

Diss. ETH No. 22013

**Stochastic Biochemical Networks in Random Environments:
Probabilistic Modeling and Inference**

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES OF ETH ZURICH
(Dr. sc. ETH Zurich)

presented by

CHRISTOPH ZECHNER

Dipl.-Ing., Graz University of Technology, Austria
born on 06.07.1986
citizen of Austria

accepted on the recommendation of

Prof. Dr. Heinz Koepl (examiner)
Prof. Dr. John Lygeros (co-examiner)
Prof. Dr. Peter Swain (co-examiner)

2014

For

Hanna

Acknowledgments

I would like to express my deepest appreciation to Heinz Koepl for believing in me and enabling me to do my PhD in this outstanding academic environment. His exceptional availability for in-depth scientific discussions, feedback and support were crucial in finding my research directions and interests and allowed me to perform my doctoral research with maximum efficiency. Furthermore, he helped me to connect within the scientific community by funding a multitude of scientific meetings and conferences around the world and even supported a six-month research internship abroad. I am grateful for the many scientific skills I could learn from him in the course of my PhD. I highly appreciated his easygoing, friendly and down-to-earth mentality and I hope I could adopt a large portion of that mindset.

I would like to greatly thank my co-advisor and head of the Automatic Control Lab John Lygeros for the assessment of my thesis, all the fruitful scientific collaborations and the excellent work environment during my doctoral research. I could strongly profit from being exposed to a diversity of group members within and outside my field of research. I always appreciated his valuable and inspiring comments, especially during the close collaborations with Jakob Ruess. I am also very grateful for having been part of his committed teaching team. I believe the amount of time and effort he invested there sets the best example of how teaching should be done. Also, I am very thankful to him for enabling me to stay at the lab also after Heinz Koepl left ETH.

I am sincerely thankful also to my second co-advisor Peter Swain for the stimulating and enriching discussions during our meetings. As an experienced pioneer in my field of research, his feedback meant a big deal to me and helped me to find “the right” questions. I am also grateful to him for investing his time to read and assess my thesis and for traveling to Zürich to attend my PhD examination.

Being part of the BISON group was a very exciting and productive experience and the fruitful interactions and team spirit were highly appreciated. I thank my long-term office mates Michael Unger, Tatjana Petrov, Preetam Nandy and Sunil Kumar for the peaceful and friendly atmosphere, intensive white-board sessions and spontaneous discussions.

None of my achievements would have been possible without the strong support from my scientific collaborators. For instance I thank Michael Unger, Peter Krenn, Serge Pelet, Anders Hansen and Matthias Peter for their intensive experimental efforts, substantially boosting our work’s relevance to molecular biology. I am greatly thankful to my research mate Jakob Ruess. Since we started our PhDs at the same time in the same field of research, we were companions from the very first day. I highly ap-

preciated all the ideas and thoughts that emerged during our countless discussions. I also want to thank him for the enjoyable and successful collaborations and I really hope those will continue in the future. However, I thank him most for being a good friend during all the years and for the great times we had hanging out together in our spare time. My research internship in Cambridge, MA was an invaluable experience for me. The scientists from Merrimack Pharmaceuticals and Peter Sorger's lab showed me how the systems biology approach can aid to understand and treat human diseases. I specifically thank Paul Kopesky, Emily Pace, Birgit Schöberl, Mario Niepel and Peter Sorger for their supervision and feedback. I also thank Maria Rodriguez Martinez and Matthias Reumann from IBM Research, Rüschlikon for the exciting collaboration over the past years. The inspiring discussions with Jan Hasenauer and Andreas Hilfinger helped me to broaden my horizon and their valuable feedback on manuscript drafts was very useful.

I am more than grateful for all the good friends I made in my PhD at the Automatic Control Lab or elsewhere. They provided me with the necessary mental backing that was required during all stages of my doctoral studies. I want to particularly thank Stephan Huck, Christian Conte and Alexander Domahidi from the Automatic Control Lab and Janick Cardinale from the MOSAIC group for their ongoing loyalty, honesty and reliability. I will always remember the parties, the late-night movie sessions, the rock'n'roll and the video game evenings.

Finally I want to express my deepest gratitude to my family members for always standing behind me and providing me unprecedented support in the past 28 years. A million of thanks go to my wife Hanna, the person being most exposed to all the ups and downs during the last years. Her constant love, backup and tolerance were crucial for the successful completion of my PhD.

Christoph Zechner
Zürich, October 2014

Contents

Abstract	vii
1 Introduction	1
1.1 Outline and Contributions	3
1.2 Related Publications	5
2 Preliminaries	7
2.1 Stochastic Chemical Kinetics	7
2.1.1 Kolmogorov Equations	9
2.1.2 The Random Time-Change Model	11
2.1.3 Stochastic Simulation Algorithm	12
2.1.4 Moment Dynamics	13
2.1.5 Random Environments and Extrinsic Noise	15
2.2 Bayesian Statistics	18
2.2.1 Predictive Distributions	20
2.2.2 Bayesian Filtering and Smoothing	20
2.2.3 Model Selection	22
2.3 Single-Cell Fluorescence Data	23
2.3.1 Population Snapshot Data	23
2.3.2 Time-Lapse Data	24
3 Inference from Heterogeneous Snapshot Data	25
3.1 Inference Based on the Chemical Master Equation	25
3.1.1 Statistical Modeling	26
3.2 Inference Based on Moments	27
3.2.1 Marginal Moment Dynamics	27
3.2.2 Statistical Modeling	29
3.2.3 Application to Simulated Data	31
3.2.4 Application to Hog1-Induced Gene Expression in Yeast	34
4 Inference from Heterogeneous Time-Lapse Data	41
4.1 Mathematical Modeling	41
4.1.1 A Straightforward Attempt at Inference	42
4.2 Dynamic Prior Propagation	43
4.2.1 Marginal Dynamics and the Innovation Theorem	44

4.2.2	Marginal Simulation Algorithm	48
4.2.3	Sequential Markov Chain Monte Carlo	50
4.2.4	Application to Synthetic Gene Expression Data.	59
4.2.5	Application to Experimental Gene Expression Data	62
4.3	Detecting Sources of Extrinsic Variability	71
4.3.1	Hierarchical Bayesian Modeling	71
4.3.2	Variational Inference	73
4.3.3	Extension to the Incomplete Data Scenario	76
4.3.4	Case Studies	76
4.4	Optimal Design of Temporal Perturbations	79
4.4.1	Choosing the Objective Function	80
4.4.2	The Variational Problem	81
4.4.3	Stochastic Approximation	82
4.4.4	Fast Gradient Approximation Using Importance Sampling	83
4.4.5	Case Studies	83
5	Uncoupled Analysis of Biochemical Networks in Random Environments	87
5.1	Mathematical Modeling	87
5.2	Construction of the Uncoupled Dynamics	88
5.2.1	Relation to Stochastic Filtering	89
5.2.2	Modified Marginal Simulation Algorithm	94
5.2.3	Marginal Moment Dynamics	95
5.3	Generalized Master Equations	97
5.4	Fluctuations on Different Timescales	100
5.4.1	The Effective Noise	101
5.4.2	The Slow Noise Approximation	104
6	Discussion and Outlook	111
	Bibliography	113
	Curriculum Vitae	125

Abstract

Biochemical processes within clonal cell populations exhibit a substantial degree of cell-to-cell variability. While a fraction of the overall variability is due to the *intrinsic* stochasticity of the chemical reactions, a major part is attributed to fluctuations in their surrounding environment – also referred to as *extrinsic* noise. The latter causes severe complications in the analysis and inference of biochemical networks, primarily because every cell needs to be described by a differently parameterized stochastic process, giving rise to a mixed-effect model. Therefore, straightforward approaches that aim to directly analyze such models scale only poorly with the number of considered cells.

The major goal of this thesis is to address these scalability issues and to develop an efficient mathematical framework for the analysis and inference of stochastic biochemical processes subject to extrinsic noise. We show that scalable kinetic models can be derived by *marginalizing* the mixed-effect model with respect to environmental components, yielding a parameter dimensionality independent of the population size. We first follow this idea to efficiently reconstruct biochemical processes from experimental single-cell measurements. In case of population snapshot data, we derive moment equations based on the marginalized model and use them to infer its kinetic parameters from experimental flow cytometry measurements.

Analogously, we develop a marginalized inference scheme for heterogeneous time-lapse data, in which the dynamics of individual cells can be recorded over time. The time-series nature of such measurements gives rise to a challenging *missing data* problem, which we solve using a sequential Monte Carlo algorithm. Using experimental data from an artificially controlled gene in yeast, we validate the approach and demonstrate its ability to infer kinetic parameters, unmeasured chemical species and to dissect intrinsic and extrinsic contributions of noise.

The marginal process framework is then extended for the important case of dynamically changing environmental conditions. In particular, we demonstrate that the construction of the marginal dynamics for fluctuating environments is closely connected to solving a stochastic filtering problem. We show how the latter can be tackled in practice and derive a marginal stochastic simulation algorithm, which allows to perform exact simulations of only parts of a network. We also show that the marginal process framework serves a promising analytical tool to study noise in biochemical systems.

Zusammenfassung

Biochemische Prozesse in genetisch identischen Zellen weisen ein beträchtliches Maß an Zell-zu-Zell-Variabilität auf. Davon kann nur ein Teil auf die *intrinsische* Stochastizität der chemischen Reaktionen zurückgeführt werden. Ein ebenso wesentlicher Teil entsteht durch Fluktuationen in der Umgebung eines Prozesses, welcher daher oftmals als *extrinsisch* bezeichnet wird. Dieser Teil der Variabilität sorgt für erhebliche Schwierigkeiten bei der Analyse und Rekonstruktion von biochemischen Netzwerken. Insbesondere hat er zur Folge, dass jede Zelle durch einen unterschiedlich parametrisierten stochastischen Prozess beschrieben werden muss. Ansätze die auf eine direkte Analyse der daraus resultierenden Mischeffekt-Modelle abzielen, skalieren daher schlecht mit der Anzahl der berücksichtigten Zellen.

Das Hauptziel dieser Arbeit ist es sich mit diesem Skalierungsproblem auseinanderzusetzen und ein effizientes mathematisches Grundgerüst für die Analyse und Rekonstruktion von stochastischen biochemischen Prozessen unter Berücksichtigung der extrinsischen Variabilität zu entwickeln. Wir zeigen, dass skalierbare kinetische Modelle konstruiert werden können, indem man das Mischeffekt-Modell bezüglich der Umgebungskomponenten *marginalisiert*. Im Speziellen kann dadurch eine Parameterdimensionalität erzielt werden, welche unabhängig von der Grösse der Zellpopulation ist. Zunächst verwenden wir diesen Ansatz um biochemische Prozesse effizient aus experimentellen Einzelzell-Daten zu rekonstruieren. Im Fall von "Population-Snapshot-Daten" leiten wir Gleichungen für die Momente des marginalisierten Modells her, welche uns anschliessend die Schätzung von kinetischen Parametern aus Durchflusszytometriedaten ermöglichen.

Analog dazu, entwickeln wir ein marginalisiertes Schätzverfahren für heterogene Zeitreihenmessungen, in denen die Einzelzelldynamik über die gesamte Zeitdauer eines Experiments aufgenommen werden kann. Die resultierenden Daten bringen jedoch ein kompliziertes "Missing-Data-Problem" mit sich, welches wir mit Hilfe eines sequentiellen Monte Carlo Algorithmus lösen. Wir validieren den Ansatz unter Zuhilfenahme von experimentellen Daten eines künstlich kontrollierten Gens in Hefe. Im Speziellen zeigen wir, dass man damit kinetische Parameter und die Anzahl von experimentell unzugänglichen Molekülen schätzen kann. Des Weiteren lässt sich damit feststellen, wie viel der gesamten Zell-zu-Zell-Variabilität auf intrinsische beziehungsweise extrinsische Ursachen zurückzuführen sind.

Anschliessend erweitern wir das marginalisierte Prozessmodell für den wichtigen Fall von sich dynamisch verändernden Umgebungsbedingungen. Im Besonderen demonstrieren wir, dass die Konstruktion der Marginaldynamik in engem Zusammenhang mit

der Lösung eines stochastischen Filterproblems steht. Wir zeigen wie dieses Problem praktisch gelöst werden kann und leiten einen marginalen Simulationsalgorithmus her, welcher es erlaubt, lediglich einen kleinen Teil eines stochastischen Netzwerkes exakt zu simulieren. Wir zeigen ebenfalls, dass das marginalisierte Prozessmodell ein vielversprechendes analytisches Instrument zur Untersuchung von Unsicherheiten in biochemischen Systemen darstellt.

1 Introduction

With the advance of single-cell techniques, realization has grown that biochemical networks are characterized by a non-negligible degree of stochasticity [74]. Particularly in the context of gene expression, cells seemingly act “by chance”, for instance in the stress-induced activation of a transcriptional program [94]. Analyses based on only population averages can therefore yield imperfect or even spurious results [108]. Especially processes involving low-copy molecules are far off any deterministic regime [64], demanding for mathematical models that can account for their *intrinsic* stochasticity. Continuous-time Markov chains (CTMCs) were shown to provide a reasonable approximation to the discrete dynamics of stochastic reaction networks [37, 74, 76]. In particular, they allow to assess how a biochemical process and its uncertainty evolve over time, for instance through a probability distribution of the process being in a particular molecular configuration at a particular time point. In practice, however, the analysis of CTMC models, for instance through the chemical master equation (CME), turns out to be highly challenging. With the increasing pertinence of stochastic kinetic models, the development of efficient techniques for their analysis has thus evolved to a central research thread within the field of quantitative biology.

Despite the analytical complexity of CTMCs, sample paths are straightforward to generate through stochastic simulation algorithms (SSA) [26, 44]. Due to their simplicity, SSA-based approaches are extensively used to study noise in biochemical networks, for instance, by computing Monte Carlo estimates of probability distributions or moments thereof. However, such estimates are characterized by a slow convergence and often rely on an overwhelmingly large number of Monte Carlo samples. Alternative approaches aim at directly solving the CME by truncating the infinite space of molecular configurations to the ones which are reached with a reasonably high probability [81, 131]. While those approaches can guarantee high accuracies, they turn out to be prohibitive for systems involving more than only a few chemical species.

Higher efficiencies can typically be achieved through analytical approximations of the CME. For instance, the diffusion approximation [37], *van Kampen’s* system size expansion [125] and associated techniques such as the *linear noise approximation* (LNA) [31, 61] are commonly employed. The latter yields a system of ordinary differential equations (ODEs) describing only means and variances of the chemical species whose solution is straightforward to compute numerically. Along those lines, the CME can be used to derive an ODE system describing arbitrary moments of the molecular species [37, 125]. Since the resulting equations are generally infinite-dimensional, *moment-closure* approximations need to be employed to obtain a closed moment system up to

a desired order [52, 106, 128]. All these methods have in common that they are highly efficient but also that their approximation accuracy is unknown beforehand and thus, needs to be assessed using additional simulations (e.g., through SSA).

In the context of modeling, one may distinguish between two elementary problems – neither being less important for quantitatively assessing a network’s dynamics. The first one is referred to as the *forward problem*, in which the behavior of a model is investigated either analytically [17, 35, 53, 67, 96, 111] or using extensive simulations [74, 110, 123] – for instance with the assistance of the aforementioned techniques. Some studies aim to relate certain biologically meaningful properties to the kinetic parametrization of a model. The authors from [35, 96, 111] derive analytical expressions of the mRNA- or protein distributions and provide expressions for characterizing the underlying kinetics. Those approaches have the advantage that they reveal simple and amenable biophysical quantities such as the burst-size and frequency of a particular gene. On their downside, however, they rely on strong assumptions or approximations which might not always represent well what happens in reality. In such cases, Monte-Carlo-based techniques may be considered favorable, since they can deal with arbitrarily complex and nonlinear networks – putting aside their limitations in terms of computational cost. Typical applications can be found in the analysis of a network’s robustness or sensitivity with respect to variations in kinetic parameters [95, 100, 123].

The second problem is the so-called *inverse problem*, which is centered around the computational reverse-engineering of dynamical systems and their properties from quantitative data. It therefore represents the critical link between mathematical models and experimental measurements. The reverse-engineering – or inference – can be performed either with respect to the network topology [10, 70, 90] or the parametrization and states of a given network [47, 82, 130]. In the context of stochastic reaction systems, the former has received only limited attention, most likely because the substantial degree of noise in such systems does not permit reliable reconstructions using current experimental protocols. In contrast, the inference of biophysical parameters of a presumably known stochastic model from experimental single-cell data has been intensively studied over the past years. A natural classification of existing approaches may be performed by the type of single-cell data they act on. Population snapshot data such as revealed by flow cytometry- or mRNA FISH (fluorescence *in situ* hybridization) provide a handle on the mRNA- or protein distributions at a sequence of time points. Correspondingly, associated inference techniques are typically centered around the CME and approximations thereof [82, 88, 105, 136]. Time-lapse microscopy techniques allow to track living cells over time and therefore, provide information about the molecular fluctuations *inside* a single cell. However, exploiting this additional source of information for the purpose of inference turns out to be computationally challenging. The partial and noisy observations of the stochastic molecular processes give rise to a complicated *missing data problem* [39, 129], whose solution is generally intractable. A variety of numerical schemes for the solution of such problems based on either analytical approximations of the underlying process [34, 79, 92] or stochas-

tic simulation [5, 46, 47, 135] have been proposed recently. Other studies focus on the combination of statistical inference with optimal design principles [22], meaning that the experimental conditions (e.g., such as the concentration of an inhibitor) are chosen such as to yield the most revealing measurements – for instance with regard to unknown process parameters or states. Most commonly, the informativeness of an experiment is quantified based on either the *Shannon-* [13, 19, 69] or *Fisher* information [9, 61, 105] and optimized with respect to the design parameters of an experiment. Such approaches were shown to yield a substantial reduction in the estimation uncertainty and can even resolve practical non-identifiabilities between pairs of parameters.

A majority of existing studies are based on the assumption that every considered cell of a population corresponds to an independent realization of one and the same stochastic model. During recent years, however, there has been increasing evidence that homogeneous models that solely account for intrinsic noise cannot explain the large degree of phenotypic variability that is commonly observed in single-cell experiments [23, 33, 99]. This stems from the fact that for complexity reasons, such models typically involve only a handful of molecular components but neglect that they are just a fractional part of a large entirety. More specifically, fluctuations in a networks environment form an additional source of stochasticity – commonly referred to as extrinsic noise. Recent dual-reporter techniques allow to quantify the extent to which the expression of a gene is driven by intrinsic or extrinsic noise [33, 53, 120]. In fact, it often turns out that the latter even dominates [23, 33]. Recent attempts to account for extrinsic noise are based on augmenting a kinetic model by certain extrinsic quantities [53, 110] such as for instance the ribosomal abundance, giving rise to a *heterogeneous* Markov model. While such models were shown to predict well the experimentally observed variability, they suffer from the increased dimensionality. In particular, since every cell's dynamics depends on a different realization of the environmental conditions, the associated models scale poorly with the population size. For the purpose of inference – for instance – reliable process reconstructions can only be obtained by incorporating a large number of individual cells. At the same time, however, the number of unknowns increases with every considered cell. As a consequence, straightforward approaches, which aim to perform inference directly on such augmented models [34, 59, 135] are likely to be prohibitive when considering more than only very few cells at once.

1.1 Outline and Contributions

The main purpose of this thesis is to develop a mathematical framework permitting an efficient analysis and reconstruction of biochemical processes in random environments. Accordingly, dealing with the associated scalability issues forms a major part of this work and runs like a common thread through Chapters 2, 3, 4 and 5.

In Chapter 2 we introduce the background concepts that the subsequent chapters rely on. For instance, we discuss a probabilistic modeling and simulation framework for chemical kinetics and introduce the notion of random environments. We further

provide a brief primer in Bayesian inference and related techniques such as Bayesian filtering and smoothing, model selection and so forth.

The succeeding two chapters largely focus on the Bayesian inference of biophysical parameters and states from experimental single-cell measurements. In Chapter 3 we propose a statistical framework for analyzing population snapshot data. We first extend the standard CME to a heterogeneous counterpart which can in principle account for random variations in a networks environment. However, we show that approaches based on full distributions are often impracticable for the purpose of inference and propose an alternative scheme, which describes a heterogeneous cell population by moments up to a certain order. We validate the approach using synthetic and experimental flow cytometry data and demonstrate that only a few moments may be sufficient to fully identify the parameters of a stochastic reaction network.

In Chapter 4, we lay out a corresponding inference framework for time-lapse microscopy measurements. In order to fully exploit the information present in that data, the measurement trajectories are modeled as heterogeneous hidden Markov models. In Section 4.2 we propose an efficient algorithm (i.e., *dynamic prior propagation*) to perform inference of such models. In contrast to previous approaches [34, 59, 135], the scheme allows to marginalize the model with respect to the random environmental conditions and therefore, achieves a parameter dimensionality *independent* of the population size. Correspondingly, we derive a stochastic process, which no longer depends on the particular realization of the extrinsic variables, providing a coherent dynamic description of all cells in population at once. The algorithm is tested using simulated measurements and subsequently used to reconstruct the transcriptional dynamics of an artificially controlled gene in yeast. In Section 4.3, the process marginalization is combined with a sparse Bayesian learning procedure [14, 87, 122] to determine which parts of a reaction network are significantly modulated by extrinsic noise. In particular, the scheme provides an automated and data-driven way of deciding which of the reaction rates are likely to vary from cell to cell. Another extension of the inference algorithm is provided in Section 4.4, wherein it is used to assess the informativeness of a given experiment. In conjunction with an efficient stochastic approximation algorithm [58, 65], this allows us to optimally design experiments for the purpose of parameter inference from time-lapse microscopy measurements. More specifically, we apply the algorithm to calculate temporal perturbation profiles which can for instance be synthesized using novel microfluidic techniques [124].

In Chapter 5 the marginal process framework is revisited and extended for the important case of dynamically changing environmental conditions [53, 110]. Technically, this corresponds to marginalizing a CTMC model with respect to certain dimensions which are considered extrinsic. Chapter 5 is analytical in nature and – in contrast to the preceding parts – largely concerned with the aforementioned forward problem. For instance, we calculate analytical expressions to quantify a network’s ability to suppress environmental noise. Furthermore, we derive a simple and widely applicable approximation (i.e., the *slow noise approximation*) for modeling protein distributions in random

environments.

1.2 Related Publications

The research of this thesis is interdisciplinary and was performed in close collaboration with other theorists and experimentalists. In the following, we provide a list of publications, underlying the results of this work.

Chapter 2 includes results from:

H. Koepl*, C. Zechner*, A. Ganguly, S. Pelet, and M. Peter. Accounting for extrinsic variability in the estimation of stochastic rate constants. *Int J Robust Nonlin*, 22(10):1103–1119, 2012. [59]

*Remark: H. Koepl and C. Zechner contributed equally to this work.

Chapter 3 summarizes the results published in:

C. Zechner*, J. Ruess*, P. Krenn, S. Pelet, M. Peter, J. Lygeros, and H. Koepl. Moment-based inference predicts bimodality in transient gene expression. *Proc Natl Acad Sci USA*, 109(21):8340–8345, 2012. [136]

*Remark: C. Zechner and J. Ruess contributed equally to this work.

Chapter 4 is based on the publications:

C. Zechner, M. Unger, S. Pelet, M. Peter, and H. Koepl. Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings. *Nat Methods*, 11(2):197–202, 2014. [137]

C. Zechner, S. Deb, and H. Koepl. Marginal dynamics of stochastic biochemical networks in random environments. In *Proceedings of the European Control Conference (ECC), 2013*, pages 4269–4274, 2013. [132]

C. Zechner, F. Wadehn, and H. Koepl. Sparse learning of Markovian population models in random environments. In *Proceedings of the 19th IFAC World Congress, 2014*, pages 1723–1728. IFAC, 2014. [138]

C. Zechner, S. Pelet, M. Peter, and H. Koepl. Recursive Bayesian estimation of stochastic rate constants from heterogeneous cell populations. In *Proceedings of the IEEE Conference on Decision and Control, 2011*, pages 5837–5843, 2011. [135]

C. Zechner, P. Nandy, M. Unger, and H. Koepl. Optimal variational perturbations for the inference of stochastic reaction dynamics. In *Proceedings of the IEEE Conference on Decision and Control, 2012*, pages 5336–5341. IEEE, 2012. [134]

Chapter 5 is mainly based on the paper:

C. Zechner and H. Koepl. Uncoupled analysis of stochastic reaction networks in fluctuating environments. *Plos Comp Biol*, 2014. In press. [133]

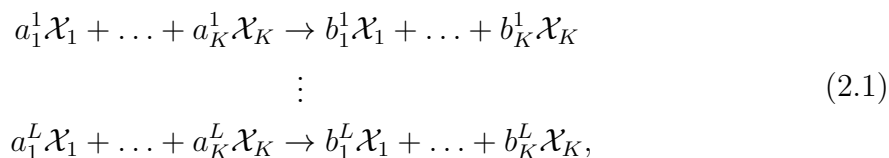
Related publications not explicitly discussed:

P. Nandy, M. Unger, C. Zechner, and H. Koepl. Optimal perturbations for the identification of stochastic reaction dynamics. In *Proceedings of the 16th IFAC Symposium on System Identification*, pages 686–691, 2012. [86]

2 Preliminaries

2.1 Stochastic Chemical Kinetics

Cells implement their complex behavior and functionality through a variety of dynamically interacting biomolecules. In the following we denote those molecules – or *species* – as $\mathcal{X}_1, \dots, \mathcal{X}_K$. The way the species can interact is defined by a set of L chemical reactions of the form



with $a_i^j, b_i^j \in \mathbb{N}$. We further define matrices

$$A = \begin{pmatrix} a_1^1 & \dots & a_K^1 \\ \vdots & \ddots & \vdots \\ a_1^L & \dots & a_K^L \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} b_1^1 & \dots & b_K^1 \\ \vdots & \ddots & \vdots \\ b_1^L & \dots & b_K^L \end{pmatrix}$$

and denote by $S = B^T - A^T = (\nu_1, \dots, \nu_L)$ the *stoichiometry matrix* of system (2.1). Accordingly, the vector ν_i specifies the net change in the molecular abundances associated with reaction i . In order to describe the temporal evolution of a chemically reacting system, we next introduce the time-dependent system state $X(t) = (X_1(t), \dots, X_K(t))^T \in \mathbb{N}^K$, collecting the molecular abundances of the species $\mathcal{X}_1, \dots, \mathcal{X}_K$ at time t^1 . Assuming that the molecular system is well-mixed, it can be shown that $X(t)$ follows a continuous-time Markov chain (CTMC) [24], which we denote by X . A CTMC is characterized by integer-valued state trajectories exhibiting discontinuities (i.e., *jumps*) whenever a reaction fires. The transition dynamics of such a CTMC are defined through the probabilities of a reaction happening within a small amount of time together with the probability that the system remains in its current state, i.e.,

$$P(X(t + \Delta t) = x + \nu_i \mid X(t) = x) = h_i(x, c_i) \Delta t + o(\Delta t) \quad \forall i = 1, \dots, L \tag{2.2}$$

$$P(X(t + \Delta t) = x \mid X(t) = x) = 1 - \sum_{i=1}^L h_i(x, c_i) \Delta t - o(\Delta t), \tag{2.3}$$

¹We will generally follow the convention to denote random quantities and their realization as upper and lower case letters, respectively. This also applies for greek symbols.

with $h_i(x, c_i)$ as the *hazard function* and $c_i \in \mathbb{R}$ as the kinetic parameter of reaction i^2 . The term $o(\Delta t)$ refers to the probability that multiple reactions happen within Δt , which for $\Delta t \rightarrow 0$ goes to zero much faster than $h_i(x, c_i)\Delta t$. Throughout this work we assume mass-action kinetics [7] such that $h_i(x, c_i) = c_i g_i(x)$ with

$$g_i(x) = \prod_{k=1}^K \binom{x_k}{a_k^i}.$$

A central property of CTMCs (and jump processes in general) are the random waiting times between two consecutive reactions. Assuming that a CTMC is in state x at time t , we may characterize the waiting time W_i for the i -th reaction channel by a cumulative distribution function (CDF) $P(W_i < w \mid X(t) = x)$. In order to find that CDF, we first note that the probability that W_i is larger than some small Δw is given by $P(W_i > \Delta w \mid X(t) = x) = 1 - h_i(x, c_i)\Delta w - o(\Delta w)$. Furthermore, due to the Markov property of X , the transition probabilities only depend on the current state $X(t)$ and not on how long the chain has already remained in that state. As a consequence, the probability $P(W_i > w + \Delta w \mid X(t) = x)$ factorizes as

$$\begin{aligned} P(W_i > w + \Delta w \mid X(t) = x) \\ &= P(W_i > w + \Delta w \mid X(t) = x, W_i > \Delta w)P(W_i > \Delta w \mid X(t) = x) \\ &= P(W_i > w \mid X(t) = x)(1 - h_i(x, c_i)\Delta w - o(\Delta w)) \end{aligned} \quad (2.4)$$

where the fact that $P(W_i > w + \Delta w \mid X(t) = x, W_i > \Delta w) = P(W_i > w \mid X(t) = x)$ is referred to as the *memoryless* property of CTMCs. The derivative of $P(W_i > w \mid X(t) = x)$ is then given by

$$\begin{aligned} &\frac{d}{dw}P(W_i > w \mid X(t) = x) \\ &= \lim_{\Delta w \rightarrow 0} \frac{P(W_i > w + \Delta w \mid X(t) = x) - P(W_i > w \mid X(t) = x)}{\Delta w} \\ &= \lim_{\Delta w \rightarrow 0} \frac{P(W_i > w \mid X(t) = x)(1 - h_i(x, c_i)\Delta w + o(\Delta w)) - P(W_i > w \mid X(t) = x)}{\Delta w} \\ &= P(W_i > w \mid X(t) = x) \lim_{\Delta w \rightarrow 0} (-h_i(x, c_i) + o(\Delta w)) \\ &= -h_i(x, c_i)P(W_i > w \mid X(t) = x), \end{aligned} \quad (2.5)$$

which means that $P(W_i > w \mid X(t) = x)$ satisfies a linear homogeneous ODE with a constant coefficient. With the initial condition $P(W_i > 0 \mid X(t) = x) = 1$, the waiting time distribution is found to be

$$P(W_i < w \mid X(t) = x) = 1 - P(W_i > w \mid X(t) = x) = 1 - e^{-h_i(x, c_i)w}, \quad (2.6)$$

²In this section we assume all kinetic parameters to be given. For the sake of simplicity, they are not considered as parameters of the respective probabilities and densities. However, the reader should keep in mind that (2.2) and (2.3) as well as derived quantities do depend on those parameters.

or equivalently,

$$W_i \mid (X(t) = x) \sim \text{Exp}(h_i(x, c_i)) = p_i(w \mid x), \quad (2.7)$$

with $p(w \mid x) = P(W_i \in [w, w + dw] \mid X(t) = x)$ as the probability density function (PDF) of W_i . Exponential waiting-time distributions are in fact a distinguishing property of CTMCs when compared to more general process classes such as renewal processes [57, 85].

2.1.1 Kolmogorov Equations

A natural quantity in the context of stochastic reaction dynamics is the probability of finding the system in state x at time t given that it was in some other state x' at time $t' < t$ denoted by $P(X(t) = x \mid X(t') = x')$. Although this probability cannot be found explicitly in all but the simplest scenarios, its temporal evolution may be given in form of an ordinary differential equation, which in principle, can be solved numerically. We realize that $P(X(t + \Delta t) = x \mid X(t') = x')$ can be factorized as

$$\begin{aligned} P(X(t + \Delta t) = x \mid X(t') = x') &= \underbrace{\sum_{i=1}^L \left(h_i(x - \nu_i, c_i) \Delta t + o(\Delta t) \right) P(X(t) = x - \nu_i \mid X(t') = x')}_{\text{Probability of moving from } x - \nu_i \text{ to } x \text{ within } \Delta t} \\ &+ \underbrace{\left(1 - \sum_{i=1}^L h_i(x, c_i) \Delta t + o(\Delta t) \right) P(X(t) = x \mid X(t') = x')}_{\text{Probability that system was already in state } x \text{ at time } t}. \end{aligned} \quad (2.8)$$

Using the definition of the differentiation operator we can then show that $P(X(t) = x \mid X(t') = x')$ satisfies the differential equation [24, 37]

$$\begin{aligned} \frac{d}{dt} P(X(t) = x \mid X(t') = x') &= \lim_{\Delta t \rightarrow 0} \frac{P(X(t + \Delta t) = x \mid X(t') = x') - P(X(t) = x \mid X(t') = x')}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{\sum_{i=1}^L \left(h_i(x - \nu_i, c_i) \Delta t + o(\Delta t) \right) P(X(t) = x - \nu_i \mid X(t') = x')}{\Delta t} \\ &\quad - \lim_{\Delta t \rightarrow 0} \frac{\left(\sum_{i=1}^L h_i(x, c_i) \Delta t + o(\Delta t) \right) P(X(t) = x \mid X(t') = x')}{\Delta t} \\ &= \sum_{i=1}^L h_i(x - \nu_i, c_i) P(X(t) = x - \nu_i \mid X(t') = x') \\ &\quad - \sum_{i=1}^L h_i(x, c_i) P(X(t) = x \mid X(t') = x'), \end{aligned} \quad (2.9)$$

with initial condition $P(X(t') = x | X(t') = x') = \delta_{xx'}$, whereby $\delta_{xx'}$ indicates the Kronecker delta function in x and x' . Eq. (2.9) is most commonly known as the chemical master equation (CME). Within the theory of Markov chains, it is also referred to as *Kolmogorov-forward equation*, since it determines how the probability mass is propagated as time increases. It is sometimes useful to write down a differential equation also for the marginal probabilities

$$P(X(t) = x) = \sum_{x' \in \mathbb{X}} P(X(t) = x | X(t') = x') P(X(t') = x'). \quad (2.10)$$

Differentiation with respect to time yields

$$\begin{aligned} \frac{d}{dt} P(X(t) = x) &= \sum_{x' \in \mathbb{X}} \frac{d}{dt} P(X(t) = x | X(t') = x') P(X(t') = x') \\ &= \sum_{i=1}^L h_i(x - \nu_i, c_i) \underbrace{\sum_{x' \in \mathbb{X}} P(X(t) = x - \nu_i | X(t') = x') P(X(t') = x')}_{P(X(t)=x-\nu_i)} \\ &\quad - \sum_{i=1}^L h_i(x, c_i) \underbrace{\sum_{x' \in \mathbb{X}} P(X(t) = x | X(t') = x') P(X(t') = x')}_{P(X(t)=x)} \\ &= \sum_{i=1}^L h_i(x - \nu_i, c_i) P(X(t) = x - \nu_i) - \sum_{i=1}^L h_i(x, c_i) P(X(t) = x), \end{aligned} \quad (2.11)$$

which further implies that $P(X(t) = x)$ satisfies the CME when the initial distribution is set to $P(X(t') = x')$.

In contrast to the forward equation, the so-called *Kolmogorov-backward equation* is centered around the question: if one finds the system in state x at time t , how likely was the system in state x' at time $t' < t$?³ Technically, it can be derived similarly to the forward equation (2.9), whereas the dynamics are computed with respect to the initial variables x' and t' instead of x and t [37], i.e.,

$$\begin{aligned} \frac{d}{dt'} \tilde{P}(X(t) = x | X(t') = x') &= \sum_{i=1}^L h_i(x', c_i) \tilde{P}(X(t) = x | X(t') = x') \\ &\quad - \sum_{i=1}^L h_i(x', c_i) \tilde{P}(X(t) = x | X(t') = x' + \nu_i). \end{aligned} \quad (2.12)$$

We remark that if the initial condition of the forward equation coincides with the terminal condition of the backward equation, i.e., $\tilde{P}(X(t') = x | X(t') = x') = P(X(t') = x |$

³Alternatively, this can be expressed by the question: how does the probability of being in state x at time t depend on the initial state x' at time t' .

$X(t') = x') = \delta_{xx'}$, we have that $\tilde{P}(X(t) = x \mid X(t') = x') = P(X(t) = x \mid X(t') = x')$ for all $t > t'$. Consequently, the probability of finding a reaction network in a particular state at a particular time satisfies both the forward- and backward equations. The latter also provides a means to compute backward probabilities $P(X(t') = x' \mid X(T) = x) \propto \tilde{P}(X(T) = x \mid X(t') = x')$ for $T > t'$ by solving (2.12) backward in time with initial condition $\tilde{P}(X(T) = x \mid X(T) = x') = \delta_{xx'}$. However, one should keep in mind that the result is not a valid probability distribution in x' and thus, requires a proper normalization.

2.1.2 The Random Time-Change Model

While the CME represents one of the most important and widely used modeling approaches for stochastic biochemical systems, it is characterized by a few major drawbacks. For instance, it typically suffers from a so-called state space explosion, meaning that the number of reachable states scales very badly with the number of chemical species and their abundance – even if one considers only the states which are reached with a reasonably high probability [81]. Furthermore, it does not offer a complete description of a CTMC or realizations thereof, since it only contains information about the molecular state $X(t)$, but not about the sequence of reactions through which that state was reached. Especially when dealing with time-lapse measurements, mathematical descriptions of the entire realizations – or *paths* – turn out to be of vital importance (see e.g., Chapters 4).

This closely resembles the case of real-valued Markov processes whose Kolmogorov-forward equation is given by a PDE (i.e., the *Fokker-Planck equation* [37]), which due to its analytical and numerical complexity, is barely used for process analysis. Instead, it often appears beneficial to work with the corresponding *process equations* (i.e., stochastic differential equations, SDEs), facilitating the use of highly developed techniques from stochastic calculus. Although far less prevalent in the field of systems biology, such process equations exist also in case of CTMCs. In particular, it can be shown that the solution of the CTMC defined in (2.2) and (2.3) is given through the *random time-change model* (RTM) as

$$X(t) = X(0) + \sum_{i=1}^L Y_i \left(\int_0^t h_i(X(s), c_i) ds \right) \nu_i, \quad (2.13)$$

where Y_1, \dots, Y_L are independent Poisson processes with intensity one⁴ [7]. Eq. (2.13) is a stochastic integral equation, where the state $X(t)$ enters the r.h.s. through the reaction counters Y_i . More specifically, the original time argument t is replaced by a state-dependent function $\int_0^t h_i(X(s), c_i) ds$. Intuitively this can be understood as randomly changing the speed at which the time of a unit Poisson process passes by. Note that eq. (2.13) can also be written in SDE-form. Via the *Doob-Meyer* decomposition

⁴ Y_i is then said to be a unit Poisson process.

theorem, it can be shown that the Poissonian reaction counters Y_i can be decomposed into a predictable part and a martingale, i.e.,

$$Y_i \left(\int_0^t h_i(X(s), c_i) ds \right) = \int_0^t h_i(X(s), c_i) ds + \tilde{Q}_i \left(\int_0^t h_i(X(s), c_i) ds \right), \quad (2.14)$$

with $\mathbb{E} \left[\tilde{Q}_i \left(\int_0^t h_i(X(s), c_i) ds \right) \right] = 0$. To keep the notation compact, we further define $Q_i(t) := \tilde{Q}_i \left(\int_0^t h_i(X(s), c_i) ds \right)$, meaning that the state-dependent transformation $\int_0^t h_i(X(s), c_i) ds$ is subsumed into $Q_i(t)$. In differential notation eq. (2.14) reads

$$dY_i(t) = h_i(X(t^-), c_i) dt + dQ_i(t), \quad (2.15)$$

where $X(t^-)$ denotes the state of the system immediately before the next jump happens⁵. The solution of the CTMC X thus satisfies the SDE

$$dX(t) = \left(\sum_{i=1}^L h_i(X(t^-), c_i) \nu_i \right) dt + \sum_{i=1}^L dQ_i(t) \nu_i. \quad (2.16)$$

In the following, we will denote by $\mathbf{X}_t \in \mathcal{S}[0, t]$ the stochastic path of the CTMC X on a time interval $[0, t]$, consisting of the random jump-times and types of events. In cases where all the stoichiometric vectors ν_1, \dots, ν_L are distinct, this path can be equivalently defined as $\mathbf{X}_t = \{X(s) \in \mathbb{X} \mid s \in [0, t]\}$. Accordingly, we can assign \mathbf{X}_t a path density which describes the distribution over all possible sample paths of X between time zero and t . We remark that a rigorous treatment of such a density is beyond the scope of this work (see e.g., [59]) but their interpretation as a likelihood function turns out to be straightforward. The latter is particularly important in the context of inference because it allows to quantify how likely a given path \mathbf{x}_t has been realized under a set of rate constants $c = \{c_1, \dots, c_L\}$. More specifically, it turns out that the likelihood function over a path instance \mathbf{x}_t can be written as [62, 129]

$$p(\mathbf{x}_t \mid c) = \prod_{i=1}^L c_i^{r_i} e^{-c_i \int_0^t g_i(x(s)) ds}, \quad (2.17)$$

with $r_i := r_i(\mathbf{x}_t)$ as the number of reactions.

2.1.3 Stochastic Simulation Algorithm

Although CTMCs are barely tractable analytically, it is straightforward to simulate sample paths thereof. A general algorithm for simulating arbitrary reaction networks

⁵This is required for the SDE to be causal. Technically $X(t^-)$ is given through a left-sided limit, i.e., $X(t^-) = \lim_{s \rightarrow t^-} X(s)$. In case no jump happens at t , we have that $X(t^-) = X(t)$. If a jump happens at t , $X(t^-)$ will correspond to the state of the CTMC immediately before that jump.

was proposed in [44], although equivalent schemes were published way earlier in the mathematical community (e.g., [26]). In principle, the algorithm and its variants are based on drawing exponential waiting-times for each of the reaction channels and picking the next reaction according to the smallest one. A standard implementation of this algorithm is given by the *first reaction method* and is outlined in Algorithm 1. We remark that several variants of the original algorithm have been proposed recently [42,97], which were shown to achieve a substantial reduction in computational effort.

Algorithm 1 (First reaction method). *The algorithm takes the stoichiometric change vectors ν_1, \dots, ν_L , the kinetic parameters c_1, \dots, c_L , the final time T and the initial state x_0 as input and returns the reaction types together with the associated firing times.*

```

1: Initialize variables  $t \leftarrow 0$  and  $x \leftarrow x_0$ .
2: while  $t < T$  do
3:   for  $i = 1, \dots, L$  do
4:     Draw waiting-time for the  $i$ -th reaction channel  $W_i \sim \text{Exp}(h_i(x, c_i))$ .
5:   end for
6:   Choose reaction associated with the minimal waiting-time  $k \leftarrow \underset{i=1, \dots, L}{\text{argmax}} W_i$ .
7:   Update state  $x \leftarrow x + \nu_k$ .
8:   Update time  $t \leftarrow t + W_k$ .
9:   Output  $t$  and  $x$ .
10: end while

```

2.1.4 Moment Dynamics

We have indicated in Section 2.1.2 that analyzing stochastic models based on probability distributions suffers from severe scalability issues. In many cases – however – already a few summary statistics might describe a dynamical system sufficiently well [82,136]. To this end, we realize that the moments of a stochastic kinetic model are straightforward to derive from the CME. Multiplying both sides of (2.9) with a polynomial $f : \mathbb{R}^K \mapsto \mathbb{R}$ and summing over all possible states yields

$$\begin{aligned}
\sum_{x \in \mathbb{X}} f(x) \frac{d}{dt} P(X(t) = x) &= \frac{d}{dt} \mathbb{E}[f(x)] \\
&= \sum_{i=1}^L \sum_{x \in \mathbb{X}} f(x) h_i(x - \nu_i, c_i) P(X(t) = x - \nu_i) - \sum_{i=1}^L \sum_{x \in \mathbb{X}} f(x) h_i(x, c_i) P(X(t) = x).
\end{aligned} \tag{2.18}$$

Using a variable transformation $\tilde{x} = x - \nu_i$, we further obtain

$$\begin{aligned} \frac{d}{dt} \mathbb{E}[f(X(t))] &= \sum_{i=1}^L \sum_{\tilde{x} \in \mathbb{X}} f(\tilde{x} + \nu_i) h_i(\tilde{x}, c_i) P(X(t) = \tilde{x}) \\ &\quad - \sum_{i=1}^L \sum_{x \in \mathbb{X}} f(x) h_i(x, c_i) P(X(t) = x) \\ &= \sum_{i=1}^L \mathbb{E}[f(X(t) + \nu_i) h_i(X(t), c_i)] - \sum_{i=1}^L \mathbb{E}[f(X(t)) h_i(X(t), c_i)]. \end{aligned} \quad (2.19)$$

We remark that (2.19) is not generally *closed*, which means that the r.h.s. may depend on moments of higher order than $\mathbb{E}[f(X(t))]$. In particular, this is the case as soon as a kinetic model is non-linear or equivalently, if it involves reactions of order two and above. As a consequence, (2.19) gives rise to an infinite-dimensional system of ODEs, demanding for suitable approximations – known as *moment-closure* schemes (see e.g., [52] and Section 3.2).

In contrast, linear models generally close at order two, such that the dynamics can be described by a finite-dimensional ODE system. The following example illustrates the calculation of moments by means of a simple birth process.

Example 1 (Moments of a birth process (CME)). *The CME for a linear birth process with rate constant c is given by*

$$\frac{d}{dt} P(X(t) = x) = cP(X(t) = x - 1) - cP(X(t) = x). \quad (2.20)$$

With $f(x) = x$ we obtain for the mean

$$\begin{aligned} \frac{d}{dt} \mathbb{E}[X(t)] &= c \sum_{x=0}^{\infty} x P(X(t) = x - 1) - c \sum_{x=0}^{\infty} x P(X(t) = x) \\ &= c \sum_{\tilde{x}=0}^{\infty} (\tilde{x} + 1) P(X(t) = \tilde{x}) - c \sum_{x=0}^{\infty} x P(X(t) = x) \\ &= c \mathbb{E}[X(t)] + c - c \mathbb{E}[X(t)] \\ &= c \end{aligned} \quad (2.21)$$

and with $f(x) = x^2$ for the second-order moment

$$\begin{aligned} \frac{d}{dt} \mathbb{E}[X^2(t)] &= c \sum_{x=0}^{\infty} x^2 P(X(t) = x - 1) - c \sum_{x=0}^{\infty} x^2 P(X(t) = x) \\ &= c \sum_{\tilde{x}=0}^{\infty} (\tilde{x} + 1)^2 P(X(t) = \tilde{x}) - c \sum_{x=0}^{\infty} x^2 P(X(t) = x) \\ &= c \mathbb{E}[X^2(t)] + 2c \mathbb{E}[X(t)] + c - c \mathbb{E}[X^2(t)] \\ &= 2c \mathbb{E}[X(t)] + c. \end{aligned} \quad (2.22)$$

Although the CME-based technique to derive moment equations appears to be standard in the field of computational biology, we would like to briefly discuss an alternative approach, which relies on the path-wise CTMC description from Section 2.1.2. In particular it is based on the extension of Ito's lemma for counting processes (see e.g., [63]). Although a general treatment is possible, we explain the approach using the following simple example.

Example 2 (Moments of a birth process (SDE)). *The solution of a linear birth process with rate constant c satisfies the SDE*

$$dX(t^-) = cdt + dQ(t) \quad (2.23)$$

where Q is a martingale. Taking the expectation on both sides and dividing by dt yields

$$\frac{d}{dt} \mathbb{E}[X(t)] = c. \quad (2.24)$$

In order to calculate the second-order moment, we first need an SDE describing $X^2(t)$. For that sake, we make use of Ito's formula for counting processes, i.e.,

$$dF(X(t)) = [F(X(t^-) + 1) - F(X(t^-))] dX(t). \quad (2.25)$$

With $F(X) = X^2$, we further obtain

$$\begin{aligned} dX^2(t) &= [(X^2(t^-)) + 2X(t^-) + 1 - X^2(t^-)] dX(t) \\ &= [2X(t^-) + 1] dX(t) \\ &= [2X(t^-) + 1] [cdt + dQ(t)] \\ &= c [2X(t^-) + 1] dt + [2X(t^-) + 1] dQ(t). \end{aligned} \quad (2.26)$$

Since the expected values of all terms multiplied by $dQ(t)$ are zero, we obtain for the second order moment

$$\frac{d}{dt} \mathbb{E}[X^2(t)] = 2c\mathbb{E}[X(t)] + c, \quad (2.27)$$

coinciding with the results from Example 1.

2.1.5 Random Environments and Extrinsic Noise

Although CTMC models of stochastic reaction dynamics can be derived from biophysical principles, they are often unable to explain well the cell-to-cell variability observed experimentally. To some extent, such discrepancies may stem from wrong or oversimplified assumptions on the reaction kinetics. The main reason – however – was shown to be the so-called *extrinsic* contribution to the overall variability, arising from unmodeled stochastic effects in a cell's microenvironment [33, 120]. Mathematically, one can account for extrinsic noise by introducing a CTMC that on top of the kinetic parameters c_1, \dots, c_L , also depends on certain environmental variables [53, 59, 120, 136, 137],

which we refer to as *extrinsic factors*. Examples for such factors are the number of ribosomes, the cell-cycle stage, the cell-size and so forth. Most extrinsic factors are believed to dynamically change over time [53]. However, when considering only short time intervals they are often assumed to stay roughly the same, for instance over the duration of an experiment [136, 137]. Constant extrinsic factors can substantially simplify the analysis of kinetic models and – from a practical viewpoint – may be distinguished from the more complicated class of temporally fluctuating environments. Nonetheless, the former can be considered a special case of the latter, which allows us to introduce a general mathematical modeling framework that will be used throughout Chapters 3-5.

2.1.5.1 Mixed-Effect Modeling

We assume that a stochastic biochemical network can be described by a *conditional* Markov chain $X \mid (S, Z)$ with $S = \{S_1, \dots, S_I\}$ and $Z = \{Z_1, \dots, Z_J\}$. The shared parameters could for instance include kinetic parameters associated with elementary reactions that are believed to be the same among cells [59]. In contrast, the extrinsic factors Z are assumed to vary between individual cells. More specifically, we assume that for each cell $m = 1, \dots, M$, the corresponding extrinsic factors are drawn from a common distribution $Z^m \mid (A = a) \sim p(\cdot \mid a)$, with A as a set of *extrinsic statistics* controlling $p(\cdot \mid a)$. As a consequence, the population of M is described by an ensemble of conditional Markov chains $X^m \mid (S, Z^m)$ for $m = 1, \dots, M$. The resulting models have gained more and more attention in the past years and are commonly known as *mixed-effect-* or *heterogeneous Markov models*. From a Bayesian point of view, such models fall into the class of hierarchical Bayesian networks [15].

A central problem of such models is that they do not scale well with the population size M , since for every considered cell m , one introduces another Markov chain $X^m \mid (S, Z^m)$. In contrast, a homogeneous population can be considered as M realizations of exactly the same Markov chain $X \mid S$. A significant portion of this thesis deals with overcoming such scalability issues for the purpose of inference (Chapter 3 and 4) and analysis (Chapter 5).

In the following, we will briefly explain two possible approaches for modeling extrinsic noise in biochemical networks.

2.1.5.2 Variability over Conserved Species

The first approach is based on the idea that at a given time point, one cell has a different molecular configuration than another cell. For instance, when considering an enzymatic reaction, a cell having many enzyme copies will facilitate a higher substrate-product turnover than a cell bearing only little enzyme. Accordingly, a heterogeneous cell population could be modeled through a randomly drawn enzyme abundance. In a straight-forward attempt, this could be implemented by assigning an initial distribution over the enzyme. However, we realize that the enzyme may exist in different states, i.e., most of the time, a fraction will be bound by the substrate and thus, only the

total number of enzyme molecules is constant or *conserved*. We therefore, assign the variability to the total number of enzyme molecules, regardless of their respective state. The following approach is based on extending this idea to the general case, where extrinsic variability enters a biochemical network through arbitrary conservation laws.

We first need a principled way to identify the conservation laws of a biochemical reaction network. For that sake, we realize that the state space of any network satisfies $\mathbb{X} = \{X \in \mathbb{N}^K : NX = B\}$, with $N \in \mathbb{N}^{l \times K}$ as the smallest positive integer basis of the null space of the stoichiometry matrix S . Matrix N then represents the l conservation laws of the reaction network. The corresponding conservation constants are collected in $B \in \mathbb{N}^l$ and determine the total number of species for each each of the l conservation laws. Considering a population of size M , each cell m would have a different conservation constant B^m . Assuming that all heterogeneity comes from a variability in conserved species, we can then set $Z^m \equiv B^m$ and define a multivariate probability distribution $p(z^m | a) = p(\cdot | a)$ for $m = 1, \dots, M$ accordingly. The shared parameters are then simply the kinetic rate constants, i.e., $S = \{C_1, \dots, C_L\}$.

In order to perform inference, we need an explicit dependency of the CTMC model on the random conservation constants Z . This dependency will be manifested in the hazard functions involving some of the conserved species. In particular, the conservation laws allow us to express some of the chemical species as a function of some other species and the conservation constants. Assuming a certain partitioning of the state space, we can write

$$NX = (\tilde{N} \quad \bar{N}) \begin{pmatrix} \tilde{X} \\ \bar{X} \end{pmatrix} = \tilde{N}\tilde{X} + \bar{N}\bar{X} = Z, \quad (2.28)$$

with $\tilde{N} \in \mathbb{N}^{l \times (K-l)}$ and $\bar{N} \in \mathbb{N}^{l \times l}$. Reformulating with respect to \bar{X} further yields

$$\bar{X} = \bar{N}^{-1}(Z - \tilde{N}\tilde{X}), \quad (2.29)$$

where the partitioning of the state vector X needs to be chosen such that \bar{N} has full rank. Relation (2.29) allows us to rewrite the hazard functions as $h_i(X, C_i) \equiv h_i(\tilde{X}, Z, C_i)$ for $i = 1, \dots, L$ such that we obtain an explicit dependency on the extrinsic factors Z . Modeling extrinsic variability through conserved species for the purpose of inference is demonstrated [59, 136] and Section 3. We remark, however, that the obtained hazard functions may be of complex form such that resulting inference problems become analytically complex.

2.1.5.3 Variability over Rate Constants

A more widely used and significantly simpler way of modeling extrinsic noise is to assume that some of the kinetic rate constants C_1, \dots, C_L randomly vary between cells [34, 50, 135] or fluctuate over time [53, 110]. Without loss of generality, we assume a particular ordering of the reactions such that $C = \{Z_1, \dots, Z_J, S_1, \dots, S_I\}$. Since we assume mass-action kinetics throughout this work, this modeling approach has the

advantage that the environment enters the target network X linearly, which will allow us to perform several important calculations analytically (see Chapters 4 and 5). Additionally, it readily extends to the case of fluctuating environments.

2.2 Bayesian Statistics

In contrast to traditional *frequentist* statistics, Bayesian modeling approaches allow the inclusion of a-priori knowledge about *latent* variables⁶. In particular, those variables are considered random themselves and are therefore, associated with respective *prior distributions*. Once a measurement is obtained, this prior distribution is used in conjunction with *Bayes theorem* to calculate the *posterior distribution*, describing the updated belief about the latent variable. In a subsequent experiment, this posterior distribution is used as a prior distribution and so forth. The more measurements are included, the more accurate the posterior distribution becomes.

Let us consider a general Bayesian model where a set of latent variables $X \in \mathbb{X}$ is observed through measurements $Y \in \mathbb{Y}$, i.e.,

$$Y \mid (X = x) \sim p(y \mid x),$$

with $p(y \mid x)$ as the conditional measurement density. In the context of inference, the latter is often referred to as the *likelihood function*, although in that case, x is considered the independent variable instead of y . The latent variables X are associated with a prior distribution, i.e., $X \sim p(x)$. The statistical model is then characterized by its joint probability distribution

$$p(x, y) = p(y \mid x)p(x). \quad (2.30)$$

Observing that $p(x, y) = p(x \mid y)p(y)$ and rearranging with respect to $p(x \mid y)$, we obtain the well-known Bayes' formula

$$p(x \mid y) = \frac{p(y \mid x)p(x)}{p(y)} = \frac{p(y \mid x)p(x)}{\int_{\mathbb{X}} p(y \mid x)p(x)dx}, \quad (2.31)$$

with $p(x \mid y)$ as the posterior distribution and $p(y) = \int_{\mathbb{X}} p(y \mid x)p(x)dx$ the *evidence-* or *marginal likelihood* function. The latter plays a central role in several areas of Bayesian statistics such as model selection (see Section 2.2.3) or sparse learning problems (see Section 4.3) [14, 102]. In the case of parameter inference – however – it often turns out to be of minor importance since it can be considered a scaling constant independent of x . Example 3 demonstrates the computation of the posterior distribution over kinetic parameters given a complete sample path of a CTMC.

⁶Latent means that the variable is not observed directly, but rather through some other variables of the model.

Example 3 (Posterior distribution over kinetic parameters). We know from eq. (2.17) that the likelihood function of a sample path \mathbf{x}_t with respect to the kinetic parameters c_1, \dots, c_L is given by

$$p(\mathbf{x}_t \mid c_1, \dots, c_L) = \prod_{i=1}^L c_i^{r_i} e^{-c_i \int_0^t g_i(x(s)) ds}. \quad (2.32)$$

We assume now that we have prior knowledge about the kinetic parameters in form of independent Gamma distributions, i.e.,

$$p(c_1, \dots, c_L) = \prod_{i=1}^L \Gamma(c_i; \alpha_i, \beta_i), \quad (2.33)$$

with α_i and β_i as the shape- and inverse scale parameters of the respective distribution. The posterior distribution then becomes [129]

$$\begin{aligned} p(c_1, \dots, c_L \mid \mathbf{x}_t) &\propto p(\mathbf{x}_t \mid c_1, \dots, c_L) p(c_1, \dots, c_L) \\ &= \prod_{i=1}^L c_i^{r_i} e^{-c_i \int_0^t g_i(x(s)) ds} \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} c_i^{\alpha_i-1} e^{-\beta_i c_i} \\ &= \prod_{i=1}^L \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} c_i^{\alpha_i+r_i-1} e^{-c_i(\beta_i + \int_0^t g_i(x(s)) ds)} \\ &\propto \prod_{i=1}^L \Gamma\left(c_i; \alpha_i + r_i, \beta_i + \int_0^t g_i(x(s)) ds\right), \end{aligned} \quad (2.34)$$

i.e., is again given by a product of independent Gamma distributions. This further implies that the Gamma distribution is a conjugate prior [39] to the path likelihood of a CTMC with linearly parameterized propensity functions.

We remark that in most practically relevant scenarios, (2.31) cannot be found in closed form such that suitable numerical methods need to be employed. Those can generally be divided into two subgroups, i.e., analytical- and sampling-based methods [14]. The former are typically characterized by low computational costs but on the other hand, only reveal approximations of the true posterior distribution. This will be demonstrated in more detail in Section 4.3 where we apply a variational approximation to the inference of a high-dimensional posterior distribution [11]. However, most parts of this thesis will rely on sampling-based approaches, since in general they allow to draw samples from the exact target posterior distribution. An extensive discussion of such computation schemes is not in the scope of this work, but further details will be provided in the respective sections of this thesis. Additional information can for instance be found in [14].

2.2.1 Predictive Distributions

Once a posterior distribution has been obtained through an initial measurement Y_1 , it can be used to predict the outcome of the next experiment. In particular, the probability distribution of the second measurement Y_2 is given through the so-called *predictive distribution*, i.e.,

$$Y_2 \mid (Y_1 = y_1) \sim p(y_2 \mid y_1) = \int_{\mathbb{X}} p(y_2 \mid x)p(x \mid y_1)dx. \quad (2.35)$$

For the j -th measurement the predictive distribution is given by

$$p(y_j \mid y_1, \dots, y_{j-1}) = \int_{\mathbb{X}} p(y_j \mid x)p(x \mid y_1, \dots, y_{j-1})dx, \quad (2.36)$$

with $p(x \mid y_1, \dots, y_{j-1})$ as the posterior distribution obtained from the $(j - 1)$ -th experiment. Note that if $j = 1$, the predictive distribution corresponds to the marginal likelihood of (2.31). Eq. (2.36) turns out to be very useful to assess the predictive power of models that have been trained against data. In particular, due to the integration over the previous posterior distribution, it takes into account the parameter uncertainty that remains a-posteriori. As a consequence, large models are typically less predictive than simple models, since they involve many (unknown) parameters that need to be estimated from data. In a Bayesian sense, highly predictive models will thus represent a compromise between model accuracy and complexity.

2.2.2 Bayesian Filtering and Smoothing

In many practical scenarios we are interested in estimating a time-dependent rather than a static quantity such as for instance the state of a continuous-time Markov process $X(t)$, characterized by a transition kernel

$$X(t) \mid (X(s) = x') \sim P(X(t) = x \mid X(s) = x'), \quad (2.37)$$

with $s \neq t$ and $P(X(t) = x \mid X(s) = x')$ given through the Kolmogorov-forward and backward equations, according to whether $s < t$ or $s > t$, respectively. Typically, we can obtain noisy measurements of the state at discrete time points which are characterized through a measurement density, i.e.,

$$Y_l \mid (X(t_l) = x) \sim p(y \mid x) \quad (2.38)$$

for $l = 1, \dots, N$. Eqs. (2.37) and (2.38) define a so-called general state space model. If $X(t)$ is Markovian, they are also referred to as hidden Markov models (HMMs, see Figure 2.1) and their inference is a central topic in the statistical- and machine learning disciplines [8, 15, 56].

Let us assume we want to estimate the state $X(t_n)$ from a sequence of $n < N$ measurements taken within an interval $[0, t_n]$. In a Bayesian scenario, this is achieved

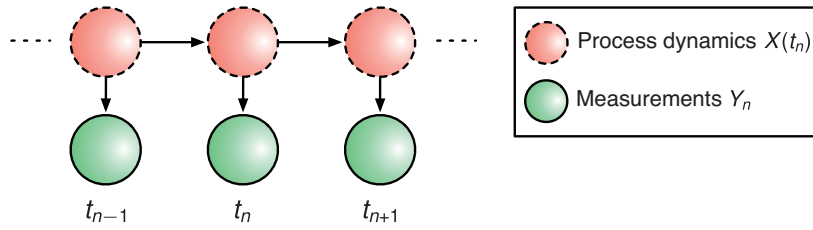


Figure 2.1: Bayesian network of a hidden Markov model. The latent process $X(t)$ (dashed circles) can be observed through noisy measurements Y_n taken discrete at time points t_n .

by computing the posterior distribution $P(X(t_n) = x \mid y_{1:n})$ with $y_{1:n} = \{y_1, \dots, y_n\}$. The latter is also referred to as the *filtering distribution* and using Bayes formula, it is straightforward to show that it satisfies the recurrence relation

$$P(X(t_n) = x \mid y_{1:n}) \propto p(y_n \mid x) \sum_{x' \in \mathbb{X}} P(X(t_n) = x \mid X(t_{n-1}) = x') P(X(t_{n-1}) = x' \mid y_{1:n-1}), \quad (2.39)$$

with $P(X(t_n) = x \mid X(t_{n-1}) = x')$ given by the Kolmogorov-forward equation and $P(X(t_{n-1}) = x' \mid y_{1:n-1})$ as the filtering distribution at time t_{n-1} . Note that the filtering distribution is straightforward to compute for time points later than the last measurement time, i.e.,

$$P(X(t) = x \mid y_{1:n}) = \sum_{x' \in \mathbb{X}} P(X(t) = x \mid X(t_n) = x') P(X(t_n) = x' \mid y_{1:n}), \quad (2.40)$$

for $t > t_n$. Eq. (2.40) represents the knowledge about the state at time t without incorporating additional measurements and therefore, its computation is often referred to as the *prediction* step. After a new measurement has been obtained at t_{n+1} we again apply (2.39) such that the distribution exhibits an instantaneous jump (i.e., it is *updated*).

Inference schemes based on (2.39) and (2.40) are generally referred to as *recursive Bayesian estimation* techniques. We remark that the sum in (2.39) is analytically tractable for only a few special cases. The most famous analytical solution of the Bayesian recursion is known as the *Kalman filter* which applies if both the measurement- and process distributions are Gaussian.

The computation of the filtering distribution is particularly useful for on-line (or tracking) applications where the observations are measured and processed one after each other. In contrast, if all N measurements are collected beforehand, better estimates can be achieved using the *smoothing distribution* $p(X(t_n) = x \mid y_{1:N})$, since it incorporates all information available (i.e., also future measurements). Importantly, the smoothing distribution can be written as a product of two independent filters [18]

$$P(X(t_n) = x \mid y_{1:N}) = \frac{P(X(t_n) = x \mid y_{1:n-1}) p(y_{n:N} \mid X(t_n) = x)}{p(y_{n:N} \mid y_{1:n-1})}, \quad (2.41)$$

where $P(X(t_n) = x \mid y_{1:n-1})$ is just the filtering distribution from (2.40) and $p(y_{n:N} \mid X(t_n) = x)$ is given through the *backward information filter*

$$\begin{aligned} p(y_{n:N} \mid X(t_n) = x) \\ = p(y_n \mid x) \sum_{x' \in \mathbb{X}} P(X(t_{n+1}) = x' \mid X(t_n) = x) p(y_{n+1:N} \mid X(t_{n+1}) = x'), \end{aligned} \quad (2.42)$$

with $p(y_{n+1:N} \mid X(t_{n+1}) = x')$ as the result from time point t_{n+1} . We note that the sum is a function in x and t_n and therefore, can be computed using the Kolmogorov-backward equation together with an initial condition of $p(y_{n+1:N} \mid X(t_{n+1}) = x')$. Finally, the smoothing distribution is obtained by multiplying both filters together and rescaling it such that it sums up to one.

The above distributions allow to optimally estimate the state $X(t)$ for any t . However, as discussed in Section 2.1.2, distributions over the state alone do not fully characterize a stochastic process. For instance, in the context of stochastic chemical kinetics, we might be interested in computing timing statistics of certain reactions, which require a path-wise description of the underlying process. In the Bayesian context, this would mean that we search for posterior distributions over entire paths $\mathbf{X}_{1:N} = \{X(t) \in \mathbb{X} \mid t \in [t_1, t_N]\}$, i.e.,

$$\begin{aligned} p(\mathbf{x}_{1:N} \mid y_{1:N}) &= \frac{\prod_{n=1}^N p(y_n \mid x_n) p(\mathbf{x}_{1:N})}{p(y_{1:N})} \\ &= \frac{p(y_N \mid x_N) p(\mathbf{x}_{N-1:N} \mid x_{N-1}) p(\mathbf{x}_{1:N-1} \mid y_{1:N-1})}{p(y_N \mid y_{1:N-1})}, \end{aligned} \quad (2.43)$$

with $p(\mathbf{x}_{1:N})$ as the path density and $x_n = x(t_n)$. This distribution can also be interpreted as a smoothing distribution over paths, which can be established recursively from its ancestor at time t_{N-1} .

2.2.3 Model Selection

In many biological systems, large parts of biochemical interactions and mechanisms are yet to be discovered. Accordingly, a problem at least as central as the inference of unknown model quantities is the discrimination between different model variants. Formally, when considering candidate models $\mathcal{M} = 1, \dots, K$, one could assign a posterior probability to all the candidates using

$$P(\mathcal{M} = i \mid y) = \frac{p(y \mid \mathcal{M} = i) P(\mathcal{M} = i)}{\sum_{i=1}^K p(y \mid \mathcal{M} = i) P(\mathcal{M} = i)} \quad \forall i = 1, \dots, K, \quad (2.44)$$

where $P(\mathcal{M} = i)$ is the model prior and $p(y \mid \mathcal{M} = i)$ is model evidence. Note that when assuming a fixed model, the latter precisely coincides with the denominator in

Bayes' formula (2.31). More specifically, it is obtained by averaging the measurement likelihood over the unknown model quantities $X_i \in \mathbb{X}^i$, i.e.,

$$\begin{aligned} p(y | \mathcal{M} = i) &= \mathbb{E} [p(y | x, \mathcal{M} = i) | \mathcal{M} = i] \\ &= \int_{\mathbb{X}^i} p(y | x, \mathcal{M} = i) p(x | \mathcal{M} = i) dx. \end{aligned} \quad (2.45)$$

In general, the expectation in (2.45) is analytically intractable and therefore, one needs to resort to the aforementioned approximation techniques. For instance, if the dimensionality of the latent space \mathbb{X}^i is moderately low, one could directly sample from the prior distribution $p(x | \mathcal{M} = i)$ and estimate the evidence as

$$p(y | \mathcal{M} = i) \approx \frac{1}{M} \sum_{m=1}^M p(y | x^{(m)}, \mathcal{M} = i), \quad (2.46)$$

with $x^{(1)}, \dots, x^{(M)}$ as i.i.d. samples from $p(x | \mathcal{M} = i)$. When two models are compared pairwise, it is common to compute their *Bayes factor*, which is defined by

$$\mathcal{K}_{i,k} = \frac{p(y | \mathcal{M} = i)}{p(y | \mathcal{M} = k)} \quad (2.47)$$

and often given in units of deciban (dB), e.g.,

$$\mathcal{K}_{i,k}^{dB} = \log_{10} \mathcal{K}_{i,k}. \quad (2.48)$$

2.3 Single-Cell Fluorescence Data

A major part of this thesis deals with calibrating stochastic biochemical models using experimental single-cell fluorescence data. In particular, we consider two classes of data, which – even when considering the same biochemical networks – require a completely different computational handling. In the following we briefly discuss both types of data from the viewpoint of inference but remark that a description of the respective experimental techniques is not in the scope of this thesis.

2.3.1 Population Snapshot Data

The first class is population snapshot data such as revealed by flow-cytometry [89, 136] or mRNA FISH [82, 88]. Characteristic for such data is that individual cells cannot be followed over multiple time points – either because the cells do not survive the acquisition (e.g., mRNA FISH) or because the cells enter the device in a random sequential order such that any associations between time points are necessarily lost (e.g., fluorescence-activated cell sorting, FACS). Mathematically, we assume hence that the M measurements at every time point t_l are statistically independent samples from

each other⁷, i.e., $Y_l^m \mid (\Theta = \theta) \sim p(y_l^m \mid \theta)$ with $m = 1, \dots, M$ and Θ as a complete parametrization of the model. Under those assumptions, it holds that

$$p(y_1^1, \dots, y_1^M, \dots, y_L^1, \dots, y_L^M \mid \theta) = \prod_{l=1}^L \prod_{m=1}^M p(y_l^m \mid \theta). \quad (2.49)$$

In the context of parameter inference, one tries to find a θ , under which the temporal distributions $p(y_l^m \mid \theta)$ or their respective statistics best explain the data (see Section 3).

2.3.2 Time-Lapse Data

The second type of single-cell measurements are so-called time-lapse or live-cell data, where individual cells can be tracked over time. Using fluorescence microscopy techniques, their dynamics are monitored in a frame-by-frame fashion such that suitable image segmentation- and tracking algorithms allow to reconstruct each cell's abundance trajectory. Consequently, such data provides additional information on how the abundance of a particular cell evolves over multiple time points. This will be reflected by a statistical dependency between individual data points of a trajectory, i.e., $Y_1^m, \dots, Y_L^m \mid (\Theta = \theta) \sim p(y_1^m, \dots, y_L^m \mid \theta)$. Therefore, independence is preserved only across cells such that we obtain

$$p(y_1^1, \dots, y_1^M, \dots, y_L^1, \dots, y_L^M \mid \theta) = \prod_{m=1}^M p(y_1^m, \dots, y_L^m \mid \theta). \quad (2.50)$$

Denoting by $Y_{1:L}^m = \{Y_l^m \mid l = 1, \dots, L\}$ the trajectory of cell m , we can compactly write (2.50) as

$$p(y_{1:L}^1, \dots, y_{1:L}^M \mid \theta) = \prod_{m=1}^M p(y_{1:L}^m \mid \theta). \quad (2.51)$$

We realize from (2.51) that time-lapse data provides more information than population snapshot data, since a parametrization θ not only needs to explain the entirety of measurements at the respective time points, but also their evolution along the entire measurement interval. Similarly, if we consider a stochastic process, it seems much more difficult to infer its time-scale (e.g., in terms of its autocorrelation) from a sequence probability distributions than from a set of sample paths. The latter provide a direct handle on the time-scale of the process and therefore, often appear favorable from the viewpoint of inference.

⁷In reality, if the same cells are measured multiple times (e.g., using a cell sorter), there will be correlated effects between time points. However, we assume those to be negligible due to the randomization of the cell indices $m = 1, \dots, M$.

3 Inference from Heterogeneous Snapshot Data

In this chapter, we develop a Bayesian inference scheme for analyzing population snapshot data such as revealed by flow cytometry [136] or mRNA FISH [82, 88]. In particular, we consider *time-lapse* variants thereof, meaning that multiple snapshots are recorded over a sequence of time points. Although such recordings provide a handle on the dynamics of the considered biochemical system, they cannot resolve any temporal correlation on a single-cell level (see Section 2.3). In contrast, they allow to display how the probability distribution over the fluorescence protein evolves over time. As a consequence, mathematical models of such data are typically based on the chemical master equation, since its transient solution corresponds to what is measured experimentally. Here we put focus on the case where a population of cells exhibits a non-negligible degree of extrinsic variability. As pointed out in Section 2.1.5, the associated mixed-effect models scale poorly with the number measured cells. This turns out to be specifically problematic when considering flow cytometry techniques, which are able to capture thousands of cells at once.

In Section 3.1 we introduce the mathematical problem formulation and discuss a straightforward attempt at inference using heterogeneous CMEs. Since that approach is of limited practical use, we subsequently introduce a *moment-based* scheme (see Section 3.2), whose computational complexity is *independent* of the population size and therefore, ideally suited for analyzing heterogeneous snapshot data.

3.1 Inference Based on the Chemical Master Equation

We consider an ensemble of heterogeneous CTMCs $X^m \mid (S, Z^m)$ for $m = 1, \dots, M$ with Z^m as a set of extrinsic factors. Throughout this section, we assume those factors to be constant over time and that $Z^m \mid (A = a) \sim p(z \mid a)$ with a as a set of extrinsic statistics. Consequently, a cell's molecular abundance can be described by a conditional probability

$$P(X^m(t) = x \mid S = s, Z^m = z) = P(x, t \mid s, z), \quad (3.1)$$

which can be understood as the solution of a conditional master equation of the form

$$\frac{d}{dt}P(x, t \mid s, z) = \sum_{i=1}^L h_i(x - \nu_i, s, z)P(x - \nu_i, t \mid s, z) - \sum_{i=1}^L h_i(x, s, z)P(x, t \mid s, z), \quad (3.2)$$

with an initial condition $P(x, 0 | s, z)$ that possibly depends on the s and z . Similarly, the hazard functions h_i of that conditional master equation are defined such that they may depend on any of the parameters in z and s . An exemplary scenario would be that s corresponds to the set of kinetic rate constants c , while z could be the ribosomal abundance that varies from cell to cell.

When considering experimental snapshot data, every acquired fluorescence value stems from an individual cell – each having different extrinsic factors Z^m . Therefore, such data cannot be appropriately described by a single CME model. In contrast, it represents a mixture of M CMEs with variable parameterization. Mathematically, this corresponds to a *marginalization*, i.e.,

$$\begin{aligned} P(x, t | s, a) &= \int_{\mathcal{Z}} P(x, t | s, z) p(z | a) dz \\ &= \mathbb{E}[P(x, t | s, Z) | a], \end{aligned} \quad (3.3)$$

which means that the conditional probabilities are “averaged” with respect to the extrinsic factors. The resulting marginal probabilities directly depend on the extrinsic statistics a . In theory, (3.3) is maximally scalable with respect to M , meaning that it is independent of the population size.

3.1.1 Statistical Modeling

Let us for the moment assume that for a given $S = s$ and $A = a$, we can evaluate the marginal distribution (3.3). We have now given a temporal sequence of snapshot measurements at times t_1, \dots, t_L at which the molecular state has been partially observed. We denote the state and corresponding measurement of the m -th cell at time t_l as $X_l^m = X^m(t_l)$ and $Y_l^m | (X_l^m = x, \Omega = \omega) \sim p(y | x, \omega)$ with Ω as a (possibly unknown) set of parameters characterizing the acquisition noise. The l -th snapshot consists of M_l single-cell measurements $Y_l = \{Y_l^m | m = 1, \dots, M_l\}$.

The goal is to infer posterior distributions over S , A and Ω . The likelihood function factorizes as

$$\begin{aligned} p(y_1, \dots, y_L | s, a, \omega) &= \prod_{l=1}^L \prod_{m=1}^{M_l} p_l(y_l^m | s, a, \omega) \\ &= \prod_{l=1}^L \prod_{m=1}^{M_l} \sum_{x \in \mathbb{X}} p(y_l^m | x, \omega) P(x, t_l | s, a) \\ &= \prod_{l=1}^L \prod_{m=1}^{M_l} \mathbb{E}[p(y_l^m | X(t_l), \omega) | s, a], \end{aligned} \quad (3.4)$$

and hence,

$$p(s, a, \omega | y_1, \dots, y_L) \propto \prod_{l=1}^L \prod_{m=1}^{M_l} \mathbb{E}[p(y_l^m | X(t_l), \omega) | s, a] p(s) p(a) p(\omega). \quad (3.5)$$

Eq. (3.5) allows us to evaluate the posterior distribution up to a normalizing constant such that in principle, we could sample from it using a suitable MCMC scheme. Nevertheless, the expected value in (3.5) is taken with respect to the marginal probability distribution which we have assumed to be known so far. In reality – however – it is barely tractable and thus, would require numerical approximations. For instance, the marginalization could be performed by Monte Carlo averaging multiple solutions of the conditional CME with randomly drawn Z . Alternatively, one could extend the state space by a discretized variant of Z with $p(z | a)$ as initial condition [50] and numerically solve the extended CME (e.g., using FSP [81]). The solution of this equation would then yield the joint distribution over $X(t)$ and Z , which can be marginalized subsequently by evaluating a finite sum. In the context of parameter estimation, both approaches are practically infeasible for all but the smallest systems, since most inference schemes rely on iterative algorithms that require many numerical integrations of the CME. In the next section, we introduce an alternative and significantly more efficient approach.

3.2 Inference Based on Moments

We have seen in Section 2.1.4 that the dynamics of a biochemical network can be described by a set of coupled moment equations, which are straightforward to derive from the CME. This section is centered around the question if and how one can leverage those moment equations for the purpose of inference. Intuitively speaking, we try to find a set of parameters such that the moments of the biochemical network agree well with the moments obtained experimentally.

3.2.1 Marginal Moment Dynamics

We again consider the heterogeneous biochemical network from Section 3.1. We remark that the moments of such a model are different from those derived in Section 2.1.4. More specifically, we require the *marginal* moments of $X(t)$, i.e.,

$$\begin{aligned} \mathbb{E}[f(X(t))] &= \sum_{x \in \mathbb{X}} f(x) P(x, t | s, a) \\ &= \sum_{x \in \mathbb{X}} f(x) \int_{\mathcal{Z}} P(x, t | s, z) p(z | a) dz, \end{aligned} \tag{3.6}$$

with $f : \mathbb{X} \times \mathcal{Z} \mapsto \mathbb{R}$ as a polynomial in x and z . Taking the time derivative of (3.6), we further obtain

$$\begin{aligned}
\frac{d}{dt} \mathbb{E} [f(X(t), Z)] &= \sum_{x \in \mathbb{X}} f(x, z) \int_{\mathcal{Z}} \frac{d}{dt} P(x, t | s, z) p(z | a) dz \\
&= \sum_{x \in \mathbb{X}} f(x, z) \int_{\mathcal{Z}} \left[\sum_{i=1}^L h_i(x - \nu_i, s, z) P(x - \nu_i, t | s, z) \right. \\
&\quad \left. - \sum_{i=1}^L h_i(x, s, z) P(x, t | s, z) \right] p(z | a) dz \\
&= \sum_{i=1}^L \int_{\mathcal{Z}} \sum_{x \in \mathbb{X}} f(x + \nu_i, z) h_i(x, s, z) P(x, t | s, z) p(z | a) dz \\
&\quad - \sum_{i=1}^L \int_{\mathcal{Z}} \sum_{x \in \mathbb{X}} f(x, z) h_i(x, s, z) P(x, t | s, z) p(z | a) dz \\
&= \sum_{i=1}^L \mathbb{E} [\mathbb{E} [f(X(t) + \nu_i, Z) h_i(X(t), s, Z) | Z]] \\
&\quad - \sum_{i=1}^L \mathbb{E} [\mathbb{E} [f(X(t), Z) h_i(X(t), s, Z) | Z]] \\
&= \sum_{i=1}^L \mathbb{E} [f(X(t) + \nu_i, Z) h_i(X(t), s, Z)] - \sum_{i=1}^L \mathbb{E} [f(X(t), Z) h_i(X(t), s, Z)].
\end{aligned} \tag{3.7}$$

Depending on the particular form of h_i and f , the r.h.s. of (3.7) may also contain higher-order moments (and cross-moments) of the extrinsic factors Z and the state $X(t)$. We therefore require a suitable moment-closure [52, 128] scheme in order to get a finite-dimensional system of ODEs. For that sake, we introduce the vector of non-central moments (and cross-moments) up to a certain order n as $\tilde{\mu}(t)$ as well as the vector of all moments of order higher than n as $\bar{\mu}(t)$. The moment system can then be written as

$$\frac{d}{dt} \tilde{\mu}(t) = A(s, a) \tilde{\mu}(t) + B(s, a) \bar{\mu}(t). \tag{3.8}$$

The idea of moment-closure is to replace the dependency of the moments of order higher than n by a dependency of the moments of order up to n . In other words, we try to find a closure function q such that $\hat{q}(\tilde{\mu}(t)) \approx \bar{\mu}(t)$. This is illustrated in the following example using a non-linear death-process.

Example 4 (Gaussian closure). *Let us consider a homogeneous one-dimensional death process $X(t)$ with death-rate $X^2(t)$ and some initial condition $X(0) = x_0$. The non-*

central moments of up to order two are given by

$$\begin{aligned}\frac{d}{dt}\mathbb{E}[X(t)] &= -\mathbb{E}[X^2(t)] \\ \frac{d}{dt}\mathbb{E}[X^2(t)] &= \mathbb{E}[X^2(t)] - 2\mathbb{E}[X^3(t)].\end{aligned}\tag{3.9}$$

In this case, the second-order moment depends on the third-order moment and so forth. Hence, we try to approximate $\mathbb{E}[X^3(t)]$ as a function of the first- and second-order moment, i.e., $\mathbb{E}[X^3(t)] \approx \hat{q}(\mathbb{E}[X(t)], \mathbb{E}[X^2(t)])$. For instance, if we assume that $P(x, t)$ is approximately Gaussian, we obtain for the third-order moment $\mathbb{E}[X^3(t)] = -2\mathbb{E}[X(t)]^3 + 3\mathbb{E}[X(t)]\mathbb{E}[X^2(t)]$ and furthermore,

$$\begin{aligned}\frac{d}{dt}\mathbb{E}[X(t)] &= -\mathbb{E}[X^2(t)] \\ \frac{d}{dt}\mathbb{E}[X^2(t)] &= 4\mathbb{E}[X(t)]^3 + \mathbb{E}[X^2(t)] - 6\mathbb{E}[X(t)]\mathbb{E}[X^2(t)].\end{aligned}\tag{3.10}$$

With eq. (3.10) we have found an approximative closed moment system of the nonlinear death-process. However, additional simulations need to be performed to judge whether the closure can accurately describe the first- and second-order moments.

Several different closure techniques have been proposed recently [52, 114, 115]. They all have in common that they rely on certain assumptions on the underlying distribution. The authors in [114] propose a closure scheme termed *derivative matching* corresponding to a log-normal assumption. When using the *zero-cumulants* closure, the $(n+1)$ -th order cumulants of the abundance distribution are set to zero, yielding a closed moment system of order n [52]. In this section we focus on the aforementioned closure-types but remark that further closures have been proposed in the literature (see e.g., [115]).

3.2.2 Statistical Modeling

Although an extension to the general case is straightforward, we assume - for the sake of clarity - that only a single species is measured from a cell population at time points $t_l, l \in \{1, \dots, L\}$. We further assume that the technical noise is negligibly small or that it has been deconvolved from the measurements beforehand (see e.g., Section 3.2.4). We define $\gamma = \{s, a\}$ and denote the approximate time evolution of the measured species' k -th order moment as $t \mapsto \tilde{\mu}_k(t, \gamma)$ for $k \in \{1, \dots, n\}$. With the k -th order experimental moments $\hat{\mu}_k = \{\hat{\mu}_k(t_l) \mid l \in \{1, \dots, L\}\}$, the posterior distribution over γ is generally given by

$$p(\gamma \mid \hat{\mu}_1, \dots, \hat{\mu}_n) \propto \prod_{l=1}^L p(\hat{\mu}_1(t_l), \dots, \hat{\mu}_n(t_l) \mid \gamma) p(\gamma),\tag{3.11}$$

with $p(\gamma)$ as the prior distribution over γ . For simplicity, we further assume that the experimental moments of different order are statistically independent of each other,

corresponding to the assumption that they have been estimated using distinct and independent sets of samples (e.g., using bootstrapping). In this case, (3.11) simplifies to

$$p(\gamma \mid \hat{\mu}_1, \dots, \hat{\mu}_n) \propto \prod_{k=1}^n \prod_{l=1}^L p(\hat{\mu}_k(t_l) \mid \gamma) p(\gamma). \quad (3.12)$$

Furthermore, the large-sample case encountered in population snapshot measurements, allows us to make use of the central limit theorem and assume that $p(\hat{\mu}_k(t_l) \mid \gamma) = \mathcal{N}(\tilde{\mu}_k(t_l, \gamma), \sigma_k^2(t_l))$, where $\sigma_k^2(t_l)$ refers to the estimated uncertainty of the respective moment. For instance, if we consider the first two central moments (i.e., mean and variance), asymptotically unbiased estimators are given by

$$\hat{\mu}(t_l) = \frac{1}{N} \sum_{i=1}^N x_i(t_l)$$

and

$$\hat{\mu}_2(t_l) = \frac{1}{N} \sum_{i=1}^N (x_i(t_l) - \hat{\mu}_1(t_l))^2.$$

and the corresponding variances can be estimated as

$$\sigma_1^2(t_l) = \frac{1}{N} \hat{\mu}_2(t_l) \quad \text{and}$$

$$\sigma_2^2(t_l) = \frac{1}{N} \left(\hat{\mu}_4(t_l) - \frac{N-3}{N-1} \hat{\mu}_2(t_l)^2 \right),$$

respectively [105, 136]. Under those assumptions, the posterior from eq. (3.12) can be evaluated using available techniques. Here, we employ a standard Metropolis-Hastings (M-H) sampler [15] to draw samples from the posterior distributions. For all the subsequent experiments, we assume flat prior distributions over parameters $\gamma_j \in \gamma$ (with zero probability for negative values). In the M-H scheme, for each of the J_P parameters in γ , we use independent log-normal proposal distributions such that $q(\gamma^{new} \mid \gamma^{old}) = \prod_{j=1}^{J_P} q(\gamma_j^{new} \mid \gamma_j^{old})$ with $q(\gamma_j^{new} \mid \gamma_j^{old}) = \mathcal{LN}(\ln \gamma_j^{old}, v_j^2)$. Proposed parameter samples are then accepted with probability

$$a = \min \left\{ 1, \frac{p(\gamma^{new} \mid \hat{\mu}_1, \dots, \hat{\mu}_n) q(\gamma^{old} \mid \gamma^{new})}{p(\gamma^{old} \mid \hat{\mu}_1, \dots, \hat{\mu}_n) q(\gamma^{new} \mid \gamma^{old})} \right\}.$$

From the resulting samples, it is straightforward to extract point estimates, for instance by selecting the parameter configuration which maximizes the posterior distribution (i.e., the maximum a posteriori (MAP) estimate). Whether the moments of the measured distributions carry enough information to jointly determine the intrinsic parameters and the extrinsic statistics, in general has to be answered by performing an identifiability analysis of the moment system [12, 51, 105].

Table 3.1: Reference and inferred model parameters.

Parameter	c_1	c_2	c_3	c_4
Reference	$1.500 \cdot 10^{-2}$	$8.000 \cdot 10^{-4}$	$1.000 \cdot 10^{-3}$	$4.000 \cdot 10^{-1}$
MAP	$1.380 \cdot 10^{-2}$	$7.050 \cdot 10^{-4}$	$9.865 \cdot 10^{-4}$	$3.988 \cdot 10^{-1}$
Unit	s^{-1}	s^{-1}	s^{-1}	s^{-1}

3.2.3 Application to Simulated Data

To test whether the moment-based inference scheme can identify parameters even in the case of multi-modal process distributions, we studied the four-species model depicted in Figure 3.1a, which can be thought of as a simple model of transiently induced gene expression. Degradation of A serves as a simplistic mechanism to model a temporal window of transcription factor activity. During this temporal window, the gene B manages to switch into a state, where protein C is produced only in a fraction of the cells. The corresponding moment equations of order one and two can be found in [136]. For simplicity, we assume that the initial conditions of the species are known. If in some application the initial conditions are unknown, they can be included as unknown parameters in the parameter search and estimated along with the other parameters. For species A we set the initial amount to 50 molecules, B was initialized at 1 (corresponding to the gene being initially in the inactive state), whereas all other species were initialized at 0 molecules. The reference data was generated by stochastic simulation [44] using $M = 20000$ sample paths of length $T = 10000s$. We computed means and variances of species C and their corresponding uncertainties each $2000s$, treated them as experimental measurements and computed maximum a posteriori (MAP) estimates as described in Section 3.2.2. The scaling parameters of the log-normal proposal densities were set to $v_j = 0.01$. After a burn-in period, we recorded around 10000 samples of the M-H algorithm from which the MAP values were extracted. This inference was performed multiple times using random initial parameter values, drawn from a log-normal distribution $\mathcal{LN}(\ln 0.002, 2^2)$. Each time, the inference scheme ended up with equivalent MAP estimates (up to small random deviations, introduced by the randomized parameter search).

The inferred and reference parameters are given in Table 3.1 and the corresponding moments are depicted in Figure 3.1c. Even though in this case, mean and variance do not paint a full picture of the underlying multi-modal protein distribution, all parameters - and thus the protein distributions - were estimated accurately up to a small approximation error (see Figure 3.1b).

3.2.3.1 Second-Order Moment Resolves Non-Identifiability

For the simple model of transient gene activation, the population mean alone does not provide enough information to uniquely determine the four model parameters. This is demonstrated in Figure 3.2, where we compared two different model configurations,

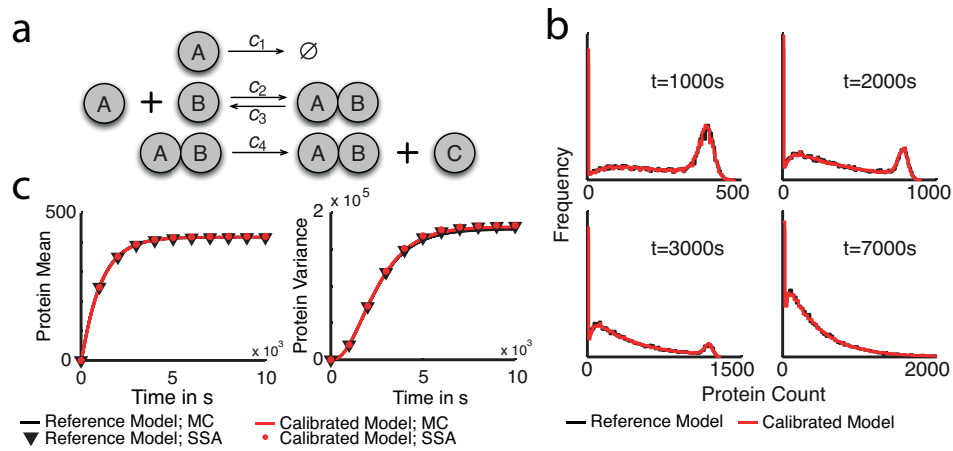


Figure 3.1: Moment-based inference using synthetic data. (a) A simple model of transient gene activation: The binding of A to the target gene B aggregates all necessary steps involved in gene activation such as the binding of additional transcription factors, polymerase binding or chromatin remodeling. Also protein synthesis is reduced to the simplest possible model, i.e., a first order production, abstracting messenger RNA (mRNA) transcription and degradation, translation and protein folding. (b) The protein distributions predicted by the calibrated model (red) compared to the distributions generated from the reference model (black) at four representative time points. Estimates of the distributions were obtained by stochastic simulation (20000 runs). (c) The time evolutions of the approximate protein mean and variance obtained from moment closure (MC) differ only little from approximations computed by stochastic simulation (SSA). Therefore, the model parameters can be inferred up to negligibly small deviations.

each of them found by running the same MCMC algorithm with different initial conditions. In both cases, the estimated means fit well the reference mean. In contrast, the variances significantly differ from each other. Neither parameter set can reproduce the underlying distribution. For one of them, the distribution is even unimodal (see Figure 3.3).

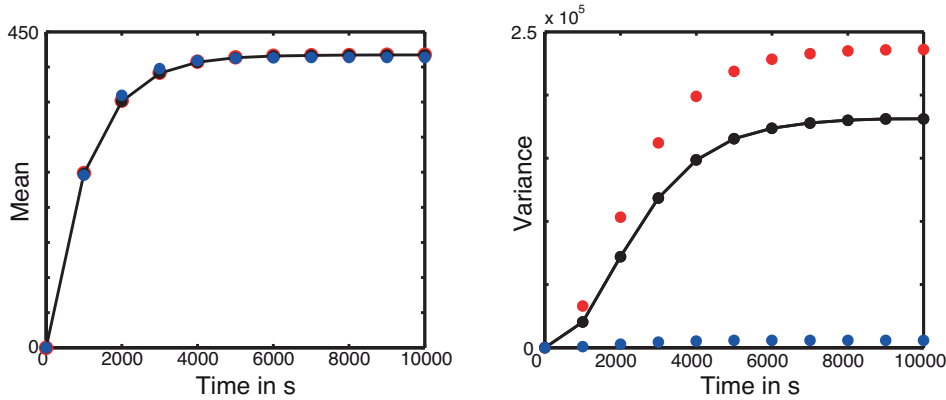


Figure 3.2: The plot shows results for two parameter configurations that achieved almost equivalent mean values but strongly differ in the variance. Red, Blue: Calibrated models; Black: Reference model

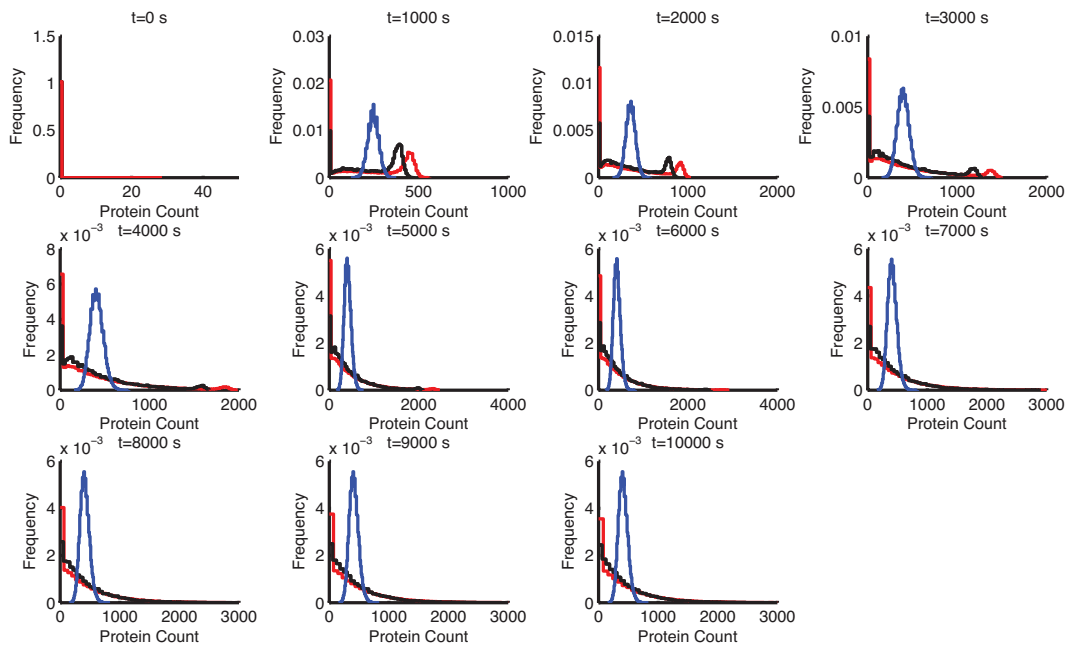


Figure 3.3: Protein distributions for two parameter configurations. As the blue distribution is unimodal for all time points, it follows that the multi-modality cannot be predicted from the protein mean alone. Red, Blue: Calibrated models; Black: Reference model.

3.2.4 Application to Hog1-Induced Gene Expression in Yeast

The moment-based inference scheme allowed us to study gene expression, activated by the HOG signaling pathway in budding yeast [54]. Upon hyper-osmotic shock yeast cells induce the MAPK Hog1 signaling cascade. The role of this kinase is two-fold. In the cytoplasm, Hog1 phosphorylates its substrate to increase the internal concentration of glycerol in the cell. In parallel, a large fraction of the active Hog1 translocates to the nucleus where it triggers the activation of a transcriptional program leading to the upregulation of roughly 300 genes [38]. Once the internal glycerol concentration allows to balance the external osmotic pressure, the HOG pathway is deactivated leading to loss of active MAPK and a rapid termination of the transcriptional process. To quantify the amount of transcription induced by this pathway a fluorescent expression reporter was generated using the promoter pSTL1 (promoter of the sugar transporter-like protein 1), a well characterized Hog1 expression target driving the expression of a fluorescent protein construct (quadrupleVenus - qV). It was shown in [94] that the transient activation of the MAPK Hog1 in conjunction with a slow step in the transcription activation process of the promoter results in a bimodality in the expression profiles of this fluorescent expression reporter. The pSTL1-qV reporter abundance was quantified by flow cytometry at 9 different time points for NaCl concentrations of 0M, 0.1M, 0.12M and 0.2M. To determine the nuclear enrichment of Hog1, we used the fluorescence microscopy data published in [94].

3.2.4.1 Kinetic Modeling

Components involved in activation and translocation of Hog1 are present in high abundance (e.g., around 6800 Hog1 molecules per cell [40]). Consequently, intrinsic fluctuations of active Hog1 are relatively small. Experimental results from [83, 94] support this and also that Hog1 signaling is robust against cell-to-cell variations. Motivated by this, we assume Hog1 signaling to be deterministic rather than stochastic and that the mean dynamics reflect well the signaling behavior [72, 140]. Continuous-time functions of nuclear Hog1 were obtained from the experimental data by linear regression with radial basis functions [15] across different NaCl concentrations (see Figure 3.4).

Several transcription factors such as Sko1 or Hot1 are under control of active Hog1 once it translocates to the nucleus as shown in [21]. This and the experimentally observed switch-like induction of fluorescent reporter expression, suggest a high cooperativity in the pSTL1 promoter dynamics. In a purely stochastic mass-action model, one way to model cooperativity is to require multiple Hog1 copies to bind to the promoter before messenger RNAs (mRNA) can be transcribed. However, the previous high copy-number considerations allow us to simplify this step into transforming the fitted Hog1 abundance curves using a Hill-function with tunable parameters, i.e.,

$$\tilde{u}(t) = \frac{V_{max}(u(t) + u_0)^{n_H}}{K_d^{n_H} + (u(t) + u_0)^{n_H}}, \quad (3.13)$$

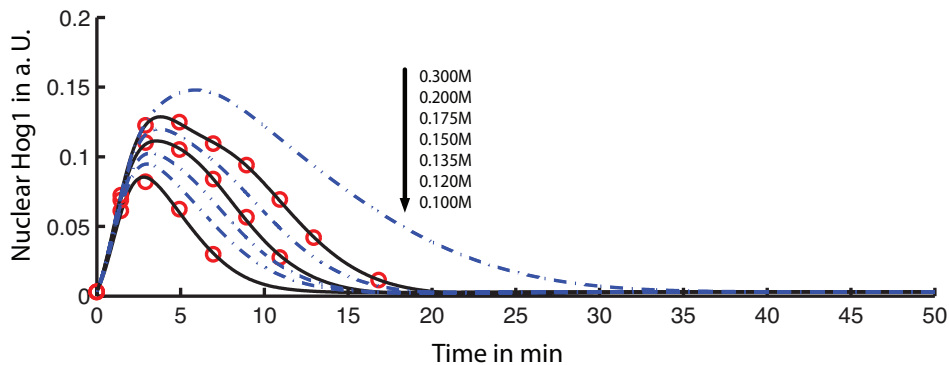


Figure 3.4: Extracted nuclear Hog1 enrichment over time. Black: Fitted using linear RBF regression; blue: interpolated; red: mean abundance obtained from microscopy.

with $u(t)$ as the fitted nuclear Hog1 abundance and u_0 , n_H , V_{max} and K_d as unknown parameters, estimated during model calibration. The output of this function is then treated as a time-varying kinetic parameter modulating the gene activation rate, i.e., $c_1(t) = \tilde{u}(t)$. Efficient transcription of mRNA requires interaction of the active gene with chromatin remodeling complexes (generic remodeler denoted as CR) [94]. Translation is modeled as a one-step linear production event, depending on the number of ribosomes.

We model extrinsic variability at two different stages of the proposed model. First, we assume variability in chromatin remodeling, because it depends on a variety of different complexes which might be subject to cell-to-cell variations (such as RSC or the SAGA complex, [55, 89]). Chromatin remodeling is modeled by recruitment of species C by the active gene B . Thus, variability in the total number of CR ($Z_1 = [C] + [D]$) leads to variability in the remodeling efficiency. Furthermore, we assume heterogeneity in the translation efficiency [23], which is reflected by a variability in the proxy species $Z_2 = [E]$. Mean and covariance matrix of the random vector $Z = [Z_1, Z_2]^T$ are defined as

$$\mathbb{E}[Z] = \begin{pmatrix} \alpha_1^1 \\ \alpha_2^1 \end{pmatrix} \quad (3.14)$$

and

$$\mathbb{E}[(Z - \mathbb{E}[Z])(Z - \mathbb{E}[Z])^T] = \begin{pmatrix} \alpha_{11}^2 & \alpha_{12}^2 \\ \alpha_{12}^2 & \alpha_{22}^2 \end{pmatrix} \quad (3.15)$$

The extrinsic statistics then enter the moment equations in terms of the initial conditions for the respective first and second order moments. The extrinsic statistics were assumed to be unknown and inferred from the measurements as explained in Section 3.2.2. Note that within a moment-based approach, no assumptions on the distribution $p(z)$ are required. However, once a comparison between the protein distributions is

desired, realizations z have to be drawn from $p(z)$ for each SSA run. As $p(z)$ is not fully characterized by mean and variance only, further assumptions need to be made. Here, we restrict the shape of $p(z)$ to be log-normal [36] and compute its parameters from the inferred extrinsic statistics (i.e., first and second order moments). A schematic illustration of the model is given in Figure 3.5a. The corresponding moment equations of order one and two can be found in [136].

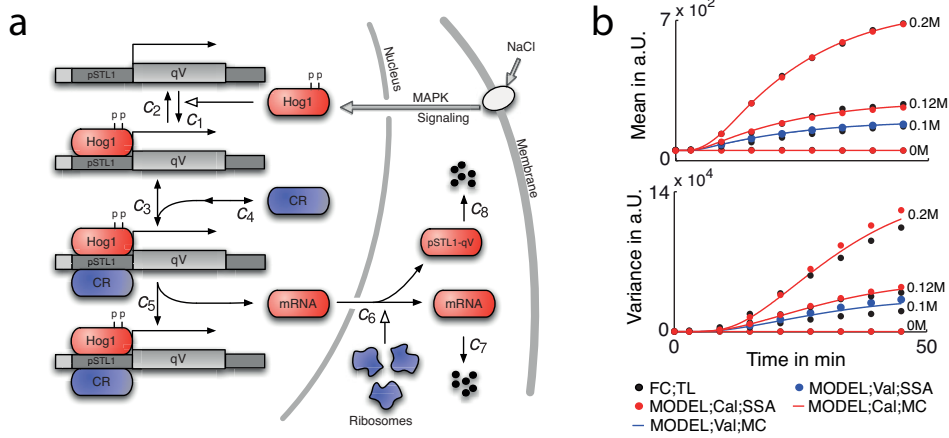


Figure 3.5: MAPK Hog1 induced pSTL1-qV expression. (a) Osmotic pressure is sensed at the membrane, and results in the activation of the MAPK signaling cascade. Once active, double-phosphorylated MAPK Hog1 translocates to the nucleus, where it can bind via transcription factors to the pSTL1 promoter. Remodeling of the chromatin structure then allows for efficient transcription of mRNA, which is exported from the nucleus to yield expression of the fluorescent reporter pSTL1-qV. Blue-shaded entities denote species subject to extrinsic variability. (b) Comparison of pSTL1-qV mean and variance obtained after calibration (Cal) and validation (Val) of the model using moment closure (MC) and 20000 stochastic simulation runs (SSA) with the experimental estimates obtained from the time-lapsed (TL) flow cytometry (FC) data (~ 20000 cells).

3.2.4.2 Modeling Fluorescence Intensities

We further assumed that the measured fluorescence intensity for a given cell is proportional to the number of fluorescent proteins [89], i.e., $I_{Tot}(t_l) = \epsilon \hat{I}_{Tot}(t_l)$ with scaling parameter ϵ . Furthermore, we assumed that the reporter abundance $I_R(t_l)$ is corrupted by autofluorescence and measurement artifacts, modeled as an additive random variable $I_{AF}(t_l)$, independent of the reporter abundance, i.e., $I_{Tot}(t_l) = I_R(t_l) + I_{AF}(t_l)$. Mean and variance of $I_{AF}(t_l)$ were estimated from the flow cytometry data for 0M NaCl, collected over the measurement time points. As this allows very accurate estimates (N in the order of hundreds of thousands), the uncertainty of those estimates can be well neglected. The experimental means and variances of pSTL1-qV abundance at a given measurement time point were calculated as $\hat{\mu}_R^k(t_l) = \hat{\mu}_{Tot}^k(t_l) - \mu_{AF}^k(t_l)$ for $k \in \{1, 2\}$. Note that moment-based inference and analysis of the model can be carried out without any assumptions on the autofluorescence distribution. In order to compare protein

distributions from the model with experimentally obtained distributions, we sampled autofluorescence values from the measured flow cytometry distribution for 0M NaCl.

3.2.4.3 Model Calibration and Validation

Alltogether, the model comprises the parameters u_0 , n_H , K_d , V_{max} , c_2 , c_3 , c_4 , c_5 , c_6 , c_7 , c_8 , α_1^1 , α_2^1 , α_{11}^2 , α_{22}^2 and α_{12}^2 , which were estimated from the time courses of the experimental means and variances (see Figure 3.5b) using NaCl concentrations 0M, 0.12M and 0.2M.

Note that the propensity of a translation event is proportional to the product $c_6 \cdot [E]$. Thus, the parameters c_6 , α_2^1 and α_{22}^2 are structurally unidentifiable [12, 51]; hence we estimated the products $c_6\alpha_2^1$ and $c_6^2\alpha_{22}^2$. Given those products, statistics of the number of ribosomes could for instance be estimated by setting c_6 to values from literature and quantifying the remaining part.

Due to the high-dimensional state and parameter space and the fact that our flow cytometry experiments only captured distributions of a single protein, we expected the resulting posterior distributions to be multi-modal. Therefore, a standard M-H scheme was performed 50 times, each time with randomly drawn initial parameter configurations. Note that in general, this is likely to give parameter configurations for which the nonlinear moment system is numerically unstable and thus, we first selected a stable parameter configuration γ_s (see Table 3.2). Then, each parameter value was randomly initialized around those initial parameter values, i.e., $\gamma_j^0 \sim \mathcal{LN}(\ln \gamma_{j,s}, 0.5^2)$. The scaling parameters v_j of the log-normal proposal densities are given in Table 3.2. We then sorted the 50 parameter sets according to their maximum a-posteriori probabilities and selected the best five parameter sets for further inspection.

To quantitatively assess the goodness-of-fit we additionally computed a distance measure between the predicted and experimental distributions. Among the various metrics proposed in literature (see [41] for an overview) we chose the uniform or Kolmogorov metric (measuring the maximum deviation between the cumulative distributions). As it is scale-invariant and bounded by one it appears more amenable to interpretation than other distance measures. Moreover, it was already used in the context of stochastic simulations of chemical kinetics [20]. For each of the 50 parameter sets we computed the Kolmogorov distance for all NaCl concentrations and time points between the empirical and the predicted pSTL1-qV distributions. The final parameter set was selected such as to minimize the total distance, i.e., the sum of individual Kolmogorov distances over all time points and the three concentrations used for fitting (i.e., 0M, 0.12M and 0.2M of NaCl). Figure 3.6 depicts the Kolmogorov distance for the best performing parameter set which is given in Table 3.2.

Using the final parameter set, the pSTL1-qV expression profiles for each measurement time point and NaCl concentration were computed from the calibrated model using stochastic simulation. A comparison between the experimental and the predicted distributions is shown in Figure 3.8a. Even though only means and variances were used

Table 3.2: Inferred model parameters and Metropolis-Hastings setup. The $\gamma_{j,s}$ denote the initial parameter values, chosen such as to obtain stable moment dynamics. The inferred MAP estimates are denoted as $\gamma_{j,MAP}$. The scaling parameters of the log-normal proposal distributions used in the Metropolis-Hastings algorithm are denoted as v_j .

Parameter	$\gamma_{j,s}$	$\gamma_{j,MAP}$	Unit	v_j
u_0	$3.000 \cdot 10^{-2}$	$1.581 \cdot 10^{-2}$	$a.U.$	0.01
n_H	3.000	6.130	1	0.01
K_d	$2 \cdot 10^{-1}$	$1.418 \cdot 10^{-1}$	$a.U.$	0.01
V_{max}	1.000	1.025	s^{-1}	0.01
c_2	1.000	1.384	s^{-1}	0.02
c_3	$4.000 \cdot 10^{-4}$	$6.669 \cdot 10^{-4}$	s^{-1}	0.02
c_4	$1.000 \cdot 10^{-3}$	$1.469 \cdot 10^{-2}$	s^{-1}	0.02
c_5	1.000	$2.825 \cdot 10^{-1}$	s^{-1}	0.02
c_7	$1.000 \cdot 10^{-3}$	$5.476 \cdot 10^{-4}$	s^{-1}	0.02
c_8	$1.000 \cdot 10^{-4}$	$1.283 \cdot 10^{-4}$	s^{-1}	0.02
α_1^1	$3.300 \cdot 10^1$	$2.250 \cdot 10^2$	1	0.01
$c_6 \alpha_2^1$	$3.300 \cdot 10^{-2}$	$5.663 \cdot 10^{-3}$	s^{-1}	0.01
α_1^2	$1.900 \cdot 10^3$	$7.809 \cdot 10^3$	1	0.02
$c_6 \alpha_2^2$	$1.900 \cdot 10^{-3}$	$3.098 \cdot 10^{-6}$	s^{-2}	0.02
$c_6 \alpha_{12}^2$	$1.100 \cdot 10^{-1}$	$8.935 \cdot 10^{-2}$	s^{-1}	0.01

in the inference, the bimodal distributions are accurately predicted by the model.

We further validated the model using an additional snapshot data set from [94], where the pSTL1-qV reporter abundance was measured for several other NaCl concentrations between 0M and 0.3M, 45 minutes upon osmotic shock. From the model predictions and the measured distributions, we computed the coefficient of variation (CV) and a dose-response as functions of the NaCl concentration (Figure 3.8b). The area around 0.1M NaCl, where the CV is large and the dose-response curve is rising, indicate the NaCl concentration interval where the expression is in a bimodal regime. Note that also at 0.3M NaCl, a concentration much larger than the concentrations that were used in the inference, the CV is predicted accurately.

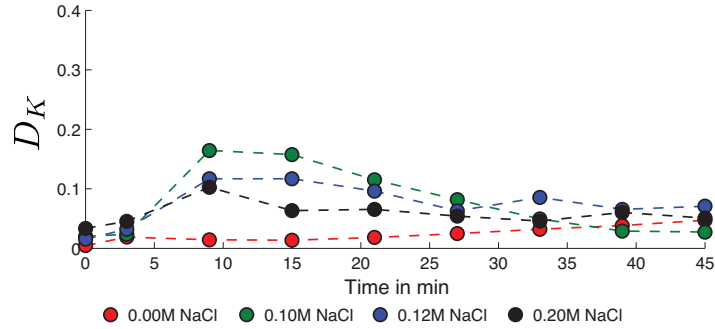


Figure 3.6: Distance in distribution between model predictions and data. The Kolmogorov distance D_K was computed between the empirical and the predicted pSTL1-qV distribution for each time point and concentration. Note that the distance D_K is bounded by one.

To study the stochastic pSTL1-qV induction, we simulated the model to estimate the average number of cells that (a) never activate the pSTL1 promoter, (b) activate the promoter at least once and (c) induce transcription. Our model predicts that for all NaCl concentrations except 0M all cells manage to activate the promoter and therefore, that the bimodality has to be caused by the subsequent - and comparably slow - chromatin remodeling step (see Figure 3.7).

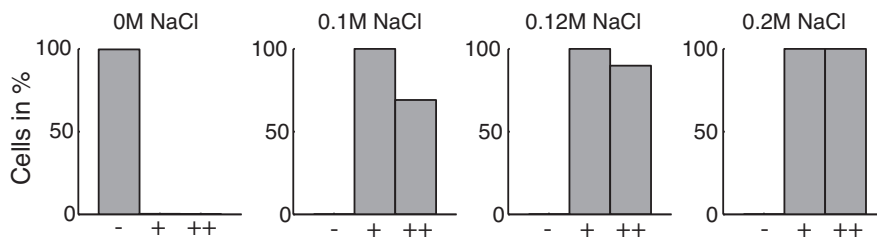


Figure 3.7: Statistical analysis of the transcriptional activation in pSTL1-qV expression. Bars indicate the percentage of cells that never activated the gene (-), that activated the gene at least once (+) and cells that initiated transcription (++). Statistics were computed from 1000 traces obtained using stochastic simulation with the inferred parameters.

Further, we performed an *in silico* knock-down of CR, by rescaling each cell’s amount of CR by a hand-tuned factor, such that the percentage of responding cells saturated around 60% as measured in the experiment (see Figure 3.8b). We found that the transition between the non- and all-responding domain is shifted to higher NaCl values and that the slope of the transition edge is decreased.

3.2.4.4 In Silico Homogenization of the Cell Population

After calibrating the model, we switched off extrinsic variability, by setting each cell’s extrinsic condition to the inferred mean value. We then recomputed estimates of the pSTL1-qV distributions using stochastic simulation. The resulting average cell can be interpreted as a homogenized version of the measured population. Again CV and dose-response were computed and plotted in Figure 3.8b. Interestingly, we find that extrinsic variability does not affect the dose-response behavior in pSTL1-qV induction. In contrast, the homogenized population shows significant differences in the CV. In particular, for larger stress levels the CV is relatively small compared to the heterogeneous counterpart, indicating less variability in pSTL1-qV reporter expression. For intermediate stress levels the homogenized population still shows a bimodal response.

To study the extent to which extrinsic variability can be reduced by cell-gating, we re-estimated the extrinsic statistics using the time-lapsed flow cytometry data set for gates of different size, applied on the forward scatter channel (FSC, often used as a proxy for cell volume). We found that the variability in the translation efficiency is significantly reduced for small gating diameters. In contrast, no significant trend was found in the estimated variability in CR (see Figure 3.9).

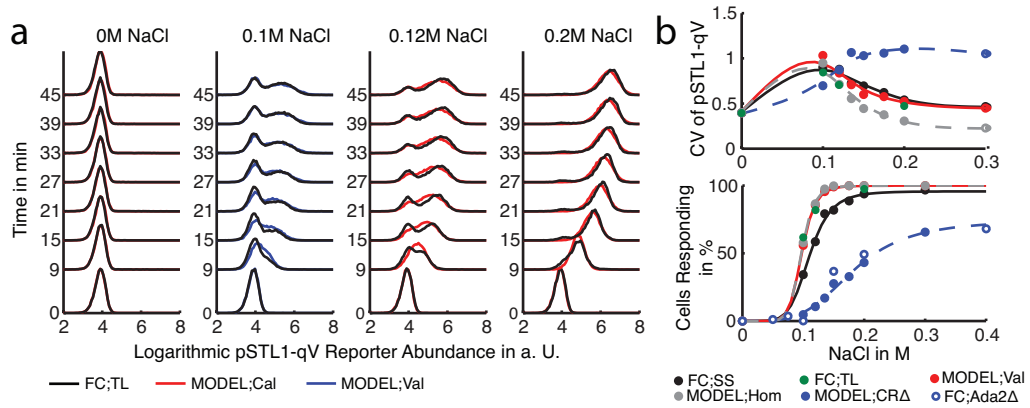


Figure 3.8: Population mean and variance predict bimodal pSTL1-qV response. (a) Model calibration (Cal) and validation (Val). The parameters were inferred using time-lapsed (TL) flow cytometry (FC) data of the pSTL1-qV reporter at three NaCl concentrations (i.e., 0M, 0.12M and 0.2M) and used to predict the expression profiles at a NaCl concentration of 0.1M. (b) Model validation using flow cytometry snapshot data (SS), recorded 45 minutes after osmotic shock for 0M, 0.1M, 0.12M, 0.135M, 0.15M, 0.175M, 0.2M and 0.3M of NaCl. Left: Coefficient of variation (CV) of pSTL1-qV intensity as a function of NaCl. Right: Dose-response comparison. All curves indicate a Hill-type relation. The calibrated model was homogenized (Hom), giving rise to the average cell's CV and dose-response curves. Additionally, we studied the suppression of the chromatin remodeling *in silico* by reducing the amount of CR, such that the percentage of responding cells saturated around 60% (CR Δ) and compared the model predictions to the results reported in [94], where the authors performed a knock-down of the transcription adapter 2 (Ada2) to demonstrate the impact of chromatin remodeling in pSTL1-qV induction (Ada2 Δ).

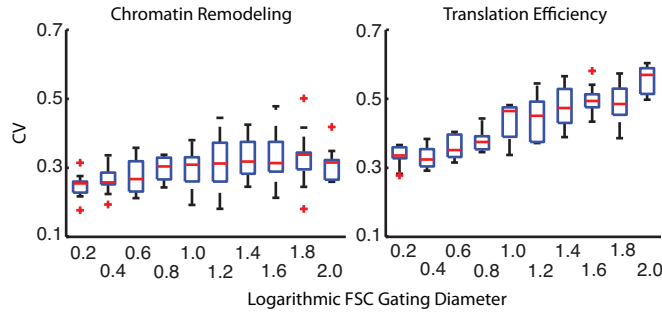


Figure 3.9: Cell-gating eliminates only a fraction of extrinsic variability. To study the influence of cell-gating on cell-to-cell variability we fixed all parameters but the extrinsic statistics \tilde{a} to their previously inferred values. The extrinsic statistics were re-estimated from the time-lapsed flow cytometry data with increasing FSC gating diameters using MCMC sampling with 100 randomly drawn initial conditions. Statistics were computed and visualized over the ten runs that achieved the highest posterior values using boxplots with MATLAB's (2009b, The Mathworks, Natick, MA) default settings. The CV of the translation efficiency increases with the logarithmic gating diameter (right), whereas the CV of the chromatin remodeling remains more or less constant (left).

4 Inference from Heterogeneous Time-Lapse Data

In contrast to population snapshot measurements, fluorescence microscopy techniques allow to record movies of cell populations from which single-cell time-lapse data can be extracted (see Section 2.3). Such data do not only provide a handle on the protein distribution at particular time points, but also, on how the abundance evolves over the measurement interval. In other words, they also contain information about the *temporal behavior* of the underlying molecular process inside a single cell. We remark that the algorithms from Chapter 3 could be readily applied to this type of data by simply discarding the additional information. However, it turns out time-lapse data provides a highly informative source of information and can even resolve non-identifiabilities of parameters (see Section 4.2.4.2). In order to fully exploit all information present in time-lapse data, we need to follow a completely different strategy. In particular, suitable approaches are centered around general *state space models* [15] and associated inference techniques such as Bayesian filtering and smoothing [8].

4.1 Mathematical Modeling

We again consider a heterogeneous CTMC $X \mid (S, Z)$ describing the time evolution of a reaction network with L reaction channels and associated kinetic parameters $C = \{C_1, \dots, C_L\} = \{Z_1, \dots, Z_J, S_1, \dots, S_I\}$. Experimentally we can retrieve noisy measurements of a few molecular species for M cells at different measurement times t_l with $l = 1, \dots, N$. The acquisition error associated with the experimental technique is characterized by a conditional measurement density, i.e.,

$$Y_l^m \mid (X_l^m = x_l^m, \Omega = \omega) \sim p(y_l^m \mid x_l^m, \omega), \quad (4.1)$$

with Ω an unknown distribution parameter such as the acquisition noise variance. Furthermore, we define the state trajectory of cell m between the l -th and the k -th measurement time as $\mathbf{X}_{l:k}^m$ and denote by $Y_{l:k}^m$ the corresponding set of measurements. Image-based single-cell techniques can additionally capture morphological features of cells such as their volume or shape (see Figure 4.1a), which were shown to correlate well with extrinsic factors [101, 116]. They are incorporated by introducing *morphological covariates* V^m and hypothesizing a statistical dependency between these covariates and the extrinsic factors Z^m . This dependency is described by a conditional density

$$V^m \mid (Z^m = z^m, B = b) \sim p(v^m \mid z^m, b), \quad (4.2)$$

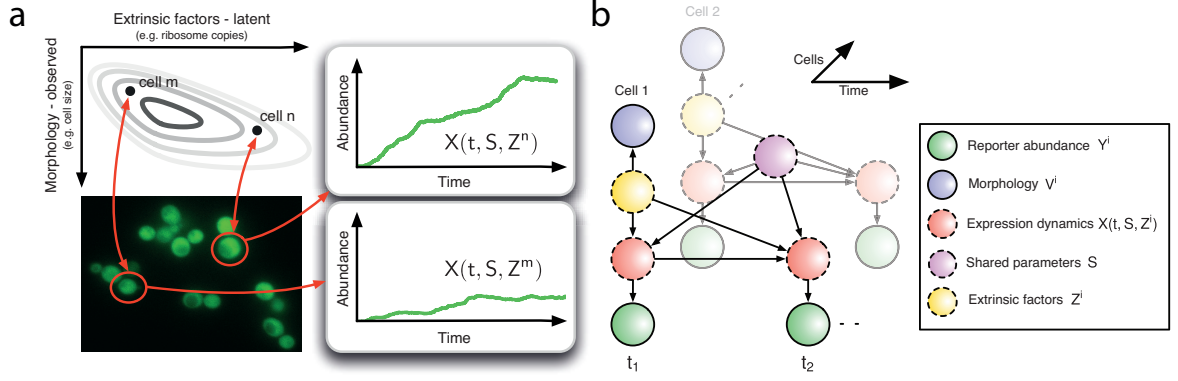


Figure 4.1: Modeling heterogeneous microscopy data. (a) Schematic generative model of the experimental data. On top of intrinsic fluctuations, extrinsic factors and their morphological covariates render individual cells different. The gene expression dynamics $X(t, S, Z^i)$ are hence characterized by a parameter set S that is shared across cells and a set of individual (i.e. extrinsic) parameters Z^i . (b) Corresponding Bayesian mixed-effect model. Nodes denote random variables and statistical dependency (and causality) is indicated by directed edges. Nodes with solid borders correspond to experimentally accessible quantities, whereas dashed nodes refer to unobserved variables. Extrinsic factors Z^i are assumed to be drawn from a common distribution.

with b a set of shape parameters characterizing this conditional density. The resulting mixed-effect state space model is depicted in Figure 4.1b.

4.1.1 A Straightforward Attempt at Inference

Formally, the posterior distribution over all unknown model quantities is given by

$$\begin{aligned}
 & p(a, b, \omega, s, z^1, \dots, z^M, \mathbf{x}_{1:N}^1, \dots, \mathbf{x}_{1:N}^M \mid y_{1:N}^1, \dots, y_{1:N}^M, v^1, \dots, v^M) \\
 & \propto p(a, b, \omega, z^1, \dots, z^M, \mathbf{x}_{1:N}^1, \dots, \mathbf{x}_{1:N}^M, y_{1:N}^1, \dots, y_{1:N}^M, v^1, \dots, v^M) \\
 & = \left[\prod_{m=1}^M \left(\prod_{l=1}^N p(y_l^m \mid x_l^m, \omega) \right) p(v^m \mid z^m, b) p(\mathbf{x}_{1:N}^m \mid s, z^m) p(z^m \mid a) \right] \\
 & \times p(a) p(b) p(\omega) p(s),
 \end{aligned} \tag{4.3}$$

i.e., it is given by the joint distribution up to a normalizing constant that depends only on the measurements. In that sense, any suitable MCMC sampler could be applied to draw samples from (4.3). For instance, we could run a Metropolis-within-Gibbs scheme and iteratively resample parameters and states using the full conditional distributions

$$\begin{aligned}
 & p(\mathbf{x}_{1:N}^1, \dots, \mathbf{x}_{1:N}^M \mid \omega, s, z^1, \dots, z^M, y_{1:N}^1, \dots, y_{1:N}^M) \\
 & \text{and } p(a, b, \omega, s, z^1, \dots, z^M \mid \mathbf{x}_{1:N}^1, \dots, \mathbf{x}_{1:N}^M, y_{1:N}^1, \dots, y_{1:N}^M, v^1, \dots, v^M).
 \end{aligned}$$

Via the chain rule of probability, we further note that

$$p(\mathbf{x}_{1:N}^1, \dots, \mathbf{x}_{1:N}^M \mid \omega, s, z^1, \dots, z^M, y_{1:N}^1, \dots, y_{1:N}^M) = \prod_{m=1}^M p(\mathbf{x}_{1:N}^m \mid \omega, s, z^m, y_{1:N}^m) \tag{4.4}$$

and

$$\begin{aligned}
& p(a, b, \omega, s, z^1, \dots, z^M \mid \mathbf{x}_{1:N}^1, \dots, \mathbf{x}_{1:N}^M, y_{1:N}^1, \dots, y_{1:N}^M, v^1, \dots, v^M) \\
&= p(s \mid \mathbf{x}_{1:N}^1, \dots, \mathbf{x}_{1:N}^M) \prod_{m=1}^M p(z^m \mid a, b, v^m, \mathbf{x}_{1:N}^m) \\
&\quad \times p(a, b \mid \mathbf{x}_{1:N}^1, \dots, \mathbf{x}_{1:N}^M, v^1, \dots, v^M) \\
&\quad \times p(\omega \mid \mathbf{x}_{1:N}^1, \dots, \mathbf{x}_{1:N}^M, y_{1:N}^1, \dots, y_{1:N}^M).
\end{aligned} \tag{4.5}$$

Hence, conditional on the measurements, the distribution in (4.5) factorizes into individual parts, from which sampling is fairly straightforward (e.g., using a Metropolis-Hastings criterium). However, in reference to Section 2.2.2, we realize that the individual terms within the product of (4.4) correspond to smoothing distributions, whose computation is generally challenging. Additionally, we are confronted with the usual scalability issues that arise when considering multiple cell measurements subject to extrinsic variability (see Section 2.1.5). Therefore, a straightforward approach like the one above is hardly able to deal with realistic problem settings. In the following section, we develop an inference algorithm which accounts for those limitations and hence, permits an application to experimental microscopy measurements.

4.2 Dynamic Prior Propagation

We can see from (4.3) that every considered cell $1, \dots, M$ increases the dimensionality of the posterior distribution due to the distinct extrinsic factors Z^1, \dots, Z^M . Picking up the marginalization concept from Chapter 3, we ask whether we can again integrate out those extrinsic factors from the posterior in order to obtain a parameter dimensionality that is independent of the population size M . For the moment, let us consider the distribution over all sample paths in a population of M heterogeneous cells, i.e.,

$$p(\mathbf{x}_t^1, \dots, \mathbf{x}_t^M \mid s, z^1, \dots, z^M) = \prod_{m=1}^M p(\mathbf{x}_t^m \mid s, z^m), \tag{4.6}$$

with \mathbf{x}_t^m as a sample path on a time interval $[0, t]$ corresponding to cell m . Assuming that $Z^m \mid (A = a) \sim p(z \mid a)$, the marginalization over extrinsic factor is performed as

$$\begin{aligned}
p(\mathbf{x}_t^1, \dots, \mathbf{x}_t^M \mid s, a) &= \int_{\mathcal{Z}} \cdots \int_{\mathcal{Z}} p(\mathbf{x}_t^1, \dots, \mathbf{x}_t^M, z^1, \dots, z^M \mid s, a) dz^1 \cdots dz^M \\
&= \prod_{m=1}^M \int_{\mathcal{Z}} p(\mathbf{x}_t^m \mid s, z^m) p(z^m \mid a) dz^m \\
&= \prod_{m=1}^M p(\mathbf{x}_t^m \mid s, a),
\end{aligned} \tag{4.7}$$

with $p(\mathbf{x}_t^m \mid s, a)$ as the marginal path density. Importantly, we realize that after the marginalization, all cells are associated with precisely the same path density showing a direct dependency on the extrinsic statistics a . Although conceptually similar, the above marginalization is different to that from Chapter 3, since it considers entire realizations of the heterogeneous processes – not just time-point-wise readouts thereof. An important question that arises in that context is whether there exists a *marginal process*, which admits the path density $p(\mathbf{x}_t^m \mid s, a)$. If yes, it would provide a coherent mathematical description of all heterogeneous cells at once.

4.2.1 Marginal Dynamics and the Innovation Theorem

Our goal is to find a dynamic description of the marginal process $X \mid (S, A)$, where the extrinsic parameters have been integrated out, such that their randomness is “self-contained” in the resulting process. Similar constructions have been performed within the theory of counting processes and they are based on the so-called *innovation theorem* [1–4]. For instance, the author of [2] studied the marginal dynamics of a simple three-state Markov Chain with random intensities. The following proposition shows how to construct the marginal dynamics of an arbitrary reaction network with propensity functions linear in Z (e.g., such as for mass-action kinetics).

Proposition 1. *The propensities of the reactions with index $j = 1, \dots, J$ of the marginal process $X \mid (S, A)$ are given by*

$$h_j(\mathbf{x}_t, t) = \mathbb{E}[Z_j \mid \mathbf{x}_t, a] g_j(x(t)), \quad (4.8)$$

with $\mathbb{E}[Z_j \mid \mathbf{x}_t, a]$ as the conditional expectation of Z_j given a sample path $\mathbf{x}_t = \{x(s) \mid s \in [0, t]\}$ and the extrinsic statistics a . All other propensities, i.e., those corresponding to reaction indices $i = J + 1, \dots, L$ remain unchanged.

Proof. Let $P(X(t + dt) = x(t) + \Delta_j \mid x(t), z_j)$ denote the probability that reaction j fires within the interval $[t, t + dt)$ given the current state $X(t) = x$, where Δ_j corresponds to the stoichiometric change vector of reaction j . Then, for the marginal jump probability we obtain

$$\begin{aligned} P(X(t + dt) = x(t) + \Delta_j \mid \mathbf{x}_t, a) &= \int_{\mathcal{Z}^J} P(X(t + dt) = x(t) + \Delta_j, z \mid \mathbf{x}_t, a) dz \\ &= \int_{\mathcal{Z}^J} P(X(t + dt) = x(t) + \Delta_j \mid x(t), z_j) p(z \mid \mathbf{x}_t, a) dz \\ &= \left(\int_{\mathcal{Z}^J} z_j p(z \mid \mathbf{x}_t, a) dz \right) g_j(x(t)) dt \\ &= \mathbb{E}[Z_j \mid \mathbf{x}_t, a] g_j(x(t)) dt \\ &= h_j(\mathbf{x}_t, t) dt. \end{aligned} \quad (4.9)$$

□

Remark 1 (MGF Representation). *We know from Section 2.1.2 and [62, 129] that the likelihood function of a path \mathbf{x}_t with respect to the parameters z is given by*

$$p(\mathbf{x}_t | z) \propto \prod_{i=1}^J z_i^{r_i} \exp \left\{ - \sum_{i=1}^J \left(z_i \int_0^t g_i(x(s)) ds \right) \right\}, \quad (4.10)$$

where r_i counts the number of reactions of type i . Then, by applying Bayes' formula, the conditional expectation in (4.8) can be written as

$$\begin{aligned} \mathbb{E}[Z_j | \mathbf{x}_t, a] &= \int_{\mathcal{Z}} z_j p(z_j | \mathbf{x}_t, a) dz_j \\ &= \int_{\mathcal{Z}} z_j \frac{p(\mathbf{x}_t | z_j) p(z_j | a)}{p(\mathbf{x}_t | a)} dz_j \\ &= \int_{\mathcal{Z}^J} z_j \frac{p(\mathbf{x}_t | z) p(z | a)}{p(\mathbf{x}_t | a)} dz. \end{aligned} \quad (4.11)$$

Hence, the marginal reaction hazard can be reformulated as

$$\begin{aligned} h_j(\mathbf{x}_t, t) &= \mathbb{E}[Z_j | \mathbf{x}_t, a] g_j(x(t)) \\ &= \left(\int_{\mathcal{Z}^J} z_j \frac{p(\mathbf{x}_t | z) p(z | a)}{p(\mathbf{x}_t | a)} dz \right) g_j(x(t)) \\ &= \left(\int_{\mathcal{Z}^J} z_j \frac{p(\mathbf{x}_t | z) p(z | a)}{\int_{\mathcal{Z}^J} p(\mathbf{x}_t | z) p(z | a) dz} dz \right) g_j(x(t)) \\ &= \frac{\mathbb{E}[z_j p(\mathbf{x}_t | z) | a]}{\mathbb{E}[p(\mathbf{x}_t | z) | a]} g_j(x(t)) \\ &= \frac{\mathbb{E} \left[z_j \prod_{i=1}^J z_i^{r_i} \exp \left\{ - \sum_{i=1}^J \left(z_i \int_0^t g_i(x(s)) ds \right) \right\} \mid a \right]}{\mathbb{E} \left[\prod_{i=1}^J z_i^{r_i} \exp \left\{ - \sum_{i=1}^J \left(z_i \int_0^t g_i(x(s)) ds \right) \right\} \mid a \right]} g_j(x(t)). \end{aligned} \quad (4.12)$$

We note that the expectations in the numerator and denominator can be rewritten as higher-order partial derivatives of the moment generating function (MGF)¹ of $Z | (A = a)$ [2] and we arrive at

$$h_j(\mathbf{x}_t, t) = \left[\frac{\partial \prod_{i=1}^J \partial^{r_i}}{\partial \sigma_j \prod_{i=1}^J \partial \sigma_i^{r_i}} G_{Z|a}(\sigma_1, \dots, \sigma_J) \left(\frac{\prod_{i=1}^J \partial^{r_i}}{\prod_{i=1}^J \partial \sigma_i^{r_i}} G_{Z|a}(\sigma_1, \dots, \sigma_J) \right)^{-1} \right] g_j(x(t)), \quad (4.13)$$

with $G_{Z|a}$ as the MGF of $Z | (A = a)$ and $\sigma_i \equiv - \int_0^t g_i(x(s)) ds$.

¹The moment generating function of a random vector $Z = (Z_1, \dots, Z_n)$ is defined as $\mathbb{E}[e^{\sigma_1 Z_1 + \dots + \sigma_n Z_n}]$.

In order to illustrate the calculation of the marginal hazard functions, a few examples for continuous as well as discrete extrinsic factors are provided in the following (Examples 5–8).

Example 5 (Univariate Gamma Distribution). *We assume a one-dimensional Gamma-distributed extrinsic parameter $Z \mid (A = a) \sim \mathcal{G}(\alpha, \beta)$ with $a = \{\alpha, \beta\}$ and reaction index 1. The MGF is known to be*

$$G_{Z|a}(\sigma) = \frac{\beta^\alpha}{(\beta - \sigma)^\alpha}$$

and the i -th derivative becomes

$$\frac{d^i}{d\sigma^i} G_{Z|a}(\sigma) = \frac{\Gamma(\alpha + i)}{\Gamma(\alpha)} \frac{\beta^\alpha}{(\beta - \sigma)^{\alpha+i}}.$$

Hence, by substituting into (4.13) the marginal hazard function is given by

$$h_1(\mathbf{x}_t, t) = \frac{\alpha + r_1}{\beta + \int_0^t g_1(x(s)) ds} g_1(x(t)). \quad (4.14)$$

Example 6 (Conditioning on Covariates). *Let us assume the case where additional covariates V of Z – e.g., morphological features – are obtained experimentally (see Section 4.1). We assume knowledge of a measurement density such that $V \mid (Z = z, B = b) \sim p(v \mid z, b)$. In this case, the marginal hazard functions become*

$$h_j(\mathbf{x}_t, v, t) = \left[\frac{\partial \prod_{i=1}^J \partial^{r_i}}{\partial \sigma_j \prod_{i=1}^J \partial \sigma_i^{r_i}} G_{Z|v,a,b}(\sigma_1, \dots, \sigma_J) \left(\frac{\prod_{i=1}^J \partial^{r_i}}{\prod_{i=1}^J \partial \sigma_i^{r_i}} G_{Z|v,a,b}(\sigma_1, \dots, \sigma_J) \right)^{-1} \right] g_j(x(t)),$$

where $G_{Z|v,a,b}$ is the MGF of $Z \mid (V = v, A = a, B = b) \sim p(z \mid v, a, b) \propto p(v \mid z, b)p(z \mid a)$. In case of Example 5, i.e., $Z \mid (A = a) \sim \mathcal{G}(\alpha, \beta)$ and $V \mid (Z = z, B = b) \sim \mathcal{G}(\rho, \phi z)$ with $b = \{\rho, \phi\}$, we obtain

$$\begin{aligned} p(z \mid v, a, b) &\propto \frac{\beta^\alpha}{\Gamma(\alpha)} z^{\alpha-1} \exp\{-\beta z\} \frac{(\phi z)^\rho}{\Gamma(\rho)} v^{\rho-1} \exp\{-\phi z v\} \\ &= \frac{\beta^\alpha \phi^\rho v^{\rho-1}}{\Gamma(\alpha)\Gamma(\rho)} z^{\alpha+\rho-1} \exp\{-z(\beta + \phi v)\}, \end{aligned}$$

and hence, $Z \mid (V = v, A = a, B = b) \sim \mathcal{G}(\alpha + \rho, \beta + \phi v)$. Then, the marginal reaction hazard is given by

$$h_1(\mathbf{x}_t, v, t) = \frac{\alpha + \rho + r_1}{\beta + \phi v + \int_0^t g_1(x(s)) ds} g_1(x(t)). \quad (4.15)$$

Example 7 (Poisson Distribution). *In many practical scenarios it might be the case that the reaction hazards are modulated by a discrete-valued extrinsic variable, e.g., due to a variability in copy numbers of certain species. For instance, if Z follows a Poisson distribution with rate parameter λ , the MGF is given by*

$$G_{Z|a}(\sigma) = \exp\{\lambda(e^\sigma - 1)\} \quad (4.16)$$

and its j -th derivative is found to be

$$\frac{d^j}{d\sigma^j} G_{Z|a}(\sigma) = G_{Z|a}(\sigma) \left(\sum_{k=1}^j d_{j,k} (\lambda e^\sigma)^k \right) \quad (4.17)$$

with the coefficients $d_{j,k}$ following the recursive relation

$$\begin{aligned} d_{j+1,1} &= 1 \\ d_{j+1,k} &= d_{j,k-1} + k d_{j,k} \quad \forall k = 2, \dots, j \\ d_{j+1,j+1} &= 1 \end{aligned} \quad (4.18)$$

The marginal hazard function can then be written as

$$h_i(\mathbf{x}_t, t) = \frac{\sum_{k=1}^{r_i+1} d_{r_i+1,k} (\lambda e^\sigma)^k}{\sum_{k=1}^{r_i} d_{r_i,k} (\lambda e^\sigma)^k} g_i(x(t)). \quad (4.19)$$

Example 8 (Bernoulli Distribution). *Consider a cell population that is randomly partitioned into two subgroups. For some reason (maybe due to different cell-cycle stages), one group can express a certain gene efficiently while the other group cannot. This can be modeled by assuming that the corresponding kinetic rate constants are drawn from a Bernoulli distribution such that each cell's rate constant is $Z = \kappa_0$ with probability p or $Z = \kappa_1$ with probability $q = 1 - p$. In this case, the MGF is given by*

$$G_{Z|a}(\sigma) = p e^{\sigma \kappa_0} + q e^{\sigma \kappa_1}. \quad (4.20)$$

Taking the r_i -th and the $(r_i + 1)$ -th derivative yields the marginal reaction hazard

$$h_i(\mathbf{x}_t, t) = \frac{p e^{\sigma \kappa_0} \kappa_0^{r_i+1} + q e^{\sigma \kappa_1} \kappa_1^{r_i+1}}{p e^{\sigma \kappa_0} \kappa_0^{r_i} + q e^{\sigma \kappa_1} \kappa_1^{r_i}} g_i(x(t)). \quad (4.21)$$

Note that the above derivations are straight-forward to extend for discrete probability distributions with more than two outcomes.

Although the marginal process framework applies to arbitrary scenarios, we will largely focus on the case where the extrinsic factors are distributed according to a Gamma distribution (i.e, Examples 5 and 6), which provides a reasonable compromise between analytical tractability and flexibility. More specifically, it represents a very versatile distribution on the positive orthant, ranging from over-dispersed and right-tailed, to under-dispersed and symmetric distributions. It was further shown to arise in the context of stochastic gene expression [35, 121].

4.2.2 Marginal Simulation Algorithm

We remark that the marginal hazard functions depend on the full process history \mathbf{X}_t through the conditional expectation from eq. (4.8), i.e., the resulting dynamics are non-Markovian. However, the Markov property can be enforced by virtually extending the state space by the summary statistics $\{r_1, \int_0^t g_1(x(s))ds\}, \dots, \{r_J, \int_0^t g_J(x(s))ds\}$ [137]. As a result, one can simulate sample paths of the marginal process using standard methods that can account for the explicit time-dependency of the hazard functions such as the first reaction method (see [6] and Section 2.1.3). In general, such algorithms rely on the generation of random waiting-times for each of the reaction channels. All reactions that are independent of the extrinsic factors will retain their exponentially distributed waiting-times. In contrast, the waiting-times W_1, \dots, W_J associated with the heterogeneous reactions are distributed according to

$$P(W_k < w \mid \mathbf{x}_t) = 1 - e^{-\int_0^w h_k(\mathbf{x}_{t+T, t+T})dT}, \quad (4.22)$$

with $k = 1, \dots, J$. Assuming that we have a means to sample from (4.22), it is straightforward to modify the first reaction method from Algorithm 1 by replacing the exponential waiting-time distribution by (4.22) for all heterogeneous reactions (see Algorithm 2).

Algorithm 2 (Marginal simulation algorithm). *The algorithm takes the stoichiometric change vectors ν_1, \dots, ν_L , the kinetic parameters c_{J+1}, \dots, c_{J+I} , the final time T and the initial state x_0 as input and returns the reaction types together with the associated firing times.*

-
- 1: Initialize variables $t \leftarrow 0$ and $x \leftarrow x_0$.
 - 2: **while** $t < T$ **do**
 - 3: **for** $i = 1, \dots, L$ **do**
 - 4: **if** $i \leq J$ **then**
 - 5: Draw waiting-time for the i -th reaction channel $W_i \sim (4.22)$.
 - 6: **else**
 - 7: Draw waiting-time for the i -th reaction channel $W_i \sim \text{Exp}(h_i(x, c_i))$.
 - 8: **end if**
 - 9: **end for**

-
- 10: Choose reaction associated with the minimal waiting-time $k \leftarrow \underset{i=1,\dots,L}{\operatorname{argmax}} W_i$.
- 11: Update state $x \leftarrow x + \nu_k$.
- 12: Update time $t \leftarrow t + W_k$.
- 13: Output t and x .
- 14: **end while**
-

We remark that in order to draw the non-exponential waiting-times for the i -th reaction (i.e., line 5 in Algorithm 2), we require the corresponding path statistics $\{r_i, \int_0^t g_i(x(s))ds\}$, which – for efficiency reasons – should be updated recursively as the algorithm evolves. Furthermore, the waiting-time distribution (4.22) may depend on additional hyperparameters a as can be seen from (4.8). For the particular case of Gamma-distributed extrinsic factors, the associated waiting-times turn out to be distributed according to a *Lomax*- or *Pareto-type-II* distribution such as shown in Example 9.

Example 9 (Waiting-time distribution). *The waiting-time distribution at time t for the i -th reaction with Gamma-type heterogeneity is given by*

$$\begin{aligned}
P(W_i < w \mid \mathbf{x}_t) &= 1 - \exp \left\{ - \int_0^w h_i(\mathbf{x}_{t+T}, t + T) dT \right\} \\
&= 1 - \exp \left\{ - (\alpha + r_i) \left(\ln \left[\beta + \int_0^{t+w} g_i(x(T)) dT \right] \right. \right. \\
&\quad \left. \left. - \ln \left[\beta + \int_0^t g_i(x(T)) dT \right] \right) \right\}
\end{aligned} \tag{4.23}$$

and with $G_i(t) := \int_0^t g_i(x(T)) dT$ we arrive at

$$\begin{aligned}
P(W_i < w \mid \mathbf{x}_t) &= 1 - \exp \left\{ - (\alpha + r_i) \left(\ln [\beta + G_i(t) + w g_i(x(t))] \right. \right. \\
&\quad \left. \left. - \ln [\beta + G_i(t)] \right) \right\} = 1 - \left[\frac{\frac{\beta + G_i(t)}{g_i(x(t))}}{\frac{\beta + G_i(t)}{g_i(x(t))} + w} \right]^{(\alpha + r_i)},
\end{aligned} \tag{4.24}$$

which we recognize as the CDF of a Pareto Type II - or “Lomax” distribution such that

$$W_i \mid \mathbf{x}_t \sim \operatorname{Lomax} \left(\frac{\beta + G_i(t)}{g_i(x(t))}, \alpha + r_i \right). \tag{4.25}$$

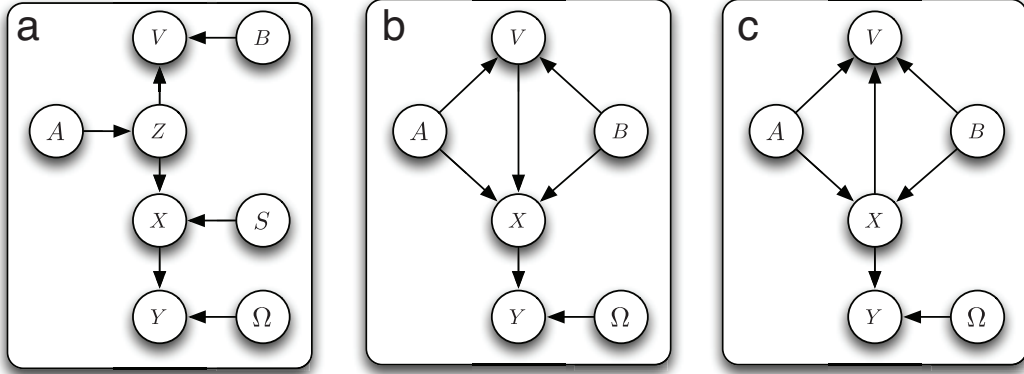


Figure 4.2: Graphical explanation of the marginalization. (a) Original Bayesian network. (b, c) Marginalized Bayesian networks. Both models represent valid Bayesian networks for the marginalized model (i.e., are mathematically equivalent), whereas they differ in the causality between V and X . Note that for clarity, individual time points are not represented separately in the above illustration.

The mean waiting-time is given by

$$\mathbb{E}[W_i | \mathbf{x}_t] = \frac{\beta + G_i(t)}{g_i(x(t))(\alpha + r_i - 1)}, \quad (4.26)$$

which for large α and β , converges to $\frac{\beta}{g_i(x(t))\alpha}$, which corresponds to the mean waiting-time obtained for a homogeneous reaction with rate constant $\frac{\alpha}{\beta}$.

4.2.3 Sequential Markov Chain Monte Carlo

While the marginal process framework permits an integration – and thus elimination – of the unknown extrinsic factors, we are still confronted with the second computational bottleneck, i.e., the inference of a general state space model given partial observations. More specifically, we aim to compute posterior distributions over states and the remaining (i.e., non-marginalized) parameters, giving rise to a simplified but still challenging inference problem. In the following, we develop a sequential Monte Carlo (SMC) algorithm, which allows to efficiently perform inference on the marginalized process model.

We remark that the marginalization also applies to the shared kinetic parameters modeled through mass-action, such that we can construct a stochastic process that depends on only very few unknown parameters, i.e., the extrinsic statistics A , the morphological shape parameters B and the acquisition noise parameter Ω (see Figure 4.2). Note that marginalized inference schemes generally profit from a variance reduction of the desired posterior statistics [28]². It can thus be considered a rule of thumb to marginalize analytically where ever possible.

²Such schemes are also denoted *Rao-Blackwellized* inference schemes since their reduction in variance can be explained through the Rao-Blackwell theorem.

The joint distribution of the marginalized model is given by

$$p(a, b, \omega, \mathbf{x}_{1:N}^1, \dots, \mathbf{x}_{1:N}^M, y_{1:N}^1, \dots, y_{1:N}^M, v^1, \dots, v^M) = \left[\prod_{m=1}^M \left(\prod_{l=1}^N p(y_l^m | x_l^m, \omega) \right) p(v^m | \mathbf{x}_{1:N}^m, b, a) \right] p(\mathbf{x}_{1:N}^1, \dots, \mathbf{x}_{1:N}^M | a, b) p(a) p(b) p(\omega). \quad (4.27)$$

We are interested in sampling from the posterior distribution

$$\begin{aligned} & p(\mathbf{x}_{1:N}^1, \dots, \mathbf{x}_{1:N}^M, a, b, \omega | \{y_{1:N}^1, v^1\}, \dots, \{y_{1:N}^M, v^M\}) \\ & \propto \left[\prod_{m=1}^M \left(\prod_{l=1}^N p(y_l^m | x_l^m, \omega) \right) p(v^m | \mathbf{x}_{1:N}^m, b, a) \right] p(\mathbf{x}_{1:N}^1, \dots, \mathbf{x}_{1:N}^M | a, b) p(a) p(b) p(\omega) \\ & = \left[\prod_{m=1}^M \left(\prod_{l=1}^N p(y_l^m | x_l^m, \omega) \right) p(v^m | b, a) \right] p(\mathbf{x}_{1:N}^1, \dots, \mathbf{x}_{1:N}^M | a, b, v^1, \dots, v^M) \\ & \quad \times p(a) p(b) p(\omega), \end{aligned} \quad (4.28)$$

where $\{\mathbf{x}_{1:N}^1, \dots, \mathbf{x}_{1:N}^M, a, b, \omega\}$ denote the unknown (i.e. *latent*) quantities. As sampling from the full latent space is practically impossible, we resort to a recursive Bayesian inference procedure, where the posterior distribution at time t_l is computed from the posterior distribution at time t_{l-1} as

$$\begin{aligned} & p(\mathbf{x}_{1:l}^1, \dots, \mathbf{x}_{1:l}^M, a, b, \omega | \{y_{1:l}^1, v^1\}, \dots, \{y_{1:l}^M, v^M\}) \\ & \propto \left[\prod_{m=1}^M p(y_l^m | x_l^m, \omega) p(\mathbf{x}_{l-1:l}^m | x_{l-1}^m, T_{l-1}^m, a, b, v^m) \right] \\ & \quad \times p(\mathbf{x}_{1:l-1}^1, \dots, \mathbf{x}_{1:l-1}^M, a, b, \omega | \{y_{1:l-1}^1, v^1\}, \dots, \{y_{1:l-1}^M, v^M\}). \end{aligned} \quad (4.29)$$

By exploiting the recursive relation of the posterior distribution, the original inference problem breaks up into a sequence of smaller problems which are easier solve. In reference to Section 2.2.2, we realize that (4.29) can be understood as a smoothing distribution. Sampling-based algorithms for the computation of such distribution go under the name of *sequential Monte Carlo* (SMC) methods [27, 28]. When applied to combined parameter- and state inference problems, standard SMC methods are likely to suffer from particle degeneracy, since the static parameters cannot take values different from their initialization at time t_1 . One strategy to overcome such problems is to randomly perturb the static parameters at each time step, in order to maintain diversity among the particles [118]. Practically, this means that the parameter values of a drawn particle are discarded and newly sampled. Ideally, the resampling should be performed such that the new parameter values are again a valid sample from the

posterior distribution. This is generally achieved by applying an *invariant kernel* to the original parameter values.

For that sake, note that the posterior can be written as

$$\begin{aligned} p(\mathbf{x}_{1:l}^1, \dots, \mathbf{x}_{1:l}^M, a, b, \omega \mid \{y_{1:l}^1, v^1\}, \dots, \{y_{1:l}^M, v^M\}) \\ = p(a, b, \omega \mid \{\mathbf{x}_{1:l}^1, y_{1:l}^1, v^1\}, \dots, \{\mathbf{x}_{1:l}^M, y_{1:l}^M, v^M\}) \\ \times p(\mathbf{x}_{1:l}^1, \dots, \mathbf{x}_{1:l}^M \mid \{y_{1:l}^1, v^1\}, \dots, \{y_{1:l}^M, v^M\}). \end{aligned} \quad (4.30)$$

Hence, diversified samples $\{\tilde{\mathbf{x}}_{1:l}^1, \dots, \tilde{\mathbf{x}}_{1:l}^M, \tilde{a}, \tilde{b}, \tilde{\omega}\}$ can be drawn by first sampling trajectories

$$\{\tilde{\mathbf{x}}_{1:l}^1, \dots, \tilde{\mathbf{x}}_{1:l}^M\} \sim p(\mathbf{x}_{1:l}^1, \dots, \mathbf{x}_{1:l}^M \mid \{y_{1:l}^1, v^1\}, \dots, \{y_{1:l}^M, v^M\}) \quad (4.31)$$

and subsequently drawing

$$\{\tilde{a}, \tilde{b}, \tilde{\omega}\} \sim p(a, b, \omega \mid \{\tilde{\mathbf{x}}_{1:l}^1, y_{1:l}^1, v^1\}, \dots, \{\tilde{\mathbf{x}}_{1:l}^M, y_{1:l}^M, v^M\}). \quad (4.32)$$

Within an SMC framework it is straight-forward to sample from the marginal distribution (4.31), given that the full posterior at a certain time step is available as a set of (possibly weighted) particles. Once a sample is drawn, marginalization is carried out by only considering the variables of interest (i.e., $\{\tilde{\mathbf{x}}_{1:l}^1, \dots, \tilde{\mathbf{x}}_{1:l}^M\}$). However, it appears to be complicated to directly sample from (4.32) using standard techniques such as the Metropolis-Hastings (M-H) algorithm. More specifically, such sampling methods rely on appropriate proposal distributions that are typically hard to find - especially if the sampling space is large. In such cases it can be beneficial to use a Gibbs-like Metropolis-Hastings sampler [77], where subsets of variables are resampled conditionally on the other variables, i.e., are drawn from the respective *full conditional* distributions³. Those distributions satisfy the invariance requirement [103] and can be easily identified from the underlying Bayesian network. On the one hand, some of the full conditional distributions might be of standard form, such that samples can be drawn straight away (e.g., if the distribution is Gaussian). On the other hand, it is often less challenging to find good proposal distributions for a single quantity such that better acceptance ratios can be achieved. Assume that the set of all variables in a network is partitioned into subsets $U = \{U_1, \dots, U_J\}$. Then the subset U_j is independent of all other variables in the Bayesian network when conditioned on its Markov blanket $\mathcal{MB}(U_j)$, defined as the set of parents, children and the children's parents of U_j [60]. Hence, the full conditional distribution over U_j is given by $p(U_j \mid \bar{U}_j) = p(U_j \mid \mathcal{MB}(U_j))$. Defining subsets $U_1 = \{\Omega\}$ and $U_2 = \{A, B\}$ we obtain

$$\tilde{\omega} \sim p(\omega \mid \{\tilde{x}_1^1, y_1^1\}, \dots, \{\tilde{x}_l^M, y_l^M\}) \quad (4.33)$$

$$\{\tilde{a}, \tilde{b}\} \sim p(a, b \mid \{\tilde{\mathbf{x}}_{1:l}^1, v^1\}, \dots, \{\tilde{\mathbf{x}}_{1:l}^M, v^M\}). \quad (4.34)$$

In the following, we will discuss in detail how to resample the individual quantities in (4.31), (4.33) and (4.34).

³Such a scheme can be understood as component-wise modification of the standard M-H sampler.

4.2.3.1 Resampling the Dynamic States

As indicated before, we sample the dynamic states by first simulating a particle from (4.30) and ignoring its static parameters which are resampled subsequently. In principle, sampling from (4.30) can be carried out by applying a M-H criterion. However, with the same argument as above, this will suffer from low acceptance ratios if many cells are in place and therefore, we again apply a component-wise modification, where each cell's trajectory is updated individually. In other words, we incorporate single measurements one after each other by additionally exploiting the recursive posterior structure over cells. To illustrate this, assume we have given the posterior distribution at time t_{l-1} over all cells. In order to incorporate the next measurement for the first cell at time t_l , we compute the posterior distribution

$$\begin{aligned} p(\mathbf{x}_{1:l}^1, \mathbf{x}_{1:l-1}^2, \dots, \mathbf{x}_{1:l-1}^M, a, b, \omega \mid \{y_{1:l}^1, v^1\}, \dots, \{y_{1:l-1}^M, v^M\}) \\ \propto p(y_l^1 \mid x_l^1, \omega) p(\mathbf{x}_{l-1:l}^1 \mid x_{l-1}^1, a, b, v^1, T_{l-1}^1) \\ \times p(\mathbf{x}_{1:l-1}^1, \dots, \mathbf{x}_{1:l-1}^M, a, b, \omega \mid \{y_{1:l-1}^1, v^1\}, \dots, \{y_{1:l-1}^M, v^M\}). \end{aligned} \quad (4.35)$$

Note that we can easily sample from (4.35) using a M-H criterion. If we propose samples

$$\begin{aligned} \{\tilde{\mathbf{x}}_{1:l}^1, \tilde{\mathbf{x}}_{1:l-1}^2, \dots, \tilde{\mathbf{x}}_{1:l-1}^M, \tilde{a}, \tilde{b}, \tilde{\omega}\} \sim p(\mathbf{x}_{l-1:l}^1 \mid x_{l-1}^1, a, b, v^1, T_{l-1}^1) \\ \times p(\mathbf{x}_{1:l-1}^1, \dots, \mathbf{x}_{1:l-1}^M, a, b, \omega \mid \{y_{1:l-1}^1, v^1\}, \dots, \{y_{1:l-1}^M, v^M\}), \end{aligned} \quad (4.36)$$

the acceptance probability reduces to

$$\gamma_x^1 = \min \left\{ 1, \frac{p(y_l^1 \mid \tilde{x}_l^1, \tilde{\omega})}{p(y_l^1 \mid x_l^1, \omega)} \right\}. \quad (4.37)$$

We proceed analogously for the remaining cells until we have obtained the full posterior distribution over all cells at time t_l . Note that samples from $p(\mathbf{x}_{l-1:l}^m \mid x_{l-1}^m, a, b, v^m, T_{l-1}^m)$ are easily obtained using a stochastic simulation algorithm like the one from Section 4.2.2. For all algorithms considered in this work, we adopt this choice of proposal distribution.

4.2.3.2 Resampling the Measurement Parameters

In most realistic cases, statistics of the measurement noise are unknown and hence, need to be included into the inference. We note that the measurement parameters Ω are independent of all other nodes in the network given the state at the sampling points x_i^m and measurements y_i^m for all $m = 1, \dots, M$, i.e., conditional on the set $\mathcal{MB}(\Omega) = \{\{X_i^m, Y_i^m\} \mid i = 1, \dots, l \wedge m = 1, \dots, M\}$. Therefore, the full conditional can be written as

$$p(\omega \mid \{x_1^1, y_1^1\}, \dots, \{x_l^M, y_l^M\}) \propto \left(\prod_{m=1}^M \prod_{i=1}^l p(y_i^m \mid x_i^m, \omega) \right) p(\omega). \quad (4.38)$$

Depending on the specific structure of measurement likelihood function $p(y_i^m | x_i^m, \omega)$, the corresponding unknown parameters Ω and the prior $p(\omega)$, eq. (4.38) might be of standard form. For instance, for a normally or log-normally distributed measurement noise with unknown scaling parameter $\Omega := \sigma$, gamma priors $\Omega \sim \mathcal{G}(\alpha_\omega, \beta_\omega)$ can be used. In particular, samples from the resulting full conditional are obtained as $\tilde{\omega} = \sqrt{1/\tilde{v}}$ with

$$\tilde{v} \sim \mathcal{G} \left(\frac{lM}{2} + \alpha_\omega, \frac{\sum_{m=1}^M \sum_{i=1}^l (x_i^m - y_i^m)^2}{2} + \beta_\omega \right) \quad (4.39)$$

in case of a normal measurement noise and with

$$\tilde{v} \sim \mathcal{G} \left(\frac{lM}{2} + \alpha_\omega, \frac{\sum_{m=1}^M \sum_{i=1}^l (\ln x_i^m - \ln y_i^m)^2}{2} + \beta_\omega \right) \quad (4.40)$$

in case of a log-normal measurement noise.

4.2.3.3 Resampling the Extrinsic Statistics and Morphological Shape Parameters

We know from (4.34) that A and B can be resampled conditional on $\{\mathbf{X}_{1:l}^1, V^1\}, \dots, \{\mathbf{X}_{1:l}^M, V^M\}$. The corresponding full conditional distribution takes the form

$$p(a, b | \{\tilde{\mathbf{x}}_{1:l}^1, v^1\}, \dots, \{\tilde{\mathbf{x}}_{1:l}^M, v^M\}) = \frac{1}{Z} \left(\prod_{m=1}^M p(\tilde{\mathbf{x}}^m | v^m, a, b) p(v^m | a, b) \right) p(a) p(b). \quad (4.41)$$

Evaluation of eq. (4.41) requires knowledge of the marginal path likelihood functions $p(\mathbf{x} | v, a, b^m)$. We assume the same configuration as in Example 6, i.e., one-dimensional extrinsic factors Z and corresponding covariates V . The path likelihood given the shared parameters S and extrinsic factors Z is given by

$$\begin{aligned} p(\mathbf{x} | s, z) &\propto \left(\prod_{i=J+1}^L s_i^{r_i} \exp \left\{ -s_i \int_0^t g_i(x(s)) ds \right\} \right) \times z^{r_1} \exp \left\{ -z \int_0^t g_1(x(s)) ds \right\} \\ &= \prod_{i=J+1}^L f_i(\mathbf{x}, s_i) \times f_1(\mathbf{x}, z). \end{aligned} \quad (4.42)$$

Due to the product form of (4.42), only the function f_1 corresponding to the extrinsic

factor Z will depend on A and B , i.e., the marginalized likelihood function becomes

$$\begin{aligned} f_1(\mathbf{x}, a, b, v) &= \int_{\mathcal{Z}} f_1(\mathbf{x}, z) p(z | a) p(v | z, b) dz \\ &\propto \int_{\mathcal{Z}} f_1(\mathbf{x}, z) p(z | v, a, b) dz \\ &= \mathbb{E}. [f_1(\mathbf{x}, z) | v, a, b] \end{aligned} \quad (4.43)$$

In case of Gamma-distributed extrinsic factors $Z | (A = a) \sim \mathcal{G}(\alpha, \beta)$ and covariates $V | (Z = z, B = b) \sim \mathcal{G}(\rho, \phi z)$, we know that $Z | (V = v, A = a, B = b) \sim \mathcal{G}(\alpha + \rho, \beta + \phi v)$. Hence, we further obtain

$$\begin{aligned} f_1(\mathbf{x}, a, b, v) &= \frac{(\beta + \phi v)^{\alpha + \rho}}{\Gamma(\alpha + \rho)} \int_{\mathcal{Z}} z^{(r_1 + \alpha + \rho - 1)} \exp \left\{ -z \left[\beta + \phi v + \int_0^t g_1(x(s)) ds \right] \right\} dz \\ &= \frac{(\beta + \phi v)^{\alpha + \rho} \Gamma(\alpha + \rho + r_1)}{\Gamma(\alpha + \rho)} \left[\beta + \phi v + \int_0^t g_1(x(s)) ds \right]^{-(\alpha + \rho + r_1)}. \end{aligned} \quad (4.44)$$

Similarly, the density $p(v | a, b)$ is found to be

$$p(v | a, b) = \int_{\mathcal{Z}} p(v | z, b) p(z | a) dz = \frac{\beta^\alpha \phi^\rho v^{\rho - 1} \Gamma(\alpha + \rho) (\beta + \phi v)^{-(\alpha + \rho)}}{\Gamma(\alpha) \Gamma(\rho)}. \quad (4.45)$$

Finally, we can rewrite (4.41) as

$$p(a, b | \{\tilde{\mathbf{x}}_{1:t}^1, v^1\}, \dots, \{\tilde{\mathbf{x}}_{1:t}^M, v^M\}) = \frac{1}{\tilde{Z}} \left(\prod_{m=1}^M f_1(\mathbf{x}_t^m, a, b, v^m) p(v^m | a, b) \right) p(a) p(b). \quad (4.46)$$

Due to the complicated structure of (4.44) and (4.45), we cannot directly sample from the full conditional such that we again make use of a M-H step, where proposed samples $\{\tilde{a}, \tilde{b}\} \sim q(\cdot, \cdot)$ are accepted with probability

$$\gamma_{A,B} = \min \left\{ 1, \frac{\prod_{m=1}^M f_1(\mathbf{x}_t^m, \tilde{a}, \tilde{b}, v^m) p(v^m | \tilde{a}, \tilde{b}) p(\tilde{a}) p(\tilde{b}) q(a, b)}{\prod_{m=1}^M f_1(\mathbf{x}_t^m, a, b, v^m) p(v^m | a, b) p(a) p(b) q(\tilde{a}, \tilde{b})} \right\}. \quad (4.47)$$

For the case studies considered later in this work, we chose q to be a multivariate log-normal distribution.

4.2.3.4 Full Posterior Reconstruction

Since we have marginalized the joint distribution with respect to certain variables, execution of the inference scheme can only deliver marginal posterior distributions.

More specifically, the algorithm returns a set of P particles consisting of the variables $\{\mathbf{X}_{1:l-1}^1, \dots, \mathbf{X}_{1:l-1}^M, A, B, \Omega\}$, while the shared and extrinsic parameters S and Z^1, \dots, Z^M are not included. However, the particle distribution over all variables can be easily reconstructed via the law of conditional probability, i.e.,

$$\begin{aligned} p(s, z^1, \dots, z^M, \mathbf{x}_{1:l}^1, \dots, \mathbf{x}_{1:l}^M, a, b, \omega \mid \{y_{1:l}^1, v^1\}, \dots, \{y_{1:l}^M, v^M\}) \\ = p(s \mid \mathbf{x}_{1:l}^1, \dots, \mathbf{x}_{1:l}^M) \left[\prod_{m=1}^M p(z^m \mid \mathbf{x}_{1:l}^m, v^m, b) \right] \\ \times p(\mathbf{x}_{1:l}^1, \dots, \mathbf{x}_{1:l}^M, a, b, \omega \mid \{y_{1:l}^1, v^1\}, \dots, \{y_{1:l}^M, v^M\}). \end{aligned} \quad (4.48)$$

This implies that a particle from the full posterior distribution can be constructed by first drawing a particle from the marginal distribution and subsequently sampling S and Z^1, \dots, Z^M conditional on that particle. If we assume Gamma-type prior distributions for each of the kinetic parameters, i.e., $S_i \sim \mathcal{G}(\kappa_i, \chi_i)$, the corresponding conditional distribution $p(s_i \mid \mathbf{x}_{1:l}^1, \dots, \mathbf{x}_{1:l}^M)$ is again Gamma, i.e., $\mathcal{G}(\kappa_i + R_i, \chi_i + G_i)$ with $R_i = \sum_{m=1}^M r_i^m$ and $G_i = \sum_{m=1}^M \int_0^{t_i} g_i(x^m(s)) ds$.

Moreover, the marginal parameter posterior can be written as a multivariate compound Gamma distribution

$$p(s \mid \{y_{1:l}^1, v^1\}, \dots, \{y_{1:l}^M, v^M\}) = \mathbb{E} [p(s \mid \mathbf{x}_{1:l}^1, \dots, \mathbf{x}_{1:l}^M)] \quad (4.49)$$

with $p(s \mid \mathbf{x}_{1:l}^1, \dots, \mathbf{x}_{1:l}^M)$ given by the product of the individual conditional distributions, i.e.,

$\prod_{i=J+1}^L \mathcal{G}(\kappa_i + R_i, \chi_i + G_i)$. Note that R_i and G_i are functions of the sample paths and that the expectation in (4.49) is taken with respect to the smoothing distribution

$$p(\mathbf{x}_{1:l}^1, \dots, \mathbf{x}_{1:l}^M \mid \{y_{1:l}^1, v^1\}, \dots, \{y_{1:l}^M, v^M\}) \approx \frac{1}{P} \sum_{p=1}^P \mathbb{1}_{\Lambda_S^{(p)}}(\mathbf{x}_{1:l}^1, \dots, \mathbf{x}_{1:l}^M), \quad (4.50)$$

with $\Lambda_S^{(p)} = \{\mathbf{x}_{1:l}^1, \dots, \mathbf{x}_{1:l}^M\}^{(p)}$ collecting the M sample paths of the p -th particle. Consequently - within the finite particle representation - (4.49) is approximated as a sum of Gamma distributions.

Analogously, we can perform the reconstruction of the marginal (and joint) posterior for the extrinsic factor Z , which for clarity is again assumed to be one-dimensional. Note however, that according to the product form in (4.48), this can be carried out independently for each cell. For the m -th cell we compute the compound density

$$p(z^m \mid \{y_{1:l}^1, v^1\}, \dots, \{y_{1:l}^M, v^M\}) = \mathbb{E} [p(z^m \mid \mathbf{x}_{1:l}^m, v^m, a, b)], \quad (4.51)$$

where the expectation is taken with respect to the distribution

$$p(\mathbf{x}_{1:l}^m, a, b \mid \{y_{1:l}^1, v^1\}, \dots, \{y_{1:l}^M, v^M\}) \approx \frac{1}{P} \sum_{p=1}^P \mathbb{1}_{\Lambda_Z^{(p)}}(\mathbf{x}_{1:l}^m, a, b), \quad (4.52)$$

with $\Lambda_Z^{(p)} = \{\mathbf{x}_{1:l}^m, a, b\}^{(p)}$. The full conditional distribution inside the expectation factorizes as

$$p(z^m | \mathbf{x}_{1:l}^m, v^m, a, b) \propto p(\mathbf{x}_{1:l}^m | z^m)p(v^m | z, b)p(z^m | a) \quad (4.53)$$

and within the setup of Example 6, is again given by a Gamma distribution, i.e., $\mathcal{G}(\alpha + r_1^m + \rho, \beta + \int_0^{t_l} g_1(x^m(s))ds + \phi v^m)$. It follows that the marginal posterior distributions over the extrinsic factors Z^m can again be approximated by sums of Gamma distributions.

4.2.3.5 Implementation Aspects

Although the algorithm structure is mathematically characterized by the foregoing derivations, several variants thereof are possible. It turns out that certain implementation details may have significant impact on the achieved performance. For instance, resampling of the static parameters can be carried out before or after sampling the dynamic states. Both strategies are mathematically correct, however, the former generally achieves better results as every particle that is drawn from the distribution at t_{l-1} receives a newly sampled value. In contrast, when sticking to the latter strategy, some of the resampled parameters are immediately lost as the corresponding particles are never drawn again at the subsequent time step. Furthermore - unlike classical sequential importance sampling methods - the proposed algorithm requires a short burn-in period at each time iteration. In all simulation studies, we discarded around 10 percent of the particles.

Finally, we mention that one is free to choose the order of the recursive updates, meaning that single measurements can be incorporated first over time and then over cells or vice versa (i.e., *time-point-first*- and *cells-first* mode). If one is mainly interested in parameter estimation, we recommend to use the latter strategy, whereas the sequence of processed cells at a particular time-step should be chosen randomly. This is likely to produce diversified summary statistics and in turn smooth posterior distributions over parameters. In contrast, processing entire cell trajectories one after each other appears to be beneficial if one aims to perform a state reconstruction. In the same context, we would like to discuss an interesting modification of the algorithm. In particular, it is based on the idea to resample and update individual cells simultaneously, i.e., without updating the posterior between consecutive cells (i.e., *simultaneous* mode).

While a theoretical analysis of that algorithm shall be performed in the future, it has proven to perform well for both state reconstruction and parameter inference while getting along with comparably few particles per time instance. An exemplary implementation of the marginal SMCMC algorithm (i.e., the *cells-first* variant) is summarized in Algorithm 3.

Algorithm 3 (Marginal SMCMC). *We assume that we have given the most recent posterior distribution as a set of particles. Then, the updated posterior distribution*

incorporating the next measurement of cell m is obtained by:

-
- 1: **for** particle $p = 1, \dots, P$ **do**
 - 2: Select the p -th particle from the particle distribution at time t_{l-1} .
 - 3: Resample the measurement parameters $\tilde{\omega}$ using (4.39) or (4.40).
 - 4: Resample the extrinsic statistics and morphological shape parameters $\tilde{\xi} = \{\tilde{\alpha}, \tilde{\beta}, \tilde{\rho}, \tilde{\phi}\}$ using a M -H step with proposal density $q(\tilde{\xi}) = \prod_{i=1}^4 \mathcal{LN}(\ln \xi_i, \sigma_{\xi}^2)$ and acceptance probability (4.47).
 - 5: Propose a sub-trajectory $\hat{\mathbf{x}}_{l-1:l}^m \sim p(\mathbf{x}_{l-1:l}^m \mid x_{l-1}^m, \tilde{a}, \tilde{b}, v^m, T_{l-1}^m)$ by simulating the marginal dynamics on $[l-1, l]$ using the resampled parameters \tilde{a} and \tilde{b} .
 - 6: Merge the sub-trajectories to obtain a full sample path $\hat{\mathbf{X}}_{1:l}^m = \{\mathbf{x}_{1:l-1}^m, \hat{\mathbf{x}}_{(l-1,l)}^m\}$.
 - 7: **if** $p = 1$ **then**
 - 8: Accept particle with probability 1.
 - 9: **else**
 - 10: Accept particle with probability

$$\gamma_x^m = \min \left\{ 1, \frac{p(y_l^m \mid \hat{\mathbf{x}}_l^m, \tilde{\omega})}{p(y_l^m \mid \hat{\mathbf{x}}_l^m, \omega)} \right\},$$
 where ω denotes the measurement noise parameter of the previous particle.
 - 11: **end if**
 - 12: Update p -th particle of the posterior distribution at time t_l using the newly sampled quantities.
 - 13: **end for**
-

Note that the case where no morphological features are used for the inference can be understood as a special instance (or simplification) of the described algorithm. Hence, we do not provide additional equations for this scenario - in particular as they are straight-forward to obtain from the provided derivations.

4.2.3.6 Bayes Factor Computation

We perform model selection by calculating the *Bayes factor* for two competing models $\mathcal{M} = 1$ and $\mathcal{M} = 2$ with equal prior probability $P(\mathcal{M} = 1) = P(\mathcal{M} = 2) = 0.5$. We consider pooled time-course measurements from a heterogeneous population but for simplicity exclude the morphological features from the analysis. Note that in this case, only the parameters A and Ω are required to specify the model. The Bayes factor is then given by

$$\mathcal{K}_{1,2} = \frac{p(y_{1:l}^1, \dots, y_{1:l}^M \mid \mathcal{M} = 1)}{p(y_{1:l}^1, \dots, y_{1:l}^M \mid \mathcal{M} = 2)}. \quad (4.54)$$

Within the SMC MC framework, the marginal likelihood (i.e., the *model-evidence*) computations in (4.54) turn out to be straight forward as they can be carried out recursively

Table 4.1: Inferred model parameters and credible intervals for the synthetic case study.

Name	Reference	Unit	Prior			Posterior		
			q_5	q_{95}	Mean	q_5	q_{95}	Mean
c_2	$5.00e-03$	s^{-1}	$3.40e-04$	$2.03e-02$	$6.67e-03$	$3.79e-03$	$6.85e-03$	$5.23e-03$
c_3	$1.00e+00$	s^{-1}	$5.13e-02$	$2.98e+00$	$1.00e+00$	$8.43e-01$	$1.18e+00$	$1.01e+00$
c_4	$2.00e-02$	s^{-1}	$1.78e-03$	$9.72e-02$	$3.33e-02$	$1.70e-02$	$2.67e-02$	$2.17e-02$
α	$3.00e+00$	–	$2.62e-01$	$7.16e+00$	$2.25e+00$	$5.26e-01$	$5.10e+00$	$2.22e+00$
β	$1.00e+02$	–	$3.20e+01$	$8.54e+02$	$2.71e+02$	$1.47e+01$	$1.28e+02$	$5.39e+01$
ρ	$5.00e+00$	–	$7.04e-01$	$7.81e+01$	$1.96e+01$	$9.14e-01$	$1.09e+01$	$4.19e+00$
ϕ	$1.00e+03$	–	$7.74e+01$	$7.32e+03$	$1.98e+03$	$1.22e+02$	$1.91e+03$	$6.93e+02$
ω	$1.50e-01$	–	$6.50e-02$	$2.36e-01$	$1.26e-01$	$1.24e-01$	$1.57e-01$	$1.39e-01$

as

$$\begin{aligned}
 p(y_{1:l}^1, \dots, y_{1:l}^M \mid \mathcal{M} = k) &= p(y_1^1, \dots, y_1^M \mid \mathcal{M} = k) \\
 &\times \prod_{i=2}^N p(y_i^1, \dots, y_i^M \mid y_{1:i-1}^1, \dots, y_{1:i-1}^M, \mathcal{M} = k).
 \end{aligned} \tag{4.55}$$

The individual terms in (4.55), i.e., the predictive densities, are given by

$$p(y_i^1, \dots, y_i^M \mid y_{1:i-1}^1, \dots, y_{1:i-1}^M, k) = \mathbb{E} \left[\prod_{m=1}^M p(y_i^m \mid x_i^m, \omega) \right] \tag{4.56}$$

where the expectation is taken with respect to $p(\mathbf{x}_{1:i}^1, \dots, \mathbf{x}_{1:i}^M, a, \omega \mid y_{1:i-1}^1, \dots, y_{1:i-1}^M)$, which is simply obtained by drawing a particle $\{\mathbf{x}_{1:i-1}^1, \dots, \mathbf{x}_{1:i-1}^M, a, \omega\}^{(p)}$ from the previous time step and extending the dynamic states until t_i using the parameters $a^{(p)}$ and $\omega^{(p)}$ from that particle.

4.2.4 Application to Synthetic Gene Expression Data.

We first studied the proposed inference framework using simulated data of a simple two-state gene expression model [96] given in Figure 4.3a under realistic measurement conditions. We assume that the target gene can be activated by stimulation with an extra-cellular signal which results in the translocation of a transcription factor. Extrinsic variability was simulated by introducing a Gamma-distributed variability in the protein translation rate. Data was collected for 20 cells, on which we applied DPP using 10,000 Monte Carlo samples per measurement time instance (for further details, see caption of Figure 4.3a-b). The inferred posterior distributions over kinetic parameters, extrinsic statistics and acquisition noise parameter – characterizing the measurement variability – are depicted in Figure 4.3b. Table 4.1 summarizes the inferred parameter values and their respective credible intervals.

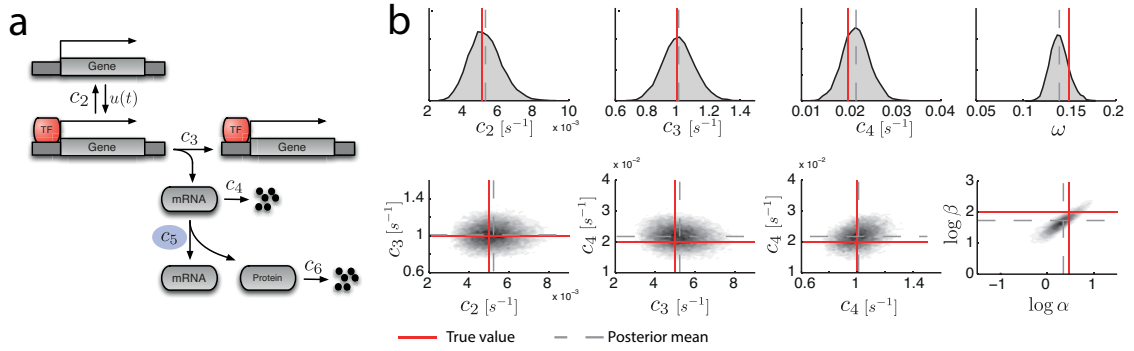


Figure 4.3: Parameter inference using simulated measurements. (a) Two-state gene expression model. The gene activation event is controlled by a time-varying rate $u(t)$ and includes transcription initiation events such as the binding of transcription factors, RNA polymerase and possibly chromatin remodeling. Upon activation of the gene, mRNA can be transcribed to further yield new proteins. All reactions are modeled according to mass-action. Altogether, the model comprises four species and six reactions, where we assume a Gamma-distributed heterogeneity in the translation efficiency, i.e., c_5 drawn from a Gamma distribution $\mathcal{G}(\alpha, \beta)$, with the extrinsic statistics α and β . We assume that 10 noisy (log-normally distributed with unknown scaling parameter) measurements of the protein abundance can be obtained at equally spaced time points within a total interval of roughly 40 minutes. (b) Parameter inference from protein time series using 20 cells. Inference results are shown for the three kinetic parameters (c_2 , c_3 and c_4), the extrinsic statistics α and β and the scaling parameter ω of the acquisition noise. The lower panel shows 2D posterior density plots, visualizing the a-posterior correlations between pairs of parameters. DPP was performed using 10,000 samples per time instance using the *simultaneous* mode.

The inferred posterior distribution over unobserved states can also be used to reconstruct activation and transcription events at the promoter. In general, the inverse problem of reconstructing promoter activation states from the slow protein dynamics is considerably ill-posed and we expect the posterior density in sequence space to be close to degenerate. However, for the simulation study, Figure 4.4 indicates that accurate detection of the promoter state is indeed possible within a realistic scenario. Using the expression model as described in Figure 4.3, we simulated a double-pulsed induction of gene expression. Noisy versions of the simulated protein abundance at sparse time points were taken as our available measurements and mRNA and promoter dynamics was reconstructed using DPP. We remark that constructing a sequence of events by simply determining the maximum of the posterior distribution time-point-wise (referred to as MMAP in Figure 4.4) does not yield a valid promoter activation sequence and hence cannot be used to extract timing statistics, such as promoter on/off times. Consequently, we determined the most likely posterior (MAP) sequence among the set of all possible promoter sequences from which timing statistics can be recovered correctly.

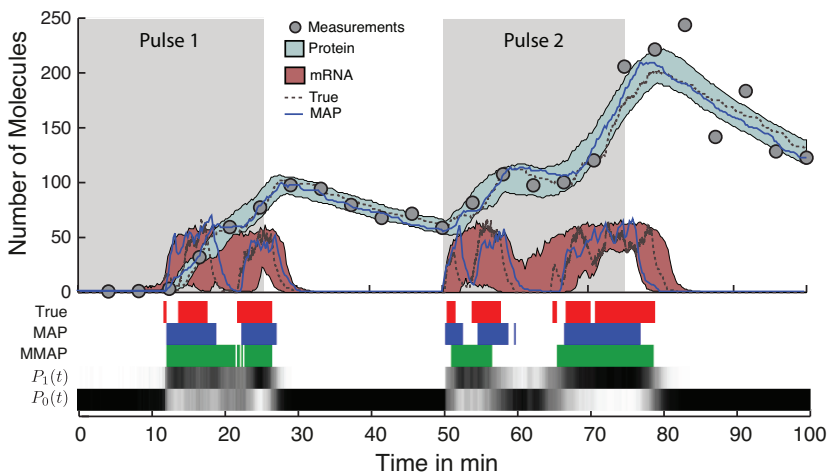


Figure 4.4: Reconstructed gene expression dynamics for a double-pulsed induction (gray boxes) from simulated protein measurements (circles). Shaded areas denote 5%- and 95%-quantiles of the posterior distributions over states and the true values (dashed) and the maximum a posteriori (MAP) trajectory (blue) are shown. Also shown are the posterior probabilities for the promoter to be inactive or active, $P_0(t)$ and $P_1(t)$, respectively; black denotes probability one. For reference, we also show the time-point-wise MAP (MMAP) estimate for the active promoter state (green) based on $P_0(t)$ and $P_1(t)$.

4.2.4.1 Improved Identifiability via Pooled Recordings

Joint inference of multiple rate constants from single trajectories often yields practical non-identifiabilities. However, the ill-posedness of such problems can be drastically reduced by pooling recordings over multiple cells. This is demonstrated in the following simulation study where two parameters of the model from Figure 4.3a (i.e., c_2 and c_4) are jointly estimated using (i) one and (ii) ten single-cell trajectories. For both rate constants, we assumed prior distributions of the form $\mathcal{G}(1, 10)$. Density plots of the prior and the posteriors for case (i) and (ii) are depicted in Figure 4.5.

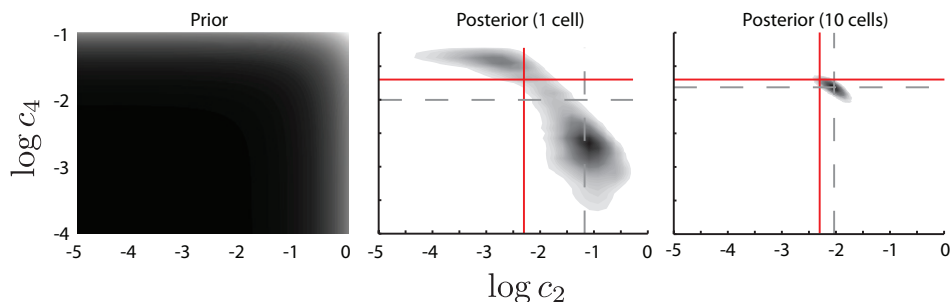


Figure 4.5: Improved identifiability via pooled single-cell trajectories; prior distribution (left), posterior distribution based on data from one cell (middle) and posterior based on ten cells (right); true parameter values (red lines) and posterior mean parameter estimates (dashed lines).

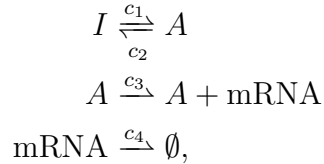
The left panel of Figure 4.5 demonstrates a weakly informative prior distribution.

As a result, a single cell is hardly enough to jointly identify both parameters, i.e., the posterior mass is distributed over roughly four orders of magnitude in both dimensions. When increasing the number of measured trajectories to ten, this non-identifiability is widely resolved such that both parameters can be inferred accurately.

4.2.4.2 Comparison between Time-Lapse and Population Snapshot Data

Recent inference approaches were developed for the use with population snapshot data and thus, do not include temporal information on a single-cell level, such as provided by time-lapse microscopy data (see Section 2.3). We indicated in the beginning of this chapter that using such an approach on the data we provide means that a significant portion of information is discarded. This will in turn lead to less accuracy or possibly even non-identifiabilities in the resulting parameter inference.

Dealing with stochastic models, there exists a lower bound on the prediction uncertainty that just corresponds to the process uncertainty itself (i.e. present even in the case of complete knowledge of the parameters). Any positive deviation from this bound stems from further parameter uncertainty represented by the posterior distribution. Assuming finite data records, there exist again fundamental lower bounds for this uncertainty (e.g., the minimum mean squared error). While some inference methods may exploit all features of the data, some others may not. Accordingly, the former will achieve this minimal posterior uncertainty bound while the latter will not. Hence, one anticipates a higher variance in the predictive distribution for methods that do not exploit the temporal correlation structure in the data. To confirm this expectation, we performed a simple case study using the transcriptional model proposed in [96], i.e.,



where I and A refer to the molecular states where the promoter is inactive or active (see Figure 4.6).

4.2.5 Application to Experimental Gene Expression Data

DPP was used to reconstruct the expression dynamics of an artificially controlled gene expression system in yeast. A widely used system to control the expression of genes under a GAL1 promoter in *Saccharomyces cerevisiae* is based on the hormone-dependent activation of the chimeric transcription factor GAL4DBD.ER.VP16 (GEV) [71, 75]. GEV consists of a strong transcriptional activator, made by fusing the GAL4 DNA binding domain (GAL4DBD) with the hormone-binding domain of the human estrogen receptor (ER) [71] and the transcription activating domain of the herpes simplex virus protein VP16 [107]. In its inactive state, GEV associates with the Hsp90 chaperone

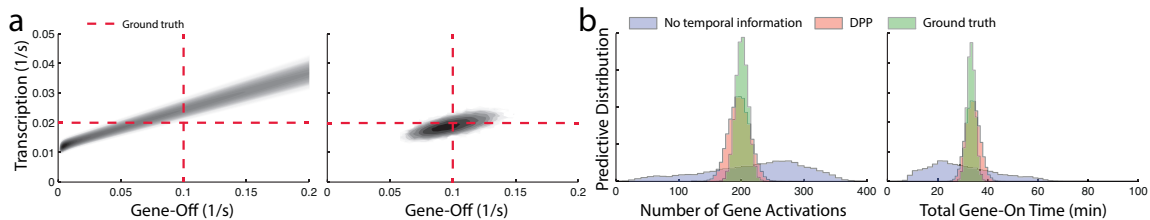


Figure 4.6: Predictions obtained using an inference scheme that neglects temporal information within single-cells (e.g., [88]) and DPP. Measurements of 20 cells were simulated at 20 equally spaced time points between 0 and 100min using a log-normal measurement model with $\sigma = 0.15$. Inference was performed with respect to gene-off- and transcription rates. Panel (a) shows the respective posterior distributions (left: population snapshot, right: DPP). Panel (b) shows the predictive distributions for two quantities characterizing the underlying process: (left) the total number the gene is activated within 100min. (right) the total time the gene is in its active state within 100min. The predictions are compared to the distributions obtained using the true kinetic parameters (ground truth).

complex and resides in the cytoplasm. Upon addition of the exogenous hormone β -estradiol to the extracellular medium, β -estradiol diffuses through the cell membrane and binds to the GEV’s ER. Thereby, Hsp90 disassociates from the complex and active GEV translocates to the nucleus where its GAL4DBD recognizes and binds to GAL promoter regions. VP16 then activates transcription of the downstream gene. [137]

We engineered a strain that allows a combined readout of GEV translocation and β -estradiol-induced gene expression. A GEV-mCherry construct in combination with a nuclear marker allows to compute the ratio of nuclear to cytoplasmic GEV. In the same strain, a destabilized [48, 126] version of the Venus fluorescent protein (Y-Venus) was placed under control of a GAL1 promoter, which allows a more accurate tracking of the gene expression dynamics.

The fluorescence microscopy experiments were performed using a flow chamber that allowed us to rapidly exchange the extracellular media and apply a 30min pulse of 50nM β -estradiol. The time-lapse microscopy movies were automatically analyzed [93] to quantify the change in nuclear localization of the GEV-mCherry protein and the fluorescent levels of the Y-Venus protein in individual cells. Additional calibration experiments with reference strains were performed to map recorded intensities to total protein abundances. Based on a one-sided Kolmogorov-Smirnov test we determined and corrected for a time delay in the protein measurements, that arises from an aggregation of unmodeled sequential events, such as mRNA export, post-transcriptional/translational modifications and reporter maturation. We then used 20 single cell trajectories of Y-Venus abundance within the subsequent analyses. Assuming that the translocation event occurs uniformly across cells and considering the high-abundance of GEV, we take the average translocation curve as input to our gene expression models. For fuller details on the experimental techniques, the reader may refer to [137].

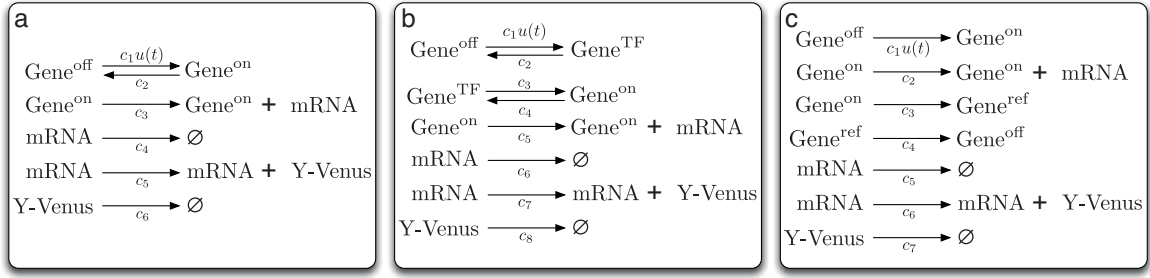


Figure 4.7: Three candidate models for β -estradiol-induced gene expression.

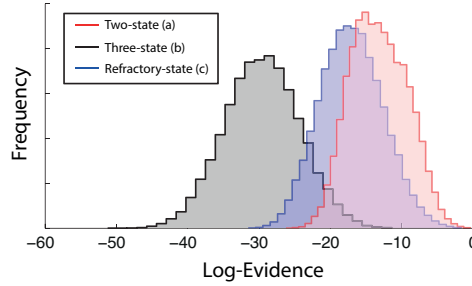


Figure 4.8: Robustness evaluation of Bayesian model selection with respect to models **a**, **b** and **c**. All histograms were shifted by the same constant for a better visualization.

4.2.5.1 Modeling pGAL1 Y-Venus Expression and Acquisition

For the GAL1 expression system we investigated three different models of eukaryotic gene expression and determined their empirical evidence by Bayesian model selection. Next to the canonical two-state model [96] (see Figure 4.7a) of a promoter, we considered a model with a third refractory state [49, 119] (see Figure 4.7c) and a three-state variant of a model proposed in [16], where the initiation-complex assembly is followed by a slow activation step representing either RNA polymerase (RNAP) binding or chromatin remodeling (see Figure 4.7b). Subsequent mRNA and protein synthesis are modeled as first order events. Note that the obtained model rankings are subject to sampling variance. Hence, in case of the three kinetic models, we computed the model-evidences five times in order to check the robustness of the obtained results. We know from section 4.2.3.6 that the evidence is computed as a product of predictive densities, each corresponding to a particular time-point. Hence, in order to approximately assess the variability over different runs, we randomly combined the individual terms over the five repeats and computed histograms over the resulting model-evidences. The results (shown in Figure 4.8) demonstrate that the two-state model **a** ranked best. Although there is a significant overlap with the histogram of the three-state model **c**, the results indicate no need for using the more complicated model **c**. Moreover, the figure indicates little evidence for model **b** compared to **a** and **c**. We then computed the average model-evidence and the respective Bayes factors, i.e., around $6.8dB$ when comparing **a** to **c** and $51.8dB$ when comparing **a** to **b**.

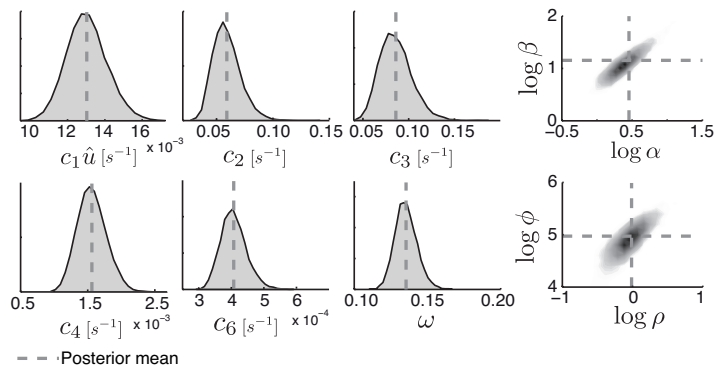


Figure 4.9: Parameter inference for the β -estradiol-induced Y-Venus expression in yeast. Posterior distributions over unknown parameters. Based on 20,000 samples for the first time-iteration and 10,000 samples for the subsequent iterations DPP (*simultaneous* mode) was performed with respect to all kinetic parameters (c_1, \dots, c_4, c_6), the acquisition noise parameter ω , the extrinsic statistics (α, β) and the morphological shape parameters (ρ, ϕ) characterizing the hypothesized dependency of volume increase and translation rate. The marginal posterior for the gene-on rate is shown for $c_1 \hat{u}$, with \hat{u} as the temporal average over the modulating GEV intensity.

Model selection was also performed for two competing measurement noise models (i.e., i.i.d. normal and log-normal). Strong evidence was found for log-normally distributed measurement noise (>100 dB). We next describe the choice of the prior distributions for the winning model **a**.

For c_1 , c_2 and c_3 we used weakly informative exponential priors, i.e., $p(\cdot) = \mathcal{G}(1, 10)$, with their quantiles shown in Table 4.2. In case of the mRNA and protein degradation rates c_4 and c_6 , the priors were chosen such that roughly 95 percent of the probability mass were within the ranges $[3, 40]$ and $[15, 2000]$ minutes expected half life (i.e., $C_4 \sim \mathcal{G}(3, 2000)$ and $C_6 \sim \mathcal{G}(1, 5000)$). Furthermore, for the hyperparameters α and β , a two-dimensional log-normal prior distribution $\mathcal{LN}(\mu_A, \Sigma_A)$ with

$$\mu_A = \begin{pmatrix} 2.71 \\ 5.70 \end{pmatrix} \text{ and } \Sigma_A = \begin{pmatrix} 0.20 & 0.00 \\ 0.00 & 0.20 \end{pmatrix}$$

was used such as to be consistent with the results found in [89]. In order to avoid a biased inference, we again picked a weakly informative log-normal prior distribution over the morphological parameters ρ and ϕ , i.e., $\mathcal{LN}(\mu_B, \Sigma_B)$ with

$$\mu_B = \begin{pmatrix} 2.30 \\ 10.82 \end{pmatrix} \text{ and } \Sigma_B = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

Parameter inference, state reconstruction and promoter activity detection were then performed as described for the synthetic case study. Figure 4.9 shows the posterior distribution over unknown parameters. The prior- and posterior statistics of the parameters are shown in Table 4.2.

Table 4.2: Inferred model parameters and credible intervals for the β -estradiol induced gene expression system.

Name	Unit	Prior			Posterior		
		q_5	q_{95}	Mean	q_5	q_{95}	Mean
c_1	s^{-1}	$5.35e-03$	$3.03e-01$	$1.00e-01$	$1.08e-01$	$1.41e-01$	$1.24e-01$
c_2	s^{-1}	$5.40e-03$	$3.03e-01$	$1.00e-01$	$3.98e-02$	$5.93e-02$	$4.92e-02$
c_3	s^{-1}	$5.00e-03$	$2.96e-01$	$1.00e-01$	$6.72e-02$	$9.27e-02$	$7.99e-02$
c_4	s^{-1}	$4.14e-04$	$3.16e-03$	$1.50e-03$	$1.30e-03$	$1.71e-03$	$1.51e-03$
c_6	s^{-1}	$1.13e-05$	$6.07e-04$	$2.00e-04$	$2.30e-04$	$2.87e-04$	$2.57e-04$
α	—	$8.78e+00$	$3.83e+01$	$2.04e+01$	$3.87e+00$	$3.83e+01$	$1.41e+01$
β	—	$1.60e+02$	$1.56e+03$	$6.29e+02$	$3.88e+01$	$4.01e+02$	$1.45e+02$
ρ	—	$7.38e+00$	$7.84e+02$	$2.05e+02$	$7.87e-01$	$1.42e+01$	$4.72e+00$
ϕ	—	$3.52e+04$	$3.98e+06$	$1.02e+06$	$6.69e+04$	$3.32e+06$	$1.02e+06$
ω	—	$8.94e-02$	$2.44e-01$	$1.48e-01$	$1.13e-01$	$1.38e-01$	$1.25e-01$

4.2.5.2 Model Predicts Mild Bursting in the GEV-pGAL1 System

The order of the estimated mRNA half-life of around 10min and mRNA synthesis rate of 6 molec/min are inline with previous findings [139]. The specific value of the latter is above most reported rates for constitutively expressing genes [139] which appears consistent with the fact that we use the strong VP-16 transcription activator. This synthesis rate together with the length of 850 bp for the Y-Venus protein and a reported elongation speed of 2kb/min [73] for GAL-driven genes indicates that there need to be roughly three RNAPs on average on the gene. Figure 4.10a shows the reconstructed states for two cells with different Y-Venus abundances. The predicted timing statistics of the promoter activation sequences indicates that for successful initiations on average around 2.5 transcripts per active promoter state are produced, suggesting that transcription re-initiation and thus mild bursting takes place in this expression system. We measured additional pulse experiments of 25nM and 100nM β -estradiol which we used to validate the inferred results. More specifically, the calibrated model was forward-simulated to check consistency with the performed measurements. Figure 4.11a shows that the model predictions agree well with the experimentally obtained data across different concentrations of β -estradiol.

4.2.5.3 Noise Contributions in pGAL1 Y-Venus Expression

Figure 4.10a also indicates to which extent a cell's expression level is explained by extrinsic and intrinsic factors. Although cell 1 shows mRNA levels similar to cell 2, the former expresses significantly more Y-Venus due to a larger translation rate c_5 . By forward simulating the inferred model we dissected and quantified the different sources of variability in the measured Y-Venus abundance to separate intrinsic, extrinsic and

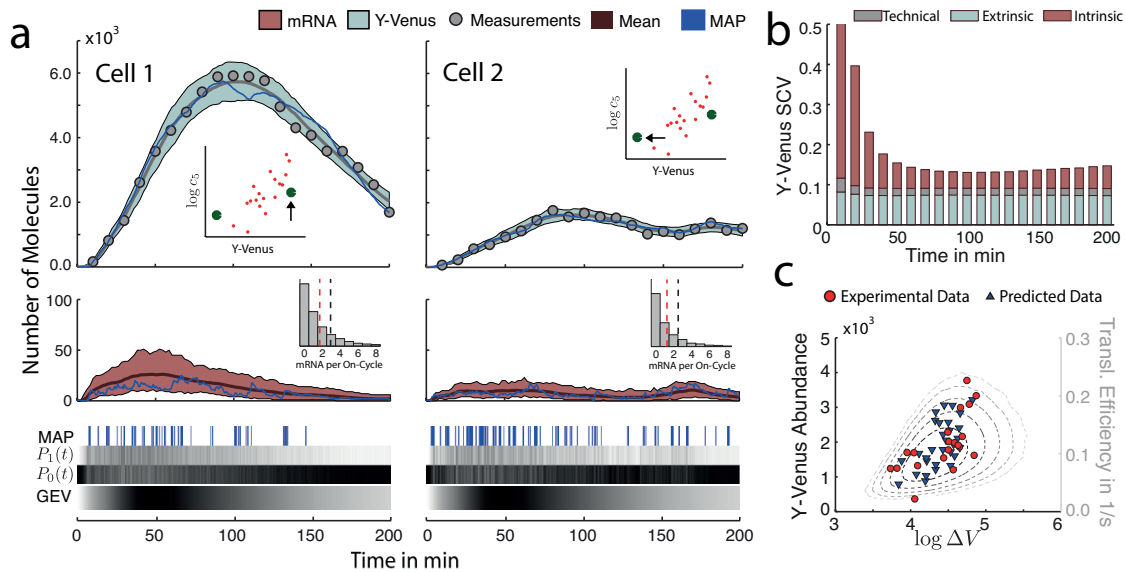


Figure 4.10: State reconstruction of heterogeneous Y-Venus dynamics. (a) Inferred dynamics for two exemplary cells showing different Y-Venus abundance. The temporal GEV-induction is shown as an intensity map, where white and black coloring denote minimal and maximal abundance, respectively. The inset scatter plots indicate the inferred expression regime, where each dot represents a single cell and the arrows point towards the respective cell; the inset x- and y-axes correspond to the logarithm of the temporal mean of Y-Venus and the mean posterior estimate of translation rate c_5 , respectively. The panel below shows the inferred mRNA dynamics. Therein, the inset with the transcripts per on cycle distribution over all posterior activation sequences shows a mean around one (gray dashed) and mean around 2.5 taking only successful initiation events in account (red dashed). (b) Sources of cell-to-cell variability in Y-Venus expression. The inferred model was used to compute the squared coefficient of variation (SCV) of the Y-Venus abundance. The total SCV was decomposed into technical, intrinsic and extrinsic components. (c) Dependency between volume increase and Y-Venus abundance. Multiple cell's intensity trajectories, their volume increase and translation efficiency (i.e., c_5) were computed via forward-simulation of the inferred model. The plot shows the Y-Venus abundance at time 200min versus the volume increase from 0min until 200min for the predicted (blue) and experimental (red) data. The inferred statistical dependency between the volume increase and the translation efficiency is indicated by the dashed iso-lines (gray).

technical contributions to the overall variability. In particular, it holds that

$$\text{SCV}[Y_l] = \underbrace{\frac{\mathbb{E}[\mathbb{E}[\text{Var}[Y_l | X_l] | Z]]}{\mathbb{E}[Y_l]^2}}_{\text{technical}} + \underbrace{\frac{\mathbb{E}[\text{Var}[\mathbb{E}[Y_l | X_l] | Z]]}{\mathbb{E}[Y_l]^2}}_{\text{intrinsic}} + \underbrace{\frac{\text{Var}[\mathbb{E}[\mathbb{E}[Y_l | X_l] | Z]]}{\mathbb{E}[Y_l]^2}}_{\text{extrinsic}}.$$

For a derivation of (4.57) the reader might refer to [137]. In order to perform the decomposition, the model parameters were set to their mean posterior values and the individual quantities were obtained via forward simulation (see Figure 4.10b). In the technical contribution any systematic bias introduced, for instance by the image segmentation algorithm, is not considered. The inferred model predicts that the variability

in pGAL1 Y-Venus expression is significantly driven by extrinsic factors. This is further supported by the results from Figure 4.12, which demonstrate that a model that just accounts for intrinsic and technical noise, fails to predict the cell-to-cell variability in the data.

We further validated our predictions using an independent dataset from a dual-reporter experiment (YFP, CFP) of the same promoter under identical experimental conditions. Intrinsic and extrinsic noise were quantified according to [99]. We remark that the latter approach does not account for technical variability (e.g., due to image segmentation errors), which will hence be subsumed in the intrinsic and/or extrinsic parts. The predicted noise contributions are in good agreement with the variance decomposition from the dual-reporter experiment across different concentrations of β -estradiol (Figure 4.11b). The decomposition is also shown across time points for the 50nM pulse experiment. While the overall noise characteristics are well captured by the model predictions, deviations are visible at early time points. Although background fluorescence levels have been estimated and subtracted from the dual-reporter data, correlated residuals will persist due to estimation uncertainties. In case of very low abundances (i.e., at early time points), such residuals are likely to dominate and cause an overestimation of extrinsic contributions. This effect vanishes at later time points, since the actual protein abundances outweigh those residuals.

4.2.5.4 Morphological Features and Extrinsic Variability

On top of the protein measurements, DPP allows to incorporate additional single-cell readouts such as morphological features. More specifically, it is able to quantify statistical dependencies between such readouts and a population's extrinsic factors.

Here we hypothesized a dependency between volume increase during the observation time interval and the extrinsic factor (i.e., the translation efficiency) and quantified it using DPP. The found covariation is depicted in Figure 4.10. Consistent with [23] we find that volume increase positively correlates with translation efficiency but that it does not explain all extrinsic variability present in the intensity trajectories. Hence, this provides more evidence for the fact that simple normalization through morphological features (e.g. forward scattering in flow cytometry data [136]) can not sufficiently correct for extrinsic variability in data.

4.2.5.5 Comparison between Homogeneous and Heterogeneous Kinetic Models

Most state-of-the-art approaches for parameter inference in biochemical networks do not account for extrinsic variability – that means – they rely on the assumption that the recorded measurements stem from a homogeneous cell population. If the considered biochemical system is characterized by significant heterogeneity, such models and their respective inference will yield biased results – either in the resulting parameter estimates or subsequent predictions. In order to demonstrate this issue, we refitted the model under the assumption of homogeneity using the same prior distributions over the

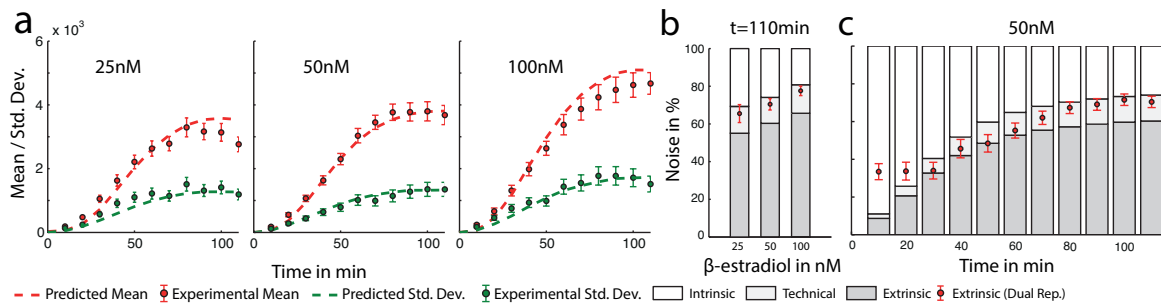


Figure 4.11: Model validation using additional datasets. (a) Prediction of reporter dynamics. Y-Venus trajectories were recorded upon application of β -estradiol pulses of 25, 50 and 100nM. The plot shows a comparison between the predicted and experimental means (upper panels) and standard deviations (lower panels) of the Y-Venus dynamics. The whiskers indicate the standard errors of the experimentally obtained quantities. (b-c) Noise contributions. An additional dual-reporter dataset was recorded for the same induction system and experimental conditions as in (a). Intrinsic and extrinsic noise contributions were estimated using a traditional variance decomposition [99]. Whiskers indicate standard errors of the experimentally obtained quantities. In order to account for differences in the reporter lifetimes, the degradation rate of the model was set to the previously reported YFP half-life of the dual-color strain [136]. (b) Comparison between dual-reporter experiments and model predictions at the final time point (i.e., 110min) across different concentrations of β -estradiol. (c) Noise decomposition across different time points for the 50nM pulse experiment.

kinetic and measurement parameters. Subsequently, that model was used to predict the three pulse experiments. Figure 4.12 shows the predicted and experimental means and standard deviations of the Y-Venus abundance for the different concentrations. For completeness, also the predictions from the original (i.e., heterogeneous) model are shown. We find that in this case, a model that neglects extrinsic variability is not able to explain the large variability present in the data. While the mean dynamics are captured well for all concentrations, the standard deviation is consistently underestimated, highlighting the importance of inference schemes that can account for extrinsic variability.

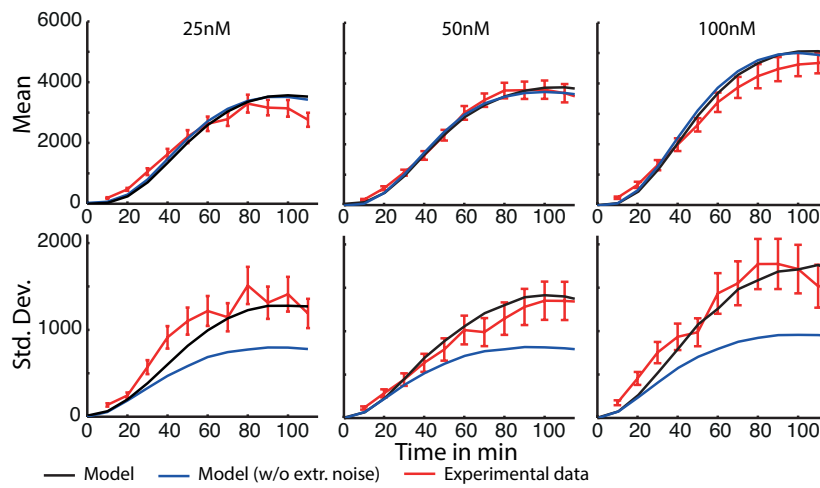


Figure 4.12: Predictions obtained using an inference scheme that neglects extrinsic variability. The results are compared to the experimental data as well as to the model predictions obtained via DPP. Whiskers indicate standard errors of the experimentally obtained quantities.

4.3 Detecting Sources of Extrinsic Variability

While DPP was developed to account for extrinsic noise, it relies on pre-specified assumptions on where extrinsic noise enters a kinetic model – e.g., through its translation rate. However, in most practical scenarios making such assumptions is challenging because the true sources and their strength for a specific cell line are yet to be identified. In this section we lay out an empirical Bayesian inference scheme, which automatically determines the heterogeneous reactions of a biochemical network using time-lapse single-cell data. The scheme is based on a hierarchical Bayesian model whose evidence is iteratively maximized using a variational Bayesian expectation-maximization algorithm [14, 87, 122]. We demonstrate and analyze the approach using a simple model of eukaryotic gene expression. In order to reduce the number of unknown parameters, we again make use of the marginalized process introduced in Section 4.2. For a simple demonstration of the main idea, the outlined method is described and analyzed for the case of complete and noise-free measurements but describe in Section 4.3.3 how the method can be generalized to the realistic scenario of partially observed- and noisy data. The following results represent a first step towards a model-based understanding of how and which concurrent processes modulate a specific cellular process under study *in vivo*.

4.3.1 Hierarchical Bayesian Modeling

We now consider the special case where every reaction is potentially modulated by extrinsic factors such that $C = \{C_1, \dots, C_L\} = \{Z_1, \dots, Z_L\}$. Moreover, we assume that the extrinsic factors are distributed according to independent Gamma distribution, i.e.,

$$p(z^1, \dots, z^L | a) = \prod_{i=1}^L \mathcal{G}(z^i; \alpha_i, \beta_i), \quad (4.57)$$

with $a = \{\{\alpha_i, \beta_i\} | i = 1, \dots, L\}$ as a set of extrinsic statistics. Assume we have given measurements of M cells of a heterogeneous population, i.e., \mathbf{x}^m for $m = 1, \dots, M$. The extrinsic variability of each reaction channel i can be quantified by inferring the extrinsic statistics $\{\alpha_i, \beta_i\}$ from those measurements. According to a Bayesian scenario, this is equivalent to finding the posterior distribution

$$\begin{aligned} p(a | \mathbf{x}^1, \dots, \mathbf{x}^M) &\propto \prod_{m=1}^M p(\mathbf{x}^m | a) p(a) \\ &= \prod_{m=1}^M \left(\prod_{i=1}^L p(\mathbf{x}^m | \alpha_i, \beta_i) p(\alpha_i, \beta_i) \right), \end{aligned} \quad (4.58)$$

which hence, factorizes such that

$$p(a | \mathbf{x}^1, \dots, \mathbf{x}^M) = \prod_{i=1}^L p(\alpha_i, \beta_i | \mathbf{x}^1, \dots, \mathbf{x}^M). \quad (4.59)$$

Naively, one could just evaluate the individual terms in (4.59) and check whether the corresponding values of α_i are below a certain threshold, indicating heterogeneity of the associated reaction. However, since those values are only accessible through the noisy measurements \mathbf{x}^m , it is not clear how to choose such a threshold in order to obtain maximally robust results. For instance, the heterogeneity stemming from the intrinsic molecular fluctuations should be "filtered out" and yield a negative detection result. Positive detections are only desired if there is significant evidence in the data. Technically, this can be understood as a *sparse* Bayesian learning problem [14, 87, 122]. The key step to achieve sparsity in empirical Bayesian models is to assign suitable prior - and hyperprior distributions to the model quantities. Since detection of heterogeneity is based on only α_i , we chose

$$p(\alpha_i, \beta_i) = p(\alpha_i | \lambda_i) p(\beta_i), \quad (4.60)$$

where λ_i controls the shape of $p(\alpha_i | \lambda_i)$ and $p(\beta_i)$ is assumed to be flat over the positive domain, such that $p(\alpha_i, \beta_i) \propto p(\alpha_i | \lambda_i)$. The goal is to define $p(\alpha_i | \lambda_i)$ such that the heterogeneity is forced to zero unless there is significant evidence in the data. Accordingly, suitable distributions will emphasize SCVs around zero while also permitting high values. Here we choose $p(\alpha_i | \lambda_i)$ such that $p(\eta_i) = \text{Exp}(\lambda_i)$ with $\eta_i = \alpha_i^{-1}$ as the SCV. A transformation of random variables yields

$$p(\alpha_i | \lambda_i) = \frac{\lambda_i}{\alpha_i^2} e^{-\frac{\lambda_i}{\alpha_i}}. \quad (4.61)$$

The resulting prior distributions over α_i are illustrated in Figure 4.13 for different values of λ_i .

While standard Bayesian approaches rely on *given* prior knowledge, empirical Bayes techniques aim to infer parameters as well as their hyperparameters from data. In our case, this means that in addition to α_i and β_i , also the hyperparameters λ_i are assumed to be unknown and need to be estimated. In order to obtain a fully Bayesian model, we need to specify hyperprior distributions $p(\lambda_i)$. Again, we assume $p(\lambda_i)$ to be flat but remark that an extension to arbitrary distributions is possible. With the model parameters a and their hyperparameters $\rho = \{\lambda_i | i = 1, \dots, L\}$, we aim to compute the posterior distribution

$$\begin{aligned} p(a, \rho | \mathbf{x}^1, \dots, \mathbf{x}^M) & \\ & \propto \prod_{i=1}^L \left(\prod_{m=1}^M p(\mathbf{x}^m | \alpha_i, \beta_i) \right) p(\alpha_i | \lambda_i) p(\lambda_i) \\ & = \prod_{i=1}^L p(\alpha_i, \beta_i, \lambda_i | \mathbf{x}^1, \dots, \mathbf{x}^M) \end{aligned} \quad (4.62)$$

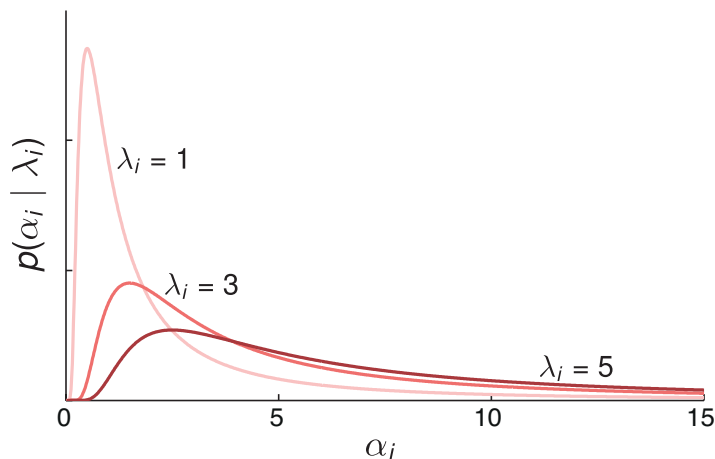


Figure 4.13: Prior distributions over α_i for different values of the hyperparameter λ_i . The distributions show a peak for low values of α_i and become more heavy-tailed with increasing λ_i .

where the r.h.s. of (4.62) is just the joint distribution over all model quantities. Unfortunately, it turns out that the posterior (4.62) is challenging to solve analytically. In the next section we will develop a variational inference scheme for its approximate solution.

4.3.2 Variational Inference

Variational inference schemes aim to approximate some target posterior $p(z | y)$ by some other distribution $q(z)$. More specifically, one chooses $q(z)$ such as to minimize the Kullback-Leibler divergence (KL) between $q(z)$ and the true distribution. For that sake, note that for every q , the log-evidence function satisfies the decomposition [11, 14]

$$\ln p(z) = \mathcal{L}[q(z)] + KL[q(z) \| p(z | y)], \quad (4.63)$$

where $\mathcal{L}[q(z)]$ forms a lower bound on $\ln p(z)$ which is given by

$$\mathcal{L}[q(z)] = \int q(z) \ln \frac{p(z, y)}{q(z)} dz. \quad (4.64)$$

Accordingly, minimizing the KL with respect to q is the same as maximizing its counterpart $\mathcal{L}[q(z)]$, i.e.,

$$q^*(z) = \operatorname{argmax}_{q(z) \in \mathcal{Q}} \mathcal{L}[q(z)]. \quad (4.65)$$

It can be seen from (4.63) and (4.64) that $\mathcal{L}[q(z)]$ is maximal if and only if $q(z) = p(z | y)$. In order to obtain a tractable $q(z)$, one typically imposes further constraints on its structure. Most commonly, individual components of z are assumed to be independent

of each other, i.e.,

$$q(z) = \prod_{l=1}^L q(z_l), \quad (4.66)$$

also known as the *mean-field* approximation [11]. In this case, it can be shown that the optimal variational solution of the individual factors $q(z_i)$ is determined by

$$\ln q^*(z_i) = \mathbb{E}_{j \neq i} [\ln p(z, y)] + \text{const.} \quad (4.67)$$

where $\mathbb{E}_{j \neq i} [\ln p(z, y)]$ denotes the expectation of the logarithm of the joint distribution, taken with respect to all factors $q(z_j)$ except $q(z_i)$. Since the optimal solution of a particular q -factor depends on all other factors, the mean-field approximation typically induces an iterative inference scheme, where the individual factors are updated in a round-robin fashion. Such schemes stand in close relation with traditional expectation-maximization (EM) algorithms [25] and accordingly, are often referred to as variational Bayesian EM (VBEM) algorithms [11, 113].

In practice, eq. (4.67) might still be intractable, in which case it is necessary to further restrict the corresponding q -factor. For instance, one could assume $q(z_i)$ to be some parameterized distribution (e.g., a Gaussian with mean and variance) and determine its parameters θ as

$$\theta^* = \operatorname{argmax}_{\theta \in \Theta} \mathbb{E} [\ln p(z, y)], \quad (4.68)$$

whereas in this case, the expectation is taken with respect to all q -factors. For instance, if one is interested solely in maximum a-posterior (MAP) estimates, $q(z_i)$ can be chosen to be a Dirac-delta function with unknown position.

We will now use the VBEM framework to derive an approximate iterative inference algorithm for the hierarchical Bayesian model from Section 4.3.1. The goal is to compute an approximate posterior distribution $q(a, \rho)$ for which we assume that it factorizes as

$$q(a, \rho) = \prod_{i=1}^L q(\alpha_i, \beta_i) q(\lambda_i). \quad (4.69)$$

We remark that in the complete-data scenario considered here, also the true posterior factors over the individual reaction channels $i = 1, \dots, L$, however, not over $\{\alpha_i, \beta_i\}$ and λ_i . For analytical simplicity, we further assume $q(\lambda_i) := \delta(\lambda_i - \hat{\lambda}_i)$ with $\hat{\lambda}_i$ as an unknown position parameter. The factor $q(\alpha_i, \beta_i)$ for the i -th reaction channel is determined by

$$\ln q^*(\alpha_i, \beta_i) = \mathbb{E}_{\lambda_i} [\ln p(a, \rho, \mathbf{x}^1, \dots, \mathbf{x}^M)] + \text{const.}, \quad (4.70)$$

which becomes

$$\begin{aligned} \ln q^*(\alpha_i, \beta_i) &= \sum_{m=1}^M \ln p(\mathbf{x}^m \mid \alpha_i, \beta_i) \\ &+ \mathbb{E}_{\lambda_i} [\ln p(\alpha_i \mid \lambda_i)] + \text{const.} \end{aligned} \quad (4.71)$$

when taking into account the r.h.s. of eq. (4.62). Together with the marginal path-likelihood function from eq. (4.44), we further obtain

$$\begin{aligned}
& \ln q^*(\alpha_i, \beta_i) \\
&= \sum_{m=1}^M \alpha_i \ln \beta_i + \ln \Gamma(\alpha_i + r_i^m) - \ln \Gamma(\alpha_i) \\
&\quad - (\alpha_i + r_i^m) \ln \left(\beta_i + \int_0^T g_i(x^m(t)) dt \right) \\
&\quad - \frac{\hat{\lambda}_i}{\alpha_i} - 2 \ln \alpha_i + \text{const.},
\end{aligned} \tag{4.72}$$

where we have used the fact that

$$\mathbb{E}_{\lambda_i} [\lambda_i] = \int \lambda_i \delta(\lambda_i - \hat{\lambda}_i) d\lambda_i = \hat{\lambda}_i.$$

Although eq. (4.72) is not of standard form, it can be evaluated analytically or using a suitable sampling algorithm.

The q -factor corresponding to λ_i is found by solving the parametric (instead of variational) optimization

$$\begin{aligned}
\lambda_i^* &= \operatorname{argmax}_{\hat{\lambda}_i \in \mathbb{R}} \mathbb{E} [\ln p(a, \rho, \mathbf{x}^1, \dots, \mathbf{x}^M)] \\
&= \operatorname{argmax}_{\lambda_i \in \mathbb{R}} \mathbb{E} [\ln p(\alpha_i | \lambda_i)].
\end{aligned} \tag{4.73}$$

The expectation inside the maximum operator is given by

$$\mathbb{E} [\ln p(\alpha_i | \lambda_i)] = -\hat{\lambda}_i \mathbb{E}_{\alpha_i} [\alpha_i^{-1}] + \ln \hat{\lambda}_i - 2 \mathbb{E}_{\alpha_i} [\ln \alpha_i], \tag{4.74}$$

whose maximum is found to be

$$\lambda_i^* = \frac{1}{\mathbb{E}_{\alpha_i} [\alpha_i^{-1}]}. \tag{4.75}$$

4.3.2.1 Implementation Aspects

As mentioned earlier, the VBEM scheme leads to an iterative algorithm, where all q -factors are estimated successively, given the most recent estimates of all other q -factors. For a particular reaction channel i , this means that we first determine $q(\alpha_i, \beta_i)$ given the most recent value of $\hat{\lambda}_i$ and subsequently re-estimate $\hat{\lambda}_i$ given $q(\alpha_i, \beta_i)$ and so forth. Since $q(\alpha_i, \beta_i)$ is not of standard form, we can compute its required statistics either via numerical integration or Monte Carlo sampling. Here we focus on the latter approach and employ a standard M-H sampler with log-normal proposal distributions to draw samples from $q(\alpha_i, \beta_i)$. Those samples are also used for updating the corresponding

hyperparameters λ_i , i.e., the expectation in eq. (4.75) is replaced by a Monte Carlo average. Moreover, we found that replacing $\mathbb{E}_{\alpha_i}[\alpha_i^{-1}]$ by $\mathbb{E}_{\alpha_i}[\alpha_i]^{-1}$ yields a similar estimation performance, while significantly reducing the number of required divisions per iteration.

Note that the parameters corresponding to the homogeneous reaction channels will be driven to infinity, which in theory, causes the algorithm to diverge. Practically – however – one can check whether α_i (or λ_i) is above a critical threshold (e.g., around $10e5$), in which case the i -th reaction is considered homogeneous and excluded from the remaining analysis. Algorithm 4 summarizes the main structure of the proposed scheme.

Algorithm 4 (VBEM algorithm). *VBEM algorithm for detecting heterogeneity in stochastic interaction networks.*

```

1: Initialize  $\hat{\lambda}_i$  for  $i = 1, \dots, L$ 
2: while not converged do
3:   for  $i = 1, \dots, L$  do
4:     Draw samples from  $q(\alpha_i, \beta_i)$  using eq. (4.72) and the current value of  $\hat{\lambda}_i$ 
5:     Update  $\hat{\lambda}_i$  using eq. (4.75)
6:   end for
7: end while

```

4.3.3 Extension to the Incomplete Data Scenario

In principle, the above algorithm can be easily extended for the incomplete data scenario, i.e., if the measurements consist of sparse and noisy readouts Y_n of the Markov chain X at times t_n . Intuitively, this can be understood as adding another layer on top of the states \mathbf{x}^m in the hierarchical Bayesian model. In this case it turns out that the variational expressions from Section 4.3.2 also involve expectations with respect to smoothing distributions of the form $p(\mathbf{x}^m \mid y_1^m, \dots, y_N^m, a)$ when considering the m -th cell. As mentioned in the beginning of this chapter, their computation is a challenging task on its own and a variety of numerical and analytical approaches have been proposed (e.g., [5, 92, 137] or Section 4.2). Apart from that, the VBEM framework can be readily applied to the more complicated case of incomplete and noisy measurements.

4.3.4 Case Studies

We performed several simulation studies in order to demonstrate and evaluate the proposed method. For each of the case studies, we used the simple reaction network of eukaryotic gene expression illustrated in Figure 4.14a. Exemplary trajectories of such a model are shown in Figure 4.14b. The model comprises six reaction channels

with kinetic parameters c_1, \dots, c_6 , which are either homogenous or heterogeneous – depending on the particular case study. Unless otherwise specified, the mean values of the kinetic parameters are chosen according to Table 4.3.

Table 4.3: Mean values of the kinetic parameters.

Parameter	c_1	c_2	c_3	c_4	c_5	c_6
Mean (s^{-1})	0.5	0.05	0.1	0.001	0.03	0.008

We first analyzed convergence of the VBEM algorithm using the network from Figure 4.14a and assuming a heterogeneity over three out of the six parameters (i.e., c_3 , c_5 and c_6). The results from Figure 4.14c indicate that the algorithm is able to correctly identify the extrinsic noise parameters α_i and β_i in presence of heterogeneity. In case of the homogeneous reactions, both α_i and β_i diverge towards infinity, corresponding to a CV of zero and a finite mean of α_i/β_i . Furthermore, we find that in case of the heterogeneous reactions, only very few iterations are necessary until convergence is achieved.

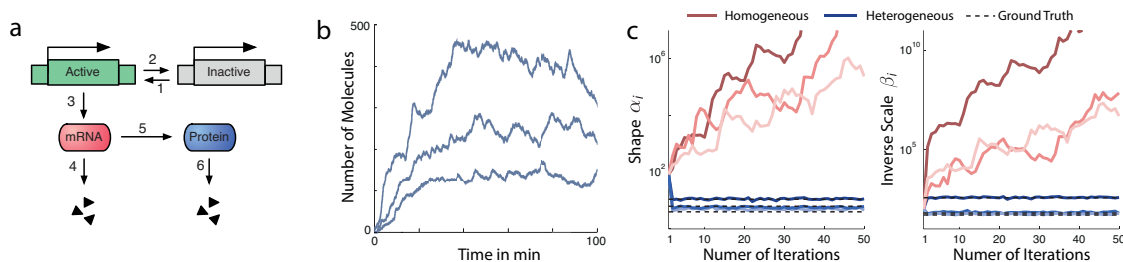


Figure 4.14: (a) A simple model of eukaryotic gene expression. The numbered arrows indicate chemical events taking place: upon activation of the gene (arrow 1), mRNA can be transcribed (arrow 3) which in turn gets translated (arrow 5) into protein. The remaining arrows indicate gene-deactivation (arrow 2) and degradation events (arrows 4 and 6). (b) Exemplary protein traces of a heterogeneous network. In this case, heterogeneity was simulated by introducing a Gamma-type variability in the translation rate. (c) Convergence of the VBEM algorithm. The algorithm was applied to $M = 30$ cell trajectories between zero and $200min$ with c_3 , c_5 and c_6 being heterogeneous with CVs 0.5, 0.3 and 0.4, respectively. The algorithm was ran for 50 update iterations. The curves correspond to expected values of the respective quantity (i.e., α_i , β_i).

Correct identification of the heterogeneous reactions depends on several parameters such as the population size M or the degree of intrinsic noise. In Figure 4.15a we analyze the detection robustness of a single reaction (i.e., the gene-activation event) as a function M . In particular, we computed the ratio of positive detections using 20 independent runs (see figure caption for fuller details). In accordance with our expectations, the results demonstrate that a robust detection of extrinsic variability is possible only if enough cells are in place (e.g. around $M > 100$ in this case).

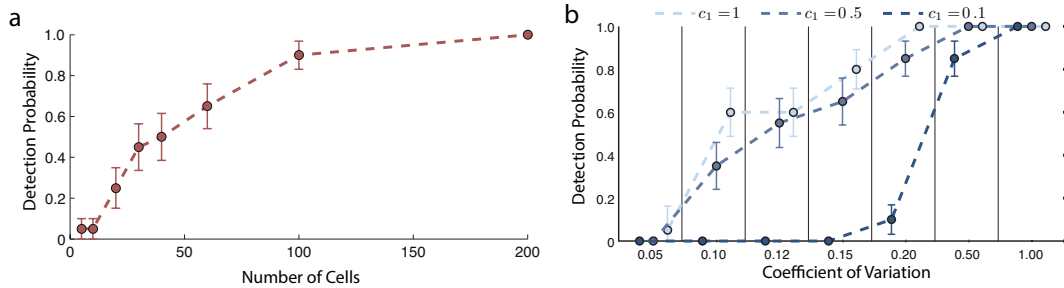


Figure 4.15: Detection robustness. (a) Detection robustness as a function of the population size. Probabilities for correct detections were computed for different population sizes (i.e., between 5 and 200 cells) using 20 independent runs. Circles denote mean values and whiskers indicate their standard errors (SEM). (b) Detection robustness as a function of intrinsic and extrinsic noise. We computed the rate of positive detection for different values of c_1 yielding different levels of intrinsic noise. For each c_1 we computed the detection robustness for several degrees of extrinsic variability (i.e., CVs between 0.05 and 1) using 20 independent runs. Circles denote mean values and whiskers indicate their standard errors (SEM).

Similarly, Figure 4.15b shows the probability of successful detection as a function of both *intrinsic* and *extrinsic* variability. Note that intrinsic noise of a reaction firing process scales inversely with its kinetic parameter. Again considering the gene-activation reaction, we computed the detection probabilities for three different values of c_1 (i.e., the intrinsic noise of the expression system) and several degrees of heterogeneity (see figure caption for further details). The parameters c_2 corresponding to the gene-deactivation event was adjusted such as to yield a constant ratio c_1/c_2 . We found that in presence of significant intrinsic noise and only moderate degrees of extrinsic noise, the algorithm facilitates the sparsity constraint and hence, yields negative results. In contrast, when decreasing the level of intrinsic noise, the algorithm is widely able to detect the heterogeneity (see Figure 4.15b).

4.4 Optimal Design of Temporal Perturbations

Up to now, the experimental design underlying a set of time-lapse recordings was treated as *given*. However, different experimental conditions can in fact yield completely different experimental outcomes and therefore, have a strong impact on the accuracy at which inference can be performed. For instance, if the transcriptional activation of a certain gene is recorded upon application of a step-like external perturbation, cells may adapt almost instantaneously such that only a few informative measurements can be obtained. Consequently, the information gain of such an experiment is comparably low, which is reflected in a low decrease in uncertainty between prior and posterior. In contrast, novel microfluidic techniques allow to synthesize sophisticated temporal perturbations, which can in turn increase the informativeness of the recorded data. In this section we briefly outline a mathematical framework to design optimal perturbations for the inference of stochastic reaction dynamics. To avoid the numerous technicalities that arise in the presence of extrinsic noise, the framework is illustrated by designing perturbations for only a *single* cell, assuming that those perturbations are close-to-optimal also in the presence of extrinsic noise. However, an extension to models including cellular heterogeneity is straightforward – for instance using the concepts from Section 4.2.

We consider a homogeneous CTMC X parameterized by a set of kinetic parameters C that we aim to estimate from experimental time-lapse data. Furthermore we assume that the reaction dynamics can be controlled by an exogenous perturbation $u \in \mathcal{U}$ on the acquisition interval $[0, T]$, giving rise to a conditional Markov chain $X \mid (C; u)$, whereas u is considered a deterministic function and is thus separated from the random quantities through a semicolon. In particular, we assume that the perturbation allows to modulate the hazard function of a particular reaction, e.g., $h_k(x(t), c_k, u(t)) = c_k u(t) g_k(x(t))$, where $u(t)$ denotes the value of the perturbation at time t . As in the previous chapters, we define sample paths of the Markov chain on the canonical interval $[0, T]$ as $\mathbf{X} = \{X(t) \mid t \in [0, T]\}$. A sequence of N measurements can be acquired at discrete time points t_1, \dots, t_N within $[0, T]$, which are assumed to be distributed according to

$$Y_l \mid (X(t_l; u) = x) \sim p(y \mid x), \quad (4.76)$$

with $p(y \mid x)$ as a known measurement density. The joint distribution over all model quantities is then given by

$$p(y_1, \dots, y_N, \mathbf{x}, c) = \left(\prod_{l=1}^N p(y_l \mid x_l) \right) p(\mathbf{x} \mid c; u) p(c). \quad (4.77)$$

The goal of the optimal input design is to choose u such as to generate maximally *informative* measurements. In the next section, we will discuss how informativeness can be expressed and formalized mathematically.

4.4.1 Choosing the Objective Function

We consider a measurement trajectory $Y_{1:N} = \{Y_1, \dots, Y_N\}$, which is recorded upon application of a perturbation u . A natural measure for the informativeness of an experiment is given by the expectation of the Kullback-Leibler divergence between the prior- and posterior distribution [13]

$$J(u) = \mathbb{E} [KL [p(c | y_{1:N}; u) || p(c)]], \quad (4.78)$$

where the expectation is calculated with respect to the marginal distribution of $Y_{1:N}$, i.e.,

$$p(y_{1:N}; u) = \int \left(\int \prod_{l=1}^N p(y_l | x_l) p(\mathbf{x} | c; u) d\mathbf{x} \right) p(c) dc, \quad (4.79)$$

where the prior $p(c)$ is assumed to be known, for instance from a previous experiment. Although it allows for an elegant information theoretic interpretation, the Kullback-Leibler divergence is often difficult to handle because of its intricate analytical form - such as in case of stochastic chemical kinetics. A more tractable objective function for experimental design purposes is the expected logarithm of the generalized posterior variance [22], i.e. the expected logarithm of the determinant of the variance-covariance matrix

$$\begin{aligned} J(u) &= \mathbb{E} [\log |\Sigma|] \\ &= \int \log |\Sigma| p(y_{1:N}; u) dy_{1:N} \\ &= \int \log |\mathbb{E} [cc^T | y_{1:N}; u] - \mu\mu^T| p(y_{1:N}; u) dy_{1:N} \\ &= \int \log \left| \int cc^T p(c | y_{1:N}; u) dc - \mu\mu^T \right| p(y_{1:N}; u) dy_{1:N} \\ &= \int \log \left| \int \left(\int cc^T p(c | \mathbf{x}) dc \right) p(\mathbf{x} | y_{1:N}; u) d\mathbf{x} - \mu\mu^T \right| p(y_{1:N}; u) dy_{1:N} \end{aligned} \quad (4.80)$$

with

$$\mu = \int \left(\int cp(c | \mathbf{x}) dc \right) p(\mathbf{x} | y_{1:N}; u) d\mathbf{x} \quad (4.81)$$

Recall that for a multivariate normal distribution the entropy is the logarithm of its generalized variance. Consequently, it can be shown that the minimizer u^* of (4.80) converges to the maximizer of (4.78), if $p(c | y_{1:N}, u)$ approaches a Gaussian distribution. Throughout this work, we chose equation (4.80) as the objective function.

Note that even though (4.80) exhibits simpler expressions than (4.78), it involves complicated expectations that have to be evaluated using Monte Carlo simulation. The most expensive expectation is the one taken with respect to $p(\mathbf{x} | y_{1:N}; u)$, since it relies on the solution of a smoothing problem. The numerical solution of the latter was already discussed in Section 4.2 and we will resort to the sampling algorithms

developed therein. As can be seen from (4.79), also the outer expectation from (4.80) involves a complicated integration with respect to the path \mathbf{x} , requiring a substantial amount of SSA runs. As a possible solution, certain simplifications can be made to reduce the problem complexity. For instance, the expectation over the sample paths and measurements could be “moved into the computation” of Σ , for instance to compute the generalized log-variance under the expected measurement. We want to point out that in this case, the resulting perturbation design cannot not account for any process and acquisition variability and consequently, does not provide a viable alternative. During our simulation studies from Section 4.4.5, we observed that that a reasonable compromise between efficiency and accuracy is achieved by moving only the expectation with respect to the observation noise into the calculation of Σ . Mathematically, this would correspond to approximate $J(u)$ as

$$J(u) \approx \int \log \left| \int \left(\int cc^T p(c | \mathbf{x}) dc \right) p(\mathbf{x} | \bar{y}_{1:N}; u) d\mathbf{x} - \bar{\mu} \bar{\mu}^T \right| p(\bar{\mathbf{x}}; u) d\bar{\mathbf{x}} \quad (4.82)$$

with

$$\bar{\mu} = \int \left(\int cp(c | \mathbf{x}) dc \right) p(\mathbf{x} | \bar{y}_{1:N}; u) d\mathbf{x} \quad (4.83)$$

and

$$\bar{y}_l = \int y_l p(y_l | \bar{x}_l) dy_l. \quad (4.84)$$

Note that the symbol $\bar{\mathbf{x}}$ is needed because a second integration over the path \mathbf{x} is used inside the logarithm. If the measurement density $p(y | x)$ has mean x (e.g., if the measurement noise is additive with mean zero), we further have that $\bar{y}_l = \bar{x}_l$.

4.4.2 The Variational Problem

Given an objective function $J(u)$, we define the optimization problem as

$$\begin{aligned} \min J(u) : \{u \in \mathcal{U}\} \\ \text{s.t. } X(t) = X(0) + \sum_{j=1}^L Y_j \left(\int_0^t h_j(X(s), c_j, u) ds \right) \nu_j, \end{aligned} \quad (4.85)$$

where the dynamic constraint in (4.85) is just a perturbation-dependent version of the random time-change model described in Section 2.1.2. Further, we restrict the perturbations to be positive and to fulfill an L^p constraint, i.e.,

$$\mathcal{U} = \{u \in L^p([0, T], \mathbb{R}) \mid u \geq 0 \wedge \|u\|_p = E\}. \quad (4.86)$$

Without further simplifying the variational problem (4.85) turns out to be intractable. Thus, we assume the perturbation to be a parameterized function, i.e., $u \equiv u(\theta)$ with $\theta \in \mathbb{R}^q$ as a set of q perturbation parameters. In particular, we assume $u(\theta)$ to be

an equally spaced, piece-wise constant function, with θ specifying the q perturbation levels, i.e.,

$$u(\theta, t) = \sum_{i=1}^q \theta_i \mathbb{1}_{\mathcal{T}_i}(t), \quad (4.87)$$

with $\mathcal{T}_i = \{t \in [0, T] \mid (i-1)\Delta \leq t < i\Delta\}$ and $\Delta = T/q$. Here we restrict our analysis to the case of $p = 1$, which - in conjunction with the positivity constraint - yields the following discrete optimization problem

$$\min J(\theta) : \{\theta \in \mathcal{G}\} \quad (4.88)$$

with $\mathcal{G} = \{\theta \in \mathbb{R}^q \mid \theta \geq 0 \wedge \|\theta\|_1 = E\Delta^{-1}\}$ as the feasible set. Note that for compactness, the dynamic constraint was omitted in (4.88).

4.4.3 Stochastic Approximation

In the following, we propose an efficient gradient-based algorithm for numerically minimizing $J(u)$ based on stochastic approximation [58, 65]. Although direct evaluation of $J(\theta)$ is impossible, it is straight forward to obtain noisy estimates $\widehat{J}(\theta)$ using Monte Carlo integration, such that we can compute the i -th component of the gradient of $J(\theta)$ as a one-sided finite difference

$$\widehat{\nabla}_i(J) = \frac{\widehat{J}(\theta + h_n e_i) - \widehat{J}(\theta)}{h_n}, \quad (4.89)$$

with $e_i \in \mathbb{R}^q$ as the i -th canonical base vector and $h_n \in \mathbb{R}$ as the discretization step size. The main idea of the constrained stochastic approximation algorithm is to iteratively update the perturbation parameters as

$$\theta^{n+1} = \mathcal{P} \left(\theta^n - \alpha_n \widehat{\nabla}(J) \right), \quad (4.90)$$

where the sequences α_n and h_n need to be chosen such that $\sum_{n=0}^{\infty} \alpha_n = \infty$, $\sum_{n=0}^{\infty} \alpha_n^2 / h_n^2 < \infty$, $\lim_{n \rightarrow \infty} \alpha_n = 0$ and $\lim_{n \rightarrow \infty} h_n = 0$ to ensure convergence [58]. For all simulations in Section 4.4.5, we choose

$$\alpha_n = \frac{a_0}{A + n^\rho} \text{ and } h_n = \frac{h_0}{n^\gamma},$$

whereas the individual parameters were tuned for each of the problems individually. Function \mathcal{P} projects θ back to the nearest point in the feasible region \mathcal{G} (by means of the L^2 -metric). In general, such a projection might be tedious to compute. Note however that in our particular case \mathcal{G} defines a canonical simplex in \mathbb{R}^q , for which the projection can be solved efficiently within a finite number of steps. In this work, we use the algorithm proposed in [78].

4.4.4 Fast Gradient Approximation Using Importance Sampling

Note that the one-sided gradient estimate rests upon $q+1$ Monte Carlo integrations over the path space, which might lead to slow convergence for large q or high-dimensional reaction dynamics. However - as demonstrated in [109] - the number of required SSA runs can be significantly reduced using importance sampling concepts. Let us now assume that a set of sample paths $\mathbf{x}^{(i)}$ for $i = 1, \dots, P$ has been simulated conditional on a particular θ to obtain an estimate $\widehat{J}(\theta)$. Then, instead of newly sampling paths $\mathbf{x}'^{(i)}$ for a new parameter set θ' (and corresponding perturbation $u(\theta')$), we can efficiently draw them from the mixture distribution

$$\mathbf{X} \mid (c; \theta') \sim \frac{1}{\sum_{i=1}^P w_i} \sum_{i=1}^P w_i \mathbb{1}_{\mathbf{x}^{(i)}}(\mathbf{x}) \quad (4.91)$$

$$\text{with } w_i = \frac{p(\mathbf{x}^{(i)} \mid c; u(\theta'))}{p(\mathbf{x}^{(i)} \mid c; u(\theta))},$$

where w_i is referred to as the *importance weight* of sample $\mathbf{x}^{(i)}$. The latter only requires an evaluation of the path likelihood function with respect to the parameters $u(\theta')$, which is typically faster than simulating a new path. For instance, if one assumes that $u(\theta)$ corresponds to the k -th rate constant, i.e., $c_k := u(\theta)$, the likelihood function is obtained by computing (2.17) from Section 2.1.2 for each time interval \mathcal{T}_i and multiplying together the individual terms.

Practically, one can sample from (4.91) by drawing an index i from the discrete distribution defined by the normalized important weights. Then, a valid path $\mathbf{x}^{(i)}$ is given by the path associated with index i , i.e., $\mathbf{x}^{(i)}$. In theory, the set of sample paths only needs to be simulated once in order to run the optimization. However - as with all importance sampling techniques - finite sample effects become more significant for larger deviations between θ and θ' . In this work, we simulate new sample paths to obtain $\widehat{J}(\theta)$ and then, use equation (4.91) to compute $\widehat{J}(\theta + h_n e_i)$ for all $i = 1, \dots, q$.

4.4.5 Case Studies

In the following, we perform simulation studies based on two simple reaction networks. For all the simulations, we assume prior knowledge over the rate constants, e.g., obtained from previous experiments. In particular, we assume a prior of the form

$$p(c) = \prod_{j \neq k} \Gamma(c_j; a_j, b_j)$$

with $a_j = 20c_j$, $b_j = 20$ and c_j as the true parameter. Furthermore - for simplicity - we assume initial conditions to be known. In all case studies, a simple step perturbation was used as a starting point for the numerical optimization. Estimates of the objective $\widehat{J}(u)$ were computed using 40 – 120 sample paths.

Table 4.4: Parameter configuration for the birth-death model.

Parameter	c_1	c_2	q	E
Value	$u(\theta, t)$	0.30	15	200
Unit	s^{-1}	s^{-1}	-	1

We studied the algorithm performance under the realistic setting of discrete-time and noisy measurement. We assume the measurements to be corrupted by additive Gaussian noise with zero-mean and standard deviation $\sigma_Y = 4$. Before we applied the algorithm to a more complicated, nonlinear reaction network, we studied the perturbation design for a simple birth-death process (see Figure 4.16a), for which the results are easier to interpret. We assume that the birth rate can be controlled by an external perturbation (i.e., $c_1 \equiv u(\theta, t)$), which is optimized such as to minimize the expected logarithm of the posterior variance of c_2 .

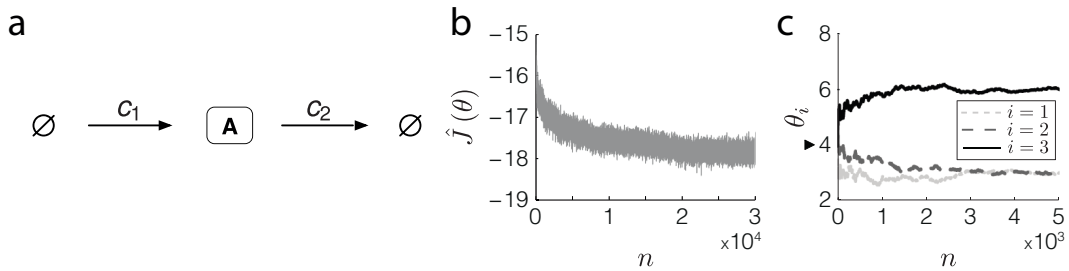


Figure 4.16: (a) A simple birth-death process. (b-c) Illustration of the stochastic approximation algorithm. (b) Exemplary minimization of $\hat{J}(\theta)$ over 30000 update iterations and (c) convergence of a three-dimensional perturbation to the optimum. The perturbation was initialized to a step function, i.e., $\theta_i = 4$ for $i = 1, 2, 3$, indicated by the black triangle.

Figures. 4.16b and 4.16c illustrate an exemplary minimization of $\hat{J}(\theta)$ over the number of update iterations for a three-level input profile applied to the birth-death process. The achieved decrease of the objective function corresponds to roughly two orders of magnitude of the posterior variance.

We then calculated optimal perturbations for the case of one (see Figure 4.17a) and two (see Figure 4.17b) measurement time points of species A, whose time evolution is denoted $A(t)$. Details on the parameter configuration used in the following simulations are summarized in Table 4.4. Interestingly, high perturbation amplitudes arise immediately before the measurement time points, which we interpret as follows: first - as true in the general - perturbations yielding measurements during a dynamic transient are preferable to measurements close to a stationary state. Second - in the particular case of the first order death reaction - strong excitation close before the acquisition time will accumulate many of the events at regions where they - conditional on that excitation - can be inferred or “located” more accurately. It can be seen from Figure 4.17 that while

the process mean is significantly increased, the standard deviation remains more or less unchanged. In contrast, when considering the step perturbation, most of the transient is missed during acquisition and furthermore, degradation events will be spread over a wider time interval. We also want to point out the significant difference between the perturbations obtained for the incomplete and complete case. For the latter, it was shown in [86] that the expected posterior variance is minimal for the case $\theta_1 = E\Delta^{-1}$ and $\theta_{i \neq 1} = 0$.

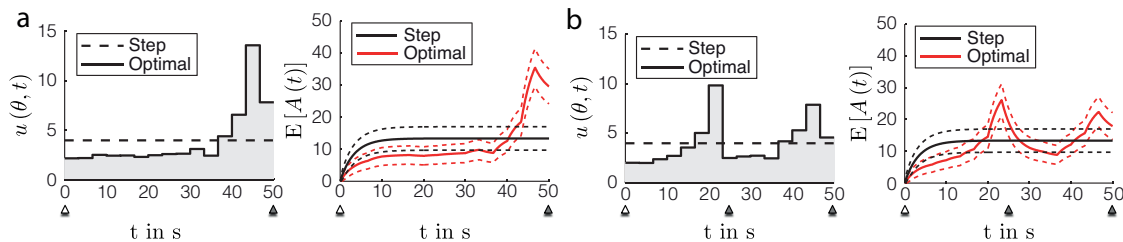


Figure 4.17: Optimal perturbations and mean process dynamics. Mean dynamics (solid) and the $\pm\sigma$ confidence bounds (dashed) for the step (black) and optimal (red) perturbation were computed by integrating the moment ODEs. The triangles indicate the initial (white) and observation (gray) time points used for perturbation design.

We repeated the two-observation experiment for a nonlinear model of transiently induced transcriptional activation. Often cells react to changing environmental conditions, by activating particular transcriptional programs (see e.g., [94]). Sensed at the cell membrane, the stimulus or stress is mediated to the nucleus by a translocation of certain transcription factors, which are activated by the signaling cascade. Once in the nucleus, the signaling proteins can initiate transcription of the target genes. After the cell has adapted, the transcription factors relocate to the cytoplasm, giving rise to only a short time period of gene activity. A minimalistic model of the transiently induced transcriptional activation is depicted in Figure 4.18a, whereas all reactions are modeled according to mass-action kinetics. We further assume that the intracellular dynamics can be perturbed by means of the rate $c_2 \equiv u(\theta, t)$ (see Table 4.5 for details). The initial abundances of A, B, AB and C are initialized at 0, 5, 0 and 0 copies, respectively. We assume that we can obtain noisy measurements of C at two time points with standard deviation $\sigma_Y = 4$. The optimal perturbation was computed for the case of jointly estimating c_1 (degradation) and c_5 (protein synthesis). The resulting perturbation as well as the mean process dynamics of species A and C are depicted in Figure 4.18b. Compared to the step response, the optimized perturbation results in a strong initial up-regulation of species A, followed by a period where it decreases again. Intuitively, it seems important to have a high transcription factor abundance during the early time points, such that (a) many degradation events have appeared at the measurement time points, and (b) maximize the temporal window of gene activity, such that many new proteins can be synthesized. This is supported by the increased mean of species C as shown in Figure 4.18b. However, we want to stress that such explanations are difficult

5 Uncoupled Analysis of Biochemical Networks in Random Environments

The results from the previous chapter give rise to the question whether the process marginalization can also be performed if the environmental conditions are not static, but dynamically changing [53, 110]. In that sense, it would be possible, to marginalize a stochastic model with respect to the fluctuations that are due to the surrounding components such as the ribosomal abundance and so forth. Intuitively speaking, we aim to find a proper dynamical description of just the system of interest as if it was still embedded into the whole environment. This would correspond to a model in which all the extrinsic fluctuations are “self-contained” meaning that it summarizes all system behaviors attainable under all possible realizations of the environmental noise. As a consequence, such models could be used to perform an *uncoupled analysis* of a reaction network subject to extrinsic noise. A historical toy example of such a self-contained model is given by the Polya urn scheme [30], which is known to be equivalent to a standard Bernoulli urn, marginalized over random (and correspondingly extrinsic) success rates [84]. In the former model, the player draws from an urn containing balls of two colors. After each draw, the player returns the ball together with a fixed number of new balls of the same color. Apparently, such scheme exhibits a self-excitation, meaning that the outcome of the player’s draw will impact his future chances. This stands in contrast with the standard Bernoulli model, where the number of balls of each color is randomly drawn at the beginning but held fixed throughout the subsequent trials. Nevertheless, looking at the number of successful draws over repeated trials, both models turn out to be equivalent. From a biophysical point of view, the Polya urn gives the desired uncoupled model since it provides a correct and self-contained description of only the number of wins.

In this chapter, we will follow up on the marginal process from Chapter 4 and demonstrate how it can be constructed and analyzed in the more general case of dynamically changing environments.

5.1 Mathematical Modeling

We again consider a continuous-time Markov chain (CTMC) X , which we assume to depend on another multivariate Markov process Z through its hazard functions in the form

$$h_i(x, z) = c_i(z)g_i(x), \quad (5.1)$$

with c_i some positive function and g_i a polynomial determined by the law of mass-action, for instance. For reactions independent of Z , we thus have $c_i(z) \equiv c_i$. Typically, Z is another jump or diffusion process corresponding to a set of modulating *environmental* species or conditions that are considered extrinsic to the system of interest, whereas the species in X represent the actual system of interest. For example, Z could be the fluctuating ribosome copy numbers affecting the kinetics of a gene regulatory network represented by X . Although a more general treatment is possible, we assume a feed-forward structure between Z and X , which means that Z modulates X but not vice-versa. Consequently, the dynamics of the joint system $Y(t) = (Z(t), X(t))$ can be described by a marginal Markov process Z together with a conditional Markov chain $X | Z$.

5.2 Construction of the Uncoupled Dynamics

Mathematical descriptions of the joint system $Y(t)$ are readily obtained using available techniques for modeling Markovian dynamics [53,59]. Practically – however – we prefer models that allow to properly describe *only* the interesting components $X(t)$. As a result, we obtain a jump process which - in contrast to the conditional process $X | Z$ - no longer depends on the environmental species in Z . We remark that a straightforward marginalization of the joint master equation of Z and X generally leads to intractable propensities [53,98]. Fortunately, it turns out that the innovation theorem [1,3,4] readily applies to fluctuating environments such that the hazard functions of the uncoupled process are formally given by

$$h_i(\mathbf{x}_t, t) = \mathbb{E} [c_i(Z(t)) | \mathbf{x}_t] g_i(X(t)), \quad (5.2)$$

where the expectation is taken with respect to the conditional distribution $\pi(z, t | \mathbf{x}_t)$ ¹. The latter describes the conditional probability of the environmental process $Z(t)$ given the entire history of process X until time t . Using the expected value of that distribution, the feed-forward influence of Z on the hazard functions of X can be replaced by a deterministic function of X , which no longer depends on the actual state of Z . Instead, the marginal process X becomes *self-exciting*, meaning that it exerts a feedback on itself. Hence, given that we can evaluate eq.(5.2), we have a means to simulate X while bypassing the need to draw realizations of Z . This has for instance been exploited for the exact simulation of diffusion-driven Poisson processes [43]. As has been observed in Section 4.2, the uncoupled process X is no longer Markovian, since the conditional expectation - and hence the hazard functions - depend on the full process history \mathbf{X}_t . A schematic illustration of that uncoupling is given in Figure 5.1.

¹Note that π is either a probability or a density, depending on whether Z is discrete- or real-valued, respectively.

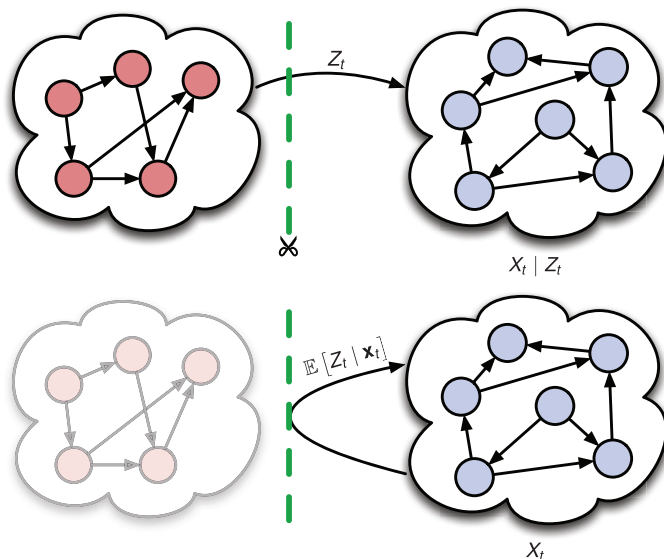


Figure 5.1: Uncoupled stochastic dynamics. The environmental process Z modulates the dynamics of the process under study X , e.g., through one of its hazard functions. Marginalization with respect to Z yields the uncoupled dynamics of X , whereas the original dependency on the environment Z is replaced by its optimal estimator given the history of X . Consequently, the marginal process X is self-exciting, i.e., it exerts a feedback on itself.

5.2.1 Relation to Stochastic Filtering

Although the construction of the uncoupled dynamics is general, any practical implementation thereof will depend on an explicit computation of the conditional expectation in (5.2). This expectation estimates the environmental state $Z(t)$ given the full history of the uncoupled process \mathbf{X}_t and therefore, can be understood as the solution to a stochastic filtering problem [3]. As discussed in Section 2.2.2, filtering techniques deal with optimally reconstructing a hidden stochastic process at time t from noisy observations of that process up to time t . In the situation considered here, the hidden process corresponds to the environment $Z(t)$, which gets reconstructed from the “observed” history \mathbf{X}_t through the conditional mean in (5.2).

We assume that the environment $Z(t)$ admits a probability distribution $p(z, t)$ described by a Kolmogorov-forward equation of the form

$$\frac{\partial}{\partial t} p(z, t) = \mathcal{A} p(z, t), \quad (5.3)$$

where \mathcal{A} represents the temporal change of $p(z, t)$, i.e., is the infinitesimal generator of Z . For instance, if Z is a diffusion process, \mathcal{A} corresponds to the Fokker-Planck operator, while in case of a CTMC, \mathcal{A} is given by the difference operator of the chemical master equation (CME). In terms of filtering, (5.3) corresponds to the process model of Z . Furthermore, we know that at a given time t , the solution of X can be written as a sum of independent but time-transformed Poisson processes [7], each of them

corresponding to a particular reaction channel. Consequently, the observation model is given by a set of Poisson counting observations with the hazard functions given in (5.1). This is closely related to Markov-modulated Poisson processes [117] and their corresponding optimal filtering [32].

5.2.1.1 Derivation of the Normalized Filtering Distribution

The following derivation of the filtering distribution is based on [29] and is demonstrated using a discrete-valued environmental process $Z(t)$ (i.e., a CTMC). Furthermore, we assume that $Z(t)$ is a one-dimensional process and c_k a linear function of $Z(t)$, i.e., $c_k(z)g_k(x) = c_k z g_k(x)$. However, analogous derivations are possible for multivariate or real-valued $Z(t)$.

Let us assume now that we have given the posterior distribution over the environmental process at time t , i.e., $P(Z(t) = z \mid \mathbf{x}_t)$. Then, the one-step posterior for a sufficiently small Δt can be generally written as

$$\begin{aligned} P(Z(t + \Delta t) = z \mid \Delta R_k = \{0, 1\}, \mathbf{x}_t) \\ = \frac{P(\Delta R_k = \{0, 1\} \mid Z(t + \Delta t) = z)P(Z(t + \Delta t) = z \mid \mathbf{x}_t)}{\sum_{z=0}^{\infty} P(\Delta R_k = \{0, 1\} \mid Z(t + \Delta t) = z)P(Z(t + \Delta t) = z \mid \mathbf{x}_t)}, \end{aligned} \quad (5.4)$$

where ΔR_k indicates whether a reaction has happened in Δt or not. The likelihood term in (5.4) is given by the Poissonian observation model with propensity $h_k(x(t), z) = c_k z g_k(x(t))$, i.e.,

$$P(\Delta R_k = 0 \mid Z(t + \Delta t) = z) = e^{-h_k(x(t), z)\Delta t} \quad (5.5)$$

$$P(\Delta R_k = 1 \mid Z(t + \Delta t) = z) = \Delta t h_k(x(t), z) e^{-h_k(x(t), z)\Delta t}. \quad (5.6)$$

For small Δt , we can write the one-step prior distribution as

$$P(Z(t + \Delta t) = z \mid \mathbf{x}_t) \approx P(Z(t) = z \mid \mathbf{x}_t) + \Delta t \mathcal{A}P(Z(t) = z \mid \mathbf{x}_t). \quad (5.7)$$

Hence, in case no reaction happens, we obtain

$$\begin{aligned} P(Z(t + \Delta t) = z \mid \Delta R_k = 0, \mathbf{x}_t) \\ = \frac{e^{-c_k z g_k(x(t))\Delta t} [P(Z(t) = z \mid \mathbf{x}_t) + \Delta t \mathcal{A}P(Z(t) = z \mid \mathbf{x}_t)]}{\sum_{z=0}^{\infty} e^{-c_k z g_k(x(t))\Delta t} P(Z(t + \Delta t) = z \mid \mathbf{x}_t)} \\ = \frac{e^{-c_k z g_k(x(t))\Delta t} P(Z(t) = z \mid \mathbf{x}_t)}{\sum_{z=0}^{\infty} e^{-c_k z g_k(x(t))\Delta t} P(Z(t + \Delta t) = z \mid \mathbf{x}_t)} \\ + \frac{\Delta t \mathcal{A}P(Z(t) = z \mid \mathbf{x}_t)}{\sum_{z=0}^{\infty} e^{-c_k z g_k(x(t))\Delta t} P(Z(t + \Delta t) = z \mid \mathbf{x}_t)} \end{aligned} \quad (5.8)$$

Taking the limit yields the temporal change in P , i.e.,

$$\begin{aligned}
& \lim_{\Delta t \rightarrow 0} \frac{P(Z(t + \Delta t) = z \mid \Delta R_k = 0, \mathbf{x}_t) - P(Z(t) = z \mid \mathbf{x}_t)}{\Delta t} \\
&= \lim_{\Delta t \rightarrow 0} \frac{P(Z(t) = z \mid \mathbf{x}_t) \left[e^{-c_k z g_k(x(t)) \Delta t} - \sum_{z=0}^{\infty} e^{-c_k z g_k(x(t)) \Delta t} P(Z(t + \Delta t) = z \mid \mathbf{x}_t) \right]}{\Delta t \sum_{z=0}^{\infty} e^{-c_k z g_k(x(t)) \Delta t} P(Z(t + \Delta t) = z \mid \mathbf{x}_t)} \\
&+ \lim_{\Delta t \rightarrow 0} \frac{\mathcal{A} P(Z(t) = z \mid \mathbf{x}_t)}{\sum_{z=0}^{\infty} e^{-c_k z g_k(x(t)) \Delta t} P(Z(t + \Delta t) = z \mid \mathbf{x}_t)} \\
&= P(Z(t) = z \mid \mathbf{x}_t) [-c_k z g_k(x(t)) + c_k g_k(x(t)) \mathbb{E}[Z(t) \mid \mathbf{x}_t]] + \mathcal{A} P(Z(t) = z \mid \mathbf{x}_t) \\
&= \mathcal{A} P(Z(t) = z \mid \mathbf{x}_t) + c_k g_k(x(t)) [z - \mathbb{E}[Z(t) \mid \mathbf{x}_t]] P(Z(t) = z \mid \mathbf{x}_t).
\end{aligned} \tag{5.9}$$

Using a simpler notation, we can write the differential change in case no reaction happens as

$$\frac{d}{dt} P(Z(t) = z \mid \mathbf{x}_t) = \mathcal{A} P(Z(t) = z \mid \mathbf{x}_t) - c_k g_k(x(t)) [z - M_1(t)] P(Z(t) = z \mid \mathbf{x}_t), \tag{5.10}$$

with $M_1(t)$ as the posterior expectation of Z . If one reaction happens, we can write the one-step posterior distribution as

$$\begin{aligned}
& P(Z(t + \Delta t) = z \mid \Delta R_k = 1, \mathbf{x}_t) \\
&= \frac{c_k z g_k(x(t)) e^{-c_k z g_k(x(t)) \Delta t} P(Z(t) = z \mid \mathbf{x}_t)}{\sum_{z=0}^{\infty} c_k z g_k(x(t)) e^{-c_k z g_k(x(t)) \Delta t} P(Z(t + \Delta t) = z \mid \mathbf{x}_t)} \\
&+ \frac{\Delta t \mathcal{A} P(Z(t) = z \mid \mathbf{x}_t)}{\sum_{z=0}^{\infty} c_k z g_k(x(t)) e^{-c_k z g_k(x(t)) \Delta t} P(Z(t + \Delta t) = z \mid \mathbf{x}_t)}.
\end{aligned} \tag{5.11}$$

Since the posterior jumps instantaneously when the reaction happens, the derivatives are not defined. Instead, we compute the increments if Δt approaches zero. This yields

$$\begin{aligned}
& \lim_{\Delta t \rightarrow 0} P(Z(t + \Delta t) = z \mid \Delta R_k = 1, \mathbf{x}_t) - P(Z(t) = z \mid \mathbf{x}_t) \\
&= \lim_{\Delta t \rightarrow 0} \frac{c_k z g_k(x(t)) e^{-c_k z g_k(x(t)) \Delta t} P(Z(t) = z \mid \mathbf{x}_t)}{\sum_{z=0}^{\infty} c_k z g_k(x(t)) e^{-c_k z g_k(x(t)) \Delta t} P(Z(t + \Delta t) = z \mid \mathbf{x}_t)} - P(Z(t) = z \mid \mathbf{x}_t) \\
&= \lim_{\Delta t \rightarrow 0} \frac{z e^{-c_k z g_k(x(t)) \Delta t} P(Z(t) = z \mid \mathbf{x}_t)}{\sum_{z=0}^{\infty} z e^{-c_k z g_k(x(t)) \Delta t} P(Z(t + \Delta t) = z \mid \mathbf{x}_t)} - P(Z(t) = z \mid \mathbf{x}_t) \\
&= \left[\frac{z - M_1(t)}{M_1(t)} \right] P(Z(t) = z \mid \mathbf{x}_t)
\end{aligned} \tag{5.12}$$

Finally, denoting by $\pi(z, t) = P(Z(t) = z \mid \mathbf{x}_t)$ the normalized filtering distribution, we obtain

$$\begin{aligned}
& d\pi(z, t) \\
&= \left(\mathcal{A} - c_k g_k(x(t^-)) [z - M_1(t^-)] \right) \pi(z, t^-) dt + \left[\frac{z - M_1(t^-)}{M_1(t^-)} \right] \pi(z, t^-) dR_k(t),
\end{aligned} \tag{5.13}$$

with $dR_k(t)$ as the firing process for reaction type k . Note that this equation depends on the conditional mean of $Z(t)$, such that it becomes tedious to solve directly. Dropping the normalizing constant in (5.4) and repeating the same derivation as above, it is straightforward to show that the dynamics of the unnormalized filtering distribution follow

$$d\tilde{\pi}(z, t) = [\mathcal{A} - c_k g_k(x(t^-))z] \tilde{\pi}(z, t^-) dt + [z - 1] \tilde{\pi}(z, t^-) dR_k(t), \quad (5.14)$$

with $\pi(z, t) \equiv \xi(t)\tilde{\pi}(z, t)$ and $\xi(t)$ a time-dependent normalizing factor independent of z . Thus, (5.14) describes a scaled version of the exact filtering distribution. Once we have numerically solved for $\tilde{\pi}$, it can be easily rescaled such that it integrates (or sums up) to one for all t . Nevertheless, also the normalized distribution is important, since it permits a simple computation of the posterior moments of Z as we shall see later in this section. Note that (5.14) is a stochastic partial differential equation (SPDE) in case Z describes a diffusion process or a stochastic difference-differential equation (SDDE) if Z is a CTMC. In the latter case and if $g_k \equiv 1$, the solution of (5.14) can be compactly written as

$$\tilde{\Pi}(t) = e^{(Q - c_k \Lambda)(t - t_{R_k(t)})} \left(\prod_{l=1}^{R_k(t)} \Lambda e^{(Q - c_k \Lambda)(t_l - t_{l-1})} \right) \Pi_0 \quad (5.15)$$

with t_l as the l -th firing-time of $R(t)$, $\tilde{\Pi}(t) = (\tilde{\pi}(0, t), \dots, \tilde{\pi}(N-1, t))^T$, N the number of reachable states of Z , $\Lambda = \text{diag}(0, \dots, N-1)$, $\Pi_0 \in \mathbb{R}^N$ the initial distribution over Z and $Q \in \mathbb{R}^{N \times N}$ the generator matrix of Z^2 .

In order to evaluate (5.2), we only require the mean (i.e., the first moment) of the filtering distribution, i.e., $M_1(t) = \mathbb{E}[Z(t) | \mathbf{x}_t]$. In general, however, the mean also depends on the second-order moment, which in turn depends on the third-order moment and so forth. The i -th order non-central moment is computed by multiplying both sides of (5.13) with z^i and summing (or integrating) over all $z \in \mathcal{Z}$, i.e.,

$$\begin{aligned} \sum_{z \in \mathcal{Z}} z^i d\pi(z, t) &= \sum_{z \in \mathcal{Z}} z^i \left(\mathcal{A}\pi(z, t^-) - c_k g_k(x(t^-)) [z - M_1(t^-)] \pi(z, t^-) \right) dt \\ &+ \sum_{z \in \mathcal{Z}} z^i \left[\frac{z - M_1(t^-)}{M_1(t^-)} \right] \pi(z, t^-) dR_k(t) \\ &= [\mathcal{D}_i(t^-) - c_k g_k(x(t^-))(M_{i+1}(t^-) - M_1(t^-)M_i(t^-))] dt \\ &+ \frac{M_{i+1}(t^-) - M_1(t^-)M_i(t^-)}{M_1(t^-)} dR_k(t), \end{aligned} \quad (5.16)$$

²Note that we define Q to be a left stochastic matrix, i.e., its rows sum up to zero.

with $\mathcal{D}_i(t^-) = \sum_{z \in \mathcal{Z}} z^i \mathcal{A}\pi(z, t^-)$. Therefore, the moment system is given by

$$\begin{aligned}
dM_1(t) &= \left[\mathcal{D}_1(t^-) - c_k g_k(x(t^-)) \left(M_2(t^-) - M_1(t^-) M_1(t^-) \right) \right] dt \\
&\quad + \frac{M_2(t^-) - M_1(t^-) M_1(t^-)}{M_1(t^-)} dR_k(t) \\
&\quad \vdots \\
dM_i(t) &= \left[\mathcal{D}_i(t^-) - c_k g_k(x(t^-)) \left(M_{i+1}(t^-) - M_1(t^-) M_i(t^-) \right) \right] dt \\
&\quad + \frac{M_{i+1}(t^-) - M_1(t^-) M_i(t^-)}{M_1(t^-)} dR_k(t).
\end{aligned} \tag{5.17}$$

Although (5.17) is generally infinite-dimensional, there are several relevant scenarios, for which the moment dynamics are *closed*, i.e., only depend on higher-order moments up to a certain order. This is for instance the case, if $Z(t)$ is a Cox-Ingersoll-Ross process [66] or any finite state Markov chain. On the other hand, if the conditional process $Z(t) \mid \mathbf{x}_t$ cannot be fully characterized by a finite number of moments, one can employ suitable moment-closure approximations (see [52] and Chapter 3). Although such techniques may yield excellent approximation performances in certain practical scenarios, there is no principled way of assessing the quality of a particular closure “beforehand”, i.e., without performing extensive stochastic simulations. Those problems are inherent if one aims to approximate $Z(t)$ without including further information. In contrast, the conditional process $Z(t) \mid \mathbf{x}_t$ turns out to be easier to approximate. This stems from the fact that due to the conditioning, the distribution over $Z(t)$ will generally be more informative than the unconditional distribution. This can be understood via Bayes’ theorem: a complicated and broad prior distribution is significantly harder to approximate by a simple distribution (e.g., a Gaussian) than a posterior distribution that is obtained after observing data. The more data (i.e., information) is included, the tighter and symmetric it is. While such arguments appear largely qualitative, they can be rigorously formulated using concepts from asymptotic theory such as large sample properties of Bayesian estimators [39]. In fact, our simulations indicated that the approximation accuracy of the uncoupled dynamics often shows little sensitivity with respect to the particular closure function. Intuitively, this can be understood by analyzing (5.15): especially for large c_k , the solution is predominantly driven by the term $e^{-c_k \Lambda \int_0^t g_k(x(s)) ds} \Lambda^{R_k(t)}$, suggesting that it can be well approximated by a Gamma-distribution. We note that the Gamma-distribution is fully characterized by two parameters – or equivalently – its first two moments $M_1(t)$ and $M_2(t)$. As a consequence, we may express the third order moment as a function of the first two moments, i.e., $M_3(t) = -M_1(t)M_2(t) + 2M_2^2(t)/M_1(t)$, such that the second conditional

moment closes as

$$\begin{aligned} dM_2(t) = & \left[\mathcal{D}_2(t^-) - 2c_k g_k(x(t^-)) \frac{M_2(t^-)}{M_1(t^-)} (M_2(t^-) - M_1^2(t^-)) \right] dt \\ & + 2 \left[\frac{M_2^2(t^-)}{M_1^2(t^-)} - M_2(t^-) \right] dR_k(t). \end{aligned} \quad (5.18)$$

For univariate environments, we consistently used this type of closure. For the multivariate case, we applied the second-order zero-cumulants closure [52] in which the third order moments are approximated by the first- and second-order moments as

$$\mathbb{E}[ABC] = \mathbb{E}[A] \mathbb{E}[BC] + \mathbb{E}[B] \mathbb{E}[AC] + \mathbb{E}[C] \mathbb{E}[AB] - 2\mathbb{E}[A] \mathbb{E}[B] \mathbb{E}[C]. \quad (5.19)$$

The numerical solution of the filtering problem is illustrated in Figure 5.2. The plot from Figure 5.2a illustrates that the conditional expectation has discontinuities only at the firing times of reaction k , whereas the original environment $Z(t)$ fluctuates significantly faster. However, when considered as hazard functions of the k -th reaction, both yield the equivalent dynamics in $X(t)$. Note that when simulating the marginal process, the true environment $Z(t)$ is not revealed. Hence, for the sake of this study, we simulated the joint model $(Z(t), X(t))$ and subsequently applied the filter on the realization of $X(t)$. Figure 5.2b shows a comparison between the exact solution of the filtering distribution and the one obtained under the Gamma-closure from eq. (5.18).

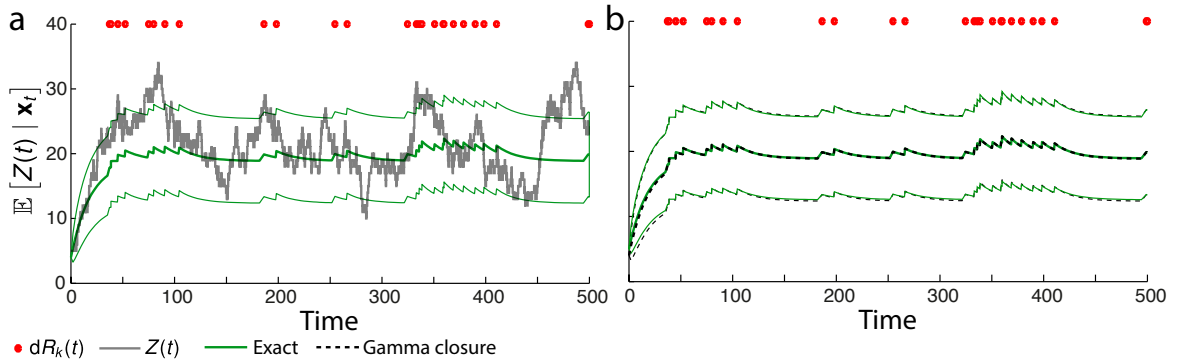


Figure 5.2: Numerical solution of the filtering problem. The environment $Z(t)$ was modeled as a birth-death process which linearly modulates the rate of the k -th reaction of X . (a) Comparison between the true environment $Z(t)$ and the corresponding conditional expectation. In order to compute the mean of the filtering distribution, we used eq. (5.15) in conjunction with a state space truncation. Since the latter was chosen to be conservative, the approximation error is negligible and the method considered practically exact. The red dots indicate the firing times of the modulated reaction k . (b) Comparison between the exact filter (green) and approximate filter (black) based on the Gamma assumption, i.e., using eq. (5.18).

5.2.2 Modified Marginal Simulation Algorithm

Although the uncoupled dynamics of X are non-Markovian, the Markov property can be enforced by virtually extending the state space by the filtering moments, which

summarize the history of X (see Section 4.2.2 and [137]). In particular, the time W_k that passes until a reaction of type k happens is distributed according to

$$P(W_k < w \mid \mathbf{x}_t) = 1 - e^{-\int_0^w h_k(\mathbf{x}_t, T) dT}, \quad (5.20)$$

with $h(\mathbf{x}_t, w) = c_i \mathbb{E}[Z(t+w) \mid \mathbf{x}_{t,w}]$, where $\mathbf{x}_{t,w}$ extends \mathbf{x}_t by a time-interval w assuming that no reaction of type k happens in w . The corresponding density is given by

$$p_k(w \mid \mathbf{x}_t) = h_k(\mathbf{x}_t, w) e^{-\int_0^w h_k(\mathbf{x}_t, T) dT}. \quad (5.21)$$

We note that as long as no reaction of type k happens, $dR_k(t)$ is zero and hence, $c_i \mathbb{E}[Z(t+w) \mid \mathbf{x}_{t,w}]$ is found by solving the ordinary differential equation from (5.17). Since that solution is not generally known in closed form, we cannot directly sample from (5.21). However, several efficient solutions to that problem have been developed in the context of inhomogeneous Poisson processes, e.g., such as the method of *thinning* [68]. In order to simulate from (5.21), we use the method of thinning as given in Algorithm 5.

Algorithm 5 (Method of thinning). *At a given time t , the algorithm requires the conditional mean $M_1(t)$ and a real-valued constant $\hat{\lambda} \geq M_1(w)g_k(x(w))$.*

-
- 1: Initialize $W \leftarrow 0$ and $u \leftarrow \infty$.
 - 2: **while** $u > M_1(t+W)g_k(x(t+W))/\hat{\lambda}$ **do**
 - 3: Simulate $\hat{W} \sim \text{Exp}(\hat{\lambda})$.
 - 4: Set $W \leftarrow W + \hat{W}$.
 - 5: Simulate $u \sim \mathcal{U}(0, 1)$.
 - 6: **end while**
 - 7: Output W .
-

Note that the tuning parameter $\hat{\lambda}$ has to be chosen such that $\hat{\lambda} \geq M_1(w)g_k(\mathbf{x}(w))$ for all $w \in [0, T]$. Once a reaction has fired, the filter moments need to be updated by the terms multiplying the firing process $dR_k(t)$ in (5.17) (i.e., they exhibit a discontinuity).

Evidently, simulation from (5.20) comes at higher cost than simulating from an exponential distribution (e.g., such as performed in standard SSA algorithms), since in general, it relies on a numerical integration of an ODE. However, reactions associated with the environmental part no longer need to be simulated, which yields a significant reduction in computational effort as soon as the environmental network is large and expensive to simulate.

5.2.3 Marginal Moment Dynamics

We want to highlight that the moments of the marginal dynamics precisely coincide with those derived from the marginal CME in Section 3.2.1. This is in fact a direct

consequence of the marginal CME being a lower-dimensional projection of the full path density associated with the marginal process. Although a general proof is straightforward, we prefer to illustrate this on a tractable and well-understood example, i.e., the Poisson process with a Gamma-distributed rate (see Example 10).

Example 10 (Marginal moment dynamics). *We consider a conditional Poisson process $X | Z$, where Z corresponds to a randomly drawn rate constant $Z \sim \mathcal{G}(\alpha, \beta)$. We know that the conditional distribution over $X(t)$ is given by a Poisson distribution, i.e.,*

$$P(x, t | z) = P(X(t) = x | X(0) = 0, Z = z) = \frac{(zt)^x}{x!} e^{-zt}. \quad (5.22)$$

Multiplying with $p(z)$ and integrating over z , we obtain for the marginal distribution over $X(t)$

$$\begin{aligned} P(x, t) &= \int_0^\infty \frac{(zt)^x}{x!} e^{-zt} \frac{\beta^\alpha}{\Gamma(\alpha)} z^{\alpha-1} e^{-\beta z} dz \\ &= \mathcal{NB} \left(x; \alpha, \frac{\beta}{\beta + t} \right), \end{aligned} \quad (5.23)$$

i.e., a negative binomial distribution. The first two moments of this distribution are known to be

$$\mathbb{E}[X(t)] = \frac{\alpha t}{\beta} \quad (5.24)$$

$$\mathbb{E}[X^2(t)] = \frac{\alpha t(\beta + t + \alpha t)}{\beta^2}. \quad (5.25)$$

If we use formula (3.7) from Section 3.2.1, we obtain a three-dimensional ODE system for computing the marginal moments, i.e.,

$$\begin{aligned} \frac{d}{dt} \mathbb{E}[X(t)] &= \mathbb{E}[Z] \\ \frac{d}{dt} \mathbb{E}[X^2(t)] &= \mathbb{E}[Z] + 2\mathbb{E}[ZX(t)] \\ \frac{d}{dt} \mathbb{E}[ZX(t)] &= \mathbb{E}[Z^2]. \end{aligned}$$

With initial conditions $\mathbb{E}[X(0)] = \mathbb{E}[X^2(0)] = \mathbb{E}[ZX(0)] = 0$, the solution of the moments in $X(t)$ are given by

$$\mathbb{E}[X(t)] = \mathbb{E}[Z] t \quad (5.26)$$

$$\mathbb{E}[X^2(t)] = \mathbb{E}[Z] t + \mathbb{E}[Z^2] t^2, \quad (5.27)$$

which with $\mathbb{E}[Z] = \alpha/\beta$ and $\mathbb{E}[Z^2] = \alpha(\alpha + 1)/\beta$ coincide with (5.24) and (5.25).

The marginal process dynamics can be written in SDE form as

$$dX(t) = M_1(t^-)dt + dQ(t) \quad (5.28)$$

with $M_1(t^-) = \frac{\alpha + X(t^-)}{\beta + t}$ and $Q(t)$ as a martingale. Taking the expectation yields

$$\frac{d}{dt} \mathbb{E}[X(t)] = \frac{\alpha}{\beta + t} + \frac{1}{\beta + t} \mathbb{E}[X(t)]. \quad (5.29)$$

For computing the second order moment, we make use of Ito's formula and find that

$$dX^2(t) = (2X(t^-) + 1)dX(t) = 2X(t^-)M_1(t^-) + M_1(t^-) + (2X(t^-) + 1)dQ(t) \quad (5.30)$$

and therefore,

$$\frac{d}{dt} \mathbb{E}[X^2(t)] = \frac{2\alpha}{\beta + t} \mathbb{E}[X(t)] + \frac{2}{\beta + t} \mathbb{E}[X^2(t)] + \frac{\alpha}{\beta + t} + \frac{1}{\beta + t} \mathbb{E}[X(t)]. \quad (5.31)$$

Interestingly, the first and second order moments can now be described using only two ODEs, whereas the latter show an explicit time dependency. However, the solution of those equations again yields the correct expressions.

5.3 Generalized Master Equations

Since the uncoupled process is generally non-Markovian, it does not satisfy a conventional master equation. Nevertheless, it can be described by a non-Markovian modification thereof, giving rise to a *generalized* master equation (GME). Typically, a GME is given in the form of an integro-differential equation, where the integral part stems from a time-convolution representing the *memory effects* of the system³. Alternatively, such master equations may be transformed into a time-convolutionless form [112], which are typically easier to handle analytically and numerically. While GMEs are barely used in the context of biology, they are frequently applied in the field of quantum- and statistical mechanics [112, 141]. In the following we will provide a brief discussion on how the uncoupled dynamics relate to the GME formalism.

An important property of a (possibly non-Markovian) jump process are the waiting-time distributions associated with each type of reaction (or transition) that may take place at a given time and molecular configuration. We assume that the k -th reaction is modulated by the environmental network Z . Accordingly, the k -th reaction will be associated with a non-exponential waiting-time distribution given in (5.21). For simplicity, the following considerations will be restricted to the case where the waiting-time distribution does not depend on the full process history \mathbf{X}_t but only on the current state $X(t)$ and time t , i.e.,

$$p_k(w | \mathbf{x}_t) = p_k(w | x, t) = h_k(x, t) e^{-\int_0^w h_k(x, t+s) ds}.$$

³In contrast, Markovian dynamics are known to be memoryless.

The remaining reaction channels $i \neq k$ are associated with exponential waiting-time distributions of the form $p_i(w | x) = h_i(x, c_i) \exp(-h_i(x, c_i)w)$. For instance, this situation arises if a static environmental network modulates X through a zero-order reaction (see e.g., Section 5.4.2). We remark that a more general analysis is not in the scope of this thesis but shall be subject to future research.

Following [45] or [57], we can use the network's waiting-time distributions to formulate a generalization of the Chapman-Kolmogorov equation for jump processes with non-exponential waiting-times. The probability of finding the system in a particular state x at time t is determined by considering two cases. If $x \neq X(0) = x_0$, we know that we have reached x from another state $x - \nu_i$ via a reaction of type i . The corresponding probability is finally obtained by summing over all possible reactions that may have moved the system to a state x and all possible times at which the reaction may have fired. The latter will be reflected by the aforementioned time-integral. A special case arises if $x = x_0$, since we need an additional term accounting for the case where the initial state x_0 has never been left until t . More specifically, it can be shown that $P(x, t) = P(X(t) = x | X(0) = x_0)$ satisfies

$$P(x, t) = S(t | x, 0) \mathbb{1}_{x_0}(x) + \sum_{i=1}^L \int_0^t P(x - \nu_i, t - w) Q_i(x - \nu_i, t - w) h(x - \nu_i, t - w) S(w | x, t - w) dw, \quad (5.32)$$

with $h(x, t) = h_k(x, t) + \sum_{l \neq k} h_l(x, c_l)$ as the total hazard,

$$Q_i(x, t) = \begin{cases} \frac{h_k(x, t)}{h(x, t)} & i = k \\ \frac{h_i(x, c_i)}{h(x, t)} & i \neq k. \end{cases} \quad (5.33)$$

as the probability next reaction will be the i -th one and $S(w | x, t)$ the probability that the system remains in x between t and $t + w$, i.e.,

$$S(w | x, t) = e^{-\int_0^w h(x, t+s) ds}. \quad (5.34)$$

[45, 57, 133]. The latter gives rise to the probability that *any* reaction happens between t and $t + w$, i.e., $P(W < w | X(t) = x) = 1 - S(w | x, t)$ with density $p(w | x, t) = h(x, t + w) S(w | x, t)$.

If $p(w | x, t) = p(w | x)$ (and hence, $Q_i(x, t) = Q_i(x)$), the differential form of eq. (5.32) is known to be

$$\frac{d}{dt} P(x, t) = \sum_{i=1}^L \int_0^t \left[P(x - \nu_i, t - w) Q_i(x - \nu_i) \phi(w | x - \nu_i) - P(x, t - w) Q_i(x) \phi(w | x) \right] dw, \quad (5.35)$$

where the *memory function* $\phi(w | x)$ is related to $p(w | x)$ through the *Montroll-Weiss* equation [80]

$$\phi(u | x) = \frac{up(u | x)}{1 - p(u | x)}, \quad (5.36)$$

where $\phi(u | x)$ and $p(u | x)$ denote the Laplace transforms of $p(w | x)$ and $\phi(w | x)$, respectively. From this relation, it is straightforward to verify that the master equation becomes memory-less in case of exponential waiting-time distributions (i.e., the memory function is given by a dirac-delta function) [45, 57]. In the more general case of time-dependent waiting-time distributions, a corresponding differential equation could for instance be obtained through direct differentiation of (5.32) [91, 133]. However, since the type and time of the next reaction only depend on the current state x and time t , a differential master equation is readily obtained by considering the uncoupled dynamics X as a time-inhomogeneous Markov chain. Similar to Section 2.1.1 it is then sufficient to characterize the probability of *moving* and *staying* within a small time interval between t and $t + \Delta t$. Those probabilities are again obtained through the respective waiting-time distributions, i.e.,

$$\underbrace{P(X(t + \Delta t) = x + \nu_i | X(t) = x)}_{\text{Probability of moving}} = P(W_i < \Delta t | X(t) = x) + f_i(\Delta t) \quad (5.37)$$

$$\underbrace{P(X(t + \Delta t) = x | X(t) = x)}_{\text{Probability of staying}} = 1 - \sum_{i=1}^L P(W_i < \Delta t | X(t) = x) - f_i(\Delta t),$$

where $f_i(\Delta t)$ corresponds to the probability that the state $x + \nu_i$ is reached through multiple reactions. However, for decreasing Δt , $f_i(\Delta t)$ goes to zeros much faster than $P(W_i < \Delta t | X(t) = x)$. The probability of being in a certain state at time $t + \Delta t$ is given by

$$P(x, t + \Delta t) = \underbrace{\sum_{i=1}^L \left(P(W_i < \Delta t | X(t) = x - \nu_i) + f_i(\Delta t) \right) P(x - \nu_i, t)}_{\text{Probability of moving in } \Delta t} \quad (5.38)$$

$$+ \underbrace{\left(1 - \sum_{i=1}^L P(W_i < \Delta t | X(t) = x) - f_i(\Delta t) \right) P(x, t)}_{\text{Probability of staying in } x \text{ in } \Delta t}.$$

The temporal change of $P(x, t)$ is then found to be

$$\begin{aligned} \frac{d}{dt}P(x, t) &= \lim_{\Delta t \rightarrow 0} \frac{P(x, t + \Delta t) - P(x, t)}{\Delta t} \\ &= \sum_{i=1}^N \lim_{\Delta t \rightarrow 0} \frac{P(W_i < \Delta t \mid X(t) = x - \nu_i)}{\Delta t} P(x - \nu_i, t) \\ &\quad - \sum_{i=1}^N \lim_{\Delta t \rightarrow 0} \frac{P(W_i < \Delta t \mid X(t) = x)}{\Delta t} P(x, t). \end{aligned} \quad (5.39)$$

The limit terms in (5.39) precisely coincide with the definition of a (possibly time-dependent) hazard function [4] and therefore,

$$\begin{aligned} \frac{d}{dt}P(x, t) &= \sum_{i \neq k} h_i(x - \nu_i, c_i) P(x - \nu_i, t) - \sum_{i \neq k} h_i(x, c_i) P(x, t) \\ &\quad + h_k(x - \nu_k, t) P(x - \nu_k, t) - h_k(x, t) P(x, t). \end{aligned} \quad (5.40)$$

Eq. (5.40) corresponds to the aforementioned time-convolutionless type of master equation.

5.4 Fluctuations on Different Timescales

The impact of environmental fluctuations on a dynamical system of interest is as diverse as the timescale on which they operate. For instance, extrinsic noise in the context of gene expression might be slowly varying (e.g., correlates well with the cell-cycle [104, 127]), while fluctuations in transcription factor abundance might be significantly faster than the expression kinetics downstream. From a technical point of view, timescales range from constant environmental conditions that are random but fixed [137] to regimes where the fluctuations are very fast, such that quasi-steady-state (QSS) assumptions become applicable [98]. A QSS-based approach for simulating a system X in the presence of extrinsic noise Z corresponds to simulating the conditional CTMC $X \mid Z$, where Z is replaced by the mean of Z . The simulation of the joint system (X, Z) become prohibitive if extrinsic fluctuations are fast, while with (5.17) the complexity of the marginal process simulation is invariant with respect to the time-scale of the environment. Alternatively, one may try to replace a fluctuating environment Z through a random but fixed environment of same variance but this leads to an overestimation of the process variance in X [53]. To investigate the two above simplifying assumptions and compare them to the exact solution obtained via SSA and via the marginal process, we performed a simulation study on a linear three-stage birth-death model given in Figure 5.3a, where only species C is considered of interest in this case. Accordingly, the uncoupled dynamics of C are obtained by integrating the dynamics over the A and B . The results are shown in Figure 5.3b and Figure 5.3c.

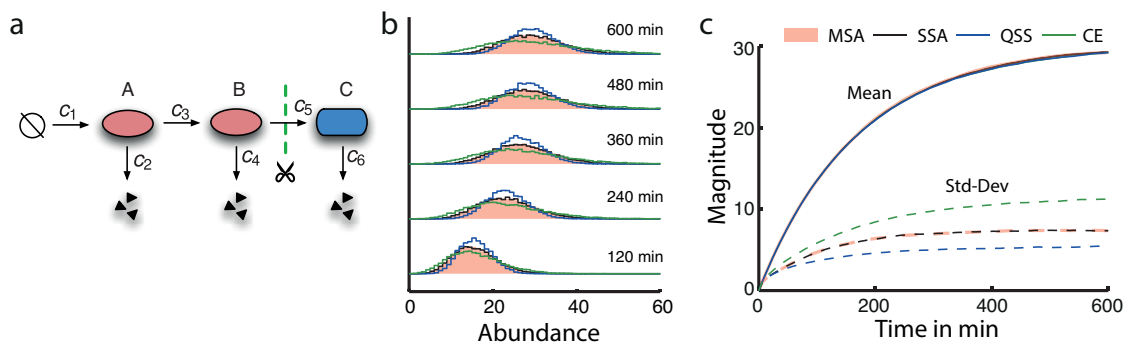


Figure 5.3: Marginal simulation algorithm. (a) Simple three-stage model. Species A, B and C are modeled as coupled linear birth-death processes, where the coupling is realized by linearly modulating the birth rates of B and C (rate constants $c_1 = 0.003$, $c_2 = 0.001$, $c_3 = 0.05$, $c_4 = 5e-4$, $c_5 = 1e-5$, $c_6 = 1e-4$). The uncoupled marginal dynamics of C are obtained by integrating over fluctuations of species A and B (7000 sample paths were used). (b, c) Evaluation of the marginal simulation algorithm. Simulations based on the QSS-approximation neglect a significant portion of variability as opposed to assuming a constant environment (CE) in which case the variability is overestimated. In contrast, the uncoupled dynamics correctly predict the fluctuations on the protein level, while yielding a reduction in computational effort when compared to standard SSA (20min simulation time instead of 46min); correspondingly higher speedup can be achieved for a larger time-scale separation of processes (A,B) versus C.

5.4.1 The Effective Noise

Several recent studies [17, 53, 99, 120] are centered around the separation of different noise contributions in biochemical networks. Typically, the law of total variance is employed to decompose the fluctuations of $X(t)$ into parts that are intrinsic to X and parts that come from Z (i.e., are extrinsic to X). Here we found that performing such an analysis on Z instead of X – in conjunction with our decoupling approach – provides a novel way to study how stochasticity is propagated through biochemical networks. Using the law of total variance, we can decompose the total (or unconditional) variance of $Z(t)$ as

$$\text{Var}[Z(t)] = \mathbb{E}[\text{Var}[Z(t) | \mathbf{x}_t]] + \text{Var}[\mathbb{E}[Z(t) | \mathbf{x}_t]]. \quad (5.41)$$

The two terms on the r.h.s. can be interpreted as follows. Assume we can observe Z only through X . Since X is intrinsically stochastic, a part of the variability of Z is not carried over to X . In (5.41), this part (i.e., the *suppressed noise*) corresponds to the first term on the r.h.s. since it quantifies the uncertainty about $Z(t)$ that remains after observing \mathbf{X}_t . The second term determines how accurate Z can be reconstructed from trajectories of \mathbf{X} . Alternatively, it can be understood as the amount of noise in Z that effectively impacts X (i.e., the *effective noise*). For instance, the environmental process could be characterized by a large variance, but still have only marginal impact on $X(t)$ – depending on the timescale of Z and X .

In order to quantify those terms, we note that the conditional variance within in the first term precisely coincides with the second-order central moment of the filtering distribution from (5.14). This further implies that it can be computed “on-the-fly”

when simulating $X(t)$ using the marginal simulation algorithm which allows an efficient estimation of its expectation. However, in some biologically relevant cases, the effective noise can be determined even analytically, which we demonstrate in the following.

While a more general treatment might be possible, we assume in the following that X is modulated by a one-dimensional process Z through a single zero-order reaction. In order to compute the suppressed noise, we rewrite the conditional moments in terms of *central* instead of non-central moments. In particular, we obtain for the mean and variance

$$\begin{aligned} dM_1(t) &= \left(\mathcal{D}_1(t^-) - c_k S_2(t^-) \right) dt + \frac{S_2(t^-)}{M_1(t^-)} dR_k(t) \\ dS_2(t) &= \left(\tilde{\mathcal{D}}_2(t^-) - c_k S_3(t^-) \right) dt + \left[\frac{S_3(t^-)}{M_1(t^-)} - \frac{S_2^2(t^-)}{M_1^2(t^-)} \right] dR_k(t), \end{aligned} \quad (5.42)$$

with $S_2(t)$ as the conditional variance $\text{Var}[Z(t) | \mathbf{x}_t]$ and $\tilde{\mathcal{D}}_2(t)$ as the unconditional central moment dynamics of order two. We next need to compute the expected value of $S_2(t)$. Decomposing $dR_k(t)$ into a predictable part and a martingale, i.e., $dR_k(t) = c_k M_1(t^-) dt + dQ_k(t)$, we can rewrite (5.42) as

$$\begin{aligned} dM_1(t) &= \mathcal{D}_1(t^-) dt + \frac{S_2(t^-)}{M_1(t^-)} dQ_k(t) \\ dS_2(t^-) &= \left(\tilde{\mathcal{D}}_2(t^-) - c_k \frac{S_2^2(t^-)}{M_1(t^-)} \right) dt + \left[\frac{S_3(t^-)}{M_1(t^-)} - \frac{S_2^2(t^-)}{M_1^2(t^-)} \right] dQ_k(t) \end{aligned} \quad (5.43)$$

Taking the expectation of (5.43), all terms involving $dQ_k(t)$ become zero and we obtain

$$\begin{aligned} \frac{d}{dt} \mathbb{E}[M_1(t)] &= \mathbb{E}[\mathcal{D}_1(t)] \\ \frac{d}{dt} \mathbb{E}[S_2(t)] &= \mathbb{E}[\tilde{\mathcal{D}}_2(t)] - c_k \mathbb{E} \left[\frac{S_2^2(t)}{M_1(t)} \right]. \end{aligned} \quad (5.44)$$

Although (5.44) is fully general, it might often be hard to evaluate the expectation $\mathbb{E}[S_2^2(t)/M_1(t)]$ (and possibly further terms stemming from the prior dynamics). We realize that the mean in (5.44) is just the unconditional mean of $Z(t)$, while the derivative of the expected variance shows an additional negative term, causing it to be smaller than the unconditional variance.

5.4.1.1 Effective Noise of a Cox-Ingersoll-Ross Process

Let us consider the case where $Z(t)$ follows a Cox-Ingersoll-Ross (CIR) process governed by the SDE

$$dZ(t) = \theta(\mu - Z(t))dt + \sigma_Z \sqrt{Z(t)} dW_D(t), \quad (5.45)$$

with θ , μ and σ_Z as real process parameters and $W_D(t)$ as a standard Wiener process. The expected central moments are then governed by

$$\begin{aligned}\frac{d}{dt}\mathbb{E}[M_1(t)] &= \theta(\mu - \mathbb{E}[M_1(t)])dt \\ \frac{d}{dt}\mathbb{E}[S_2(t)] &= -2\theta\mathbb{E}[S_2(t)] + \sigma_Z^2\mathbb{E}[M_1(t)] - c_k\mathbb{E}\left[\frac{S_2^2(t)}{M_1(t)}\right].\end{aligned}\quad (5.46)$$

The only term that remains to be specified is the expectation $\mathbb{E}[S_2^2(t)/M_1(t)]$. Fortunately, it turns out that for the a Gamma-type conditional distribution, this expectation simplifies to $\mathbb{E}[S_2^2(t)/M_1(t)] = \mathbb{E}[S_2(t)]^2/\mathbb{E}[M_1(t)]$. A derivation of that fact can be performed using Ito's lemma. However, since it involves a multitude of technicalities that are not in the scope of this study, we skip the individual steps. Instead we provide a heuristic but substantially simpler explanation based on the fact that the CIR process is *conjugate* to the Poissonian reaction channel. In particular, we consider the case of a Gamma distributed random variable $Z \sim \mathcal{G}(\alpha, \beta)$, with α and β as shape- and inverse scale parameters. The random variable is observed through a Poissonian measurement $X | (Z = z) \sim \text{Poiss}(z)$. After observing X , the conditional distribution over Z is given by

$$p(z | X = x) = \mathcal{G}(z; \alpha + x, \beta + 1). \quad (5.47)$$

Furthermore, the conditional mean and variance are

$$\begin{aligned}M_1 &= \frac{\alpha + x}{\beta + 1} \\ S_2 &= \frac{\alpha + x}{(\beta + 1)^2}\end{aligned}\quad (5.48)$$

and the ratio thereof becomes

$$\frac{S_2^2}{M_1} = \frac{\alpha + x}{(\beta + 1)^3}. \quad (5.49)$$

Taking the expectation with respect to x then yields

$$\frac{S_2^2}{M_1} = \frac{\alpha + \mathbb{E}[X]}{(\beta + 1)^3}. \quad (5.50)$$

We now compare this expression to $\mathbb{E}[S_2]^2/\mathbb{E}[M_1]$. In particular, we obtain for the two expectations

$$\begin{aligned}\mathbb{E}[M_1] &= \frac{\alpha + \mathbb{E}[X]}{\beta + 1} \\ \mathbb{E}[S_2] &= \frac{\alpha + \mathbb{E}[X]}{(\beta + 1)^2},\end{aligned}\quad (5.51)$$

and therefore, the both expressions will coincide. The expected moments – and hence the suppressed noise then can be found by solving

$$\begin{aligned}\frac{d}{dt}\mathbb{E}[M_1(t)] &= \theta(\mu - \mathbb{E}[M_1(t)])dt \\ \frac{d}{dt}\mathbb{E}[S_2(t)] &= -2\theta\mathbb{E}[S_2(t)] + \sigma_Z^2\mathbb{E}[M_1(t)] - c_k\frac{\mathbb{E}[S_2(t)]^2}{\mathbb{E}[M_1(t)]}.\end{aligned}\tag{5.52}$$

In order to find an expression at stationarity, we set the l.h.s. to zero and solve for M_1 and S_2 .

$$\begin{aligned}M_1^\infty &= \mu \\ S_2^\infty &= \frac{\sqrt{\mu^2(c_k\sigma_Z^2 + \theta^2)} - \mu\theta}{c_k}.\end{aligned}\tag{5.53}$$

Hence, the suppressed and effective noise terms of $Z(t)$ at stationarity are given by

$$\begin{aligned}\mathbb{E}[\text{Var}[Z(t) | \mathbf{x}_t]] &= \frac{\sqrt{\mu^2(c_k\sigma_Z^2 + \theta^2)} - \mu\theta}{c_k} \\ \text{Var}[\mathbb{E}[Z(t) | \mathbf{x}_t]] &= \text{Var}[Z(t)] - \frac{\sqrt{\mu^2(c_k\sigma_Z^2 + \theta^2)} - \mu\theta}{c_k}.\end{aligned}\tag{5.54}$$

Furthermore, the relative effective noise is found by dividing the effective noise by the total noise, i.e.,

$$\frac{\text{Var}[\mathbb{E}[Z(t) | \mathbf{x}_t]]}{\text{Var}[Z(t)]} = 1 + 2\frac{v^2}{c_k} \left(1 - \sqrt{\frac{c_k}{v^2} + 1}\right),\tag{5.55}$$

with $v = \theta/\sigma_Z$ as the normalized timescale of $Z(t)$. The computation of the effective noise and its dependency on the environmental timescale is illustrated in Figure 5.4.

5.4.2 The Slow Noise Approximation

The effective noise can be understood as a measure of how strong Z impacts X . Only in the special case of a very slow or constant environment, i.e., $\mathcal{D}_i \approx 0$, we see from (5.44) that for large t , $\text{Var}[\mathbb{E}[Z(t) | \mathbf{x}_t]] \rightarrow \text{Var}[Z(t)]$, i.e., all variability in Z is transferred to X . Hence, a more noisy but fluctuating environment may induce a similar (or even the same) effective noise in X than a random but fixed environment of the same variance. Consequently, when looking at only snapshot data for X one can generally not infer whether the environment is constant or fluctuating. On the other hand, this suggests that we may well approximate the impact of a complicated and dynamically changing environment by a simple random variable of appropriate variance. More specifically, we demand for an equivalent constant environment \bar{Z} such that $\text{Var}[\bar{Z}] \equiv \text{Var}[\mathbb{E}[Z(t) | \mathbf{x}_t]]$, where $\text{Var}[\mathbb{E}[Z(t) | \mathbf{x}_t]] = \sigma^2$ is the effective noise of the original, fluctuating environment Z at stationarity. Let us again consider the birth-death process

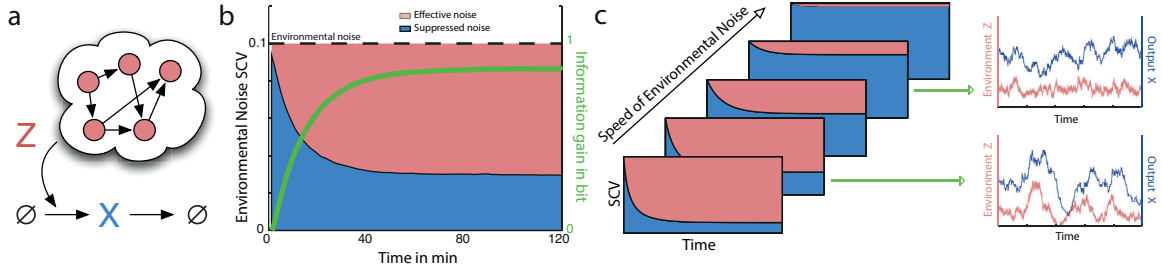


Figure 5.4: Propagation and suppression of environmental fluctuations. (a) Linear birth-death process in a fluctuation environment. The birth-rate is assumed to be linearly modulated by an environmental stochastic process Z . (b) Calculation of suppressed and effective noise. Individual components were computed analytically by solving the ordinary differential equation from (5.52) with $\mu = 0.1$, $\sigma_Z = 0.003$, $\theta = 1e - 4$ and $c_k = 0.05$. For orientation, we also show the information gain between $Z(t)$ and $Z(t) | \mathbf{x}_t$, computed using the marginal simulation algorithm (green); it can be understood as the gain in information about Z through observing X and it exhibits a monotone relationship with the effective noise. (c) Relation between the effective noise and the speed of the environmental fluctuations. Noise contributions were computed by numerically solving the ODE from (5.44) for different values of θ (i.e., timescales).

of Figure 5.4a and set the birth rate to one such that any scaling is subsumed in the environmental process Z . With $X_0 = 0$, the abundance of the birth death process at any time is given by $X(t) = R_b(t) - R_d(t)$ with $R_b(t)$ and $R_d(t)$ as counting processes for the birth and death reaction, respectively.

The slow noise approximation is based on two critical assumptions:

1. The conditional process $Z(t) | \mathbf{x}_t$ can be well represented by a Gamma distribution with time-varying parameters. We highlight that this assumption does *not* mean that the unconditional process $Z(t)$ needs to be approximately Gamma-distributed.
2. The impact of a fluctuating environment on a system can be well “mimicked” by a static environment \bar{Z} with a suitably chosen variance (e.g., the effective noise).

Under the Gamma-assumption, the conditional expectation $M_1(t) = \mathbb{E}[Z(t) | \mathbf{x}_t]$ is governed by the differential equations

$$\begin{aligned}
 dM_1(t) &= [\mathcal{D}_1(t^-) - (M_2(t^-) - M_1^2(t^-))] dt + \frac{M_2(t^-) - M_1^2(t^-)}{M_1(t^-)} dR_b(t) \\
 dM_2(t) &= \left[\mathcal{D}_2(t^-) - 2 \frac{M_2(t^-)}{M_1(t^-)} (M_2(t^-) - M_1^2(t^-)) \right] dt + 2 \left[\frac{M_2^2(t^-)}{M_1^2(t^-)} - M_2(t^-) \right] dR_b(t).
 \end{aligned} \tag{5.56}$$

Assuming a very slow environmental dynamics, we have that $\mathcal{D}_1(t^-)$ and $\mathcal{D}_2(t^-)$ become

zero and hence,

$$\begin{aligned} dM_1(t) &= -(M_2(t^-) - M_1^2(t^-))dt + \frac{M_2(t^-) - M_1^2(t^-)}{M_1(t^-)}dR_b(t) \\ dM_2(t) &= -2\frac{M_2(t^-)}{M_1(t^-)}(M_2(t^-) - M_1^2(t^-))dt + 2\left[\frac{M_2^2(t^-)}{M_1^2(t^-)} - M_2(t^-)\right]dR_b(t). \end{aligned} \quad (5.57)$$

The solution of (5.57) immediately before the next jump at time t is given by

$$\begin{aligned} M_1(t^-) &= \frac{M_1^2(0)}{M_1(0) + (M_2(0) - M_1^2(0))t} \\ M_2(t^-) &= \frac{M_1^2(0)M_2(0)}{[M_1(0) + (M_2(0) - M_1^2(0))t]^2}. \end{aligned} \quad (5.58)$$

Adding the the jump term $[M_2(t^-) - M_1(t^-)^2]/M_1(t^-)$ to $M_1(t^-)$ further yields

$$M_1(t) = \frac{M_1^2(0) + (M_2(0) - M_1^2(0))}{M_1(0) + (M_2(0) - M_1^2(0))t}. \quad (5.59)$$

Repeating the above procedure for the subsequent jumps, we obtain

$$\begin{aligned} M_1(t) &= \frac{M_1^2(0) + (M_2(0) - M_1^2(0))R_b(t)}{M_1(0) + (M_2(0) - M_1^2(0))t} \\ &= \frac{\mu^2}{\mu + \sigma^2 t} + \frac{\sigma^2}{\mu + \sigma^2 t}R_b(t) \\ &= h_b(R_b(t), t), \end{aligned} \quad (5.60)$$

with $\mu = \mathbb{E}[\bar{Z}]$ and $\sigma^2 = \text{Var}[\bar{Z}]$ as mean and variance of the approximate environment \bar{Z} . Importantly, we find that the conditional mean – and therefore the hazard function only depends on the time t and the number of birth reactions $R_b(t)$. If we additionally characterize the system in terms of $R_b(t)$ and $R_d(t)$ instead of $X(t)$, the time-convolutionless master equation from (5.40) applies.

Using the fact that $X(t) = R_b(t) - R_d(t)$ and assuming a death-hazard of the form

$$h_d(X(t), c_d) = c_d X(t) = c_d(R_b(t) - R_d(t)),$$

we obtain the time-convolutionless master equation for $P(r_b, r_d, t) = P(R_b(t) = r_b, R_d(t) = r_d \mid R_b(0) = 0, R_d(0) = 0)$, i.e.,

$$\begin{aligned} \frac{d}{dt}P(r_b, r_d, t) &= h_b(r_b - 1, t)P(r_b - 1, r_d, t) + c_d[r_b - r_d + 1]P(r_b, r_d - 1, t) \\ &\quad - h_b(r_b, t)P(r_b, r_d, t) - c_d[r_b - r_d]P(r_b, r_d, t). \end{aligned} \quad (5.61)$$

In the following, we derive a solution of this equation using the concept of generating functions [37]. In particular, we employ certain properties of the probability generating function

$$\gamma(\eta_b, \eta_d, t) = \sum_{r_d=0}^{\infty} \sum_{r_b=0}^{\infty} \eta_b^{r_b} \eta_d^{r_d} p(r_b, r_d, t), \quad (5.62)$$

which allow us to transform the difference-differential equation (5.61) into a PDE. It is straightforward to show that the discrete shifts in $P(r_b, r_d, t)$ map to partial derivatives of the probability generating function yielding

$$\begin{aligned} \frac{d}{dt}\gamma(\eta_b, \eta_d, t) &= \frac{\mu^2}{\mu + \sigma^2 t} (\eta_b - 1) \gamma(\eta_b, \eta_d, t) \\ &+ \eta_b \left[\frac{\sigma^2}{\mu + \sigma^2 t} (\eta_b - 1) + c_d (1 - \eta_d) \right] \frac{\partial}{\partial \eta_b} \gamma(\eta_b, \eta_d, t) \\ &+ c_d \eta_d (\eta_d - 1) \frac{\partial}{\partial \eta_d} \gamma(\eta_b, \eta_d, t). \end{aligned} \quad (5.63)$$

We realize that the above equation is a linear PDE with time-varying coefficients, which we aim to solve using the method of characteristics. This method is based on describing the PDE by means of so-called *characteristic curves* that are given through a set of coupled ODEs – each of them corresponding to a particular dimension of the PDE (i.e., η_b , η_d and γ). Considering a general linear first-order PDE with three independent variables, i.e.,

$$\frac{d}{dt}\gamma(\eta_b, \eta_d, t) = a(\eta_b, \eta_d, t, \gamma) \frac{\partial}{\partial \eta_b} \gamma(\eta_b, \eta_d, t) + b(\eta_b, \eta_d, t, \gamma) \frac{\partial}{\partial \eta_d} \gamma(\eta_b, \eta_d, t) + c(\eta_b, \eta_d, t, \gamma) \quad (5.64)$$

the characteristic equations are given by

$$\frac{d}{dt}\eta_b(t) = -a(\eta_b(t), \eta_d(t), t, \gamma(t)) \quad (5.65)$$

$$\frac{d}{dt}\eta_d(t) = -b(\eta_b(t), \eta_d(t), t, \gamma(t)) \quad (5.66)$$

$$\frac{d}{dt}\gamma(t) = c(\eta_b(t), \eta_d(t), t, \gamma(t)). \quad (5.67)$$

In the special case of eq. (5.63), we have that

$$\frac{d}{dt}\eta_b(t) = -\eta_b(t) \left(\eta_b(t) \frac{\sigma^2}{\mu + \sigma^2 t} - c_d \eta_d(t) - \frac{\sigma^2}{\mu + \sigma^2 t} + c_d \right) \quad (5.68)$$

$$\frac{d}{dt}\eta_d(t) = -c_d \eta_d(t) (\eta_d(t) - 1) \quad (5.69)$$

$$\frac{d}{dt}\gamma(t) = \frac{\mu^2}{\mu + \sigma^2 t} \gamma(t) (\eta_b(t) - 1), \quad (5.70)$$

whose solution is given by

$$\eta_b(t) = \frac{Ac_d (Be^{c_d t} - B + 1) (\mu + \sigma^2 t)}{-AB\sigma^2 e^{c_d t} + ABC_d \sigma^2 t e^{c_d t} + AB\sigma^2 + A\sigma^2 e^{c_d t} - A\sigma^2 + c_d \mu e^{c_d t}} \quad (5.71)$$

$$\eta_d(t) = \frac{Be^{c_d t}}{Be^{c_d t} - B + 1} \quad (5.72)$$

$$\gamma(t) = e^{-\frac{c_d \mu^2 t}{\sigma^2}} \left(\frac{e^{c_d t} (c_d (AB\sigma^2 t + \mu) - A(B-1)\sigma^2) + A(B-1)\sigma^2}{c_d (\mu + \sigma^2 t)} \right)^{\frac{\mu^2}{\sigma^2}}, \quad (5.73)$$

with $\eta_b(0) = A$, $\eta_d(0) = B$ and $\gamma(0) = 1$. In order to obtain the general solution of γ , we need to express the initial conditions A and B as functions of η_b and η_d using (5.71) and (5.72). The probability generating function γ is finally given by

$$\gamma(\eta_b, \eta_d, t) = \left(\frac{c_d \mu}{c_d (\mu + \sigma^2 t (1 - \eta_b \eta_d)) + \sigma^2 \eta_b (\eta_d - 1) (e^{c_d t} - 1)} \right)^{\frac{\mu^2}{\sigma^2}}. \quad (5.74)$$

Back-transformation then yields the joint probability distribution over r_b and r_d , i.e.,

$$\begin{aligned} P(r_b, r_d, t) &= \frac{1}{r_b! r_d!} \left. \frac{\partial^{r_b+r_d}}{\partial \eta_b^{r_b} \partial \eta_d^{r_d}} \gamma(\eta_b, \eta_d, t) \right|_{\eta_b=0, \eta_d=0} \\ &= \frac{\left(\frac{\mu}{\mu + \sigma^2 t} \right)^{\frac{\mu^2}{\sigma^2}} \Gamma\left(\frac{\mu^2}{\sigma^2} + r_b\right)}{r_d! (r_b - r_d)! \Gamma\left(\frac{\mu^2}{\sigma^2}\right)} \left(\frac{\sigma^2 (e^{c_d t} - 1)}{c_d (\mu + \sigma^2 t)} \right)^{r_b} \left(\frac{c_d t - e^{c_d t} + 1}{e^{c_d t} - 1} \right)^{r_d}, \end{aligned} \quad (5.75)$$

from which we compute the distribution in X as

$$\begin{aligned} P(x, t) &= \sum_{r_b=x}^{\infty} P(r_b, r_b - x, t) \\ &= \mathcal{NB} \left(x; \frac{\mu^2}{\sigma^2}, \frac{c_d \mu e^{c_d t}}{c_d \mu e^{c_d t} + (e^{c_d t} - 1) \sigma^2} \right), \end{aligned} \quad (5.76)$$

i.e., a negative binomial distribution. Furthermore, it is straightforward to show that marginally, both r_b and r_d have negative binomial distributions. We remark that the slow noise approximation is exact in the case of infinitely slow and Gamma distributed—as well as infinitely fast fluctuations. Eq. (5.76) provides a surprisingly simple approximate solution for the transient probability distribution of birth-death processes in a fluctuating environment. In order to check its validity, we compared the analytical approximate distributions to the ones obtained through SSA for a gene expression model, where the environmental fluctuations are assumed to be due to the mRNA dynamics (see Figure 5.5a). More specifically, we computed the Kolmogorov distance between the resulting protein distributions as a function of the environmental timescale. Apart from the high accuracy for the limiting time-scales, Figures 5.5b,c indicate that the SNA provides a good approximation also for intermediate regimes.

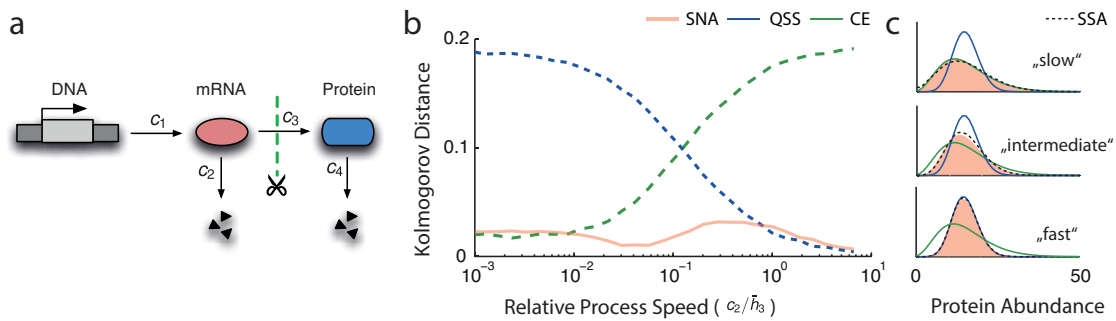


Figure 5.5: Analytical protein distributions through the slow noise approximation. (a) Two-stage gene expression model. Transcription and translation are modeled through mass-action kinetics with reaction rate constants c_1 - c_4 . Fluctuations on the mRNA are considered environmental and hence, integrated out in order to obtain a one-dimensional stochastic process describing only the protein. (b) Accuracy of the slow noise approximation. The SNA was compared to the QSS- and CE-approximations by means of the Kolmogorov distance between the respective approximate and exact distribution (SSA) as a function of the relative speed of the mRNA fluctuations, i.e., c_2/\bar{h}_3 and $\bar{h}_3 = c_3c_1/c_2$. QSS- and CE approximations break down for slow or fast environmental fluctuations respectively, whereas the SNA yields accurate distributions regardless of the mRNA’s timescale. (c) Exemplary distributions obtained through the different approaches in three different regimes (slow, intermediate, fast).

6 Discussion and Outlook

There is increasing evidence that mathematical models of stochastic reaction networks need to account for both intrinsic and extrinsic noise. The latter stems from the stochasticity in a network’s microenvironment such as the ribosomal abundance in the context of gene expression. Previous modeling approaches either completely neglect the extrinsic contributions of noise or account for them by *conditioning* a Markov process on certain environmental quantities which are assumed to be randomly distributed. In the latter models – commonly termed mixed-effect models – every considered cell depends on its individual parametrization, significantly complicating their analysis and inference.

A central goal of this thesis was to develop a general yet efficient modeling framework of stochastic reaction networks subject to extrinsic noise. We have demonstrated that the mathematically proper construction of a coherent and scalable model is achieved through marginalization of the extrinsic factors (see e.g., Sections 3.1 or 4.2). In Chapter 3, this marginalization was performed directly on the CME and its moments, yielding the desired description of the heterogeneous protein distributions. We followed a similar strategy in Chapters 4 and 5, although therein, the marginalization was performed on the process level, meaning that the path measure was integrated with respect to the extrinsic factors, instead of the CME. Importantly, this allows to construct a *marginal* stochastic process, which provides a complete dynamic description of a biochemical network with unknown extrinsic factors. In that sense, the two marginal models from Chapters 3 and 4 differ from each other, but by construction, yield the same marginal probabilities and moments over chemical species. Consequently, the marginal CME models can be understood as “subsets” of the more general models targeting the path measure of the marginal process.

In Chapter 3, this allowed us to derive a scalable statistical model for analyzing population snapshot data such as revealed by flow cytometry. Since the samples of such data are considered statistically independent between subsequent time points, a distribution-based approach – e.g., based on the CME – appears natural for the purpose of inference. However, numerically integrating a CME – for instance using the FSP algorithm – is feasible only for small networks. Beyond that, a straightforward integration of the CME with respect to the extrinsic factors is analytically intractable and thus, an alternative strategy had to be followed. In particular, we proposed a moment-based description of the reaction network, for which the marginalization can be performed analytically. In conjunction with suitable moment-closure techniques, this yields a low-dimensional description of the heterogeneous dynamics, which compared

to a CME-based approach, has excellent scalability with respect to dimensionality and abundance. On its downside – however – it is characterized by two critical limitations. First, suitable closure functions have to be identified through trial and error, meaning that for the inferred parameters, the obtained approximate moments are compared to the ones obtained by Monte Carlo simulation. Second, it might not always be able to uniquely identify all parameters of a network since the information present in the data is not fully exploited by using only a finite number of moments. In the case studies considered in Sections 3.2.3 and 3.2.4, however, mean and variances carried indeed enough information to correctly infer the parameters of the respective reaction networks. In that context, the reader might refer to [105], where the authors propose a principled mathematical framework for assessing the parameter identifiability for given experimental setup.

In Chapter 4 we laid out a marginal inference scheme that can take into account the additional temporal information that is provided by time-lapse microscopy data. It turns out that the inference of the resulting mixed-effect hidden Markov models is highly challenging, since on top of the kinetic parameters, it relies on reconstructing the partially observed sample paths of individual cells. By integrating over extrinsic factors, we first constructed a marginal hidden Markov model and developed a corresponding Bayesian inference scheme. In particular, we made use of a sequential MCMC algorithm, which constructs the posterior distribution over states and parameters recursively over time points and cells such that the full inference problem is split up into a sequence of smaller problems. Marginalized inference schemes appear generally favorable because they profit from a so-called *Rao-Blackwellization* [28], i.e., they can achieve a substantial variance reduction of the desired posterior statistics. We used the algorithm to infer the transcriptional dynamics of an inducible promoter in yeast and show that it is able to correctly dissect the intrinsic and extrinsic contributions to the total variability. We believe that with the increasing throughput and resolution of imaging-based protocols, automated approaches like the one from Section 4.3 will become more and more important. The main idea is to extend the marginal inference framework for the case where the true sources of extrinsic variability are not known beforehand. Due to a sparse Bayesian learning procedure, the method is able to detect on its own which reactions are likely to be targeted by environmental variability. Especially if many cells are considered, such approaches provide a promising means to obtain a comprehensive understanding if and how parts of a biochemical network are affected by extrinsic variability.

Chapter 5 provides a broader view on the marginal dynamics of biochemical networks in random environments. Extending the construction of the marginal process to the case of dynamically changing environments permits an *uncoupled* analysis of a subnetwork that is considered a part of a possibly large entirety. We have shown that the corresponding marginal hazard functions are strictly related to the solution of a stochastic filtering problem and discussed how the latter can be solved in practice. The resulting stochastic process depends on its full history and therefore, cannot be han-

dled through Markovian modeling techniques. For instance, we demonstrated that the uncoupled dynamics do not satisfy a conventional CME but instead, can be modeled through a non-Markovian extension thereof (i.e., a GME). The latter is characterized by either a time-convolution reflecting the memory effects or an explicit time-dependency of the probability in- and out-fluxes. We remark that the hence obtained GMEs must coincide with the solution obtained by performing a straightforward marginalization of a joint CME (e.g., [98]). However, to the best of our knowledge the latter yields intractable hazard functions straightaway, although a detailed comparison of the two approaches is yet to be done. Based on the uncoupled dynamics, we were able to analytically compute how extrinsic noise is transmitted over Poissonian reaction channels as a function of the environmental time scale. More specifically, we derived the amount of variance that is effectively sensed at the Poissonian channel, which we refer to as the effective noise. In this context, we believe that a combination of the marginal process framework with recent information theoretic approaches [17] might help to better understand the complex but robust signal processing within single cells. Furthermore, the framework gave rise to a simple approximation of protein expression in fluctuating environments, which could for instance be used to analyze intrinsic and extrinsic fluctuations of high-throughput single-cell datasets.

In general, we believe that the marginal process framework provides a promising perspective on the dynamics of biological systems subject to extrinsic noise. We have demonstrated its validity using several examples from inference and analysis, but believe it will have further consequences that are not known at present. For instance, the process could be used to quantify and analyze the robustness of a biochemical process with respect to changes in the environmental fluctuations (or statistics thereof). In the context of synthetic biology, this could aid to design circuits, that largely retain their properties when they are embedded into a random or fluctuating environment. We also believe that the framework has the potential to yield further analytical results. For instance, one may examine the uncoupled dynamics for the limiting regime where the same environment impacts a large number (e.g., infinitely many) reactions in a network. Such approach might permit to compute a limiting stochastic process from which a general approximation of biochemical networks in fluctuating environments could be derived.

Bibliography

- [1] O. Aalen. Nonparametric inference for a family of counting processes. *Ann Stat*, 6(4):701–726, 1978.
- [2] O. Aalen. Mixing distributions on a Markov chain. *Scand J Statist*, 14(4):281–289, 1987.
- [3] O. Aalen. Dynamic description of a Markov chain with random time scale. *MATH. SCI.*, 13(2):90–103, 1988.
- [4] O. O. Aalen, Borgan, and H. K. Gjessing. *Survival and event history analysis: a process point of view*. Springer Verlag, 2008.
- [5] M. Amrein and H. R. Künsch. Rate estimation in partially observed Markov jump processes with measurement errors. *Stat Comput*, 22(2):513–526, 2012.
- [6] D. F. Anderson. A modified next reaction method for simulating chemical systems with time dependent propensities and delays. *J Chem Phys*, 127(21):214107, Dec. 2007.
- [7] D. F. Anderson and T. G. Kurtz. Continuous time Markov chain models for chemical reaction networks. In *Design and Analysis of Biomolecular Circuits*, pages 3–42. Springer, 2011.
- [8] A. Bain and D. Crisan. *Fundamentals of stochastic filtering*. Springer (New York), 2009.
- [9] S. Bandara, J. P. Schöder, R. Eils, H. G. Bock, and T. Meyer. Optimal experimental design for parameter estimation of a cell signaling model. *PLoS computational biology*, 5(11):e1000558, 2009.
- [10] N. A. Barker, C. J. Myers, and H. Kuwahara. Learning genetic regulatory network connectivity from time series data. *Computational Biology and Bioinformatics, IEEE/ACM Transactions on*, 8(1):152–165, 2011.
- [11] M. J. Beal. *Variational algorithms for approximate Bayesian inference*. PhD-thesis, University College London, 2003.
- [12] R. Bellman and K. Astrom. On structural identifiability. *Mathematical Biosciences*, 7(3-4):329–339, 1970.
- [13] J. M. Bernardo. Expected information as expected utility. *The Annals of Statistics*, pages 686–690, 1979.
- [14] C. M. Bishop. *Pattern recognition and machine learning*, volume 4. Springer, New York, 2007.
- [15] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 1st ed. 2006. corr. 2nd printing edition, Oct. 2007.

- [16] W. J. Blake, M. Kaern, C. R. Cantor, and J. J. Collins. Noise in eukaryotic gene expression. *Nature*, 422(6932):633–7, Apr. 2003.
- [17] C. G. Bowsher, M. Voliotis, and P. S. Swain. The fidelity of dynamic signaling by noisy biomolecular networks. *Plos Comp Biol*, 9(3):e1002965, 2013.
- [18] M. Briers, A. Doucet, and S. Maskell. Smoothing algorithms for state–space models. *Annals of the Institute of Statistical Mathematics*, 62(1):61–89, 2010.
- [19] A. G. Busetto, A. Hauser, G. Krummenacher, M. Sunnåker, S. Dimopoulos, C. S. Ong, J. Stelling, and J. M. Buhmann. Near-optimal experimental design for model selection in systems biology. *Bioinformatics*, 29(20):2625–2632, 2013.
- [20] Y. Cao and L. Petzold. Accuracy limitations and the measurement of errors in the stochastic simulation of chemically reacting systems. *Journal of Computational Physics*, 212(1):6–24, Feb. 2006.
- [21] A. P. Capaldi, T. Kaplan, Y. Liu, N. Habib, A. Regev, N. Friedman, and E. K. O’Shea. Structure and function of a transcriptional network activated by the MAPK Hog1. *Nat Genet*, 40(11):1300–6, Nov. 2008.
- [22] K. Chaloner and I. Verdinelli. Bayesian Experimental Design: A Review. *Statistical Science*, 10(3):273–304, 1995.
- [23] A. Colman-Lerner, A. Gordon, E. Serra, T. Chin, O. Resnekov, D. Endy, C. G. Pesce, and R. Brent. Regulated cell-to-cell variation in a cell-fate decision system. *Nature*, 437(7059):699–706, 2005.
- [24] G. D. A rigorous derivation of the chemical master equation. *Physica A: Statistical Mechanics and its Applications*, 188(1-3):404–425, 1992.
- [25] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38, 1977.
- [26] J. L. Doob. Markoff chains—denumerable case. *Transactions of the American Mathematical Society*, 58(3):pp. 455–473, 1945.
- [27] A. Doucet, N. De Freitas, and N. Gordon, editors. *Sequential Monte Carlo methods in practice*. Springer, 2001.
- [28] A. Doucet, N. Freitas, K. Murphy, and S. Russell. Rao-Blackwellised particle filtering for dynamic bayesian networks. *The 16th Annual Conference on Uncertainty in Artificial Intelligence*, pages 176–183, 2000.
- [29] U. T. Eden and E. N. Brown. Continuous-time filters for state estimation from point process models of neural data. *Statistica Sinica*, 18(4):1293, 2008.
- [30] F. Eggenberger and G. Polya. Über die Statistik verketteter Vorgänge. *Zeitschrift für angewandte Mathematik und Mechanik*, 3(4):279–289, 1923.
- [31] J. Elf and M. Ehrenberg. Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. *Genome research*, 13(11):2475–2484, 2003.
- [32] R. J. Elliott and W. P. Malcolm. General smoothing formulas for markov-modulated poisson observations. *Automatic Control, IEEE Transactions on*, 50(8):1123–1134, 2005.

- [33] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–6, Aug. 2002.
- [34] B. Finkenstädt, D. J. Woodcock, M. Komorowski, C. V. Harper, J. R. Davis, M. R. White, D. A. Rand, et al. Quantifying intrinsic and extrinsic noise in gene transcription using the linear noise approximation: An application to single cell data. *The Annals of Applied Statistics*, 7(4):1960–1982, 2013.
- [35] N. Friedman, L. Cai, and X. S. Xie. Linking stochastic dynamics to population distribution: An analytical framework of gene expression. *Phys. Rev. Lett.*, 97:168302, Oct 2006.
- [36] C. Furusawa, T. Suzuki, A. Kashiwagi, T. Yomo, and K. Kaneko. Ubiquity of log-normal distributions in intra-cellular reaction dynamics. *Biophysics*, 1:25–31, 2005.
- [37] C. W. Gardiner et al. *Handbook of stochastic methods*, volume 3. Springer Berlin, 1985.
- [38] A. P. Gasch, P. T. Spellman, C. M. Kao, C. O. Harel, M. B. Eisen, G. Storz, D. Botstein, and P. O. Brown. Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell*, 11(12):4241–4257, Dec. 2000.
- [39] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian data analysis*. CRC press, 2013.
- [40] S. Ghaemmaghami, W.-K. Huh, K. Bower, R. W. Howson, A. Belle, N. Dephoure, E. K. O’Shea, and J. S. Weissman. Global analysis of protein expression in yeast. *Nature*, 425(6959):737–741, Oct. 2003.
- [41] A. L. Gibbs and F. E. Su. On choosing and bounding probability metrics. *International Statistical Review*, 70(3):419–435, 2002.
- [42] M. A. Gibson and J. Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *The journal of physical chemistry A*, 104(9):1876–1889, 2000.
- [43] K. Giesecke, H. Kakavand, and M. Mousavi. Exact simulation of point processes with stochastic intensities. *Operations research*, 59(5):1233–1245, 2011.
- [44] D. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of computational physics*, 22(4):403–434, 1976.
- [45] D. Gillespie. Master equations for random walks with arbitrary pausing time distributions. *Physics Letters A*, 64(1):22–24, 1977.
- [46] A. Golightly and D. J. Wilkinson. Bayesian inference for stochastic kinetic models using a diffusion approximation. *Biometrics*, 61(3):781–8, Sept. 2005.
- [47] A. Golightly and D. J. Wilkinson. Bayesian parameter inference for stochastic biochemical network models using particle Markov chain Monte Carlo. *Interface Focus*, 1(6):807–820, 2011.
- [48] E. A. Hackett, K. R. Esch, S. Maleri, and B. Errede. A family of destabilized cyan fluorescent proteins as transcriptional reporters in *s. cerevisiae*. *Yeast*, 23(5):333–

- 349, 2006.
- [49] C. V. Harper, B. Finkenstädt, D. J. Woodcock, S. Friedrichsen, S. Semprini, L. Ashall, D. G. Spiller, J. J. Mullins, D. a. Rand, J. R. E. Davis, and M. R. H. White. Dynamic analysis of stochastic transcription cycles. *PLoS Biol*, 9(4):e1000607, Apr. 2011.
 - [50] J. Hasenauer, S. Waldherr, M. Doszczak, N. Radde, P. Scheurich, and F. Allgöwer. Identification of models of heterogeneous cell populations from population snapshot data. *BMC Bioinformatics*, 12(1):125, 2011.
 - [51] S. Hengl, C. Kreutz, J. Timmer, and T. Maiwald. Data-based identifiability analysis of non-linear dynamical models. *Bioinformatics*, 23(19):2612–2618, 2007.
 - [52] J. Hespanha. Moment closure for biochemical networks. In *Communications, Control and Signal Processing, 2008. ISCCSP 2008. 3rd International Symposium on*, pages 142–147. IEEE, 2008.
 - [53] A. Hilfinger and J. Paulsson. Separating intrinsic from extrinsic fluctuations in dynamic biological systems. *Proc Natl Acad Sci USA*, 108(29):12167–12172, July 2011.
 - [54] S. Hohmann. Osmotic stress signaling and osmoadaptation in yeasts. *Microbiology and molecular biology reviews : MMBR*, 66(2):300–372, June 2002.
 - [55] F. C. Holstege, E. G. Jennings, J. J. Wyrick, T. I. Lee, C. J. Hengartner, M. R. Green, T. R. Golub, E. S. Lander, and R. a. Young. Dissecting the regulatory circuitry of a eukaryotic genome. *Cell*, 95(5):717–28, Nov. 1998.
 - [56] S. M. Kay. *Fundamentals of statistical signal processing, volume i: Estimation theory (v. 1)*. Prentice Hall, 1993.
 - [57] V. Kenkre, E. Montroll, and M. Shlesinger. Generalized master equations for continuous-time random walks. *Journal of Statistical Physics*, 9(1):45–50, 1973.
 - [58] J. Kiefer and J. Wolfowitz. Stochastic estimation of the maximum of a regression function. *The Annals of Mathematical Statistics*, 23:462–466, 1952.
 - [59] H. Koepl, C. Zechner, A. Ganguly, S. Pelet, and M. Peter. Accounting for extrinsic variability in the estimation of stochastic rate constants. *Int J Robust Nonlin*, 22(10):1103–1119, 2012.
 - [60] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press, Cambridge, 1 edition, July 2009.
 - [61] M. Komorowski, M. J. Costa, D. A. Rand, and M. P. Stumpf. Sensitivity, robustness, and identifiability in stochastic chemical kinetics models. *Proceedings of the National Academy of Sciences*, 108(21):8645–8650, 2011.
 - [62] U. Kuechler and M. Sorensen. *Exponential families of stochastic processes*. Springer, New York, 1997.
 - [63] T. Kurtz. Lectures on stochastic analysis. *Department of Mathematics and Statistics, University of Wisconsin, Madison, WI*, pages 53706–1388, 2001.
 - [64] T. G. Kurtz. The relationship between stochastic and deterministic models for chemical reactions. *The Journal of Chemical Physics*, 57(7):2976–2978, 1972.

- [65] H. Kushner. A projected stochastic approximation method for adaptive filters and identifiers. *IEEE Transactions on Automatic Control*, 25(4):836–838, 1980.
- [66] V. Leijdekker and P. Spreij. Explicit computations for a filtering problem with point process observations with applications to credit risk. *Probability in the Engineering and Informational Sciences*, 25:393–418, 7 2011.
- [67] I. Lestas, G. Vinnicombe, and J. Paulsson. Fundamental limits on the suppression of molecular fluctuations. *Nature*, 467(7312):174–178, 2010.
- [68] P. A. W. Lewis and G. S. Shedler. Simulation of nonhomogeneous Poisson processes by thinning. *Naval Research Logistics Quarterly*, 26(3):403–413, 1979.
- [69] J. Liepe, S. Filippi, M. Komorowski, and M. P. Stumpf. Maximizing the information content of experiments in systems biology. *PLoS computational biology*, 9(1):e1002888, 2013.
- [70] J. Liepe, P. Kirk, S. Filippi, T. Toni, C. P. Barnes, and M. P. Stumpf. A framework for parameter estimation and model selection from experimental data in systems biology using approximate bayesian computation. *Nature protocols*, 9(2):439–456, 2014.
- [71] J. F. Louvion, B. Havaux-Copf, and D. Picard. Fusion of GAL4-VP16 to a steroid-binding domain provides a tool for gratuitous induction of galactose-responsive genes in yeast. *Gene*, 131(1):129–34, Sept. 1993.
- [72] J. Macia, S. Regot, T. Peeters, N. Conde, R. Solé, and F. Posas. Dynamic signaling in the Hog1 MAPK pathway relies on high basal signal transduction. *Science signaling*, 2(63):ra13, Jan. 2009.
- [73] P. B. Mason and K. Struhl. Distinction and relationship between elongation rate and processivity of rna polymerase ii in vivo. *Mol Cell*, 17(6):831–841, 2005.
- [74] H. H. McAdams and A. Arkin. Stochastic mechanisms in gene expression. *Proc Natl Acad Sci USA*, 94(3):814–819, 1997.
- [75] R. S. McIsaac, S. J. Silverman, M. N. McClean, P. A. Gibney, J. Macinskas, M. J. Hickman, A. A. Petti, and D. Botstein. Fast-acting and nearly gratuitous induction of gene expression and protein depletion in *Saccharomyces cerevisiae*. *Mol Biol Cell*, 22(22):4447–59, Nov. 2011.
- [76] D. A. McQuarrie. Stochastic approach to chemical kinetics. *Journal of Applied Probability*, 4(3):413–478, 1967.
- [77] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.
- [78] C. Michelot. A finite algorithm for finding the projection of a point onto the canonical simplex of \mathbb{R}^n . *Journal of Optimization Theory and Applications*, 50(1):195–200, 1986.
- [79] P. Milner, C. S. Gillespie, and D. J. Wilkinson. Moment closure based parameter inference of stochastic kinetic models. *Statistics and Computing*, 23(2):287–295, 2013.

- [80] E. W. Montroll and G. H. Weiss. Random walks on lattices. ii. *Journal of Mathematical Physics*, 6(2):167–181, 1965.
- [81] B. Munsky and M. Khammash. The finite state projection algorithm for the solution of the chemical master equation. *J Chem Phys*, 124:044104, 2006.
- [82] B. Munsky, B. Trinh, and M. Khammash. Listening to the noise: random fluctuations reveal gene network parameters. *Mol Syst Biol*, 5(1), 2009.
- [83] D. Muzzey, C. a. Gómez-Uribe, J. T. Mettetal, and A. van Oudenaarden. A systems-level analysis of perfect adaptation in yeast osmoregulation. *Cell*, 138(1):160–71, July 2009.
- [84] S. K. N. Johnson. *Urn Models and Their Application*. Wiley & Sons, New York, 1977.
- [85] G. O. N. Limnios. *Semi-Markov processes and reliability*. Birkhäuser, Boston, 2001.
- [86] P. Nandy, M. Unger, C. Zechner, and H. Koepl. Optimal perturbations for the identification of stochastic reaction dynamics. In *Proceedings of the 16th IFAC Symposium on System Identification*, pages 686–691, 2012.
- [87] R. M. Neal. *Bayesian Learning for Neural Networks*. Springer, New York, 1996.
- [88] G. Neuert, B. Munsky, R. Z. Tan, L. Teytelman, M. Khammash, and A. van Oudenaarden. Systematic identification of signal-activated stochastic gene regulation. *Science*, 339(6119):584–587, 2013.
- [89] J. R. S. Newman, S. Ghaemmaghami, J. Ihmels, D. K. Breslow, M. Noble, J. L. DeRisi, and J. S. Weissman. Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature*, 441(7095):840–6, June 2006.
- [90] U. Nodelman, C. R. Shelton, and D. Koller. Continuous time bayesian networks. In *Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence*, pages 378–387. Morgan Kaufmann Publishers Inc., 2002.
- [91] J. R. Norris. *Markov chains*. Cambridge university press, 1998.
- [92] M. Opper and G. Sanguinetti. Variational inference for Markov jump processes. *Adv. in Neural Information Proc. Systems 20*, 2007.
- [93] S. Pelet, R. Dechant, S. S. Lee, F. van Drogen, and M. Peter. An integrated image analysis platform to quantify signal transduction in single cells. *Integr Biol*, 4(10):1274–82, 2012.
- [94] S. Pelet, F. Rudolf, M. Nadal-Ribelles, E. de Nadal, F. Posas, and M. Peter. Transient activation of the HOG MAPK pathway regulates bimodal gene expression. *Science*, 332(6030):732–5, May 2011.
- [95] S. Plyasunov and A. P. Arkin. Efficient stochastic sensitivity analysis of discrete event systems. *Journal of Computational Physics*, 221(2):724–738, 2007.
- [96] A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol*, 4(10):e309, Oct. 2006.
- [97] R. Ramaswamy, N. González-Segredo, and I. F. Sbalzarini. A new class of highly efficient exact stochastic simulation algorithms for chemical reaction networks.

- The Journal of chemical physics*, 130(24):244104, 2009.
- [98] C. V. Rao and A. P. Arkin. Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the gillespie algorithm. *The Journal of Chemical Physics*, 118(11):4999–5010, 2003.
- [99] J. M. Raser and E. K. O’Shea. Control of stochasticity in eukaryotic gene expression. *Science*, 304(5678):1811–1814, 2004.
- [100] M. Rathinam, P. W. Sheppard, and M. Khammash. Efficient computation of parameter sensitivities of discrete stochastic chemical reaction networks. *The Journal of chemical physics*, 132(3):034103, 2010.
- [101] R. Rinott, A. Jaimovich, and N. Friedman. Exploring transcription regulation through cell-to-cell variability. *Proc Natl Acad Sci USA*, 108(15), Mar. 2011.
- [102] C. P. Robert. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation (Springer Texts in Statistics)* by. Springer-Verlag New York, 2001.
- [103] G. O. Roberts and J. S. Rosenthal. Harris recurrence of Metropolis-within-Gibbs and trans-dimensional Markov chains. *Ann. Appl. Probab.*, 16(4):2123–2139, 2006.
- [104] N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain, and M. B. Elowitz. Gene regulation at the single-cell level. *Science*, 307(5717):1962–1965, 2005.
- [105] J. Ruess, A. Miliadis-Argeitis, and J. Lygeros. Designing experiments to understand the variability in biochemical reaction networks. *Journal of The Royal Society Interface*, 10(88), 2013.
- [106] J. Ruess, A. Miliadis-Argeitis, S. Summers, and J. Lygeros. Moment estimation for chemically reacting systems by extended Kalman filtering. *J Chem Phys*, 135:165102, 2011.
- [107] I. Sadowski, J. Ma, S. Triezenberg, and M. Ptashne. GAL4-VP16 is an unusually potent transcriptional activator. *Nature*, 335:563–64, 1988.
- [108] M. S. Samoilov and A. P. Arkin. Deviant effects in molecular reaction pathways. *Nature biotechnology*, 24(10):1235–1240, 2006.
- [109] W. Sandmann. Simultaneous stochastic simulation of multiple perturbations in biological network models. In *Computational Methods in Systems Biology*, pages 15–31. Springer, 2007.
- [110] V. Shahrezaei, J. F. Ollivier, and P. S. Swain. Colored extrinsic fluctuations and stochastic gene expression. *Mol Syst Biol*, 4(196):196, 2008.
- [111] V. Shahrezaei and P. S. Swain. Analytical distributions for stochastic gene expression. *Proceedings of the National Academy of Sciences*, 105(45):17256–17261, 2008.
- [112] F. Shibata, Y. Takahashi, and N. Hashitsume. A generalized stochastic liouville equation. non-markovian versus memoryless master equations. *Journal of Statistical Physics*, 17(4):171–187, 1977.
- [113] D. Shutin, C. Zechner, S. R. Kulkarni, and H. V. Poor. Regularized Variational

- Bayesian Learning of Echo State Networks with Delay & Sum Readout. *Neural Comput*, 24(4):967–995, 2012.
- [114] A. Singh and J. P. Hespanha. A derivative matching approach to moment closure for the stochastic logistic model. *Bulletin of mathematical biology*, 69(6):1909–1925, 2007.
- [115] P. Smadbeck and Y. N. Kaznessis. A closure scheme for chemical master equations. *Proceedings of the National Academy of Sciences*, 110(35):14261–14265, 2013.
- [116] B. Snijder and L. Pelkmans. Origins of regulated cell-to-cell variability. *Nat Rev Mol Cell Biol*, 12(2):119–25, 2011.
- [117] D. L. Snyder and M. I. Miller. *Random Point Processes in Time and Space*. Wiley & Sons, New York, 1975.
- [118] G. Storvik. Particle filters for state-space models with the presence of unknown static parameters. *Trans. Sig. Proc.*, 50(2):281–289, Feb. 2002.
- [119] D. M. Suter, N. Molina, D. Gatfield, K. Schneider, U. Schibler, and F. Naef. Mammalian genes are transcribed with widely different bursting kinetics. *Science*, 332(6028):472–4, Apr. 2011.
- [120] P. S. Swain, M. B. Elowitz, and E. D. Siggia. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences*, 99(20):12795–12800, 2002.
- [121] Y. Taniguchi, P. J. Choi, G. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie. Quantifying E. coli proteome and transcriptome with Single-Molecule sensitivity in single cells. *Science*, 329(5991):533–538, 2010.
- [122] M. E. Tipping. Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, 1:211–244, 2001.
- [123] T. Toni and B. Tidor. Combined model of intrinsic and extrinsic variability for computational network design with application to synthetic biology. *PLoS Comput Biol*, 9(3):e1002960, 03 2013.
- [124] M. Unger, S. Lee, M. Peter, and H. Koepl. Pulse Width Modulation of Liquid Flows: Towards Dynamic Control of Cell Microenvironments. *μ TAS 11*, pages 1567–1569, 2011.
- [125] N. G. Van Kampen. *Stochastic processes in physics and chemistry*, volume 1. Elsevier, 1992.
- [126] A. Varshavsky. The N-end rule: Functions, mysteries, uses. *Proc Natl Acad Sci USA*, 93(22):12142–9, Oct. 1996.
- [127] D. Volfson, J. Marciniak, W. J. Blake, N. Ostroff, L. S. Tsimring, and J. Hasty. Origins of extrinsic variability in eukaryotic gene expression. *Nature*, 439(7078):861–864, 2006.
- [128] P. Whittle. On the use of the normal approximation in the treatment of stochastic processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 19(2):268–281, 1957.

-
- [129] D. J. Wilkinson. *Stochastic Modelling for Systems Biology*. Chapman and Hall/CRC, London, Apr. 2006.
- [130] D. J. Wilkinson. Parameter inference for stochastic kinetic models of bacterial gene regulation: a Bayesian approach to systems biology. *Bayesian Statistics*, 9(Wilkinson 2009):679–705, 2011.
- [131] V. Wolf, R. Goel, M. Mateescu, and T. Henzinger. Solving the chemical master equation using sliding windows. *BMC systems biology*, 4:42, Jan. 2010.
- [132] C. Zechner, S. Deb, and H. Koepl. Marginal dynamics of stochastic biochemical networks in random environments. In *Proceedings of the European Control Conference (ECC), 2013*, pages 4269–4274, 2013.
- [133] C. Zechner and H. Koepl. Uncoupled analysis of stochastic reaction networks in fluctuating environments. *Plos Comp Biol*, 2014. In press.
- [134] C. Zechner, P. Nandy, M. Unger, and H. Koepl. Optimal variational perturbations for the inference of stochastic reaction dynamics. In *Proceedings of the IEEE Conference on Decision and Control, 2012*, pages 5336–5341. IEEE, 2012.
- [135] C. Zechner, S. Pelet, M. Peter, and H. Koepl. Recursive Bayesian estimation of stochastic rate constants from heterogeneous cell populations. In *Proceedings of the IEEE Conference on Decision and Control, 2011*, pages 5837–5843, 2011.
- [136] C. Zechner, J. Ruess, P. Krenn, S. Pelet, M. Peter, J. Lygeros, and H. Koepl. Moment-based inference predicts bimodality in transient gene expression. *Proc Natl Acad Sci USA*, 109(21):8340–8345, 2012.
- [137] C. Zechner, M. Unger, S. Pelet, M. Peter, and H. Koepl. Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings. *Nat Methods*, 11(2):197–202, 2014.
- [138] C. Zechner, F. Wadehn, and H. Koepl. Sparse learning of Markovian population models in random environments. In *Proceedings of the 19th IFAC World Congress, 2014*, pages 1723–1728. IFAC, 2014.
- [139] D. Zenklusen, D. R. Larson, and R. H. Singer. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat Struct Mol Biol*, 15(12):1263–71, Dec. 2008.
- [140] Z. Zi, W. Liebermeister, and E. Klipp. A quantitative study of the Hog1 MAPK response to fluctuating osmotic stress in *Saccharomyces cerevisiae*. *PloS one*, 5(3):e9522, Jan. 2010.
- [141] R. Zwanzig. On the identity of three generalized master equations. *Physica*, 30(6):1109–1123, 1964.

Curriculum Vitae

Christoph Zechner

born July 6th, 1986 in Austria.

- 2010 – 2014 Doctorate at the Automatic Control Laboratory,
ETH Zürich, Switzerland (Dr. sc. ETH Zürich)
- 2008 – 2010 Master studies in Telematics,
Graz University of Technology, Austria (MSc.)
- 2005 – 2008 Bachelor studies in Telematics,
Graz University of Technology, Austria (BSc.)
- 1996 – 2004 Gymnasium Fürstenfeld, Austria (High School, Matura)