Reconstructing dynamic molecular states from single-cell time series

Lirong Huang1, Loic Pauleve4, Christoph Zechner2, Michael Unger3, Anders S. Hansen5 and Heinz Koeppl6,7

1Institute of Molecular Systems Biology, D-BIOL, 2Department of Biosystems Science and Engineering, and 3Institute of Biochemistry, ETH Zurich, Zurich, Switzerland
4Laboratoire de Recherche en Informatique, Université Paris-Sud, Orsay, France
5Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA
6Department of Electrical Engineering and Information Technology, and 7Department of Biology, Technische Universität Darmstadt, Darmstadt, Germany

The notion of state for a system is prevalent in the quantitative sciences and refers to the minimal system summary sufficient to describe the time evolution of the system in a self-consistent manner. This is a prerequisite for a principled understanding of the inner workings of a system. Owing to the complexity of intracellular processes, experimental techniques that can retrieve a sufficient summary are beyond our reach. For the case of stochastic biomolecular reaction networks, we show how to convert the partial state information accessible by experimental techniques into a full system state using mathematical analysis together with a computational model. This is intimately related to the notion of conditional Markov processes and we introduce the posterior master equation and derive novel approximations to the corresponding infinite-dimensional posterior moment dynamics. We exemplify this state reconstruction approach using both in silico data and single-cell data from two gene expression systems in Saccharomyces cerevisiae, where we reconstruct the dynamic promoter and mRNA states from noisy protein abundance measurements.

1. Introduction

Today’s experimental techniques of molecular biology give us some insight into biological cells but provide a far from complete picture of the inner workings of such cells or even of any of their subcomponents. With the advances in quantitative single-cell technologies, the generation of calibrated models of particular cellular processes such as the expression of a gene becomes feasible [1–7]. A calibrated model can then be simulated forward to explore the a priori behaviours that one can expect to observe experimentally. However, such forward simulations are not useful if one asks which of those behaviours are compatible with the actual measurement of a particular experiment. For instance, a stochastic gene expression model can give rise to various mRNA and protein trajectories but a model alone cannot be used to determine those mRNA dynamics that are compatible with particular protein measurement trajectories. In general, the problem is to reconstruct the dynamics of experimentally inaccessible states of a process that best match the trajectories of the observable states of the process in a particular experimental run. In other words, the observations allow us to filter the a priori behaviours into compatible posterior behaviours. Mathematically, we condition the inaccessible states onto those observations. Such conditioning or filtering has a long tradition in mathematics and engineering [8,9]. The key is to obtain governing master equations, such as the Kushner and the Zakai equations [10,11], that describe the time evolution of the conditional probability distribution. The best known of such governing equations is the Kalman filter, which yields a finite parametrization of the posterior distribution by considering the case that states evolve according to a linear stochastic differential equation and measurements are

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.3457479.
Gaussian distributed [12]. In most other cases (e.g. nonlinear stochastic differential equation), no finite parametrization of the posterior distribution can be found and a plethora of approximations has been proposed in recent decades [13–15]. In comparison with the other approximation methods, such as the popular extended Kalman filter, particle filters [16,17] do not rely on any local linearization techniques while, as a Monte Carlo method, they are computationally expensive and do not scale well to high state dimensions.

In this work, we complement the chemical master equation used to describe the a priori dynamics of a stochastic reaction network with its posterior counterpart. It refactors the seminal result of Wonham obtained for the case of continuous-time Markov processes [18]. The posterior master equation and its exact posterior moment equation exhibit the same scalability problems as their traditional a priori counterparts and we present scalable approximations of the posterior process. More specifically, the main contribution of this work is a novel approximation approach obtained by specific adaptations of moment closure techniques to the practical scenario where observations are available only in continuous time [19], we mainly focus on the practical scenario where observations are available only at discrete time points.

2. Conditional Markov processes

We consider a well-mixed reaction system of $d$ species and $r$ different reaction channels. The state $X(t)$ comprising the integer abundance of the species at time $t$ shows Markovian dynamics when state transitions take place according to the change vectors $v_i \in \mathbb{Z}^d$ of each reaction and where the reaction’s propensity is given by the function $\alpha_i(x)$ for $x \in \mathcal{X} \subset \mathbb{N}^d$. With that, a reaction system with $X(0)$ copies of each species at time zero will have

$$X(t) = X(0) + \sum_{j=1}^{r} N_j(t)v_j, \quad N_j(t) = \xi_j \left( \int_0^t \alpha_j(X(s)) \, ds \right),$$

copies at time $t$, where $\xi_j(t)$ are independent Poisson processes of unit rate [20]. Starting with some initial distribution $P(X(0) = x)$, the distribution $P(X(t) = x)$ at time $t$ evolves according to the chemical master equation [21]. Following our nomenclature, we refer to this equation as the unconditional or prior master equation as it determines the probability over species abundances if no further information or measurements on the system are provided. The traditional way in mathematics to formalize measurements [19] is to assume some $l$-dimensional covariate process $Y(t)$ of $X(t)$, for instance of the form

$$Y(t) = \int_0^t g(X(s)) \, ds + BW(t), \quad (2.1)$$

where $B$ is a full-rank matrix and $W(t)$ is the standard $l$-dimensional Brownian motion independent of $X(t)$. For example, for $l = 1$ and $g(x) = x$, one can think of a reaction system where the $i$-th species is fluorescently labelled and $Y(t)$ corresponds to the integrated fluorescence intensity measured at the microscope. Observation model (2.1) could be appropriate in the context of fluorescence correlation spectroscopy [22] where a photon count trace at the photo-multipliers under high arrival rates admits a diffusion approximation of the form (2.1). Having such observations available one can ask for the probability over species abundance in the presence of that information. That is, the conditional probability $P(X(t) = x | y_s, s \in (0, \tau])$, where $\tau$ denotes the time up to which measurements are available. Accordingly, one distinguishes between filtering and smoothing for $\tau = t$ and $\tau > t$, respectively. As conditioning usually reduces variance, the measurements from a single cell result in less uncertainty in the dynamic states of the reaction system than with the traditional (prior) chemical master equation. Interestingly, the process $X(t)$ conditioned on such covariate information is still Markovian [8]. Refactoring the seminal work of Wonham for optimal filtering of continuous-time Markov chains [18], the resulting conditional chemical master equation for $\tau = t$ reads

$$d\pi(x,t) = \sum_{j=1}^{r} \left[ \pi(x - v_j,t)\alpha_j(x - v_j) - \pi(x,t)\alpha_j(x) \right] dt + \pi(x,t)[g(x) - \tilde{g}(t)]^T[BB^T]^{-1}[dy(t) - \tilde{g}(t) \, dt], \quad (2.2)$$

with $\pi(x,t) = P(X(t) = x \mid y_s, s \in (0, t])$, $\pi(x,0) = P(X(0) = x)$ and $\tilde{g}(t) = \mathbb{E}[g(X(t))]$, where expectation is taken with respect to $\pi(x,t)$. Owing to its dependency on that expectation, (2.2) is cumbersome to solve and one often resorts to an equation for the unnormalized version of $\pi(x,t)$, performing normalization after numerical integration (electronic supplementary material, S.3). Equation (2.2) is a special case of the general class of optimal filtering equations [19].

In most live-cell imaging applications observation at continuous time are unrealistic due to experimental constraints such as phototoxicity and bleaching. In practice one is faced with observations at discrete times $y_1 = y(t_1), \ldots, y_N = y(t_N)$ with $0 = t_0 < t_1 < \ldots < t_N = T$ usually admitting a conditional distribution of the form $p(y_{1:k} | X(s), s \in [0,t_n]) = p(y_k | X(t_n))$. Given the states at observation times, the observations are assumed to be independent. The conditional probability $P(X(t) = x | y_1, \ldots, y_N)$ with $n = \max(n \in \mathbb{N} \mid t_n \leq t)$ satisfies the unconditional master equation

$$\frac{d\pi(x,t)}{dt} = \sum_{j=1}^{r} \left[ \pi(x - v_j,t)\alpha_j(x - v_j) - \pi(x,t)\alpha_j(x) \right], \quad (2.3)$$

with $\pi(x,0) = P(X(0) = x)$ together with the reset conditions $\pi(x,t_i) = C_i^{-1} \pi(x,t_i^-)p(y_i | x)$,

$$\pi(x,t_i) = C_i^{-1} \pi(x,t_i^-)p(y_i | x), \quad (2.4)$$

at the observation times $t_i \leq t_n$, with the normalizing constant $C_i = \sum_{x \in \mathcal{X}} \pi(x,t_i^+)p(y_i | x)$. For the smoothing case, the conditional probability $P(X(t) = x \mid y_1, \ldots, y_N)$ for any $t \in [0,T]$ denoted by $\tilde{\pi}(x,t)$ admits the factorization through the Markov property of the form

$$\tilde{\pi}(x,t) = \beta(x,t) \pi(x,t), \quad (2.5)$$

where $\beta(x,t) = Z_k^{-1} p(y_k, \ldots, y_N | X(t) = x)$ with $k = \min (k \in \mathbb{N} \mid t_k > t)$ and normalizer $Z_k = p(y_k, \ldots, y_N | y_1, \ldots, y_{k-1})$. The probability of future observations given the current state $\beta(x,t)$ satisfies the backward master equation

$$\frac{d\beta(x,t)}{dt} = \sum_{j=1}^{r} [\beta(x,t) - \beta(x + v_j,t)]\alpha_j(x), \quad (2.6)$$

with reset conditions $\beta(x,t_n^-) = C_i^{-1} \beta(x,t_i) p(y_i | x)$,

$$\beta(x,t_i^+) = C_i^{-1} \beta(x,t_i^-) p(y_i | x), \quad (2.7)$$

and terminal conditions $\beta(x,T) = 1$ for all $x \in \mathcal{X}$. By solving the
backward and the forward conditional master equation, one can
determine the smoothing probability \( \pi(x, t) \) through (2.5). Exam-
ining this factorization and the fact that \( \beta(x, T) = 1 \), we conclude
that our knowledge of the underlying system state \( X(t) \) is less
uncertain when all measurements up to time \( T \) are incorporated
compared with the filtering case where measurements up to time
\( t \) are taken into account. Differentiating (2.5) yields an evolution
equation for \( \pi(x, t) \) directly

\[
\frac{d\pi(x, t)}{dt} = \sum_{j=1}^{r} \hat{a}_j(x-v_j,t)\pi(x-v_j,t) - \hat{a}_i(x,t)\pi(x,t),
\]

which we term the posterior or smoothing master equation
(2.8), where \( \hat{a}_i \) are the time-varying propensities functions of the form

\[
\hat{a}_i(x,t) = a_i(x) \frac{\beta(x + v_i,t)}{\beta(x,t)}. \tag{2.9}
\]

Hence, the prior propensities \( a_i(x) \) are modulated by a
time-varying fraction that steers the process towards future
measurements. The expression for the posterior propensities provides
the means to draw sample paths of the posterior smoothing process through a stochastic simulation scheme adapted to time-varying propensities [23,24].

As an illustrative example, we consider the smoothing problem for a birth–death process \( 0 \rightarrow X \rightarrow \omega \) with respective rates \( c_1, c_2 \) and with a single noise-free observation \( y_t = X(t_1) = 0 \) at \( t_1 = T \) and the deterministic initial condition \( X(0) = 0 \). This set-up corresponds to the classical bridging problem or to the problem of endpoint conditioned sampling of Markov chains [25] and can be solved explicitly for this case. We aim to compute the smoothing distribution \( \pi(x,t) = P(X(t) = x | X(0) = X(T) = 0, \forall t \in [0, T] \) coincides with the filtering distribution within \( t \in [0, T] \) and admits a representation in terms of a Poisson distribution with time-varying rate

\[
\lambda(t) = 1 - e^{-\lambda(t)},
\]

where \( \lambda(t) \) is a monotonically decreasing function in \( t \) and reaches zero at \( t = T \). Hence, in order to reach the state \( X(T) = 0 \) with probability one, the posterior birth rate converges to zero while the death rate becomes unbounded as \( t \to T \). Some sample paths of the posterior birth–death process with time-varying rates (2.10) are shown in figure 1.

3. Posterior moment equations

The presented posterior master equations inherit the same
stability problems as the original chemical master equation
due to the combinatorial increase in the cardinality of \( \mathcal{X} \)
with the number of species \( d \). Traditional approaches that can approximately capture the stochastic dynamics of the prior process are the van Kampen expansion [21] and moment closure techniques [28–30]. Similar techniques can be applied to the posterior master equations (2.3) and also to the backward equation (2.6). For the latter, a linear noise approximation was performed in [31]. We follow the moment closure approach and subsequently derive novel approximate posterior moment dynamics. Throughout we will consider propensity functions of the form \( a_i(x) = c_i g_i(x) \)
with \( c_i \in \mathbb{R}_{>0} \), the stochastic rate constant of the reaction \( j \) and \( g_i(x) \) any polynomial function of bounded degree. Using a multi-index \( \eta = (\eta_1, \ldots, \eta_d) \) with \( |\eta| = \eta_1 + \cdots + \eta_d \) for \( \eta \in \mathbb{N}^d \) [30], we can compactly write \( a_i(x) = \sum_{|\eta| \leq \eta} a_{ij} x^\eta \)
with the shorthand \( x^\eta = x_1^{\eta_1} \cdots x_d^{\eta_d} \). Similar to the moment expansion for the traditional master equation [30], one can develop a moment expansion of the filtering master equation (2.3). Denoting its moment of order \( \eta \) by \( M^n(t) \) the generally

\[
\frac{dM^n(t)}{dt} = \sum_{j=1}^{r} \sum_{|\eta| \leq \eta} a_{ij} \sum_{\eta_0 < \eta} v_j^\eta \left( \frac{\eta}{\xi} \right) M^{n-\eta + \eta_j}(t),
\]

with the reset conditions at the observation time \( t_i \leq t_0 \)
following the Kallianpur–Striebel formula [19]

\[
M^n(t_i) = C_i^{-1} \sum_{x \in \mathcal{X}} \pi(x, t_i) x^\eta p(y_i | x).
\]

In terms of computational complexity, solving the moment equations together with (3.2) does not show any advantage compared with directly solving (2.3) because (3.2) still involves the filtering distribution \( \pi(x,t) \). The main contribution of this work is to propose an approximate approach to the reset condition (3.2) for the posterior moments.

4. Approximate posterior moments dynamics

In the following, we derive novel approximate posterior moment dynamics exploiting a special form of the tower property for conditional expectations

\[
E[s(X(t_0)) | y_1, \ldots, y_{n-1}]
= E[E[s(X(t_0)) | y_1, \ldots, y_{n-1}, y_1, \ldots, y_{n-1}]],
\]

for integrable functions \( s \) and \( u \). Additionally, we make use of traditional moment closure techniques. We begin with a univariate system (i.e. \( d = 1 \)) and assume that observations are subject to additive noise \( y_t = X(t) + \epsilon_t \), where \( \epsilon_t \) are i.i.d. random variables with bounded moments up to order four. For \( \eta \geq 1 \), we obtain the following approximation to the
reset conditions (3.2) for the posterior filtering moments (electronic supplementary material, S.6):
\[ M^n(t) = \sum_{q=0}^{\eta} \partial_{\eta q} y_q, \]  
(4.1)

where the coefficients \( \partial_{\eta q} \) are determined by the linear equations obtained from the tower property (4.1) (electronic supplementary material, S.6)

\[ \sum_{q=0}^{\eta} \partial_{\eta q} \sum_{j=0}^{\eta} \left( q + j \right) M^{q+j}(t_{i}) \mathbb{E}[w_{i}^{q+j-k}] \]
\[ = \sum_{j=0}^{\eta} \left( j \right) M^{n+j}(t_{i}) \mathbb{E}[w_{i}^{j-k}], \]
(4.2)

with \( j = 0, 1, \ldots, \eta \), which, however, involve moment \( M^{q}(t_{i}) \) with order \( q \) up to \( 2\eta \). To cope with this problem, one can employ moment closure techniques and approximate the higher order moments by functions of lower order moments up to order \( \eta \); see, for example, [32–34]. With that, the linear set of equations can be solved for \( \partial_{\eta q} \) and (4.1) can serve as an approximation to (3.2). We provide explicit expressions for (4.1) in the case of normal, lognormal and modified normal moment closure techniques in the electronic supplementary material, S.6.1. It is worth noting that the update formula of the celebrated Kalman filter turns out to be a special case of our approximation approach for the case of a normal closure combined with an additive Gaussian measurement model (electronic supplementary material, S.6.1).

We also provide an alternative approximation to posterior variance for that case. It is natural to assume that the reset step, which incorporates the information of a measurement, leads to a variance \( \Sigma(t) \) smaller than \( \Sigma(t_{i}) \). That is, at measurement point \( t_{i} \)
\[ \Sigma(t) = M^{2}(t_{i}) - |M^{1}(t_{i})|^{2} = m \Sigma(t_{i}), \]  
(4.3)

and hence \( M^{2}(t) = |M^{1}(t)|^{2} + m \Sigma(t_{i}) \), where coefficient \( m \) can be obtained by using (4.1) (see electronic supplementary material, S.6.4). For a multi-variate system, we define the first and the second filtering moments \( M(t) = \mathbb{E}[X(t)] \) and \( \Sigma(t) = \mathbb{E}[X(t)^{2}(t)] \) \( \in \mathbb{R}^{d} \) and \( M^{2}(t) = \mathbb{E}[X(t)X^{T}(t)] \) \( \in \mathbb{R}^{d \times d} \), respectively, together with the filtering covariance \( \Sigma(t) = M^{2}(t) - M(t)M^{T}(t) \). The corresponding moment dynamics for these quantities are detailed in the electronic supplementary material, S.7. The proposed approach can be applied to other observation models such as a lognormal noise model, which will subsequently be used in a gene expression model.

and the smoothing covariance by \( \hat{\Sigma}(t) = \hat{M}^{T}(t) - \hat{M}(t) \)
\[ \hat{M}(t) \in \mathbb{R}^{d \times d}. \]

Let us consider the first-order reaction networks. The prior mean \( M(t) \) of such a system is given by (equation (S48) with \( L = 1 \) in the electronic supplementary material, S.5.1)
\[ \frac{d \hat{M}(t)}{dt} = \hat{C}_{1} \hat{M}(t) + b_{1}. \]
(5.1)

For the gene expression model introduced below, one gets
\[ \hat{C}_{1} = \left( \begin{array}{c} c_{1} + c_{2} \\ c_{3} \\ c_{4} - c_{5} \end{array} \right) \quad \text{and} \quad b_{1} = \left( \begin{array}{c} c_{1} \\ 0 \end{array} \right). \]
(5.2)

Let \( \Psi(t, \tau) \) be the state transition matrix corresponding to \( \hat{C}_{1} \), that is, \( \frac{d}{d\tau} \Psi(t, \tau) = -\Psi(t, \tau) \hat{C}_{1} \) and \( \Psi(t, \tau) \big|_{t=t} = I \), which yields \( \Psi(t, \tau) = e^{(t-\tau)\hat{C}_{1}} \) for all \( t \geq \tau \). According to (2.5), the posterior smoothing and filtering moments have to coincide at the endpoint \( t_{N} \). Consequently, we initialize our approximation to \( M^{1}(t) \) by the approximate filtering moments at \( t_{N} \) obtained through (4.1) and (4.2). The resulting RTS smoother (see electronic supplementary material, S.8) is then given by the backward equation
\[ \hat{M}(t) = M(t) - K_{smt}(t)[M(t_{n+1}) - \hat{M}(t_{n+1})] \]
\[ \hat{\Sigma}(t) = \hat{\Sigma}(t) - K_{smt}(t)[\hat{\Sigma}(t_{n+1}) - \hat{\Sigma}(t_{n+1})]K_{smt}^{T}(t), \]

with smoother gain \( K_{smt}(t) = \hat{\Sigma}(t)\Psi^{T}(t_{n+1}, t)[\hat{\Sigma}(t_{n+1})]^{-1} \) and \( K_{smt}(t_{n+1}) = I \).

6. Application to gene expression models

Consider the standard two-state gene expression model consisting of six reactions that involve four different species: \( G_{1} \) and \( G_{0} \) for the active and inactive promoter, respectively, \( M \) for mRNA and \( P \) for the expressed protein
\[ G_{0} \xrightarrow{c_{1}} G_{1}, \quad G_{1} \xrightarrow{c_{2}} G_{1} + M, \]
\[ M \xrightarrow{c_{3}} P, \quad P \xrightarrow{c_{4}} \emptyset, \quad P \xrightarrow{c_{5}} \emptyset. \]

Through the conservation of active and inactive promoters, the state of the system can be represented by \( X = (X_{1} X_{2} X_{3})^{T} \in \mathbb{N}^{3} \), where \( X_{1} \), \( X_{2} \) and \( X_{3} \) are the amounts of \( G_{1} \), \( M \) and \( P \), respectively. A key problem in gene expression is to reconstruct the inaccessible states such as the mRNA abundance from noisy measurements of the protein abundance dynamics (e.g. through fluorescent labeling). Such state reconstruction is of particular interest for transient induction of genes, where the time-varying inducer can be modelled by a time-varying promoter activation rate \( c_{1} = c_{1}(t) \). Throughout this section, we assume lognormal measurement noise on the protein dynamics. That is,
\[ y_{i} = e^{y_{i}}X_{i}(t_{i}) = e^{y_{i}}FX_{i}(t_{i}), \]

with \( \mathbb{E}[y_{i}] = 0, \mathbb{E}[w_{i}^{2}] = \sigma_{i}^{2} \) and the corresponding observation matrix \( F = (0 \ 0 \ 1) \). In the following, we show that for synthetic data and for real single-cell data from \textit{Saccharomyces cerevisiae} the proposed method allows one to reconstruct robustly the mRNA abundance and true protein abundance from such noisy measurement trajectories. Before the next observation instant \( t_{i} \), the propagation of the moments is given by the prior moment equations (see, for example, [37] and also equation (S48) with \( L = 2 \) in the

5. An Rauch–Tung–Striebel approximation to the smoothing moments

Based on the proposed moment approximation to the filtering distribution, one can derive approximations to the moments of the smoothing distribution leveraging existing results for the Rauch–Tung–Striebel (RTS) or the modified Bryson–Frazier (MBF) smoother [35]. Here, we consider the RTS smoother because the RTS smoother has some desired properties such as stability and several other smoothers are based upon it [35,36]. Similarly, a reaction system with \( d \geq 1 \) denotes the first and the second smoothing moments by \( \hat{M}(t) = \mathbb{E}[X(t)] \) and \( \hat{\Sigma}(t) = \mathbb{E}[X(t)X^{T}(t)] \in \mathbb{R}^{d \times d} \)
The filtering and the smoothing moments are also computed using the RTS approximate for the smoothing moments. For reference, we compute the filtering moments and the exact ones are given in figure 2, where the prior moments are also given for comparison. Note that the measurement noise standard deviation $\sigma$ is larger than those in the experimental data used in this study, which were identified as $\sigma = 0.15$ in the Msn2 case and $\sigma = 0.125$ in the GEV case below, respectively. The approximation to the filtering moments and to the smoothing moments is in good agreement with the exact results. Moreover, one can observe that the actual mRNA and protein dynamics corresponding to the actual measured data points can accurately be tracked by the respective posterior means. From the above discussion, it is evident that the traditional prior dynamics cannot provide this single-trajectory resolution. In the electronic supplementary material, S.7.3, we apply the proposed moment approximation to another in silico problem that exhibits bimodal distributions. It indicates that the posterior distribution of multi-modal systems can often be unimodal due to the conditioning and can hence be approximated well by low-order moment equations.

6.2. Gene expression systems in yeast

We apply our proposed state reconstruction approach to two different inducible gene expression systems in *Saccharomyces cerevisiae*. Both systems can be described by the above gene expression model with a time-varying promoter activation rate $c_t$. In the first system, a microfluidic device is used to control the expression model with a time-varying promoter activation rate caused by the nuclear translocation of an inducer.

In the first system, a microfluidic device is used to control the nuclear–cytoplasmic translocation dynamics of the transcription factor Msn2 by modulating the levels of the small molecule 1-NM-PP1 to control the expression of a fluorescent reporter protein (see [6] for a detailed description; subsequently referred to as the Msn2 system). The second system is an artificial gene expression system centred around the chimeric transcription factor Msn2. We apply our proposed approach to the two systems and obtain the filtering and the smoothing moments. The comparison between the approximate moments and the exact ones is given in figure 2, where the prior moments are also given for comparison. Note that the measurement noise standard deviation $\sigma$ is larger than those in the experimental data used in this study, which were identified as $\sigma = 0.15$ in the Msn2 case and $\sigma = 0.125$ in the GEV case below, respectively. The approximation to the filtering moments and to the smoothing moments is in good agreement with the exact results. Moreover, one can observe that the actual mRNA and protein dynamics corresponding to the actual measured data points can accurately be tracked by the respective posterior means. From the above discussion, it is evident that the traditional prior dynamics cannot provide this single-trajectory resolution. In the electronic supplementary material, S.7.3, we apply the proposed moment approximation to another in silico problem that exhibits bimodal distributions. It indicates that the posterior distribution of multi-modal systems can often be unimodal due to the conditioning and can hence be approximated well by low-order moment equations.

6.1. In silico experiments

We first apply our proposed approach to the gene expression model using simulated non-stationary data and, for simplicity, assume a constant reaction rate $c_t$. Based on the obtained filtering moments, we compute the RTS approximate for the smoothing moments. For reference, the filtering and the smoothing moments are also computed exactly by integration of the corresponding conditional master equation. The comparison between the approximate moments and the exact ones is given in figure 2, where the prior moments are also given for comparison. Note that the measurement noise standard deviation $\sigma$ is larger than those in the experimental data used in this study, which were identified as $\sigma = 0.15$ in the Msn2 case and $\sigma = 0.125$ in the GEV case below, respectively. The approximation to the filtering moments and to the smoothing moments is in good agreement with the exact results. Moreover, one can observe that the actual mRNA and protein dynamics corresponding to the actual measured data points can accurately be tracked by the respective posterior means. From the above discussion, it is evident that the traditional prior dynamics cannot provide this single-trajectory resolution. In the electronic supplementary material, S.7.3, we apply the proposed moment approximation to another in silico problem that exhibits bimodal distributions. It indicates that the posterior distribution of multi-modal systems can often be unimodal due to the conditioning and can hence be approximated well by low-order moment equations.
factor GAL4DBD.ER.VP16 (GEV). The GEV translocation is again modulated using a microfluidic device controlling the supply of the hormone β-oestradiol. The nuclear transcription factor GEV activates the transcription of genes under a GAL1 promoter, where we placed a fluorescent reporter protein as a readout (see [7] for a detailed description; subsequently referred to as the GEV system). In both model systems, fluorescent time-lapse microscopy is used to monitor the nuclear–cytoplasmic translocation of the respective transcription factor fused to a fluorescent protein, as well as the expression of a fluorescent protein induced by the respective transcription factor in individual cells.

The two case studies are successively complicated. In the Msn2 case, we associate for every single-cell trace a separate parameter set. This is feasible due to a sufficient number of observations for each trace. The problem thus corresponds to the in silico study from above and no extrinsic noise model is assumed that couples parameter sets from different traces. In most scenarios, however, one needs to pool together heterogeneous traces in order to achieve sufficient estimation accuracy. To illustrate this complication we show, for the GEV case, how an extrinsic noise model can be incorporated into the proposed filtering or smoothing method. The extension is in line with the observation that the GEV expression variability shows a large extrinsic component [7].

6.3. Approximate state reconstruction for Msn2 system

To demonstrate the effectiveness of the proposed method, we reconstruct the mRNA dynamics based on the noisy fluorescent readout of the protein level. For the first case, we used single-cell traces from [6] to estimate the parameters of the kinetic expression model and the measurement noise model using the algorithm given in [7]. The temporal profile of the promoter activation rate is estimated from the nuclear concentration of the transcription factor (see also [6,7] for details). We remark that this induction happens quite rapidly (figure 3). Generally, it is challenging to perform state estimation for such fast varying systems.

Figure 3 shows the reconstruction results for two exemplary single-cell traces of the dataset in [6]. Applying finite-state-projection techniques [38,39], we were also able to compute the exact moments from solving posterior master equations (2.3) and (2.8) as a reference. However, for larger systems, solving the corresponding master equation quickly becomes computationally infeasible—motivating our approximate approach. Figure 3 indicates that the proposed approximation to the filtering moments works well. As it only involves solving a set of ordinary differential equations with reset conditions the approach is scalable to large reaction systems.

Similarly, we applied the proposed smoothing algorithm to the single-cell trajectories. In particular, we used the presented RTS approximation to the smoothing moments. However, it is observed that the traditional RTS approximation does not always work for the considered experimental data. Figure 4 shows that the approximation works well for trajectory B while it fails for trajectory A where one can observe a collapse of the smoothing covariance for the approximate method even though the involved approximate filtering moments are accurate (cf. figure 3). This indicates that novel approximate methods for the smoothing moments are needed to overcome the limitations of traditional RTS schemes.

6.4. State reconstruction in the presence of extrinsic noise

Apart from the inherent randomness of biomolecular reactions, gene expression was shown to exhibit a substantial degree of extrinsic noise [5,40–42], stemming from various factors in a cell’s microenvironment. The considered GEV system showed
significant extrinsic noise [7] and we therefore minimally extend the standard gene expression system by a random protein translation rate \( c_4 \) assumed to be a gamma distribution characterized by shape and rate parameters \( a \) and \( b \). In contrast with the Msn2 case study, where we assumed the model parameters to be given, we aim to estimate states and parameters—for the latter in particular the population heterogeneity captured by \((\tilde{a}, \tilde{b})\).

Treating a parameter as just another state that also follows a distribution serves as a Bayesian point estimate. After receiving the data, the prior heterogeneity can be turned into a posterior by conditioning on all recorded time traces. To obtain such a posterior over states, we employ Markov chain Monte Carlo techniques to sample

\[
p(a, b | y_{1:N}^{1}, \ldots, y_{L:N}^{1}) \propto \prod_{m=1}^{L} p(y_{1:N}^{m} | a, b) p(a, b),
\]

with \( y_{1:N}^{m} = \{y_{1}^{m}, \ldots, y_{N}^{m}\} \) the measured trace corresponding to cell \( m \), measured at \( N \) time points. The mean value of this distribution serves as a Bayesian point estimate \((\tilde{a}, \tilde{b})\) that we can then use to determine the posterior moments. When receiving a new single-cell trajectory \( y_{1:N}^{1} \) one can now ask for its most likely mRNA and protein dynamics given this posterior heterogeneity. That is, we aim to compute

\[
\tilde{M}^{n}(t) = E[E[X^{n}(t) | c_4, y_{1:N}^{1}],\]

where the outer expectation is over \(c_4 \sim \Gamma(\tilde{a}, \tilde{b})\). To approximate these posterior moments, we combine our proposed smoothing approach with [3] to obtain a posterior marginal moment equation. Thereby, we follow again the RTS ansatz and use the exact moments obtained from integrating the smoothing master equation (2.8) as a reference. Conceptually, the resulting marginal moments are equivalent to averaging the traditional smoothing moments for random \( c_4 \) drawn from the gamma distribution. The results for one exemplary single-cell trajectory are given in figure 5. It is observed that the RTS approximation to the smoothing moments is accurate for the considered data of the GEV system. Also evident from the comparison in figure 5 is the significant reduction of variance of the posterior moments with respect to the prior moment dynamics.

### 7. Conclusion

Our capacity to decipher the inner workings of a cellular process strongly depends on the dimensionality of the available molecular readout. For time-resolved single-cell analysis the number of simultaneous readouts remains limited and biologists are trained to qualitatively infer the behaviour of unobservable states of the process. However, with the rise of the computational models that can quantitatively capture the behaviour processes, one can now improve on this qualitative inference. We remark, again, that estimating the most likely latent state of the process for a given observation is different from just computing the solution to a calibrated model. The theory of optimal filtering offers the general solution to the problem on how to combine data with a dynamic model to predict such states. However, it is known that solving the exact filtering or smoothing problem.

---

**Figure 4.** Comparison between the RTS approximations and the exact smoothing moments based on two exemplary trajectories from the experimental dataset of [6]: (a,b) protein and mRNA dynamics corresponding to trajectory A; (c,d) protein and mRNA dynamics corresponding to trajectory B.
is computationally costly and we show that for a biochemical network it is at least as costly as integrating the chemical master equation.

To this end, we develop an approximate but scalable approach to filtering by exploring the fundamental relationship [43] and combining it with traditional moment closure techniques. We verify the effectiveness of the proposed method through single-cell experimental data and through \textit{in silico} experiments. Based on the approximation to the filtering moments obtained by the method, one can further compute the RTS approximation to the smoothing moments. Although the RTS approximation often works well (see figures 4c,d and 5), it also show significant deviations (e.g. figure 4), even when the proposed approximation to the filtering moments performs well (figure 3). This lack of robustness indicates that a novel approximation to the smoothing moments needs to be developed for the case of stochastic reaction networks with lognormal measurement noise.

We observe that the imposed conditioning often yields posterior processes with confined, unimodal distributions even if the corresponding prior process exhibits multi-modal distributions. Hence, approximations through moment closure or state-space truncation of posterior processes for complex systems such as multi-stable or oscillatory networks [44] appears feasible and is a promising research question. Moreover, the proposed method can be extended to a hybrid framework (e.g. [45–47]) where a diffusion approximation can be performed for some states and reactions. This is especially interesting for multi-scale cellular processes, for instance in signal transduction coupled to gene expression where different abundance scales of molecules are involved. The proposed state reconstruction approach can profit from such a hybridization and would lead to even more scalable algorithms. Because it is well known that the moment closure techniques can also fail [33,48], theoretical analysis, such as the computation of error bounds, for the special case of posterior moments is a promising future research topic. The developed moment approximations of the filtering distribution can also be used for the stochastic decoupling of networks as proposed in [49].

8. Methods and experimental protocols

8.1. Mathematical methods and algorithms

Details on mathematical derivations of the posterior master equations, the posterior moment equations and details on the discussed case studies together with more corresponding simulation results are given in the electronic supplementary material. The MATLAB codes used to generate all results in this paper are available at http://www.bcs.tu-darmstadt.de/media/bcs/Reconstructing_dynamic_molecular_states.zip.

8.2. Calibrating YFP fluorescence to absolute numbers of molecules

The previous work [6] quantified induction of Msn2 target genes as the mean fluorescence intensity per pixel in arbitrary units (arb. units) through a fast-maturing YFP reporter protein, mCitrineV163A. However, in order to apply the state reconstruction approach, it is necessary to calibrate these measurements to obtain absolute numbers of YFP molecules per cell. A calibration relationship was developed by measuring the mCitrineV163A fluorescence of five yeast proteins of known abundance [50] using the same exposure conditions as in the original study [6]. The five yeast genes were: YGP117C (1280 molecules per cell), TMA108 (5110 molecules per cell), HOG1 (6780 molecules per cell), TDA1 (10 200 molecules per cell) and CAR1 (42 800 molecules per cell). Each gene was C-terminally tagged with mCitrineV163A (in a
pKT-vector; available from AddGene as no. 64685) followed by a His-marker by transforming a polymerase chain reaction-generated mCitrineV163A-HIS construct into the original haploid S288C Saccharomyces cerevisiae strain used by Ghaemmaghami et al. [51] (EY0986, MAT a, ATCC201388, his3Δ1, leu2Δ0, met150, ura3Δ0, S288C) and selecting on SD-HIS plates. To minimize experimental variability, we picked and measured four independent clones for each of the five genes. We used the untagged wild-type strain EY0986 to determine the autofluorescence background. To measure fluorescence intensity from each clone, we closely followed the protocol described by Ghaemmaghami et al. Briefly, we picked a single colony to inoculate a flask containing yeast extract peptone dextrose medium. Cells were grown overnight until an OD600 of approximately 0.7. Cells were fixed by incubating 0.9 ml of culture with 0.1 ml 10% buffered formalin solution (Sigma-Aldrich HT5011) for 5 min with occasional mixing. Cells were spun down and washed with 0.1 M KH2PO4 pH 8.5 and then 1.2 M sorbitol in KH2PO4 pH 8.5. Cells were resuspended in 20 μl 1.2 M sorbitol in KH2PO4 pH 8.5. Then 2 μl of this solution was loaded onto a microscope slide, a coverslip added and sealed with nail polish. Cells were then immediately imaged using the exact same exposure conditions as described in [6]. Images were then analysed and fluorescence quantified as previously described [6,51]. After quantifying the fluorescence intensity per cell for each of the five genes, we then fitted a simple line to the data and found that each YFP molecule contributed about 100.8 arb. units fluorescence. The expression levels of the YFP-tagged proteins were measured with illumination conditions similar to those used for the Y-Venus imaging. See [7] for a detailed description.

8.3. Fluorescence microscopy and image analysis in pGAL1 Y-Venus expression

The experiments were performed on the same epifluorescence microscope (Eclipse Ti, Nikon Instruments), 60x (NA1.4) oil objective and specific (CFP/YFP/mCherry) excitation and emission filters located in an incubation chamber set to maintain 30°C. Imaging conditions and parameters were kept constant for all experiments. Single colonies of the respective yeast strain were picked, inoculated in synthetic (SD) medium and grown overnight at 30°C. The saturated cultures were then diluted and grown in log phase for at least two doubling times (more than 4 h). Before they were loaded into the imaging chambers, the cell suspensions were diluted again (OD600 = 0.01) and briefly sonicated. Single-cell traces were recorded by fluorescence microscopy with a 30 min induction pulse of 25, 50 and 100 nM β-oidiol. The pulses were done by switching between two hydrostatic pressure (1 p.s.i.) driven flows (SD-full and SD-full + β-oidiol) using a three-way solenoid valve (The Lee Company) connected to the cell chamber (μ-Slide VI; Ibidi). All microscopy images were analysed with the Yeast-Quant platform. The GEV relocation and Venus expression time-lapse movies were segmented on the basis of the nuclear CFP image from the HTA2-CFP marker. The expression of the Y-Venus protein was quantified as the total intensity in the cell. The expression levels of the YFP-tagged proteins were measured with illumination conditions similar to those used for the Y-Venus imaging. See [7] for a detailed description.

References


