Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Modeling of stochastic biological processes with non-polynomial propensities using non-central conditional moment equation

Atefeh Kazeroonian, Fabian J. Theis, and Jan Hasenauer

Institute of Computational Biology, Helmholtz Zentrum München, 85764 Neuherberg, Germany

Department of Mathematics, Technische Universität München, 85748 Garching, Germany

(e-mail: {atefeh.kazeroonian, fabian.theis, jan.hasenauer}@helmholtz-muenchen.de).

Abstract: Biological processes exhibiting stochastic fluctuations are mainly modeled using the Chemical Master Equation (CME). As a direct simulation of the CME is often computationally intractable, we recently introduced the Method of Conditional Moments (MCM). The MCM is a hybrid approach to approximate the statistics of the CME solution. In this work, we provide a more comprehensive formulation of the MCM by using non-central conditional moments instead of central conditional moments. The modified formulation allows for additional insight into the model structure and for extensions to higher-order reactions and non-polynomial propensity functions. The properties of the non-central MCM are analyzed using a model for the regulation of pili formation on the surface of bacteria, which possesses rational propensity functions.

Keywords: stochastic modeling, chemical master equation, moment equations

1. INTRODUCTION

Gene expression, signal transduction and even cell fate decisions have been shown to be subject to stochastic fluctuations [Raser and O'Shea, 2004, Eldar and Elowitz, 2010]. These stochastic fluctuations are often due to the low abundance of DNAs, mRNAs and proteins [Taniguchi et al., 2010]. For decades it was assumed that these fluctuations are a nuisance and disturb correct information processing in cells. However, in recent years it has been shown that stochastic fluctuations are essential for functioning as well as robustness of many processes [Eldar and Elowitz, 2010]. Furthermore, fluctuations can be employed to unravel the underlying signaling mechanisms [Munsky et al., 2009, 2012].

A multitude of approaches have been proposed to model stochastic dynamics in biological systems. Discrete-state continuous-time Markov chains (CTMCs) are the gold standard as they capture the discreteness of the ensemble sizes of chemical species $(S_1, S_2, \ldots, S_{n_s})$ as well as the discreteness of chemical reactions,

$$\mathrm{R}_j: \quad \sum_{i=1}^{n_s} \nu_{ij}^- S_i \rightarrow \sum_{i=1}^{n_s} \nu_{ij}^+ S_i, \quad j = 1, \ldots, n_r.$$

The stoichiometric coefficients ν_{ij}^-, ν_{ij}^+ and $\nu_{ij} = \nu_{ij}^+ - \nu_{ij}^-$ denote the number of molecules of species S_i consumed, produced and net produced, respectively, when the reaction R_j takes place. Accordingly, ν_j^-, ν_j^+ and ν_j describe the overall stoichiometry of reaction R_j.

CTMCs describe the time evolution of the ensemble state $X_t = (X_{1,t}, \ldots, X_{n_s,t}) \in \mathbb{N}_0^{n_s}$ of the species $S_1, S_2, \ldots, S_{n_s}$ as a jump process. X_t remains constant as long as no reaction occurs. If R_j takes place, the ensemble sizes change according to the stoichiometry of R_j, $X_t \rightarrow X_t + \nu_j$. The index j of the next reaction and the time to the next reaction are random with distributions determined by the propensity functions $a_j : \mathbb{N}_0^{n_s} \rightarrow \mathbb{R}_+$, $j = 1, \ldots, n_r$ [Feller, 1940]. The statistics of the process, i.e. the probabilities $p(x|t) = P(X_t = x)$ that X_t occupies a certain state x at time t, are described by the chemical master equation (CME) [van Kampen, 2007],

$$\frac{\partial}{\partial t} p(x|t) = \sum_{\substack{j=1 \\ x \geq \nu_j^+}}^{n_r} a_j(x - \nu_j) p(x - \nu_j|t) - \sum_{j=1}^{n_r} a_j(x) p(x|t), \quad (1)$$

in which the inequality constraint $x \geq \nu_j^+$ ensures positivity. Associated propensities a_j are "proper", meaning that if $\exists i \in \{1, \ldots, n_s\} : X_{i,t} \ngeq \nu_{ij}^-$ then $a_j(X_t) = 0$.

The CME is a system of linear ordinary differential equations (ODEs) which describes the dynamics of CTMCs. Jahnke and Huisinga [2007] derived a closed-form solution of the CME in the case of monomolecular reactions. If the process contains nonlinear propensity functions, in general, numerical approximations are necessary. A multitude of approximation methods have been proposed over the last decades, e.g., error-aware state truncation [Munsky and Khammash, 2006], inexact integration [Sidje et al., 2007], product approximations [Jahnke, 2011], approximation of the CME by the Fokker-Planck equation [Gardiner,

2011], or modeling of the statistical moments of the CME solution [Engblom, 2006]. However, these methods often fail if low- as well as high-copy number species are involved in the biochemical process.

In recent years, several hybrid methods have been introduced to circumvent these shortcomings. These hybrid methods are based on decomposing the system into fast and slow reactions [Haseltine and Rawlings, 2002], or low- and high-copy number species [Hellander and Lötstedt, 2007, Henzinger et al., 2010, Jahnke, 2011, Menz et al., 2012]. For the latter we recently proposed a generalization, the *method of conditional moments (MCM)* [Hasenauer et al., 2013]. The MCM provides a fully stochastic description for the low-copy number species and a moment-based description for the medium/high-copy number species. Thus, it combines concepts from *hybrid stochastic-deterministic modeling* [Jahnke, 2011, Menz et al., 2012] and *moment-based modeling* [Engblom, 2006]. We showed that this allows for an improved approximation quality for common models of transcription-translation process.

In this manuscript, we generalize the MCM to include reactions with rates not obeying the law of mass action. This allows for the consideration of activation and inhibition mechanisms possessing Michaelis-Menten-like characteristics. In addition to this generalization, we state the MCM in terms of non-central moments. This improves the readability and interpretability compared to the central MCM [Hasenauer et al., 2013]. To enhance the MCM further for systems with nonlinear propensity functions, we propose the use of Taylor series expansion (TSE) together with the low-dispersion closure scheme. This approach is evaluated using a model for PapI regulation in *E. coli* [Munsky and Khammash, 2006].

2. APPROACH

Single-molecule fluorescence microscopy techniques, such as fluorescence *in situ* hybridization, revealed that the copy numbers of chemical species spread over several orders of magnitude. In *E. coli*, the mean number of a protein is in general 100- to 1000-fold higher than the mean number of the corresponding mRNA [Taniguchi et al., 2010]. Such naturally occurring scale separations can be exploited to accelerate the simulation of stochastic biochemical processes. Therefore, species $S_1, \dots, S_{n_s}$ are classified as either low- or medium/high-copy number species. The abundances of low-copy number species are collected in Y_t, while the abundances of medium/high-copy number species are collected in Z_t. Thus, without lost of generality $X_t = (Y_t, Z_t)$ and $p(x|t) = p(y, z|t)$.

The CME describes the evolution of the full joint distribution $p(y, z|t)$. In contrast, the MCM employs the decomposition

$$p(y, z|t) = p(z|y, t)p(y|t) \tag{2}$$

which follows from the multiplication axiom. $p(y|t)$ denotes the marginal probability of the low-copy number species being in state y, while $p(z|y, t)$ denotes the conditional probability of the medium/high-copy number species being in state z given that the low-copy number species are in state y. Using this decomposition, the CME can be rewritten as

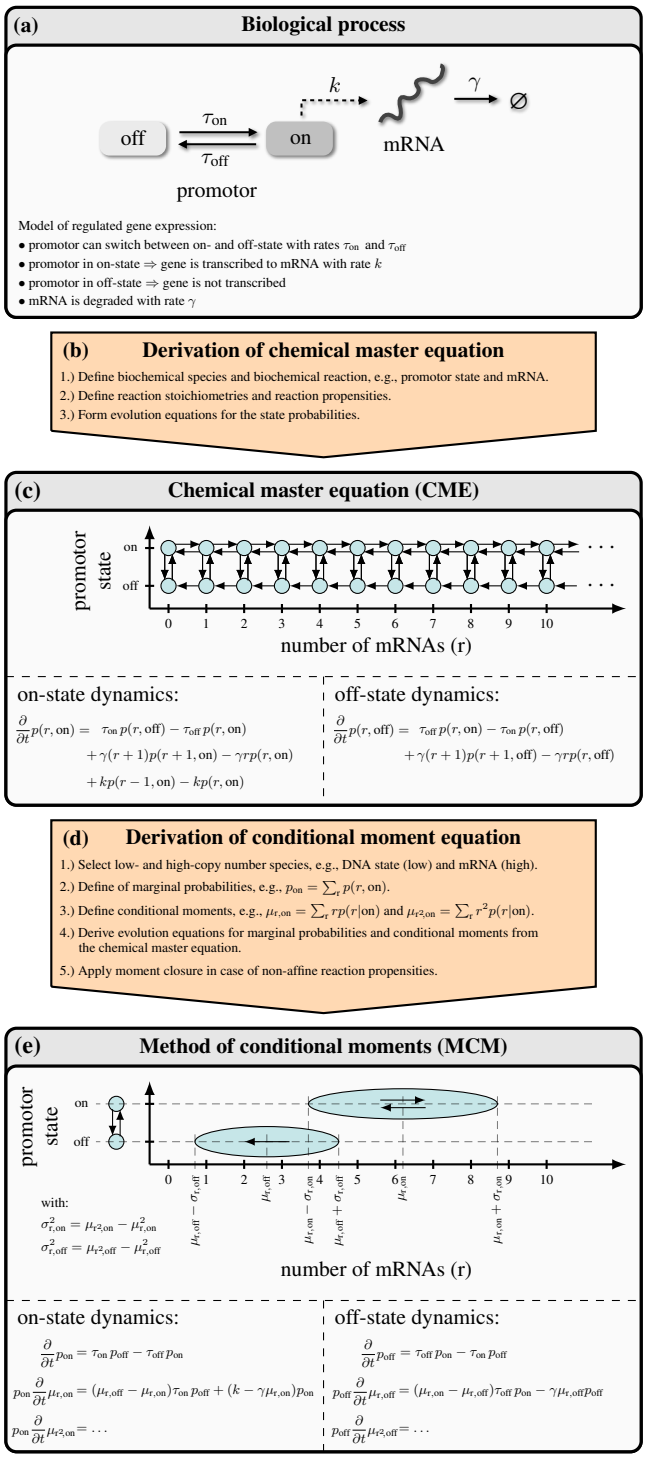

Fig. 1. Illustration of the method of conditional moments using a two-state model for gene expression. (a) Model of gene expression accounting for two promotor states [Munsky et al., 2012]. (b) Procedure to derive the CME. (c) CME of the gene expression model. The discrete state space is visualized along with the possible transitions. Note that we skip the dependence on the time t to simplify the notation. (d) Procedure to derive the conditional moment equation from the CME. (e) Conditional moment equation when modeling the promotor state as low-copy number species and the mRNA as medium/high-copy number species.

$$\frac{\partial}{\partial t}p(y,z|t) = -\sum_{j=1}^{n_r} a_j(y,z)p(z|y,t)p(y|t) + \sum_{\substack{j=1 \\ y\geq\nu_{j,y}^+ \\ z\geq\nu_{j,z}^+}}^{n_r} a_j(y-\nu_{j,y}, z-\nu_{j,z})p(z-\nu_{j,z}|y-\nu_{j,y},t)p(y-\nu_{j,y}|t). \tag{3}$$

This decomposition suggests that the dynamics of $p(y|t)$ and $p(z|y,t)$ can be modeled separately [Haseltine and Rawlings, 2005, Hasenauer et al., 2013]. In the MCM, the distribution of the low-copy number species is described in terms of marginal probabilities,

$$p(y|t) = \sum_{z\geq 0} p(y,z|t). \tag{4}$$

For medium/high-copy number species, the non-central moments of $p(z|y,t)$ are considered,

$$\mu_{I,z}(y,t) = \mathbb{E}_z\left[Z^I \middle|\, y,t\right] = \sum_{z\geq 0} z^I p(z|y,t), \tag{5}$$

with I being a non-negative integer-valued vector of length $n_{s,z}$ and $Z^I := \prod_{i=1}^{n_{s,z}} Z_i^{I_i}$. The conditioning on y can be important if transitions between different low-copy number states are slow. The marginal probabilities of discrete states together with the corresponding conditional moments can be used to determine the overall moments of the process, $\bar{\mu}_{I,z}(t)$, i.e. the moments independent of the stochastic states, via

$$\bar{\mu}_{I,z}(t) = \mathbb{E}\left[Z^I|t\right] = \sum_{y\geq 0} \mu_{I,z}(y,t)p(y|t).$$

In the following, we provide exact and approximate evolution equations for $p(y|t)$ and $\mu_{I,z}(y,t)$. Figure 1 provides a visual outline of the method.

3. NON-CENTRAL CONDITIONAL MOMENT EQUATIONS

Upon decomposition of state vector, $x = (y,z)$, evolution equations for marginal probabilities of low-abundance species, $p(y|t)$, as well as non-central moments of high-abundance species conditioned on the state of the low-abundance species, $\mu_{I,z}(y,t)$, have to be determined. Therefore, a governing equation for expectation of an arbitrary polynomial test-function $T(Z)$ is derived to provide MCM equations as special cases.

Lemma 1. Let $p(y,z|t) = p(z|y,t)p(y|t)$ satisfy a proper CME (3) ($\forall x \not\geq \nu_j^- : a_j(x) = 0$), then, for any polynomial test-function $T : \mathbb{N}_0^{n_z} \times \mathbb{R}_+ \to \mathbb{R}$,

$$\frac{\partial}{\partial t}\left(\mathbb{E}_z[T(Z)|\, y,t]\, p(y|t)\right) = \sum_{\substack{j=1 \\ y\geq\nu_{j,y}^+}}^{n_r} \mathbb{E}_z[T(Z+\nu_{j,z})a_j(y-\nu_{j,y},Z)|\, y-\nu_{j,y},t]\, p(y-\nu_{j,y}|t) - \sum_{j=1}^{n_r} \mathbb{E}_z[T(Z)a_j(Z,y)|\, y,t]\, p(y|t). \tag{6}$$

Note that Lemma 1 is only valid if the expectation $\mathbb{E}_z[T(Z)a_j(Z,y)|\, y,t]$ exists. This is generally true for reasonable models of biological processes.

Proof. The time derivative of $\mathbb{E}_z[T(Z)|\, y,t]\, p(y|t)$ is

$$\frac{\partial}{\partial t}\left(\mathbb{E}_z[T(Z)|\, y,t]\, p(y|t)\right) = \frac{\partial}{\partial t}\left(\sum_{z\geq 0} T(z)p(z|y,t)p(y|t)\right) = \sum_{z\geq 0} T(z)\frac{\partial}{\partial t}p(z,y|t) + \sum_{z\geq 0} p(z,y|t)\frac{\partial}{\partial t}T(z).$$

The second term vanishes as the time derivative of $T(z)$ is zero. Similar to the proof by Hasenauer et al. [2013], $\frac{\partial}{\partial t}p(z,y|t)$ is substituted according to the CME (3), the order of summations is changed, and z is replaced by $z+\nu_{j,z}$ in the first sum to obtain

$$\frac{\partial}{\partial t}\left(\mathbb{E}_z[T(Z)|\, y,t]\, p(y|t)\right) = \sum_{\substack{j=1 \\ y\geq\nu_{j,y}^+}}^{n_r} \sum_{z\geq\nu_{j,z}^-} T(z+\nu_{j,z})a_j(y-\nu_{j,y},z)p(z|y-\nu_{j,y},t) \times p(y-\nu_{j,y}|t) - \sum_{j=1}^{n_r}\sum_{z\geq 0} T(z)a_j(y,z)p(z|y,t)p(y|t).$$

The lower bound $z \geq \nu_{j,z}^-$ can be replaced by $z \geq 0$ as for $z \not\geq \nu_{j,z}^- : a_j(z) = 0$ (due to propensities being proper). Utilizing the definition of conditional expectation $\mathbb{E}_z[T(z)|\, y,t] = \sum_{z\geq 0} T(Z)p(z|y,t)$, the expression above simplifies to the evoluation equation stated in Lemma 1, which concludes the proof. □

Setting $T(Z)$ to 1 and Z^I, Lemma 1 yields the governing equations for $p(y|t)$ and $\mu_{I,z}(y,t)$ respectively.

Theorem 2. Let $p(y,z|t) = p(z|y,t)p(y|t)$ satisfy a proper CME (3), the evolution equations for marginal probabilities, $p(y|t)$, and non-central conditional moments, $\mu_{I,z}(y,t)$, are given by the system

$$\begin{aligned}
\frac{\partial}{\partial t}p(y|t) &= -\sum_{j=1}^{n_r} \mathbb{E}_z[a_j(Z,y)|\, y,t]\, p(y|t) + \sum_{\substack{j=1 \\ y\geq\nu_{j,y}^+}}^{n_r} \mathbb{E}_z[a_j(y-\nu_{j,y},Z)|\, y-\nu_{j,y},t]\, p(y-\nu_{j,y}|t), \\
p(y|t)\frac{\partial}{\partial t}\mu_{I,z}(y,t) + \mu_{I,z}(y,t)\frac{\partial}{\partial t}p(y|t) &= \sum_{\substack{j=1 \\ y\geq\nu_{j,y}^+}}^{n_r} \mathbb{E}_z\left[a_j(y-\nu_{j,y},Z)(Z+\nu_{j,z})^I \middle|\, y-\nu_{j,y},t\right] \times p(y-\nu_{j,y}|t) - \sum_{j=1}^{n_r} \mathbb{E}_z\left[a_j(Z,y)Z^I \middle|\, y,t\right] p(y|t).
\end{aligned} \tag{7}$$

The MCM equations can be written for moments of arbitrary order. In contrast to the central conditional moment equations [Hasenauer et al., 2013], no distinction between first and higher-order moments is necessary, yielding a more compact set of equations. Also this presentation of the MCM is a generalization of the

central MCM since it removes the assumption that the propensity functions should allow for a decomposition of the form $a_j(x,t) = c\, g_j(y,t)h_j(z,t)$. Therefore, the non-central MCM provides a simpler and more general formulation and thus facilitates further investigations.

The resulting set of evolution equations is a DAE system. Initial conditions for $p(y|t)$, $\mu_{I,z}(y,t)$, $\dot{p}(y|t)$ and $\dot{\mu}_{I,z}(y,t)$ can be calculated via (7) given that $\forall y : p(y|t_0) \neq 0$. If this is not fulfilled, the procedure introduced by Hasenauer et al. [2013] can be adopted.

The simulation of the conditional moment equations requires the evaluation of expectations $\mathbb{E}_z[a_j(Z,y)|\, y,t]$ and $\mathbb{E}_z\big[a_j(Z,y)Z^I\big|\, y,t\big]$. This can be done by employing the Taylor series expansion of the propensity function $a_j(z,y)$. More specifically, since the expectation with respect to the random variable Z is sought, $a_j(z,y)$ is merely expanded with respect to z. In principle, any expansion point can be selected for the TSE, however, the vector of conditional means of z, $\mu'_z(y,t) = \sum_{z\geq 0} z\; p(z|y,t) = (\mu_{e_1,z}(y,t),\ldots,\mu_{e_{n_s},z}(y,t))$, is considered in the following:

$$\begin{aligned} a_j(z,y) &= a_j(\mu'_z(y,t),y) \\ &+ \sum_{k=1}^{n_{s,z}} \frac{\partial a_j(\mu'_z(y,t),y)}{\partial z_k}(z_k - \mu_{e_k,z}(y,t)) \\ &+ \frac{1}{2}\sum_{k,l=1}^{n_{s,z}} \frac{\partial^2 a_j(\mu'_z(y,t),y)}{\partial z_k \partial z_l}(z_k - \mu_{e_k,z}(y,t))(z_l - \mu_{e_l,z}(y,t)) \\ &+ \ldots. \end{aligned} \tag{8}$$

The expectation $\mathbb{E}_z[a_j(Z,y)|\, y,t]$ follows as

$$\begin{aligned} \mathbb{E}_z[a_j(Z,y)|\, y,t] &= a_j(\mu'_z(y,t),y) \\ &+ \frac{1}{2}\sum_{k,l=1}^{n_{s,z}} \frac{\partial^2 a_j(\mu'_z(y,t),y)}{\partial z_k \partial z_l} C_{e_k+e_l,z}(y,t) + \ldots, \end{aligned} \tag{9}$$

in which e_i denotes the i^{th} unit vector and $C_{I,z}(y,t) = \sum_{z\geq 0}(z-\mu'_z(y,t))^I\, p(z|y,t)$ represent the central moments. The central moments can be replaced by their equivalent expressions in terms of non-central moments, e.g., $C_{e_k+e_l,z}(y,t) = \mu_{e_k+e_l,z}(y,t) - \mu_{e_k,z}(y,t)\mu_{e_l,z}(y,t)$. In case the TSE (9) is finite, $\mathbb{E}_z\big[a_j(Z,y)Z^I\big|\, y,t\big]$ can be evaluated in a similar manner by writing the TSE of $a_j(Z,y)Z^I$. However, if the TSE (9) is infinite, or intractably high-order, it may be truncated at a specific order N. This truncation introduces a degree of freedom in choosing either of the following approaches for evaluating $\mathbb{E}_z\big[a_j(Z,y)Z^I\big|\, y,t\big]$.

Truncate-multiply approach. To approximate the expectation $\mathbb{E}_z\big[a_j(Z,y)Z^I\big|\, y,t\big]$, first the TSE of $a_j(Z,y)$ (8) is truncated at order N, and then it is multiplied by Z^I. The expectation of the resulting product is

$$\begin{aligned} \mathbb{E}_z\big[a_j(Z,y)Z^I\big|\, y,t\big] &= a_j(\mu'_z(y,t),y)\mu_{I,z}(y,t) \\ &+ \sum_{k=1}^{n_{s,z}} \frac{\partial a_j(\mu'_z(y,t),y)}{\partial z_k}\mathbb{E}_z\big[(Z_k - \mu_{e_k,z}(y,t))Z^I\big|\, y,t\big] \\ &+ \frac{1}{2}\sum_{k,l=1}^{n_{s,z}} \frac{\partial^2 a_j(\mu'_z(y,t),y)}{\partial z_k \partial z_l} \\ &\times \mathbb{E}_z\big[(Z_k - \mu_{e_k,z}(y,t))(Z_l - \mu_{e_l,z}(y,t))Z^I\big|\, y,t\big] + \ldots. \end{aligned} \tag{10}$$

The expectation terms in (10) can easily be expressed in terms of non-central moments.

Multiply-truncate approach. In the multiply-truncate approach, the order of operations is changed. First Z^I is multiplied by $a_j(Z,y)$, then the TSE of the product $a_j(Z,y)Z^I$ is obtained and, if necessary, truncated, yielding the expectation

$$\begin{aligned} \mathbb{E}_z\big[Z^I a_j(Z,y)\big|\, y,t\big] &= (\mu'_z(y,t))^I\, a_j(\mu'_z(y,t),y) \\ &+ \frac{1}{2}\sum_{k,l=1}^{n_{s,z}} \frac{\partial^2\left((\mu'_z(y,t))^I\, a_j(\mu'_z(y,t),y)\right)}{\partial z_k \partial z_l} C_{e_k+e_l,z}(y,t) \\ &+ \ldots. \end{aligned} \tag{11}$$

In the truncate-multiply approach, if the TSE is truncated at order N, (10) contains moments up to order $N+\sum_i I_i$, whereas in the multiply-truncate approach, with the TSE of order N, (11) contains moments up to order N. Thus, for the multiply-truncate approach it may be more plausible to have the truncation order N equal to or greater than the moment order, i.e. $N \geq \sum_i I_i$. In this way, the evolution equation for a moment $\mu_{I,z}(y,t)$ depends on moments of the same order.

4. CLOSURE OF THE CONDITIONAL MOMENT EQUATIONS

The evolution equations for moments up to order M, i.e. $\forall I : \sum_i I_i \leq M$, in general depend on moments of orders $> M$. To simulate the conditional moment equations, these higher-order moments have to be approximated using moment closure. Also, if the propensities are non-polynomial, their TSEs are generally infinite and need to be truncated. Accordingly, the accuracy of conditional moment equations is determined by (1) the error introduced by truncating Taylor series of the propensity functions $a_j(z,y)$, and (2) the error introduced by the moment closure scheme. In the following, these two sources of error are discussed for polynomial and non-polynomial propensity functions.

4.1 Polynomial propensities

If the kinetics obey the law of mass action, all propensities are polynomial functions and their TSEs are finite and their truncation is not necessary. However, higher-order moments still appear.

Under certain conditions the higher-order moments cancel out, yielding a closed set of equations [Hasenauer et al., 2013]. However, in general, closure schemes have to be employed. Moment closure schemes approximate higher-order moments as functions of the lower-order moments, e.g., using distributional assumptions [Engblom, 2006, Singh and Hespanha, 2011]. For instance, the simplest and also most commonly used moment closure is low-dispersion closure which relies on the assumption that the distribution is tightly clustered around the mean, implying that the higher-order central moments are negligible. Accordingly, all higher-order central moments are set to zero,

$$\forall I \text{ with } \sum_i I_i > M : C_{I,z}(y,t) = 0. \tag{12}$$

In case of polynomial propensities, different moment closure schemes can be used with either of the truncate-multiply and multiply-truncate approaches. The error of the approximation is then directly related to the validity of the assumptions made by the closure.

4.2 Non-polynomial propensity functions

In case of non-polynomial propensity functions, the corresponding TSEs are infinite and need to be truncated. In this case, both the errors introduced by the truncation of the TSE and by the moment closure affect the approximation quality of the MCM.

TSEs are truncated by discarding higher-order terms in (9) and (10) or (11). In the truncate-multiply approach, higher-order terms in (9) and (10) are of different natures, i.e. the former are the higher-order central moments while the latter are combinations of non-central moments. Thus, setting them to zero implies different, and inconsistent, assumptions about the moments. However, in the multiply-truncate approach, truncations of the TSEs (9) and (11) both correspond to the same assumption, i.e. that the higher-order central moments are zero. This is conceptually similar to the low-dispersion moment closure, which also sets higher-order central moments to zero. Hence, in the approximation of conditional moment equations with non-polynomial propensities, the low dispersion closure together with the multiply-truncate approach is a promising choice as it ensures consistency.

Interestingly, it can be shown that using the low-dispersion closure, the truncate-multiply approach is identical to the multiply-truncate approach, given that the order of the TSE truncation at least equals the moment order, i.e. $N \geq M$. However, the two approaches are different if $N < M$, or if another moment closure scheme is applied.

5. EXAMPLE: PAPI REGULATION MODEL

In this section, the performance of non-central MCM is assessed using a biological system that describes the regulation of Pap pili formation on the surface of *E. coli* [Munsky and Khammash, 2006]. This biological process involves low- as well as medium/high-copy number species. Therefore, it is challenging for simulation methods that do not account for the differences in the abundance of the species. Furthermore, it demands handling of non-polynomial propensity functions.

Several simulations based on MCM with different setups are carried out and the results are compared to the results obtained by finite state projection (FSP). As shown by Munsky and Khammash [2006], the results of FSP can be assumed to be exact for this problem.

5.1 Biological system

The PapI regulation model (Figure 2) comprises a *pap* operon and two regulatory proteins. The regulatory protein LRP can reversibly bind to either or both of the binding sites on the *pap* operon. The states g_1 to g_4 represent the four possible configurations of the *pap* operon. Pili production can only take place if the operon is in state g_2. Protein PapI decreases the unbinding rate of LRP from

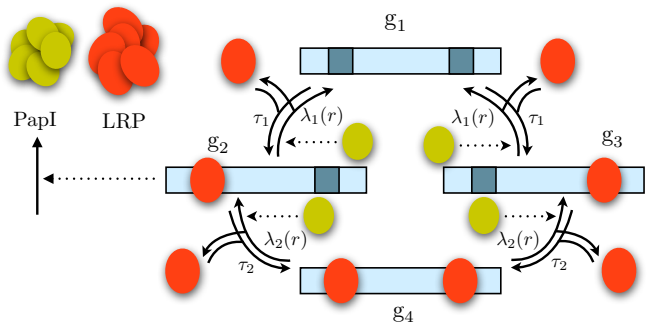

Fig. 2. Schematic of the PapI regulation model. Arrows represent the binding and unbinding of LRP to/from the operon. Dotted arrows indicate the influence of PapI on the reaction rates.

Table 1. Reactions and reaction propensities for the PapI regulation model.

reaction number	stoichiometry	rate
R_1	$g_1 + l \to g_2$	$\tau_1 = c_1$
R_2	$g_2 \to l + g_1$	$\lambda_1 = c_3 - c_4 \frac{r}{r+1}$
R_3	$g_1 + l \to g_3$	$\tau_1 = c_1$
R_4	$g_3 \to l + g_1$	$\lambda_2 = c_5 - c_6 \frac{r}{r+1}$
R_5	$g_2 + l \to g_4$	$\tau_2 = c_2$
R_6	$g_4 \to l + g_2$	$\lambda_2 = c_5 - c_6 \frac{r}{r+1}$
R_7	$g_3 + l \to g_4$	$\tau_2 = c_2$
R_8	$g_4 \to l + g_3$	$\lambda_1 = c_3 - c_4 \frac{r}{r+1}$
R_9	$g_2 \to g_2 + r$	k_r
R_{10}	$r \to \emptyset$	γ_r

the operon, and therefore establishes a positive feedback loop for the production of pili. The total number of LRP molecules (denoted by l) is constant, while the count of PapI molecules (denoted by r) is variable. Reactions and kinetic rates of the model are provided in Table 1.

The operon states are modeled as low-abundance species as there is only a single operon. PapI and LRP proteins are found in relatively larger amounts, therefore, they are considered as medium/high-copy number species and represented by the moments of their distributions. Furthermore, to obtain the MCM equations, the nonlinear kinetic rates, i.e. those in reactions R_2, R_4, R_6 and R_8, should be approximated as polynomials by means of TSE.

5.2 Simulation study

To analyze the impact of the approximation errors of moment closure and truncation of TSE (in either of the truncate-multiply and multiply-truncate approaches) on the accuracy of the MCM simulation, several simulations are carried out. We use the notation MCMi/j to refer to different simulations where i denotes the highest moment order (previously mentioned as M) and j denotes the order of the TSE (previously mentioned as N). For all the simulations, parameter values $(c_1, c_2, c_3, c_4, c_5, c_6, k_r, \gamma_r) = (1, 0.01, 2.5, 2.25, 1.2, 0.2, 10, 1)$ and initial conditions $l = 100, r = 5$, and $p(g_1) = 1$ are used.

Using the truncate-multiply approach (Figure 3), we find that all MCM simulations generally agree with the FSP in resolving marginal probabilities and conditional moments. However, as Figure 4 shows, there is no consistent trend in

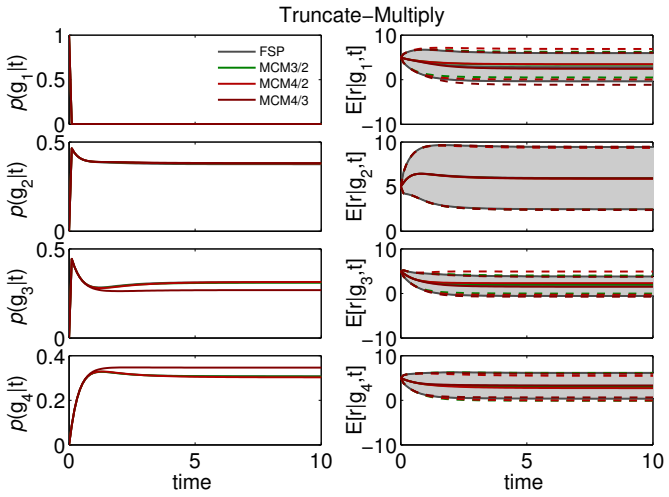

Fig. 3. Marginal probabilities of states of the *pap* operon (left) and conditional means and 1-σ intervals of PapI (right) for FSP, MCM3/2, MCM4/2, and MCM4/3 with the truncate-multiply approach.

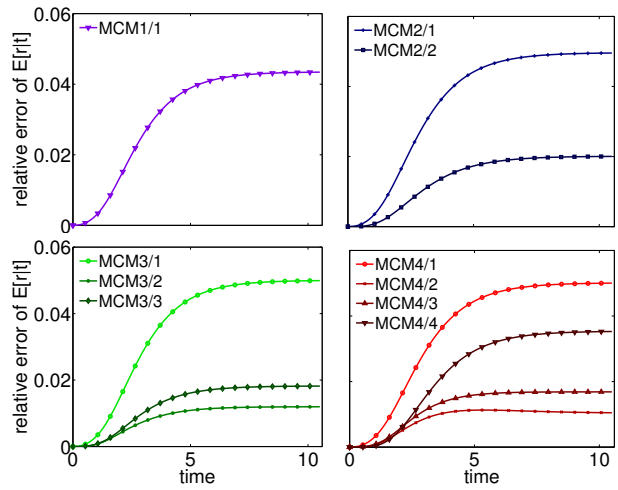

Fig. 4. Relative errors of the mean number of PapI molecules predicted by different MCM simulations with the truncate-multiply approach.

the impact of the moment order and the truncation order on the accuracy of the simulation results.

The results for the overall mean of PapI (Figure 4) suggest that, for most cases, applying a truncation order smaller than the moment order leads to improved approximation quality. For this example, the TSE of order 2 yields the smallest error. Although increasing the moment order improves the results when the truncation order is equal to/greater than two, this is not always the case.

For instance, MCM1/1 performs better than MCM2/1, MCM3/1, and MCM4/1 (Figure 4). Relative errors in Figure 4 are computed with respect to FSP simulation, e.g., $\text{error}_{\text{MCM2/2}} = \text{abs}(\mathbb{E}[r|t]_{\text{MCM2/2}} - \mathbb{E}[r|t]_{\text{FSP}})/\mathbb{E}[r|t]_{\text{FSP}}$.

The same study is repeated for the multiply-truncate approach. In this approach, the highest moment order that appears in the MCM equations corresponds to the minimum of the truncation order and the moment order,

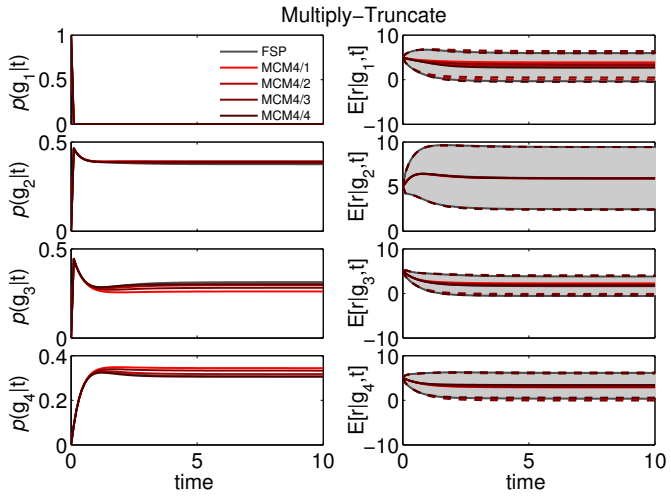

Fig. 5. Marginal probabilities of states of the *pap* operon (left) and conditional means and 1-σ intervals of PapI (right) for FSP, MCM4/1, MCM4/2, MCM4/3, and MCM4/4 with the multiply-truncate approach.

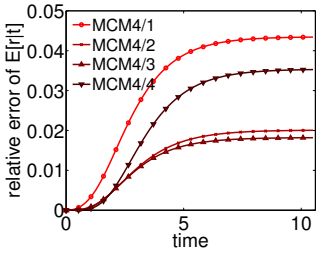

Fig. 6. Relative errors of the mean number of PapI molecules predicted by different MCM simulations with the multiply-truncate approach.

i.e. $\min(M, N)$. Therefore, given that $j < i$, all simulations MCMi/j with the same j are identical. Therefore, only the effect of the truncation order on the accuracy of the MCM simulation has to be investigated. Unfortunately, we again do not find a consistent trend (Figure 6).

To summarize, this example illustrates how MCM can be used to approximate the statistics of stochastic processes with non-polynomial reaction propensities. Surprisingly, no consistent trend was found in the impact of the order of TSE and the order of moment closure on the accuracy of the MCM simulation when low dispersion closure was used.

6. DISCUSSION

In this work, we presented the non-central conditional moment equations, a reformulation and extension of the central MCM [Hasenauer et al., 2013]. Being a hybrid simulation method for systems of stochastic dynamics, the MCM combines stochastic and moment-based descriptions depending on copy-numbers of species. Reformulation in terms of non-central moments facilitated the extension of the MCM to include reactions with non-polynomial kinetic rates. We proposed the use of Taylor series expansion for the approximation of non-polynomial propensity functions. As the truncation of the TSE introduces degrees of

freedom, we compared two alternative approaches for the approximation of the conditional moment equations.

To evaluate the performance of non-central MCM, a model for regulation of Pap pili formation on the surface of *E. coli* was analysed. Our study demonstrated that non-central MCM can handle non-polynomial propensity functions by means of Taylor series expansion. Surprisingly, we found that increasing the order of Taylor series expansion does not always improve the accuracy of simulation.

In situations where the low-dispersion assumption is not physically plausible, the compatibility of more sophisticated closure techniques [Gillespie, 2009, Singh and Hespanha, 2011] with the TSE has to be analyzed. Also, to further enhance the approximation quality, approximation approaches such as sigma-point expansion methods, instead of Taylor series expansion, might be used.

If all propensities are rational, the approach introduced by Milner et al. [2011] for moment equations can also be adapted for the MCM. In this approach, a polynomial system is obtained by multiplying the original system by the product of the propensity denominators, and the TSE can be avoided.

The approximation of the statistics of stochastic processes by the MCM can be used in a variety of applications. In particular, parameter estimation, experimental design and control of stochastic processes can be rendered more efficient [Zechner et al., 2012].

ACKNOWLEDGEMENTS

The authors would like to thank Justin Feigelman and Fabian Fröhlich for proofreading the article. The authors acknowledge financial support from the German Federal Ministry of Education and Research (BMBF) within the Virtual Liver project (Grant No. 0315766) and LungSys II (Grant No. 0316042G), and the European Union within the ERC grant "LatentCauses".

REFERENCES

A. Eldar and M. B. Elowitz. Functional roles for noise in genetic circuits. *Nature*, 467(9):1–7, Sept. 2010. doi: 10.1038/nature09326.

S. Engblom. Computing the moments of high dimensional solutions of the master equation. *Appl. Math. Comp.*, 180:498–515, 2006.

W. Feller. On the integro-differential equation of purely discontinous Markoff processes. *Trans. of the American Mathematical Society*, 48:4885–4915, 1940.

C. W. Gardiner. *Handbook of stochastic methods: For physics, chemistry and natural sciences*. Springer Series in Synergetics. Springer Berlin / Heidelberg, 4th edition, 2011.

C. S. Gillespie. Moment-closure approximations for mass-action models. *IET Syst. Biol.*, 3(1):52–58, 2009.

E. L. Haseltine and J. B. Rawlings. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *J. Chem. Phys.*, 117(15):6959–6969, 2002.

E. L. Haseltine and J. B. Rawlings. On the origins of approximations for stochastic chemical kinetics. *J. Chem. Phys.*, 123(164115), 2005.

J. Hasenauer, V. Wolf, A. Kazeroonian, and F. J. Theis. Method of conditional moments (MCM) for the chemical master equation. *J. Math. Biol.*, pages 1–49, 2013.

A. Hellander and P. Lötstedt. Hybrid method for the Chemical Master Equation. *J. Comput. Phys.*, 227:100–122, 2007.

T. A. Henzinger, L. Mikeev, M. Mateescu, and V. Wolf. Hybrid numerical solution of the chemical master equation. In *Proceedings of the 8th International Conference on Computational Methods in Systems Biology*, pages 55–65, New York, NY, USA, 2010. ACM.

T. Jahnke. On reduced models for the chemical master equation. *Multiscale Model. Simul.*, 9(4):1646–1676, 2011.

T. Jahnke and W. Huisinga. Solving the chemical master equation for monomolecular reaction systems analytically. *J. Math. Biol.*, 54(1):1–26, 2007.

I. Lestas, G. Vinnicombe, and J. Paulsson. Fundamental limits on the suppression of molecular fluctuations. *Nat.*, 467:174–178, Sept. 2010.

S. Menz, J. C. Latorre, C. Schütte, and W. Huisinga. Hybrid stochastic deterministic solution of the Chemical Master Equation. *SIAM J. on Multiscale Model. Simul.*, 10(4):1232—1262, 2012.

P. Milner, C. S. Gillespie, and D. J. Wilkinson. Moment closure approximations for stochastic kinetic models with rational rate laws. *Mathematical Biosciences*, 231 (2):99 – 104, 2011.

B. Munsky and M. Khammash. The finite state projection algorithm for the solution of the chemical master equation. *J. Chem. Phys.*, 124(4):044104, 2006.

B. Munsky, B. Trinh, and M. Khammash. Listening to the noise: random fluctuations reveal gene network parameters. *Mol. Syst. Biol.*, 5(318), 2009.

B. Munsky, G. Neuert, and A. von Oudenaarden. Using gene expression noise to understand gene regulation. *Science*, 336(6078):183–187, 2012.

J. M. Raser and E. K. O'Shea. Control of stochasticity in eukaryotic gene expression. *Science*, 304(5678):1811–1814, 2004.

R. Sidje, K. Burrage, and S. MacNamara. Inexact uniformization method for computing transient distributions of Markov chains. *SIAM J. Sci. Comput.*, 29(6): 2562–2580, 2007.

A. Singh and J. P. Hespanha. Approximate moment dynamics for chemically reacting systems. *IEEE Trans. Autom. Control*, 56(2):414–418, 2011.

Y. Taniguchi, P. J. Choi, G.-W. 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.

N. G. van Kampen. *Stochastic processes in physics and chemistry*. North-Holland, Amsterdam, 3rd revised edition, 2007.

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