Adaptive moment closure for parameter inference of biochemical reaction networks

Sergiy Bogomolov[1], Thomas A. Henzinger[1], Andreas Podelski[2], Jakob Ruess[1] and Christian Schilling[2]

[1] IST Austria, Austria
[2] University of Freiburg, Germany

Abstract. Continuous-time Markov chain (CTMC) models have become a central tool for understanding the dynamics of complex reaction networks and the importance of stochasticity in the underlying biochemical processes. When such models are employed to answer questions in applications, in order to ensure that the model provides a sufficiently accurate representation of the real system, it is of vital importance that the model parameters are inferred from real measured data. This, however, is often a formidable task and all of the existing methods fail in one case or the other, usually because the underlying CTMC model is high-dimensional and computationally difficult to analyze. The parameter inference methods that tend to scale best in the dimension of the CTMC are based on so-called moment closure approximations. However, there exists a large number of different moment closure approximations and it is typically hard to say a priori which of the approximations is the most suitable for the inference procedure. Here, we propose a moment-based parameter inference method that automatically chooses the most appropriate moment closure method. Accordingly, contrary to existing methods, the user is not required to be experienced in moment closure techniques. In addition to that, our method adaptively changes the approximation during the parameter inference to ensure that always the best approximation is used, even in cases where different approximations are best in different regions of the parameter space.

Keywords: Stochastic reaction networks, continuous-time Markov chains, parameter inference, moment closure

Acknowledgements. This work was partly supported by the German Research Foundation (DFG) as part of the Transregional Collaborative Research Center "Automatic Verification and Analysis of Complex Systems" (SFB/TR 14 AVACS[3]), by the European Research Council (ERC) under grant 267989 (QUAREM) and by the Austrian Science Fund (FWF) under grants S11402-N23 (RiSE) and Z211-N23 (Wittgenstein Award). J.R. acknowledges support from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement no. 291734.

[3] `http://www.avacs.org/`

1 Introduction

With the advancement of measurement technologies for biochemical processes in the last decades, quantitative mathematical modeling of biochemical reaction networks has continuously increased in importance [1,14,19]. Chemical reactions inside cells, where some of the reacting species may be present in very low amounts of molecules, are inherently driven by random fluctuations [6,12,20]. Accordingly, an accurate mathematical model should take this stochasticity into account. The most widely used class of stochastic models in this context are continuous-time Markov chains (CTMCs) [5]. The advantage of these models is that they are easy to formulate and can be justified based on first principles [4]. The major drawback is that their analytical or computational analysis can be extremely difficult, especially when more than just a few different chemical species play a role for the reaction network. This is because the chemical master equation (CME), which governs the time evolution of the probability distribution of the CTMC, cannot be solved for anything but the simplest systems and even approximation techniques [13,23] tend to fail when the CTMC is high-dimensional. In such cases, an alternative is to focus only on some low-order moments of the probability distribution. Ordinary differential equations that describe the time evolution of these moments can be derived from the CME [2], but their solution typically requires some kind of approximation [18,21]. These approximations, known as moment closure, are usually based on an assumption of the underlying probability distribution and exist in many different varieties [8]. Often, for a given system and given model parameters, some of these approximations provide good results whereas others fail to be sufficiently accurate or fail entirely. Unfortunately, there exists no approach for determining a priori which moment closure technique will provide the best approximation. In general, the only approach that is guaranteed to provide at least statistically exact results is to simply simulate the CTMC using a stochastic simulation algorithm (SSA) [3] and to compute Monte Carlo estimates of the system output of interest based on the simulation results. To obtain precise estimates, however, a large number of simulations may be required, leading to a high computational cost. For the forward analysis of a system, i.e. when the model parameters are known, this is not a serious problem. For the reverse engineering task of identifying the model parameters from measured data, however, the CTMC needs to be analyzed for many different parameter values in order to determine those in best agreement with the measured data. Accordingly, for this task the computational cost of approaches based on stochastic simulation [10] is often prohibitively large.

In this paper, we propose an approach for parameter inference based on moment closure that is complemented by stochastic simulation. In particular, the parameter inference is performed based on the computationally cheap moment closure approximation, whereas the stochastic simulation is employed whenever new regions in the parameter space are explored, either to ensure that the approximation is still sufficiently accurate, or to propose a new approximation that outperforms the previously used one. With this approach we are able to combine the computational advantages of moment closure with the statistical exactness of SSA and obtain a method that is both scalable and does not require a priori knowledge of the performance of different moment closure techniques. Importantly, the method is completely automated and chooses and adapts the approximation from a precomputed library of moment closure methods. Thus, the user only has to specify the model and supply the data and, contrary to previous approaches [9,16,24], no expertise in the analysis of CTMCs is required.

The remaining paper is structured as follows. In Section 2, we introduce biochemical reaction networks, the chemical master equation and moment closure methods. In Section 3, we formulate a maximum-likelihood estimation problem for the model parameters and describe previously published moment-based methods for solving these problems. In Section 4, we propose our automated adaptive parameter inference method. In Section 5, we study the performance of our method for some benchmark reaction networks. Finally, in Section 6, we discuss our results and provide some concluding remarks.

2 Stochastic modeling of biochemical reaction networks

Consider a biochemical reaction network consisting of m different chemical species $X^1, \ldots, X^m$ that interact according to K different reactions:

$$\nu_{1k}^{'} X^1 + \ldots + \nu_{mk}^{'} X^m \xrightarrow{\theta_k} \nu_{1k}^{''} X^1 + \ldots + \nu_{mk}^{''} X^m, \quad k = 1, \ldots, K, \tag{1}$$

where the coefficients $\nu_{ik}^{'}$ and $\nu_{ik}^{''}$ determine how many molecules of the i-th species are consumed and produced in the k-th reaction, respectively. Under the assumption that the reaction network is well-stirred and in thermal equilibrium, it can be described by a continuous-time Markov chain $X(t,\theta) = \left[X^1(t,\theta) \cdots X^m(t,\theta)\right]^T$ that takes states $x = [x^1 \cdots x^m]^T \in \mathbb{N}_0^m$ [4]. The transition probabilities of this CTMC are determined by the reaction parameters $\theta = [\theta_1 \cdots \theta_K]^T \in \left(\mathbb{R}_0^+\right)^K$ and the kinetic rate law of the reactions. Here, we restrict our attention to mass action kinetics and elementary chemical reactions (i.e. reactions of order at most 2). These assumptions simplify the computation of moments of the CTMC. It should be noted, however, that they are not strictly necessary for the results of this paper and are mainly imposed because it is very unlikely that, in a three-dimensional space, more than two molecules meet at exactly the same time. Accordingly, any more complicated biochemical reaction can essentially be decomposed into a series of elementary reactions whose reaction rates are governed by the law of mass action. These assumptions lead to transition probabilities of the CTMC that are determined by propensity functions of the form $a_k(x,\theta) = \theta_k h_k(x), k = 1, \ldots, K$, where $h_k(x)$ are at most quadratic polynomials in x. The time evolution of the probability distribution of $X(t,\theta)$ can then be described by the chemical master equation:

$$\dot{p}(x,t) = -p(x,t) \sum_{k=1}^{K} a_k(x,\theta) + \sum_{k=1}^{K} p(x-\nu_k,t) a_k(x-\nu_k,\theta), \tag{2}$$

where $\nu_k = [\nu_{1k} \cdots \nu_{mk}]^T$, $\nu_{ik} = \nu_{ik}^{''} - \nu_{ik}^{'}$, $i = 1, \ldots, m$, and $p(x,t) := P(X(t,\theta) = x)$ is the probability that x molecules of the m chemical species are present at time t.
Since $X(t,\theta)$ has a countably infinite state space, computing the probabilities $p(x,t)$ requires solving an infinite system of coupled ordinary differential equations, which is generally not possible. Approximate solutions can be obtained in some cases, for instance by projection to a finite state space [13,23], but we will not discuss these approaches here.
An alternative is to focus only on some low-order moments of the probability distribution. Ordinary differential equations describing their time evolution can be derived from the CME [2] and written as

$$\dot{\eta}(t) = A(\theta)\eta(t) + B(\theta)\bar{\eta}(t), \tag{3}$$

where $\eta(t)$ is a vector containing the (uncentered) moments up to some desired order L and $\bar{\eta}(t)$ contains moments of order $L+1$. Eq. (3) shows that the time evolution of $\eta(t)$ depends on moments of higher order; hence $\eta(t)$ cannot be computed without knowledge of $\bar{\eta}(t)$. Accordingly, the open system of equations Eq. (3) is typically replaced by an approximate closed system of equations

$$\dot{\tilde{\eta}}(t) = A(\theta)\tilde{\eta}(t) + B(\theta)f(\tilde{\eta}(t)), \tag{4}$$

where $\tilde{\eta}(t)$ are approximations of $\eta(t)$. The function f is usually chosen according to an assumption on the underlying probability distribution. Typical examples are to assume that the centered moments (or cumulants) of order $L+1$ are zero [22,11], or to choose f according to a log-normal distribution [21]. In general, the choice of f is made rather arbitrarily without actual knowledge of the underlying distribution. Furthermore, whether a given closure will provide good approximations depends on the system that is being studied, the model parameters, and the order L at which the moment equations are closed. This makes it practically impossible for someone who is not an expert in the use of these methods to choose an appropriate closure. Despite all this, moment closure methods have been successfully applied for analyzing CTMCs, and specifically also for parameter inference [16,24]. The choice of the closure method used in these references, however, was based on trial and error and the success of the performed studies accordingly required a portion of luck.

An alternative approach for analyzing biochemical reaction networks is by using a stochastic simulation algorithm (SSA). It is straightforward to generate statistically exact sample paths $x_1(t), \ldots, x_n(t)$ of $X(t, \theta)$ in this way. From these sample paths, estimators of any system output, for instance some moments or the entire probability distribution at a certain time point, can be constructed. While such an approach is easy to implement and can always be used, it comes with the major drawback that often a large number of sample paths n is required to obtain precise estimates. This can make the use of stochastic simulation for reverse engineering tasks computationally prohibitively expensive.

3 Moment-based parameter inference

In this section, we formulate the parameter inference problem and review previous methods that have been developed to solve it. The goal in this paper is to estimate the reaction rate constants θ from measured data that is of the form $y = \left\{x_1^j(t_s), \ldots, x_n^j(t_s),\ s = 1, \ldots, S\right\}$ and corresponds to measuring the number of molecules of the j-th chemical species in n cells at each measurement time point $t_s, s = 1, \ldots, S$ (extension to more than one measured chemical species is straightforward but requires more complicated expressions for the likelihood in Eq. 6 as shown in [17]). We assume that all the collected measurements are statistically independent. This is for instance the case for flow cytometry data where the cells are discarded after being measured so that two different measurements can never come from the same cell. The task of identifying the model parameters from this data can be posed as a maximum-likelihood estimation problem

$$\theta_{\mathrm{MLE}}(y) = \arg\max_{\theta} \mathcal{L}(y, \theta), \tag{5}$$

where y is the measured data and $\mathcal{L}(y, \theta) = p(y|\theta)$ is the likelihood of the parameters θ, i.e. the probability (density) of the data given that θ are the model parameters. Analytically computing the likelihood is usually impossible, and accordingly, the optimization

problem Eq. (5) is typically solved by iterative numerical evaluation of $\mathcal{L}(y,\theta)$ for many different values of θ. Unfortunately, evaluating the likelihood for given parameters θ requires solving the CME with these parameters, which, as discussed in the previous section, is often impossible or computationally expensive itself. For this reason, one option is to use sample moments of the data as measurements instead of the entire data [24]. For instance, one can compute sample means $\hat{\mu}_1(t_s)$ and sample variances $\hat{\mu}_2(t_s), s = 1,\ldots,S$ from the data y and treat the vector $\hat{\mu} := [\hat{\mu}(t_1)\cdots\hat{\mu}(t_S)]^T$, where $\hat{\mu}(t_s) := [\hat{\mu}_1(t_s)\ \hat{\mu}_2(t_s)]$, as new data. In earlier publications [24,17], we have shown that the probability density function $p(\hat{\mu}|\theta)$ of $\hat{\mu}$ is given by

$$p(\hat{\mu}|\theta) = \prod_{s=1}^{S} p(\hat{\mu}(t_s)|\theta), \quad \text{where} \quad p(\hat{\mu}(t_s)|\theta) = \mathcal{N}(M(t_s),\Sigma(t_s)) \quad \text{and} \tag{6}$$

$$M(t_s) = \begin{bmatrix}\mu_1(t_s)\\ \mu_2(t_s)\end{bmatrix} \quad \text{and} \quad \Sigma(t_s) = \frac{1}{n}\begin{bmatrix}\mu_2(t_s) & \mu_3(t_s)\\ \mu_3(t_s) & \mu_4(t_s) - \frac{n-3}{n-1}\left(\mu_2(t_s)\right)^2\end{bmatrix},$$

where $\mathcal{N}$ stands for the normal distribution, $\mu_1(t_s) = \mu_1(t_s,\theta)$ is the mean and $\mu_i(t_s) = \mu_i(t_s,\theta), i = 2,3,4$ are the centered moments of the measured species $X^j(t_s,\theta)$ at time t_s for model parameters θ. Since these moments can be computed from the solution of Eq. (4), we can use this result to approximately compute the likelihood $\mathcal{L}(\hat{\mu},\theta) = p(\hat{\mu}|\theta)$ without having to solve the CME. Accordingly, we can solve the optimization problem in Eq. (5) using $\hat{\mu}$ instead of y to compute the maximum-likelihood estimator $\theta_{\mathrm{MLE}}(\hat{\mu})$. However, the fact that moments up to order four are required to evaluate the covariance matrices $\Sigma(t_s)$ means that moment closure of order at least $L = 4$ is necessary. To avoid this, one can estimate the covariance matrices $\Sigma(t_s)$ from the data by computing empirical estimates of the moments up to order four and plugging them into the above equation. Throughout this paper, we will follow such a strategy and denote by μ_{data} the moments up to fourth order of the data, i.e. $\mu_{\mathrm{data}} := [\mu_{\mathrm{data}}(t_1)\cdots\mu_{\mathrm{data}}(t_S)]^T$, where $\mu_{\mathrm{data}}(t_s) := [\hat{\mu}_1(t_s)\ \hat{\mu}_2(t_s)\ \hat{\mu}_3(t_s)\ \hat{\mu}_4(t_s)]$ contains the first four centered empirical moments of the data set at time t_s. This strategy is appropriate whenever sufficiently many cells are measured so that the moments up to order four can be estimated with reasonable precision. For flow cytometry data, the number of cells measured per time point typically ranges in the order of thousands or even tens of thousands; hence sufficing precision is always guaranteed.

4 Adaptive approach for parameter inference

The drawback of the approach described in the previous section is that a moment closure method has to be chosen in advance and this closure will be used throughout the entire parameter search. This leads to the problems that, on the one hand, it is a priori very difficult to choose the best closure and, on the other hand, which closure is best may also be different for different parts of the parameter space. The main idea of the method that we propose in the following is to use a small number of simulated trajectories of the system that are generated using a stochastic simulation algorithm (SSA) in order to test different approximations during the parameter space exploration. Specifically, whenever the parameter search leaves a certain area in parameter space, defined as an ϵ-neighborhood around the point at which the last SSA run was carried out, new simulations are performed and all closure methods from a predefined library are evaluated by comparing the different approximations at the current point in parameter space to the simulation results. Importantly, all the approximate moment systems,

corresponding to closures of different types and degrees, are precomputed only once, and thus new derivations of the moment equations are not required during the search. To generate these systems we make use of Hespanha's StochDynTools toolbox [7].

Pseudocode of our approach is given in Algorithm 1. The inputs of the algorithm are the CTMC model $X(t, \theta)$, parametrized by the reaction rate constants θ, a set of ODE systems $CL = \{c_1(\theta), \ldots, c_q(\theta)\}$ corresponding to different approximations of the moment dynamics obtained through various closures of different types and degrees, the centered moments up to the fourth order μ_{data} of a measured data set Y, and a maximal number of iterations $i_{\max}$ that determines for how many steps in parameter space the search is performed. The algorithm returns the maximum likelihood estimator θ_{MLE}. The core idea of our approach works independently from the actual parameter search technique used in the background. Thus, it can be applied in conjunction with any standard optimization scheme used to minimize some distance between model output and data (for instance simple gradient descent). Accordingly, we focus on the adaptive update of the closure method while abstracting from the actual details of the parameter search for a fixed approximation by the function NextParameter (line 18). It takes the current values of the parameters θ_i and the chosen approximate ODE system $c_{\text{best}}(\theta_i)$ and moves the search to the new parameters θ_{i+1} according to some criteria. In our implementation, we instantiate it with a Markov chain Monte Carlo method and a Metropolis-Hastings sampler, based on the likelihood in Eq. (6) [24]. Additionally, this function also takes care of updating the value of the maximum likelihood estimator θ_{MLE} based on the likelihood of the new parameters θ_{i+1}. The remaining pseudocode describes how and when the used closure method is adjusted. We first check whether the current parameter values θ_i are still within the ϵ-neighborhood $N_\epsilon(\theta_{\text{ref}})$, where θ_{ref} are the parameters at which the previous simulation was performed (line 5). In our implementation, we choose a neighborhood in the form of a hyperrectangle of relative size $N_\epsilon(\theta_{\text{ref}}) = \{\theta \mid |\theta - \theta_{\text{ref}}|_k \leq \epsilon \cdot |\theta_{\text{ref}}|_k, k = 1, \ldots, K\}$. If $\theta_i \in N_\epsilon(\theta_{\text{ref}})$, we directly proceed with the standard inference method in line 18, relying on the ODE system $c_{\text{best}}(\theta_i)$ from the most recent evaluation. Otherwise, stochastic simulation is employed with the current parameter values θ_i to compute estimates of the moments $\mu_{\text{SSA}}(\theta_i)$ using the function $\text{Compute}_{\text{SSA}}$ (line 6), for which we utilize a standard implementation of Gillespie's SSA in our implementation. These estimates are then compared to the approximations $\mu_{\text{ODE}}(\theta_i)$ obtained with all the different closure methods using the function $\text{Compute}_{\text{ODE}}$ which numerically computes the solution of the system of ODEs $c(\theta_i) \in CL$ (lines 8-15). The best approximate system $c_{\text{best}}(\theta_i)$ is chosen as the one that minimizes some distance Dist between estimation and approximation. In general, this distance could be determined in many different ways. In our implementation, we choose Dist as the likelihood of the estimated moments for the measured species X^j (Eq. 6), i.e. we measure the performance of the approximations by evaluating how precise the approximated moments of the system output (not of the entire state) are. Finally, we update the reference point θ_{ref} to θ_i (line 16) and the search continues in the standard way until the next ϵ-neighborhood is left.

5 Case studies

We applied our inference method to several benchmark stochastic reaction networks. In this section, we report some exemplary results. For all examples, to generate the set of approximate ODE systems CL we used *derivative matching* (dm), *zero cumulants* (zc),

Algorithm 1 Adaptive moment-based parameter inference algorithm

Input: CTMC $X(t,\theta)$, where $\theta \in (\mathbb{R}_0^+)^K$, set of approximate moment systems $CL = \{c_1(\theta), \ldots, c_q(\theta)\}$ obtained using different closure methods, data μ_{data}, and maximum number of iterations $i_{\max}$

Output: Maximum likelihood estimator θ_{MLE}

1: $\theta_1 :=$ random initial parameter values
2: $\theta_{\text{MLE}} := \theta_1$
3: $\theta_{\text{ref}} := +\infty$
4: **for** $i := 1$ **to** $i_{\max}$ **do**
5: **if** $\theta_i \notin N_\epsilon(\theta_{\text{ref}})$ **then**
6: $\mu_{\text{SSA}}(\theta_i) := \textsc{ComputeSSA}(X(t,\theta_i))$
7: $d_{\text{best}} := +\infty$
8: **for all** $c(\theta_i) \in CL$ **do**
9: $\mu_{\text{ODE}}(\theta_i) := \textsc{ComputeODE}(c(\theta_i))$
10: $d := \textsc{Dist}(\mu_{\text{SSA}}(\theta_i), \mu_{\text{ODE}}(\theta_i))$
11: **if** $d < d_{\text{best}}$ **then**
12: $d_{\text{best}} := d$
13: $c_{\text{best}}(\theta_i) := c(\theta_i)$
14: **end if**
15: **end for**
16: $\theta_{\text{ref}} := \theta_i$
17: **end if**
18: $\langle \theta_{i+1}, \theta_{\text{MLE}} \rangle := \textsc{NextParameter}(\theta_i, c_{\text{best}}(\theta_i), \mu_{\text{data}}, \theta_{\text{MLE}})$
19: **end for**
20: **return** θ_{MLE}

zero variance (zv) moment closure, each with degree 2, 3, and 4, and *low dispersion* (ld) moment closure with degree 3 and 4 (see [8] for details).

Example 1. The first network is a model that has recently been used to describe agricultural pests [15] but can also be regarded as a model of gene expression in which the produced protein is positively regulated by the current amount of protein and negatively regulated (through an increased degradation rate) by past amounts of protein (i.e. species N could be regarded as an abstraction of a slow process that is activated by C and leads to the production of proteases that degrade C). It is given by the following reactions:

$$\emptyset \xrightarrow{\theta_1} N + C \qquad N \xrightarrow{\theta_2} 2N + C$$
$$N + C \xrightarrow{\theta_3} C \qquad C \xrightarrow{\theta_4} \emptyset.$$

We assume that $N(0) = C(0) = 0$ and that the true parameters are given by $\theta_1 = 0.03$, $\theta_2 = 0.012$, $\theta_3 = 0.25 \cdot 10^{-4}$ and $\theta_4 = 0.003$, and that 5,000 cells are measured at the time points $t_1 = 10, \ldots, t_{90} = 900$. As settings for our algorithm we used $\epsilon = 0.2$ and performed 200 simulations whenever the search leaves an ϵ-neighborhood, i.e. in line 6 of Algorithm 1.

An exemplary run of our parameter search for $i_{\max} = 1{,}000$ iterations, started from random initial parameter values, is shown in Figure 1. It can be seen that all the inferred parameters, i.e. the maximum-likelihood estimates $\theta_{\text{MLE}}(\hat{\mu})$, agree with the

true parameter values up to negligible errors with basically no uncertainty. The former is a sign that a precise moment closure method exists for this example, whereas the latter stems from the large number of measurements that we assumed to be available. Figure 2 shows that also the model predictions, computed with the inferred parameters $\theta_{\mathrm{MLE}}(\hat{\mu})$ and the best closure method, agree well both with the data and with SSA estimates of mean and variance obtained with the inferred parameters. We can conclude that the moment closure approximation is very precise and can match the data up to very small errors.

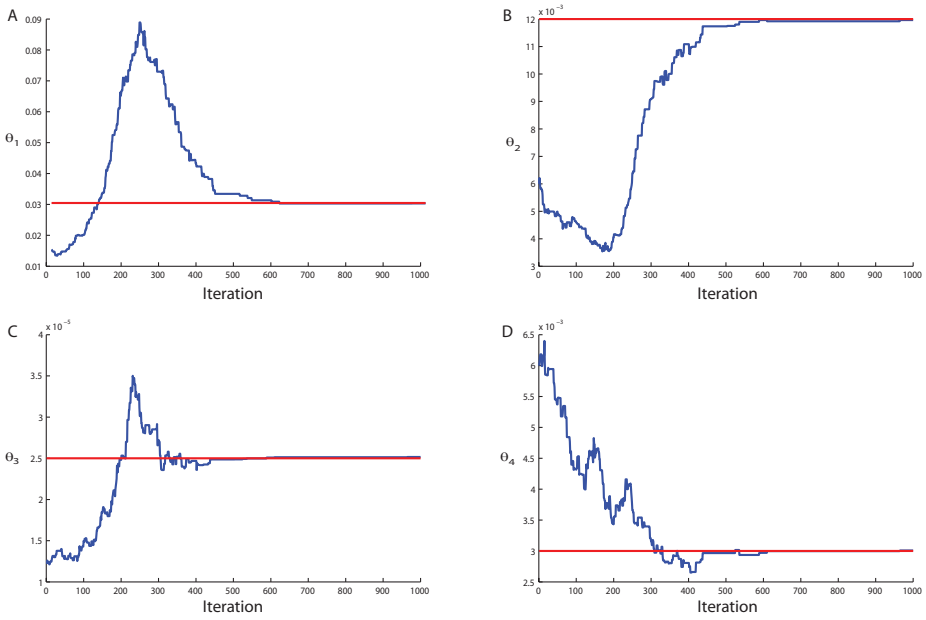

Fig. 1. Parameter search for Example 1. The panels show the values of the parameters in the search as a function of the iteration (blue). It can be seen that after approximately 600 iterations the search is very close to the true values (red lines) for all parameters and retains these values.

To evaluate on the one hand how important it is to choose a good approximation, and on the other hand whether it is necessary to adaptively change the closure method during the search, we performed the parameter inference with the same data and the same algorithm, but fixed an initial closure method and did not allow the search to switch between different approximations (i.e. by choosing $\epsilon = +\infty$). Table 1(a) compares the error in the inferred parameters obtained from our approach to the error in the results when the closure is fixed. It can be seen that for some of the fixed closure approaches the error in the parameter estimates is very large (specifically for all of the zero variance closures). Other methods provide more precise results, but overall all methods with fixed approximation are outperformed by our adaptive approach. Only

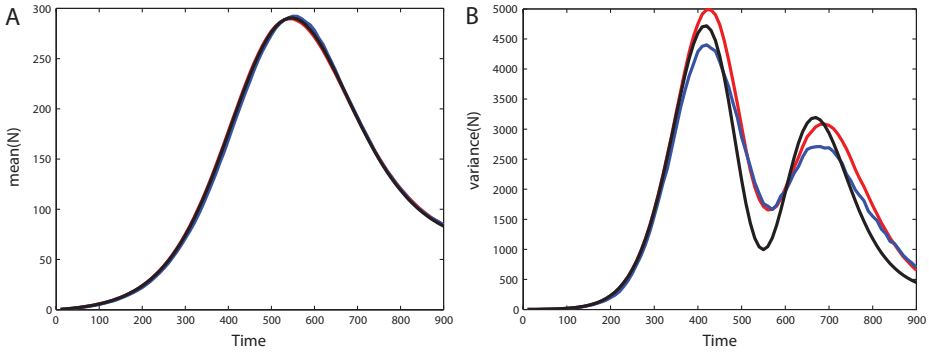

Fig. 2. Model output and data for the inferred parameters. (A) The mean computed with the best closure method (black) and the inferred parameters agrees very well both with the data (red) and the results of stochastic simulation with the inferred parameters (blue). (B) Also all the variances agree very well. The color coding is the same as in (A).

the fourth order zero cumulants (zc4) closure was more precise than our approach for two of the four parameters. However, for our case study this closure was also computationally the most expensive one and the parameter search with fixed zc4 closure actually took twice as long as the adaptive search, despite the additional stochastic simulations and evaluations of all closure methods needed here.

To further test our results, we investigated how often the approximation was changed during the run of our algorithm and which closure methods were used most often. Table 1(b), column Ex 1, shows how often the different closure methods were chosen as best. It can be seen that some approximations were never chosen (for instance all of the zero variance closures but also the third order low dispersion closure) whereas derivative matching and zero cumulants closures are chosen most often. Overall, high order closures are preferred over low order closures. This was to be expected, since these usually provide more precise results at the cost of an increased computational effort. Also we highlight that the option to switch the approximation was often used (in 19 out of 23 evaluations), and, compared to a pure simulation-based approach, we needed to employ stochastic simulation only 23 times (instead of 1,000 times).

Further examples. In addition to Example 1 we applied our algorithm to two further reaction networks and performed the same comparisons. Specifically, we considered the model of transient gene expression reported in reference [24] (termed here Example 2) and the first case study in reference [18] (termed here Example 3). The results were overall similar to those obtained for Example 1 and we only report in Table 1(b), columns Ex 2 and Ex 3, how often the different closure methods were used by our adaptive search. It can be seen that in Example 3 the second order zero cumulants and the fourth order derivative matching closure were chosen exclusively, whereas in Example 2, different zero cumulants and low dispersion closures were used most often and there was no noticeable preference for higher order closures.

(a) Example 1

closure	θ_1	θ_2	θ_3	θ_4
adapt	0.44	0.31	**0.65**	**0.29**
dm2	4.45	2.74	2.68	4.32
zc2	11.02	6.11	3.23	2.93
zv2	281.09	74.85	45.72	76.29
dm3	2.54	1.23	1.85	3.55
zc3	9.72	4.80	0.86	2.87
zv3	285.55	79.96	49.01	83.41
ld3	9.08	4.30	6.75	9.63
dm4	3.43	1.33	4.17	9.54
zc4	**0.35**	**0.19**	3.77	9.29
zv4	292.60	78.89	46.60	71.90
ld4	14.44	3.80	12.31	28.06

(b) Search statistics

closure	Ex 1	Ex 2	Ex 3
dm2	15	13	0
zc2	10	23	50
zv2	0	0	0
dm3	10	0	0
zc3	5	16	0
zv3	0	0	0
ld3	0	23	0
dm4	15	0	50
zc4	35	6	0
zv4	0	0	0
ld4	10	19	0
switch	19	30	11
sim tot	23	44	46
$i_{\max}$	1,000	1,000	2,000

Table 1. (a) Relative distance (in percent) between true and inferred parameters obtained from our adaptive algorithm (adapt) and the different closure methods on their own. The smallest distance is marked in bold. (b) Statistics of the used closure methods for the three considered reaction networks. Columns correspond to the different networks (Ex stands for example), rows report in percent how often each of the closure methods was chosen as best in our adaptive search. The bottom block of rows show how often the used approximation was changed as our search progressed through the parameter space (switch), how often stochastic simulation was performed, i.e. how often ϵ-neighborhoods were left and all the closure methods were tested (sim tot), and the total number of iterations in the search ($i_{\max}$).

6 Discussion

Using mathematical models to help in the understanding of complex biological systems is the core idea of systems biology. Up to some years ago, the main bottleneck in the identification of models was the availability of sufficiently precise and abundant data. Recently, measurement technologies have been improving at an amazing pace and nowadays enable us to simultaneously observe the dynamics of many different chemical species at single cell resolution. As these developments continue, we will gain access to data that is sufficiently informative to allow us to infer mathematical models of complex reaction networks from the measurements. However, for stochastic kinetic models that capture the inherent randomness of chemical reactions, this leads to a new bottleneck: the chemical master equation becomes intractable for high-dimensional models and especially the reverse engineering task of identifying model parameters from the measured data quickly becomes computationally infeasible. Parameter inference methods based on moment closure offer a solution to this problem but come with their own drawbacks. The goal of this paper was to address these drawbacks and to provide an automated moment-based inference method that can be used without in-depth knowledge of moment closure. To this end, we interfaced previously proposed approaches with a

stochastic simulation algorithm by continuously checking the quality of the approximations and adaptively adjusting the used closure method to the best one available. Accordingly, our approach is generally applicable whenever a sufficiently accurate approximation in the generated library of moment closure methods exists. Importantly, since the approach can adapt the used closure during the exploration of the parameter space, it is not required that a unique closure method provides good approximations for the entire parameter space. Naturally, these benefits come with an increased computational cost compared to most standard moment-based inference approaches. This increase can primarily be attributed to the additional stochastic simulation and the evaluation of all the closure methods that is performed whenever the parameter search leaves an ϵ-neighborhood around the point in parameter space where the last simulation was performed. Accordingly, the parameter ϵ provides a trade-off between computational cost and guarantees that a good approximation is used. For $\epsilon \to \infty$ our approach becomes a standard moment-based inference method, whereas $\epsilon \to 0$ essentially leads to a method akin to those based entirely on stochastic simulation. We believe that this flexibility will prove to be valuable and allow us to investigate a large variety of different reaction networks with one unified inference method.

As future work, we plan to include and test more moment closure methods (e.g. the linear noise approximation), to apply our algorithm to larger and more challenging reaction networks, and to make a complete toolbox for moment-based parameter inference publicly available. In addition to this, in order to speed up our algorithm, we plan to introduce a trade-off between precision and computational cost of the different approximations such that the more expensive high order closure methods are only chosen when the low order closures do not provide acceptable precision.

References

1. Bertaux, F., Stoma, S., Drasdo, D., Batt, G.: Modeling Dynamics of Cell-to-Cell Variability in TRAIL-Induced Apoptosis Explains Fractional Killing and Predicts Reversible Resistance. PLOS Computation Biology 10(10), e1003893 (2014)
2. Engblom, S.: Computing the moments of high dimensional solutions of the master equation. Applied Mathematics and Computation 180(2), 498–515 (2006)
3. Gillespie, D.: A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Journal of Computational Physics 22(4), 403–434 (1976)
4. Gillespie, D.: A rigorous derivation of the chemical master equation. Physica A 188(1-3), 404–425 (1992)
5. Goutsias, J., Jenkinson, G.: Markovian dynamics on complex reaction networks. Physics Reports 529, 199–264 (2013)
6. Hasty, J., Pradines, J., Dolnik, M., Collins, J.: Noise-based switches and amplifiers for gene expression. Proceedings of the National Academy of Sciences of the USA 97(5), 2075–2080 (2000)
7. Hespanha, J.: StochDynTools - a MATLAB toolbox to compute moment dynamics for stochastic networks of bio-chemical reactions. Available at http://www.ece.ucsb.edu/∼hespanha/software (2006)
8. Hespanha, J.: Moment closure for biochemical networks. Proceedings of the 3rd International Symposium on Communications, Control and Signal Processing (IEEE). St Julians, Malta. pp. 142–147 (2008)

9. Kügler, P.: Moment fitting for parameter inference in repeatedly and partially observed stochastic biological models. PLoS ONE 7(8), e43001 (2012)
10. Lillacci, G., Khammash, M.: The signal within the noise: efficient inference of stochastic gene regulation models using fluorescence histograms and stochastic simulations. Bioinformatics 29(18), 2311–2319 (2013)
11. Matis, T., Guardiola, I.: Achieving moment closure through cumulant neglect. The Mathematica Journal 12 (2010)
12. McAdams, H., Arkin, A.: Stochastic mechanisms in gene expression. Proceedings of the National Academy of Sciences of the USA 94(3), 814–819 (1997)
13. Munsky, B., Khammash, M.: The finite state projection algorithm for the solution of the chemical master equation. Journal of Chemical Physics 124, 044104 (2006)
14. Neuert, G., Munsky, B., Tan, R., Teytelman, L., Khammash, M., van Oudenaarden, A.: Systematic identification of signal-activated stochastic gene regulation. Science 339, 584–587 (2013)
15. Parise, F., Lygeros, J., Ruess, J.: Bayesian inference for stochastic individual-based models of ecological systems: an optimal pest control case study. Frontiers in Environmental Science 3, 42 (2015)
16. Ruess, J., Lygeros, J.: Moment-based methods for parameter inference and experiment design for stochastic biochemical reaction networks. ACM Transactions on Modeling and Computer Simulation (TOMACS) 25(2), 8 (2015)
17. Ruess, J., Milias-Argeitis, A., Lygeros, J.: Designing experiments to understand the variability in biochemical reaction networks. Journal of the Royal Society Interface 10(88), 20130588 (2013)
18. Ruess, J., Milias-Argeitis, A., Summers, S., Lygeros, J.: Moment estimation for chemically reacting systems by extended Kalman filtering. The Journal of Chemical Physics 135, 165102 (2011)
19. Ruess, J., Parise, F., Milias-Argeitis, A., Khammash, M., Lygeros, J.: Iterative experiment design guides the characterization of a light-inducible gene expression circuit. Proceedings of the National Academy of Sciences of the USA (2015)
20. Samoilov, M., Arkin, A.: Deviant effects in molecular reaction pathways. Nature Biotechnology 24(10), 1235–1240 (2006)
21. Singh, A., Hespanha, J.: Lognormal moment closures for biochemical reactions. 45th IEEE Conference on Decision and Control pp. 2063–2068 (2006)
22. Whittle, P.: On the use of the normal approximation in the treatment of stochastic processes. Journal of the Royal Statistical Society Series B Statistical Methodology 19, 268–281 (1957)
23. Wolf, V., Goel, R., Mateescu, M., Henzinger, T.: Solving the chemical master equation using sliding windows. BMC Systems Biology 4, 42 (2010)
24. Zechner, C., Ruess, J., Krenn, P., Pelet, S., Peter, M., Lygeros, J., Koeppl, H.: Moment-based inference predicts bimodality in transient gene expression. Proceedings of the National Academy of Sciences of the USA 109(21), 8340–8345 (2012)