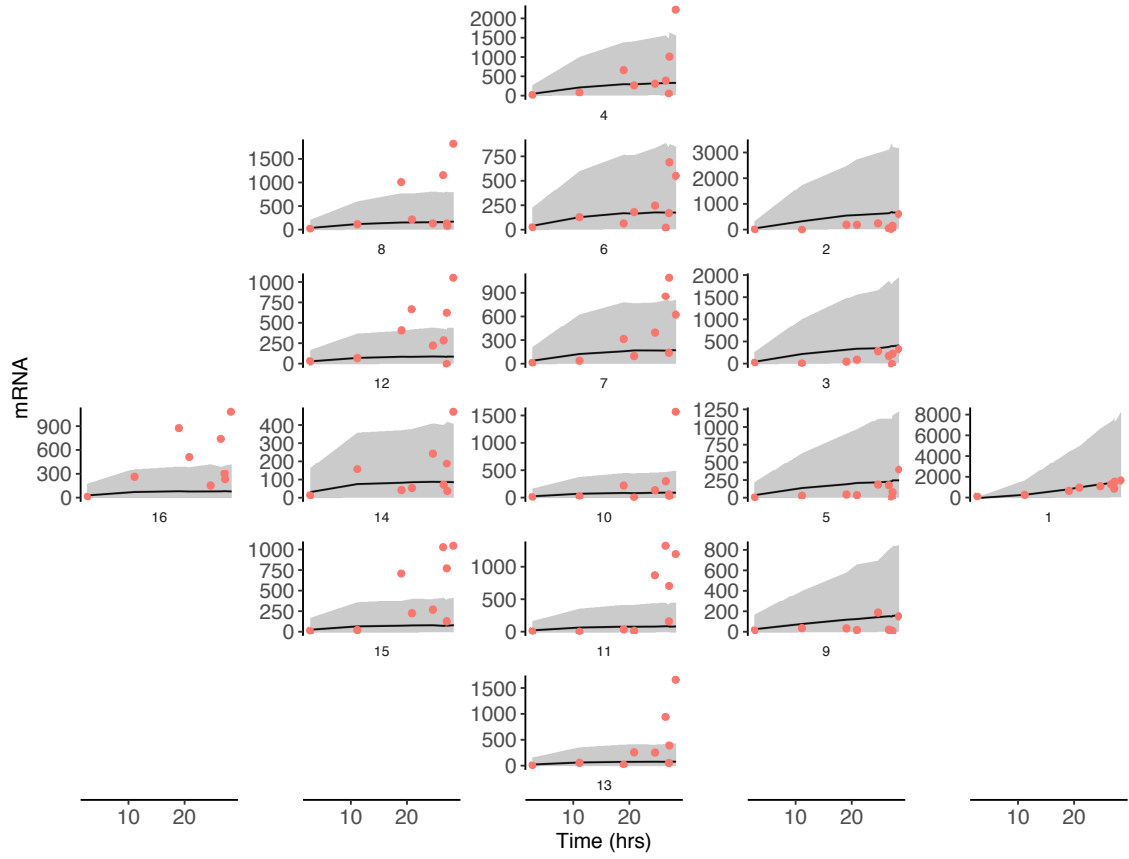


Figure 6.13. Posterior predictive distribution for the over-expression mutant based on the ODE model (6.1) fitted to wild type data. Model predictions (the grey ribbons) show good agreement with observed data (red points) for cells close to the oocyte. Cells further from the oocyte, such as cell 16, show discrepancy between model predictions and the observed data, indicating that there is some biological mechanism not captured by the model. The observed data for the over-expression mutant is shown as red dots, with the grey envelope showing a 95% credible interval of predictions from the model via the posterior predictive distribution. The subplots are arranged to highlight distance from the oocyte within the network of connections between nurse cells shown in Figure 3.3c).

6.9 Localization of mRNA in the oocyte of the over-expression mutant reveals robustness

Quasi-steady-state analysis of the ODE model in Equation (6.1) predicts that mRNA complex numbers in the oocyte, y_1 , grow linearly at a rate proportional to $15a$ (Section 6.6.1). Therefore, in addition, the model predicts that in the over-expression mutant, which has an increased production rate of γa , the number of mRNA complexes localized to the oocyte at a given developmental time should increase by a factor of γ compared to wild type. However, rather than this predicted increase, we observe experimentally the number of mRNA complexes in the over-expression mutant oocytes to be similar to that in wild type (Figure 6.14). This results demonstrates an incredible robustness of the system in the sense that the same amount of *gurken* mRNA is localized to the oocyte, despite large changes to the production rate. A key question is now what mechanism gives rise to the observed robustness?

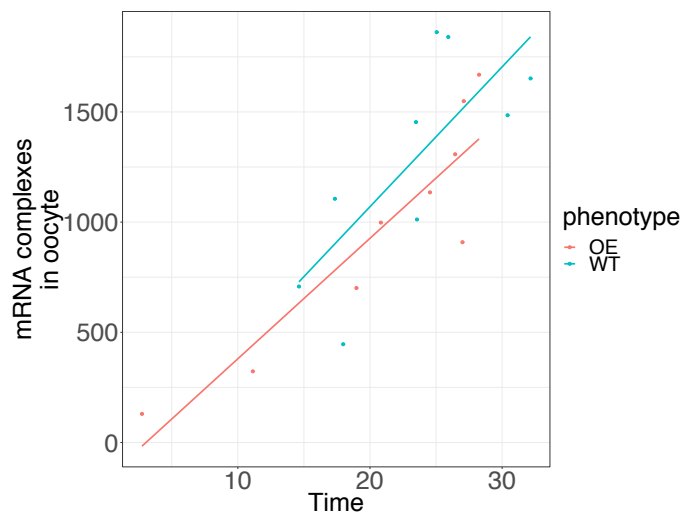


Figure 6.14. The number of mRNA complexes in the oocyte increases linearly over time for both the wild type and the over-expression mutant. Strikingly, the rate of increase of mRNA in the oocyte (shown by the gradient of the red and blue lines) is approximately equal in each case, whereas we would have expected this to be $\gamma \approx 2$ times greater for the over-expression mutant. The tight regulation of this process biologically ensures the right amount of RNA reaches the oocyte at the right time, even under biological perturbation. The points show observed data on number of mRNA complexes, and the lines show linear models fit to these data.

6.10 Biological mechanisms to explain the over-expression data can be represented as alternative models

We propose several plausible biological mechanisms to explain the behaviour of the localization process in the over-expression mutant, as listed below.

Inhomogeneous production. Previously we assumed equal production rates of mRNA complexes across all nurse cells. However, the over-expression of *gurken* in nurse cells is driven by the GAL4-UAS system, which can result in patchy expression across cells. We now relax the assumption of equal production rates, and estimate the production rates of individual nurse cells by quantifying the number of nascent *gurken* transcripts (see Section C.2 for details on the methods used to obtain these estimates). The measurements of nascent transcription in units of arbitrary fluorescence are shown in Figure 6.15. The resulting estimates for transcription rates

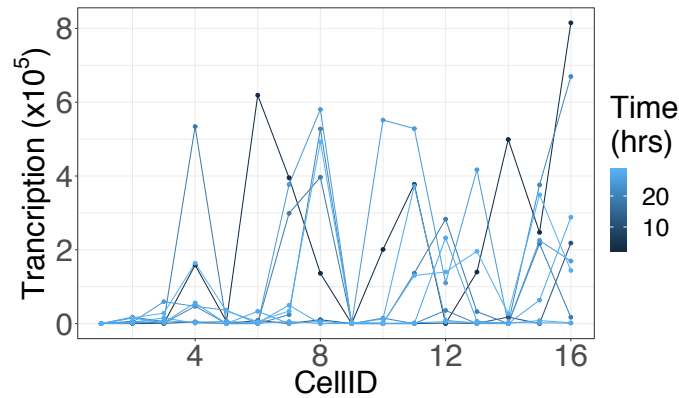


Figure 6.15. Nascent transcription measured in the nuclei of nurse cells from *Drosophila* egg chambers when a *gurken* over-expression phenotype. Each line corresponds to a different egg chamber, coloured to indicate the age of the egg chamber in hours. The index for each cell shown on the x axis corresponds to the numbering of the graph in Figure 3.3. The transcription shown on the y axis is in units of fluorescence.

in each cell are shown Figure 6.16, assuming that each cell has a characteristic individual transcription rate which is shared across egg chambers.

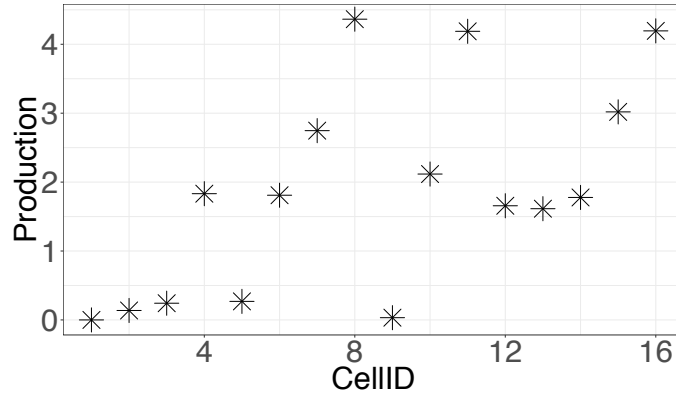


Figure 6.16. Estimated inhomogeneous production values across cells based on measurements of nascent transcription and a mixture model. Points shown are median estimates for the production in each cell. Posterior distributions are tight around these median values. Note that the normalized production values are shown such that an homogeneous model for the over-expressor would have production of $\gamma = 2$ across all nurse cells, and the wild type would have production of 1 in each nurse cell.

Blocking or queuing at ring canals. With increased levels of *gurken* in the over-expression mutant, the environment inside the nurse cells is more crowded. We hypothesize that the transport of mRNA complexes through ring canals could be blocked or restricted due to crowding. We can represent this hypothesis in the model by changing appropriate entries in the matrix $\mathbf{B}$ to zero: manual examination of the data was used to predict which ring canals are blocked.

Density dependent transport. The final hypothesis we consider is that increases in *gurken* in the over-expression mutant lead to saturation of the transport mechanism. This hypothesis is motivated by observations of a build up of mRNA at nuclear pore complexes in the over-expression mutant, suggesting that nuclear export or the ability to form competent mRNA particles for transport may be saturated. Instead of assuming that transport is linearly proportional to the number of mRNA complexes, we assume a saturating transport term of the form

$$f(y) = \frac{y}{1 + \beta y},$$

where β is a parameter describing the density dependence.

A collection of models. We represent the hypotheses detailed above as a collection of models, $\mathcal{M} = \{M0, M1, \dots, M7\}$, that together represent all possible combinations of the additional mechanisms (see Table 6.1 for details). Models including the crowding-induced blocking at ring canals and inhomogeneous production mechanisms can be forward simulated using parameters from the posterior distribution based on wild type data, as in Figure 6.13. Models including density dependent transport must be fitted to the wild type data to estimate the parameter β .

6.11 Model comparison

We perform model comparison via leave-one-out cross validation [127, 149, 150]. From the family of models under consideration, we make testable predictions for the behaviour of a genetic mutant and compare these with experimental data. We compute pseudo-BMA+ (Bayesian Model Averaging) weights [148] for each model, where the largest weight will give the model closest in KL divergence to the data generating model. However, these weights can potentially be misleading in the case where the true model is not contained in the set of models considered, as is the case here. In this $\mathcal{M}$ -open case, stacking weights [148] offer a more conservative choice of models. We present pseudo-BMA+ and stacking weights, since both are informative. For further details on these model comparison methods, see Sections 2.3.6 and 2.3.8.

6.12 Model comparison reveals blocking of ring canals as the most plausible mechanism

We use model comparison approaches (Section 6.11) to evaluate the ability of the different models to explain the biological observations. Table 6.1 indicates strong support for the blocking of ring canals as a mechanism to explain the observed over-

	Production	Blocking	Density dependent transport	Pseudo BMA+ weights	Stacking weights
M0	0	0	0	0.04	0.00
M1	0	1	0	0.94	0.40
M2	0	0	1	0.00	0.08
M3	1	0	0	0.01	0.23
M4	1	0	1	0.00	0.00
M5	1	1	0	0.00	0.29
M6	0	1	1	0.00	0.00
M7	1	1	1	0.00	0.00

Table 6.1. A description of the collection of models $\mathcal{M} = \{M0, M1, \dots, M7\}$, together with the weights generated using model comparison (see Section 6.11 for details).

expression data: Model M1 has a pseudo-BMA+ weight close to 1, indicating that of the models considered this is the most plausible model. The more conservative stacking weights approach also weights several other models highly, however these are predominantly models that involve blocking as well as additional mechanisms. These results, shown also in Figure 6.17, suggest that none of the models considered perfectly describe the data generating process (to be expected since the model is an incredibly simplified description of a very complex biological process). However, predictions of mRNA complex distributions for the over-expression mutant made using M1 (which incorporates only the blocking of ring canals) are substantially improved compared to the predictions of the basic model. Taken together, our results reveal crowding-driven blocking of ring canals as the most plausible explanation for our experimental observations.

6.13 Reexamining microscopy data provides evidence of queuing at ring canals

A natural question to ask is then whether it is possible to validate this prediction experimentally? By re-examining microscopy data from the over-expression mutant, we demonstrate evidence for the crowding-related blocking of ring canals in smFISH

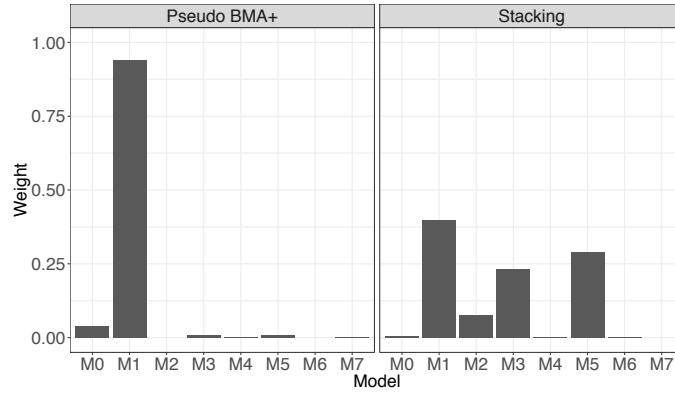


Figure 6.17. Model comparison based on predictions for the over-expression mutant provides weights for each of the models, which represent combinations the proposed biological mechanisms: inhomogeneous production, blocking of ring canals and density dependent transport. Two different methods are used to compute weights for the models. The Pseudo BMA+ weights assume the true model comes from the collection of models considered ($\mathcal{M}$ -closed). The stacking weights assume that the true model is not contained among the models considered ($\mathcal{M}$ -open), giving a more conservative weighting of models. The weights computed support blocking of ring canals (M1) as the best explanation for the observed over-expression data.

data: we observe clusters of mRNA complexes around ring canals (Figure 6.18). Similar observations of a build up of mRNA complexes near ring canals have also been reported upon injection of *gurken* mRNA in nurse cells [182]. Together, these results suggest that, while the transport of mRNA complexes within nurse cells may be relatively fast, the efficiency of the transport process as a whole can suffer when transit through ring canals is blocked by crowding.

6.14 Model generalizations support the same conclusions about the blocking mechanism

We consider generalizations of the simple model (Equation (6.1)) to incorporate inhomogeneous variance across cells in our observations or a decay term with decay of RNA at rate δ . Our conclusions about the blocking mechanism hold also for these more general descriptions. The simple model is favoured by model comparison when

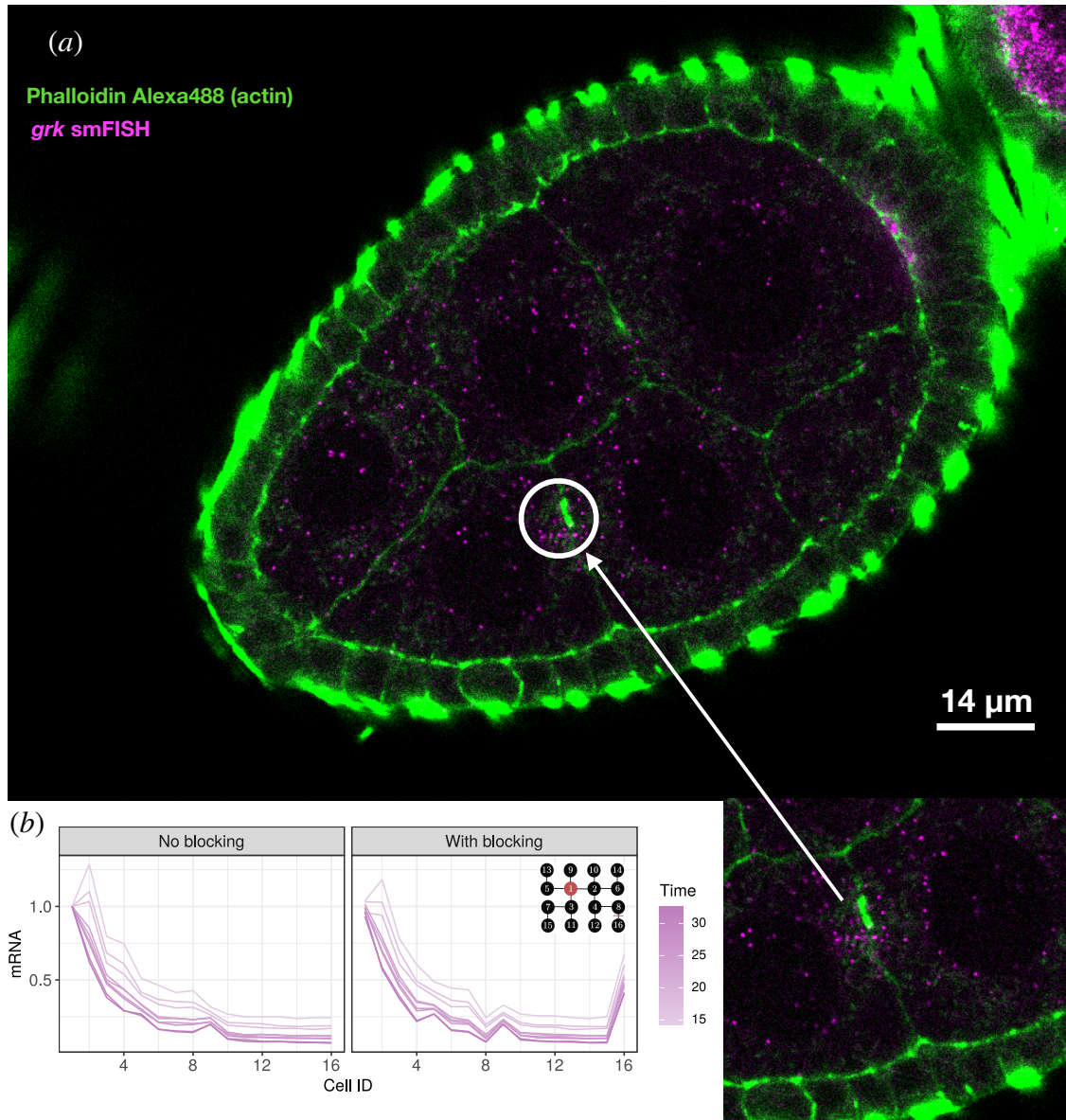


Figure 6.18. Crowding-induced blocking of ring canals results in significant accumulation of mRNA in nurse cells far from the oocyte. a) A clustered distribution of mRNA complexes queuing near the ring canal (arrow). b) A comparison between simulations from the model with all ring canals transporting RNA equally, and a situation where the ring canal between cells 8 and 16 is blocked. The schematic inset in b) shows the position of the blocked ring canal within the tissue.

compared to these models, which do not generalize as well to predictions for the over-expression data.

6.14.1 Extension to model with mRNA decay

We consider a generalization of the model described in Equation (6.1) and Section 6.4 that accounts for degradation of mRNA. We assume that mRNA degrades at rate δ in each of the cells. If there is no degradation or decay of mRNA this will be captured by this model with a very low decay rate, δ . Including decay in the model fundamentally changes the dynamics of the system of equations. In the quasi-steady-state, the mRNA levels in the nurse cells will be constant when decay is included whereas without decay mRNA levels in the nurse cells increase linearly over time.

The resulting system of equations with degradation of mRNA at rate δ is as follows:

$$\frac{d\mathbf{y}}{dt} = a\mathbf{v} + b\mathbf{B}(\nu)\mathbf{y} - \delta\mathbf{y}. \quad (6.9)$$

Setting the same prior for the decay rate, δ , as for the other rate parameters of $\mathcal{N}(0, 10^2)$ in units of $[\text{hr}^{-1}]$, and otherwise using the same priors described in Section 6.4.1, we can perform inference for the model with degradation given by Equation (6.9). We obtain a 95% credible interval for δ of $[0.01, 0.33]$ with a median of 0.14. Samples from the posterior predictive distribution for this model are shown in Figure 6.19. Conclusions drawn from our analyses are unaffected by using the model with decay.

6.14.2 Analysis of inhomogeneous variance model

Previously we have assumed a homogeneous noise model across cells in the egg chamber, in that the overdispersion parameter, σ , in the negative binomial measurement model in Equation 6.3 is the same for every cell. The observed data, shown as dots in Figures 6.8 and 6.13, seems to exhibit greater variation in certain cells than others. We attempt to account for this variation by relaxing the assumption that the

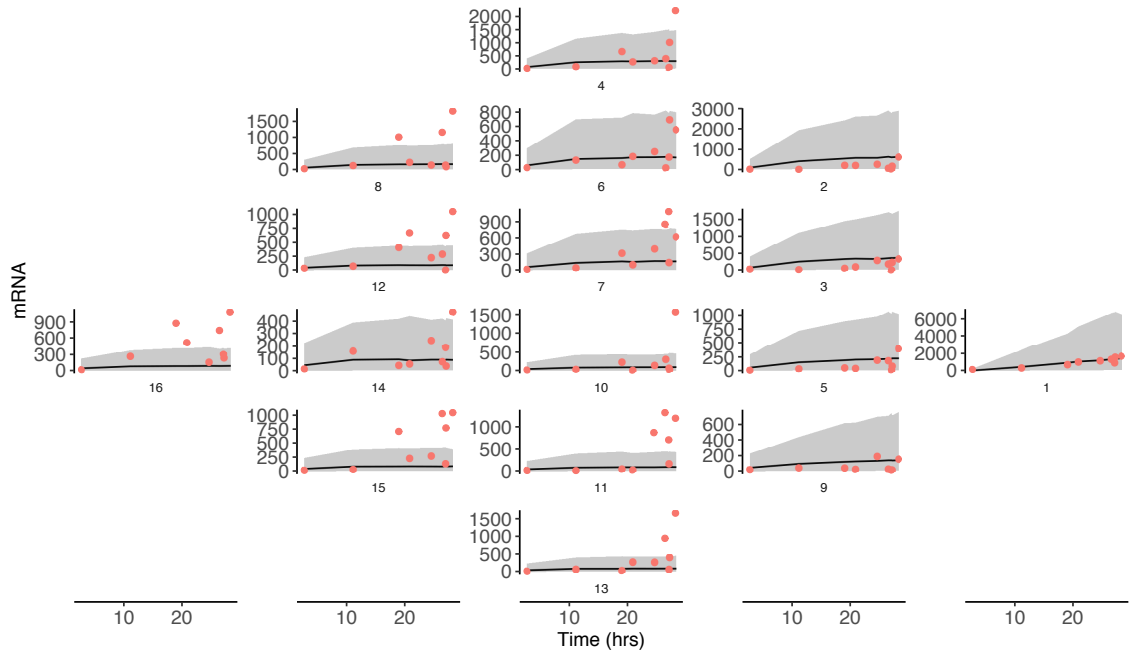


Figure 6.19. Posterior predictive distribution for the over-expression mutant using the model with mRNA decay at rate δ . The grey envelope shows a 95% credible interval for predictions, while the red points correspond to observed data.

overdispersion parameter is the same for all cells and instead letting it vary across the tissue. Predictions for the over-expression mutant from the simple model, M0, but with inhomogeneous overdispersion, σ , are shown in Figure 6.20. The biggest difference is a smaller range of predictions for the oocyte. This is consistent with the robustness of the amount of mRNA localized in the oocyte. The disparity between the prediction for the over-expression mutant and the observed levels of mRNA localization in the oocyte is clearly apparent from these results also.

Further to this we considered a comparison of the collection of models $\mathcal{M} = \{M0, M1, \dots, M7\}$ between the homogeneous and inhomogeneous overdispersion cases. We computed weights jointly for this collection of 16 models together. The results are shown in Figure 6.21 and demonstrate that in general the homogeneous overdispersion parameter models should be favoured over the inhomogeneous overdispersion parameter models. This justifies using the model with homogeneous overdispersion across cells. Moreover, the conclusion that crowding-induced block-

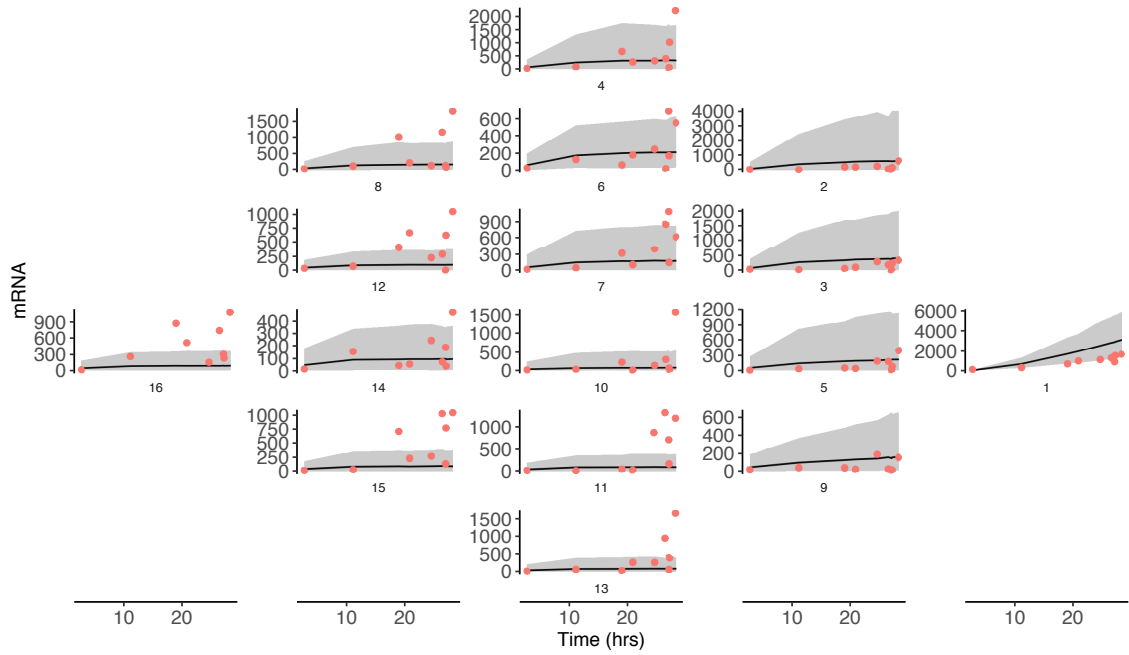


Figure 6.20. Posterior predictive distribution for the over-expression mutant using the overdispersion model, where a separate overdispersion parameter has been fitted in the measurement model for each cell in the egg chamber. The grey envelope shows a 95% credible interval for predictions, while the red points correspond to observed data.

ing of ring canals is the most plausible model to explain the observed data for the over-expression mutant still holds when jointly comparing this larger collection of models. We note that the inhomogeneous overdispersion model is a more flexible class of model, but that this overfits to the wild type data and is less able to make predictions relevant for the over-expression mutant. This could indicate further biological differences between the wild type and over-expression mutant, which could be investigated in future studies.

6.15 Discussion

Combining very simple mechanistic mathematical models and Bayesian inference approaches together with experimental evidence, we have investigated the dynamics of mRNA localization in *Drosophila* egg chambers, and made quantitative pre-

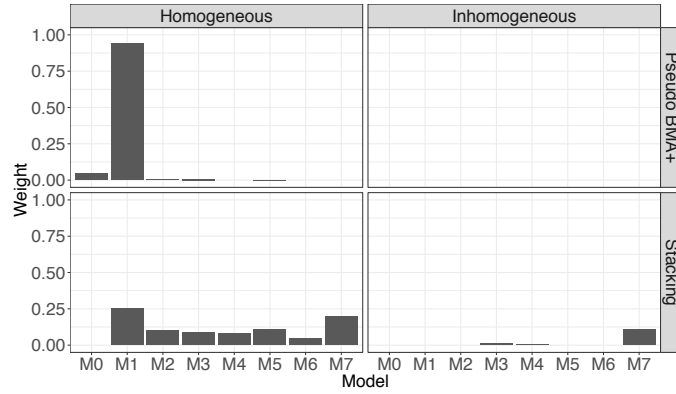


Figure 6.21. Model comparison weights between homogeneous and inhomogeneous overdispersion models based on a joint comparison of 16 models. Weights are computed via pseudo-BMA+ and stacking methods.

dictions of the dynamics of the localization process. Estimates of the production and transport rates of *gurken* mRNA indicate that these processes are balanced in mRNA localization. To confer robustness on the system, large perturbations must be buffered by the localization mechanism to ensure that the ‘correct’ amount of RNA is localized at the right time in development. We have shown that crowding-driven blocking of ring canals is sufficient to regulate the localization of mRNA in the oocyte.

In addition, our results suggest that transport of mRNA complexes through ring canals is strongly biased. The characteristic network of connections between the nurse cells specifies a given network structure that we represent via the matrix $\mathbf{B}$ in Equation (6.1). By inferring the bias parameter, ν , we have shown that the observed distribution of mRNA complexes across the egg chamber is consistent with strongly biased transport of mRNA complexes towards the oocyte. Biologically, this biased transport is due to active transport of mRNA complexes on microtubules, and reflects bias in the orientation of the microtubule network within a cell, although ν is not a direct quantification of the bias in the microtubule network.

We have also provided measurements and estimates for the assembly of higher order mRNA complexes in the oocyte following remodelling upon transit through the

ring canals between the oocyte and the nurse cells. Measurements of RNA content for *bicoid*, *oskar* and *nanos* mRNA complexes have been performed previously [229, 274], although largely for later stages of development, and we note that the estimates given here are consistent with similar measurements on the distribution of multimeric mRNA complexes of *oskar* upon entry to the oocyte at stage 10b [274]. This process of remodelling of mRNA complexes upon transit into the oocyte may play a role within wider changes in how RNA is packaged in the oocyte compared to the nurse cells.

We attempted to validate the mathematical model (Equation (6.1)) by making predictions about behaviour of an over-expression mutant. The predictions of this simple model revealed a discrepancy in the accumulation of mRNA in the nurse cells furthest from the oocyte compared with the observed experimental data: in some cases, far more mRNA is observed in these cells than predicted by the model. In addition, the observed rate of localization of mRNA in the oocyte is similar in the over-expression mutant as compared to wild type, contrary to predictions from the simple model. We resolved this discrepancy by using model comparison approaches to compare the ability of a range of biological mechanisms to replicate the observed data. Our results suggest crowding-induced blocking of ring canals as a plausible explanation for the discrepancy between model prediction and experimental observations, and more generally that the transport process becomes overloaded in the over-expression mutant. We therefore argue that the crowding-induced blocking of ring canals is effectively a regulatory mechanism that helps ensure the robustness of *gurken* localization in *Drosophila* oocytes. Precise control of gene expression during development can be facilitated by indirect methods [287]; blocking of ring canals is an example of such indirect regulation.

Finally, we highlight that the conclusions drawn as a result of this combined modelling-experiment study were made possible through use of Bayesian inference approaches that allowed us to interrogate and interpret quantitative data, making

full use of the information contained therein. It is important to note also that our conclusions were reached through the use of an incredibly simple, analytically tractable, coarse-grained model, that allowed us to abstract many of the intricate details, and focus on salient mechanisms. The fine-grained stochastic model of Chapter 3 provided a detailed description of the mechanisms of mRNA transport within a cell; the simple representation presented in this chapter relates to the same underlying process on different spatial and temporal scales. Previously in Chapter 3, we provided evidence that transport within a single cell is not rate-limiting for the mRNA localization process, whereas here we suggest that production and transport are in tightly regulated balance. These results are nonetheless consistent, since the transport process considered in this chapter includes also transit through ring canals between cells in addition to transport within a single cell.

Chapter 7

Discussion

Quantitative approaches to interrogating biological hypotheses with data are crucial in utilising complex data sources such as imaging data. In this work, we have developed and applied Bayesian inference methods to aid in understanding processes in developmental and cell biology. By using simple models together with experimental data, we have gained insight into regulation of mRNA localization. In this chapter, we review the progress made in this thesis and set out areas for further work in the future.

7.1 Discussion

The contribution of this thesis has been in several parts, spanning methodological improvements to approximate Bayesian inference algorithms (Chapter 5, novel application of existing modelling methods to a new biological application area to answer biological questions (Chapters 3 and 6), computational investigation into the effect of experimental design on inference (Chapter 4), and integrated modelling and Bayesian inference to shine light on the regulation of mRNA localization (Chapter 6). We hope that this portfolio of work can help guide future studies in similar areas.

7.1.1 Chapter 3

In particular, we developed a microscopic, stochastic model in Chapter 3 for the movement of individual mRNA complexes within a nurse cell. This model allowed us to study the timescale for transport from the nucleus of the nurse cells to ring canals with other cells. By fitting this model to experimental data via ABC, we have made an experimentally testable prediction regarding the behaviour of a half-speed Dynein mutant.

Localization of mRNA is crucial in controlling the expression of proteins in space and time. In this work, we have addressed questions about the regulatory mechanisms and bottlenecks governing mRNA localization. By using a microscopic

stochastic model of mRNA transport within a single nurse cell, we have demonstrated that the timescale of transport of mRNA complexes within a single cell is not a bottleneck for localization of mRNA, and this timescale is strongly dependent on the structure and dynamics of the cytoskeleton.

7.1.2 Chapter 4

Motivated by this work in Chapter 4, we have examined the effects of experimental design parameters such as the temporal sampling frequency of collecting imaging data, or the observation noise, on parameter estimation within a Bayesian framework. When using certain models of biological transport, such as the VJP model, and linking this to data observed at discrete times, estimates of parameters can be biased. Although we must still make approximations to formulate the VJP model within a hidden states framework to enable us to use particle MCMC methods for inference, this approximation can be understood and controlled.

7.1.3 Chapter 5

Although one of the appealing aspects of the ABC algorithm is its simplicity in how easy it is to understand, computational performance of the method can depend highly on its implementation. In Chapter 5, we focused on the role of summary statistics in ABC and how to combine them within the ABC distance function to make best use of the information contained therein. We developed an algorithm that chooses weights for a collection of summary statistics adaptively by maximizing the distance between the prior and approximate posterior distribution. We analyzed this method thoroughly both from a theoretical point of view and by applying it to a suite of test problems. To enable a wider audience of computational scientists to make use of inference methods such as ABC, it may be beneficial to have robust, automated methods of selecting key implementation parameters such as summary statistics and tolerance schedules for ABC-SMC. Although the adaptive method

presented in Algorithm 5 does not remove completely the need for a collection of summary statistics, it can help avoid the need for hand-crafted summary statistics. The concept of an automated statistician [288] may be one way of incorporating quantitative approaches in science by removing the burden of choice of implementation parameters or even choice of algorithm from the practitioner.

7.1.4 Chapter 6

Finally, we have investigated the nature of robustness during development, emphasizing the role of robustness during localization of mRNA in *Drosophila* oogenesis. We developed a coarse-grained differential equation model, and related this model to smFISH data via Bayesian inference methods. By making testable predictions from our model, we showed that our initial understanding of the dynamics was not sufficient to explain the data observed for an over-expression mutant. Model comparison between competing hypotheses, which were represented quantitatively as alternative models, uncovered a crowding-induced blocking mechanism at ring canals as the most plausible hypothesis to explain the observed behaviour. This framework for thorough model-based analysis, through representing competing hypotheses as models and performing comparison between them, is a promising approach to answering biological questions in a quantitative manner.

By comparing predictions from a simple differential equation model of mRNA transport and localization with experimental data from an over-expression mutant, we revealed robustness in how much mRNA is localized at different stages of development. We highlighted crowding-induced bottlenecks in transport through ring canals between nurse cells as a possible regulatory mechanism involved in maintaining this robustness.

7.2 Future work

7.2.1 Further work to extend inference methods

Biological phenomena generally involve processes taking place on multiple spatial and temporal scales. It can therefore be informative to build models that focus on events taking place on each of these scales and combine the models into a hierarchy in which detail from a finer scale model can be represented in a simplified way in a model on a more coarse scale, as in Chapters 3 and 6. We currently lack methodologies to conduct parameter inference within such hierarchies of models. The key difficulty lies in uniting the model descriptions within a single framework in a general way to allow for parameter inference informed jointly by behaviour across scales. Sequential applications of Bayesian analyses [226] can fulfil this role if the models are nested appropriately. Similarly Multifidelity Monte Carlo methods [114] go some way to addressing this by using cheaper or surrogate models to assist inference with a more expensive model of interest. Tools to enable parameter estimation within a hierarchical, multiscale modelling framework integrating models and data across scales could be developed in future work.

7.2.2 Further work to test current biological hypotheses

As a result of developing the microscopic, stochastic model of mRNA transport within a single nurse cell in Chapter 3, we have made testable predictions about the behaviour of a Dynein mutant and the effect of this on FPTs, which described the timescale for transport within a nurse cell. These predictions should be investigated experimentally to assess and validate the predictions of the model.

In Chapter 3, we investigated different summary statistics, including population based summary statistics, and combined summary statistics using both microscopic and macroscopic data. These proved to be effective in simulations on synthetic data. For further investigation of models of this type, we recommend performing injection or photo-activation experiments to make use of these summary statistics.

In photo-activation experiments, a region of interest is switched from green to red by excitation with radiation at a particular wavelength. Thus far, such experiments have been challenging due to the sensitivity of the photo-activatable protein to further light used during imaging. In further imaging the sample to observe changes after a period of time, much of the remaining sample will also switch, making interpretation of results challenging. As improved photo-activation technologies become available, experiments such as this may become more feasible.

In Chapter 6, we proposed a crowding-induced blocking mechanism to explain accumulation of mRNA in nurse cells for an over-expression mutant. Although other regulatory mechanisms such as saturation of rate-limiting transit through ring canals could account for such a build up of mRNA, by performing model comparison we showed that the observed data favour the blocking model. To further assess the crowding-induced blocking mechanism, we suggest directly measuring transit through ring canals in the wild type and over-expression mutant conditions. This is nontrivial and would require extended time lapse imaging to track movements of complexes through representative subsets of all ring canals.

7.2.3 Further work to extend modelling of mRNA localization

The models presented in Chapters 3 and 6 offer different perspectives on parts of the same biological process. From a modelling point of view, it would be desirable to unite these into a single model of the entire *Drosophila* egg chamber capturing the microscopic detail of mRNA transport within and between cells, up to localization. This detailed model would allow different regulatory mechanisms to be explored in ways that cannot be captured by the models outlined in this work. Before starting with this ambitious project to model many cells with microscopic detail, it is important to build, test and validate simpler models of transport within a single cell, and at coarser scales in space and time, and we have done so in this thesis.

One detail of the geometry used by the fine-grained microscopic model described in Chapter 3 that could be further investigated is the positioning of the ring canal and the number of ring canals. In this work, we assumed that the nurse cell of

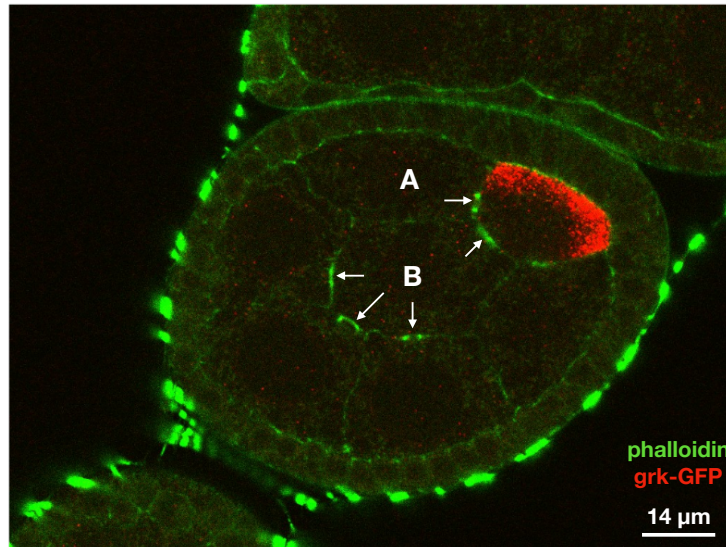


Figure 7.1. Stage 6 *Drosophila* egg chamber showing the positioning of ring canals in two different nurse cells. The nurse cell labelled A has only a single ring canal connecting it to the oocyte and none of the other nurse cells. In contrast, the nurse cell labelled B has connections to the oocyte and to 3 other nurse cells. The ring canals are indicated by white arrowheads.

interest contained only a single ring canal connecting it to other cells, as in cell A in Figure 7.1. It would be interesting to assess the effects of several ring canals within a modelling framework on the transport dynamics, for example cell B in Figure 7.1 has four ring canals. Due to the graph structure of ring canals connecting different nurse cells, certain nurse cells will have stereotypic numbers of ring canals connecting them to other cells. Since the cytoskeleton is connected directly to the ring canal, the presence of a ring canal does affect the cytoskeleton locally and further experimental studies of the cytoskeleton structure in nurse cells may be required before modelling assumptions can be made regarding multiple ring canals.

7.2.4 Further work on software and reproducibility

Modern computational science works most effectively when analyses can be reproduced easily by other researchers, but there are currently problems in this area in terms of the availability of raw data and computer codes [289, 290], stemming from a lack of appropriate training and incentives for researchers.

To help replicate the work presented in this thesis, we have provided code repositories hosted on Github with key parts of code used. Code for Chapter 3 is available at <https://github.com/shug3502/mRNA> Code for Chapter 4 is available at https://github.com/shug3502/pmmc_inference_for_vjps Code for Chapter 5 is available at <https://github.com/shug3502/CombiningSumStats> Code for Chapter 6 is available at https://github.com/shug3502/rstan_analysis. This code is intended to help other researchers interested in replicating or extending the research presented here, rather than providing a fully reproducible analysis.

7.2.5 Details on computational methods used

To aid with further studies, we outline here the languages and computational tools used in each chapter of this thesis and describe the computational running time in each case. For Chapter 3, code to simulate the fine-grained model and perform inference with ABC is written in MATLAB. The inference with ABC takes hours to around a day in duration. For Chapter 4, code is primarily written in R with parts of the bootstrap particle filter written in C++ to accelerate inference with the particle MCMC algorithm. To create figures such as Figure 4.10, MCMC chains were run in parallel for different values of Δt using a python script. The computational cost is dominated by the smallest Δt values, which take up to around 3 days to run. The ABC analysis in this chapter takes hours to around a day to complete and a computational comparison between the ABC and particle MCMC methods is presented in Section 4.3.7. For Chapter 5, code to implement the adaptive algorithm to combine summary statistics is written in MATLAB, and also to implement the

test problems described in this chapter. In comparing with other work, we use our own MATLAB implementation of the algorithm of Barnes et al. [36], and use the R package `abctools` [272] for the method of Fearnhead and Prangle [37]. For Chapter 6, code to analyse the data is written in R with the model implemented via Stan. The Bayesian inference takes minutes to hours to run. Image analysis involved spot detection using a MATLAB package (FISH-quant [207]) accessible via a GUI, and manual segmentation of cells through a custom python GUI [284].

7.2.6 Conclusions

Many of the most ambitious problems currently being tackled in biology and medicine require quantitative data and quantitative models to aid the analysis of these data, and interrogate biological hypotheses. Complex models become most useful when we have a good understanding of their parameters [9], but this can present numerous challenges. In this work, we have addressed some of these challenges and have applied mathematical modelling and Bayesian inference methods in the context of mRNA localization and transport. We have used the predict-test-refine loop to connect computational models with the underlying mRNA localization process via testable experimental predictions. We hope that the work contained in this thesis provides a foundation and an exemplar for future studies in this area.

Appendix A

Supplementary Material for Chapter 3

A.1 Experimental methods and materials

A.1.1 Fly strains and tissue preparation

Stocks were raised on standard cornmeal-agar medium at 25°C. The wild type was Oregon R (OrR). Over expression of *gurken* RNA was obtained by crossing *UASp grk3A* (based upon genomic sequence DS02110, which includes the full 3' and 5' UTRs [291]) to maternal tubulin driver *TubulinGal4-VP16*. Tagging of *gurken* RNA for live imaging was achieved using the MS2-MCP(FP) system: *gurken*-(MS2)¹² [205], P[*nos-NLS-MCP-mCherry*] [191]; P[*w+nanos-5'-NLS-MCP-Dendra2 K10-3'*] (Rippe Hayashi). GFP protein trap lines: *Moesin::GFP* [292]; *Tau-GFP 65/167* (D. St Johnston, University of Cambridge). Oocytes were isolated for live imaging as described in Parton et al. [293].

A.1.2 Live-Cell imaging

Live imaging of isolated *Drosophila* oocytes was done at 21°C with a GE Healthcare DeltaVision CORE integrated wide field imaging system (Lumencor 7-line illumination source upgrade) using an Olympus 100x SI 1.35 NA lens, standard filter sets and a Cascade-II EMCCD camera.

Appendix B

Supplementary Material for Chapter 5

B.1 Theoretical justification

We will require the following results:

Lemma 3. *Póczos and Schneider [267]*

Suppose that $k \geq 2$ and that $\mathcal{M} = \text{supp}(p)$. Let $\gamma = 1 - \alpha$. Assume that (a) q is bounded above, (b) p is bounded away from zero, (c) p is uniformly Lebesgue approximable¹, (d) $\exists \delta_0$ such that $\forall \delta \in (0, \delta_0)$ $\int_{\mathcal{M}} H(x, p, \delta, 1/2) p(x) dx < \infty$, (e) $\int_{\mathcal{M}} \|x - y\|^\gamma p(y) dy < \infty$ for almost all $x \in \mathcal{M}$, $\int \int_{\mathcal{M}^2} \|x - y\|^\gamma p(y) p(x) dy dx < \infty$, where $H(x, p, \delta, \psi) =$

$$\sum_{j=0}^{k-1} \left(\frac{1}{j!}\right)^\psi \Gamma\left(\frac{1}{2} + j\psi\right) \left(\frac{p(x)+\delta}{p(x)-\delta}\right)^{j\psi} (p(x) - \delta)^{-(\frac{1}{2})} ((1 - \delta)\psi)^{-(\frac{1}{2})-j\psi}.$$

Then

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\left(\hat{D}_h(X_{1:N} || Y_{1:N}) - D_h(p || q) \right)^2 \right] = 0. \quad (\text{B.1})$$

The proof of this results relies on constructing an integrable function as a bound such that Lebesgue's dominated convergence theorem can be applied. See Póczos and Schneider [267] for details. Lemma 3 specifies L_2 consistency of the nearest neighbour estimator, which ensures that the estimates of the distance between p and q become more concentrated around the true values as more samples are used.

Lemma 4. *Let X_n and X be random variables in $\mathbb{R}^d$.*

If

$$\lim_{n \rightarrow \infty} \mathbb{E} [||X_n - X||^2] = 0,$$

then, for any $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P} (||X_n - X|| > \epsilon) = 0.$$

That is, L_2 convergence implies convergence in probability.

Proof. By Chebyshev's inequality, for any $\epsilon > 0$

$$\mathbb{P} (||X_n - X|| > \epsilon) \leq \frac{\mathbb{E} [||X_n - X||^2]}{\epsilon^2}.$$

1

Definition (Uniformly Lebesgue approximable function). Let $g \in L_1(E)$ for $E \subset \mathbb{R}^d$. g is uniformly Lebesgue approximable on E if, for any sequence $R_n \rightarrow 0$ and any $\delta > 0$, $\exists n = n_0(\delta) \in \mathbb{Z}^+$ (independent of x) such that if $n > n_0$, then for almost all $x \in E$,

$$g(x) - \delta < \frac{\int_{\mathcal{B}(x, R_n) \cap E} g(t) dt}{\mathcal{V}(\mathcal{B}(x, R_n) \cap E)} < g(x) + \delta,$$

where $\mathcal{B}(x, R)$ is the closed ball around point $x \in \mathbb{R}^d$ with radius R , and $\mathcal{V}(\mathcal{B}(x, R))$ is the volume of the ball.

□

Theorem 5. Assume E is finite, $|E| = \chi$. Assume the conditions of Lemma 3 hold for distributions p and $q^{(\mathbf{w})}$. Assume that a unique $\mathbf{w}^*$ maximizes $D_h(p||q^{(\mathbf{w})})$ and arrange parameter values $\mathbf{w}^j$ for $j \in \{1, \dots, \chi\}$ in order such that they are descending in $D_h(p||q^{(\mathbf{w})})$. That is $\mathbf{w}^1 = \mathbf{w}^*$, using $\mathbf{w}^2$ gives the next biggest value and so on. Then

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\operatorname{argmax}_{\mathbf{w} \in S} \hat{D}_h \left(X_{1:N} || Y_{1:N}^{(\mathbf{w})} \right) = \mathbf{w}^* \right) = 1. \quad (\text{B.2})$$

Proof. Let $\epsilon > 0$. Take $\delta < D_h(p||q^{(\mathbf{w}^*)}) - D_h(p||q^{(\mathbf{w}^2)})$. Using Lemma 3, we have L_2 convergence for the estimator $\hat{D}_h$ and, via Lemma 4, this implies convergence in probability. Therefore $\exists M \in \mathbb{N}$ such that $\forall n \geq M$

$$\mathbb{P} \left(\left| \hat{D}_h \left(X_{1:N} || Y_{1:N}^{(\mathbf{w})} \right) - D_h(p||q^{(\mathbf{w})}) \right| > \delta \right) < \epsilon.$$

Therefore $\forall N \geq M$

$$\mathbb{P} \left(\operatorname{argmax}_{\mathbf{w} \in E} \hat{D}_h \left(X_{1:N} || Y_{1:N}^{(\mathbf{w})} \right) = \mathbf{w}^* \right) > 1 - \epsilon.$$

□

If we make the (generally unrealistic) assumption that the space of possible parameters, E , is finite, then we are able to show that, in the limit of a large number of samples, we can recover the optimum parameters with probability 1. We next explore how this can be extended to a compact, continuous space of parameters, E , provided we make assumptions about the structure of dependence of the distance estimator on parameters $\mathbf{w} \in E$. We show that assumptions about this dependence structure of the estimator on the parameters can be justified by considering some of the details of Algorithm 5.

Lemma 6. Consider the estimator of the Hellinger distance as a function of parameters, $\mathbf{w}$, such that

$$L(\mathbf{w}) = \hat{D}_h \left(X_{1:N} || Y_{1:N}^{(\mathbf{w})} \right).$$

Then $L(\mathbf{w})$ is piecewise constant with respect to parameters $\mathbf{w} \in E$, with finitely many discontinuities.

Proof. Suppose that, according to Algorithm 5, at generation t , we have generated pseudo data $\{\mathbf{x}^{i*}\}_{i=1}^N$, which we summarize via summary statistics $s(\mathbf{x}^{i*}) = (s_1, \dots, s_\kappa) \in \mathbb{R}^\kappa$ and κ is the number of summary statistics used to summarise the model output. The parameters $\mathbf{w} \in \mathbb{R}^\kappa$ are summary statistic weights, and these are used within a weighted Euclidean distance function

$$\begin{aligned} d_{\mathbf{w}}(s(\mathbf{x}), s(\mathbf{y})) &= \sum_{j=1}^{\kappa} w_j^2 (s_j - s_j(\mathbf{y}))^2 \\ &= (s(\mathbf{x}) - s(\mathbf{y}))^\top \Sigma_w^\top \Sigma_w (s(\mathbf{x}) - s(\mathbf{y})) \\ &= (\Sigma_w s(\mathbf{x}) - \Sigma_w s(\mathbf{y}))^\top (\Sigma_w s(\mathbf{x}) - \Sigma_w s(\mathbf{y})), \end{aligned}$$

where $\Sigma_w = \text{diag}(\mathbf{w})$, to compare the pseudo data with observed data. We note that this is equivalent to stretching the space in which the pseudo data lies via the matrix Σ_w , and using the usual Euclidean distance.

Consider a small perturbation in parameter space $\mathbf{w} = \mathbf{w}_0 + \epsilon$, with $\|\epsilon\| \ll 1$. Then

$$\Sigma_{\mathbf{w}} = \text{diag}(\mathbf{w}_0 + \epsilon) = \text{diag}(\mathbf{w}_0) + \text{diag}(\epsilon) = \Sigma_{\mathbf{w}_0} + \Sigma_\epsilon.$$

Using this decomposition of $\Sigma_{\mathbf{w}}$ gives, for the weighted Euclidean distance,

$$\begin{aligned} d_{\mathbf{w}_0 + \epsilon}(s(\mathbf{x}), s(\mathbf{y})) &= ((\Sigma_{\mathbf{w}_0} + \Sigma_\epsilon) s(\mathbf{x}) - (\Sigma_{\mathbf{w}_0} + \Sigma_\epsilon) s(\mathbf{y}))^\top \\ &\quad ((\Sigma_{\mathbf{w}_0} + \Sigma_\epsilon) s(\mathbf{x}) - (\Sigma_{\mathbf{w}_0} + \Sigma_\epsilon) s(\mathbf{y})) \\ &= (\Sigma_{\mathbf{w}_0} s(\mathbf{x}) - \Sigma_{\mathbf{w}_0} s(\mathbf{y}))^\top (\Sigma_{\mathbf{w}_0} s(\mathbf{x}) - \Sigma_{\mathbf{w}_0} s(\mathbf{y})) \\ &\quad + (\Sigma_\epsilon s(\mathbf{x}) - \Sigma_\epsilon s(\mathbf{y}))^\top (\Sigma_{\mathbf{w}_0} s(\mathbf{x}) - \Sigma_{\mathbf{w}_0} s(\mathbf{y})) \\ &\quad + (\Sigma_{\mathbf{w}_0} s(\mathbf{x}) - \Sigma_{\mathbf{w}_0} s(\mathbf{y}))^\top (\Sigma_\epsilon s(\mathbf{x}) - \Sigma_\epsilon s(\mathbf{y})) + \mathcal{O}(\epsilon^2) \\ &= d_{\mathbf{w}_0}(s(\mathbf{x}), s(\mathbf{y})) + A + A^\top + \mathcal{O}(\epsilon^2), \end{aligned}$$

where $A = (\Sigma_{\mathbf{w}_0} s(\mathbf{x}) - \Sigma_{\mathbf{w}_0} s(\mathbf{y}))^\top (\Sigma_\epsilon s(\mathbf{x}) - \Sigma_\epsilon s(\mathbf{y}))$ which is linear in ϵ . Suppose we order the pseudo-data such that $\mathbf{x}^{j*}$ is the j th closest point to the observed data based on $d_{\mathbf{w}_0}$. Provided that

$$d_{\mathbf{w}_0}(s(\mathbf{x}^{n*}), s(\mathbf{y})) - d_{\mathbf{w}_0}(s(\mathbf{x}^{(n+1)*}), s(\mathbf{y})) > A + A^\top + \mathcal{O}(\epsilon^2), \quad (\text{B.3})$$

then making this perturbation in $\mathbf{w}$ will not change which parameter samples are selected, as the same pseudo data will remain closest to the observed data, $\mathbf{y}$. If the same parameter samples are selected, then the value of $L(\mathbf{w})$ will remain constant as a function of $\mathbf{w}$ under the perturbation $\mathbf{w} = \mathbf{w}_0 + \epsilon$.

In cases where Equation (B.3) does not hold, there will be a jump discontinuity in $L(\mathbf{w})$ as different parameter samples are selected. This will occur finitely many

times corresponding to the finite number, N , of points in the sample of pseudo data, $\{\mathbf{x}^{i*}\}_{i=1}^N$. $\square$

We confirm computationally that $L(\mathbf{w})$ is piecewise constant for the test problem described in Section 5.4.1, and show this in Figure B.1.

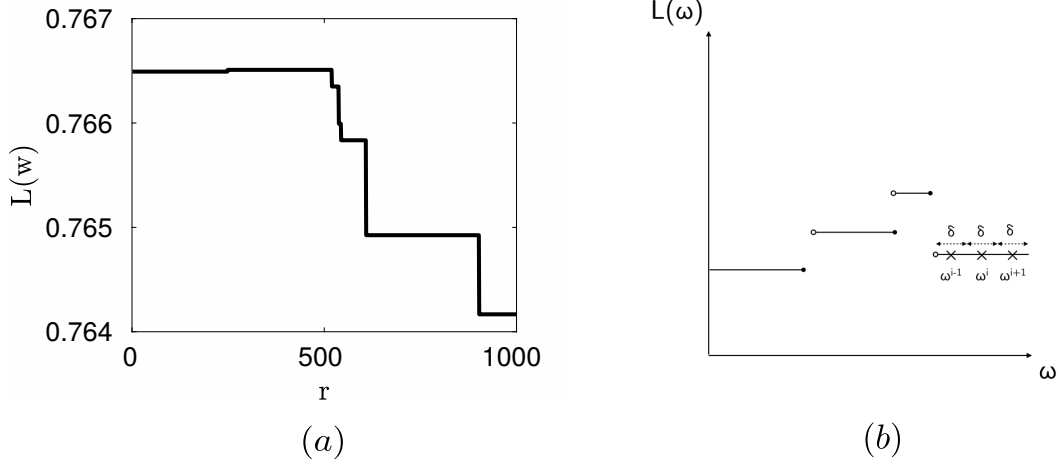


Figure B.1. The estimator $L(\mathbf{w}) = \hat{D}_h \left(X_{1:N} || Y_{1:N}^{(\mathbf{w})} \right)$ is piecewise constant for the toy model described in Section 5.4.1 when $X_{1:N}$ are samples from the prior and $Y_{1:N}^{(\mathbf{w})}$ are samples from the approximate posterior generated via ABC with summary statistics weights $\mathbf{w}$. After a single generation of ABC-SMC, we optimize $L(\mathbf{w})$ as a function of $\mathbf{w}$ to find a local maximum $\mathbf{w}^*$. In (a), we then consider the value of $L(\mathbf{w})$ on a line of parameters in parameter space $\mathbf{w} = \mathbf{w}^* + 10^{-4}r\vec{\eta}$, where $\vec{\eta} \sim \mathcal{N}(\mathbf{0}, I_\kappa)$ is a random choice of direction, I_κ is the $\kappa \times \kappa$ identity matrix and r parameterizes a line in this direction. $L(\mathbf{w})$ is piecewise constant in $\mathbf{w}$ as shown in Lemma 6. For a piecewise constant function $L(\mathbf{w})$, we can choose δ and $\mathbf{w}^j$ for $j \in \{1, \dots, \kappa\}$ such that $L(\mathbf{w})$ is locally constant. We illustrate such a choice of δ and $\mathbf{w}^j$ in (b).

Lemma 7. Assume $E \subset \mathbb{R}^\kappa$ is compact and that a unique $\mathbf{w}^* \in E$ maximizes $D_h(p||q^{(\mathbf{w})})$. Assume the conditions of Lemma 3 hold for distributions p and $q^{(\mathbf{w})}$. Suppose $L(\mathbf{w})$ is piecewise constant in $\mathbf{w}$ with finitely many jump discontinuities. Then

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\operatorname{argmax}_{\mathbf{w} \in E} \hat{D}_h \left(X_{1:n} || Y_{1:n}^{(\mathbf{w})} \right) = \mathbf{w}^* \right) = 1.$$

Proof. We can choose a $\delta > 0$ and finitely many $\mathbf{w}^j$, $j \in \{1, \dots, \chi\}$ such that every point $\mathbf{w}$ is within a ball of radius δ from some $\mathbf{w}^j$ (see Figure B.1 and note that L is locally constant). Then $\forall \mathbf{w} \in E \ ||\mathbf{w} - \mathbf{w}^j|| < \delta \implies L(\mathbf{w}) = L(\mathbf{w}^j) = L_j$. Since there are finitely many values $\mathbf{w}^j$ corresponding to distinct unique values L^j , we can apply the result from Theorem 5 to give the required result. $\square$

Theorem 8. Suppose that $k \geq 2$ and that $\mathcal{M} = \text{supp}(p)$. Assume that

- (a) q is bounded above,
- (b) p is bounded away from zero,
- (c) p is uniformly Lebesgue approximable,
- (d) $\exists \delta_0$ such that $\forall \delta \in (0, \delta_0)$

$$\int_{\mathcal{M}} H(x, p, \delta, 1/2) p(x) dx < \infty,$$

- (e) $\int_{\mathcal{M}} \|x - y\|^\gamma p(y) dy < \infty$ for almost all $x \in \mathcal{M}$,
- $\int \int_{\mathcal{M}^2} \|x - y\|^\gamma p(y) p(x) dy dx < \infty$,
- (f) $E \subset \mathbb{R}^\kappa$ is compact,
- (g) a unique $\mathbf{w}^* \in E$ maximizes $D_h(p||q^{(\mathbf{w})})$.

Then

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\operatorname{argmax}_{w \in E} \hat{D}_h \left(X_{1:N} || Y_{1:N}^{(\mathbf{w})} \right) = \mathbf{w}^* \right) = 1. \quad (\text{B.4})$$

Proof. Apply Lemma 6 to show that $L(\mathbf{w})$ is piecewise constant in $\mathbf{w}$. Apply Lemma 7 to give the desired result. $\square$

Theorem 9. Prangle [111]

Suppose that the following conditions hold:

- (a) $\theta \in \mathbb{R}^\lambda$, $s(\mathbf{x}) \in \mathbb{R}^\kappa$ and these random variables have joint density $\pi(\theta, s(\mathbf{x}))$ with respect to the Lebesgue measure.
- (b) The sets $A_t = \{s(\mathbf{x}) \mid d_{\mathbf{w}_t^*}(s(\mathbf{x}), s(\mathbf{y})) < \epsilon_t \forall t\}$ are Lebesgue measurable, where $\mathbf{w}^*$ are the optimal weights at generation t , and ϵ_t are the tolerances at each generation.
- (c) $\pi(s(\mathbf{y})) > 0$.
- (d) $\lim_{t \rightarrow \infty} m(A_t) = 0$, where $m(\cdot)$ represents Lebesgue measure.
- (e) The sets A_t have bounded eccentricity. That is for any A_t , $\exists$ a ball $B_t = \{s(\mathbf{x}) \mid \|s(\mathbf{x}) - s(\mathbf{y})\|_2 \leq r_t\}$ about $s(\mathbf{y})$ with radius r_t such that $A_t \subset B_t$ and $m(A_t) \geq c m(B_t)$, where $\|\cdot\|$ is the Euclidean or L_2 norm and $c > 0$ is a constant. (This property implies that A_t is contained within a ball B_t , but that its measure is comparable to the measure of the ball [294]).

Then the posterior approximations converge:

$$\lim_{t \rightarrow \infty} p_{ABC,t}(\theta | s(\mathbf{y})) = p(\theta | s(\mathbf{y})) \text{ for almost all } (\theta, s(\mathbf{y})),$$

where $p_{ABC,t}(\theta | s(\mathbf{y}))$ is the ABC posterior defined as:

$$p_{ABC,t}(\theta | s(\mathbf{y})) \propto \int p(s | \theta) \pi(\theta) \mathbb{1}[d_{\mathbf{w}_t^*}(s(\mathbf{x}), s(\mathbf{y})) \leq \epsilon_t] d\mathbf{x}.$$

Proof.

$$\begin{aligned}
\lim_{t \rightarrow \infty} p_{\text{ABC},t}(\theta|s(\mathbf{y})) &= \lim_{t \rightarrow \infty} \frac{\int p(\theta, s) \mathbf{1}[s \in A_t] \, ds}{\int p(\theta, s) \mathbf{1}[s \in A_t] \, ds d\theta} \\
&= \lim_{t \rightarrow \infty} \frac{\int_{s \in A_t} p(\theta, s) \, ds}{\int_{s \in A_t} p(\theta, s) \, ds d\theta} \\
&= \frac{\lim_{t \rightarrow \infty} \frac{1}{m(A_t)} \int_{s \in A_t} p(\theta, s) \, ds}{\lim_{t \rightarrow \infty} \frac{1}{m(A_t)} \int_{s \in A_t} p(s) \, ds} \\
&= \frac{p(\theta, s(\mathbf{y}))}{p(s(\mathbf{y}))} \text{ almost everywhere} \\
&= p(\theta|s(\mathbf{y})).
\end{aligned}$$

The fourth equality above is due to the Lebesgue differentiation theorem [294], which requires bounded eccentricity and $\lim_{t \rightarrow \infty} m(A_t) = 0$. $\square$

Test problem	κ	$\mathbf{A}(\mathbf{0})$	T	θ^*	N	α	Repeats	Prior interval
Uniform toy model	10	-	-	10	2000	0.5	1	$\log_{10} \theta \sim U[-4, 4]$
Bimodal toy model	2	-	-	$(\pi/4, 5\pi/4)$	2000	0.5	1	$\theta \sim U[0, 2\pi]$
Gaussian toy model	2	-	-	(0,0)	2000	0.5	1	$\theta \sim N(0, 100)$
Death process	9	10	20	(0.1, 0.01)	2000	0.5	1	$\log_{10} \theta \sim U[-4, 4]$
Dimerization	24	$(10^5, 0, 0)$	100	$(1, 0.04, 0.002, 0.5)$	2000	0.5	1	$\log_{10} \theta \sim U([-2, 2] \times [-3, 1] \times [-5, -1] \times [-3, 1])$
Diffusion	64	$10 * \mathbb{1}_{x < 0}$	20	0.1	2000	0.5	1	$\log_{10} \theta \sim U[-4, 0]$

Table B.1. Summary of hyperparameters used in simulations.

Appendix C

Supplementary Material for Chapter 6

C.1 Experimental methods and materials

C.1.1 Fly strains and tissue preparation

Stocks were raised on standard cornmeal-agar medium at 25°C. The wild type was Oregon R (OrR). Over expression of *gurken* RNA was obtained by crossing *UASp grk3A* (based upon genomic sequence DS02110, which includes the full 3' and 5' UTRs [291]) to maternal tubulin driver *TubulinGal4-VP16*. Tagging of *gurken* RNA for live imaging was achieved using the MS2-MCP(FP) system: *gurken*-(MS2)₁₂ [205], P[*nos-NLS-MCP-mCherry*] [191]; P[*w+nanos-5'-NLS-MCP-Dendra2 K10-3'*] (Rippe Hayashi). GFP protein trap lines: *Moesin::GFP* [292]; *Tau-GFP 65/167* (D. St Johnston, University of Cambridge). Oocytes were isolated for fixation and single molecule fluorescence in situ hybridization (smFISH), as described in Davidson et al. [295].

C.1.2 Single molecule FISH

smFISH detection of *gurken* RNA was performed as described in Davidson et al. [295]. Stellaris Oligonucleotide probes 20nt in length complementary to the *gurken* transcript (CG17610; 48 probes) conjugated to CAL Fluor Red 590 were obtained from Biosearch technologies). Labelled oocytes were mounted on slides in 70% Vectashield (Vector laboratories).

C.1.3 Fixed-cell imaging

Confocal imaging of fixed *Drosophila* oocytes was performed using an inverted Olympus FV3000 six laser line spectral confocal system fitted with high sensitivity gallium arsenide phosphide (GaAsP detectors) and using a 60x SI 1.3 NA lens. For smFISH detection of *gurken* transcripts, settings were optimized with a pinhole of 1.2 airy units and increasing laser power over the depth of the sample (between 30-50µm from the coverslip) to compensate for signal attenuation.

C.2 Quantifying production via nascent transcription

Previously in Section 6.3, we assumed equal production of mRNA in each of the nurse cell nuclei. Here we relax this assumption and consider estimating directly how much production occurs in each nurse cell nucleus. The UAS system is used as a genetic driver to drive the over-expression of *gurken*, and this may not operate identically in all the nurse cells. This justifies considering inhomogeneous production for the over-expression mutant, but nonetheless assuming equal production across nurse cells in the wild type.

Challenges arise in the nature of the data available in that the smFISH data is from fixed specimens capturing a snapshot of the dynamics at a particular point in time. Transcription is a noisy process and there is some evidence that time averaging of transcription helps regulate gene expression [287]. By considering only a single sample in time from each nurse cell, we lack information to make sound inferences that can extrapolate to other time points. Instead, we assume that there is a characteristic distribution of transcription across cells within the egg chamber.

To quantify transcription in each nurse cell, we measure the brightness of nascent transcripts in the nuclei of nurse cells. We perform these measurements based on an average projection in the z direction through the nucleus of each nurse cell, and apply a threshold to remove background signal using FIJI (ImageJ V1.51d; *fiiji.sc*, [223]). We note that the resulting distributions of fluorescence across cells for different over-expression egg chambers are very noisy, as shown in Figure 6.15. In particular, a notable minority of cells have a measured nascent transcription value of zero. To overcome problems caused from these zero observations, we use a mixture model and treat the data as missing with some probability, p , in which case we observe zero for the nascent transcription in a given cell; otherwise with probability $1 - p$ we use a poisson model for the nascent transcription. Suppose $\mathbf{a} = (a_2, \dots, a_{16})$ are the production rates in each cell (we assume $a_1=0$ since the oocyte is transcriptionally

silent), and suppose also that $\tau = (\tau_2, \dots, \tau_{16})$ are the observed nascent transcription values. The model to estimate the rates of inhomogeneous production in each cell takes the following form:

$$\tau \sim (1 - p) \text{Poisson}(\tilde{\mathbf{a}}) + p \delta(x), \quad (\text{C.1})$$

where $\delta(x)$ is a delta function, and $\tilde{\mathbf{a}}$ are the unnormalized production rates such that $\mathbf{a} = 15\gamma\tilde{\mathbf{a}}/\sum_i \tilde{a}_i$. We assume a prior uniform on $(0, 1)$ for the missing data probability, p , and a Half-Cauchy(0, 5) prior with support on $(0, \infty)$ for the inhomogeneous production rate $\tilde{\mathbf{a}}$. The resulting estimates for inhomogeneous production, $\mathbf{a}$, across nurse cells of the egg chamber are shown in Figure 6.16 and are used in Section 6.10 for the inhomogeneous production model.

Bibliography

- [1] J U Harrison and R E Baker. The impact of temporal sampling resolution on parameter inference for biological transport models. *PLoS Computational Biology*, 14(6):e1006235, 2018.
- [2] J U Harrison and R E Baker. An automatic adaptive method to combine summary statistics in approximate Bayesian computation. *arXiv preprint arXiv:1703.02341*, 2017.
- [3] J U Harrison, R M Parton, I Davis, and R E Baker. Testing models of mRNA localization reveals robustness regulated by reducing transport between cells. *BioRxiv preprint BioRxiv:533133v1*, 2019.
- [4] A Mogilner, R Wollman, and W F Marshall. Quantitative modeling in cell biology: what is it good for? *Developmental Cell*, 11(3):279–287, 2006.
- [5] C J Tomlin and J D Axelrod. Biology by numbers: mathematical modelling in developmental biology. *Nature Reviews Genetics*, 8(5):331, 2007.
- [6] P M Altrock, L L Liu, and F Michor. The mathematics of cancer: integrating quantitative models. *Nature Reviews Cancer*, 15(12):730, 2015.
- [7] S Ekins, J Mestres, and B Testa. In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. *British Journal of Pharmacology*, 152(1):9–20, 2007.
- [8] G R Mirams, M R Davies, Y Cui, P Kohl, and D Noble. Application of cardiac

- electrophysiology simulations to pro-arrhythmic safety testing. *British Journal of Pharmacology*, 167(5):932–945, 2012.
- [9] A C Babbie and M P H Stumpf. How to deal with parameters for whole-cell modelling. *Journal of The Royal Society Interface*, 14(133):20170237, 2017.
- [10] F Hartig, J M Calabrese, B Reineking, T Wiegand, and A Huth. Statistical inference for stochastic simulation models—theory and application. *Ecology Letters*, 14(8):816–827, 2011.
- [11] S N Wood. Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, 466(7310):1102–1104, 2010.
- [12] C C Drovandi, A N Pettitt, and M J Faddy. Approximate Bayesian computation using indirect inference. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 60(3):317–337, 2011.
- [13] L F Price, C C Drovandi, A Lee, and D J Nott. Bayesian synthetic likelihood. *Journal of Computational and Graphical Statistics*, 2017.
- [14] J K Pritchard, M T Seielstad, A Perez-Lezaun, and M W Feldman. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution*, 16(12):1791–1798, 1999.
- [15] M A Beaumont, W Zhang, and D J Balding. Approximate Bayesian computation in population genetics. *Genetics*, 162(4):2025–2035, 2002.
- [16] T Toni, D Welch, N Strelkowa, A Ipsen, and M P H Stumpf. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, 6(31):187–202, 2009.
- [17] M A Beaumont. Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution and Systematics*, 41:379–405, 2010.

- [18] M Sunnåker, A G Busetto, E Numminen, J Corander, M Foll, and C Dessi-moz. Approximate Bayesian computation. *PLOS Computational Biology*, 9(1):e1002803, 2013.
- [19] E O Buzbas and N A Rosenberg. AABC: Approximate approximate Bayesian computation for inference in population-genetic models. *Theoretical Population Biology*, 99:31–42, 2015.
- [20] J R Karr, J C Sanghvi, D N Macklin, M V Gutschow, J M Jacobs, B Bolival, N Assad-Garcia, J I Glass, and M W Covert. A whole-cell computational model predicts phenotype from genotype. *Cell*, 150(2):389–401, 2012.
- [21] O Purcell, B Jain, J R Karr, M W Covert, and T K Lu. Towards a whole-cell modeling approach for synthetic biology. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 23(2):025112, 2013.
- [22] J C Sanghvi, S Regot, S Carrasco, J R Karr, M V Gutschow, B Bolival Jr, and M W Covert. Accelerated discovery via a whole-cell model. *Nature Methods*, 10(12):1192–1195, 2013.
- [23] R Erban, M B Flegg, and G A Papoian. Multiscale stochastic reaction–diffusion modeling: application to actin dynamics in filopodia. *Bulletin of Mathematical Biology*, 76(4):799–818, 2014.
- [24] M Mak, T Kim, M H Zaman, and R D Kamm. Multiscale mechanobiology: computational models for integrating molecules to multicellular systems. *Integrative Biology*, 7:1093–1108, 2015.
- [25] R Erban. Coupling all-atom molecular dynamics simulations of ions in water with Brownian dynamics. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 472(2186), 2016.

- [26] E Rolls, Y Togashi, and R Erban. Varying the resolution of the Rouse model on temporal and spatial scales: application to multiscale modeling of DNA dynamics. *Multiscale Modeling & Simulation*, 15(4):1672–1693, 2017.
- [27] J M Osborne, A Walter, S K Kershaw, G R Mirams, A G Fletcher, P Pathmanathan, D J Gavaghan, O E Jensen, P K Maini, and H M Byrne. A hybrid approach to multi-scale modelling of cancer. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 368(1930):5013–5028, 2010.
- [28] C A Yates and M B Flegg. The pseudo-compartment method for coupling partial differential equation and compartment-based models of diffusion. *Journal of The Royal Society Interface*, 12(106):20150141, 2015.
- [29] F Spill, P Guerrero, T Alarcon, P K Maini, and H M Byrne. Hybrid approaches for multiple-species stochastic reaction–diffusion models. *Journal of Computational Physics*, 299:429–445, 2015.
- [30] J U Harrison and C A Yates. A hybrid algorithm for coupling partial differential equation and compartment-based dynamics. *Journal of The Royal Society Interface*, 13(122):20160335, 2016.
- [31] H M Byrne, T Alarcon, M R Owen, S D Webb, and P K Maini. Modelling aspects of cancer dynamics: a review. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 364(1843):1563–1578, 2006.
- [32] B Lambert, A L MacLean, A G Fletcher, A N Combes, M H Little, and H M Byrne. Bayesian inference of agent-based models: a tool for studying kidney branching morphogenesis. *Journal of Mathematical Biology*, 76(7):1673–1697, 2018.

- [33] G G Powathil, M Swat, and M A Chaplain. Systems oncology: Towards patient-specific treatment regimes informed by multiscale mathematical modelling. In *Seminars in Cancer Biology*, volume 30, pages 13–20. Elsevier, 2015.
- [34] T D Lewin, P K Maini, E G Moros, H Enderling, and H M Byrne. The evolution of tumour composition during fractionated radiotherapy: implications for outcome. *Bulletin of Mathematical Biology*, 80(5):1207–1235, 2018.
- [35] M G Blum, M A Nunes, D Prangle, and S A Sisson. A comparative review of dimension reduction methods in approximate Bayesian computation. *Statistical Science*, 28(2):189–208, 2013.
- [36] C Barnes, S Filippi, M P H Stumpf, and T Thorne. Considerate approaches to constructing summary statistics for ABC model selection. *Statistics and Computing*, 22(6):1181–1197, 2012.
- [37] P Fearnhead and D Prangle. Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation. *Journal of the Royal Statistical Society: Series B*, 74(3):419–474, 2012.
- [38] A P Browning, S W McCue, and M J Simpson. A Bayesian computational approach to explore the optimal duration of a cell proliferation assay. *Bulletin of Mathematical Biology*, 79(8):1888–1906, 2017.
- [39] D Silk, P W Kirk, C P Barnes, T Toni, and B Rost. Model selection in systems biology depends on experimental design. *PLOS Computational Biology*, 10(6):1–14, 2014.
- [40] J Kursawe, R E Baker, and A G Fletcher. Impact of implementation choices on quantitative predictions of cell-based computational models. *Journal of Computational Physics*, 2018.

- [41] A Parker, M J Simpson, and R E Baker. The impact of experimental design choices on parameter inference for models of growing cell colonies. *Royal Society Open Science*, 5(8):180384, 2018.
- [42] J Owen, D J Wilkinson, and C S Gillespie. Scalable inference for Markov processes with intractable likelihoods. *Statistics and Computing*, 25(1):145–156, 2015.
- [43] A N Becalska and E R Gavis. Lighting up mRNA localization in *Drosophila* oogenesis. *Development*, 136(15):2493–2503, 2009.
- [44] R M Parton, A Davidson, I Davis, and T T Weil. Subcellular mRNA localisation at a glance. *Journal of Cell Science*, 127(10):2127–2133, 2014.
- [45] A R Buxbaum, G Haimovich, and R H Singer. In the right place at the right time: visualizing and understanding mRNA localization. *Nature Reviews Molecular Cell Biology*, 2014.
- [46] J Stelling, U Sauer, Z Szallasi, F J Doyle III, and J Doyle. Robustness of cellular functions. *Cell*, 118(6):675–685, 2004.
- [47] K E Hines. A primer on Bayesian inference for biophysical systems. *Biophysical Journal*, 108(9):2103–2113, 2015.
- [48] C Robert and G Casella. *Monte Carlo statistical methods*. Springer Science & Business Media, 2013.
- [49] M B Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008.
- [50] M B Giles. Multilevel Monte Carlo methods. *Acta Numerica*, 24:259, 2015.
- [51] C Lester, C A Yates, M B Giles, and R E Baker. An adaptive multi-level simulation algorithm for stochastic biological systems. *The Journal of Chemical Physics*, 142(2):024113, 2015.

- [52] C Lester, C A Yates, and R E Baker. Robustly simulating biochemical reaction kinetics using multi-level Monte Carlo approaches. *Journal of Computational Physics*, 375:1401–1423, 2018.
- [53] B Peherstorfer, K Willcox, and M Gunzburger. Optimal model management for multifidelity Monte Carlo estimation. *SIAM Journal on Scientific Computing*, 38(5):A3163–A3194, 2016.
- [54] A B Owen. Randomly permuted (t, m, s) -nets and (t, s) -sequences. In *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, pages 299–317. Springer, 1995.
- [55] A B Owen. Scrambling Sobol’ and Niederreiter–Xing points. *Journal of Complexity*, 14(4):466–489, 1998.
- [56] A Hellander. Efficient computation of transient solutions of the chemical master equation based on uniformization and quasi-Monte Carlo. *The Journal of Chemical Physics*, 128(15):154109, 2008.
- [57] C H Beentjes and R E Baker. Quasi-Monte Carlo methods applied to tau-leaping in stochastic biological systems. *Bulletin of Mathematical Biology, Special Issue: Gillespie and his Algorithms*:1–29, May 2018.
- [58] A Gelman, J B Carlin, H S Stern, D B Dunson, A Vehtari, and D B Rubin. *Bayesian Data Analysis*, volume 2. CRC press Boca Raton, FL, 2014.
- [59] N Metropolis, A W Rosenbluth, M N Rosenbluth, A H Teller, and E Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.
- [60] G O Roberts, A Gelman, and W R Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*, 7(1):110–120, 1997.

- [61] G O Roberts and J S Rosenthal. Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, 16(4), 2001.
- [62] M Betancourt. A conceptual introduction to Hamiltonian Monte Carlo. *arXiv preprint arXiv:1701.02434*, 2017.
- [63] R M Neal. MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 2(11):2, 2011.
- [64] M Betancourt, S Byrne, S Livingstone, and M Girolami. The geometric foundations of Hamiltonian Monte Carlo. *Bernoulli*, 23(4A):2257–2298, 2017.
- [65] B Carpenter, A Gelman, M Hoffman, D Lee, B Goodrich, M Betancourt, M Brubaker, J Guo, P Li, and A Riddell. Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1):1–32, 2017.
- [66] D Tran, A Kucukelbir, A B Dieng, M Rudolph, D Liang, and D M Blei. Edward: A library for probabilistic modeling, inference, and criticism. *arXiv preprint arXiv:1610.09787*, 2016.
- [67] M D Hoffman and A Gelman. The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1):1593–1623, 2014.
- [68] B Carpenter, M Hoffman, M Brubaker, D Lee, P Li, and M Betancourt. The stan math library: Reverse-mode automatic differentiation in C++. *arXiv preprint arXiv:1509.07164*, 2015.
- [69] G O Roberts and J S Rosenthal. General state space Markov chains and MCMC algorithms. *Probability Surveys*, 1:20–71, 2004.
- [70] J M Flegal, M Haran, and G L Jones. Markov chain Monte Carlo: Can we trust the third significant figure? *Statistical Science*, 23(2):250–260, 2008.

- [71] J Gabry, D Simpson, A Vehtari, M Betancourt, and A Gelman. Visualization in Bayesian workflow. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(2):389–402, 2019.
- [72] O J Maclaren, A Parker, C Pin, S R Carding, A J M Watson, A G Fletcher, H M Byrne, and P K Maini. A hierarchical Bayesian model for understanding the spatiotemporal dynamics of the intestinal epithelium. *PLOS Computational Biology*, 13(7):e1005688, 2017.
- [73] S Talts, M Betancourt, D Simpson, A Vehtari, and A Gelman. Validating Bayesian inference algorithms with simulation-based calibration. *arXiv preprint arXiv:1804.06788*, 2018.
- [74] A Bouchard-Côté, S J Vollmer, and A Doucet. The bouncy particle sampler: A nonreversible rejection-free Markov chain Monte Carlo method. *Journal of the American Statistical Association*, 113(522):855–867, 2018.
- [75] C Wu and C P Robert. Generalized bouncy particle sampler. *arXiv preprint arXiv:1706.04781*, 2017.
- [76] J Bierkens, P Fearnhead, and G Roberts. The zig-zag process and super-efficient sampling for Bayesian analysis of big data. *The Annals of Statistics*, 47(3):1288–1320, 2019.
- [77] A Doucet, S Godsill, and C Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10(3):197–208, 2000.
- [78] P Del Moral, A Doucet, and A Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436, 2006.
- [79] N J Gordon, D J Salmond, and A F Smith. Novel approach to nonlinear

- non-Gaussian Bayesian state estimation. In *IEEE Proceedings F*, volume 140, pages 107–113. IET, 1993.
- [80] O A Chkrebtii, D A Campbell, B Calderhead, and M A Girolami. Bayesian solution uncertainty quantification for differential equations. *Bayesian Analysis*, 11(4):1239–1267, 2016.
- [81] J Moore, H Ahmed, J Jia, R Akondy, R Ahmed, and R Antia. What controls the acute viral infection following yellow fever vaccination? *Bulletin of Mathematical Biology*, 80(1):46–63, 2018.
- [82] S P Pearce, M Heil, O E Jensen, G W Jones, and A Prokop. Curvature-sensitive Kinesin binding can explain microtubule ring formation and reveals chaotic dynamics in a mathematical model. *Bulletin of Mathematical Biology*, 80(11):3002–3022, 2018.
- [83] E F Keller and L A Segel. Traveling bands of chemotactic bacteria: a theoretical analysis. *Journal of Theoretical Biology*, 30(2):235–248, 1971.
- [84] James D Murray. *Mathematical Biology: I. An Introduction (interdisciplinary applied mathematics)(Pt. 1)*. New York, Springer, 2007.
- [85] R Bellman and K J Åström. On structural identifiability. *Mathematical Biosciences*, 7(3-4):329–339, 1970.
- [86] R N Gutenkunst, J J Waterfall, F P Casey, K S Brown, C R Myers, and J P Sethna. Universally sloppy parameter sensitivities in systems biology models. *PLoS Computational biology*, 3(10):e189, 2007.
- [87] A F Villaverde, A Barreiro, and A Papachristodoulou. Structural identifiability of dynamic systems biology models. *PLOS Computational Biology*, 12(10):e1005153, 2016.

- [88] K E Hines, T R Middendorf, and R W Aldrich. Determination of parameter identifiability in nonlinear biophysical models: A Bayesian approach. *The Journal of General Physiology*, 143(3):401–416, 2014.
- [89] A A Sher, K Wang, A Wathen, P J Maybank, G R Mirams, D Abramson, D Noble, and D J Gavaghan. A local sensitivity analysis method for developing biological models with identifiable parameters: Application to cardiac ionic channel modelling. *Future Generation Computer Systems*, 29(2):591–598, 2013.
- [90] D V Raman, J Anderson, and A Papachristodoulou. Delineating parameter unidentifiabilities in complex models. *Physical Review E*, 95(3):032314, 2017.
- [91] E Dufresne, H A Harrington, and D V Raman. The geometry of sloppiness. *Journal of Algebraic Statistics*, 9:30–68, 2018.
- [92] D A Campbell and S R Lele. An ANOVA test for parameter estimability using data cloning with application to statistical inference for dynamic systems. *Computational Statistics & Data Analysis*, 70:257–267, 2014.
- [93] S R Lele, B Dennis, and F Lutscher. Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. *Ecology Letters*, 10(7):551–563, 2007.
- [94] I Siekmann, J Sneyd, and E J Crampin. MCMC can detect nonidentifiable models. *Biophysical Journal*, 103(11):2275–2286, 2012.
- [95] A C Daly, D J Gavaghan, C Holmes, and J Cooper. Hodgkin–Huxley revisited: reparametrization and identifiability analysis of the classic action potential model with approximate Bayesian methods. *Royal Society Open Science*, 2(12):150499, 2015.

- [96] J Liepe, P Kirk, S Filippi, T Toni, C Barnes, and M P H Stumpf. A framework for parameter estimation and model selection from experimental data in systems biology using approximate Bayesian computation. *Nature Protocols*, 9(2):439–456, 2014.
- [97] P Joyce and P Marjoram. Approximately sufficient statistics and Bayesian computation. *Statistical Applications in Genetics and Molecular Biology*, 7(1), 2008.
- [98] M G Blum and O François. Non-linear regression models for approximate Bayesian computation. *Statistics and Computing*, 20(1):63–73, 2010.
- [99] W Li and P Fearnhead. Convergence of regression-adjusted approximate Bayesian computation. *Biometrika*, 105(2):301–318, 2018.
- [100] D T Frazier, C P Robert, and J Rousseau. Model misspecification in ABC: consequences and diagnostics. *arXiv preprint arXiv:1708.01974*, 2017.
- [101] G S Rodrigues, D Prangle, and S A Sisson. Recalibration: A post-processing method for approximate Bayesian computation. *Computational Statistics & Data Analysis*, 126:53–66, 2018.
- [102] D Prangle, M G Blum, G Popovic, and S A Sisson. Diagnostic tools for approximate Bayesian computation using the coverage property. *Australian & New Zealand Journal of Statistics*, 56(4):309–329, 2014.
- [103] P Marjoram, J Molitor, V Plagnol, and S Tavaré. Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 100(26):15324–15328, 2003.
- [104] O Ratmann, O Jørgensen, T Hinkley, M H P Stumpf, S Richardson, and C Wiuf. Using likelihood-free inference to compare evolutionary dynamics of

- the protein networks of *H. pylori* and *P. falciparum*. *PLoS Computational Biology*, 3(11):e230, 2007.
- [105] R D Wilkinson. Approximate Bayesian computation (ABC) gives exact results under the assumption of model error. *Statistical Applications in Genetics and Molecular Biology*, 12(2):129–141, 2013.
 - [106] S A Sisson, Y Fan, and M M Tanaka. Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 104(6):1760–1765, 2007.
 - [107] M A Beaumont, J Cornuet, J Marin, and C P Robert. Adaptive approximate Bayesian computation. *Biometrika*, 96(4):983–990, 2009.
 - [108] S Filippi, C P Barnes, J Cornebise, and M P H Stumpf. On optimality of kernels for approximate Bayesian computation using sequential Monte Carlo. *Statistical Applications in Genetics and Molecular Biology*, 12(1):87–107, 2013.
 - [109] J Alsing, B D Wandelt, and S M Feeney. Optimal proposals for approximate Bayesian computation. *arXiv preprint arXiv:1808.06040*, 2018.
 - [110] D Silk, S Filippi, and M P H Stumpf. Optimizing threshold-schedules for sequential approximate Bayesian computation: applications to molecular systems. *Statistical Applications in Genetics and Molecular Biology*, 12(5):603–618, 2013.
 - [111] D Prangle. Adapting the ABC distance function. *Bayesian Analysis*, 12(1):289–309, 2017.
 - [112] D Prangle. Lazy ABC. *Statistics and Computing*, 26(1-2):171–185, 2016.
 - [113] S Wiqvist, U Picchini, and J L Forman. Accelerating delayed-acceptance Markov chain Monte Carlo algorithms. *arXiv preprint arXiv:1806.05982*, 2018.

- [114] T P Prescott and R E Baker. Multifidelity approximate Bayesian computation. *arXiv preprint arXiv:1811.09550*, 2018.
- [115] D M Blei, A Kucukelbir, and J D McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518): 859–877, 2017.
- [116] C Gourieroux, A Monfort, and E Renault. Indirect inference. *Journal of Applied Econometrics*, 8(S1):85–118, 1993.
- [117] Z An, D J Nott, and C Drovandi. Robust Bayesian synthetic likelihood via a semi-parametric approach. *arXiv preprint arXiv:1809.05800*, 2018.
- [118] R D Wilkinson. Accelerating ABC methods using Gaussian processes. *arXiv preprint arXiv:1401.1436*, 2014.
- [119] G Papamakarios and I Murray. Fast ε -free inference of simulation models with Bayesian conditional density estimation. In *Advances in Neural Information Processing Systems*, pages 1028–1036, 2016.
- [120] G Papamakarios, D C Sterratt, and I Murray. Sequential neural likelihood: fast likelihood-free inference with autoregressive flows. *arXiv preprint arXiv:1805.07226*, 2018.
- [121] A Pesah, A Wehenkel, and G Louppe. Recurrent machines for likelihood-free inference. *arXiv preprint arXiv:1811.12932*, 2018.
- [122] C Andrieu, A Doucet, and R Holenstein. Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B*, 72(3):269–342, 2010.
- [123] D A Rasmussen, O Ratmann, and K Koelle. Inference for nonlinear epidemiological models using genealogies and time series. *PLOS Computational Biology*, 7(8):e1002136, 2011.

- [124] A Golightly and D J Wilkinson. Bayesian parameter inference for stochastic biochemical network models using particle Markov chain Monte Carlo. *Interface Focus*, 1(6):807, 2011.
- [125] A Golightly, D A Henderson, and C Sherlock. Delayed acceptance particle MCMC for exact inference in stochastic kinetic models. *Statistics and Computing*, 25(5):1039–1055, 2015.
- [126] A Vehtari, D P Simpson, Y Yao, and A Gelman. Limitations of “Limitations of Bayesian leave-one-out cross-validation for model selection”. *Computational Brain & Behavior*, 2(1):22–27, 2019.
- [127] J Piironen and A Vehtari. Comparison of Bayesian predictive methods for model selection. *Statistics and Computing*, 27(3):711–735, 2017.
- [128] H Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723, 1974.
- [129] H Akaike. Information theory and an extension of the maximum likelihood principle. In *Selected papers of Hirotugu Akaike*, pages 199–213. Springer, 1998.
- [130] G Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.
- [131] K P Burnham and D R Anderson. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media, 2003.
- [132] R E Kass and A E Raftery. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, 1995.

- [133] D Warne, R E Baker, and M J Simpson. Using experimental data and information criteria to guide model selection for reaction–diffusion problems in mathematical biology. *bioRxiv*, page 444679, 2018.
- [134] D J Spiegelhalter, N G Best, B P Carlin, and A Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639, 2002.
- [135] D J Spiegelhalter, N G Best, B P Carlin, and A Linde. The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(3):485–493, 2014.
- [136] S Watanabe. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11:3571–3594, 2010.
- [137] H Jeffreys. *The Theory of Probability*. OUP Oxford, 1998.
- [138] S Shao, P E Jacob, J Ding, and V Tarokh. Bayesian model comparison with the Hyvärinen score: computation and consistency. *Journal of the American Statistical Association*, pages 1–33, 2018.
- [139] N Lartillot and H Philippe. Computing Bayes factors using thermodynamic integration. *Systematic Biology*, 55(2):195–207, 2006.
- [140] C Barnes, D Silk, and M P H Stumpf. Bayesian design strategies for synthetic biology. *Interface Focus*, 1(6):895–908, 2011. ISSN 2042-8898.
- [141] V Vyshemirsky and M A Girolami. Bayesian ranking of biochemical system models. *Bioinformatics*, 24(6):833–839, 2008.
- [142] C P Robert, J Cornuet, J Marin, and N S Pillai. Lack of confidence in approximate Bayesian computation model choice. *Proceedings of the National Academy of Sciences*, 108(37):15112–15117, 2011.

- [143] C Barnes, D Silk, X Sheng, and M P H Stumpf. Bayesian design of synthetic biological systems. *Proceedings of the National Academy of Sciences*, 108(37):15190–15195, 2011.
- [144] K Kamary, K Mengersen, C P Robert, and J Rousseau. Testing hypotheses via a mixture estimation model. *arXiv preprint arXiv:1412.2044*, 2014.
- [145] C P Robert. The expected demise of the Bayes factor. *Journal of Mathematical Psychology*, 72:33–37, 2016.
- [146] A Vehtari and J Ojanen. A survey of Bayesian predictive methods for model assessment, selection and comparison. *Statistics Surveys*, 6:142–228, 2012.
- [147] C P Robert. Intrinsic losses. *Theory and Decision*, 40(2):191–214, 1996.
- [148] Y Yao, A Vehtari, D Simpson, and A Gelman. Using stacking to average Bayesian predictive distributions. *Bayesian Analysis*, 2018.
- [149] A Vehtari, A Gelman, and J Gabry. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5):1413–1432, 2017.
- [150] A Vehtari, J Gabry, Y Yao, and A Gelman. loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models, 2018. URL <https://CRAN.R-project.org/package=loo>. R package version 2.0.0.
- [151] A Gelman, J Hwang, and A Vehtari. Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 24(6):997–1016, 2014.
- [152] J A Hoeting, D Madigan, A E Raftery, and C T Volinsky. Bayesian model averaging: a tutorial. *Statistical Science*, pages 382–401, 1999.
- [153] D H Wolpert. Stacked generalization. *Neural Networks*, 5(2):241–259, 1992.
- [154] L Breiman. Stacked regressions. *Machine Learning*, 24(1):49–64, 1996.

- [155] F Dyson. A meeting with Enrico Fermi. *Nature*, 427(6972):297, 2004.
- [156] J Wei. Least square fitting of an elephant. *Chemtech*, 5(2):128–129, 1975.
- [157] Steven T Piantadosi. One parameter is always enough. *AIP Advances*, 8(9):095118, 2018.
- [158] R E Baker, J M Peña, J Jayamohan, and A Jérusalem. Mechanistic models versus machine learning, a fight worth fighting for the biological community? *Biology Letters*, 14(5):20170660, 2018.
- [159] H Tekotte and I Davis. Intracellular mRNA localization: motors move messages. *Trends in Genetics*, 18(12):636–642, 2002.
- [160] W R Jeffery, C R Tomlinson, and R D Brodeur. Localization of actin messenger RNA during early ascidian development. *Developmental Biology*, 99(2):408–417, 1983.
- [161] J B Lawrence and R H Singer. Intracellular localization of messenger RNAs for cytoskeletal proteins. *Cell*, 45(3):407–415, 1986.
- [162] G S Wilkie and I Davis. Drosophila wingless and pair-rule transcripts localize apically by dynein-mediated transport of RNA particles. *Cell*, 105(2):209–219, 2001.
- [163] K L Mowry and D A Melton. Vegetal messenger RNA localization directed by a 340-nt RNA sequence element in *Xenopus* oocytes. *Science*, 255(5047):991–994, 1992.
- [164] N Bobola, R Jansen, T H Shin, and K Nasmyth. Asymmetric accumulation of Ash1p in postanaphase nuclei depends on a myosin and restricts yeast mating-type switching to mother cells. *Cell*, 84(5):699–709, 1996.
- [165] L A Mingle, N N Okuhama, J Shi, R H Singer, J Condeelis, and G Liu. Localization of all seven messenger RNAs for the actin-polymerization nucleator

- Arp2/3 complex in the protrusions of fibroblasts. *Journal of Cell Science*, 118 (11):2425–2433, 2005.
- [166] K Nevo-Dinur, A Nussbaum-Shochat, S Ben-Yehuda, and O Amster-Choder. Translation-independent localization of mRNA in *E. coli*. *Science*, 331(6020): 1081–1084, 2011.
- [167] C E Holt and S L Bullock. Subcellular mRNA localization in animal cells and why it matters. *Science*, 326(5957):1212–1216, 2009.
- [168] R Bastock and D St Johnston. *Drosophila* oogenesis. *Current Biology*, 18(23): R1082–R1087, 2008.
- [169] J A Doudna and E Charpentier. The new frontier of genome engineering with CRISPR-Cas9. *Science*, 346(6213):1258096, 2014.
- [170] T T Weil, R M Parton, and I Davis. Making the message clear: visualizing mRNA localization. *Trends in Cell Biology*, 20(7):380–390, 2010.
- [171] F S Neuman-Silberberg and T Schüpbach. The *Drosophila* dorsoventral patterning gene *gurken* produces a dorsally localized RNA and encodes a TGF α -like protein. *Cell*, 75(1):165–174, 1993.
- [172] W Driever and C Nüsslein-Volhard. The *bicoid* protein determines position in the *Drosophila* embryo in a concentration-dependent manner. *Cell*, 54(1): 95–104, 1988.
- [173] J Kim-Ha, J L Smith, and P M Macdonald. *oskar* mRNA is localized to the posterior pole of the *Drosophila* oocyte. *Cell*, 66(1):23–35, 1991.
- [174] C Wang and R Lehmann. *Nanos* is the localized posterior determinant in *Drosophila*. *Cell*, 66(4):637–647, 1991.

- [175] D Cai, W Dai, M Prasad, J Luo, N S Gov, and D J Montell. Modeling and analysis of collective cell migration in an *in vivo* three-dimensional environment. *Proceedings of the National Academy of Sciences*, 113(15):E2134–E2141, 2016.
- [176] P Lasko. RNA sorting in *Drosophila* oocytes and embryos. *The FASEB Journal*, 13(3):421–433, 1999.
- [177] V Riechmann and A Ephrussi. Axis formation during *Drosophila* oogenesis. *Current Opinion in Genetics & Development*, 11(4):374–383, 2001.
- [178] L Wolpert, R Beddington, J Brockes, T Jessell, P Lawrence, and E Meyerowitz. *Principles of Development*. Oxford University Press, Oxford, 1998.
- [179] S Mahajan-Miklos and L Cooley. Intercellular cytoplasm transport during *Drosophila* oogenesis. *Developmental Biology*, 165(2):336–351, 1994.
- [180] L Cáceres and L A Nilson. Production of gurken in the nurse cells is sufficient for axis determination in the *Drosophila* oocyte. *Development*, 132(10):2345–2353, 2005.
- [181] S L Bullock and D Ish-Horowicz. Conserved signals and machinery for RNA transport in *drosophila* oogenesis and embryogenesis. *Nature*, 414(6864):611–616, 2001.
- [182] A Clark, C Meignin, and I Davis. A Dynein-dependent shortcut rapidly delivers axis determination transcripts into the *Drosophila* oocyte. *Development*, 134(10):1955–1965, 2007.
- [183] A C Spradling. Developmental genetics of oogenesis. In *The development of *Drosophila melongaster**, pages 1–70. Cold Spring Harbour, NY, Cold Spring Harbour Laboratory Press, 1993.

- [184] D N Robinson, K Cant, and L Cooley. Morphogenesis of *Drosophila* ovarian ring canals. *Development*, 120(7):2015–2025, 1994.
- [185] Å H Eikenes, L Malerød, A Lie-Jensen, C S Wegner, A Brech, K Liestøl, H Stenmark, and K Haglund. Src64 controls a novel actin network required for proper ring canal formation in the *Drosophila* male germline. *Development*, 142(23):4107–4118, 2015.
- [186] J I Alsous, P Villoutreix, N Stoop, S Y Shvartsman, and J Dunkel. Entropic effects in cell lineage tree packings. *Nature Physics*, 14(10):1016, 2018.
- [187] N J Pokrywka and E C Stephenson. Microtubules mediate the localization of bicoid RNA during *Drosophila* oogenesis. *Development*, 113(1):55–66, 1991.
- [188] R M Parton, R S Hamilton, G Ball, L Yang, C F Cullen, W Lu, H Ohkura, and I Davis. A PAR-1-dependent orientation gradient of dynamic microtubules directs posterior cargo transport in the *Drosophila* oocyte. *The Journal of Cell Biology*, 194(1):121–135, 2011.
- [189] V L Zimyanin, K Belaya, J Pecreaux, M J Gilchrist, A Clark, I Davis, and D St Johnston. In vivo imaging of oskar mRNA transport reveals the mechanism of posterior localization. *Cell*, 134(5):843–853, 2008.
- [190] R S Hamilton, G Ball, and I Davis. A multidisciplinary approach to RNA localisation. In *Biophysical approaches to translational control of gene expression*, pages 213–233. Springer, 2013.
- [191] T T Weil, R M Parton, B Herpers, J Soetaert, T Veenendaal, D Xanthakis, I M Dobbie, J M Halstead, R Hayashi, C Rabouille, and I Davis. *Drosophila* patterning is established by differential association of mRNAs with P bodies. *Nature Cell Biology*, 14(12):1305–1313, 2012.

- [192] S A Isaacson, D M McQueen, and C S Peskin. The influence of volume exclusion by chromatin on the time required to find specific DNA binding sites by diffusion. *Proceedings of the National Academy of Sciences*, 108(9):3815–3820, 2011.
- [193] K M Forrest and E R Gavis. Live imaging of endogenous RNA reveals a diffusion and entrapment mechanism for nanos mRNA localization in *Drosophila*. *Current Biology*, 13(14):1159–1168, 2003.
- [194] T T Weil, K M Forrest, and E R Gavis. Localization of bicoid mRNA in late oocytes is maintained by continual active transport. *Developmental Cell*, 11(2):251–262, 2006.
- [195] J Howard. *Mechanics of Motor Proteins and the Cytoskeleton*. Sinauer, Sunderland, MA, 2001.
- [196] G Vendra, R S Hamilton, and I Davis. Dynactin suppresses the retrograde movement of apically localized mRNA in *Drosophila* blastoderm embryos. *RNA*, 13(11):1860–1867, 2007.
- [197] M Amrute-Nayak and S L Bullock. Single-molecule assays reveal that RNA localization signals regulate dynein-dynactin copy number on individual transcript cargoes. *Nature Cell Biology*, 14(4):416–423, 2012.
- [198] R Delanoue and I Davis. Dynein anchors its mRNA cargo after apical transport in the *Drosophila* blastoderm embryo. *Cell*, 122(1):97–106, 2005.
- [199] E Bertrand, P Chartrand, M Schaefer, S M Shenoy, R H Singer, and R M Long. Localization of ASH1 mRNA particles in living yeast. *Molecular Cell*, 2(4):437–445, 1998.
- [200] R S Hamilton, R M Parton, R A Oliveira, G Vendra, G Ball, K Nasmyth, and I Davis. Particlestats: open source software for the analysis of particle motility

and cytoskeletal polarity. *Nucleic Acids Research*, 38(suppl 2):W641–W646, 2010.

- [201] N MacDougall, A Clark, E MacDougall, and I Davis. *Drosophila gurken* (TGF α) mRNA localizes as particles that move within the oocyte in two dynein-dependent steps. *Developmental Cell*, 4(3):307–319, 2003.
- [202] V Van De Bor, E Hartswood, C Jones, D Finnegan, and I Davis. *gurken* and the *I* factor retrotransposon RNAs share common localization signals and machinery. *Developmental Cell*, 9(1):51–62, 2005.
- [203] A M Femino, F S Fay, K Fogarty, and R H Singer. Visualization of single RNA transcripts in situ. *Science*, 280(5363):585–590, 1998.
- [204] A Raj, P Van Den Bogaard, S A Rifkin, A Van Oudenaarden, and S Tyagi. Imaging individual mRNA molecules using multiple singly labeled probes. *Nature Methods*, 5(10):877–879, 2008.
- [205] A M Jaramillo, T T Weil, J Goodhouse, E R Gavis, and T Schupbach. The dynamics of fluorescently labeled endogenous *gurken* mRNA in *Drosophila*. *Journal of Cell Science*, 121(6):887–894, 2008.
- [206] F Persson, M Lindén, C Unoson, and J Elf. Extracting intracellular diffusive states and transition rates from single-molecule tracking data. *Nature Methods*, 10(3):265–269, 2013.
- [207] F Mueller, A Senecal, K Tantale, H Marie-Nelly, N Ly, O Collin, E Basyuk, E Bertrand, X Darzacq, and C Zimmer. FISH-quant: automatic counting of transcripts in 3D FISH images. *Nature Methods*, 10(4):277–278, 2013.
- [208] S Ganguly, L S Williams, I M Palacios, and R E Goldstein. Cytoplasmic streaming in *Drosophila* oocytes varies with kinesin activity and correlates

- with the microtubule cytoskeleton architecture. *Proceedings of the National Academy of Sciences*, 109(38):15109–15114, 2012.
- [209] P K Trong, H Doerflinger, J Dunkel, D St Johnston, and R E Goldstein. Cortical microtubule nucleation can organise the cytoskeleton of *Drosophila* oocytes to define the anteroposterior axis. *eLife*, 4:e06088, 2015.
- [210] M V Ciocanel, J A Kreiling, J A Gagnon, K L Mowry, and B Sandstede. Analysis of active transport by fluorescence recovery after photobleaching. *Biophysical Journal*, 112(8):1714–1725, 2017.
- [211] G T Reeves, C B Muratov, T Schüpbach, and S Y Shvartsman. Quantitative models of developmental pattern formation. *Developmental Cell*, 11(3):289–300, 2006.
- [212] D M Umulis and H G Othmer. The role of mathematical models in understanding pattern formation in developmental biology. *Bulletin of Mathematical Biology*, 77(5):817–845, 2015.
- [213] D M Umulis, O Shimmi, M B O’Connor, and H G Othmer. Organism-scale modeling of early *Drosophila* patterning via bone morphogenetic proteins. *Developmental Cell*, 18(2):260–274, 2010.
- [214] S C Little, G Tkačik, T B Kneeland, E F Wieschaus, and T Gregor. The formation of the Bicoid morphogen gradient requires protein movement from anteriorly localized mRNA. *PLOS Biology*, 9(3):e1000596, 2011.
- [215] W Liu and M Niranjan. The role of regulated mRNA stability in establishing Bicoid morphogen gradient in *Drosophila* embryonic development. *PLOS ONE*, 6(9):e24896–e24896, 2011.
- [216] L A Goentoro, G T Reeves, C P Kowal, L Martinelli, T Schüpbach, and

- S Y Shvartsman. Quantifying the Gurken morphogen gradient in *Drosophila* oogenesis. *Developmental Cell*, 11(2):263–272, 2006.
- [217] Z Szymańska, M Parisot, and M Lachowicz. Mathematical modeling of the intracellular protein dynamics: The importance of active transport along microtubules. *Journal of Theoretical Biology*, 363:118–128, 2014.
- [218] P C Bressloff and J M Newby. Stochastic models of intracellular transport. *Reviews of Modern Physics*, 85(1):135, 2013.
- [219] A E Hafner and H Rieger. Spatial cytoskeleton organization supports targeted intracellular transport. *Biophysical Journal*, 114(6):1420–1432, 2018.
- [220] H G Othmer, S R Dunbar, and W Alt. Models of dispersal in biological systems. *Journal of Mathematical Biology*, 26(3):263–298, 1988.
- [221] E A Codling and N A Hill. Calculating spatial statistics for velocity jump processes with experimentally observed reorientation parameters. *Journal of Mathematical Biology*, 51(5):527–556, 2005.
- [222] J P Taylor-King, E van Loon, G Rosser, and S J Chapman. From birds to bacteria: generalised velocity jump processes with resting states. *Bulletin of Mathematical Biology*, 77(7):1213–1236, 2015.
- [223] J Schindelin, I Arganda-Carreras, E Frise, V Kaynig, M Longair, T Pietzsch, S Preibisch, C Rueden, S Saalfeld, and B Schmid. Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9(7):676, 2012.
- [224] A Davidson. *The mechanism of transport and translational control of gurken mRNA in the Drosophila egg chamber*. PhD thesis, University of Oxford, 2015.
- [225] D Ando, N Korabel, K C Huang, and A Gopinathan. Cytoskeletal network morphology regulates intracellular transport dynamics. *Biophysical Journal*, 109(8):1574–1582, 2015.

- [226] A P Browning, P Haridas, and M J Simpson. A Bayesian sequential learning framework to parameterise continuum models of melanoma invasion into human skin. *Bulletin of Mathematical Biology*, 2018.
- [227] N G Gurskaya, V V Verkhusha, A S Shcheglov, D B Staroverov, T V Chepurnykh, A F Fradkov, S Lukyanov, and K A Lukyanov. Engineering of a monomeric green-to-red photoactivatable fluorescent protein induced by blue light. *Nature Biotechnology*, 24(4):461, 2006.
- [228] D M Chudakov, S Lukyanov, and K A Lukyanov. Tracking intracellular protein movements using photoswitchable fluorescent proteins PS-CFP2 and Dendra2. *Nature Protocols*, 2(8):2024, 2007.
- [229] V Trovisco, K Belaya, D Nashchekin, U Irion, G Sirinakis, R Butler, J J Lee, E R Gavis, and D St Johnston. bicoid mRNA localises to the *Drosophila* oocyte anterior by random Dynein-mediated transport and anchoring. *eLife*, 5, 2016.
- [230] H C Berg and D A Brown. Chemotaxis in *Escherichia coli* analysed by three-dimensional tracking. *Nature*, 239(5374):500–504, 1972.
- [231] G Rosser, A G Fletcher, D A Wilkinson, J A de Beyer, C A Yates, J P Armitage, P K Maini, and R E Baker. Novel methods for analysing bacterial tracks reveal persistence in *Rhodobacter sphaeroides*. *PLOS Computational Biology*, 9(10):1:18, 2013.
- [232] H G Othmer and T Hillen. The diffusion limit of transport equations derived from velocity-jump processes. *SIAM Journal on Applied Mathematics*, 61(3):751–775, 2000.
- [233] E A Codling, M J Plank, and S Benhamou. Random walk models in biology. *Journal of the Royal Society Interface*, 5(25):813–834, 2008.

- [234] P M Kareiva and N Shigesada. Analyzing insect movement as a correlated random walk. *Oecologia*, 56(2):234–238, 1983.
- [235] P Bovet and S Benhamou. Spatial analysis of animals’ movements using a correlated random walk model. *Journal of Theoretical Biology*, 131(4):419–433, 1988.
- [236] I D Jonsen, J M Flemming, and R A Myers. Robust state-space modeling of animal movement data. *Ecology*, 86(11):2874–2880, 2005.
- [237] J Liepe, H Taylor, C Barnes, M Huvet, L Bugeon, T Thorne, J Lamb, M Dallman, and M P H Stumpf. Calibrating spatio-temporal models of leukocyte dynamics against in vivo live-imaging data using approximate Bayesian computation. *Integrative Biology*, 4(3):335–345, 2012.
- [238] P Jones, A Sim, H Taylor, L Bugeon, M Dallman, B Pereira, M P H Stumpf, and J Liepe. Inference of random walk models to describe leukocyte migration. *Physical Biology*, 12(6):66001–66012, 2015.
- [239] G M Viswanathan, V Afanasyev, S V Buldyrev, E J Murphy, P A Prince, and H E Stanley. Lévy flight search patterns of wandering albatrosses. *Nature*, 381(6581):413–415, 1996.
- [240] G M Viswanathan, V Afanasyev, S V Buldyrev, S Havlin, M G Da Luz, E P Raposo, and H E Stanley. Lévy flights in random searches. *Physica A*, 282(1):1–12, 2000.
- [241] A M Edwards, R A Phillips, N W Watkins, M P Freeman, E J Murphy, V Afanasyev, S V Buldyrev, M G da Luz, E P Raposo, and H E Stanley. Revisiting Lévy flight search patterns of wandering albatrosses, bumblebees and deer. *Nature*, 449(7165):1044–1048, 2007.

- [242] S Benhamou. How many animals really do the Lévy walk? *Ecology*, 88(8): 1962–1969, 2007.
- [243] K J Painter. Modelling cell migration strategies in the extracellular matrix. *Journal of Mathematical Biology*, 58(4):511–543, 2009.
- [244] J T Mandeville, R N Ghosh, and F R Maxfield. Intracellular calcium levels correlate with speed and persistent forward motion in migrating neutrophils. *Biophysical Journal*, 68(4):1207–1217, 1995.
- [245] R Pankov, Y Endo, S Even-Ram, M Araki, K Clark, E Cukierman, K Matsumoto, and K M Yamada. A Rac switch regulates random versus directionally persistent cell migration. *Journal of Cell Biology*, 170(5):793–802, 2005.
- [246] A Nicosia, T Duchesne, L P Rivest, and D Fortin. A general hidden state random walk model for animal movement. *Computational Statistics and Data Analysis*, 105:76–95, 2017.
- [247] C Andrieu and G O Roberts. The pseudo-marginal approach for efficient Monte Carlo computations. *The Annals of Statistics*, 37(2):697–725, 2009.
- [248] G Grimmett and D Stirzaker. *Probability and Random Processes*. Oxford University Press, 2001.
- [249] G Rosser. *Mathematical modelling and analysis of aspects of planktonic bacterial motility*. PhD thesis, University of Oxford, 2012.
- [250] C Sherlock, A H Thiery, G O Roberts, and J S Rosenthal. On the efficiency of pseudo-marginal random walk Metropolis algorithms. *The Annals of Statistics*, 43(1):238–275, 2015.
- [251] M K Pitt, R dos Santos Silva, P Giordani, and R Kohn. On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. *Journal of Econometrics*, 171(2):134–151, 2012.

- [252] E Bernton, P E Jacob, M Gerber, and C P Robert. Inference in generative models using the Wasserstein distance. *arXiv preprint arXiv:1701.05146*, 2017.
- [253] Q Zhao, I T Young, and J G De Jong. Photon budget analysis for fluorescence lifetime imaging microscopy. *Journal of Biomedical Optics*, 16(8):086007–086007, 2011.
- [254] H B Taylor, J Liepe, C Barthen, L Bugeon, M Huvet, P W Kirk, S B Brown, J R Lamb, M P H Stumpf, and M Dallman. P38 and JNK have opposing effects on persistence of *in vivo* leukocyte migration in zebrafish. *Immunology and Cell Biology*, 91(1):60–69, 2013.
- [255] H Weavers, J Liepe, A Sim, W Wood, P Martin, and M P H Stumpf. Systems analysis of the dynamic inflammatory response to tissue damage reveals spatiotemporal properties of the wound attractant gradient. *Current Biology*, 26(15):1975–1989, 2016.
- [256] S L Scott, A W Blocker, F V Bonassi, H A Chipman, E I George, and R E McCulloch. Bayes and big data: The consensus Monte Carlo algorithm. *International Journal of Management Science and Engineering Management*, 11(2):78–88, 2016.
- [257] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL <http://www.R-project.org/>.
- [258] A Miroshnikov and E M Conlon. ParallelMCMCcombine: an R package for Bayesian methods for big data and analytics. *PLOS One*, 9(9):e108425, 2014.
- [259] M Kumar, M S Mommer, and V Sourjik. Mobility of cytoplasmic, membrane, and DNA-binding proteins in *Escherichia coli*. *Biophysical Journal*, 98(4):552–559, 2010.

- [260] D J Wilkinson. Stochastic modelling for quantitative description of heterogeneous biological systems. *Nature Reviews Genetics*, 10(2):122–133, 2009.
- [261] B M Turner and T Van Zandt. A tutorial on approximate Bayesian computation. *Journal of Mathematical Psychology*, 56(2):69–85, 2012.
- [262] M A Nunes and D J Balding. On optimal selection of summary statistics for approximate Bayesian computation. *Statistical Applications in Genetics and Molecular Biology*, 9(1):1–16, 2010.
- [263] H Jung and P Marjoram. Choice of summary statistic weights in approximate Bayesian computation. *Statistical Applications in Genetics and Molecular Biology*, 10(1):1–23, 2011.
- [264] P Singh and A Hellander. Multi-statistic approximate Bayesian computation with multi-armed bandits. *arXiv preprint arXiv:1805.08647*, 2018.
- [265] M U Gutmann, R Dutta, S Kaski, and J Corander. Likelihood-free inference via classification. *Statistics and Computing*, 28(2):411–425, 2018.
- [266] T McKinley, A R Cook, and R Deardon. Inference in epidemic models without likelihoods. *The International Journal of Biostatistics*, 5(1):1–40, 2009.
- [267] B Póczos and J Schneider. On the estimation of alpha-divergences. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 609–617, 2011.
- [268] B Póczos, L Xiong, D J Sutherland, and J Schneider. Nonparametric kernel estimators for image classification. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2989–2996. IEEE, 2012.
- [269] D J Sutherland, L Xiong, B Póczos, and J Schneider. Kernels on sample sets via nonparametric divergence estimates. *arXiv preprint arXiv:1202.0302*, 2012.

- [270] MATLAB. *version 9.1.0.441655 (R2016b)*. The MathWorks Inc., Natick, Massachusetts, 2016.
- [271] D T Gillespie. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 81(25):2340–2361, 1977.
- [272] M A Nunes and D Prangle. abctools: an R package for tuning approximate Bayesian computation analyses. *The R Journal*, 7(2):189–205, 2015.
- [273] J Bohrmann and K Biber. Cytoskeleton-dependent transport of cytoplasmic particles in previtellogenic to mid-vitellogenic ovarian follicles of drosophila: time-lapse analysis using video-enhanced contrast microscopy. *Journal of Cell Science*, 107(4):849–858, 1994.
- [274] S C Little, K S Sinsimer, J J Lee, E F Wieschaus, and E R Gavis. Independent and coordinate trafficking of single drosophila germ plasm mRNAs. *Nature Cell Biology*, 17(5):558, 2015.
- [275] R J Hawkins, O Benichou, M Piel, and R Voituriez. Rebuilding cytoskeleton roads: Active-transport-induced polarization of cells. *Physical Review E*, 80(4):040903, 2009.
- [276] P C Bressloff and B Xu. Stochastic active-transport model of cell polarization. *SIAM Journal on Applied Mathematics*, 75(2):652–678, 2015.
- [277] J I Alsous, P Villoutreix, A M Berezhkovskii, and S Y Shvartsman. Collective growth in a small cell network. *Current Biology*, 27(17):2670–2676.e4, 09 2017.
- [278] W Tadros and H D Lipshitz. Setting the stage for development: mRNA translation and stability during oocyte maturation and egg activation in *Drosophila*. *Developmental Dynamics*, 232(3):593–608, 2005.
- [279] R C King and R G Burnett. Autoradiographic study of uptake of tritiated

glycine, thymidine, and uridine by fruit fly ovaries. *Science*, 129(3364):1674–1675, 1959.

- [280] B Schäling. *The boost C++ libraries*. Boris Schäling, 2011.
- [281] A Gelman and D B Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472, 1992.
- [282] C Allan, J M Burel, J Moore, C Blackburn, M Linkert, S Loynton, D MacDonald, W J Moore, C Neves, and A Patterson. OMERO: flexible, model-driven data management for experimental biology. *Nature Methods*, 9(3):245, 2012.
- [283] M Linkert, C T Rueden, C Allan, J M Burel, W Moore, A Patterson, B Lorange, J Moore, C Neves, and D MacDonald. Metadata matters: access to image data in the real world. *The Journal of Cell Biology*, 189(5):777–782, 2010.
- [284] M Hailstone, L Yang, D Waithe, T J Samuels, Y Arava, T Dobrzycki, R M Parton, and I Davis. Brain development: machine learning analysis of individual stem cells in live 3D tissue. *bioRxiv*, 2017.
- [285] L Yang, J Titlow, D Ennis, C Smith, J Mitchell, F L Young, S Waddell, D Ish-Horowicz, and I Davis. Single molecule fluorescence in situ hybridisation for quantitating post-transcriptional regulation in *Drosophila* brains. *Methods*, 126:166–176, 2017.
- [286] D Jia, Q Xu, Q Xie, W Mio, and W M Deng. Automatic stage identification of *Drosophila* egg chamber based on DAPI images. *Scientific Reports*, 6:18850, 2016.
- [287] S C Little, M Tikhonov, and T Gregor. Precise developmental gene expression arises from globally stochastic transcriptional activity. *Cell*, 154(4):789–800, 2013.

- [288] Z Ghahramani. Probabilistic machine learning and artificial intelligence. *Nature*, 521(7553):452, 2015.
- [289] D J Gavaghan. Problems with the current approach to the dissemination of computational science research and its implications for research integrity. *Bulletin of Mathematical Biology*, pages 1–7, 2018.
- [290] S Schnell. “Reproducible” research in mathematical sciences requires changes in our peer review culture and modernization of our current publication approach. *Bulletin of Mathematical Biology*, 80(12):3095–3105, 2018.
- [291] C Bökel, S Dass, M Wilsch-Bräuninger, and S Roth. *Drosophila* Cornichon acts as cargo receptor for ER export of the TGF α -like growth factor Gurken. *Development*, 133(3):459–470, 2006.
- [292] K A Edwards, M Demsky, R A Montague, N Weymouth, and D P Kiehart. GFP-Moesin illuminates actin cytoskeleton dynamics in living tissue and demonstrates cell shape changes during morphogenesis in *Drosophila*. *Developmental Biology*, 191(1):103–117, 1997.
- [293] R M Parton, A M Vallés, I M Dobbie, and I Davis. Pushing the limits of live cell imaging in *Drosophila*. *Live cell imaging: a laboratory manual*, pages 387–418, 2010.
- [294] E M Stein and R Shakarchi. *Real Analysis: Measure Theory, Integration, and Hilbert Spaces*. Princeton University Press, 2009.
- [295] A Davidson, R M Parton, C Rabouille, T T Weil, and I Davis. Localized translation of gurken/TGF- α mRNA during axis specification is controlled by access to Orb/CPEB on processing bodies. *Cell Reports*, 14(10):2451–2462, 2016.