Imperial College London
Department of Life Sciences

Probabilistic modelling of noise as a driving force in biological systems

Robert Andrew Johnson

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy of Imperial College London 21st September 2016

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

Abstract

Systems biology takes a mechanistic, relational approach to the study of biological processes, commonly finding expression in mathematical models. Hypotheses about systems can be tested when formulated as models, and promising avenues for further study identified. A model sufficiently faithful to the system under study can be used to guide experiments, to probe the system *in silico*, and to learn about emergent features not evident from the static picture of the system.

In this work, three contributions to the modelling community are proffered. First, a computational package is presented that implements an algorithm for the validation and parametrisation of a model. In validation, we are asking how likely we were to make some observation, given the model, or, equivalently, how able the model is to explain the data.

The subsequent two contributions concern noise in biological systems. Biological systems display inherent variability, or noise, due to the stochastic mechanisms through which biochemical processes occur. This variability can be critical to the behaviour of a system and to the fates of individual cells.

With this in mind, the second contribution is the development of a methodology to model protein-dependent population dynamics. The idea is to model cell population dynamics that result of noisy intracellular protein dynamics. The method's application is demonstrated in population-level models of a protein-dependent cell cycle and yeast antibiotic resistance.

Given an appreciation of the pivotal effects of noise, the third and final contribution is a study of the mechanism of noise propagation. I present an analysis of the contributions of biochemical reaction motifs to the creation and transmission of noise that ultimately manifest in observations of biological systems. This study points to specific processes that enhance or attenuate noise, with the aim of beginning to unravel the flow of noise through a system.

Acknowledgements

I offer my thanks and gratitude first to my supervisor, Michael Stumpf. He welcomed me into the Theoretical Systems Biology group, whose atmosphere reflects him as architect: it has been an enriching, challenging, fun environment in which to work and develop as a researcher. Michael has guided me from being a Master's student through to this point of documenting my work as a research student, and beyond to the beginnings of a scientific career.

I thank Paul Kirk for his supervision of the work in Chapter 3, but I am most grateful to him for his support and guidance, for his friendship, for laughs and silliness, and for all the cups of tea.

I thank Brian Munsky, for the collaboration and encouragement that resulted in Chapters 4 to 6, and the q-bio summer school, in which fruitful and stimulating environment the project began.

I am grateful to the Theoretical Systems Biology group: all who pass through contribute something to the unique and supportive community. I am particularly grateful to Virginia Fairclough for her companionship and her practical help in asking questions and solving problems.

I am grateful also to my friends, my family and all at FSSE and the WLM, who have upheld me throughout, and sensitively asked me/not asked me how things are going. Thank you dad and Auntie Carol, for all the years of proofreading. Thank you Martin, for steadfastness, sustenance, and unwavering belief in me.

Thank you, finally, to my mother, who has been parent, friend, challenger, anchor, helm, all.

Contents

List of Tables

List of Figures

1

Introduction

When taking a systems approach to biology, we commonly summarise our current understanding of or hypothesis about a biological system with a model (Kitano, 2002). This is a mathematical abstraction that aims to capture characteristic behaviours of the system under study. The choice of model requires fine balancing between realism and manageability, aiming to hit the sweet spot of precisely sufficient detail that the desired behaviour is exhibited (Kirk et al., 2013).

In this work, three contributions are made to the modelling community. First, a computational package is presented that aids in validating and parametrising models. Second, a method is presented for modelling population dynamics driven by molecular dynamics. Finally, a method is presented for analysing motifs that constitute models in order to determine how a reaction network's structure gives rise to noise.

1.1 Using models

When we have a model, we routinely wish to a) validate it, using experimental data, i.e. ask the question 'how well does our model explain the data?' (Kirk et al., 2013), b) inform it, using experimental data, i.e. learn about the parameters that determine the system (Toni et al., 2009), and c) exploit it to learn more about the structure and behaviours of the system (Silk et al., 2014).

To aid in answering questions (a) and (b), an accelerated computational package is first presented that implements nested sampling, a Bayesian model selection and parameter inference algorithm (Skilling, 2006).

1.2 Modelling noisy systems

Of particular interest to the present work is the modelling of noisy systems. Biological processes are inherently random (or noisy), resulting in variability of phenotype that may be seen over time or across a population of cells (Elowitz et al., 2002). This heterogeneity may ultimately result in different cell fates within a clonal cell population, for example (MacArthur et al., 2009; Balázsi et al., 2011).

Many modelling methods have been presented in the literature that aim to capture noisy dynamics (Gillespie, 2007; Munsky & Khammash, 2006; Ale et al., 2013; van Kampen, 1981). The distributions that result from these models correspond to what the models predict would be seen in a survey of a number of cells, or over a time period (Nevozhay et al., 2012).

A particular challenge is to model a population in which a cell divides at a rate depending on the value of a noisy variable. The most faithful way of doing so is to use the method of stochastic simulation, in which very many simulations of a process are generated (Gillespie, 2007). However, the computation required to generate a full distribution over both cell and molecule number can be prohibitively expensive (Wilkinson, 2009). Therefore, a new development of the finite state projection (Munsky & Khammash, 2006) is presented that can model protein-dependent cell population dynamics without resorting to exhaustive simulations.

1.3 Understanding noise

The extent of noise present in a biological system is pivotal to cell population dynamics, as well as to many ubiquitous cellular processes, such as the faithfulness with which a system may respond to a signal (Raser & O'Shea, 2005). It might be assumed, therefore, that variability is itself a regulated process, and, in order to fully understand the system, it becomes pertinent to understand how the variability arises (Paulsson, 2004).

To address this question, I use the methodology of Toni & Tidor (2013) as a basis to unpick the sources of noise and present some new findings into the propagation of noise in dynamical systems. I identify the interactions that amplify or attenuate variability in certain conditions, shedding light on the mechanistic reason for cell-to-cell variability and suggesting principles for designs of synthetic gene circuits.

1.4 Overview

In Chapter 2, concepts, methods and terminology relevant to this work are presented. In Section 2.1, the concept of modelling is introduced, followed by some commonly used modelling methods. In Section 2.2, model selection and parameter inference are explained, and some methods for these processes are surveyed. In Section 2.3, the notion of "noise" in a biological context is discussed. Two types of noise are presented: intrinsic noise, as that which arises inherently as a product of a system; and extrinsic noise, as that which is imposed from the outside (Elowitz et al., 2002). The modelling methods introduced in Section 2.1 are revisited and assessed for their ability to capture the different types of noise.

In Chapter 3, I present a computational tool I have developed for the biological community to perform model selection and parameter inference using an accelerated version of the nested sampling method of Skilling (2006). Its performance is demonstrated through application to two questions posed in the literature and assessed via a comparison with some other methods.

In Chapter 4, I present a new method for modelling a population of cells in which the individual cells have different rates of division, the rate of division depending on the level of some defined protein. The methodology is applied to two biological systems: to models of cell division in Chapter 5 (Spencer et al., 2013; Overton et al., 2014) and to a model of antibiotic resistance in Chapter 6 (Nevozhay et al., 2012).

In Chapter 7, I present a study into the decomposition of noise, demon-

strating through some simple examples how different reaction mechanisms result in different levels of cell-to-cell variability.

Finally, in Chapter 8, a summary and discussion of the work presented is given.

2

Theoretical background

In this chapter, material and literature relevant to the results chapters to follow will be introduced. To begin, I will give an overview of modelling methods, parameter inference and model selection in systems biology. These overviews will highlight the modelling methods that will be used in later chapters, and the parameter inference and model selection method that will be implemented in Chapter 3.

Thereafter, the terminology surrounding noise will be discussed in order to clarify the scope of and the assumptions made in Chapter 7, in which a method is presented for modelling intrinsic and extrinsic noise arising in biochemical processes.

2.1 Modelling methods

A model is a mathematical abstraction of a system. It is the formulation, in mathematical terms, of a hypothesis about the relationships between variables that compose the system. In a systems-biology setting, these variables might include cells, organelles, biomolecules such as DNA and proteins, time, and space. A useful model is one that accurately captures some hypothesis, and permits critique of the hypothesis through comparison with an observation of a realisation of the system (Kirk et al., 2015). A good model is one that is found to accurately capture or explain some behaviour of a biological system, and can be used to make predictions of future observations, or identify experiments that will be informative (Silk et al., 2014).

Various methods are used routinely in the systems biology community to realise models, such as stochastic simulations (Gillespie, 2007), the finite state projection (Munsky & Khammash, 2006), moment-expansion approximations (Ale et al., 2013), and the linear noise approximation (van Kampen, 1981). These are introduced in Sections 2.1.2–2.1.5, with some discussion of the scope of their application and their resultant strengths and limitations.

With a model, we describe the state of a system, given the reactions that change it and the rate at which these reactions occur (Szallasi et al., 2010). The state, $\mathbf{x}(t)$, of the system at time t typically refers to the number or concentration of one or more species. The possible reactions that occur that change the state of the system define the stoichiometry of the system,

S. S is a matrix of dimension (n_S, n), where n_S is the number of species (the dimension of $\mathbf{x}(t)$) and n is the number of reactions. The system is parametrised by the reaction rate vector, $a(\mathbf{x}(t))$.

Species evolve over time according to the equation

$$\mathbf{x}(t + dt) = \mathbf{x}(t) + S \cdot \mathcal{F}(a(\mathbf{x}(t)), dt), \tag{2.1}$$

where $\mathcal{F}(a(\mathbf{x}(t)), dt)$ is the update rule that depends on the reaction rates and the time step (Gillespie, 2007). If we take a view that a reaction event fires randomly according to its propensity to occur, then the state $\mathbf{x}(t)$ is not a deterministic function of t, S and $a(\mathbf{x}(t))$, but a probability distribution over a number of possible states (Munsky & Khammash, 2006). Different considerations of this principle have given rise to a number of different methods that aim to learn something about the state $\mathbf{x} = \mathbf{x}(t)$ given t, $a(\mathbf{x})$ and S. Some such methods are discussed below, with those employed later in the document highlighted.

2.1.1 The chemical master equation

A biological system, seen as a chemical reaction network, can be described by the chemical master equation (CME), which gives an exact expression for the evolution of the system's state over time (Gillespie, 1992). The description assumes fixed volume and temperature, and that the components of the system freely mix.

From Equation 2.1, with

$$\mathcal{F}(a_j(\mathbf{x}(t)), dt) = \begin{cases} 1 & \text{if reaction } j \text{ occurs in time } [t, t+dt) \\ 0 & \text{otherwise,} \end{cases} \tag{2.2}$$

the probability to be in state $\mathbf{x}$ at time $t + dt$ is defined

$$p(\mathbf{x}; t+dt) = p(\mathbf{x}; t)\left(1 - \sum_{j=1}^{n} a_j(\mathbf{x})dt\right) + \sum_{j=1}^{n} p(\mathbf{x} - S_j; t)a_j(\mathbf{x} - S_j)dt, \tag{2.3}$$

leading to the definition of the CME as (Gillespie, 1992)

$$\frac{p(\mathbf{x}; t)}{dt} = -p(\mathbf{x}; t)\sum_{j=1}^{n} a_j(\mathbf{x}) + \sum_{j=1}^{n} p(\mathbf{x} - S_j; t)a_j(\mathbf{x} - S_j). \tag{2.4}$$

For some examples, the CME can be solved exactly to give the probability distribution over all possible states $\mathbf{x}$ at time t:

$$p(\mathbf{x}; t) = p(\mathbf{x}; 0)\exp(At), \tag{2.5}$$

where A is a time-independent transition matrix enumerating the probability for each state to transition to every other state. However, for any system with non-finite state space, the CME is not directly solvable (Jahnke, 2011), and an alternative method is required.

2.1.2 Stochastic simulation

By simulating the system defined in Equations 2.1 and 2.2, one is effectively sampling from the distribution $p(\mathbf{x}; t)$. An efficient means to do so was presented by Gillespie, referred to as the stochastic simulation algorithm (SSA) (Gillespie, 1992). Many simulations are required to generate a good estimate of the distribution and, though a number of developments to

Gillespie's original algorithm have been presented that accelerate the computation (Gillespie, 2007; Gillespie et al., 2013), this expense remains in general prohibitively high, with no guarantee that the full distribution has been sampled. These shortcomings motivate the analytic approximations outlined below.

2.1.3 The finite state projection

The finite state projection (FSP) is an approximation to the CME (Munsky & Khammash, 2006). It functions by artificially truncating the state space, rendering Equation 2.4 solvable. Further, the loss of probability mass that results from the truncation is measured. Unlike the SSA, then, the FSP gives a direct read-out of the extent of divergence of the approximate distribution from the true distribution.

The FSP forms the basis for the method developed in Chapter 4, where it is fully expounded, and developed for application to a growing population of cells in which the protein content determines the rate of division. The SSA is used as a benchmark for its accuracy in Chapters 4–6.

2.1.4 Moment-expansion approximations

A probability distribution can be described by its sequence of moments, where the first moment is the mean, the second is the variance, the third the skewness and so on. Expressions for moments for components of a chemically interacting system can be generated directly from Equations 2.1

and 2.2 (Ale et al., 2013). Each moment is expressed as a sum including both lower- and higher-order moments.

For the majority of cases (in which there is no reason to expect that moments beyond a certain order are equal to zero), it is common practice to terminate the sum at some pre-determined point, either neglecting or approximating the higher-order terms. At present, there is no way of knowing *a priori* for a given system and a given number of moments what the divergence will be between the resulting approximation and the true distribution. There are thus two key considerations in the implementation of this method: one is the choice of number of moments to calculate; the other is how to approximate the moments that are not calculated (Lakatos et al., 2015). These can affect substantially the fidelity of the approximation to the true distribution of the model.

2.1.5 The linear noise approximation

The linear noise approximation (LNA) is an approximation to the CME (Equations 2.1 and 2.2) that provides the first two moments via a linearisation of the system. It derives from the system-size expansion, in which the dynamics of a system are separated according to the size over which they operate; namely, it employs an ansatz based on fluctuations scaling as the square root of the system size (van Kampen, 1981).

Its validity and its applicability have been widely discussed (Thomas et al., 2013; Grima, 2015). It is exact for linear systems (those that consist of at highest first-order reactions), up to the first two moments. For nonlin-

ear systems, the solution converges to the true first two moments in the limit of large molecule numbers; at other scales, its validity depends on the specifics of the system (Grima, 2015). Its applicability is limited compared to the moment-expansion approximation, which lends itself more naturally to multimodality and nonlinear effects (Ale et al., 2013). Multimodality in a biomolecule (i.e. the existence of more than one mode in the dynamic system) leads to phenotypic heterogeneity. The LNA cannot capture more than one mode, and therefore no behaviour that rests on multimodality can be described by the LNA.

The LNA is combined with the unscented transform (Wan & Van Der Merwe, 2000) in Chapter 7, where both methods are fully expounded along the lines of Toni & Tidor (2013). This machinery is then applied to archetypal motifs in order to understand how noise arises and is transmitted through a biochemical pathway.

2.2 Parameter inference and model selection

The particular choice of modelling method will depend on many criteria, including the size of the system, time and computation concerns, and expectations around the resultant dynamics (Wilkinson, 2009). The construction of a particular model, i.e. how a system is to be concretely represented in a mathematical abstraction, involves choices around which species and reactions to include (Kirk et al., 2015), and how reactions are realised mathematically. Each realisation corresponds to a different model of the system (Vyshemirsky & Girolami, 2008a).

When different models are proposed for a single system (Vyshemirsky & Girolami, 2008a), it is pertinent to ask if one model is 'better' than another – better able to explain data representative of the behaviour we are trying to model. Additionally, we are concerned that the data are not overfit, i.e. that the model is not so complex that it is has the capacity to find a parametrisation to fit any seen data (Kirk et al., 2013). Such models capture the observed data well, but fail to model unseen data due to the entrainment of the parameters to the seen data. The task of choosing between models is known as model selection (see Section 2.2.2).

We are simultaneously interested in inferring the posterior distributions of parameters that give rise to the behaviour observed, a task known as parameter inference (see Section 2.2.1) (Toni et al., 2009). Both these tasks require data with which to compare the outcome predicted from the model, and the aim is to find models, or parametrisations of models, that minimise the difference between the data and the prediction.

In Chapter 3, an existing algorithm for model selection and parameter inference, nested sampling (Skilling, 2006), will be developed and implemented as software package.

2.2.1 Parameter inference

Parameter inference denotes the task of learning the parametrisation of a model given some data. Bayesian methods for parameter inference rely on a guided exploration of the parameter space in order to construct a picture of how a model's prediction deviates from its target across the parameter

space (Toni et al., 2009). In the Bayesian formalism, this picture is known as the posterior parameter distribution, commonly referred to as the posterior. The posterior of a model indicates likely values taken by parameters, the extent of variability that preserves the output, and putative relationships betweens parameters (Gutenkunst et al., 2007).

Bayesian methods for parameter inference include Markov chain Monte Carlo (MCMC) and sequential Monte Carlo (SMC) methods. The Metropolis–Hastings algorithm, an implementation of MCMC, samples the parameter space sequentially by generating a new point from the last. The acceptance criteria for the points result in eventual convergence to the desired distribution (Hastings, 1970; Metropolis et al., 1953).

SMC methods sample from a series of related distributions that evolve towards the desired distribution via importance sampling. Each distribution is approximated by a weighted population of points from which the next generation is sampled (Del Moral et al., 2006; Toni et al., 2009).

Nested sampling (Skilling, 2006) generates samples from the posterior in the course of validating the model itself. This method therefore combines the processes of parameter inference and model selection.

2.2.2 Model selection

Numerous methods and criteria have been proposed to choose between models quantitatively and consistently according to some data. These methods have been reviewed in Calderhead & Girolami (2009); Kirk et al. (2013);

Vyshemirsky & Girolami (2008a). Model validation is primarily used in justifying a choice to retain or reject one model when compared to another. More interestingly, it can be used to demonstrate uncertainty in the choice of model with respect to the data (Babtie et al., 2014).

The key considerations in model validation are (i) the fit of the model to the data and (ii) the complexity of the model, where the ideal solution maximises the former and minimises the latter. There are two broad classes of methods: Bayesian approaches, which have a framework that automatically accounts for both; and likelihood-based approaches, which account for model complexity latterly. Some examples of the methods are outlined below.

Likelihood-based methods

Likelihood-based methods define a likelihood function, typically a function of the distance between the model prediction and the data, and identify the parameter set that minimises the distance (or, equivalently, maximises the likelihood). These maximum likelihoods are the means by which models are compared, and the methods rely on being able to identify them by navigating the (*a priori* unknown) likelihood surface (Burnham & Anderson, 2010; Cover & Thomas, 2012; Grueber et al., 2011).

Bayesian methods

Bayesian model selection methods, grounded in the formalism of assigning uncertainty to all unknown quantities, avoid the reliance on the identifica-

tion of maximum-likelihood parameters. The question of model selection from this viewpoint has been addressed in two ways: via model indicators and via the evidence.

Model indicators To choose between a number of models, a model indicator can be introduced, and is inferred in the same way as other parameters. The posterior distribution over this indicator assigns to each model some probability, quantifying how models are preferred relative to one another. The algorithm must be re-run each time new models are to be compared. Such a method was applied in Carlin & Chib (1995), where an MCMC algorithm was extended to allow for transitions between models. The approximate Bayesian computation (ABC) package of Liepe et al. (2010) similarly performs both model selection and parameter inference, in this case using sequential Monte Carlo (Toni et al., 2009).

The evidence In the Bayesian framework, we have a formulation for independently quantifying the support afforded to a model by some data. This score we call the evidence, Z, and is also known as the integrated or marginal likelihood. It appears as the normalising factor in Bayes' theorem,

$$p(\theta|D, M) = \frac{\pi(\theta|M)\ell(\theta|M)}{Z}, \tag{2.6}$$

where we define a prior distribution π over the parameter space $\Theta \ni \theta$ and a likelihood function ℓ that depends on the parameters and the data D given a model M.

It is ratios of evidences, or Bayes factors (Jeffreys, 1961), that indicate relative support for one model over another. For two models, M_1 and M_2,

the Bayes factor $K_{1,2}$ is

$$K_{1,2} = \frac{Z_1}{Z_2}, \tag{2.7}$$

where $Z_i = p(D|M_i)$ is the marginal likelihood associated with observing the data given the model M_i. If $K_{1,2} \approx 1$, neither model is clearly better. A very large $K_{1,2}$ indicates that M_1 is much preferred to M_2, with the magnitude of $K_{1,2}$ indicating the strength of the support (Jeffreys, 1961).

The evidence is the integrated likelihood over all prior space. More formally,

$$Z = \int_\Theta \ell(\theta|M)\pi(\theta|M)\,\mathrm{d}\theta. \tag{2.8}$$

The integral is, in general, analytically intractable. However, a number of methods has been proposed to estimate the evidence, thus allowing direct comparisons between models.

Estimating the evidence Methods to estimate the evidence generally sample from the parameter space and construct the evidence according how the sample was taken. Using Bayes theorem directly, for any θ,

$$Z = \frac{\pi(\theta|M)\ell(\theta|M)}{p(\theta|D,M)}. \tag{2.9}$$

Therefore, if $p(\theta|D,M)$ can be evaluated (e.g. by MCMC), Z can be estimated by calculating the likelihood and prior for some θ (Chib, 1995; Chib & Jeliazkov, 2001).

Defining it as the integrated likelihood (Equation 2.8), the evidence can be approximated through a weighted sum of points sampled from the parameter space. This class of methods makes use of importance weights that

correct for the discrepancy between the target distribution and the proposal distribution from which samples are taken (Friel & Wyse, 2012).

Importance sampling estimators include the prior arithmetic mean estimator (McCulloch & Rossi, 1992) and the posterior harmonic mean estimator (Newton & Raftery, 1994), which use the prior and from the posterior, respectively, as the proposal distribution. Annealed importance sampling finds a path from the prior to the posterior by evolving a point through a Markov chain transition kernel. Many iterations yield many samples from the posterior (Neal, 2001). A related approach is taken by Calderhead & Girolami (2009), in which population MCMC is used to generate multiple samples across all 'temperatures' whose thermodynamic integrals go from the integrated prior to the evidence.

Nested sampling Nested sampling is a method for estimating a model's evidence, while simultaneously inferring the posterior parameter distribution (Skilling, 2006). It is similar to importance sampling estimators in that each point sampled contributes to the evidence with a weighting dependent upon the sampling procedure. It has the advantage of employing more fine-grained sampling in parameter regions with higher likelihood, yielding more stable, accurate results.

Nested sampling will be the focus of Chapter 3, where it will be fully expounded and developed, and its implementation in a software package will be presented.

2.3 Variability in biological systems

Biological systems are inherently noisy (Elowitz et al., 2002). Molecules move, collide and interact unpredictably, resulting in random occurrences of reactions and random numbers of biochemical species within a cell. Such fluctuations lead to variability in cells, which might be seen across a population of genetically identical cells, or in a single cell viewed over time. Such noisiness ultimately has consequences for the fate of a cell (Rao et al., 2002; Wilkinson, 2009; Ruess et al., 2013).

A useful model of a biological system must have a meaningful grasp of the noise in the dynamics, or the effects of noise in the dynamics, for any system in which variability has a role. Each of the modelling methods outlined in Section 2.1 captures variability that arises from within the system: the chemical master equation, stochastic simulations and the finite state projection all have within their remit the full distribution over reaction species; the moment-expansion approximation includes as many moments as is desired and feasible; and the linear noise approximation gives the first two moments: the mean and the variance.

Variability that arises from within a system, due to the inherently random nature of biochemical processes, is termed *intrinsic noise*. Conversely, noise that arises from outside the system is termed *extrinsic noise* (Elowitz et al., 2002). It is thought that these two classes of noise have different mechanisms and, perhaps, consequences (Mc Mahon et al., 2015; Filippi et al., 2016).

The canonical illustration of the difference between intrinsic noise and ex-

trinsic noise is the dual-reporter experiment of Elowitz et al. (2002). In this study, two reporter genes are embedded within a plasmid and are constructed so as to have identical expression patterns. One expresses a cyan fluorescent protein, the other a yellow fluorescent protein. Both protein contents in a sample of cells are measured: the covariance between the two is indicative of variability in the shared environment (extrinsic noise); any additional variance is due to independent fluctuations in the expression of each gene (intrinsic noise). An illustration of this experiment is given in Figure D.1, panels A and B.

The appraisal given in Hilfinger & Paulsson (2011) on the uses of the terms intrinsic and extrinsic in discussion, modelling and interpreting data is extremely helpful in understanding the literature concerning extrinsic noise. Importantly for this discussion and the following work, a clear definition of extrinsic noise is presented: whatever is external to the system is extrinsic, and whatever is internal is intrinsic, which depends on the perimeter of the 'system' (Hilfinger & Paulsson, 2011; Scott et al., 2006). It is not necessarily the case that the perimeter drawn in a model will be biologically meaningful, or realisable in an experimental setting.

In the following sections, we discuss how intrinsic and extrinsic noise have been studied and presented in the literature. Much of the literature focuses on how to interpret intrinsic noise and extrinsic noise from experiments. Here, I focus on studies that use models to understand the effects of noise, and studies of how noise is transmitted in models. Chapters 4–6 add to the former, in which the combined effect of intrinsic and extrinsic noise is

behind the modelling method and results presented. Chapter 7 adds to the latter, in which a mechanistic study of the propagation of intrinsic and extrinsic noise is presented.

2.3.1 Intrinsic noise

Analytic studies of intrinsic noise have aimed to characterise the properties and functions of stochasticity in biological networks. These studies often assume a Gaussian model of noise, considering only the first two moments in their analyses.

Studies have queried how noise is propagated through biological systems, typically considering representative motifs. Komorowski et al. (2013) sought to identify how much reaction mechanisms in network motifs contributed to output noise through a decomposition of the linear noise approximation (LNA), concluding that the mechanism of decay of the molecule of interest was the greatest contributor to noise. Ozbudak et al. (2002) investigated the contributions of the rates of transcription and translation to phenotypic noise, concluding that transcriptional bursting, in which transcription occurs in pulses with multiple copies of mRNA being created in each pulse, drives the heterogeneity observed. In their two-moment model, Pedraza & Paulsson (2008) identified gestation and senescence as noise-attenuating mechanisms.

A different approach was taken by Tănase-Nicola et al. (2006), who used the power spectra resulting from the LNA to analyse how network motifs incorporate noise in external signals: whether the noise correlates, or whether

the components operate in a modular fashion and could therefore be modelled separately. A similar approach was taken by Tkačik et al. (2008) to quantify the effects of noise in transcription factor dynamics on noise in the activated protein, highlighting how upstream noise will manifest in a measured output.

2.3.2 Extrinsic noise

Studies of extrinsic noise comprise those that aim to infer it (Hilfinger & Paulsson, 2011; Hilfinger et al., 2012; Schwabe & Bruggeman, 2014; Pedraza & van Oudenaarden, 2005; Bowsher & Swain, 2012) and those that aim to model it, with the ultimate aim of combining the two. It is in the combination that the definition of extrinsic noise becomes critical (Hilfinger & Paulsson, 2011).

Modelling studies typically express extrinsic noise mathematically as variability in reaction rates. This variability is purported to correspond to varied availability of some machinery, for example ribosomes, in different cells, which aligns with interpretations of dual-reporter experiments in which extrinsic noise corresponds to any factor that is shared by the measured quantities (Elowitz et al., 2002).

Effects of noise

Stochastic simulations have been used both to study the effects of extrinsic noise on biological processes and to explore which of its aspects are most pivotal to the output. It has been shown through simulations and experi-

mentation that extrinsic noise has the capacity to degrade the information-processing capacity of biochemical signalling networks, but viewing the molecule as a time-separated multivariate response can recover it (Selimkhanov et al., 2014). In showcasing the great power and flexibility in biological systems, this study highlights the huge loss in explanatory power that is possible when these systems are oversimplified in models.

Stochastic simulation is clearly the most flexible and generalisable of the methods, but the computational expense is great (Wilkinson, 2009). For this reason, approximative methods are popularly used to study extrinsic noise. For example, two moments alone were used to quantify the robustness of oscillatory systems to extrinsic noise (Scott et al., 2006). Here, the LNA was used with one parameter perturbed at a time.

In Chapter 4, a multi-level modelling method that combines protein dynamics and population dynamics will be presented. This method is based on the finite state projection and is validated using stochastic simulations. In this formulation, the protein dynamics serve as fluctuating extrinsic noise for the population-level dynamics, and some emergent behaviours of systems modelled with this method are given in Chapters 5 and 6.

Drivers of noise

In terms of the mechanisms through which extrinsic noise acts, stochastic simulations have been used to study the effects of various aspects of extrinsic noise on variability in an output. Examples include noise in inheritance (Schwabe & Bruggeman, 2014), and the effects of nuances in extrinsic fluc-

tuations, including their relative timescales, and patterns of interference as constructive or destructive, depending on the reaction network architecture (Shahrezaei et al., 2008).

Two-moment approximations have been used to study the magnitudes of effects of cell-to-cell variability on output noise (Zechner et al., 2012). For example, cell-to-cell variability in transcription rate and burst size combine constructively with intrinsic noise (Singh & Soltani, 2013).

A similar method to incorporate cell-to-cell variability in all parameters simultaneously combined the LNA (for intrinsic noise) and the unscented transform (for extrinsic noise) (Toni & Tidor, 2013). With this method, it was shown that extrinsic factors were primarily responsible for cell-to-cell variability in a mechanistic model of ERK phosphorylation (Filippi et al., 2016).

In Chapter 7, a study into noise propagation will be presented using the method of Toni & Tidor (2013). This method assumes static extrinsic noise; that is, noise that operates on a timescale much slower than that of intrinsic fluctuations. The method will be used to identify how much reactions and motifs in a pathway contribute to output noise. Its similarities with the method of Zechner et al. (2012) will be drawn out.

3

Nested sampling for Bayesian model selection and parameter inference

In this chapter, I will introduce a new computational implementation for parameter inference and model selection using the nested sampling method of Skilling (2006). First, the algorithm is explained, and a method for its acceleration derived. Then a new package, Sysbions (Johnson et al., 2015), is introduced, following which some parameter inference and model selection results are presented. The chapter concludes with a comparison of Sysbions with two other available model selection software packages.

3.1 Background

Nested sampling is an algorithm for parameter inference and model selection. Model selection in the Bayesian framework relies on comparison of evidence values for different models, where the evidence is the probability of observing the data given the model. The algorithm iteratively samples from the prior subject to nested likelihood constraints. Through knowledge of the sampling method, the evidence is accumulated as the algorithm progresses (Skilling, 2006).

3.1.1 The theory of nested sampling

The evidence, Z, is defined as the integrated likelihood over the prior parameter space. The key to nested sampling is the transformation of the space over which we integrate. Given a prior $\pi(\theta)$ over a parameter space $\theta \in \Theta$, we define $\mathrm{d}X(\theta) = \pi(\theta)\mathrm{d}\theta$. By considering elements ($\mathrm{d}X(\theta)$) rather than areas ($\pi(\theta)\mathrm{d}\theta$) of prior mass, the integral is transformed to a one-dimensional space:

$$Z = \int_{\Theta} \ell(\theta)\pi(\theta)\,\mathrm{d}\theta \tag{3.1}$$

$$= \int_{0}^{1} \ell(\theta)\,\mathrm{d}X(\theta), \tag{3.2}$$

where $\ell(\theta)$ is the likelihood of the point θ. By the definition of $X(\theta)$,

$$X(\tilde{\theta}) = \int_{\ell(\theta)>\ell(\tilde{\theta})} \pi(\theta)\,\mathrm{d}\theta \tag{3.3}$$

is the proportion of prior mass with likelihood greater than that of $\tilde{\theta}$.

By the probability integral transform (Angus, 1994), $X(\theta)$ is uniformly distributed on $[0, 1]$. Therefore, for a sample of N points, $\theta_i \sim \pi(\theta)$, $i = 1, ..., N$, ordered according to increasing likelihood values $\ell_i = \ell(\theta_i)$, we can estimate the values $X_i = X(\theta_i)$. We can then write Equation 3.2 as a sum:

$$Z \approx \sum_{i=1}^{N} \ell_i W_i, \tag{3.4}$$

where $W_i = X_{i-1} - X_i$ is the proportion of prior mass with likelihood greater than that of point θ_{i-1} and less than that of θ_i.

As the algorithm proceeds, low-likelihood points are discarded and replaced by higher-likelihood points, so that there are always N points. Using the probability integral transform, X_i is defined as the expected largest value of N points uniformly distributed on $[0, X_{i-1}]$, and X_0 as the largest of N points uniformly distributed on $[0, 1]$. Further, we can find the expected variance of these random variables, which informs the expected variance of the final value for the evidence (Skilling, 2006) (see Appendix A).

Nested sampling is a method for generating a collection of points θ_i with likelihood ℓ_i, and associating to them the proportion of prior mass they represent, W_i.

3.1.2 The algorithm

The algorithm proceeds by discarding low-likelihood points and replacing them with higher-likelihood points. Thus in the sequence of $\{\ell_i, W_i\}$ pairs, as the algorithm continues, we collect increasing values of ℓ_i and decreasing

values of W_i. This allows for the low-likelihood regions to be coarse-grained while the high-likelihood regions are explored in finer detail.

The general algorithm, as set out by Skilling (2006), proceeds as follows:

1. Initialise $Z = 0$
2. Generate N points from $\pi(\theta)$
3. **for** $i = 1 : T$
   a) Find θ^* with lowest likelihood, ℓ^*
   b) Calculate $W_i = \exp\left(-\frac{i-1}{N}\right) - \exp\left(-\frac{i}{N}\right)$
   c) Set $Z = Z + \ell^* W_i$
   d) Resample $\theta^* \sim \pi(\theta)|_{\ell(\theta)>\ell^*}$
4. **end for**
5. Set $Z = Z + \sum_{j=1}^{N} \ell_j \exp\left(-\frac{T}{N}\right)/N$

The algorithm terminates either when it has completed a pre-specified number of iterations T, or when the rate of accumulation of the evidence falls below some pre-specified tolerance value.

3.1.3 Nested sampling in the literature

Nested sampling has been applied in biology to problems of identifying optimal protein-folding solutions from polypeptide sequences (Burkoff et al., 2012), and has been applied in systems biology to questions of model selection and of topological sensitivity in "model space" (Kirk et al., 2013; Aitken & Akman, 2013; Dybowski et al., 2013; Pullen & Morris, 2014; Babtie et al.,

2014). MultiNest is a nested sampling tool designed for and used in the astrophysics community (Feroz et al., 2009). It is written primarily in Fortran with the possibility of supplying the likelihood function in C. It samples parameters by sampling from ellipsoids defined about the current population of live points, with the possibility of defining a single ellipsoid or more than one in order to capture multiple modes of the likelihood surface.

MultiNest has a finite capacity in that the number of live points is limited, depending on the architecture of the model: for example, a six-parameter model with single-mode sampling is limited to around 15,000 points. The ellipsoidal sampling method increases the acceptance rate of points relative to samples from the full prior, but the construction of the ellipsoids risks excluding areas of genuine interest by virtue of the limited coverage of the present population.

With these considerations in mind, I have created a nested sampling tool specifically for systems biologists, presented in Section 3.2. I have used MultiNest to test some models and as a benchmark for my own programme.

3.2 Sysbions

I have created a command-line tool for systems biologists to perform nested sampling (Johnson et al., 2015). It includes an SBML (Systems Biology Markup Language, Rodriguez et al. (2007)) parser to create likelihood functions directly (Liepe et al., 2010). Written primarily in C, it is also adaptable and accessible to the computational biology community.

I have added further functionality to the package. First, I have included different methods for sampling from the prior subject to the likelihood constraint. Second, I have included a CUDA option for GPU-capable computers. Third, I have included the possibility of skipping multiple points in a single iteration to accelerate the algorithm, and, finally, I have enabled easy plotting of the results of the algorithm, demonstrated in Figures 3.2 and 3.3.

3.2.1 Sampling methods

The accuracy of the approximation in Step $3b$ of the algorithm in Section 3.1.2 relies on the precision in Step $3d$: the live points must be distributed as the prior parameter distribution subject to the likelihood constraint (Skilling, 2006). I have explored four methods for sampling from this distribution. In general, the algorithm samples from a space larger than that of interest, and accept the point sampled if it has likelihood greater than the lowest likelihood, ℓ^*, and reject it otherwise. The aim is to be able to sample from the full space $(\pi(\theta)|_{\ell(\theta)>\ell^*})$, whose boundaries are typically unknown, with maximum efficiency.

Rejection

In the rejection sampling method, the full prior is sampled. This method always satisfies the requirement of Step $3d$. However, the acceptance rate decreases as the lowest likelihood among the population increases: if the areas of interest are confined to small regions of the prior, most samples

will be rejected. This makes the method prohibitively slow.

Random walk

In the random-walk sampling method, a point is randomly chosen from among the current live points, duplicated, and perturbed. The perturbation consists of a walk of 20 steps that are scaled according to the variances of the parameters in the current population. Each step is accepted or rejected based on the likelihood criterion, and subsequent steps are further scaled in order to converge to an acceptance rate of 0.5 (Sivia & Skilling, 2006).

Ellipsoidal

In the ellipsoidal sampling method (Mukherjee et al., 2006), an ellipsoid is defined to surround the current population of live points, from which samples are taken. This method is only compatible with uniform prior parameter distributions. Restriction of the sampling space to the ellipsoid increases the acceptance rate, but it is possible for relevant regions of the prior to be precluded entirely from sampling. In such cases it no longer holds that the points are samples from the prior subject to the likelihood constraint. This is the method used in MultiNest (Feroz et al., 2009).

Population Monte Carlo

Population Monte Carlo (Cappé et al., 2004), or sampling importance resampling, uses the current population of live points to define a proposal distribution, from which a superpopulation of points are sampled, as in sequential importance sampling. From this population, weighted according

to its divergence from the prior, a new population of live points is sampled. These two steps mean that the algorithm is sampling from the correct distribution, and no regions are outrightly excluded.

In this method, points cannot be re-used as they were previously, when the algorithm discarded just one point and replaced it upon each iteration. Here, all points, rather than just one, must be re-sampled on each iteration, which makes it very computationally intensive. To mitigate the effects of this cost, the programme can be run in parallel on graphical processing units, and points can be skipped in order to cover the prior space more quickly.

3.2.2 Skipping points

For likelihood surfaces with vast low-lying regions that contribute little to the evidence, it is advantageous to leap multiple points in each iteration in order to reach the area(s) of interest more quickly. Points that are discarded together (Step $3a$ below) contribute to Z with equal weights W_i (Steps $3b$ and $3c$) and are resampled simultaneously (Step $3d$). For this I propose the following algorithm:

1. Initialise $Z = 0$
2. Generate N points from $\pi(\theta)$
3. **for** $i = 1 : T$
    a) Find the L points $\theta^*_{(1)}, ..., \theta^*_{(L)}$ with lowest likelihoods, $\ell^*_{(1)} < ... < \ell^*_{(L)}$

b) Calculate $W_i = \left(\log(1 - L(1 - e^{-\frac{1}{n}}))\right)^{i-1} - \left(\log(1 - L(1 - e^{-\frac{1}{n}}))\right)^{i}$

c) Set $Z = Z + \sum_{j=1}^{L} \ell^*_{(j)} W_i$

d) Resample $\theta^*_{(1)}, ..., \theta^*_{(L)} : \ \theta^*_{(j)} \sim \pi(\theta)|_{\ell(\theta) > \ell^*_{(L)}}$

4. **end for**

5. Add the likelihoods of the remaining points:
$Z = Z + \frac{1}{N} \sum_{j=1}^{N} \ell_j \left(\log(1 - L(1 - e^{-\frac{1}{n}}))\right)^{T}$.

To leap points adaptively, so that uninteresting regions are traversed quickly and interesting ones explored in detail, I employ a tolerance *tol*. At each iteration the algorithm calculates how many of the lowest-likelihood points together contribute less than *tol* to the evidence given all of the current live points. Formally, for N live points and a current value Z for the evidence, L is maximised such that

$$\frac{\sum_{i=1}^{L} \ell_i W_i}{Z + \sum_{i=1}^{N} \ell_i W_i} < tol. \tag{3.5}$$

When points are leaped, a new method is required to calculate the variance associated with the estimate of Z, given in Appendix A.

Graphical processing units can be used to perform calculations in parallel. More points than are required can be sampled simultaneously and the number required selected from among them. This is particularly advantageous when the rejection rate of sampled points is high (e.g. with the rejection sampling method) or when leaping schemes are used.

3.2.3 Output

A summary file of input and output information is created, documenting the number of live points, the number of iterations, the tolerance, the sampling method and the parameter ranges, followed by the evidence with standard deviation, the information and the means of all parameters and their standard deviations.

Restart files are created, one documenting input values that must persist upon restart (such as the number of live points) and one listing all points, live and discarded. These files can be used to restart the program from where it terminated. It is also possible to specify the path to where the restart files are written.

Posterior distributions of the parameters can be plotted individually as histograms and in pair-wise scatterplots using the data stored in the posterior file (see Figure 3.2 for an example). Finally, a file of trajectories is created that can be read or visualised with any plotting software and compared with the input data (see Figure 3.3 for an example). Full documentation can be found at `http://www.theosysbio.bio.ic.ac.uk/resources/sysbions/`.

3.3 Results

I demonstrate the functionality of Sysbions first in inferring the parameters for a model constructed to explain real data and second in choosing between competing models. Finally, I stay with the model selection question and compare the results of a number of methods.

3.3.1 Parameter inference

I demonstrate parameter inference using the model and data of Swameye et al. (2003) concerning the JAK-STAT signalling pathway. This model, shown in Figure 3.1, was proposed to explain STAT5 dynamics on activation by JAK2. Specifically, they propose inclusion in the model of translocation and recycling of nuclear STAT5 instead of its degradation.

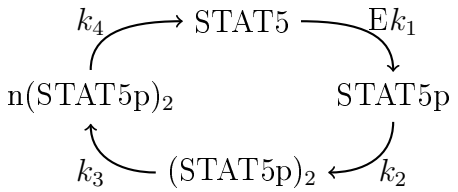

Figure 3.1: Model of the JAK-STAT pathway. STAT5 is phosphorylated by EpoR_A (E), dimerises, enters the nucleus, and is recycled back into the cytoplasm. Parameters are defined in Table 3.1.

Seven parameters make up the model. The first four (k_1, k_2, k_3, k_4) correspond to reaction rates (Figure 3.1) and the final three correspond to scaling parameters that relate the experimental measurements to concentrations (Swameye et al., 2003). The quantities measured relate to the total concentration of phosphorylated cytosolic STAT5 (y_1), the total concentration of cytosolic STAT5 (y_2), and the concentration of EpoR_A (y_3):

$$y_1 = k_5(\text{STAT5p} + 2(\text{STAT5p})_2), \tag{3.6}$$

$$y_2 = k_6(\text{STAT5} + \text{STAT5p} + 2(\text{STAT5p})_2), \tag{3.7}$$

$$y_3 = k_7\text{E}. \tag{3.8}$$

Parameters to infer are defined in Table 3.1. Using 10,000 live points, 300,000 iterations and the random-walk sampling method, I find an evidence

of $\log(Z) = 22.722 \pm 0.048$.

Table 3.1: Prior and posterior parameter distributions for the JAK-STAT signalling pathway.

Parameter	Meaning	Lower bound	Upper bound	Posterior
p1	$[STAT5]_{t=0}$	0	10	0.373 ± 0.084
p2	Variance	0	1	0.0038 ± 0.001
p3	k_5/k_2	0	10	3.944 ± 0.806
p4	k_6/k_2	0	10	2.736 ± 0.661
p5	k_1/k_7	0	10	2.295 ± 0.344
p6	k_3	0	10	5.801 ± 2.329
p7	k_4	0	10	0.365 ± 0.085

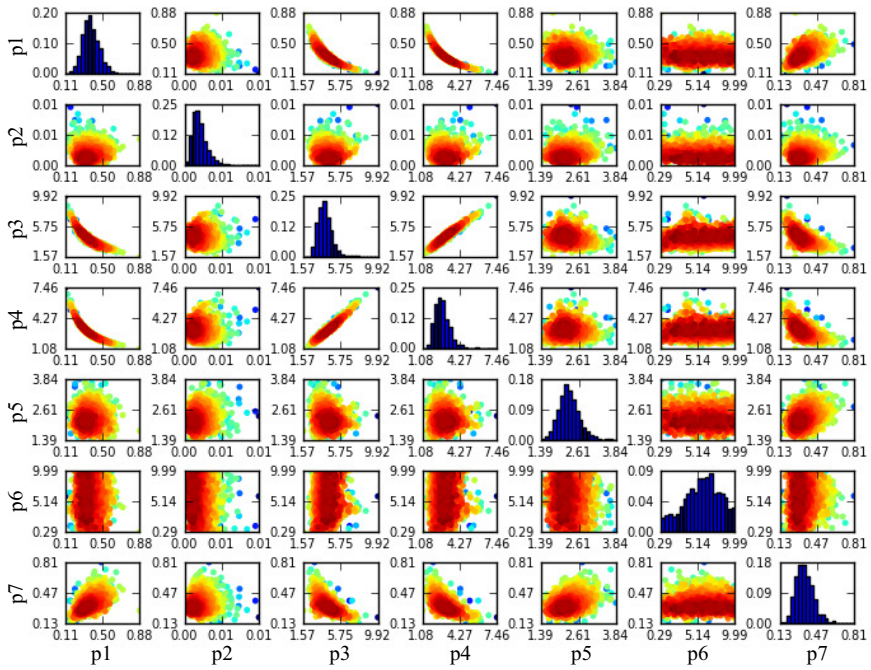

Figure 3.2: Inferred posterior parameter distributions for the JAK-STAT model and data of Swameye et al. (2003). Each scatterplot shows pairwise posterior parameter distributions where colour indicates likelihood (red being the highest likelihood), and histograms show the inferred marginal distributions of individual parameters. Parameters correspond to those of the model in Figure 3.1 and Table 3.1.

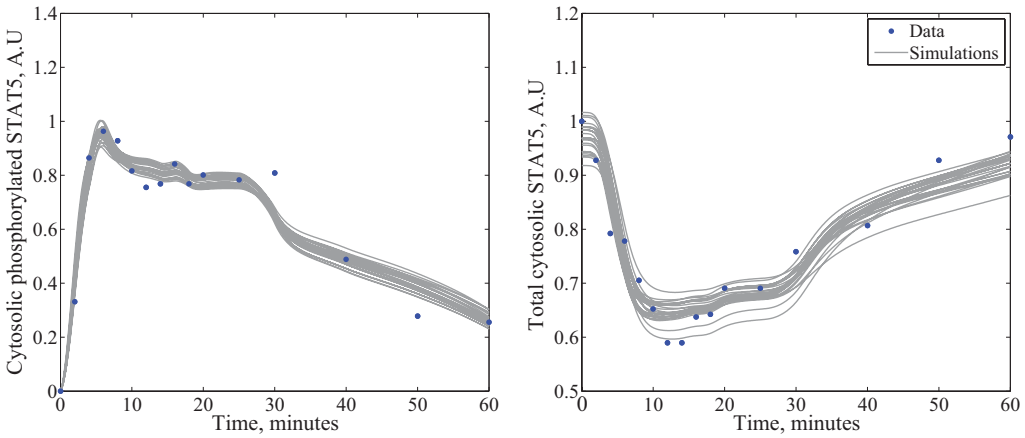

Figure 3.3: Trajectories simulated using parameter sets sampled from the posterior parameter distribution of Figure 3.2 are shown in grey. The data to which they were fit (Swameye et al., 2003) are shown in blue. Left: cytosolic phosphorylated STAT5 concentration. Right: total cytosolic STAT5 concentration.

Scatterplots and histograms of the posterior parameter distributions are shown in Figure 3.2 (a direct output of the Sysbions package). Histograms demonstrate the ranges of parameters that give rise to a good approximation to the data; pairwise scatterplots show any relationships between the parameters. A sample from this distribution can be used to simulate trajectories (Figure 3.3).

3.3.2 Model selection

To demonstrate model selection, I use the artificial signalling system of Vyshemirsky & Girolami (2008a). The authors construct four models (M_1, M_2, M_3 and M_4) and compare them to some artificial data: M_1 is used to generate the data, M_2 is a simplification of M_1, M_3 is incorrect, and M_4 is an overcomplication of M_1 (see Figure 3.4). I test the models using 1,000 live points and a tolerance of 1e-06, once following the standard routine and

once skipping 50 points.

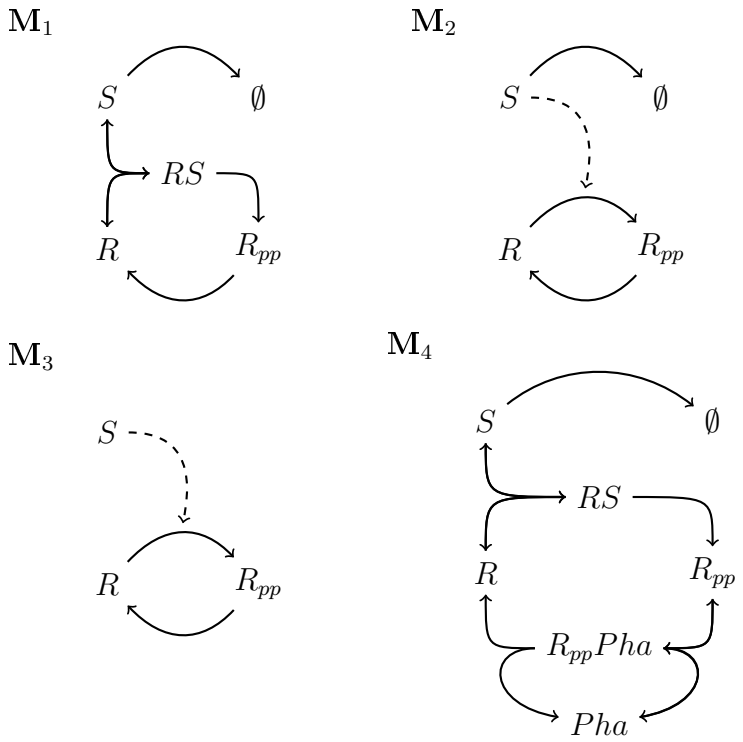

Figure 3.4: The four models proposed in Vyshemirsky & Girolami (2008a). S is a signalling molecule, $\emptyset$ its degradation, R its target, RS the heterodimer, R_{pp} the phosphorylated form of R, Pha a phosphatase, and $R_{pp}Pha$ the phosphatase bound to R_{pp}. Full arrows indicate consuming reactions and dashed arrows indicate non-consuming activation.

Table 3.2: Sysbions model selection results for the signalling system of Vyshemirsky & Girolami (2008a). The program is run natively ($L = 1$) and leaping 50 points at a time ($L = 50$).

Model	$\log(Z)$ $(L = 1)$	$\log(Z)$ $(L = 50)$
M_1	-0.510 ± 0.084	-0.556 ± 0.085
M_2	-1.304 ± 0.088	-1.363 ± 0.088
M_3	-6.708 ± 0.045	-6.792 ± 0.051
M_4	0.102 ± 0.077	0.0718 ± 0.079

Our results, shown in Table 3.2, are consistent and in agreement with each other but not with those of Vyshemirsky & Girolami (2008a) in terms of

model ranking. They rank the models as follows: $M_1 > M_4 > M_2 > M_3$ (see Table 3.3), whereas I find M_4 to be the preferred model. The Bayes factors for M_4 in preference to M_1 are 1.8441 and 1.8725, respectively, for $L = 1$ and $L = 50$. Although M_4 seems preferred, it does not conclusively outperform M_1 as it does the other two models, as Bayes factors between 0 and 5 are considered 'barely worth mentioning' (Jeffreys, 1961)

To see if the rankings are dependent on the data, I simulated ten more datasets (each consisting of seven data points) and computed each model's evidence five times for each dataset using our nested sampling implementation with parameters as detailed above. The results are shown in Figure 3.5.

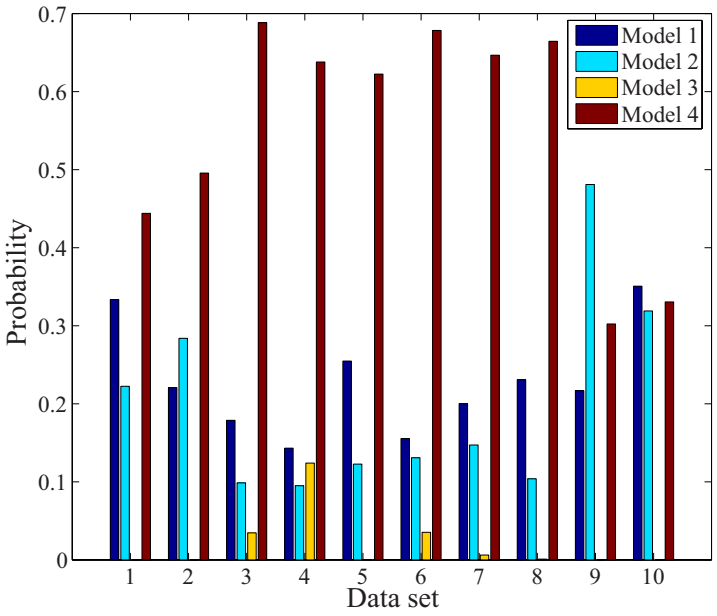

Figure 3.5: Model comparison using simulated datasets. Ten datasets are simulated from M_1. The evidence of each model given each dataset is computed five times with Sysbions. Bars show the relative probability for each model to be the true model based on the dataset.

M_4 outranks all other models for eight out of the ten datasets. The Bayes factor the preference of M_4 over M_1 does not exceed 5 for any dataset. Were we to understand these results to imply a rejection of M_1 in favour of M_4, then the results of the nested sampling algorithm would contradict Occam's razor, the principle that a more complicated model should be selected only if it provides a substantial improvement over the simpler model. Further, we would have rejected the true, data-generating model in favour of another model. This concerning conclusion is avoided only by asserting a Bayes factor of 5 as a threshold below which a realised value is not considered significant.

3.3.3 Comparison of Sysbions with other packages

The results in Table 3.2 differ considerably from those of Vyshemirsky & Girolami (2008a). I use the BioBayes package (Vyshemirsky & Girolami, 2008b) to see if it is possible to replicate their results using their settings. BioBayes is a software package that implements annealing–melting integration for estimating the evidence for a model given some data.

With the settings and prior parameter distributions quoted in Vyshemirsky & Girolami (2008a), I find the same ranking as before. Table 3.3 summarises some results, and Figure 3.6 demonstrates graphically a comparison between a number of methods. Each method finds M_4 to be the best model to explain the data, despite the penalty it incurs for having extra parameters. However, unlike Vyshemirsky & Girolami (2008a), I do not find the Bayes factors to be so convincing as to rule out three models in light of one model.

Table 3.3: BioBayes model selection results for the signalling system of Vyshemirsky & Girolami (2008a). The first column shows the published results of Vyshemirsky & Girolami (2008a). The following columns show results of the present study. For all studies, the convergence threshold is 1.21.

Study	2008	Present	Present	Present (8 runs)
Priors	$\Gamma(1,3)$	$\Gamma(1,3)$	$\Gamma(1,3)$	$\mathcal{U}(0,1)$
Chains	20	20	3	3
Samplers	40	40	20	20
Model 1	45.8 ± 0.2	-3.3 ± 0.25	-3.06 ± 0.01	-0.64 ± 0.19
Model 2	29.2 ± 0.1	-3.38 ± 0.05	-3.15 ± 0.27	-1.29 ± 0.25
Model 3	-1.1 ± 0.1	-7.23 ± 0.01	-7.34 ± 0.004	-6.7 ± 0.02
Model 4	34.8 ± 0.1	-3.06 ± 0.05	-2.77 ± 0.03	0.32 ± 0.08

In this case, the principle that the least complicated model capable of explaining the data should be ranked highest is not met. The conclusion must be that the data are insufficient to distinguish between the two models.

Figure 3.6 shows that, in general, the ellipsoidal methods estimate slightly larger evidences. It was expected that the evidence would fall as the size of the ellipsoid increases, which has not been conclusively shown with the tests run. However, as the size increases, the program becomes prohibitively slow. We see also that results are quite consistent as the number of live points changes (for MultiNest and Sysbions with random-walk sampling). The greatest agreement seems to be between BioBayes and Sysbions random walk.

3.4 Summary

To conclude, Sysbions was created for easy application of nested sampling by the systems biology community (Johnson et al., 2015). It is proposed as

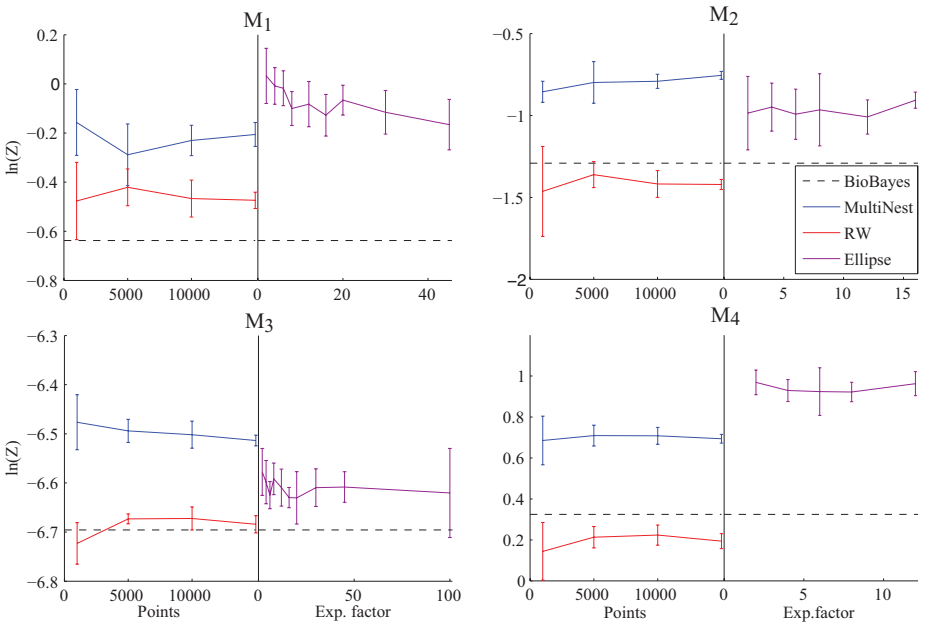

Figure 3.6: A comparison of MutliNest and Sysbions applied to the model selection problem of Vyshemirsky & Girolami (2008a). Points correspond to the means and error bars the standard deviations of five runs. My results from the BioBayes package are shown for reference in the black dashed line. The annealing–melting algorithm was used with 20 chains, three simultaneous samplers, a convergence threshold of 1.21, a posterior sample size of 20,000 and a thinning rate of 5. For all tests, all models' parameters had uniform prior distributions over [0,1] and the data had associated Gaussian noise with variance 0.01, as was used its generation. Left-hand plots show the effect of the number of live points used on the evidence, Z. In blue are the results of MultiNest, with standard (default) input. In red are the results of Sysbions using a random-walk (RW) sampling method and exit tolerances of [1e-4,1e-4,e-6,1e-7]. Right-hand plots show the effect of the ellipsoid expansion factor on Z. 2,000 live points and a tolerance of 1e-5 were used for each test.

an alternative to MultiNest (Feroz et al., 2009), where the key differences are: the accessibility to the biological community via SBML integration, the choice of language (C being more used that Fortran), and the outputs produced; its computational capacity and GPU support; the development of methods (specifically, the sampling schemes and the leaping scheme); and the modularity of the programme, allowing easy manipulation and

adaptation.

Sysbions has been applied to a large-scale model ranking study. The ability to calculate evidences quickly allows for the sensitivity of a model to its topology to be studied, wherein a model's structure is permuted and the resulting effects on the model's ability to explain some data can be quantified (Babtie et al., 2014).

The studies presented in Sections 3.3.2 and 3.3.3 suggest a need for care in interpreting evidence values. Among four competing models, one was the true, data-generating model, and another was an over-complication of that model. All nested sampling applications used in the above studies found a preference for the over-complicated model in terms of evidence values. Only through the assertion that Bayes factors must exceed 5 to be considered significant do we avoid rejection of the true model in favour of the over-complicated one.

4

Finite state projection for protein-dependent cell population growth

In this chapter, two methods are presented to model a population of dividing cells, where the rate of division is dependent on internal protein dynamics. The methodology is set out in a general format here and applied to specific biological questions in Chapters 5 and 6.

In Section 4.1, the biological motivation for the development of the methodology is outlined, followed by the method on which it is based. Two convergent methods are presented in Sections 4.2 and 4.3, and are validated and evaluated in Section 4.4.

4.1 Introduction

Cell populations arise through a sequence of cell divisions that give rise to a growing number of (nominally) genetically identical cells. The behaviours at the macro or population level emerge in part as a result of the conditions driving cell divisions at the micro or individual level. These conditions are heterogeneous due to the inherent stochasticity of biological systems. Therefore, in order to understand the mechanisms of growth at the macro level, it is necessary to understand the variability at the micro level, and how each conceivable realisation at the micro level informs cellular behaviour (Shahrezaei & Marguerat, 2015).

For example, antimicrobial resistance is becoming an increasing public-health threat (The review on antimicrobial resistance, 2016). In order to design effective strategies to prevent or avert it, it is necessary to understand how microbial populations respond to toxic conditions. Variability within the population is crucial to these dynamics, underlying the emergence of persister populations (Allen & Waclaw, 2016) and support of mutability (Tadrowski et al., 2016). There is therefore much to be gained from modelling how internal factors (which might depend on external conditions) lead to population-level dynamics (Shahrezaei & Marguerat, 2015).

In this chapter, following an overview of previous advances in modelling population dynamics, a new contribution will be presented: an adaptation of the finite state projection that allows dynamics at the protein level to drive population dynamics. This will be applied to biological questions in

the subsequent two chapters: the dynamics of a protein-dependent cell cycle in Chapter 5, and yeast population fitness in the presence of antibiotic in Chapter 6.

4.1.1 Modelling cell population growth

Numerous methods have been presented to model cell population growth. They each aim to capture something of population dynamics by employing rules around cell division. Delay differential equations impose a time lag for a newly born cell to mature before it is able to divide, and are often cited as a proxy for cell-cycle stages that are not explicitly modelled (Villasana & Radunskaya, 2003). Branching processes are used to derive analytically the existence of a stable birth-type distribution, where new cells inherit some 'type' from their mother that determines their propensities to reproduce (Taib, 1999; Alexandersson, 2001). These processes lend themselves to questions of cell differentiation (Nordon et al., 2011), for example, and take account of the type as a random variable, but do not allow for a mechanistic underpinning for this variability. Where analytic models are undesirable or indefinable, stochastic simulations, agent-based models and cellular automata have been used to emulate what populations might result from a model of growth (Charlebois & Kaern, 2013; Murphy & Walshe, 2011; Nevozhay et al., 2012; Schwabe & Bruggeman, 2014; Tadrowski et al., 2016; Altinok et al., 2011).

Physiologically structured population models define the populations to be structured according to one or several physiological attribute(s) (Banasiak

et al., 2012), such as age (Dyson et al., 2002), size (Ellermeyer & Pilyugin, 2012), stage (Massie et al., 2013), DNA content (Basse et al., 2003) and protein. In these models, the growth rate for each individual depends on their realisation of the quality under consideration. For example, in age- or size-structured models, older (or larger, respectively) cells have a greater propensity to divide. This prevents the artefact of newly born cells immediately dividing. Such models can also be used with label structuring to infer rates of death and division from fluorescence data (Luzyanina et al., 2007).

The present work most closely relates to protein-structured models, in which division depends upon protein (e.g. cyclin) content (Borges et al., 2014; Bekkal Brikci et al., 2007). In Bekkal Brikci et al. (2007), the population is structured additionally by age, whereas in Borges et al. (2014), delay differential equations are employed to prevent immediate re-entry into the cell cycle. These models typically define two stages, proliferative and quiescent. Propensities to divide and transition to quiescence depend on age and cyclin content, whereas transition to proliferation depends on population density. The result is a system of two partial differential equations, one for each sub-population.

These studies recognise the role of cell stages in population dynamics: growth dynamics fundamentally depend on which cells are dividing, and their daughters' propensities to divide. A promising development would therefore be to include many cell-cycle stages, as proteins leverage control at checkpoints throughout the cell cycle (Altinok et al., 2011) (rather than

just two: Massie et al. (2013); Borges et al. (2014); Bekkal Brikci et al. (2007)). Therefore inclusion of stages allows fine-grained modelling of the control of the protein.

Here, I present a general formulation for modelling cell population growth allowing for any number of stages. Protein dynamics can be defined for each stage, which may parametrise transition rates between stages. What is returned is a full probability distribution over stage and protein content.

4.1.2 The finite state projection

We model cell population dynamics by first modelling the noisy protein dynamics that drive it, via the finite state projection (FSP) (Munsky & Khammash, 2006). In the FSP, the chemical master equation (CME) is approximately solved by truncating the state space of interest. There is often an infinite number of states available to a Markov process describing a system that counts protein content. In reality, most of the probability mass is contained within a space bounded by some finite number of proteins. By excluding a small proportion of total probability mass, we retain a precise, tractable differential equation for the probability distribution of cell state.

For a model comprising N_S cell stages and a single protein with a maximum content of N_P, there are $N_T = N_S(N_P + 1)$ possible states a cell can be in. A vector P denotes the probability of being in each state. Transition probability matrix A operates on P to give the rate of change: $\dot{P} = AP$. Value A_{ij} is the rate of transition from state j to state i. The sum of P is 1, and the sums of the columns of A are 0 (ensuring that probability mass

that leaves one state enters another state).

The differential equation $\dot{P} = AP$ is easily solved to give the distribution over P at some time t. The solution for $P = P(t)$ has the form $P(t) = P(0)\exp(At)$.

4.1.3 The finite state projection and cell division

In a mitosis event, one cell becomes two cells. The FSP as described above cannot capture the increase in probability for divided cells relative to dividing cells without compromising conservation of probability. We present two methods for circumventing this problem. The first, presented in Section 4.2, assumes a steady state and recycle cells (and therefore probability). The second, presented in Section 4.3, keeps track of the total population, and therefore the proportions of all cells.

4.2 Method 1: cell recycling

A possible solution to the problem of probability conservation is to assume a steady-state population. This might be assumed if a) the population has reached a natural steady state, or b) the population is kept artificially at a steady state through continual removal of cells at random to maintain the same overall number.

A steady state is one in which the state variable is unchanging over time. Stationarity of cell number in a population of cells requires that the rate of mitosis is matched by the overall rate of death of all cells, so that for every

cell that is created, one is destroyed. Equivalently, to supply probability mass to cells that are created into new states, probability is drawn from all states. This recycling is represented by $\Delta(P)$, where $\Delta(P)$ is the amount to be recycled. The dependence on the state probability vector P breaks the linearity of the CME: rather than $\dot{P} = AP$, we now have $\dot{P} = A(P)P$.

4.2.1 The A matrix

The matrix A contains the transition rates between all states. For a general model, transitions out of a state in stage i with n proteins are: protein creation (with rate $\alpha_i(n)$), protein decay (with rate $\gamma_i(n)$), transition through the cell cycle stages (with rate $k_i(n)$), and apoptosis (with rate $\Delta(P)$). Mitosis events result in daughter cells with protein content n^* from cells with n proteins according to some function $g_0(n, n^*)$.

Consider a system with N_S stages, labelled $\{0, 1, ..., N_S - 1\}$, where $G, M \subseteq \{0, 1, ..., N_S - 1\}$ are the indices of daughter and mitosing stages, respectively. To implement cell recycling, we supply the total flux into daughter states from mitosing states, and then refund half of the flux out of mitosing cells from all states. The net result is that one mitosing cell goes to two daughter cells, and one random cell dies to maintain the population number.

The amount to refund a mitosing state with n proteins is

$$\delta_n(P) = \frac{1}{2} k_M(n) P_{M,n}, \tag{4.1}$$

where $k_M(n)$ is the rate of exit from that state, and $P_{M,n}$ denotes the

probability of being in a mitosing state with n proteins. Assuming one of the N_S stages is mitosing, there are N_P+1 mitosing states to refund, which require a total mass of

$$\Delta(P) = \sum_{n=0}^{N_P} \delta_n(P) \tag{4.2}$$

$$= \frac{1}{2} \sum_{n=0}^{N_P} k_M(n) P_{M,n}. \tag{4.3}$$

This probability mass is drawn from all states in the system, and represents that each state has equal probability to die.

For a cell in state j, we define the number of proteins $j_p = \lfloor \frac{j-1}{N_S} \rfloor$ and its stage $j_S = (j-1) \bmod N_S$. For clarity, an enumeration is given up to $j = 8$ in Table 4.1.

Table 4.1: Enumeration of state labels in the construction of the A matrix. The top row gives the first eight indices j. The j_S are the state indices, with the state labels in brackets. The j_p are the protein values given the indices j.

j	1	2	3	4	5	6	7	8
j_S	0 (G1)	1 (S)	2 (G2)	3 (M)	0 (G1)	1 (S)	2 (G2)	3 (M)
j_p	0	0	0	0	1	1	1	1

Then the matrix A is defined:

$$A_{ij} = \begin{cases} -k_{j_S}(j_p) - \alpha_{j_S}(j_p) - \gamma_{j_S}(j_p) - \Delta(P) & i = j, i \neq N_T + 1 \\ k_{j_S}(j_p) & i = j+1, j_S \neq M \\ k_M(j_p) g_0(j_p, i_p) & j_S = M, i_S = G, i_p \leq j_p \\ \alpha_{j_S} & i = j + N_S \\ \gamma_{j_S}(j_p) & i = j - N_S \\ -\Delta(P) & i = j = N_T + 1 \\ \delta_{i_p}(P) & j = N_T + 1, i_S = M \\ \Delta(P) & i = N_T + 1, i \neq j \end{cases} \tag{4.4}$$

An illustration of the construction of the A matrix will be given in the next subsection.

4.2.2 Four-stage cell-cycle model

Figure 4.1: Four-stage cell-cycle model. Cells begin in stage G1, progress through stages S and G2 to M, and divide, giving rise to two cells in stage G1.

We demonstrate the method with a four-stage, protein-dependent cell cycle, which will be later developed in Chapter 5. The full A matrix for the four-stage system shown in Figure 4.1, where the stages are $\{S_0, S_1, S_2, S_3\} = \{\text{G1, S, G2, M}\}$, is given by:

$$A = \begin{pmatrix} A_0 - B_0 & A_{0,1} + D_1 & \cdots & A_{0,N_P} & \mathbf{b}_0(P) \\ B_0 & A_1 - B_1 - D_1 & \cdots & A_{1,N_P} & \mathbf{b}_1(P) \\ 0 & B_1 & \cdots & A_{2,N_P} & \mathbf{b}_2(P) \\ 0 & 0 & \cdots & A_{3,N_P} & \mathbf{b}_3(P) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & A_{N_P} - B_{N_P} - D_{N_P} & \mathbf{b}_{N_P}(P) \\ \mathbf{c}_\Delta(P) & \mathbf{c}_\Delta(P) & \cdots & \mathbf{c}_\Delta(P) & -\Delta(P) \end{pmatrix}, \tag{4.5}$$

where $\mathbf{c}_\Delta(P) = [\Delta(P)\ \Delta(P)\ \Delta(P)\ \Delta(P)]$, $\mathbf{b}_n(P) = [0\ 0\ 0\ \delta_n(P)]^T$,

$$B_n = \begin{pmatrix} \alpha_0 & 0 & 0 & 0 \\ 0 & \alpha_1 & 0 & 0 \\ 0 & 0 & \alpha_2 & 0 \\ 0 & 0 & 0 & \alpha_3 \end{pmatrix}, \tag{4.6}$$

$$D_n = \begin{pmatrix} \gamma_0(n) & 0 & 0 & 0 \\ 0 & \gamma_1(n) & 0 & 0 \\ 0 & 0 & \gamma_2(n) & 0 \\ 0 & 0 & 0 & \gamma_3(n) \end{pmatrix}, \tag{4.7}$$

$$A_n = \begin{pmatrix} -k_0(n) & 0 & 0 & k_3(n)g_0(n,n) \\ k_0(n) & -k_1(n) & 0 & 0 \\ 0 & k_1(n) & -k_2(n) & 0 \\ 0 & 0 & k_2(n) & -k_3(n) \end{pmatrix} - \mathrm{diag}(\Delta(P)) \tag{4.8}$$

and

$$A_{n^*,n} = \begin{pmatrix} 0 & 0 & 0 & k_3(n)g_0(n,n^*) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{4.9}$$

Equations 4.6 and 4.7 correspond to protein creation and degradation, respectively. Equation 4.8 represents the diagonal units of the matrix corresponding to four stages with n proteins. It shows forward transitions that occur at rate $k_i(n)$, and mitosis events that occur at rate $k_3(n)$ and result in stage 0 and n^* proteins with probability $g_0(n, n^*)$ (also shown in Equation 4.9).

Refunding of mitosing states is shown in the far-right column of Equation 4.5. This is drawn from the 'null state' in the bottom-right corner. The null state is fed by all states, shown in the bottom row. Note that the flux into the null state (cell death) is equal to the flux out of it (cell birth), so it should neither accumulate nor lose probability mass (the death rate equals the birth rate and we have a steady state).

We use a null state rather than transferring mass directly from all states to mitosing states as it is computationally more efficient but mathematically equivalent. We refund mitosing states rather than daughter cells as it is more tractable, as there may be multiple possible destinations for mitosing cells.

4.3 Method 2: growth and probability redistribution

In this section a different approach to modelling population dynamics with the FSP is presented, and in Section 4.4.1 it is shown that the two methods converge.

To model the cell cycle whilst allowing for mitosis (necessitating probability redistribution) as well as the possibility of growth, we keep track of the overall population size and use the FSP architecture to track the proportions. At each time step, for each state, we are interested not only in which reactions occurred that affect that state, but also in all other reactions that occurred, as any change in cell number affects the proportions of all states.

Therefore, the rate of change of P depends on itself and the overall number of cells, N:

$$\dot{P} = A(N, P)P. \tag{4.10}$$

Suppose we have $N_T = N_S(N_P + 1)$ states indexed by $j = \{1, 2, ..., N_T\}$. Let R_j denote the number of cells in each state, and P_j the proportion of N that is in state j. So $\sum P_j = 1$, $\sum R_j = N$, and $R = NP$. Usually, we express the result of reactions occurring over some time step as

$$R_i^{(t+1)} = R_i^{(t)} + u_i(R^{(t)}),$$

where u_i is the change to R_i as a function of all the states.

In order to model the evolution of probabilities, we need to know what happens to all cells at each time point. A state's proportion P_i can change even when there has been no change in R_i. What we seek is the change in proportion, which we call v_i:

$$P_i^{(t+1)} = P_i^{(t)} + v_i(N^{(t)}, P^{(t)}). \tag{4.11}$$

The function v_i can be expressed in terms of N and P at time t:

$$v_i(N, P) = \frac{u_i(NP) - P_i \sum u_j(NP)}{N + \sum u_j(NP)}. \tag{4.12}$$

Here, v_i is the change in proportion, u_i is the change in number, N is the total number of cells, and $\sum u_j(NP)$ is the change in total cell number going to the next time step. The derivation is shown in Appendix B and applies for any N. In the next subsection the equation is illustrated with a small system.

4.3.1 Illustrative example

Consider two cell types ($N_T = 2$) with initial conditions $P_1 = P_2 = 0.5$, $N = 2$ (i.e. we begin with two cells, one of each type). There is one reaction, in which one cell type increases its population by one: $R_1^{(t+1)} = R_1^{(t)} + 1$, so $u_1(R) = 1$. The other cell type is inactive: $u_2(R) = 0$.

Table 4.2 charts the progress of the two states at each time point, t. Note that each state contributes probability mass x/y at each time step, where x is the number of cells in that state and y is the total number of cells at this time point multiplied by the number at the next time point. These two contributions are added together and supplied to the state that is accumulating probability, state R_1.

Table 4.2: Probability evolution for a simple two–cell-type system. Columns P_i show the probability of cells to be in state i at time t. The subsequent two columns show how $v_i(N, P)$ is calculated: the first shows the probability mass it gains due to any reactions affecting it that occur over the next time step. The second shows the probability mass that it loses due to all reactions that occur. The probability at the next time step, $P_i^{(t+1)}$, is calculated using these columns according to Equation 4.11.

Time	P_1	$\frac{u_1(NP)}{N+\sum u_j(NP)}$	$-\frac{P_1 \sum u_j(NP)}{N+\sum u_j(NP)}$	P_2	$\frac{u_2(NP)}{N+\sum u_j(NP)}$	$-\frac{P_2 \sum u_j(NP)}{N+\sum u_j(NP)}$
0	$1/2$	$1/3$	$-1/6$	$1/2$	0	$-1/6$
1	$2/3$	$1/4$	$-2/12$	$1/3$	0	$-1/12$
2	$3/4$	$1/5$	$-3/20$	$1/4$	0	$-1/20$
3	$4/5$	$1/6$	$-4/36$	$1/5$	0	$-1/36$

This demonstrates the general principle of the method: probability is drawn from each state according to its proportion and supplied to states accumulating mass.

4.3.2 Probability redistribution

In Equation 4.12, probability mass is drawn from all states into all other states in the term $\sum u_j = \sum u_j(NP)$. Many of the terms in $\sum u_j$ cancel out, and it is only necessary to consider probability mass into states whose changes (u_j) effect change in the total number of cells, $N^{(t+1)} - N^{(t)}$. Therefore, we can disregard most of the u_j and consider only residual changes, which we denote $\tilde{u}_j$. By definition, $\sum_j u_j = \sum_j \tilde{u}_j$. States j that are created via mitosis will have $\tilde{u}_j > 0$, and all other states, for which all terms in $\sum u_j$ cancel out, will have $\tilde{u}_j = 0$.

For a state j that results of cell division and inherits n^* proteins, we have

$$\tilde{u}_j = \sum_{n=0}^{N_P} k_M(n) g_0(n, n^*) N P_{M,n} \tag{4.13}$$

as the total flux into that state due to mitosis. This allows the total cell number N to be tracked, which can be done via a readout of mitosis and death events. We have that $N^{(t+1)} = N^{(t)} + \sum u_j$, so

$$\begin{aligned}
N^{(t+1)} - N^{(t)} &= \sum_{j=1}^{N_S(N_P+1)} \tilde{u}_j && (4.14)\\
&= \sum_{n^*=0}^{N_P} \sum_{n=0}^{N_P} k_M(n) g_0(n, n^*) N^{(t)} P_{M,n} && (4.15)\\
&= N^{(t)} \sum_{n=0}^{N_P} k_M(n) P_{M,n} \sum_{n^*=0}^{N_P} g_0(n, n^*) && (4.16)\\
&= N^{(t)} \sum_{n=0}^{N_P} k_M(n) P_{M,n} && (4.17)\\
&= 2N^{(t)} \Delta(P), && (4.18)
\end{aligned}$$

using Equations 4.13 and 4.3, and the requirement $\sum_{n^*=0}^{N_P} g_0(n, n^*) = 1$.

As with the recycling method, this quantity is drawn from all states and delivered into states that result of mitosis.

4.3.3 The A matrix

We define again the number of stages N_S, $G, M \subseteq \{0, 1, ..., N_S - 1\}$ the indices of daughter and mitosing stages, respectively, and $j_p = \lfloor \frac{j-1}{N_S} \rfloor$ the number of proteins and $j_S = (j-1) \bmod N_S$ the stage for a cell in state j. Then the matrix A, analogously to Equation 4.4 for the recycling method, is defined:

$$(1+2\Delta(P))A_{ij} = \begin{cases} -k_{j_S} - \alpha_{j_S}(j_p) - \gamma_{j_S}(j_p) - 2\Delta(P) & i = j \\ k_{j_S} & i = j+1, j_S \neq M \\ k_M g_0(j_p, i_p) & j_S = M, i_S = G, i_p \leq j_p \\ \alpha_{j_S} & i = j + N_S \\ \gamma_{j_S}(j_p) & i = j - N_S \\ \tilde{u}_i / N & i_S = G \end{cases} \tag{4.19}$$

where $2\Delta(P) = N^{(t+1)}/N^{(t)} - 1$ from Equation 4.18.

To make the model the same as the recycling model in terms of reaction rates, we can simply scale k_M by $1/2$; then, we are redistributing exactly the same amount of probability. The only differences are that the recycling method uses a null state that acts as the recycler, whereas this method has all states contributing directly to daughter states, and that this method tracks the total cell number.

In summary, the method requires:

1. Every state to contribute probability mass to mitosing states;

2. The total number of cells to be tracked and updated based on the flux out of mitosing states;

3. The transition matrix to be corrected by $N^{(t)}/N^{(t+1)}$, where $N^{(t)}$ is the number of cells in the current time step and $N^{(t+1)}$ the number of cells in the next time step.

4.3.4 Four-stage cell-cycle model

The A matrix for a cell-cycle model (analogous to Equation 4.5) can be written

$$A = \frac{1}{N + \sum u_j} \begin{pmatrix} A_0 - B_0 & A_{0,1} + D_1 & \cdots & A_{0,N_P} \\ A_{1,0} + B_0 & A_1 - B_1 - D_1 & \cdots & A_{1,N_P} \\ A_{2,0} & A_{2,1} + B_1 & \cdots & A_{2,N_P} \\ \vdots & \vdots & \ddots & \vdots \\ A_{N_P,0} & A_{N_P,1} & \cdots & A_{N_P} - B_{N_P} - D_{N_P} \end{pmatrix} \tag{4.20}$$

where B_n and D_n are defined as before in Equations 4.6 and 4.7.

For a system with up to N_P proteins and four stages, there are $4(N_P + 1)$ possible states, which we index by n_i, where n is the number of proteins and i the stage. Thus u_{n_0} refers to the changes to the state with n proteins in stage 0.

Then we have, for a four-stage model with stage 0 the daughter cell and

stage 3 the mitosing cell,

$$A_n = \begin{pmatrix} -k_0N + \tilde{u}_{n_0} & \tilde{u}_{n_0} & \tilde{u}_{n_0} & k_3Ng_0(n,n) + \tilde{u}_{n_0} \\ k_0N & -k_1N & 0 & 0 \\ 0 & k_1N & -k_2N & 0 \\ 0 & 0 & k_2N & -k_3N \end{pmatrix} - \text{diag}\left(\sum u_j\right) \tag{4.21}$$

and

$$A_{n^*,n} = \begin{pmatrix} \tilde{u}_{n_0} & \tilde{u}_{n_0} & \tilde{u}_{n_0} & \tilde{u}_{n_0} + k_3Ng_0(n,n^*) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{4.22}$$

where, again, $g_0(n, n^*)$ is some function that determines protein allocation upon partition (and is equal to zero when $n^* > n$).

On each row of the A matrix corresponding to a daughter cell, every entry has in it some $\tilde{u}_j$, which is equal to the sum total of all flux into that state from all mitosing states. This is seen in Equations 4.21 and 4.22 and performs the same function as the recycler, $\Delta(P)$, in Equations 4.8 and 4.9.

4.4 Method evaluation

In this section, the methods for modelling protein-dependent population dynamics are first compared to each other and then validated via comparison with stochastic simulation. This uses for demonstration the four-stage cell-cycle model previously introduced. Then the methods' scope is discussed: first it is shown how they can be applied to a system in which

cells can die at different rates; and, finally, their inability to incorporate population-density effects is demonstrated.

4.4.1 Comparison of the methods

The A matrix for recycling (Equation 4.4) is very similar to that for redistribution (Equation 4.19); it differs only in that Equation 4.19 has a factor of $1 + 2\Delta(P) = {}^{N^{(t+1)}}/_{N^{(t)}}$. In recycling, a steady state is assumed, so that $N^{(t+1)} = N^{(t)} \forall t$. Applied to the four-stage cell-cycle model, there is a very small discrepancy in the transition to steady-state proportions, shown in Figure 4.2. The methods are equivalent when, for the redistribution model, $N^{(t+1)} = N^{(t)}$ (at small t, when there is no flux through mitosis) and converge when N is large relative to flux through mitosis.

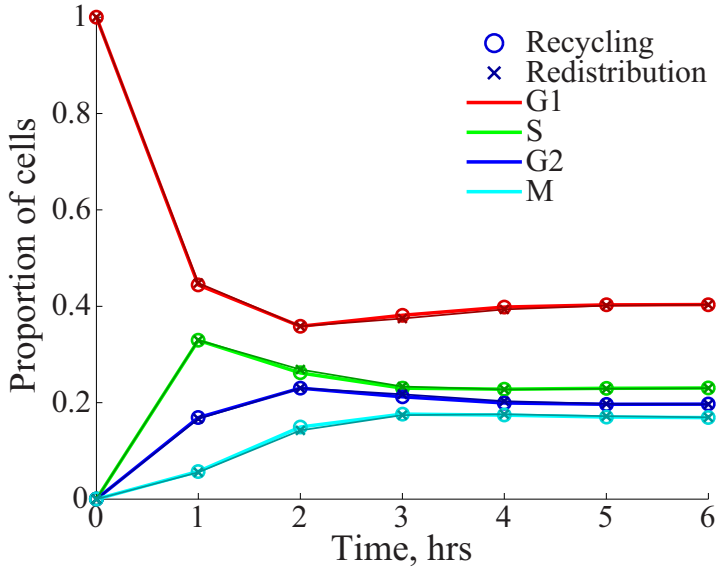

Figure 4.2: Comparison of recycling method and redistribution method, using the four-stage model of Figure 4.1. Their predicted proportions differ only very slightly in the transition to steady proportions.

Having established the near equivalence of the methods, in what follows we use recycling as representative of both methods, unless otherwise specified.

4.4.2 Validation

To validate the methods, we compare their result to many stochastic simulations. We implement the stochastic simulation algorithm (SSA) for the protein-dependent cell cycle by simulating one cell at a time. For the four-stage system, a cell has six possible reactions: the four stage transitions (of which only one at a time can be realisable), protein creation, and protein degradation. When a cell exits from M to G1, its protein content is chosen via the binomial distribution and the simulation continues. Then the sister cell is also simulated, starting at the time of the mitosis event, and with the complement protein content. This process is illustrated in Figure 4.3.

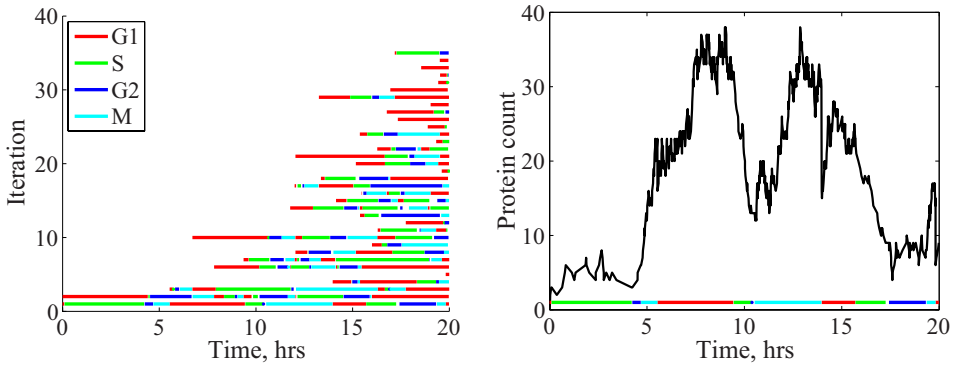

Figure 4.3: Stochastic simulation of the four-stage model. Left: results of a simulation that started with two cells, where each line represents a cell. Cells go through four stages, G1, S, G2, and M. Upon division, a new cell is started from that time point with the appropriate protein complement. E.g., at time $t \approx 5.5$, cell 1 exits mitosis, and at the same time, cell 3 begins in G1. At the end, there were 36 cells. Right: protein-content trajectory of one cell, with cell stage shown below.

Comparison of the methods with the SSA is shown in Figures 4.4 and 4.5. Parametrisation of this model is discussed in full in Chapter 5 but, for the purposes of comparison, the following parametrisation is used, following Overton et al. (2014):

$$k(n) = \left\{ \frac{35^2}{35^2 + n^2}, 1, 1, 2 \right\},$$

$$g_0(n, n^*) = \binom{n}{n^*} \frac{1}{2^n},$$

$$\alpha = \{80, 20, 40, 80\},$$

$$\gamma_i(n) = n \text{ for } i \neq 0,$$

$$\gamma_0(n) = n \left(1 + \frac{8C_G^4}{1 + C_G^4} \right),$$

and

$$C_G(n) = \frac{1.75}{10n^4/(35^4+n^4) + 0.75}.$$

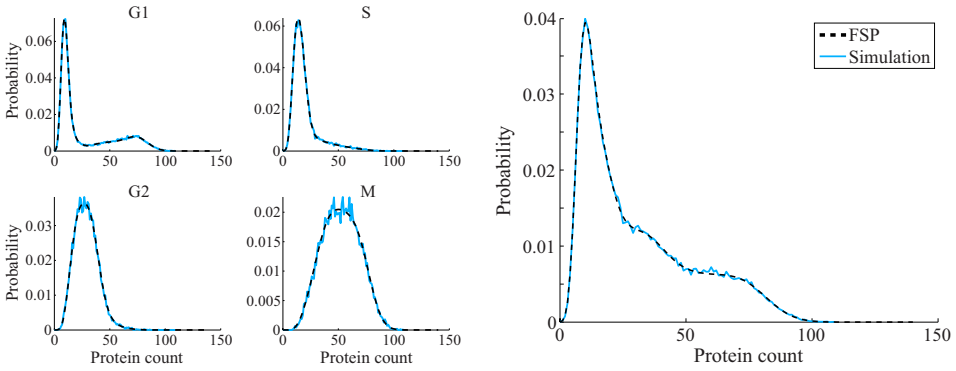

Figure 4.4: Protein distributions in each cell stage resulting of the four-stage model with protein bifurcation in stage G1. Left: protein distribution in each stage. Right: overall protein distribution. Black dashed: recycling FSP model. Blue: SSA results, starting with 500 cells and ending at time $t = 30$ with ~84,000 cells.

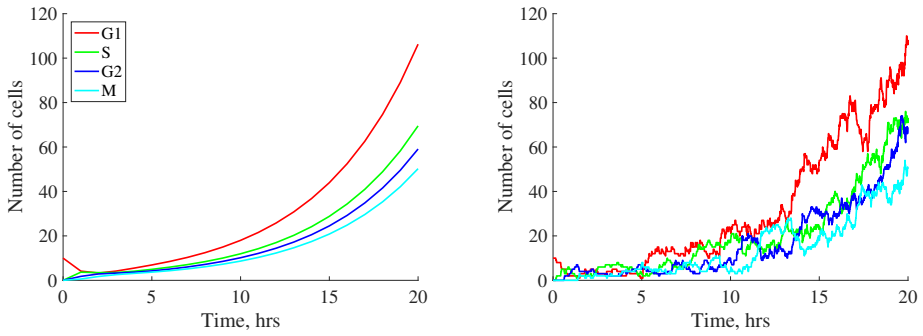

Figure 4.5: The evolution of cell number over time for the four-stage model, starting with 10 cells. Left: FSP with probability redistribution. Right: SSA.

Figures 4.4 and 4.5 demonstrate that the cell-population FSP methods accurately capture the protein and cell-stage dynamics for this population-growth model, where the SSA results are taken to represent a true account of the model's behaviours.

4.4.3 Cell death

Both methods support models in which cells in different states die at different rates. This is captured automatically by the $\tilde{u}_j$ of the redistribution method but requires a different usage of $\Delta(P)$ in the recycling method. If each state has a uniquely defined rate of death depending on its stage i and protein content n, $z_i(n)$, then the total flux through cell death is

$$Z(P) = \sum_{n=0}^{N_P} \sum_{i=0}^{N_S} z_i(n) P_{i,n}.$$

Each state then dies with a corrected rate

$$\tilde{z}_i(P, n) = z_i(n) + \Delta(P) - Z(P),$$

which is subtracted from the diagonal of the A matrix (replacing $\Delta(P)$ in Equation 4.8), and refunded to the null state through the bottom row in Equation 4.5. The A matrix otherwise remains the same.

This type of adaptation would be necessary for modelling the effects of bactericidal antibiotics on bacterial population growth, for example, where the effect of the toxin is to kill susceptible cells. Bacteriostatic antibiotics, on the other hand, prevent cell progression, growth, or division, and therefore could be modelled (at least in principle) without differential rates of cell death.

4.4.4 Cell-number distribution

The FSP methods presented are limited to models that can be written as $\dot{P} = A(P)P$; they could not implement a model of the form $\dot{P} = A(P_N, P)P$, where dynamics depend on the population size, which has some distribution described by P_N.

As an illustration, we construct a version of the four-stage cell-cycle model in which transcription diminishes as the total number of cells, N, increases:

$$\alpha_i(N) = 10\frac{40^4}{40^4 + N^4} \ \forall\, i. \tag{4.23}$$

The resulting protein dynamics are shown in Figure 4.6 (top, bottom left). Both the ODE and FSP models fail to capture protein dynamics due to averaging of the cell-number distribution. They predict a final cell number of 227; the simulations have final cell counts that range from 34 to 559 (Figure 4.6, bottom right).

Therefore, the methodology will need to be developed before it can be applied to a system in which cell number impacts dynamics. Such effects might be observed in a homeostatic system (Foley et al., 2006) or a population in which local cell density impacts growth rates, through higher-order organisation (Volfson et al., 2008), quorum sensing (Whitehead et al., 2001), or space or resource restriction (Shraiman, 2005).

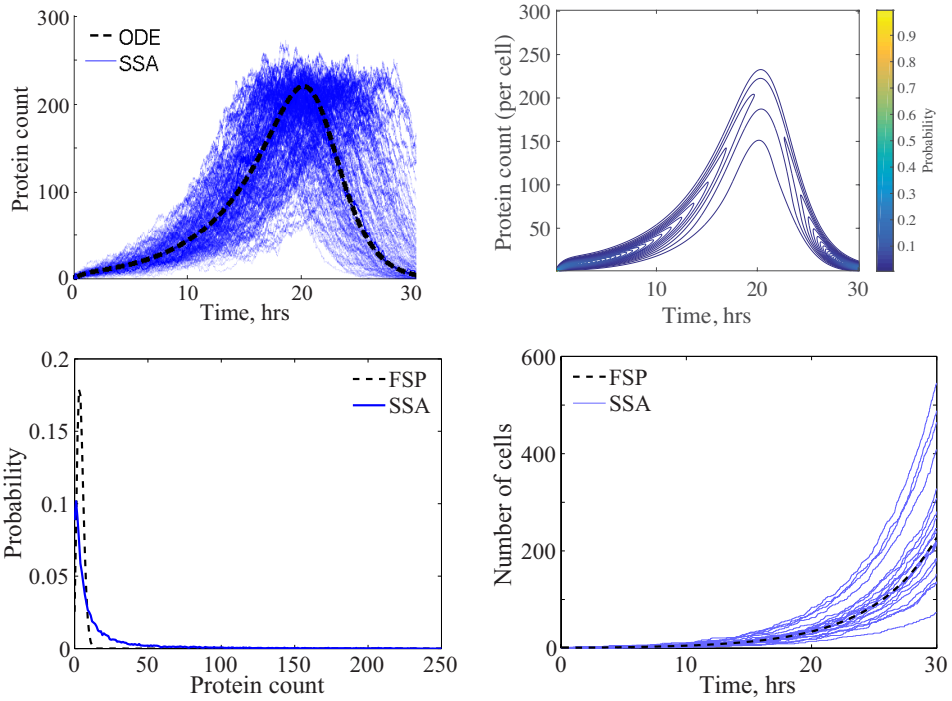

Figure 4.6: Results from a four-stage model in which protein dynamics depend on the total number of cells, N. All transition rates are 1 ($k_i = 1 \forall i$); protein decay is linear in n ($\gamma_i(n) = n \forall i$); and the rate of protein transcription declines as total cell number increases (see main text). Top: Total protein number. Left: ODE and SSA (150 trajectories). Right: redistribution FSP. Bottom left: protein number distribution at time $t = 30$. Black dashed: FSP. Blue: SSA, distribution over 20,000 simulations. The FSP mean is 4.1126 with standard deviation 2.2453, whereas the SSA mean is 17.1435 with standard deviation 31.019. Bottom right: Cell number over time. Black dashed: ODE and FSP. Blue: SSA results, 20 trajectories.

4.5 Summary

In summary, two methods have been presented that allow us to model population growth that depends on protein dynamics. They use the FSP (Munsky & Khammash, 2006), which captures noisy protein dynamics, allowing for complicated behaviours such as bimodality, which would be lost in a mean-value model. The methods are considerably quicker than performing a large number of simulations, as they require solving only a system of ODEs. Furthermore, a readout is directly given of the accuracy of the output (Munsky & Khammash, 2006).

Characterising growth rates for cells in different states, underwritten by a molecular basis for transitions between the states, will allow for a fuller picture of the range of possible growth dynamics that are driven by noisy protein dynamics. This method would find application in questions concerned with growth rates, fitness, and the effects of switching between states on population growth. In the following two chapters, the method will be applied to models of the cell cycle (Chapter 5) and a model of yeast population fitness in a toxic environment (Chapter 6).

5

Modelling p21-dependent cell cycling

In this chapter, the finite state projection method presented in Chapter 4 will be used to implement models of a protein-dependent cell cycle. Section 5.1 outlines the motivation for modelling the cell cycle with a dependence on protein, and Section 5.2 describes the biology behind the modelling choices made. Models are defined in Section 5.3 and implemented in Section 5.4. A summary and directions for future work are given in Section 5.5.

5.1 Introduction

The eukaryotic cell cycle is a fundamental determiner of cell population growth, and its dynamics are crucial to understanding various processes that operate at a cell-population level (Altinok et al., 2011). The cell cycle

is the regulated progression of a cell from its birth to division. The process has characteristic stages: a cell begins in stage G1, progresses to stage S, in which DNA is replicated, continues to stage G2, and then reaches stage M, at whose end it divides, giving rise to two cells in stage G1, and the cycle recommences (Figure 4.1) (Schafer, 1998). At division, proteins and cell organelles are partitioned randomly between the two daughter cells (Huh & Paulsson, 2011).

Transitions between stages are tightly controlled and unidirectional (Schafer, 1998). It is proposed that there exist checkpoints, beyond which a cell is committed to enter into the next stage of the cycle (Barnum & O'Connell, 2014). These checkpoints are governed by protein content and their mechanisms have not been fully elucidated to date (Barnum & O'Connell, 2014).

Abnormal cell cycle regulation and uncontrolled division are features of cancerous cells (Vermeulen et al., 2003; Malumbres & Barbacid, 2009), and population dynamics underpin the mechanism by which a cancerous cell establishes a neoplasm within a tissue (Bravo & Axelrod, 2013; MacLean et al., 2014). Thus models of the cell cycle are of particular relevance to cancer development (Bellomo et al., 2008) and tumour growth (Alarcón & Jensen, 2011).

Much work has been done to unify knowledge of cell-cycle progression into models (Tyson et al., 1995; Chiorino & Lupi, 2002; Altinok et al., 2011). Further, some have married models at the molecular level with those at the cellular level to model protein-dependent cell population growth (Bekkal Brikci et al., 2007, 2008, 2009; Borges et al., 2014).

Here, I will use the finite state projection method presented in Chapter 4 to model the full distribution over states for a protein-dependent cell cycle. Within this framework, I will make comparisons between two proposed models: one consisting of four states, and the other including a fifth state, G0, a quiescent equivalent of G1, reachable from stage M. In G0, the cell pauses its cycling, which recommences only via state G1. The four-stage model was introduced in Section 4.2.2 (Figure 4.1); the corresponding five-stage model is shown in Figure 5.1.

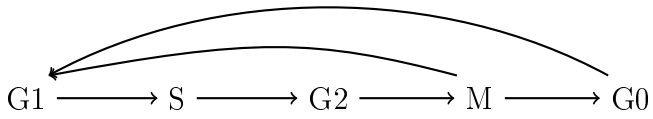

Figure 5.1: Five-stage cell-cycle model. Cells begin in stage G1, progress through stages S and G2 to M, and divide, giving rise to two cells that can be in either stage G1 or stage G0. G0 is the quiescent stage. From G0, cells transition into stage G1.

5.2 Biological background

Cell cycle progression is regulated by cyclin-dependent kinases (CDKs) (Malumbres & Barbacid, 2009). CDK2 activity was found to bifurcate at mitotic exit, marking cells for either quiescence or re-entry into the cell cycle (corresponding to G0 and G1, respectively) (Spencer et al., 2013).

In a follow-up study, it was established that this mechanism, and therefore cell fate, is determined by a CDK inhibitor, p21 (Overton et al., 2014). Low p21 (and increasing CDK2 activity) after mitosis is associated with re-entry

into the cell cycle, and high p21 (and low CDK2 activity) is associated with entry into a quiescent state (Spencer et al., 2013; Overton et al., 2014).

For the models developed in Section 5.3, p21 is chosen as the protein to model that determines cell cycle dynamics. The results presented by Spencer et al. (2013) and Overton et al. (2014) will underpin the construction and parametrisation of the models.

5.2.1 Protein dynamics

The amount of p21 in a cell varies over the course of the cell cycle (Amador et al., 2007). The dynamics result of the interplay between gene expression rate, protein decay rate and inheritance, and due to reaction rates varying across the different stages (Abbas & Dutta, 2009).

The models developed in Section 5.3 will employ constant p21 expression rates that are unique to each stage. The fine details of its regulation will be subsumed into overall expression levels characteristic of each stage (Abbas & Dutta, 2009). These rates will ultimately be fit to the data presented by Spencer et al. (2013).

In G1, p21 degradation is facilitated by SCF/Skp2 via CDK2 activity (Abbas & Dutta, 2009; Overton et al., 2014). The complementary inactivation of CDK2 by p21 results in bistability. Spencer et al. (2013) suggest that bifurcation occurs in stage M, as protein content appears to 'mark' cells out for a certain fate. A mechanism is not given, and the SCF/Skp2 complex is restricted to stage G1 (Amador et al., 2007). Instead, in stage M,

p21 degradation is facilitated by APC/C^{Cdc20}; this complex, activated by CDK1, ubiquitinates p21 (Amador et al., 2007; Abbas & Dutta, 2009), and p21 suppresses CDK1 (Kidokoro et al., 2008; Satyanarayana et al., 2008).

Therefore, the models will use two mechanisms for protein decay. The first is a linear protein decay rate, which occurs in all stages (Abbas & Dutta, 2009). The second emulates the mutual inhibition between CDK2 (or CDK1) activity and p21 that permits a bistable solution: low CDK activity and high p21, and high CDK activity and low p21 (Amador et al., 2007). This decay rate will be employed in stages M, G1 and G0, and is defined in Section 5.3.1.

5.2.2 Stage transitions

Duration in G1/0 correlates with cell-cycle duration (Spencer et al., 2013). This motivates the decision to model only stage transitions concerning G1 or G0 as dependent on protein, and all other transitions as protein independent, following precedents for a focus on the G1/S transition (Bekkal Brikci et al., 2009).

For a four-stage model (Figure 4.1), consisting of stages G1, S, G2 and M, the transition from G1 to S is protein dependent, where p21 slows transition, allowing for long sojourns in G1 that resemble quiescence. For a five-stage model (Figure 5.1), which includes the quiescent stage G0, the choice of fate at mitosis is protein dependent (as suggested by Spencer et al. (2013)), and the rate of transition from G0 to G1 is protein dependent. All other transitions are protein independent.

Other studies have considered transitions between G1 and G0 (Bekkal Brikci et al., 2009), and others still have limited these transitions to occur only once in a cell's lifetime (Borges et al., 2014). Here we implement a model of G0 being attainable only from phase M following insights from Spencer et al. (2013) and Overton et al. (2014), though to implement and compare another stage-transition structure within this framework would be a straightforward matter.

5.2.3 Mitosis

The model uses a binomial distribution to implement protein partition at mitosis (Huh & Paulsson, 2011). Given that protein content determines cell fate, and that 98% of sister cells share their fate (Spencer et al., 2013), then the number of proteins inherited will need to be on the same side of a dividing line in 98% of cases. The bimodal distribution over p21, as described above, in stage M would allow for this, as cells with high (low) p21 would likely divide to produce two cells with high (low) p21.

5.3 Models

Here I delineate the construction of the two models, one with four stages (Figure 4.1) and one with five (Figure 5.1). Protein dynamics depend on the cell stage and, in turn, determine cell-cycle transitions. Motivated by the biology outlined in Section 5.2, the mathematical formulation is presented below.

5.3.1 Four-stage model

For this model, the four stages are $\{S_0, S_1, S_2, S_3\} = \{\text{G1, S, G2, M}\}$, and the possible stage transitions are shown in Figure 4.1.

Protein dynamics

Protein production and degradation are parametrised by rates $\alpha(n)$ and $\gamma(n)$ respectively, where n is the protein count:

$$S_{i,n} \xrightarrow{\alpha_i(n)} S_{i,n+1} \quad i = 0, 1, 2, 3, \tag{5.1}$$

$$S_{i,n+1} \xrightarrow{\gamma_i(n+1)} S_{i,n} \quad i = 0, 1, 2, 3. \tag{5.2}$$

Whilst α and γ may be functions of both cell stage and protein number, I choose $\alpha(n) = [\alpha_0, \alpha_1, \alpha_2, \alpha_3]$ as a vector of constants so that each stage has its own rate of protein production. Noting the behaviour of p21 throughout the cell cycle (Amador et al., 2007), one expects its transcription rate to be highest in M and G1 and lowest in S and G2.

Decay rates $\gamma(n)$ are set as linear functions of protein number in stages S and G2: $\gamma_i(n) = n$ for $i \in \{1, 2\}$. For stage $S_0 = \text{G1}$, following Overton et al. (2014), facilitated degradation is defined as:

$$\gamma_0(n) = n\left(1 + \frac{\kappa_2 C_G^{\eta_2}}{\lambda_2^{\eta_2} + C_G^{\eta_2}}\right), \tag{5.3}$$

where $C_G = C_G(n)$ is the expected CDK2 activity in stage G1 given p21 (Overton et al., 2014):

$$C_G(n) = \frac{\kappa_1 F}{\kappa_3 n^{\eta_1} / ((\beta\lambda_1)^{\eta_1} + n^{\eta_1}) + \xi_c - \kappa_4} \tag{5.4}$$

Parameter meanings and values are given in Table 5.1.

For stage $S_3 = \mathrm{M}$, I use the same functional form as in Equation 5.3:

$$\gamma_3(n) = n\left(1 + \frac{\kappa_2 C_M^{\eta_2}}{\lambda_2^{\eta_2} + C_M^{\eta_2}}\right), \tag{5.5}$$

where $C_M = C_M(n)$ is the expected CDK1 activity given p21, following Overton et al. (2014):

$$C_M(n) = \frac{\kappa_1 F}{\kappa_3 n^{\eta_1} / ((\beta\lambda_M)^{\eta_1} + n^{\eta_1}) + \xi_c - \kappa_4}, \tag{5.6}$$

where $\lambda_M = 2\lambda_1$, as in stage M there is approximately twice the protein complement as in stage G1.

Stage transitions

Cell-cycle progression is dependent on protein number. Spencer et al. (2013) report that the majority of variability in cell-cycle duration is attributable to variability in M-to-S duration. Therefore, in these models, all transitions are protein independent with the exception of the G1/S transition. As G1 exit depends on high CDK2 activity, and CDK2 activity is inhibited by p21, the model requires low p21 for transition from G1 to S.

The rates are defined as follows:

$$k_0(n) = K_{G1\to S} = \frac{\beta^h}{\beta^h + n^h} V_{max}, \tag{5.7}$$

$$k_1 = K_{S\to G2} = V_{max}, \tag{5.8}$$

$$k_2 = K_{G2\to M} = V_{max}, \tag{5.9}$$

$$k_3 = K_{M\to G1} = 2V_{max}. \tag{5.10}$$

Mitosis

Upon division, proteins are partitioned between the two daughter cells according to the binomial distribution. Then, the rate of exit from cells in stage M with n proteins is k_3, and the rate of entry into stage G1 with n^* proteins is

$$g_0(n, n^*) = \binom{n}{n^*} \frac{1}{2^n}. \tag{5.11}$$

The A matrix

Using the recycling FSP method, the A matrix has the form given in Section 4.2.2. Parameter values and definitions are given in Table 5.1.

5.3.2 Five-stage model

A fifth stage, G0, is introduced, which is the quiescent equivalent of stage G1. Possible transitions are shown in Figure 5.1. The new stage, G0, can be attained from mitosis of cells in stage M. Cells in stage G0 transition into stage G1. The stages are denoted $\{S_0, S_1, S_2, S_3, S_4\} = \{$G1, S, G2, M, G0$\}$.

Protein dynamics

Protein dynamics depend on the stage of the cell, i, as in Equations 5.1 and 5.2. For transcription, again $\alpha = [\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4]$ is a vector of constants, and $\gamma_i(n) = n$ for $i \in \{1, 2\}$. Rates for degradation in stages G1 and G0 ($\gamma_0(n)$ and $\gamma_4(n)$) are as in Equation 5.3, and Equation 5.5 for stage M ($\gamma_3(n)$).

Table 5.1: Parameters and functions governing cell stage transitions and the relationships between CDK activity and p21 content for the four-stage and five-stage models of the cell cycle.

Symbol	Description	Value
n	p21	Equation 5.1
C_G	CDK2 activity in stage G1/0	Equation 5.4
C_M	CDK1 activity in stage M	Equation 5.6
F	Growth factor concentration	1.75
α	Rate of p21 transcription	See text
γ	Rate of p21 decay	See text
κ_1	Rate of CDK activation	1 s^{-1}
κ_2	Rate of p21 inhibition by CDK	8 s^{-1}
κ_3	Rate of p21 inhibition of CDK	10 s^{-1}
κ_4	CDK positive feedback rate	0.25 s^{-1}
η_1	Hill coefficient for p21 inhibition by CDK	4
η_2	Hill coefficient for p21 inhibition of CDK	4
λ_1	Effective affinity constant for the inhibition of CDK2 by p21 in G stages	1 nM
λ_M	Effective affinity constant for the inhibition of CDK1 by p21 in M stage	2 nM
λ_2	Effective affinity constant for the inhibition of p21 by CDK	1 nM
ξ_c	Inactivation rate of CDK	1 s^{-1}
β	Position of bifurcation	35
g_0	Rate of transition from stage M to stage G1	Equation 5.11 or 5.23
g_4	Rate of transition from stage M to stage G0	Equation 5.24
h	Hill coefficient for p21-dependent stage transition	2
k_i	Rate of transition from cell stage i	See text
V_{max}	Maximum rate of stage transition	1 s^{-1}

Stage transitions

The possible stage transitions, not including division, are:

$$S_i \xrightarrow{k_i(n)} S_{i+1} \quad i = 0, 1, 2 \tag{5.12}$$

$$S_4 \xrightarrow{k_4(n)} S_0. \tag{5.13}$$

They are parametrised, as before, as

$$k_0 = K_{G1\to S} = V_{max}, \tag{5.14}$$

$$k_1 = K_{S\to G2} = V_{max}, \tag{5.15}$$

$$k_2 = K_{G2\to M} = V_{max}, \tag{5.16}$$

$$k_3 = K_{M\to G1/0} = 2V_{max}, \tag{5.17}$$

$$k_4(n) = K_{G0\to G1} = \frac{\beta^h}{\beta^h + n^h} V_{max}, \tag{5.18}$$

where the last rate function reflects that high p21 holds the cell in the quiescent stage.

Mitosis

Upon mitosis, cells exit stage M at rate k_3 and enter stage G1 or the quiescent stage G0:

$$S_3 \xrightarrow{k_3} S_i \quad i = 0, 4. \tag{5.19}$$

The destination state depends on some function of the amount of protein of the parent, n, and the amount they receive at partitioning, n^*:

$$S_{3,n} \xrightarrow{k_3 g_i(n,n^*)} S_{i,n^*} \quad i = 0, 4. \tag{5.20}$$

The functions g_i must sum to 1 (so that all cells that leave M enter either G0 or G1) and their sum must be symmetric (so that for each cell with n proteins, the resultant cells with n^* and $n - n^*$ proteins respectively have equal probability):

$$\sum_{i=0,4} \sum_{n^*=0}^{n} g_i(n, n^*) = 1, \tag{5.21}$$

$$g_0(n, n^*) + g_4(n, n^*) = g_0(n, n - n^*) + g_4(n, n - n^*). \tag{5.22}$$

Following the principle that, at division, daughter cells enter a stage based on protein number, the functions g_i are defined according to the expectation that low-p21 cells enter stage G1 and high-p21 cells enter stage G0 (Spencer et al., 2013). A simple implementation would be that any daughters with $n^* > \beta$ go into G0 and all others go into G1:

$$g_0(n, n^*) = \begin{cases} \binom{n}{n^*}\frac{1}{2^n} & n^* \leq \beta \\ 0 & n^* > \beta \end{cases} \tag{5.23}$$

$$g_4(n, n^*) = \begin{cases} 0 & n^* \leq \beta \\ \binom{n}{n^*}\frac{1}{2^n} & n^* > \beta. \end{cases} \tag{5.24}$$

The A matrix

The matrix A is as shown in Equation 4.5, with $\mathbf{b}_n(P) = [0\,0\,0\,\delta_n(P)\,0]^T$, $\mathbf{c}_\Delta(P) = [\Delta(P)\,\Delta(P)\,\Delta(P)\,\Delta(P)\,\Delta(P)]$, and B_n and D_n equivalent five-by-five matrices. Its component parts, Equations 5.25 and 5.26, show cell transitions for a five-stage system: forward transitions that occur at rate k_i, and mitosis events that occur at rate k_3 and result in stage i with probability g_i.

$$A_n = \begin{pmatrix} -k_0 & 0 & 0 & k_3 g_0(n,n) & k_4 \\ k_0 & -k_1 & 0 & 0 & 0 \\ 0 & k_1 & -k_2 & 0 & 0 \\ 0 & 0 & k_2 & -k_3 & 0 \\ 0 & 0 & 0 & k_3 g_4(n,n) & -k_4 \end{pmatrix} - \mathrm{diag}(\Delta(P)) \tag{5.25}$$

and

$$A_{n^*,n} = \begin{pmatrix} 0 & 0 & 0 & k_3 g_0(n, n^*) & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & k_3 g_4(n, n^*) & 0 \end{pmatrix}. \tag{5.26}$$

5.4 Results

The cell cycle is modelled using the recycling method of Chapter 4. Expression rates (α) are fit for both the four-stage and the five-stage models to the data reported by Spencer et al. (2013), demonstrating that these simple models are sufficient to capture the dynamic behaviour of this population of cells.

5.4.1 Four-stage model

The model given in Section 5.3.1 and parameters in Table 5.1 were used to fit values for α. The parameter space over α was searched in order to minimise the L_1 distance between its predicted metrics and the data. The values that minimised this distance were $\alpha = [129.7, 22.8, 1.76, 204.2]$. The results of the fitting are given in Table 5.2. The proportions of cells in each stage are given in Table 5.3. While only point estimates, the values for α align with the expectation in Section 5.3.1 that expression should be higher in stages G1 and M.

Bifurcation of p21 in G1 and M leads to some cells with high p21 content

in stage G1, which corresponds to being arrested (Figure 5.2, left). Due to the bifurcation in stage M, there is a correlation between daughter cells in terms of p21 content as they exit mitosis (Figure 5.3, left).

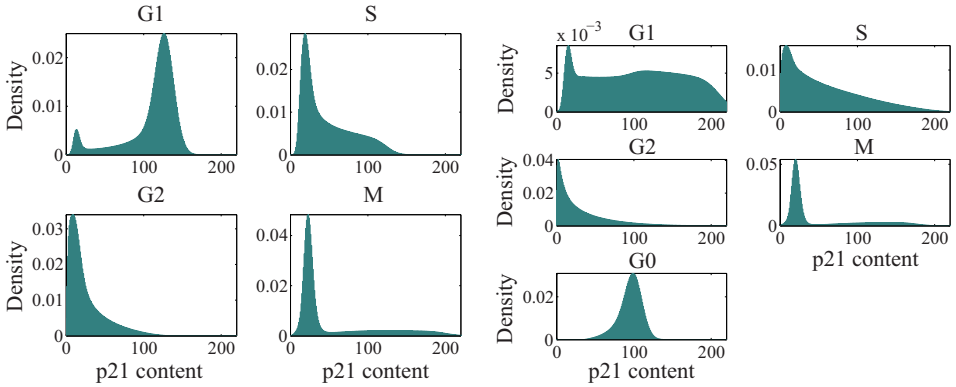

Figure 5.2: Protein distributions in each stage for the four-stage model (left) and the five-stage model (right) with bifurcations in stages G1, G0 and M.

5.4.2 Five-stage model

For the five-stage model, distance minimisation yields optimal values for expression as $\alpha = [215.8, 1.020, 0.797, 179.9, 103.3]$. The results of the fitting are given in Table 5.2.

The resulting protein distributions as shown in Figure 5.2 (right) and stage distributions in Table 5.3. The correlation between daughter cells in terms of p21 content as they exit mitosis is shown in Figure 5.3 (right).

5.4.3 Discussion

The results show that the four-stage model results in the G1 stage being highly occupied; more so than stages G1 and G0 combined for the five-stage

Table 5.2: Cell-division distributions for models with bifurcations in stages G0, G1 and M. The columns show the percentage of daughter pairs in each class at time $t = 30$, where the classes are 'low p21' and 'high p21', and two daughter cells can be in the same class or in different classes. Data are from Spencer et al. (2013).

Class	Low p21, %	High p21, %	Same class, %	Different classes, %
Four stages	75.0822	24.9178	97.8572	2.1428
Five stages	74.9972	25.0028	98.0033	1.9967
Data	75.00	25.00	98.00	2.00

Table 5.3: Cell-stage occupations for models with bifurcations in stages G0, G1 and M. The columns show the proportion of cells in each state at time $t = 30$.

Cell stage	G1	S	G2	M	G0
Four stages	0.7319	0.0965	0.0892	0.0824	–
Five stages	0.2220	0.1936	0.1687	0.1471	0.2686

model (Table 5.3). The distributions show that the high rate of expression of p21 in M stage (Figure 5.2) gives rise to a broad distribution over high-p21 levels in daughter cells for both models (Figure 5.3).

The results presented are point estimates for values of α that minimise the distance between the data and the predictions. No information is gained on the distribution of values that would support a similarly good fit to the data; nor is there any confirmation that the minimum obtained is a global, rather than local, minimum. In fact, many values for α give rise to similarly good fits to the data. The optimal values given above are not to be understood as 'true' sets of values, but rather demonstrate that a set of values exist to provide a good fit. A full exploration of the posterior, using

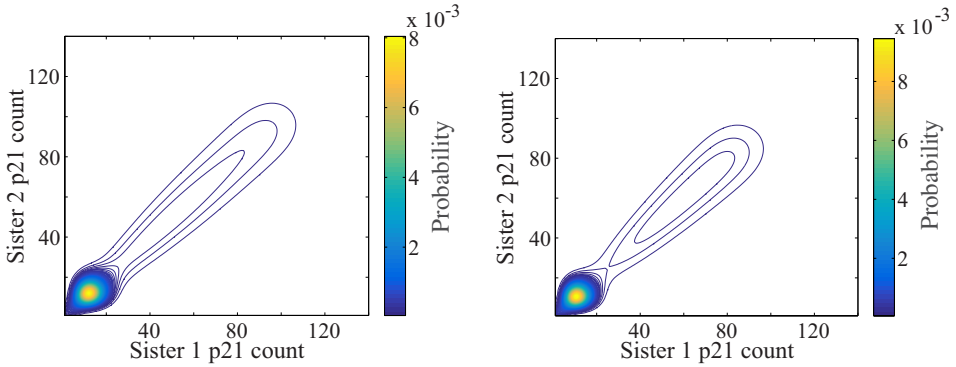

Figure 5.3: p21 distributions for daughter-cell pairs resulting of the four-stage model (left) and the five-stage model (right) with bifurcations in stages G1, G0 and M.

a likelihood-based method, would elucidate the spread of support for the model, and allow more concrete statements to be made about parameter sets.

Raw data concerning cell-stage durations and protein counts at cell division, rather than summary statistics, would allow for the definition of a likelihood function. From this it would be possible to infer a distribution over parameters that fit the data well, and therefore indicate a distribution over likely behaviours that an experimentalist might observe.

With more definition in data, it would also be possible to infer other parameters in the model. For example, it might be the case that the four-stage model favours the G1 state due to the transition rates chosen, and that other choices would find similarly good fits to the data alongside different distributions over stage occupations.

5.5 Summary & future work

The method of Chapter 4 has been applied to the cell cycle, where p21 dynamics are bimodal and determine re-entry into the cell cycle. Bifurcation of p21 due to CDK2 activity has been implemented in stages G1 and G0, and that between p21 and CDK1 in stage M.

Following a number of practical and qualitative assumptions, the model was able to fit the data. One such simplification is the assumption that CDK activity reacts spontaneously to p21 content, as implemented in Equations 5.4 and 5.6. The distribution that results, according to the p21 distributions in Figure 5.3, is bimodal, whereas data suggest that the spectrum of activity is continuous (Spencer et al., 2013). An improvement to the model might be to include CDK dynamics explicitly, or to model CDK activity in Equations 5.4 and 5.6 as random variables.

Additional information on the expression and degradation patterns of p21 throughout the cell cycle would assist the functional parametrisation of the model (Abbas & Dutta, 2009). For simplicity, zero-th order reactions were employed for protein expression, and first-order reactions for base-level degradation. These functions yielded satisfactory results without superfluous complexity.

The observation that bifurcation occurs in stage M (Spencer et al., 2013) prompts the question of how to implement the mechanism. Here, it has been assumed that CDK1 performs this function (Amador et al., 2007; Satyanarayana et al., 2008; Abbas & Dutta, 2009). The mechanism em-

ployed was the same as that for stage G1, but should be expounded with its own unique form and dynamics that reflect the biology. A model including this mechanism in stage G2 should also be considered (Abbas & Dutta, 2009).

In terms of model development, it may become pertinent to employ differential cell-death rates. For example, in their cellular automaton model, Altinok et al. (2011) define stages G1 and G2 as those that can commit to apoptosis.

In this chapter, the data to which the parameters were fit were cumulative densities (Spencer et al., 2013). With raw data, it would be possible to define a likelihood function, allowing parameters to be inferred and models to be compared via their evidences using Sysbions (Chapter 3). It would then be possible to ask which models of the cell cycle are best supported by the data.

6

Modelling antibiotic resistance in yeast

In this chapter, the finite state projection method presented in Chapter 4 will be used to implement a model of antibiotic resistance in a population of yeast cells. Section 6.1 outlines the motivation for modelling antibiotic resistance in a population, and Section 6.2 describes the biological results behind the modelling choices made. The model is defined in Section 6.3 and implemented in Section 6.4. A summary and directions for future work are given in Section 6.5.

6.1 Introduction

Antimicrobial resistance is the evolution in microbial species of an ability to withstand toxins that previously inhibited their growth, division or survival

(The review on antimicrobial resistance, 2016). Resistance emerges in an individual through mutation or horizontal gene transfer and spreads in a population through cell divisions (Allen & Waclaw, 2016).

How a gene within a genetically heterogeneous population comes to predominate in the population is a question for population genetics (MacLean et al., 2010). An important component of these dynamics is how variability within an isogenic population influences the emergence of resistance and the durability of a resistant population (de Jong et al., 2011). The ability of noisy protein dynamics to give rise to different states of resistance can be exploited by populations to survive and adapt to unfavourable environments, by resorting to persister populations (Allen & Waclaw, 2016) that support genetic mutability (Tadrowski et al., 2016).

The existence of multiple phenotypic states within isogenic populations is attributed to evolutionary adaptations to unpredictable environments (Thattai & van Oudenaarden, 2004; de Jong et al., 2011; Shahrezaei & Marguerat, 2015). The heritable trait is multimodality in protein dynamics, and the fitness associated with the trait belongs to (e.g. is averaged over) the whole population (Nevozhay et al., 2012). Theoretical and experimental studies have suggested that the rates of switching between states, driven by gene expression, optimise the population's fitness when they match the rate of environmental switching (Thattai & van Oudenaarden, 2004; Kussell & Leibler, 2005; Acar et al., 2008).

A contrasting view is presented by Nevozhay et al. (2012), who show that the optimal survival strategy of a population of yeast cells depends on the

interplay between protein dynamics and the fitness landscape, regardless of fluctuations in the environment. In their study, the authors engineer yeast cells capable of transitioning between distinct states of antibiotic resistance, where resistance allows cell division to proceed in the presence of antibiotic. However, the authors show that there is also a cost to antibiotic resistance: in the absence of antibiotic, occupying the resistant state leads to a decrease in fitness relative to the non-resistant state.

In their experimental set-up, resistance is conferred by expression of a resistance gene, which can be either 'high' or 'low'. High expression corresponds to resistance and low expression to vulnerability to antibiotic. Their synthetic gene circuit was constructed to permit the experimenters to induce occupation of the higher expression level and explore what level of induction maximises population fitness. Their finding was that the cell population as a whole had greatest fitness in antibiotic not when the gene was constitutively expressed, but at the lowest level of induction that allows the high expression state to be occupied, which they term the "sweet spot" (Nevozhay et al., 2012).

The authors presented a molecular model intended to capture the fitness costs associated with the antibiotic and expression of the resistance gene, and found that it was unable to reproduce the "sweet spot" of induction. Here, I show that a molecular description similar to that presented in Nevozhay et al. (2012), implemented using the population-growth modelling method presented in Chapter 4, may permit the observed behaviour giving rise to a "sweet spot" at a low level of induction and a penalty at

higher levels.

6.2 Biological background

In this section, material from Nevozhay et al. (2012) relevant to the mathematical model constructed in Section 6.3 is presented. It concerns an isogenic population of yeast cells containing a synthetic gene circuit.

The fitness of the population was measured in different conditions, where the fitness is the overall rate of population division, constructed as the sum over the fitnesses of all individuals in the population. Each individual's fitness is a function of their protein level and their environment (Nevozhay et al., 2012). The model therefore aims to capture how protein dynamics and their dependent cell division rates cause the population to develop.

6.2.1 Protein dynamics

The synthetic gene circuit consists of an *rtTA* gene and a *ZeoR* gene with identical promoters. These genes have the same base level of expression and are both activated by an external inducer, ATc, interacting with the rtTA protein. Their rate of expression is thus the same, with the inducer leading to high expression of both genes when the rtTA level crosses some threshold. The rate of protein decay and dilution of rtTA far exceeds that of ZeoR (Nevozhay et al., 2012).

Together, these dynamics result in bimodal distributions of rtTA and ZeoR. The two stable steady states have either low expression of both genes or

high expression of both genes. The amount of inducer provided by the experimenter determines the propensity for an individual cell to be in a given state and, therefore, viewed over the population, the proportions of cells in each state.

The ZeoR protein confers resistance to the antibiotic, Zeocin, and is labelled with green fluorescent protein. Measurements of its intensity form the basis for the analysis in Nevozhay et al. (2012). A schematic of the interactions between ATc, rtTA, ZeoR and Zeocin are shown in Figure 6.1.

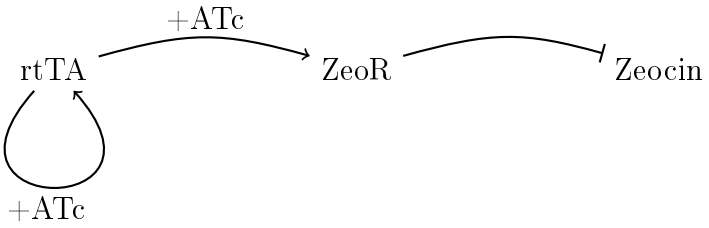

Figure 6.1: Model of interactions between inducer ATc, proteins rtTA and ZeoR, and antibiotic Zeocin. The genes for proteins rtTA and ZeoR are both activated by the action of rtTA with ATc. ZeoR confers resistance to the antibiotic Zeocin.

6.2.2 Cell population growth

Nevozhay et al. (2012) use the synthetic gene circuit to study the fitness of yeast populations in different conditions. In the absence of inducer, fitness decreases in the presence of increasing antibiotic. Similarly, in the absence of antibiotic, fitness decreases as the amount of inducer increases, which is attributed to a number of factors, including the sequestration of transcriptional machinery by rtTA (Baron et al., 1997).

Note that the effects of both the antibiotic, which is bacteriostatic, and the

inducer are to reduce the rate of cell division. Neither causes cells to die, but to become less likely to divide, almost entirely in the case of excess antibiotic (Nevozhay et al., 2012).

6.2.3 Cellular memory

The observations of fitness in antibiotic alone and inducer alone form the basis of the expectation that in the presence of antibiotic, increased inducer will increase fitness. However, the authors find that the optimal level of induction to maximise fitness in the presence of antibiotic is not high expression in every cell. Instead, the optimal level of induction is the minimum at which bimodality occurs (Nevozhay et al., 2012).

The fundamental insight of Nevozhay et al. (2012) is that the observed fitness behaviour is made possible by the mechanism of cellular memory. Cellular memory is here defined as the duration for which a particular phenotypic state is retained in a constant environment. Concerning gene expression states, it is the inverse of the spontaneous switching rate (Acar et al., 2008). This mechanism is not present in their molecular model and explains its failure to capture the observed behaviour. The model considers one variable (ZeoR), which is responsible for both antibiotic resistance and the fitness cost to the cell.

In this work, I will model these two factors separately. The cost to fitness is due to the activator, rtTA. Resistance is given by ZeoR. It is the resistance protein that holds memory. By having a much longer half life than rtTA, ZeoR is able to continue to confer resistance once in a high expression state

even when rtTA has dropped to a low expression level.

This is the sweet spot at low inducer concentrations: rtTA is low, and thus the fitness cost is low, because it has a propensity to be in the low state due to low inducer concentration. ZeoR can simultaneously be high due to rare fluctuations in rtTA that lift its expression level and its slow decay rate that causes it to remain there. In this transient position, the division rate is high. This mechanism is displayed graphically in Figure 6.2.

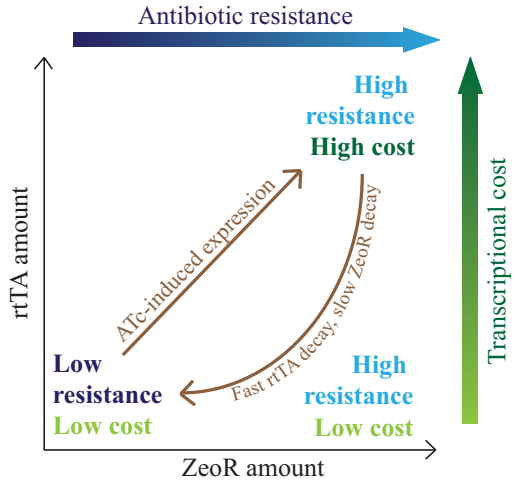

Figure 6.2: The relationships between the proteins rtTA and ZeoR and fitness. When rtTA and ZeoR are low, the cost to fitness due to rtTA transcription is low, but so is antibiotic resistance. With expression of the *rtTA* and *ZeoR* genes, the high-rtTA, high-ZeoR state is reached. This has high resistance to antibiotic, but a high cost due to rtTA. Through fast decay of rtTA and slow decay of ZeoR, the cell transiently visits the site of maximal fitness.

6.2.4 Data

The data to which the model is fit in Section 6.4 are taken from Nevozhay et al. (2012). The measurements are of population fitness and of the proportion of the cells in the low-expressing state in terms of fluorescence.

The populations were maintained at an exponential rate of growth and measurements were taken once a steady state was reached. The data are summarised in Table 6.1.

Table 6.1: Yeast fitness data from Nevozhay et al. (2012). The inducer, ATc, and the antibiotic Zeocin, Z, are the experimental inputs. Fitness is the average number of cell divisions per hour for the whole population. N_L is the proportion of cells in the low-expression state.

	ATc, ng/ml	Z, mg/ml	Fitness, h^{-1}	N_L
$Z = 0$	1.0	0.0	0.2181	
	2.5	0.0	0.2174	
	5.0	0.0	0.2115	0.8707
	10.0	0.0	0.1846	0.6041
	15.0	0.0		0.3638
	20.0	0.0	0.1527	0.2268
	25.0	0.0		0.1457
	30.0	0.0		0.1069
	50.0	0.0		0.0294
$Z = 2$	1.0	2.0	0.1885	
	2.5	2.0	0.1739	
	5.0	2.0	0.1502	
	10.0	2.0	0.1346	
	20.0	2.0	0.1247	
ATc=0	0.0	0.25	0.2371	
	0.0	0.5	0.2215	
	0.0	1.0	0.1806	
	0.0	2.0	0.0571	

6.3 Mathematical model

Given the motivations outlined above, this section enumerates how the model is formulated mathematically. The two state variables are rtTA number, x, and ZeoR number, y. The external conditions are the amounts

of inducer and antibiotic supplied.

6.3.1 Protein dynamics

In the synthetic gene circuit, the genes have identical promoters. They have base-level expression l. Upon binding the inducer ATc, the rtTA protein activates its own gene's expression and that of the Zeocin resistance gene, following a function $F(x)$. Thus both genes have the same pattern of expression (Nevozhay et al., 2012).

Both proteins have first-order decay rates, with that of rtTA (d_P) exceeding that of ZeoR (d_R) (Nevozhay et al., 2012). Function definitions, parameter values and their meanings are given in Table 6.2.

6.3.2 Cell population growth

Cell population growth is modelled exponentially (i.e. without stages or delays). Dynamics of cell growth are also excluded, and thus so is partitioning: each state divides to give rise to two of itself (rather than two cells with half the protein content).

There is a base-level fitness for cell replication, g_0, and this value is corrected according to protein content and environmental conditions (Nevozhay et al., 2012). The rate of division for a cell is defined as

$$m(x, y, C, Z) = g_0\gamma_1(y, Z)\gamma_2(x, C), \tag{6.1}$$

where x is the count of rtTA, y is the count of ZeoR, C is a multiple of inducer ATc concentration, and Z is extracellular Zeocin concentration.

Table 6.2: Parameters and functions governing cell growth and transition between phenotypes. Superscripts in the descriptions indicate the source of the value of the parameter: whether they were fit from data or taken from Nevozhay et al. (2012).

Symbol	Description	Value
x	rtTA	
y	ZeoR	
ATc	Inducer	0–20 ng/ml
C	Measure of ATc[1]	43.5845[ATc]
Z	Extracellular Zeocin	0–2 mg/ml
d_P	Rate of rtTA decay[2]	2.3404 h^{-1}
d_R	Rate of ZeoR decay[2]	0.25 h^{-1}
$k(x)$	Rate of rtTA and ZeoR transcription	$aF(x)+l$
a	Rate of activation[2]	180 h^{-1}
l	Base rate of transcription[2]	2.5 h^{-1}
$F(x)$	Activation of transcription with ATc	$\frac{v(x)^n}{v(x)^n+\varphi^n}$
n	Hill coefficient[2]	2
φ	rtTA activation point[1]	22.6281
$v(x)$	Amount of ATc-bound rtTA	$\frac{1}{2}\left(\frac{C}{d_P}+\frac{h}{b}+x\right) -\frac{1}{2}\sqrt{\left(\frac{C}{d_P}+\frac{h}{b}+x\right)^2-\frac{4Cx}{d_P}}$
h	Inducer escape[2]	1 h^{-1}
b	Inducer–repressor association rate[2]	3 h^{-1}
g_0	Maximum fitness[2]	0.2473 h^{-1}
$m(x,y,C,Z)$	Corrected fitness	$g_0\gamma_1(y,Z)\gamma_2(x,C)$
$\gamma_1(y,Z)$	Fitness cost due to Zeocin	$\frac{\chi}{Z_i(y,Z)+\chi}$
$\gamma_2(x,C)$	Fitness cost due to rtTA	$\frac{\alpha}{\alpha+\frac{xC}{C+\beta}}$
$Z_i(y,Z)$	Intracellular Zeocin	$\frac{\kappa sZ-d_Rh_z-d_Rsy}{2h_zs}+ \frac{\sqrt{(\kappa sZ-d_Rh_z-d_Rsy)^2+4h_z\kappa sd_RZ}}{2h_zs}$
χ	DNA repair rate over damage accumulation[3]	3.383e-5 mg/ml
α	Rate of fitness loss[1]	62.0266
β	Binding efficiency[1]	154.41 ng/ml
κ	Zeocin diffusion into cell[3]	1.323 mg/ml h^{-1}
h_z	Zeocin diffusion out of cell[2]	0.5 mg/ml h^{-1}
s	Binding affinity[3]	$1.0301/\chi$ h^{-1}

[1] Fit from ATc>0 data; [2] Nevozhay et al. (2012);
[3] Fit from ATc=0 data

The fitness cost associated with the antibiotic is given by $\gamma_1(y, Z)$ and that with induction by $\gamma_2(x, C)$. In Nevozhay et al. (2012), both γ_1 and γ_2 are functions of fluorescence, and therefore ZeoR. Here, these functions are instead defined as

$$\gamma_1(y, Z) = \frac{\chi}{Z_i(y, Z) + \chi} \tag{6.2}$$

and

$$\gamma_2(x, C) = \frac{\alpha}{\alpha + \frac{xC}{C+\beta}} \tag{6.3}$$

where

$$Z_i(y, Z) = \frac{sZ - d_R h_z - d_R sy + \sqrt{(sZ - d_R h_z - d_R sy)^2 + 4h_z s d_R Z}}{2h_z s} \tag{6.4}$$

is intracellular Zeocin that is not bound by ZeoR. Parameter values and meanings are given in Table 6.2.

6.3.3 Implementation with the FSP

The recycling method of Chapter 4 is used to model cell population growth dynamics that depend on protein. All cells divide to reproduce themselves, and mitosis is implemented by refunding each state the total mass that goes through division. This amount is

$$\delta_{x,y}(P) = m(x, y, C, Z)P_{x,y}, \tag{6.5}$$

where $P_{x,y}$ the probability to be in state (x, y). Then the total amount to refund is

$$\Delta(P) = \sum_{x=0}^{N_P} \sum_{y=0}^{N_R} \delta_{x,y}(P) \tag{6.6}$$

$$= \sum_{x=0}^{N_P} \sum_{y=0}^{N_R} m(x, y, C, Z)P_{x,y}, \tag{6.7}$$

where there are up to N_P rtTA proteins and N_R ZeoR proteins.

The total number of states is $N_T = (N_P + 1)(N_R + 1)$. For a cell in state j, the number of ZeoR proteins is $j_R = \lfloor \frac{j-1}{N_R+1} \rfloor$ and the number of rtTA proteins $j_P = (j-1) \bmod (N_R + 1)$. Then the matrix A is defined:

$$A_{ij} = \begin{cases} -2k(j_P) - d_P j_P - d_R j_R - \Delta(P) & i = j, i \neq N_T + 1 \\ k(j_P) & i = j + 1, j_P \neq N_P \\ k(j_P) & i = j + N_P + 1, j_R \neq N_R \\ d_P j_P & i = j - 1, j_P \neq 0 \\ d_R j_R & i = j - N_P - 1, j_R \neq 0 \\ -\Delta(P) & i = j = N_T + 1 \\ \delta_{i_P, i_R}(P) & j = N_T + 1, i \neq j \\ \Delta(P) & i = N_T + 1, i \neq j \end{cases} \tag{6.8}$$

or, equivalently,

$$A = \begin{pmatrix} A_0 - B_0 & D_1 & \cdots & 0 & \mathbf{b}_0(P) \\ B_0 & A_1 - B_1 - D_1 & \cdots & 0 & \mathbf{b}_1(P) \\ 0 & B_1 & \cdots & 0 & \mathbf{b}_2(P) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & A_{N_R} - B_{N_R} - D_{N_R} & \mathbf{b}_{N_R}(P) \\ \mathbf{c}_\Delta(P) & \mathbf{c}_\Delta(P) & \cdots & \mathbf{c}_\Delta(P) & -\Delta(P) \end{pmatrix}, \tag{6.9}$$

where $\mathbf{c}_\Delta(P) = \mathbf{1}\Delta(P)$ and $\mathbf{1}$ is a vector of ones of length $N_P + 1$, $\mathbf{b}_y(P) = [\delta_{0,y}(P)\, \delta_{1,y}(P) \cdots \delta_{N_P,y}(P)]^T$, and each A_y is an $N_P + 1$ by $N_P + 1$ matrix

defined as

$$A_y = \begin{pmatrix} -2k(0) - d_R y & d_P & \cdots & 0 \\ k(0) & -2k(1) - d_P - d_R y & \cdots & 0 \\ 0 & k(1) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -2k(N_P) - N_P d_P - d_R y \end{pmatrix} - \operatorname{diag}(\Delta(P)). \quad (6.10)$$

Finally,

$$B_y = \begin{pmatrix} k(0) & 0 & 0 & \cdots & 0 \\ 0 & k(1) & 0 & \cdots & 0 \\ 0 & 0 & k(2) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & k(N_P) \end{pmatrix}, \quad (6.11)$$

and

$$D_y = \begin{pmatrix} d_R y & 0 & 0 & \cdots & 0 \\ 0 & d_R y & 0 & \cdots & 0 \\ 0 & 0 & d_R y & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & d_R y \end{pmatrix} \quad (6.12)$$

give the rates of ZeoR creation and degradation, respectively. Functions and rates are given in Table 6.2.

6.3.4 Parameters

In this work, parameters used for the model follow the choices of Nevozhay et al. (2012). Some adjustments are necessary and, due to the differences

between the models, some are reinterpreted.

Nevozhay et al. (2012) define $C = 7[ATc]$, stating that, for this system, bifurcation occurs at ATc = 1 ng/ml. A value for φ is not given. However, it seems that bifurcation occurs for no value of φ using the parameters defined (Figure 6.3).

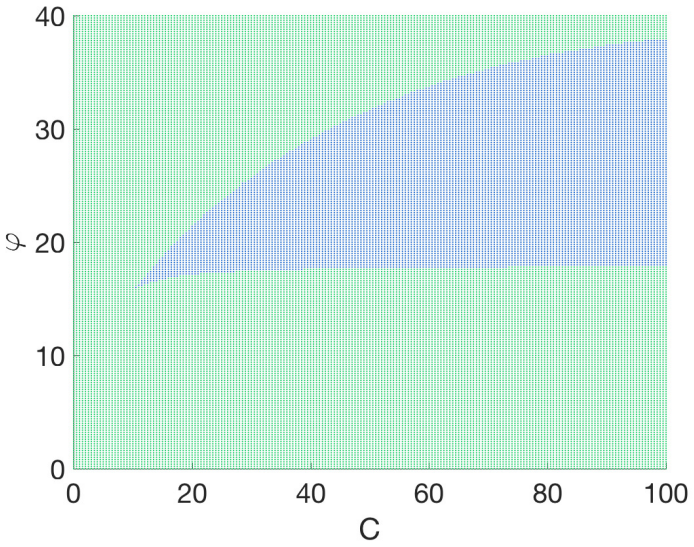

Figure 6.3: Combinations of parameters C and φ that permit one (green) or two (blue) stable steady states.

Since α and β were fit assuming that fluorescence is indicative of presence of rtTA (Nevozhay et al., 2012), whereas in this model fluorescence is indicative of ZeoR alone, I aim to fit these parameters anew. Also, φ is inferred, and the assumption is made that C takes the value at which bifurcation occurs in Figure 6.3.

These four parameters have no bearing on results obtained for experiments with ATc = 0. Therefore, these four data points (Table 6.1, bottom four

rows) are used first to fit parameters χ, κ and s, which are fixed for fitting the ATc > 0 data.

6.4 Results

Using the model outlined above, with functions and parameters based on those of Nevozhay et al. (2012), a reasonable fit to some data can be found (Figure 6.4). For these results, a distance-minimisation procedure was used to fit α, β, χ, κ, s, C and φ, with results given in Table 6.2. In the procedure, the normalised distance between the observed population fitness and the model's expected population fitness was minimised.

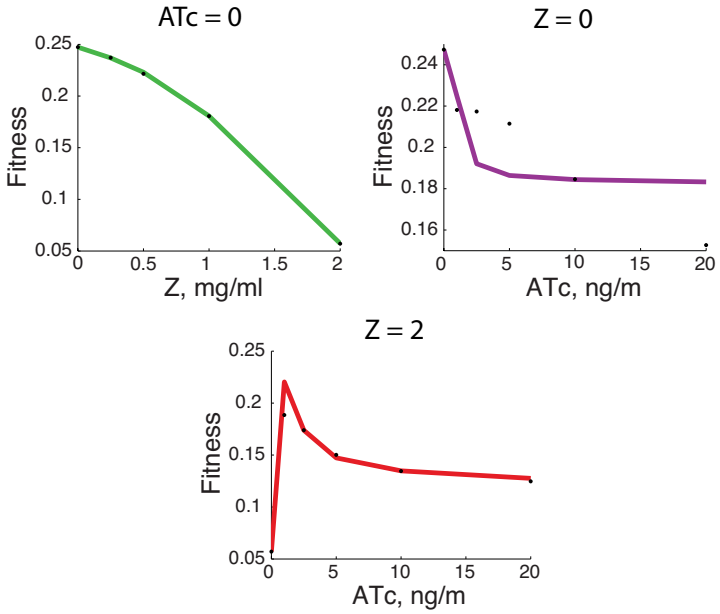

Figure 6.4: Predictions of a two-protein FSP model with population growth, fit using a distance-minimisation algorithm. Data from Nevozhay et al. (2012) shown in black.

6.4.1 Fit to data with maximal antibiotic

A good fit is found to the maximal-antibiotic ($Z = 2$) data, shown in Figure 6.4 (red). The mechanism behind the result is illustrated in Figure 6.5. When ATc is low (Figure 6.5, left), the slow decay of ZeoR relative to rtTA results in transient occupation of a low-rtTA, high-ZeoR state. It is here that cells have greatest fitness, dividing and feeding the steady states. Increasing ATc (Figure 6.5, right) causes rtTA to be more frequently in the high state, reducing the occupation of the high-fitness low-rtTA, high-ZeoR state.

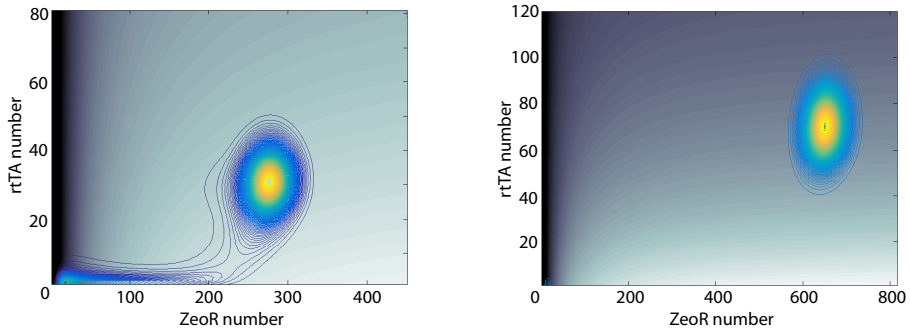

Figure 6.5: Protein distributions over fitness landscapes with Z=2 and ATc=1 ng/ml (left) and ATc=10 ng/ml (right). Background: fitness given rtTA and ZeoR content, where black is no division and white is maximal division. Overlaying contours: distribution that results of cell division model after 100 hours. Mean fitness is 0.2204 in ATc=1 ng/ml and 0.1348 in ATc=10 ng/ml, showing that additional ATc ceases to be beneficial once the bifurcation point has been passed.

6.4.2 Fit to data without antibiotic

The model fails to fit the no-antibiotic ($Z = 0$) data (Figure 6.4, purple). Having fixed the parametrisation for protein dynamics, no parametrisation for fitness (α, β, φ, and C) can fit the qualitative behaviour seen in Ta-

ble 6.1. This is likely a consequence of rtTA dynamics at high inducer concentrations (ATc $\geq$ 5 ng/ml) being near indistinguishable. Figure 6.6 shows rtTA decay rate and expression rates for different inducer concentrations. Such a qualitative behaviour is seen for any choice of φ: expression dynamics and steady states converge for high ATc.

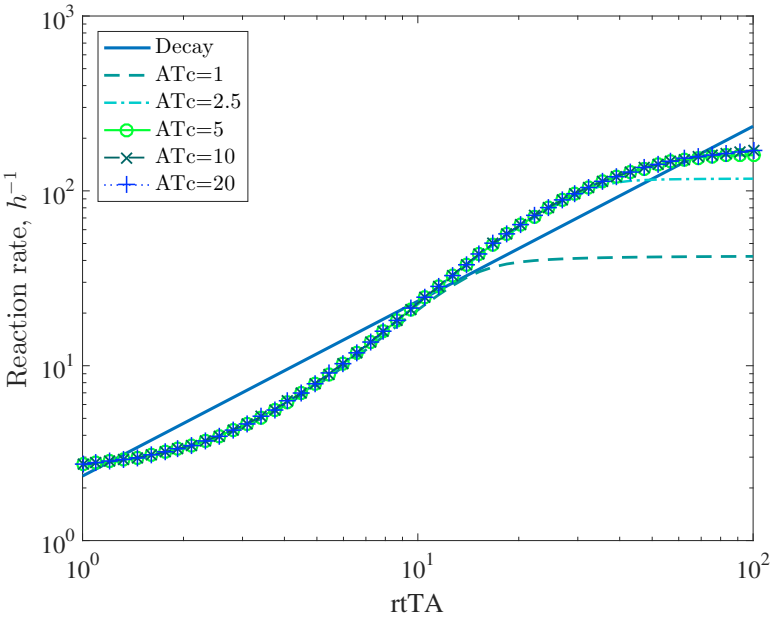

Figure 6.6: rtTA dynamics for $\varphi = 28$. The rate of decay is constant. The rate of gene expression is shown for five concentrations of ATc (ng/ml). Steady states occur at the intersections of the expression rate with the decay rate. For ATc>1 ng/ml, there are three steady states, the first stable, the second unstable, the third stable. Note that the steady states and reaction rates are very similar for ATc$\geq$ 5 ng/ml.

Furthermore, the model fails to capture the changes in the proportions of low-expressing cells as ATc changes. Table 6.1 gives the proportion of cells that were found to be in the low state for $Z = 0$, showing that the proportion decreases as ATc increases. However, for this model, all parametrisations

found through systematic searches result in a levelling out or increase in the occupation of the low state as the amount of inducer is increased.

My hypothesis is that the function given for $F(x)$ in Table 6.2, and shown in Figure 6.6, will not permit a fit to the data presented. Two analyses are presented below that give support for this.

Analysis 1: steady-state dynamics

The first analysis is a comparison between the data presented, and a model that has no division and yields the steady-state protein distribution. For the no-division model, parameters are set as $\varphi = 31$ and C as the point at which rtTA bifurcates. We are interested here in the proportion of cells that are in the low-expressing state, N_L. The model yields $N_L = 0.1929$ for ATc=10 ng/ml, and $N_L = 0.1853$ for ATc=20 ng/ml.[1] We suppose that these are the protein distributions that we would see in a population of non-dividing cells, and in this light compare them with the experimental data from Table 6.1. These data are summarised in Table 6.3.

Table 6.3 suggests that, on introducing differential cell division to the model, there is a *greater* shift from high expression to low expression for ATc=10 ng/ml than for ATc=20 ng/ml. The shift for ATc=10 ng/ml is 0.4112, and that for ATc=20 ng/ml 0.0415 (Table 6.3). The shift is due to the difference between the fitnesses of the high-expressing and low-expressing states. This difference must be greater for ATc=10 ng/ml than for ATc=20

[1] rtTA dynamics are very similar for ATc$\geq$ 5 ng/ml; the steady states are very close (Figure 6.6), as are the proportions of cells in the low-expressing state for a range of choices of C and φ.

Table 6.3: The proportion of cells in the low-expressing state, N_L, for $Z = 0$ and ATc=10 and 20 mg/ml. The first row shows the data from Nevozhay et al. (2012). The second row shows the steady-state proportion that results from the model applied to a non-dividing population. The third row shows the supposed shift from high to low expression that would be expected to result from introduction of division to the non-dividing model, according to the data.

ATc, ng/ml	10.0	20.0
Data	0.6041	0.2268
Non-dividing model	0.1929	0.1853
Shift	0.4112	0.0415

ng/ml, according to the shifts in Table 6.3. This can only be the case if fitness increases as ATc increases, which is impossible (Nevozhay et al., 2012). My conclusion is that the no-division steady states, and therefore the protein dynamics, are incorrect.

Analysis 2: average fitnesses for high- and low-expressing states

A concurrent conclusion is found with an approximation study using the $Z = 0$ data. Making the approximation that each subset (high and low) is represented by one protein-content value, it is possible to learn something about the parameters α and β. The fitness approximation is written as

$$g_{ATc} = g_0 N_L \frac{\alpha}{\alpha + P_L \frac{C}{C+\beta}} + g_0 N_H \frac{\alpha}{\alpha + P_H \frac{C}{C+\beta}} \tag{6.13}$$

where g_{ATc} and N_L are taken from Table 6.1, $N_H = 1 - N_L$, and P_L and P_H are taken as the stable steady states in Figure 6.6. Using $X = \alpha(C + \beta)$, we can rewrite this as

$$X^2 \left(1 - \frac{g_0}{g_{ATc}}\right) + X \left(C(P_L + P_H) - \frac{g_0 C}{g_{ATc}}(N_L P_H + N_H P_L)\right) + P_L P_H C^2 = 0. \tag{6.14}$$

This equation was solved using a range of C–φ combinations taken from the upper boundary in Figure 6.3. All the resulting roots imply negative values of β (such that $C + \beta < 0$). The implication is that the only way to fit the $Z = 0$ data is to have $\gamma_2(P, C)$ an increasing function of C. This suggests that there is something wrong with the model, and together these observations point to the problem residing with the expression rate of $rtTA$.

Conclusion

These analyses support the notion that for the function $F(x)$ illustrated in Figure 6.6, it is not possible to generate populations with fitnesses given in Table 6.1. This is because protein dynamics for high ATc (5, 10 and 20 ng/ml) are near indistinguishable. The only factor that could cause populations to differ in different conditions is the fitness differential between the high subset and the low subset. This differential should be greatest for highest ATc. The greater the difference, the more cells 'move' from high to low expression, and this contradicts the final column of Table 6.1.

I propose that distinguishable protein dynamics could solve this problem. Considering Figure 6.6, dynamics to improve the fit to the data might be either a steeper expression gradient for higher ATc, or a higher steady state for higher ATc, or both – with the aim to have steady-state protein distributions that already resemble the target proportions in Table 6.1.

6.5 Summary & future work

Using the FSP method developed in Chapter 4, I have shown that it is possible to use a molecular model to explain the qualitative behaviour observed

by Nevozhay et al. (2012). The key is that the fitness cost is associated with the driver (rtTA) and the gain is associated with the resistance protein (ZeoR). When ZeoR decays more slowly than rtTA, the optimal solution is to express minimal rtTA to achieve occupation in the high-expression state, so that the long memory of ZeoR confers resistance for a period sufficient to allow the population to grow.

This conclusion relies on the rate of ZeoR decay exceeding the rate of rtTA decay. What is not clear is whether the phenomenon persists in the absence of the slow-decaying appended green fluorescent protein (Corish & Tyler-Smith, 1999), and therefore if one would expect to find it *in vivo*.

In terms of the model capturing the dynamics of the system, there is a clear discrepancy in one dataset in particular (Figure 6.4, purple), and this problem remains to be solved. A possible solution has been proposed: to alter the expression function $F(x)$, so that protein dynamics at high induction are more distinct. This hypothesis is qualitatively supported by fluorescence measurements reported in Nevozhay et al. (2012), in which there is a graded response to increasing ATc.

The existence of a "sweet spot" could not be modelled by a non-dividing population: if protein dynamics alone determine fitness, then sole occupation of the high-expressing state will maximise fitness. A modelling method such at that presented in Chapter 4 is necessary to capture the population-level dynamics.

7

Extrinsic noise decomposition

In Section 2.3, the concept of noise in a biological system was introduced, along with number of ways in which it is modelled. In the previous two chapters we saw how noise in a system can give rise to interesting population-level dynamics. The focus of this chapter is what drives noise *ab initio*. I will present an analysis that demonstrates how a modeller can predict the proportion of the variability in an output species that arises from intrinsic and extrinsic sources from each reaction leading to its creation.

The modelling methods used in this chapter will be described in Section 7.1, and developed in Section 7.2 in order to show the decomposition of extrinsic noise. In Section 7.3, the decomposition will be applied to some simple systems, and some heuristics will be presented in Section 7.4.

7.1 Introduction

The heterogeneity seen across an isogenic population of cells arises as a result of the stochastic nature of biological processes within a cell, and differences in environmental circumstances across cells. The former type of effect is referred to as intrinsic noise, and the latter extrinsic noise (Elowitz et al., 2002).

In Chapters 5 and 6, we saw how the combination of intrinsic noise in gene expression and extrinsic noise in protein number inheritance led to interesting behaviour at the population level. In this chapter, I will open a study into how intrinsic and extrinsic noise are propagated through a biological system.

An overview was given in Section 2.3 of methods commonly used for modelling these types of noise either separately or in combination. Here, I will adopt the approach of defining a model that consists of stochastic reactions and variable (but fixed) parameters (Toni & Tidor, 2013). This approach naturally equates species and reactions inherent to the defined system with intrinsic noise and variability in reaction rates with extrinsic noise.

In what remains of the Introduction, two methods will be presented. Each method pertains to one type of source of noise: the linear noise approximation, described in Section 7.1.1, is used to model noise from intrinsic sources, and the unscented transform, described in Section 7.1.2, is used to model noise from extrinsic sources. Each method can be decomposed, and is subsequently used to address the overarching question: can we attribute

the variability that arises in a system to any of the processes giving rise to the output?

7.1.1 The linear noise approximation

The linear noise approximation (LNA) is a method to approximate stochastic processes via expressions for means and covariances of the species of a system as functions of time, assuming spatial homogeneity (van Kampen, 1981). The LNA is exact for purely linear systems (Grima, 2010) and a valid approximation for nonlinear systems with sufficiently large populations of species (Wallace et al., 2012). For nonlinear systems where the molecular abundance is not large, it is necessary to dissect the dynamics in order to determine how valid its linearisation is Grima (2010).

With the LNA, the mean values for all species $\mathbf{x}$ in the modelled system are defined according to the rate equation

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = S \cdot a(\mathbf{x}, \theta). \tag{7.1}$$

Here, S is the stoichiometry matrix and $a(\mathbf{x}, \theta)$ a vector of reaction propensities, where we include explicitly the dependence on the parameter vector θ. The covariance matrix C for the species evolves according to the fluctuation–dissipation theorem (van Kampen, 1981)

$$\frac{\mathrm{d}C}{\mathrm{d}t} = JC + CJ^T + D \tag{7.2}$$

where J is the Jacobian, and $D = S \cdot \mathrm{diag}(a(\mathbf{x}, \theta)) \cdot S^T$. In what follows, I will solve assuming stationarity, i.e. $\frac{\mathrm{d}C}{\mathrm{d}t} = 0$.

Equation 7.2 can be decomposed into its component parts, corresponding to reactions r_j, $j = 1, ..., n$, which are given by

$$\frac{\mathrm{d}C_{r_j}}{\mathrm{d}t} = JC_{r_j} + C_{r_j}J^T + S^{(j)} \cdot \mathrm{diag}(a(\mathbf{x}, \theta)) \cdot S^{(j)T} \tag{7.3}$$

where C_{r_j} is the contribution to the covariance matrix from the j-th reaction, and $S^{(j)}$ is the row of the stoichiometry matrix that corresponds to reaction r_j. Then $C = \sum_{j=1,...,n} C_{r_j}$ (Komorowski et al., 2013).

For an individual molecule (here, the output molecule of a system) the LNA-derived variance is written

$$\mathrm{Var}_{\mathrm{LNA}} = \mathrm{Var}_{r_1} + ... + \mathrm{Var}_{r_n}, \tag{7.4}$$

where Var_{r_i} is the variance contributed by reaction r_i (Komorowski et al., 2013).

7.1.2 The unscented transform

The unscented transform (UT) (Wan & Van Der Merwe, 2000) is used in this chapter to model extrinsic noise, in the same manner as in Toni & Tidor (2013). In this section the mathematics behind it is described, and in the following section it is shown how it can be decomposed to learn about how extrinsic noise propagates through a system.

Mathematical formulation of the unscented transform

The unscented transform is an efficient scheme for propagating a multivariate Gaussian distribution through a non-linear function, recovering a multivariate Gaussian distribution in output space (Wan & Van Der Merwe,

2000; Silk et al., 2011). A precisely defined set of points in the input space, called sigma points, are mapped via the function to the output space. Their configuration in output space enables the construction of the output distribution. It is an efficient approximation to extensive sampling or tabulation of parameter space (Scott et al., 2006).

The sigma points are chosen according to a set of rules. The first is the mean of the parameter distribution, μ_θ. Subsequent points, of which there are $2L$, where L is the dimensionality of θ, are selected to represent the spread about the mean (van der Merwe, 2004). The formulae for choosing sigma points are:

$$\begin{aligned} \theta_0 &= \mu_\theta; \\ \theta_s &= \mu_\theta + \left(\sqrt{(L+\lambda)\Sigma_\theta}\right)_s, \quad s = 1, ..., L; \\ \theta_{L+s} &= \mu_\theta - \left(\sqrt{(L+\lambda)\Sigma_\theta}\right)_s, \quad s = 1, ..., L. \end{aligned} \tag{7.5}$$

Parameters α and κ govern the spread of the points via λ, where $\lambda = \alpha^2(L+\kappa) - L$. Here I will use $\alpha = 0.001$ in order to achieve stable, consistent results, and $\kappa = 0$ following Toni & Tidor (2013). See van der Merwe (2004) for details on scaling parameters.

The mapping function, f, is evaluated at each sigma point, generating a set of points in the output space

$$\mathbf{x}_s = f(\theta_s), \quad s = 0, ..., 2L. \tag{7.6}$$

These points are used to construct the output distribution, $\mathbf{x} \sim \mathcal{N}(\mu_\mathbf{x}, \Sigma_\mathbf{x})$, according to some weights, w. For $x, y \in \mathbf{x}$, means and (co)variances are

defined as

$$\mu_x \approx \sum_{s=0}^{2L} w_s^{(m)} x_s; \tag{7.7}$$

$$\Sigma_{xy} \approx \sum_{s=0}^{2L} w_s^{(c)} (x_s - \mu_x)(y_s - \mu_y), \tag{7.8}$$

where the weights are defined to be (van der Merwe, 2004)

$$\begin{aligned} w_0^{(m)} &= \frac{\lambda}{L+\lambda}; \\ w_0^{(c)} &= \frac{\lambda}{L+\lambda} + (1 - \alpha^2 + \beta); \\ w_s^{(m)} = w_s^{(c)} &= \frac{1}{2(L+\lambda)}, \quad s = 1, ..., 2L. \end{aligned} \tag{7.9}$$

In the following work, I will use $\beta = 2$, the suggested choice for reconstructing a Gaussian distribution (van der Merwe, 2004).

Modelling extrinsic noise

In modelling biological systems with these methods, we are representing extrinsic sources of noise with a multivariate Gaussian probability distribution for the parameter set, θ: $\theta \sim \mathcal{N}(\mu_\theta, \Sigma_\theta)$.[1] The transformation of this distribution through the LNA-derived functions for the first two moments gives the output distribution, $x : x \sim \mathcal{N}(\mu_x, \Sigma_x)$ (Toni & Tidor, 2013).

Though a Gaussian input is required, the rate functions $a(\mathrm{x}, \theta)$ can involve any transformation of θ, so that other distributions might be realised. For example, with a log-normal distribution, real rates can be bounded below by zero, and the multivariate Gaussian provides the hyper-parameters.

[1] In this work, only diagonal matrices Σ_θ are considered. For implementation and interpretation of non-diagonal covariance matrices, see Appendix C.

The modelling process is shown in Figure 7.1. The LNA captures intrinsic fluctuations by mapping a parameter set to a normal output distribution of dimension n_S, where n_S is the number of species. The UT captures extrinsic noise by mapping a parameter distribution of dimension n to a normal output distribution of dimension n_S. Together, they map the parameter distribution of dimension n to a normal output distribution of dimension $n_S(n_S+3)/2$, corresponding to n_S mean values, n_S variances, and $n_S(n_S-1)$ covariances.

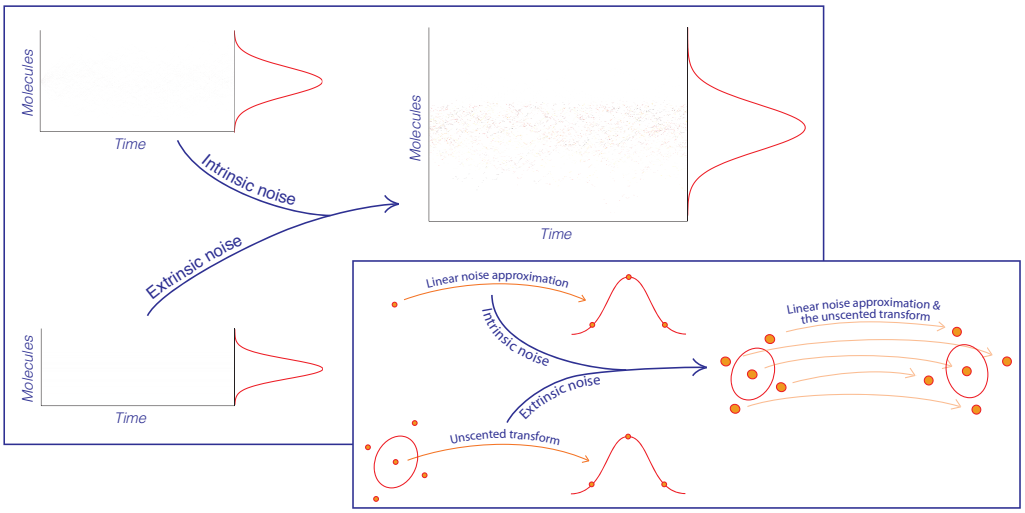

Figure 7.1: The linear noise approximation and the unscented transform for one output. Intrinsic noise in a simulation arises due to the stochastic nature of reactions; it is modelled using the linear noise approximation, which maps a parameter to the output distribution. Extrinsic noise in a simulation corresponds to differing rate parameters; it is modelled using the unscented transform, which maps a parameter distribution to the output distribution. Combining the methods yields a map from a parameter distribution to an output distribution that includes, in one dimension, the expected variance (the intrinsic noise) and, in the other dimension, the variance about the expectation (the extrinsic noise).

For example, for two variables x and y, the mean vector output by the UT is given by $\left(\mu_{\bar{x}}\ \mu_{\bar{y}}\ \mu_{\mathrm{Var}_x}\ \mu_{\mathrm{Var}_y}\ \mu_{\mathrm{Cov}_{x,y}}\right)^T$, where μ_{Var_x}, μ_{Var_y} and $\mu_{\mathrm{Cov}_{x,y}}$ give

the intrinsic noise. The covariance matrix is

$$\begin{pmatrix} \Sigma_{\bar{x},\bar{x}} & \Sigma_{\bar{x},\bar{y}} & \Sigma_{\bar{x},\mathrm{Var}_x} & \Sigma_{\bar{x},\mathrm{Var}_y} & \Sigma_{\bar{x},\mathrm{Cov}_{x,y}} \\ \Sigma_{\bar{y},\bar{x}} & \Sigma_{\bar{y},\bar{y}} & \Sigma_{\bar{y},\mathrm{Var}_x} & \Sigma_{\bar{y},\mathrm{Var}_y} & \Sigma_{\bar{y},\mathrm{Cov}_{x,y}} \\ \Sigma_{\mathrm{Var}_x,\bar{x}} & \Sigma_{\mathrm{Var}_x,\bar{y}} & \Sigma_{\mathrm{Var}_x,\mathrm{Var}_x} & \Sigma_{\mathrm{Var}_x,\mathrm{Var}_y} & \Sigma_{\mathrm{Var}_x,\mathrm{Cov}_{x,y}} \\ \Sigma_{\mathrm{Var}_y,\bar{x}} & \Sigma_{\mathrm{Var}_y,\bar{y}} & \Sigma_{\mathrm{Var}_y,\mathrm{Var}_x} & \Sigma_{\mathrm{Var}_y,\mathrm{Var}_y} & \Sigma_{\mathrm{Var}_y,\mathrm{Cov}_{x,y}} \\ \Sigma_{\mathrm{Cov}_{x,y},\bar{x}} & \Sigma_{\mathrm{Cov}_{x,y},\bar{y}} & \Sigma_{\mathrm{Cov}_{x,y},\mathrm{Var}_x} & \Sigma_{\mathrm{Cov}_{x,y},\mathrm{Var}_y} & \Sigma_{\mathrm{Cov}_{x,y},\mathrm{Cov}_{x,y}} \end{pmatrix}$$

where the top-left two-by-two matrix gives the extrinsic noise. We can therefore write the total variability for the system as the sum of the expected variance (intrinsic noise) and the variance about the mean (extrinsic noise) (Toni & Tidor, 2013):

$$\begin{pmatrix} \mathrm{Var}_{x,tot} & \mathrm{Cov}_{x,y,tot} \\ \mathrm{Cov}_{x,y,tot} & \mathrm{Var}_{y,tot} \end{pmatrix} = \begin{pmatrix} \mu_{\mathrm{Var}_x} + \Sigma_{\bar{x},\bar{x}} & \mu_{\mathrm{Cov}_{x,y}} + \Sigma_{\bar{x},\bar{y}} \\ \mu_{\mathrm{Cov}_{x,y}} + \Sigma_{\bar{x},\bar{y}} & \mu_{\mathrm{Var}_y} + \Sigma_{\bar{y},\bar{y}} \end{pmatrix}. \tag{7.10}$$

In this work, the focus will be on the total output variance of some species x, which is expressed as the sum of intrinsic and extrinsic contributions (Toni & Tidor, 2013):

$$\begin{aligned} \mathrm{Var}_{x,tot} &= \mu_{\mathrm{Var}_x} + \Sigma_{\bar{x},\bar{x}} \\ &= \mathrm{Var}_{x,in} + \mathrm{Var}_{x,ex}. \end{aligned} \tag{7.11}$$

$\mathrm{Var}_{x,in}$ corresponds to (but is not necessarily equal to) $\mathrm{Var}_{\mathrm{LNA}}$ and $\mathrm{Var}_{x,ex}$ corresponds to noise arising due to extrinsic sources. Below, the noise contributed by each reaction is decomposed into a pair of contributions, where one term corresponds to intrinsic and the other to extrinsic noise emanating from this reaction.

7.2 Extrinsic noise sources: decomposing the unscented transform

From the LNA, we have functions for the mean and the variance in terms of the parameters, $f_{mean}(\theta)$ and $f_{var}(\theta)$. In order to decompose the result of the UT, for each reaction r_i, we define

$$f^{(r_i)}_{mean}(\theta_s, \mu_\theta) = f_{mean}(\{\mu_{\theta,1}, ..., \theta_{s,i}, ..., \mu_{\theta,n}\}) \quad \text{and}$$

$$f^{(r_i)}_{var}(\theta_s, \mu_\theta) = f_{var}(\{\mu_{\theta,1}, ..., \theta_{s,i}, ..., \mu_{\theta,n}\}),$$

i.e., the function for reaction r_i is evaluated with noise in parameter θ_i alone.

The resulting normal distribution contains, in its mean vector, the expected variances given noise in each reaction alone, which may differ from the LNA prediction of the intrinsic noise. The covariance matrix contains the variances about the mean from each reaction. It also contains a readout of the accuracy of the decomposition.

For a single output, then, the total intrinsic noise is the intrinsic noise predicted by the LNA, plus any additional sources of intrinsic noise that arise due to extrinsic sources,

$$\mathrm{Var}_{x,in} = \mathrm{Var}_{\mathrm{LNA}} + \sum_{i=1}^{n} \left(\mu^{(r_i)}_{\mathrm{Var}_x} - \mathrm{Var}_{\mathrm{LNA}} \right), \tag{7.12}$$

where $\mu^{(r_i)}_{\mathrm{Var}_x}$ is the expected intrinsic noise given variability in parameter θ_i. The extrinsic noise is the sum of the individual reaction contributions and any interference terms,

$$\mathrm{Var}_{x,ex} = \sum_{i=1}^{n} \Sigma^{(r_i)}_{\bar{x},\bar{x}} + \sum_{i=1}^{n} \sum_{j \neq i}^{n} \Sigma^{(r_{i,j})}_{\bar{x},\bar{x}}, \tag{7.13}$$

where $\Sigma^{(r_i)}_{\bar{x},\bar{x}}$ is the extrinsic noise from parameter θ_i and $\Sigma^{(r_{i,j})}_{\bar{x},\bar{x}}$ corresponds to extrinsic noise interference for reactions r_i and r_j. The total variance is the sum of the variances from intrinsic and extrinsic sources, as in Equation 7.11.

The interference terms represent the extent to which the extrinsic noise cannot be decomposed. In summation, they are equal to the difference between the variance given the UT applied to the full system, and sum of variances given the UT applied to each parameter in turn. The interference terms are typically small for low variability in parameters. In such cases, extrinsic noise is traceable to individual reactions.

A detailed example of the decomposition is given in Section 7.3.1, applied to the birth–death process.

7.3 Application to simple systems

The previous section gave the derivation of the decompositions for the variance of some output, defined in Equations 7.12 and 7.13, along with Equation 7.11. In this section, the method of decomposing noise into intrinsic and extrinsic components of individual reactions is applied to some simple systems.

First, it is applied to the birth–death process, in which the full decomposition is enumerated for clarity. It is then applied to a catalytic cascade and a feed-forward loop to highlight the effects of reaction rates and network architectures. Figures 7.2A and 7.3A and B depict the motifs studied.

Details of the set-up for each system are presented in Appendix D, including μ_θ and Σ_θ, the molecular species modelled (output species in bold), the reactant and product stoichiometries (S_R and S_P), the mean value from the UT output, and each contribution to the noise using the decomposition above. Mass-action kinetics are employed, with all reactions of zero-th or first order. In Appendix D.4 it is shown that it is also possible to model the dual-reporter experiment within this framework.

7.3.1 Birth–death

In the birth–death process (Figure 7.2A), a species is created in reaction 1 (zero-th order) and destroyed in reaction 2 (1st order) (Feller, 1968). With this system, the LNA–UT is compared with an analysis of 10,000 stochastic simulations, and its decomposition to the moment expansion method of Zechner et al. (2012).

Detailed decomposition

For this process, the functions to be transformed are

$$f_{mean}(\theta_\bullet) = \theta_{\bullet,1}/\theta_{\bullet,2} \quad \text{and}$$
$$f_{var}(\theta_\bullet) = \theta_{\bullet,1}/\theta_{\bullet,2},$$

yielding four equations:

$$f^{(r_1)}_{mean}(\theta_s, \mu_\theta) = \theta_{s,1}/\mu_{\theta,2},$$
$$f^{(r_2)}_{mean}(\theta_s, \mu_\theta) = \mu_{\theta,1}/\theta_{s,2},$$
$$f^{(r_1)}_{var}(\theta_s, \mu_\theta) = \theta_{s,1}/\mu_{\theta,2},$$
$$f^{(r_2)}_{var}(\theta_s, \mu_\theta) = \mu_{\theta,1}/\theta_{s,2},$$

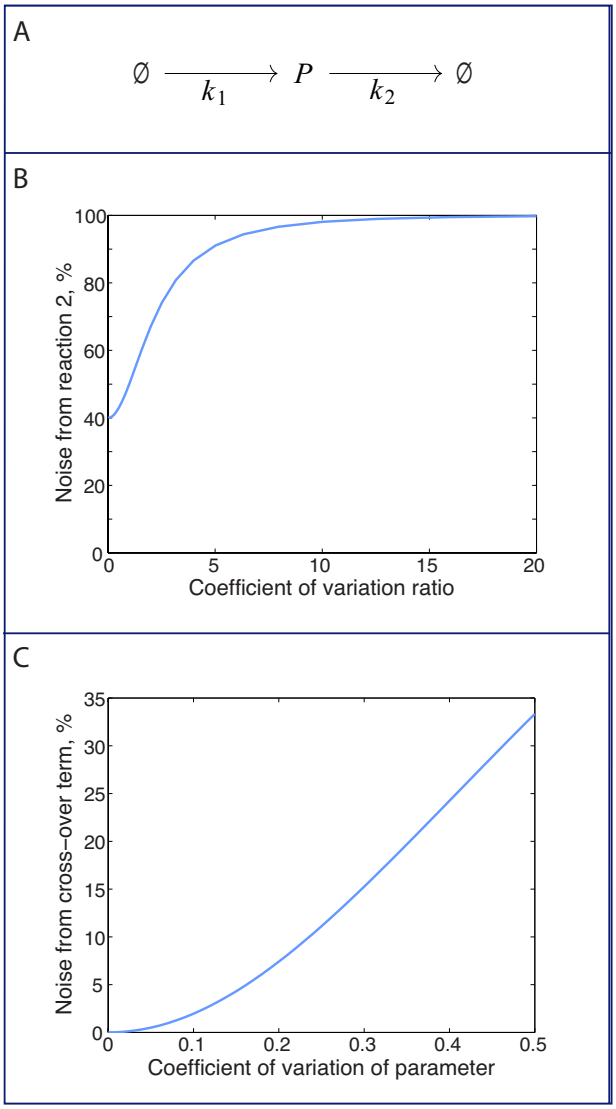

Figure 7.2: A: Schematic of the birth–death process for species P. B: The balance of noise attribution changes as the variance in the decay reaction of the birth–death process shifts. The coefficient of variation for parameter k_i is α_{k_i}. The coefficient of variation (CV) ratio is $\alpha_{k_2}/\alpha_{k_1}$: the CV of the decay parameter over the CV of the birth parameter. The percentage of noise from reaction 2 is bounded below at 40% due to intrinsic noise when the variance in the second reaction is zero. C: Amount of intrinsic noise in the second reaction that can be attributed to variance in parameter k_2 as a function of the coefficient of variation of k_2.

With the decomposed UT (dUT), an output space of the form $x_{(dUT)} \sim \mathcal{N}(\mu^{(dUT)}, \Sigma^{(dUT)})$ is generated, where

$$\mu^{(dUT)} = \begin{pmatrix} \mu_x^{(r_1)} \\ \mu_x^{(r_2)} \\ \mu_{\mathrm{Var}_x}^{(r_1)} \\ \mu_{\mathrm{Var}_x}^{(r_2)} \end{pmatrix}, \quad \Sigma^{(dUT)} = \begin{pmatrix} \Sigma_{\bar{x},\bar{x}}^{(r_1)} & \Sigma_{\bar{x},\bar{x}}^{(r_{1,2})} & \Sigma_{\bar{x},\mathrm{Var}_x}^{(r_1)} & \Sigma_{\bar{x},\mathrm{Var}_x}^{(r_{1,2})} \\ \Sigma_{\bar{x},\bar{x}}^{(r_{2,1})} & \Sigma_{\bar{x},\bar{x}}^{(r_2)} & \Sigma_{\bar{x},\mathrm{Var}_x}^{(r_{2,1})} & \Sigma_{\bar{x},\mathrm{Var}_x}^{(r_2)} \\ \Sigma_{\bar{x},\mathrm{Var}_x}^{(r_1)} & \Sigma_{\bar{x},\mathrm{Var}_x}^{(r_{1,2})} & \Sigma_{\mathrm{Var}_x\mathrm{Var}_x}^{(r_1)} & \Sigma_{\mathrm{Var}_x\mathrm{Var}_x}^{(r_{1,2})} \\ \Sigma_{\bar{x},\mathrm{Var}_x}^{(r_{2,1})} & \Sigma_{\bar{x},\mathrm{Var}_x}^{(r_2)} & \Sigma_{\mathrm{Var}_x\mathrm{Var}_x}^{(r_{2,1})} & \Sigma_{\mathrm{Var}_x\mathrm{Var}_x}^{(r_2)} \end{pmatrix} \tag{7.14}$$

and $\Sigma_{\bar{x},\bar{x}}^{(r_{i,j})}$ are the extrinsic interference terms, corresponding to the amount the extrinsic sources combine with one another.

For $\mu_\theta = (500\ 10)$ and $\Sigma_\theta = \begin{pmatrix} 50 & 0 \\ 0 & 1 \end{pmatrix}$, $\mathrm{Var}_{LNA} = 50$. The distributions for the decomposed UT and the full UT, $x_{(UT)} \sim \mathcal{N}\left(\mu^{(UT)}, \Sigma^{(UT)}\right)$, are defined as follows:

$$\mu^{(dUT)} = \begin{pmatrix} 50.0 \\ 50.5 \\ 50.0 \\ 50.5 \end{pmatrix}, \qquad \Sigma^{(dUT)} = \begin{pmatrix} 0.5 & 0.0 & 0.5 & 0.0 \\ 0.0 & 25.5 & 0.0 & 25.5 \\ 0.5 & 0.0 & 0.5 & 0.0 \\ 0.0 & 25.5 & 0.0 & 25.5 \end{pmatrix}; \tag{7.15}$$

$$\mu^{(UT)} = \begin{pmatrix} 50.5 \\ 50.5 \end{pmatrix}, \qquad \Sigma^{(UT)} = \begin{pmatrix} 26.0 & 26.0 \\ 26.0 & 26.0 \end{pmatrix}. \tag{7.16}$$

Applying Equations 7.12 and 7.13, the total extrinsic noise in the output is a sum of noise from the two reactions:

$$\begin{aligned} \mathrm{Var}_{x,in} &= \mathrm{Var}_{LNA} + \sum_i \left(\mu_{\mathrm{Var}_x}^{(r_i)} - \mathrm{Var}_{LNA}\right) \\ &= 50 + 0.0 + 0.5 \end{aligned}$$

$$
\begin{aligned}
&= \mu^{(UT)}_{\mathrm{Var}_x}; \\
\mathrm{Var}_{x,ex} &= \sum_i \Sigma^{(r_i)}_{\bar{x},\bar{x}} + \sum_i \sum_{j \neq i} \Sigma^{(r_{i,j})}_{\bar{x},\bar{x}} \\
&= 0.5 + 25.5 + 0.0 \\
&= \Sigma^{(UT)}_{\bar{x},\bar{x}}.
\end{aligned}
$$

The decomposed UT therefore equates to the full UT.

Results

The LNA–UT method concurs with the average of many stochastic simulations, and the method of Zechner et al. (2012). Using parameters $\theta = \{k_1, k_2\}$, $k_1 = 10$ and $k_2 = 0.1$, the LNA–UT predicts $x \sim \mathcal{N}(100.25, 150.375)$ (see Appendix D.1 for details). 10,000 stochastic simulations find $x \sim \mathcal{N}(100.19, 152.165)$.

In the method of Zechner et al. (2012), static extrinsic noise is incorporated directly into the expression for the output moment via calculation of cross moments. For this system, it yields $x \sim \mathcal{N}(100.25, 150.125)$. Dissection of the moment equations suggest the following expressions for the contributions from each reaction:

	Intrinsic	Extrinsic
Reaction 1	$\frac{\mu_{\theta_1}}{2\mu_{\theta_2}} = 50$	$\frac{\Sigma_{\theta_1}}{\mu^2_{\theta_2}} = 25$
Reaction 2	$\frac{\mu_{\theta_1}}{2\mu_{\theta_2}} = 50$	$\mu^2_x \Sigma_{\theta_2} = 25.125$

which agree with those of the LNA–UT (see Table D.1), with the exception of the intrinsic noise in reaction 2: the LNA–UT finds that extrinsic noise in reaction 2 adds 0.25 to its intrinsic noise.

This example highlights two points about the method that relate to the effects of the coefficients of variation of the parameters. The first is that the greater the variation, the greater the contribution to extrinsic noise. Figure 7.2B shows how noise attribution changes as the coefficient of variation in one parameter changes relative to that of the other. In what follows, I will include parameter variability with coefficients of variation equal for all parameters.

The second point to note is that as the variation increases, the interference terms in Equation 7.12, where present, increase. Where extrinsic sources of noise are small, the effects of interference are small (Figure 7.2C), and the individual contributions remain separable.

7.3.2 Catalytic cascade and feed-forward loop

In this section, the decomposition method is used to demonstrate how stoichiometry and rate in a reaction system impact on intrinsic and extrinsic noise, via comparison of a catalytic cascade and a feed-forward loop. These systems have six reactions in common, in which three reactants sequentially activate one another, and each molecular species can decay (Figures 7.3A and B). The feed-forward loop (Alon, 2007) has an additional reaction in which the first reactant can also activate the last.

Noise in the catalytic cascade

The catalytic cascade, shown in Figure 7.3A, is one in which a molecular species is created (reaction 1) and destroyed (reaction 2), and activates a

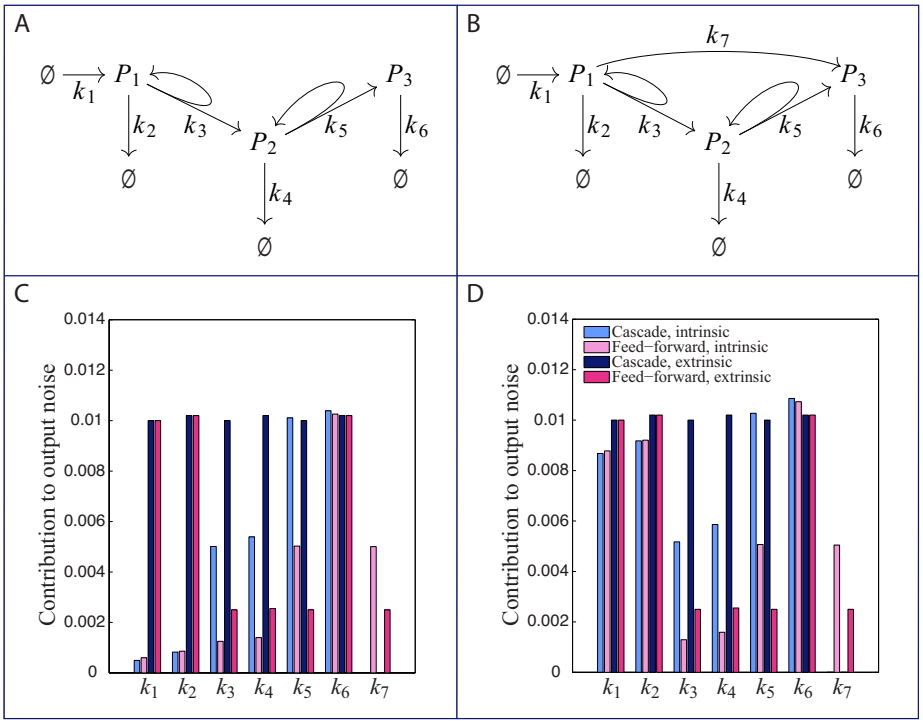

Figure 7.3: Intrinsic and extrinsic noise in a catalytic cascade and a feed-forward loop. A: Schematic motif of a catalytic cascade. B: Schematic motif of a feed-forward loop. C and D: Intrinsic and extrinsic contributions to output coefficient of variation for the two motifs where the reactions are slow (C) and fast (D). Bars show the contribution to the coefficient of variation of each rate parameter k_i.

second species (reaction 3). The second species can be destroyed (reaction 4) and activates a third species (reaction 5), which can also be destroyed (reaction 6). First, a slow cascade is compared to a fast cascade: reactions 3, 4, 5 and 6 are 100 times faster in the latter relative to the former, following Komorowski et al. (2013).

The results, shown in Figure 7.3C and D and Tables D.2 and D.3, demonstrate that intrinsic noise is transmitted from a reaction through the cascade according to the reaction's rate relative to that of the final reactions; in-

creasing all rates will not change the relative intrinsic noise resulting from any reaction. A relatively fast reaction will transmit little noise (Figure 7.3C, k_1 and k_2), as its fluctuations are averaged out by the later, slower dynamics. A relatively slow reaction will transmit more noise (Figure 7.3D, k_1 and k_2). This is recognised as output reactions acting as low-pass filters.

Extrinsic noise is transmitted to the same extent irrespective of the cascade's reaction rates, and it also arises evenly across all reactions: all reaction rate variabilities have equal impact on variability in the output.

Noise in the feed-forward loop

The feed-forward loop is shown in Figure 7.3B. To make its dynamics comparable to the catalytic cascade, the total activation of the third species is the same in both systems, i.e. the rate of reaction 5 in the cascade is equal to the sum of the rates of reactions 5 and 7 in the feed-forward loop. The loop is compared to the cascade and, to compare slow to fast kinetics, the rates of reactions 3–7 are increased 100-fold.

The results are shown in Figure 7.3C and D and Tables D.4 and D.5. The intrinsic noise in reactions 5 and 7 of the loop add up to that of reaction 5 of the cascade. However, intrinsic contributions of reactions 3 and 4 are substantially reduced compared to the cascade, and not compensated at any point. As before, when reactions 1 and 2 become relatively slow, their intrinsic noise terms become enhanced while all other noise sources remain the same (Figure 7.3D, k_1 and k_2).

Again, extrinsic noise in the loop is unaffected by changes in rate. However,

it is reduced across all reactions that form the loop in the motif (Figures 7.3C and D, k_3, k_4, k_5 and k_7). The feed-forward loop architecture therefore appears to reduce both types of noise compared to the cascade, with a more pronounced effect in extrinsic noise.

7.4 Mechanistic explanation of observations

The last section highlighted how variability in reaction rates can influence noise in the output. The LNA–UT is now dissected in order to discover the mechanism underlying the observations. For this, I return to the definition of extrinsic noise in Equation 7.8: it is the 'variance around the expected mean'. It is calculated as the perturbation to the output mean that results from perturbation to a parameter value. Therefore, the role of a parameter in determining the output mean determines its role in extrinsic noise.

For illustration, consider a single species and a simple function that defines the mean, e.g. $f_{\text{mean}}(\theta) = k_1/k_2$. To find the extrinsic noise using the UT, we find first the expected mean when the parameters are altered, one parameter at a time. We capture parameter variability by adding or subtracting a small fraction, δ. The expected mean given variability in parameter k_i, $\mu_i(\theta)$, will have the form

$$\mu_1(\theta) = w_0^{(m)}\frac{k_1}{k_2} + w_1^{(m)}\frac{k_1+\delta k_1}{k_2} + w_1^{(m)}\frac{k_1-\delta k_1}{k_2}, \tag{7.17}$$

$$\mu_2(\theta) = w_0^{(m)}\frac{k_1}{k_2} + w_1^{(m)}\frac{k_1}{k_2+\delta k_2} + w_1^{(m)}\frac{k_1}{k_2-\delta k_2}, \tag{7.18}$$

where weights $w_0^{(m)}$ and $w_1^{(m)}$ are as given in Equation 7.9.

Then, the extrinsic noise, $\Sigma_i(\theta)$, due to variability in parameter k_i is given by

$$\Sigma_1(\theta) = w_0^{(s)}\left(\mu_1 - \frac{k_1}{k_2}\right)^2 + w_1^{(s)}\left(\mu_1 - \frac{k_1+\delta k_1}{k_2}\right)^2 + w_1^{(s)}\left(\mu_1 - \frac{k_1-\delta k_1}{k_2}\right)^2, \tag{7.19}$$

$$\Sigma_2(\theta) = w_0^{(s)}\left(\mu_2 - \frac{k_1}{k_2}\right)^2 + w_1^{(s)}\left(\mu_2 - \frac{k_1}{k_2+\delta k_2}\right)^2 + w_1^{(s)}\left(\mu_2 - \frac{k_1}{k_2-\delta k_2}\right)^2. \tag{7.20}$$

Thus the effect of extrinsic sources of noise on outputs is determined by which parameters appear in the expression for the mean of the output, and what effect perturbing them has on the output. How a parameter appears in the expression for the output mean can be deduced from the pathway structure.

In what remains of this section, some typical motifs are discussed in terms of how the reaction rates influence the expression for the mean, and therefore how much they impact on extrinsic noise in the output.

7.4.1 Obligatory paths

Rates governing any (portions of) paths that are obligatory do not affect the mean and therefore do not contribute any extrinsic noise to the system output, with the exceptions of original sources and final sinks. An 'obligatory' path is one for which there is no alternative. A net contribution of zero is made to output noise because any fluctuation in the substrate is precisely negated by the fluctuation in the product.

For example, consider enzyme dynamics of the form

$$\emptyset \xrightarrow{k_1} P_0 \underset{k_3}{\overset{k_2}{\rightleftharpoons}} E \xrightarrow{k_4} P \xrightarrow{k_5} \emptyset$$

where P_0 is the substrate, P the product, and E the complex, with the enzyme omitted for simplicity. The output P has expectation k_1/k_5. All of the intervening steps are obligatory and lead only to P, so, apart from reactions 1 and 5, every reaction is both a sink and a source, and therefore their contributions to extrinsic noise in P cancel each other out.

7.4.2 Stoichiometry

The expression for the mean will typically be a quotient with the numerator comprising the sum of all sources and the denominator the sum of all sinks. Parameters can appear in the numerator, the denominator, or both; they can be multiplying factors, they can be part of a sum, and they can be raised to powers higher than one. The fundamental question is: what is the effect of a small shift δ on expressions of the form of Equations 7.17–7.20?

Sources and sinks

Nominally, numerators and denominators make similar contributions to the overall noise in the output. A rate parameter for a reaction that is a unique source (sink) for the output will manifest as a multiplying factor in the numerator (denominator). All such multiplying factors have similar effects in Equations 7.17–7.20.

For example, the final protein of the catalytic cascade of Figure 7.3A has

the expected stationary solution

$$\frac{k_1 k_3 k_5}{k_2 k_4 k_6}.$$

Terms in the numerator correspond to the creation of each species, and those in the denominator to their decay. Analysis of extrinsic noise shows that each reaction contributes similarly (Figures 7.3C and D).

Branching

Where a sink or source is not unique but occurs at a branching point of the reaction network, the rate parameters appear as sums in the expression for the mean. For example, if $\emptyset \xrightarrow{k_1} A \xrightarrow{k_2} \emptyset$ and, in addition, $A \xrightarrow{k_3} B$, the mean of A is

$$\frac{k_1}{k_2 + k_3}.$$

Here, heterogeneity in rates k_2 and k_3 will typically make small to moderate contributions to the output variability.

When the production (or decay) of a molecular species can occur through more than one reaction, the extrinsic noise contributions of the individual reactions do not sum up to extrinsic noise resulting from a single reaction, given the same overall flux. The more similar the rates of the two reactions are, the smaller the sum of their extrinsic noise contributions will be. Parallel branches are therefore a means for diminishing the effects of extrinsic noise in a chemical reaction pathway, as was seen in the feed-forward loop when compared to the catalytic cascade (Figure 7.3).

Multimerisation

A molecular species that enters a reaction in a cooperative manner contributes to the expression for the mean a term that is raised to the appropriate power (e.g. squared for a reaction involving a dimer). As a result, variability in such processes can contribute considerably to the extrinsic noise in the output.

Homodimer formation contributes (relatively) more extrinsic noise than heterodimer formation. A homodimer would have a mean of the form

$$\frac{k_1^2 k_3}{k_2^2 k_4},$$

where k_1 is rate of formation of the monomer, k_2 its decay, and k_3 and k_4 creation and decay of the dimer, respectively. The corresponding heterodimer would have mean

$$\frac{v_1 v_3 v_5}{v_2 v_4 v_6},$$

where $v_1 = v_3 = k_1$, $v_2 = v_4 = k_2$, $v_5 = k_3$ and $v_6 = k_4$. Perturbation to k_1 has greater impact than perturbation to both v_1 and v_3. For parameters $\{k_1, k_2, k_3, k_4\} = \{25, 0.5, 0.1, 0.1\}$, the homodimer contributes extrinsic noise of $\text{Var}_{hom,ex} \approx$158,000, compared to $\text{Var}_{het,ex} \approx$94,000 of the heterodimer. Given that transcription factors frequently occur as dimers, tetramers, hexamers etc. (Toni et al., 2011), it might be expected that variability in upstream transcriptional processes greatly increases extrinsic noise in the output.

7.4.3 Reaction rates

The extent of a reaction's contribution to extrinsic noise depends little on the parameter's magnitude (see Figure 7.3) and largely on its form in the mean expression, as discussed above. However, there are instances in which magnitude does play a role. As an illustration, consider a conversion cascade with three species, which is similar to the catalytic cascade but, rather than activating subsequent species (e.g. $A \rightarrow A+B$), each species is transformed into the next (e.g. $A \rightarrow B$). The output of a conversion cascade has mean

$$\frac{k_1 k_3 k_5}{(k_2 + k_3)(k_4 + k_5)k_6}, \tag{7.21}$$

where k_2, k_4 and k_6 are rates of decay of the three species, k_1 is the rate of creation of the first species, and k_3 and k_5 and are the rates of conversion (similarly to Figure 7.3A). The output of a catalytic cascade has mean

$$\frac{k_1 k_3 k_5}{k_2 k_4 k_6}. \tag{7.22}$$

For a conversion cascade with slow decay terms (k_2, k_4) relative to conversion rates (k_3, k_5), the mean expression tends to k_1/k_6. Thus, for a fast conversion cascade, there are effectively only two reactions contributing to the extrinsic noise. Conversely, with fast decay the mean expression tends to Equation 7.22 and the reaction scheme resembles the catalytic cascade, with all reactions contributing equally. The relative rates determine where a reaction system resides on the spectrum between these two extremes, which suggests control of decay rates along a pathway as a tool for tuning noise in a process output.

7.5 Discussion

In the preceding sections, a study of how sources of extrinsic and intrinsic noise affect the overall variability in the output of biomolecular reaction systems using the LNA and the UT was presented. Both the LNA and the UT capture the dynamics of the first two moments. Third moments and beyond are outside their scope, although it is in principle possible to capture higher moments using generalisations of the LNA (Grima, 2010; Thomas et al., 2013; Endres & Wingreen, 2008; Lakatos et al., 2015), and the UT can also be extended, for example by fitting a Gaussian mixture model to the distribution (Silk et al., 2013).

For the examples considered here, and perhaps more generally, intrinsic and extrinsic contributions are comparable for parameter coefficients of variation (a) of the order 0.1. For $a \ll 0.01$, intrinsic variability is likely to be the sole determiner of output noise. For $a \gg 0.1$, extrinsic variability is likely to be the sole determiner of output noise.

Propagation of intrinsic noise is sensitive to rate parameters. Reactions likely to contribute to noise in the output are those that are slower and those that are more proximate (in the reaction scheme) to the final output. All reactions propagate noise from previous reactions, and slower reactions are capable of filtering out (high-frequency) noise from upstream reactions, effectively acting as low-pass filters (Fujita et al., 2010; Mc Mahon et al., 2015).

For a pathway consisting only of conversion reactions, it is necessarily the

case that exactly half comes from the output's creation and half from its decay, as can be seen from the analytic LNA decomposition and was reported in Komorowski et al. (2013). Any steps involving catalysis, however, do not comply with this rule, and the contribution of noise intrinsic to these reactions is not bounded in relation to conversion steps, as is seen in Figure 7.3.

Unlike the propagation of intrinsic noise, that of extrinsic noise does not depend substantially on the rates of reactions (see Figures 7.3C and D). Note that parameter variabilities were set relative to their absolute values so that they themselves did not confound the analysis (Toni & Tidor, 2013). A larger factor in the pattern of transmission of sources of extrinsic noise is the structure of the reaction network. Obligatory steps contribute no extrinsic noise; increased branching reduces noise transmission; and reactions involving homomultimers contribute relatively more to extrinsic noise than those involving equivalent heteromultimers.

7.6 Conclusion

This chapter has demonstrated the decomposition of noise in a molecule of interest, in which contributions are traced to intrinsic and extrinsic sources in the different reactions that gave it rise. The sum is less straightforward than those given in Komorowski et al. (2013) and Toni & Tidor (2013) when the parameters are mapped through a non-linear transformation. The additional terms that account for the discrepancy are generally small for sufficiently small variation in the parameter (see Figure 7.2C); where they

are large, they indicate the confounding of different sources of extrinsic noise, rendering separation impossible.

Assuming a Gaussian distribution over the model parameters to capture cell-to-cell variability in a system limits the applicability of the method. While it can be effective (Filippi et al., 2016), approaches such as this (and its predecessors (Komorowski et al., 2013; Toni & Tidor, 2013)) are to be viewed as a starting point.

The heuristics derived in Section 7.4 might be most immediately applicable to design in synthetic biology. In this setting, taking into account many unknown factors via extrinsic sources of noise in the modelling process may aid in the choice of e.g. regulatory motifs (Diambra et al., 2014) or signalling components (Voliotis et al., 2014; Jovanovic et al., 2015), allowing identification of those that are more robust to or less likely to introduce extrinsic noise.

They may also find a role in model simplification or reduction (Gillespie et al., 2009; Cappelletti & Wiuf, 2014), in which choices are made regarding whether factors are modelled explicitly or by proxy. An understanding of the different implications of intrinsic and extrinsic factors would elucidate the implications of such decisions. For example, a sequence of open-conversion reactions can be faithfully replaced by a single reaction, which would not be true of a catalytic cascade.

The decomposed LNA–UT method is a quick and flexible method for studying noise behaviours in simple or simplified systems (Filippi et al., 2016;

Toni & Tidor, 2013). It has been used to show how some mechanisms may be able to enhance or attenuate different classes of noise; however, it is not sufficiently sophisticated to interpret the relative contributions of the noisy forces that drive the interesting dynamics seen in Chapters 5 and 6.

8

Concluding remarks and future work

The purpose of modelling is to learn about biological systems. Mechanistic models are routinely created in order to summarise and test our understanding of a system under study. Simultaneously, models can be used to formulate and query general principles that apply across a range of biological systems. Both these processes are founded upon some choice of modelling methodology.

Contributions to these areas have been made each in different sections of this thesis. Towards model testing, a computational tool was presented that performs model selection and parameter inference in Chapter 3. Towards theoretical principles in systems biology, a study was presented into how noise propagates in generic biochemical pathways in Chapter 7. Towards the foundations on which systems biology is built, a modelling methodology was

developed in Chapter 4. The methodology is designed specifically to model systems in which molecular dynamics interact with population dynamics, and two implementations were presented in Chapters 5 and 6.

Proposed models are compared to data in order to query how well the behaviour observed is captured by the model, or, more explicitly, the hypothesis represented by the model. With nested sampling, the support afforded to a model by some data (the evidence) is approximated algorithmically, while the posterior parameter distribution is inferred simultaneously (Skilling, 2006). Sysbions (Johnson et al., 2015), a software package implementing accelerated nested sampling, was presented in Chapter 3. This tool is designed to facilitate model parametrisation and validation in the systems biology community.

Nested sampling requires the definition of a likelihood function for the data (Skilling, 2006). This function quantifies the probability of observing the data given some model. The model must therefore yield some distribution over values corresponding to an observation. In the examples given in Chapter 3, Gaussian noise was assumed about the mean-field dynamics of the models, following mass-action kinetics.

In different settings, different modelling methods may be chosen, according to the system under study and the purpose of the model. For example, noisy protein dynamics have been proposed as the driving force behind numerous population-level phenomena resulting of cell-fate decisions (Balázsi et al., 2011). An appealing methodology to model such systems would

therefore be one that captures the full distribution over species that result from their interactions, such as the finite state projection (FSP) (Munsky & Khammash, 2006). To capture population-level phenomena, however, requires linkage of protein-level dynamics to population-level dynamics. To this end, a new modelling methodology was presented in Chapter 4 that models protein-driven population dynamics.

The methodology developed in Chapter 4 assumes an isogenic population of cells. As an implementation of the FSP, it approximates the chemical master equation (CME) via truncation of the state space (Munsky & Khammash, 2006). It therefore has within its remit the full distribution over state occupations. It has the capacity to model systems with numerous stages and biomolecules. The usefulness of the methodology is demonstrated in its ability to model population dynamics in a cell cycle and in antimicrobial resistance. However, further method development is required before it can be used to model a system in which population dynamics depend on population size.

The models of the cell cycle presented in Chapter 5 were developed to capture two hypotheses regarding the existence of a quiescent stage. The first model had four stages only, representing the hypothesis that the quiescent phenotype belongs to stage G1. The second model included a fifth, quiescent stage, G0, distinct from G1.

It was shown that both models were able to reproduce the behaviours observed in published data (Spencer et al., 2013); however, further analysis is required before it can be established what accompanying behaviours are

predicted by each model. For example, from the distance-minimisation parametrisation of the models, it seems that the four-stage model predicts that cells spend more time in stage G1 than the five-stage model does in G1 and G0. This parameter optimisation, however, does not take account of the range of parametrisations, and therefore predicted behaviours, that would give rise to an adequate fit to the data. Such an analysis could be performed with itemised data (rather than summary data) and a likelihood-based parameter inference and model validation method such as that presented in Chapter 3.

The model of antibiotic resistance in yeast, presented in Chapter 6, demonstrated that combining population- and protein-level models yields predictions that are not intuitive from considering population- or protein-level models alone. Using distance minimisation to parameterise the model, a solution was found that exhibited the desired behaviour in one experimental setting, but highlighted a problem with the model in another. Ruling out the presented model of *rtTA* expression, future work will begin with formulating a new hypothesis about rtTA dynamics.

The power of noise in driving population-level dynamics motivated the question of where variability arises and how it is transmitted. The dissection of noise sources presented in Chapter 7 began to unpick mechanisms through which biological systems accumulate and attenuate noise. A number of factors influencing the magnitude of extrinsic noise was identified, including branching in biochemical pathways and reactions involving multimerisation.

Note that these factors were found by considering only the first two moments

of the distribution over parameters and the distribution over the output; to validate these insights, it would be informative to ask the same questions using more moments, and of biological systems *in vivo*. Further, the insights pertain to systems in which the sources of extrinsic noise are static, i.e. varied but fixed across different realisations. This corresponds to factors that vary on a timescale much slower than that of the output of interest. Another question to ask then is the extent to which the observations hold as these factors vary faster.

Modelling noise in biological systems has been the common thread in this thesis. By considering noisy protein dynamics, we can model emergent behaviours at the population level. By dissecting models of noise, we can learn how variability arises in a system. By defining noise models, we can use likelihood-based methods of model validation, with which we can transparently critique the suitability of our hypotheses to explain biological phenomena. If we are to test our hypotheses through comparison of models to data, we must be confident that our modelling methodology is faithfully capturing our beliefs about how stochasticity inherent to the processes manifests in what we ultimately observe.

Bibliography

Abbas, T., & Dutta, A. (2009). p21 in cancer: intricate networks and multiple activities. *Nature Reviews Cancer*, *9*(6), 400–414.

Acar, M., Mettetal, J. T., & van Oudenaarden, A. (2008). Stochastic switching as a survival strategy in fluctuating environments. *Nature Genetics*, *40*(4), 471–475.

Aitken, S., & Akman, O. E. (2013). Nested sampling for parameter inference in systems biology: application to an exemplar circadian model. *BMC Systems Biology*, *7*(1), 72.

Alarcón, T., & Jensen, H. J. (2011). Quiescence: a mechanism for escaping the effects of drug on cell populations. *Journal of the Royal Society Interface*, *8*(54), 99–106.

Ale, A., Kirk, P., & Stumpf, M. P. H. (2013). A general moment expansion method for stochastic kinetic models. *The Journal of Chemical Physics*, *138*(17), 174101.

Alexandersson, M. (2001). On the existence of the stable birth-type distribution in a general branching process cell cycle model with unequal cell division. *Journal of Applied Probability*, *38*(3), 685–695.

Allen, R., & Waclaw, B. (2016). Antibiotic resistance: a physicist's view. *arXiv*.

Alon, U. (2007). Network motifs: theory and experimental approaches. *Nature Reviews Genetics*, *8*(6), 450–461.

Altinok, A., Gonze, D., Lévi, F., & Goldbeter, A. (2011). An automaton model for the cell cycle. *Interface Focus*, *1*(1), 36–47.

Amador, V., Ge, S., Santamaría, P. G., Guardavaccaro, D., & Pagano, M. (2007). APC/C(Cdc20) controls the ubiquitin-mediated degradation of p21 in prometaphase. *Molecular Cell*, *27*(3), 462–473.

Angus, J. E. (1994). The probability integral transform and related results. *SIAM Review*, *36*(4), 652–654.

Babtie, A. C., Kirk, P., & Stumpf, M. P. H. (2014). Topological sensitivity analysis for systems biology. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(52), 18507–18512.

Balázsi, G., van Oudenaarden, A., & Collins, J. J. (2011). Cellular decision making and biological noise: from microbes to mammals. *Cell*, *144*(6), 910–925.

Banasiak, J., Pichór, K., & Rudnicki, R. (2012). Asynchronous exponential growth of a general structured population model. *Acta Applicandae Mathematicae*, *119*(1), 149–166.

Barnum, K. J., & O'Connell, M. J. (2014). Cell cycle regulation by checkpoints. *Methods in Molecular Biology*, *1170*, 29–40.

Baron, U., Gossen, M., & Bujard, H. (1997). Tetracycline-controlled transcription in eukaryotes: novel transactivators with graded transactivation potential. *Nucleic Acids Research*, *25*(14), 2723–2729.

Basse, B., Baguley, B. C., Marshall, E. S., Joseph, W. R., van Brunt, B., Wake, G., & Wall, D. J. N. (2003). A mathematical model for analysis of the cell cycle in cell lines derived from human tumors. *Journal of Mathematical Biology*, *47*(4), 295–312.

Bekkal Brikci, F., Chiorino, G., & Boushaba, K. (2009). G1/S transition and cell population dynamics. *Networks and Heterogeneous Media*, *4*(1), 67–90.

Bekkal Brikci, F., Clairambault, J., & Perthame, B. (2008). Analysis of a molecular structured population model with possible polynomial growth for the cell division cycle. *Mathematical and Computer Modelling*, *47*(7-8), 699–713.

Bekkal Brikci, F., Clairambault, J., Ribba, B., & Perthame, B. (2007). An age-and-cyclin-structured cell population model for healthy and tumoral tissues. *Journal of Mathematical Biology*, *57*(1), 91–110.

Bellomo, N., Li, N. K., & Maini, P. K. (2008). On the foundations of cancer modelling: selected topics, speculations, and perspectives. *Mathematical Models and Methods in Applied Sciences*, *18*(04), 593–646.

Borges, R., Calsina, À., Cuadrado, S., & Diekmann, O. (2014). Delay equation formulation of a cyclin-structured cell population model. *Journal of Evolution Equations*, *14*(4-5), 841–862.

Bowsher, C. G., & Swain, P. S. (2012). Identifying sources of variation and the flow of information in biochemical networks. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(20), E1320–8.

Bravo, R., & Axelrod, D. E. (2013). A calibrated agent-based computer model of stochastic cell dynamics in normal human colon crypts useful for in silico experiments. *Theoretical Biology & Medical Modelling*, *10*(1), 66.

Burkoff, N. S., Várnai, C., Wells, S. A., & Wild, D. L. (2012). Exploring the energy landscapes of protein folding simulations with Bayesian computation. *Biophysical Journal*, *102*(4), 878–886.

Burnham, K. P., & Anderson, D. R. (2010). *Model Selection and Multimodel Inference*. A Practical Information-Theoretic Approach. Springer.

Calderhead, B., & Girolami, M. (2009). Estimating Bayes factors via thermodynamic integration and population MCMC. *Computational Statistics & Data Analysis*, *53*(12), 4028–4045.

Cappé, O., Guillin, A., Marin, J. M., & Robert, C. P. (2004). Population Monte Carlo. *Journal of Computational and Graphical Statistics*.

Cappelletti, D., & Wiuf, C. (2014). Elimination of intermediate species in multiscale stochastic reaction networks. *arXiv*.

Carlin, B. P., & Chib, S. (1995). Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

Charlebois, D. A., & Kaern, M. (2013). An Accelerated Method for Simulating Population Dynamics. *Communications in Computational Physics*, *14*(2), 461–476.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*.

Chib, S., & Jeliazkov, I. (2001). Marginal likelihood from the Metropolis–Hastings output. *Journal of the American Statistical Association*.

Chiorino, G., & Lupi, M. (2002). Variability in the timing of G1/S transition. *Mathematical Biosciences*, *177*, 85–101.

Corish, P., & Tyler-Smith, C. (1999). Attenuation of green fluorescent protein half-life in mammalian cells. *Protein Engineering*, *12*(12), 1035–1040.

Cover, T. M., & Thomas, J. A. (2012). *Elements of Information Theory*. John Wiley & Sons.

de Jong, I. G., Haccou, P., & Kuipers, O. P. (2011). Bet hedging or not? A guide to proper classification of microbial survival strategies. *BioEssays*, *33*(3), 215–223.

Del Moral, P., Doucet, A., & Jasra, A. (2006). Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *68*(3), 411–436.

Diambra, L., Raj Senthivel, V., Barcena Menendez, D., & Isalan, M. (2014). Cooperativity to increase Turing pattern space for synthetic biology. *ACS Synthetic Biology*, *4*(2), 177–186.

Dybowski, R., McKinley, T. J., Mastroeni, P., & Restif, O. (2013). Nested sampling for Bayesian model comparison in the context of Salmonella disease dynamics. *PLoS ONE*, *8*(12), e82317.

Dyson, J., Villella-Bressan, R., & Webb, G. F. (2002). Asynchronous exponential growth in an age structured population of proliferating and quiescent cells. *Mathematical Biosciences*, *177-178*, 73–83.

Ellermeyer, S. F., & Pilyugin, S. S. (2012). A size-structured model of bacterial growth and reproduction. *Journal of Biological Dynamics*, *6*(2), 131–147.

Elowitz, M. B., Levine, A. J., Siggia, E. D., & Swain, P. S. (2002). Stochastic gene expression in a single cell. *Science*, *297*(5584), 1183–1186.

Endres, R. G., & Wingreen, N. S. (2008). Accuracy of direct gradient

sensing by single cells. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(41), 15749–15754.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*. John Wiley & Sons.

Feroz, F., Hobson, M. P., & Bridges, M. (2009). MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics. *Monthly Notices of the Royal Astronomical Society*, *398*(4), 1601–1614.

Filippi, S., Barnes, C. P., Kirk, P. D. W., Kudo, T., Kunida, K., McMahon, S. S., Tsuchiya, T., Wada, T., Kuroda, S., & Stumpf, M. P. H. (2016). Robustness of MEK-ERK dynamics and origins of cell-to-cell variability in MAPK Signaling. *Cell Reports*, *15*(11), 2524–2535.

Foley, C., Bernard, S., & Mackey, M. C. (2006). Cost-effective G-CSF therapy strategies for cyclical neutropenia: mathematical modelling based hypotheses. *Journal of Theoretical Biology*, *238*(4), 754–763.

Friel, N., & Wyse, J. (2012). Estimating the evidence – a review. *Statistica Neerlandica*, *66*(3), 288–308.

Fujita, K. A., Toyoshima, Y., Uda, S., Ozaki, Y.-i., Kubota, H., & Kuroda, S. (2010). Decoupling of receptor and downstream signals in the Akt pathway by its low-pass filter characteristics. *Science Signaling*, *3*(132), ra56.

Gillespie, D. T. (1992). A rigorous derivation of the chemical master equation. *Physica A*, *188*, 404–425.

Gillespie, D. T. (2007). Stochastic simulation of chemical kinetics. *Annual Review of Physical Chemistry*, *58*, 35–55.

Gillespie, D. T., Cao, Y., Sanft, K. R., & Petzold, L. R. (2009). The subtle business of model reduction for stochastic chemical kinetics. *The Journal of Chemical Physics*, *130*(6), 064103.

Gillespie, D. T., Hellander, A., & Petzold, L. R. (2013). Perspective: Stochastic algorithms for chemical kinetics. *The Journal of Chemical Physics*, *138*(17), 170901.

Grima, R. (2010). An effective rate equation approach to reaction kinetics in small volumes: Theory and application to biochemical reactions in nonequilibrium steady-state conditions. *The Journal of Chemical Physics*, *133*(3), 035101.

Grima, R. (2015). Linear-noise approximation and the chemical master equation agree up to second-order moments for a class of chemical systems. *Physics Review E: Statistical, Nonlinear and Soft Matter Physics*, *92*(4), 042124.

Grueber, C. E., Nakagawa, S., Laws, R. J., & Jamieson, I. G. (2011). Multimodel inference in ecology and evolution: challenges and solutions. *Journal of Evolutionary Biology*, *24*(4), 699–711.

Gutenkunst, R. N., Waterfall, J. J., Casey, F. P., Brown, K. S., Myers, C. R., & Sethna, J. P. (2007). Universally sloppy parameter sensitivities in systems biology models. *PLoS Computational Biology*, *3*(10), e189.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*(1), 97–109.

Hilfinger, A., Chen, M., & Paulsson, J. (2012). Using temporal correlations and full distributions to separate intrinsic and extrinsic fluctuations in biological systems. *Physical Review Letters*, *109*(24), 248104.

Hilfinger, A., & Paulsson, J. (2011). Separating intrinsic from extrinsic fluctuations in dynamic biological systems. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(29), 12167–12172.

Huh, D., & Paulsson, J. (2011). Random partitioning of molecules at cell division. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(36), 15004–15009.

Jahnke, T. (2011). On reduced models for the chemical master equation. *Multiscale Modeling & Simulation*, *9*(4), 1646–1676.

Jeffreys, H. (1961). *The Theory of Probability*. Oxford University Press.

Johnson, R., Kirk, P., & Stumpf, M. P. H. (2015). SYSBIONS: nested sampling for systems biology. *Bioinformatics*, *31*(4), 604–605.

Jovanovic, G., Sheng, X., Ale, A., Feliu, E., Harrington, H. A., Kirk, P., Wiuf, C., Buck, M., & Stumpf, M. P. H. (2015). Phosphorelay of non-orthodox two component systems functions through a bi-molecular mechanism in vivo: the case of ArcB. *Molecular BioSystems*, *11*(5), 1348–1359.

Kidokoro, T., Tanikawa, C., Furukawa, Y., Katagiri, T., Nakamura, Y., & Matsuda, K. (2008). CDC20, a potential cancer therapeutic target, is negatively regulated by p53. *Oncogene*, *27*(11), 1562–1571.

Kirk, P., Thorne, T., & Stumpf, M. P. H. (2013). Model selection in systems and synthetic biology. *Current Opinion in Biotechnology*, *24*(4), 767–774.

Kirk, P. D. W., Babtie, A. C., & Stumpf, M. P. H. (2015). Systems biology (un)certainties. *Science*, *350*(6259), 386–388.

Kitano, H. (2002). Systems biology: a brief overview. *Science*, *295*(5560), 1662–1664.

Komorowski, M., Miękisz, J., & Stumpf, M. (2013). Decomposing noise in biochemical signaling systems highlights the role of protein degradation. *Biophysical Journal*, *104*(8), 1783–1793.

Kussell, E., & Leibler, S. (2005). Phenotypic diversity, population growth, and information in fluctuating environments. *Science*, *309*(5743), 2075–2078.

Lakatos, E., Ale, A., Kirk, P. D. W., & Stumpf, M. P. H. (2015). Multivariate moment closure techniques for stochastic kinetic models. *The Journal of Chemical Physics*, *143*(9), 094107.

Liepe, J., Barnes, C., Cule, E., Erguler, K., Kirk, P., Toni, T., & Stumpf, M. P. H. (2010). ABC-SysBio–approximate Bayesian computation in Python with GPU support. *Bioinformatics*, *26*(14), 1797–1799.

Luzyanina, T., Roose, D., Schenkel, T., Sester, M., Ehl, S., Meyerhans, A., & Bocharov, G. (2007). Numerical modelling of label-structured cell population growth using CFSE distribution data. *Theoretical Biology & Medical Modelling*, *4*(1), 26.

MacArthur, B. D., Ma'ayan, A., & Lemischka, I. R. (2009). Systems biology of stem cell fate and cellular reprogramming. *Nature Reviews Molecular Cell Biology*, *10*(10), 672–681.

MacLean, A. L., Filippi, S., & Stumpf, M. P. H. (2014). The ecology in the hematopoietic stem cell niche determines the clinical outcome in chronic myeloid leukemia. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(10), 3883–3888.

MacLean, R. C., Hall, A. R., Perron, G. G., & Buckling, A. (2010). The population genetics of antibiotic resistance: integrating molecular mechanisms and treatment contexts. *Nature Reviews Genetics*, *11*(6), 405–414.

Malumbres, M., & Barbacid, M. (2009). Cell cycle, CDKs and cancer: a changing paradigm. *Nature Reviews Cancer*, *9*(3), 153–166.

Massie, T. M., Ryabov, A., Blasius, B., Weithoff, G., & Gaedke, U. (2013). Complex transient dynamics of stage-structured populations in response to environmental changes. *The American Naturalist*, *182*(1), 103–119.

Mc Mahon, S. S., Lenive, O., Filippi, S., & Stumpf, M. P. H. (2015). Information processing by simple molecular motifs and susceptibility to noise. *Journal of the Royal Society Interface*, *12*(110), 20150597.

McCulloch, R. E., & Rossi, P. E. (1992). Bayes factors for nonlinear hypotheses and likelihood distributions. *Biometrika*, *79*(4), 663–676.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, *21*(6), 1087–1092.

Mukherjee, P., Parkinson, D., & Liddle, A. R. (2006). A nested sampling algorithm for cosmological model selection. *The Astrophysical Journal Letters*, *638*(2), L51.

Munsky, B., & Khammash, M. (2006). The finite state projection algorithm for the solution of the chemical master equation. *The Journal of Chemical Physics*, *124*(4), 044104.

Murphy, J. T., & Walshe, R. (2011). Modeling Antibiotic Resistance in Bacterial Colonies Using Agent-Based Approach. In *Understanding the Dynamics of Biological Systems*, (pp. 131–154). New York, NY: Springer New York.

Neal, R. M. (2001). Annealed importance sampling. *Statistics and Computing*, *11*(2), 125–139.

Nevozhay, D., Adams, R. M., Van Itallie, E., Bennett, M. R., & Balázsi, G. (2012). Mapping the environmental fitness landscape of a synthetic gene circuit. *PLoS Computational Biology*, *8*(4), e1002480.

Newton, M. A., & Raftery, A. E. (1994). Approximate Bayesian inference

with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

Nordon, R. E., Ko, K.-H., Odell, R., & Schroeder, T. (2011). Multi-type branching models to describe cell differentiation programs. *Journal of Theoretical Biology*, *277*(1), 7–18.

Overton, K. W., Spencer, S. L., Noderer, W. L., Meyer, T., & Wang, C. L. (2014). Basal p21 controls population heterogeneity in cycling and quiescent cell cycle states. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(41), E4386–93.

Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D., & van Oudenaarden, A. (2002). Regulation of noise in the expression of a single gene. *Nature Genetics*, *31*(1), 69–73.

Paulsson, J. (2004). Summing up the noise in gene networks. *Nature*, *427*(6973), 415–418.

Pedraza, J. M., & Paulsson, J. (2008). Effects of molecular memory and bursting on fluctuations in gene expression. *Science*, *319*(5861), 339–343.

Pedraza, J. M., & van Oudenaarden, A. (2005). Noise propagation in gene networks. *Science*, *307*(5717), 1965–1969.

Pullen, N., & Morris, R. J. (2014). Bayesian model comparison and parameter inference in systems biology using nested sampling. *PLoS ONE*, *9*(2), e88419.

Rao, C. V., Wolf, D. M., & Arkin, A. P. (2002). Control, exploitation and tolerance of intracellular noise. *Nature*, *420*(6912), 231–237.

Raser, J. M., & O'Shea, E. K. (2005). Noise in gene expression: origins, consequences, and control. *Science*, *309*(5743), 2010–2013.

Rodriguez, N., Donizelli, M., & Le Novère, N. (2007). SBMLeditor: effective creation of models in the Systems Biology Markup language (SBML). *BMC Bioinformatics*, *8*(1), 79.

Ruess, J., Milias-Argeitis, A., & Lygeros, J. (2013). Designing experiments to understand the variability in biochemical reaction networks. *Journal of the Royal Society Interface*, *10*(88), 20130588–20130588.

Satyanarayana, A., Hilton, M. B., & Kaldis, P. (2008). p21 Inhibits Cdk1 in the absence of Cdk2 to maintain the G1/S phase DNA damage checkpoint. *Molecular Biology of the Cell*, *19*(1), 65–77.

Schafer, K. A. (1998). The cell cycle: a review. *Veterinary Pathology*, *35*(6), 461–478.

Schwabe, A., & Bruggeman, F. J. (2014). Contributions of cell growth and biochemical reactions to nongenetic variability of cells. *Biophysical Journal*, *107*(2), 301–313.

Scott, M., Ingalls, B., & Kaern, M. (2006). Estimations of intrinsic and extrinsic noise in models of nonlinear genetic networks. *Chaos*, *16*(2), 026107.

Selimkhanov, J., Taylor, B., Yao, J., Pilko, A., Albeck, J., Hoffmann, A., Tsimring, L., & Wollman, R. (2014). Systems biology. Accurate information transmission through dynamic biochemical signaling networks. *Science*, *346*(6215), 1370–1373.

Shahrezaei, V., & Marguerat, S. (2015). Connecting growth with gene expression: of noise and numbers. *Current Opinion in Microbiology*, *25*, 127–135.

Shahrezaei, V., Ollivier, J. F., & Swain, P. S. (2008). Colored extrinsic fluctuations and stochastic gene expression. *Molecular Systems Biology*, *4*, 196.

Shraiman, B. I. (2005). Mechanical feedback as a possible regulator of tissue growth. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(9), 3318–3323.

Silk, D., Filippi, S., & Stumpf, M. P. H. (2013). Optimizing threshold-schedules for sequential approximate Bayesian computation: applications to molecular systems. *Statistical Applications in Genetics and Molecular Biology*, *12*(5), 603–618.

Silk, D., Kirk, P. D. W., Barnes, C. P., Toni, T., Rose, A., Moon, S., Dallman, M. J., & Stumpf, M. P. H. (2011). Designing attractive models via automated identification of chaotic and oscillatory dynamical regimes. *Nature Communications*, *2*, 489.

Silk, D., Kirk, P. D. W., Barnes, C. P., Toni, T., & Stumpf, M. P. H. (2014). Model selection in systems biology depends on experimental design. *PLoS Computational Biology*, *10*(6), e1003650.

Singh, A., & Soltani, M. (2013). Quantifying intrinsic and extrinsic variability in stochastic gene expression models. *PLoS ONE*, *8*(12), e84301.

Sivia, D., & Skilling, J. (2006). *Data Analysis*. A Bayesian Tutorial. Oxford University Press.

Skilling, J. (2006). Nested sampling for general Bayesian computation. *Bayesian Analysis*, *1*(4), 833–859.

Skilling, J. (2012). Bayesian computation in big spaces – nested sampling and Galilean Monte Carlo. *AIP Conference Proceedings*, *1443*(1), 145–156.

Spencer, S. L., Cappell, S. D., Tsai, F.-C., Overton, K. W., Wang, C. L., & Meyer, T. (2013). The proliferation-quiescence decision is controlled by a bifurcation in CDK2 activity at mitotic exit. *Cell*, *155*(2), 369–383.

Swameye, I., Muller, T. G., Timmer, J., Sandra, O., & Klingmuller, U. (2003). Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(3), 1028–1033.

Szallasi, Z., Stelling, J., & Periwal, V. (2010). *System Modeling in Cellular Biology*. From Concepts to Nuts and Bolts. MIT Press (MA).

Tadrowski, A. C., Evans, M. R., & Waclaw, B. (2016). Phenotypic switching can speed up biological evolution of microbes. *arXiv*.

Taib, Z. (1999). A branching process version of the Bell-Anderson cell population model. *Communications in Statistics. Stochastic Models*, *15*(4), 719–729.

Tănase-Nicola, S., Warren, P. B., & ten Wolde, P. R. (2006). Signal detection, modularity, and the correlation between extrinsic and intrinsic noise in biochemical networks. *Physical Review Letters*, *97*(6), 068102.

Thattai, M., & van Oudenaarden, A. (2004). Stochastic gene expression in fluctuating environments. *Genetics*, *167*(1), 523–530.

The review on antimicrobial resistance (2016). Tackling drug-resistant infections globally: final report and recommendations. Tech. rep.

Thomas, P., Matuschek, H., & Grima, R. (2013). How reliable is the linear noise approximation of gene regulatory networks? *BMC Genomics*, *14 Suppl 4*(Suppl 4), S5.

Tkačik, G., Gregor, T., & Bialek, W. (2008). The role of input noise in transcriptional regulation. *PLoS ONE*, *3*(7), e2774.

Toni, T., Jovanovic, G., Huvet, M., Buck, M., & Stumpf, M. P. H. (2011). From qualitative data to quantitative models: analysis of the phage shock protein stress response in Escherichia coli. *BMC Systems Biology*, *5*(1), 69.

Toni, T., & Tidor, B. (2013). Combined model of intrinsic and extrinsic variability for computational network design with application to synthetic biology. *PLoS Computational Biology*, *9*(3), e1002960.

Toni, T., Welch, D., Strelkowa, N., Ipsen, A., & Stumpf, M. P. H. (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, *6*(31), 187–202.

Tyson, J. J., Novak, B., Chen, K., & Val, J. (1995). Checkpoints in the cell cycle from a modeler's perspective. *Progress in Cell Cycle Research*, *1*, 1–8.

van der Merwe, R. (2004). *Sigma-Point Kalman Filters for Probabilistic Inference in Dynamic State-Space Models*. Ph.D. thesis, Oregon Health & Science University.

van Kampen, N. G. (1981). *Stochastic Processes in Physics and Chemistry*. Elsevier.

Vermeulen, K., Van Bockstaele, D. R., & Berneman, Z. N. (2003). The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer. *Cell Proliferation*, *36*(3), 131–149.

Villasana, M., & Radunskaya, A. (2003). A delay differential equation model for tumor growth. *Journal of Mathematical Biology*, *47*(3), 270–294.

Volfson, D., Cookson, S., Hasty, J., & Tsimring, L. S. (2008). Biomechanical

ordering of dense cell populations. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(40), 15346–15351.

Voliotis, M., Perrett, R. M., McWilliams, C., McArdle, C. A., & Bowsher, C. G. (2014). Information transfer by leaky, heterogeneous, protein kinase signaling systems. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(3), E326–33.

Vyshemirsky, V., & Girolami, M. (2008a). Bayesian ranking of biochemical system models. *Bioinformatics*, *24*(6), 833–839.

Vyshemirsky, V., & Girolami, M. (2008b). BioBayes: a software package for Bayesian inference in systems biology. *Bioinformatics*, *24*(17), 1933–1934.

Wallace, E. W. J., Gillespie, D. T., Sanft, K. R., & Petzold, L. R. (2012). Linear noise approximation is valid over limited times for any chemical system that is sufficiently large. *IET Systems Biology*, *6*(4), 102–115.

Wan, E. A., & Van Der Merwe, R. (2000). The unscented Kalman filter for nonlinear estimation. In *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000*, (pp. 153–158). IEEE.

Whitehead, N. A., Barnard, A. M. L., Slater, H., Simpson, N. J. L., & George P.C. Salmond (2001). Quorum-sensing in Gram-negative bacteria. *FEMS Microbiology Reviews*, *25*(4), 365–404.

Wilkinson, D. J. (2009). Stochastic modelling for quantitative description of heterogeneous biological systems. *Nature Reviews Genetics*, *10*(2), 122–133.

Zechner, C., Ruess, J., Krenn, P., Pelet, S., Peter, M., Lygeros, J., & Koeppl, H. (2012). Moment-based inference predicts bimodality in transient gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(21), 8340–8345.

Appendix A

The variance of the log-evidence in nested sampling

At each iteration the value for X, representing the proportion of prior mass contained within the likelihood contour, are estimated using the relation $X_i = tX_{i-1}$. Here, t is the shrinkage factor. In Skilling's presentation of the algorithm, t is beta-distributed with parameters $(n, 1)$ (Sivia & Skilling, 2006).

A.1 The variance when one point is discarded and replaced

The distribution $t \sim Beta(n, 1)$ describes the case in which one point is discarded and replaced at each iteration; t is the largest value of points uniformly distributed on [0,1]. With n live points, $p(t) = nt^{n-1}$. In Sivia & Skilling (2006) the expected value $\mathbb{E}(\log\ t) = -1/n$ is used for each shrinkage factor. The variance in t is $1/n$ and so, after M iterations, a standard deviation of $\sqrt{M}/n$ is accumulated (Sivia & Skilling, 2006).

Skilling recommends that the algorithm is terminated once the number of iterations greatly exceeds $n\mathcal{H}$, where $\mathcal{H}$ is the information accumulated (Sivia & Skilling, 2006). The standard deviation accompanying the final estimate of $\log\ Z$ is $\sqrt{\frac{\mathcal{H}}{n}}$.[1]

[1] It is not clear in their algorithm or in their discussions what the correspondence is

A.2 The variance when many points are discarded and replaced

It is possible for more than one point to be discarded and replaced in a single iteration. In this instance, the shrinkage factor sought is the k^{th} largest value of points uniformly distributed on [0,1], where k is the number of points to be discarded. Such a value has a distribution $t_k \sim Beta(n-k+1,k)$ (Skilling, 2012), where

$$p(t_k) = \binom{n}{k} k t_k^{n-k} (1-t_k)^{k-1}.$$

The expected value in this case is

$$\mathbb{E}(\log\ t_k) = \log(1-k(1-\exp(-1/n))).$$

As in the case in which $k=1$, $X_i = t_k X_{i-1}$.

The variance of $\log\ t_k$ is then

$$var(\log\ t_k) = \sum_{m=0}^{k-1} \frac{1}{(n-m)^2}.$$

A.3 Implications for the algorithm

The variance of a variable $t_k \sim Beta(n-k+1,k)$ increases as k increases. First: another method is required to calculate the final standard deviation. Second: a choice needs to be made as to how the number of points to be skipped is chosen given its effect on the variance of $\log\ Z$.

A.3.1 Comparing the algorithms $k=1$ and $k>1$

When points are skipped, the ground is covered more quickly. It is therefore not accurate to compare directly the case that $k=1$ with $k>1$ in terms of numbers of points discarded. It makes sense to compare them in terms of X values.

between $\mathcal{H}n$ and the point at which the algorithm exits; indeed, the algorithm iterates 1000 times and $\mathcal{H}n \approx 39$. It is unclear, then, why their variance is $\mathcal{H}/n$ and what is meant by m 'significantly exceeding' $\mathcal{H}n$.

If M_1 iterations in the $k = 1$ case take the algorithm to a point on the X axis X^*, the same point is reached in M_k iterations for $k > 1$ where

$$M_k = \frac{-M_1}{n \log(1 - k(1 - \exp(-1/n)))}. \tag{A.1}$$

A.3.2 The final variance

After M_k iterations, upon each of which k points are skipped, the variance might be estimated as the product of the number of iterations and the variance associated with each iteration:

$$var(\log\ X_{M_k}) = \sum_{j=1}^{M_k} (\psi_1(n - k + 1) - \psi_1(n + 1)) \tag{A.2}$$

$$= \sum_{j=1}^{M_k} \sum_{m=0}^{k-1} \frac{1}{(n - m)^2}. \tag{A.3}$$

Supposing k is variable:

$$var(\log\ X_{M_k}) = \sum_{j=1}^{M_k} (\psi_1(n - k_j + 1) - \psi_1(n + 1)) \tag{A.4}$$

$$= \sum_{j=1}^{M_k} \sum_{m=0}^{k_j-1} \frac{1}{(n - m)^2}, \tag{A.5}$$

where k_j is the number of points skipped on the j^{th} iteration.

A.3.3 Skilling's approximation of the variance

Skilling states that it takes $\mathcal{H}n$ steps to traverse the posterior to its bulk. To use $\sqrt{\mathcal{H}/n}$ (that is, $\sqrt{var(\log\ X_{\mathcal{H}n})}$) as the standard deviation seems to suggest that all errors arise in the transition to the bulk, and that none is associated with its traversal. Similarly, the final n live points are not included in the calculation of the standard deviation.

In Equation A.2, we need to know the number of iterations M_k that take us to the point equivalent to $\mathcal{H}n$ in the case that $k = 1$. However, M_k is typically not an integer so it is unclear how to perform this sum. In the

case that $k_j \equiv k$ it might be tempting to write

$$var(\log\ X_{M_k}) = M_k(\psi_1(n-k+1) - \psi_1(n+1)) \tag{A.6}$$

$$= M_k \sum_{m=0}^{k-1} \frac{1}{(n-m)^2}. \tag{A.7}$$

However, the variance is not a linear function of the variance of the k^{th} point; the rate of change of the variance dips and rises between leap points. We therefore use the more general formula

$$var(\log\ X_{M_k}) = \sum_{j=1}^{M_k} \sum_{m=0}^{k_j-1} \frac{1}{(n-m)^2} \tag{A.8}$$

where the final sum to k_{M_k} can terminate at the point corresponding to $X_{\mathcal{H}n}$. This is different from taking some fraction of the variance of one point corresponding to M_k. This applies also to the case where k_j varies; the variance can be summed precisely, point by point.

In terms of the algorithm: We start at the $X = 1$ end and work back through the points until we reach the first point with X smaller than $e^{-\mathcal{H}}$. Each point adds a term $\frac{1}{(n-m)^2}$ to the variance, where m is the number of preceding points with the same width. E.g., if the width changes for every point ($k \equiv 1$), $m \equiv 0$; if the width changes after 10 points, then the tenth point will have $m = 9$.

Appendix B

Derivation of Equation 4.12

To recap, we have:

- N cells
- Up to N_P proteins
- N_S cell stages
- $J = N_S(N_P + 1)$ states indexed by $j = \{1, 2, ..., J\}$
- R_j denoting the number of cells in state j, with $\sum R_j = N$
- P_j denoting the proportion of N that is in state j, with $\sum P_j = 1$ and $R = NP$.

Usually, we express the result of reactions occurring over some time step as

$$R_i^{(t+1)} = R_i^{(t)} + u_i(R^{(t)}),$$

where u_i is the change to R_i as a function of all the states. In order to model the evolution of probabilities, we need to know what happens to all cells at each time point. A state's proportion P_i can change even if there has been no change in R_i. What we seek is the change in proportion, which we call v_i:

$$P_i^{(t+1)} = P_i^{(t)} + v_i(N^{(t)}, P^{(t)}). \tag{B.1}$$

The change to the total number of cells is

$$N^{(t+1)} = N^{(t)} + \sum_{i=1}^{J} u_i(R^{(t)}). \tag{B.2}$$

A state's probability is

$$P_i^{(t+1)} = \frac{R_i^{(t+1)}}{N^{(t+1)}}. \tag{B.3}$$

Writing in terms of states at time t and then omitting the superscript, we have

$$\begin{aligned}
P_i^{(t+1)} &= \frac{R_i^{(t)} + u_i(R^{(t)})}{N^{(t)} + \sum_{j=1}^{J} u_j(R^{(t)})} && \text{(B.4)}\\
&= \frac{R_i + u_i(R)}{N + \sum_{j=1}^{J} u_j(R)} && \text{(B.5)}\\
&= \frac{N(R_i + u_i(R))}{N(N + \sum_{j=1}^{J} u_j(R))} && \text{(B.6)}\\
&= \frac{NR_i + \sum_{j=1}^{J} u_j(R)R_i + Nu_i(R) - R_i \sum_{j=1}^{J} u_j(R)}{N(N + \sum_{j=1}^{J} u_j(R))} && \text{(B.7)}\\
&= \frac{(N + \sum_{j=1}^{J} u_j(R))R_i + Nu_i(R) - R_i \sum_{j=1}^{J} u_j(R)}{N(N + \sum_{j=1}^{J} u_j(R))} && \text{(B.8)}\\
&= \frac{R_i}{N} + \frac{Nu_i(R) - R_i \sum_{j=1}^{J} u_j(R)}{N(N + \sum_{j=1}^{J} u_j(R))} && \text{(B.9)}\\
&= \frac{R_i}{N} + \frac{Nu_i(R) - NP_i \sum_{j=1}^{J} u_j(R)}{N(N + \sum_{j=1}^{J} u_j(R))} && \text{(B.10)}\\
&= \frac{R_i}{N} + \frac{u_i(R) - P_i \sum_{j=1}^{J} u_j(R)}{N + \sum_{j=1}^{J} u_j(R)} && \text{(B.11)}\\
&= P_i + \frac{u_i(R) - P_i \sum_{j=1}^{J} u_j(R)}{N + \sum_{j=1}^{J} u_j(R)} && \text{(B.12)}
\end{aligned}$$

Thus the function v_i can be expressed in terms of N and P at time point t:

$$v_i(N, P) = \frac{u_i(NP) - P_i \sum u_j(NP)}{N + \sum u_j(NP)}. \tag{B.13}$$

Here, v_i is the change in proportion, u_i is the change in number, N is the total number of cells, and $\sum u_j(NP)$ is the change in protein number going to the next time step.

Appendix C

Diagonalisation for correlated parameters

The UT decomposition relies on the construction of extrinsic noise by considering reactions independently. In cases where parameters are dependent, it is necessary to transform the equations to a new basis. Then the contributions of noise will correspond to linear combinations of parameters, i.e. relations between parameters. For example, two transcription rates might be jointly determined by the availability of transcriptional machinery.

A new set of parameters $\tilde{\theta} = M^{-1}\theta$ is constructed such that the covariance matrix for the transformed parameters $\tilde{\Sigma} = M^{-1}\Sigma M$ is diagonal. For example, for the birth–death process and $M^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$:

$$\begin{aligned}
\tilde{\theta}_1 &= a\theta_1 + b\theta_2 \\
\tilde{\theta}_2 &= c\theta_1 + d\theta_2 \\
\mathrm{Var}(\tilde{\theta}_1) &= a^2\mathrm{Var}(\theta_1) + b^2\mathrm{Var}(\theta_2) \\
\mathrm{Var}(\tilde{\theta}_2) &= c^2\mathrm{Var}(\theta_1) + d^2\mathrm{Var}(\theta_2) \\
\mathrm{Cov}(\tilde{\theta}_1, \tilde{\theta}_2) &= ac \cdot \mathrm{Var}(\theta_1) + bd \cdot \mathrm{Var}(\theta_2) + (ac + bd)\mathrm{Cov}(\theta_1, \theta_2) \\
&= 0.
\end{aligned}$$

These allow the functions $x_s = f(\theta_s)$ to be rewritten as $x_s = g(\tilde{\theta}_s)$ so that

$$f_{mean}(\theta_s) = \theta_{s,1}/\theta_{s,2} \quad \mapsto \quad g_{mean}(\tilde{\theta}_s) = \frac{d\tilde{\theta}_{s,1} - b\tilde{\theta}_{s,2}}{a\tilde{\theta}_{s,2} - c\tilde{\theta}_{s,1}},$$

$$f_{var}(\theta_s) = \theta_{s,1}/\theta_{s,2} \quad \mapsto \quad g_{var}(\tilde{\theta}_s) = \frac{d\tilde{\theta}_{s,1} - b\tilde{\theta}_{s,2}}{a\tilde{\theta}_{s,2} - c\tilde{\theta}_{s,1}}.$$

For reaction rates that are related, we can no longer say that some variance can be attributed to some reaction, because this variance has an extrinsic component which does not exclusively belong to one reaction. We no longer have a situation of 'variability in reactions' contributing variance to the output. Rather, hidden variables manifest in multiple reactions contribute to the variance.

Appendix D

Decomposed noise: model details and results

Here, details of the set-up for each system are presented, including μ_θ and Σ_θ, the molecular species modelled (output species in bold), the reactant and product stoichiometries (S_R and S_P), the mean value from the UT output, and each contribution to the noise using the decomposition above. Mass-action kinetics are employed, with all reactions of zeroth or first order.

D.1 Birth–death

$$\mu_\theta = \begin{pmatrix} 10 & 0.1 \end{pmatrix}, \quad \Sigma_{\theta_i} = a^2\mu_{\theta_i}^2, \quad a = 0.05$$
$$S_R = \begin{pmatrix} 0 & 1 \end{pmatrix}, \quad S_P = \begin{pmatrix} 1 & 0 \end{pmatrix}$$
$$\mu_x = 100.25$$

D.2 Catalytic cascade

Species: protein 1, protein 2, **protein 3**.

$$S_R = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad S_P = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Table D.1: Reaction contributions to intrinsic and extrinsic noise for the birth–death system

i	Var_{r_i}	$\mu^{(r_i)}_{\mathrm{Var}_x} - \mathrm{Var}_{LNA}$	$\Sigma^{(r_i)}_{\bar{x},\bar{x}}$	Non-zero cross-terms $\left(\Sigma^{(r_{i,j})}_{\bar{x},\bar{x}}\right)$	Total
1	50.0	0.0	25.0		75.0
2	50.0	0.25	25.125		75.375
Total	100.0	0.25	50.125	0.0	150.375

D.2.1 Slow

$$\mu_\theta = \begin{pmatrix} 50 & 1 & 0.1 & 0.1 & 0.1 & 0.1 \end{pmatrix}$$
$$\Sigma_{\theta_i} = a^2\mu^2_{\theta_i}, \quad a = 0.1$$
$$\mu_x = \begin{pmatrix} 50.5 & 51 & 51.5 \end{pmatrix}$$

Table D.2: Reaction contributions to intrinsic and extrinsic noise for slow catalysis

i	Var_{r_i}	$\mu^{(r_i)}_{\mathrm{Var}_x} - \mathrm{Var}_{LNA}$	$\Sigma^{(r_i)}_{\bar{x},\bar{x}}$	Non-zero cross-terms $\left(\Sigma^{(r_{i,j})}_{\bar{x},\bar{x}}\right)$	Total
1	1.2397	-0.0000	25.0		26.2397
2	1.2397	0.8231	25.5	0.5000 0.5000	28.5628
3	12.5	0.0248	25.0		37.5248
4	12.5	0.9812	25.5	0.5000 0.5000	39.9812
5	25.0	0.2748	25.0		50.2748
6	25.0	0.9812	25.5	0.5000 0.5000	52.4812
Total	77.4793	3.0851	151.5	3.0	235.0644

D.2.2 Fast

$$\mu_\theta = \begin{pmatrix} 50 & 1 & 10 & 10 & 10 & 10 \end{pmatrix}$$
$$\Sigma_{\theta_i} = a^2\mu^2_{\theta_i}, \quad a = 0.1$$
$$\mu_x = \begin{pmatrix} 50.5 & 51 & 51.5 \end{pmatrix}$$

Table D.3: Reaction contributions to intrinsic and extrinsic noise for fast catalysis

i	Var_{r_i}	$\mu^{(r_i)}_{\mathrm{Var}_x} - \mathrm{Var}_{LNA}$	$\Sigma^{(r_i)}_{\bar{x},\bar{x}}$	Non-zero cross-terms $\left(\Sigma^{(r_{i,j})}_{\bar{x},\bar{x}}\right)$	Total
1	21.6942	-0.0000	25.0		46.6942
2	21.6942	1.2491	25.5	0.5000 0.5000	49.4433
3	12.5000	0.4339	25.0		37.9339
4	12.5000	2.1493	25.5	0.5000 0.5000	41.1493
5	25.0000	0.6839	25.0		50.6839
6	25.0000	2.1493	25.5	0.5000 0.5000	53.6493
Total	118.3884	6.6655	151.5	3.0000	279.5539

D.3 Feed-forward loop

Species: protein 1, protein 2, **protein 3**.

$$S_R = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \quad S_P = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}$$

D.3.1 Slow

$$\mu_\theta = \begin{pmatrix} 50 & 1 & 0.1 & 0.1 & 0.05 & 0.1 & 0.05 \end{pmatrix}$$

$$\Sigma_{\theta_i} = a^2 \mu^2_{\theta_i}, \quad a = 0.1$$

$$\mu_x = \begin{pmatrix} 50.5 & 51 & 51.25 \end{pmatrix}$$

Table D.4: Reaction contributions to intrinsic and extrinsic noise for slow feed-forward loop

i	Var_{r_i}	$\mu^{(r_i)}_{\mathrm{Var}_x} - \mathrm{Var}_{LNA}$	$\Sigma^{(r_i)}_{\bar{x},\bar{x}}$	Non-zero cross-terms $\left(\Sigma^{(r_{i,j})}_{\bar{x},\bar{x}}\right)$	Total
1	1.4979	0.0	25.0		26.4979
2	1.4979	0.6484	25.5	0.2500 0.5000	28.3964
3	3.125	0.0062	6.25		9.3812
4	3.125	0.3735	6.375	0.2500 0.2500	10.3735
5	12.5	0.0687	6.25		18.8187
6	25.0	0.6546	25.5	0.5000 0.2500	51.9046
7	12.5	0.0114	6.25		18.7614
Total	59.2459	1.7628	101.125	2.0000	164.1336

D.3.2 Fast

$$\mu_\theta = \begin{pmatrix} 50 & 1 & 10 & 10 & 5 & 10 & 5 \end{pmatrix}$$
$$\Sigma_{\theta_i} = a^2\mu^2_{\theta_i}, \quad a = 0.1$$
$$\mu_x = \begin{pmatrix} 50.5 & 51 & 51.25 \end{pmatrix}$$

Table D.5: Reaction contributions to intrinsic and extrinsic noise for fast feed-forward loop

i	Var_{r_i}	$\mu^{(r_i)}_{\mathrm{Var}_x} - \mathrm{Var}_{LNA}$	$\Sigma^{(r_i)}_{\bar{x},\bar{x}}$	Non-zero cross-terms $\left(\Sigma^{(r_{i,j})}_{\bar{x},\bar{x}}\right)$	Total
1	21.9525	-0.0000	25.0		46.9525
2	21.9525	1.0617	25.5	0.2500 0.5000	49.2642
3	3.125	0.1085	6.25		9.4835
4	3.125	0.8489	6.375	0.2500 0.2500	10.8489
5	12.5	0.1710	6.25		18.9210
6	25.0	1.8291	25.5	0.5000 0.2500	53.0791
7	12.5	0.1136	6.25		18.8636
Total	100.155	4.1328	101.125	2.0000	207.4128

D.4 Dual-reporter experiments

The classical illustration of the roles of extrinsic and intrinsic noise is the dual-reporter experiment pioneered by Elowitz et al. (2002), in which two reporters share an environment (extrinsic) and have independent fluctuations (intrinsic).

In the dual-reporter experiment, contributions from intrinsic and extrinsic sources of noise to the overall variability are measured directly. $\mathrm{Var}(p_1, p_2)$ is calculated from the experimental (or here the simulated) data, shown schematically in Figure D.1B, and the coefficient of variation (CV) is defined as (Hilfinger & Paulsson, 2011)

$$\mathrm{CV}_{tot} = \frac{\mathrm{Var}_{p_1}^{tot}}{\mu_{p_1}^2} = \mathrm{CV}_{in} + \mathrm{CV}_{ex}, \tag{D.1}$$

with

$$\mathrm{CV}_{ex} = \frac{\mathrm{Var}(p_1, p_2)}{\mu_{p_1}\mu_{p_2}}. \tag{D.2}$$

D.4.1 Model set-up: two reporters

Here I use the model of Toni & Tidor (2013) with two identical mRNAs and proteins (shown schematically in Figure D.1A), the parameters of Shahrezaei et al. (2008), and the noise equations of Hilfinger & Paulsson (2011) to show how this experiment can be modelled and dissected.

Species: mRNA 1, **protein 1**, mRNA 2, **protein 2**.

$$S_R = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad S_P = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix},$$

$$\mu_\theta = \begin{pmatrix} 0.07 & 0.005 & 0.2 & 0.0004 \end{pmatrix}$$

$$\Sigma_{\theta_i} = a^2\mu_{\theta_i}^2, \quad a = 0.05$$

$$\mu_x = \begin{pmatrix} 14 & 7035 & 14 & 7035 \end{pmatrix}$$

The same parameters are used for each mRNA–protein motif.

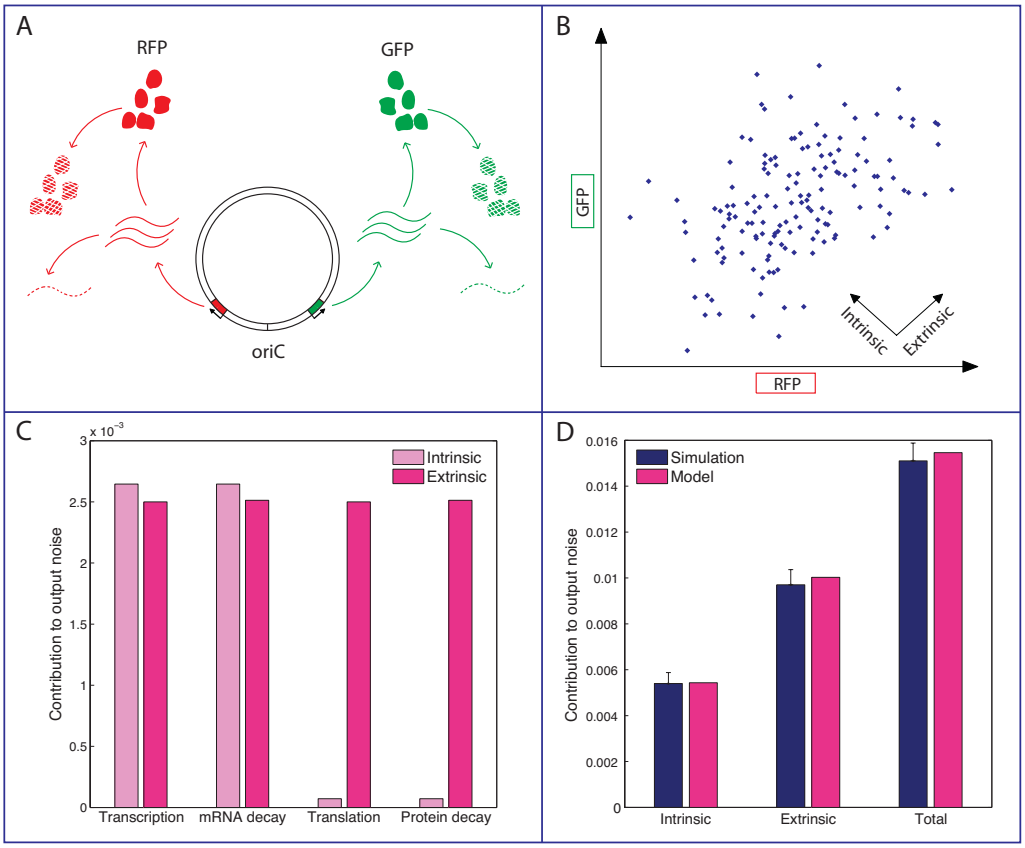

Figure D.1: The dual-reporter experiment. A: A schematic of the dual-reporter experiment. A plasmid contains two identical genes, one tagged red (RFP), the other green (GFP). For the purposes of modelling, the process of their expression consists of transcription, mRNA decay, translation and protein decay. B: Simulated measurements from cells expressing the plasmid in panel A. Correlations between RFP and GFP are due to shared environmental conditions and thus constitute extrinsic noise (i.e. spread along the $x = y$ axis). Uncorrelated variability is due to random occurrences of reactions and thus contributes to intrinsic noise. C: The intrinsic and extrinsic contributions of the four reactions to the output noise, as calculated with the LNA–UT model of a single instance of gene expression. D: Comparison of predictions of intrinsic and extrinsic noise in protein expression from simulations of the dual-reporter experiment (panels A and B) and the LNA–UT model (panel C). Simulation data consist of 20 repetitions of 512 simulations, with error bars corresponding to one standard deviation.

$\mathrm{Var}(p_1, p_2)$ is found from the LNA–UT approach as:

$$\begin{aligned}\mathrm{Var}(p_1, p_2) &= \mathrm{Var}_{in}(p_1, p_2) + \mathrm{Var}_{ex}(p_1, p_2) \\ &= \begin{pmatrix} 269500 & 0 \\ 0 & 269500 \end{pmatrix} + \begin{pmatrix} 492500 & 492500 \\ 492500 & 492500 \end{pmatrix}. \end{aligned} \tag{D.3}$$

Shown in Table D.6, the results of the model coincide well with the stochastic simulation results (Figure D.1D).

Table D.6: Noise calculations for simulated and LNA–UT modelled gene expression. Simulation data consist of 20 repetitions of 512 simulations.

	$\langle p_1 \rangle$, $\langle p_2 \rangle$	$\mathrm{Var}(p_1, p_2)$	CV_{in}	CV_{ex}	CV_{tot}
Simulation	7059, 7042	$\begin{pmatrix} 770100 & 492300 \\ 492300 & 763800 \end{pmatrix}$	0.0055	0.0099	0.0154
LNA–UT	7035, 7035	$\begin{pmatrix} 761900 & 492500 \\ 492500 & 761900 \end{pmatrix}$	0.0054	0.01	0.0154

D.4.2 Model set-up: one reporter

Here I show how the LNA–UT method can be used to reproduce the same result by modelling only one reporter.

Species: mRNA, **protein**.

$$S_R = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad S_P = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix},$$

$$\mu_\theta = \begin{pmatrix} 0.07 & 0.005 & 0.2 & 0.0004 \end{pmatrix}$$

$$\Sigma_{\theta_i} = a^2 \mu_{\theta_i}^2, \quad a = 0.05$$

$$\mu_x = \begin{pmatrix} 14 & 7035 \end{pmatrix}$$

$$\eta_{tot}^2 = 0.0154 \quad \eta_{int}^2 = 0.0054 \quad \eta_{ext}^2 = 0.01$$

The values for intrinsic, extrinsic and total noise are the same as those generated when two reporters were modelled explicitly (Table D.6).

Table D.7: Reaction contributions to intrinsic and extrinsic noise for one reporter modelling a two-reporter experiment

i	Var_{r_i}	$\mu^{(r_i)}_{\mathrm{Var}_x} - \mathrm{Var}_{LNA}$	$\Sigma^{(r_i)}_{\bar{x},\bar{x}}$	Non-zero cross-terms $\left(\Sigma^{(r_i,j)}_{\bar{x},\bar{x}}\right)$	Total
1	129629.6	0.0	122500.0		252129.6
2	129629.6	1821.5	123112.5	612.5	255176.1
3	3500.0	648.1	122500.0		126648.1
4	3500.0	717.2	123112.5	612.5	127942.2
Total	266259.2	3186.8	491225.0	1225.0	761896.0

D.4.3 One reporter, dynamic activator

Species: activator, mRNA, **protein**.

$$S_R = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad S_P = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix},$$

$$\mu_\theta = \begin{pmatrix} 0.1 & 0.001 & 0.07 & 0.005 & 0.2 & 0.0004 \end{pmatrix}$$

$$\mu_x = \begin{pmatrix} 100.25 & 14.07 & 7052.5 \end{pmatrix}$$

Table D.8: Reaction contributions to intrinsic and extrinsic noise for one reporter modelling a two-reporter experiment with activator dynamics

i	Var_{r_i}	Intrinsic	Extrinsic
1	69135.8		69135.8
2	69135.8		69135.8
3	129629.6	129629.6	
4	129629.6	129629.6	
5	3500.0	3500.0	
6	3500.0	3500.0	
Total	404530.8	266259.2	138271.6

To relate the results to the dual-reporter experiment, we relabel the activator dynamics as extrinsic in Table D.8. Thus, the totals for intrinsic and extrinsic are 266259.2 and 138271.6, respectively.