Multiscale Model. Simul.
Vol. 9, No. 4, pp. 1646–1676

ON REDUCED MODELS FOR THE CHEMICAL MASTER EQUATION*

TOBIAS JAHNKE†

Abstract. The chemical master equation plays a fundamental role for the understanding of gene regulatory networks and other discrete stochastic reaction systems. Solving this equation numerically, however, is usually extremely expensive or even impossible due to the huge size of the state space. Thus, the chemical master equation must often be replaced by a reduced model which operates with a considerably smaller number of degrees of freedom but hopefully still provides the essential information about the dynamics of the full system. We prove error bounds for two reduced models which have previously been proposed in the literature. Based on the error analysis, an alternative model reduction approach for the chemical master equation is introduced and discussed, and its advantage is illustrated by numerical examples.

Key words. chemical master equation, model reduction, hybrid models, error bounds

AMS subject classifications. 34K60, 65C20, 60J27, 92D25

DOI. 10.1137/110821500

1. Introduction.

Many processes in nature can be considered as reaction systems in which $d \in \mathbb{N}$ different species interact via $r \in \mathbb{N}$ reaction channels. The time evolution of such a system is usually modeled by the traditional *reaction-rate equation*, a set of d coupled ordinary differential equations which indicate how the concentrations of the d species change in time. This approach is simple and computationally cheap, but it fails in situations where the influence of inherent stochastic noise cannot be ignored and where certain species have to be described in terms of integer particle numbers instead of real-valued, continuous concentrations. This is the case in gene regulatory networks, viral kinetics with few infectious individuals, and many other biological systems.

The *chemical master equation* (CME) respects the discreteness and randomness of the problem and thus provides a more accurate model. The state of the system is considered as an $\mathbb{N}_0^d$-valued random variable $Z(t)$, which evolves according to a Markov jump process. If $Z(t) = z$ for some state[1] $z = (z_1, \ldots, z_d) \in \mathbb{N}_0^d$, then exactly z_i particles of the ith species exist at time t. The object of interest is the *probability* $p(t, z) = \mathbb{P}(Z(t) = z)$ that at time t the system is in state $z \in \mathbb{N}_0^d$. The probability distribution p is the solution of the CME, a special type of the Kolmogorov forward equation. Unfortunately, solving the CME numerically turns out to be a considerable challenge in most cases because the number of relevant states is so large that traditional methods cannot be applied. Explicit solution formulas are known for monomolecular reaction systems (cf. [20]) but not for general CMEs.

The CME and the reaction-rate equation describe the dynamics on the mesoscopic and the macroscopic scale, respectively. Between these two extremes a number of other

*Received by the editors January 18, 2011; accepted for publication (in revised form) October 13, 2011; published electronically December 6, 2011. The author's research was supported by the "Concept for the Future" of the Karlsruhe Institute of Technology within the framework of the German Excellence Initiative and by the DFG priority program SPP 1324 "Mathematical Methods for Extracting Quantifiable Information from Complex Systems." Some parts of this work have been summarized (without proofs) in an extended abstract in the proceedings of the 7th International Conference on Numerical Analysis and Applied Mathematics, AIP, 2009 [22].

http://www.siam.org/journals/mms/9-4/82150.html

†Fakultät für Mathematik, Institut für Angewandte und Numerische Mathematik, Karlsruhe Institute of Technology (KIT), 76133 Karlsruhe, Germany (tobias.jahnke@kit.edu).

[1]Here and below, all vectors are to be understood as column vectors.

models exist. The reaction-rate equation is supposed to approximate the mean of the solution of the CME, and an improved approximation is obtained by including differential equations for the covariance matrix or even higher moments; cf. [7]. Stochastic effects can be modeled by the *chemical Langevin equation*, a stochastic differential equation obtained by appending an appropriate noise term to the reaction-rate equation. The corresponding probability density is the solution of a high-dimensional Fokker–Planck equation. This approach differs from the CME in that the quantities of the species are real numbers instead of integers. A more elaborate discussion of various models and their relationships can be found, e.g., in [9], [26], [18], [35], [10].

Hybrid models combine different descriptions to represent different parts of the problem. This can be accomplished in various ways, and many hybrid approaches can be found in the literature. The common idea is to represent only the small but "critical" part of the system by an accurate but costly model and to use a "cheaper" model for the other part. When single trajectories are to be generated by stochastic simulation, the goal is usually to remove the fast time-scales from the dynamics (cf., e.g., [14], [32], [5], [1], [33], [36], [6], [2] and references therein), but this is *not* the objective of this article. Our goal is to approximate the associated probability distribution directly (i.e., not via stochastic simulation) and to use hybrid approaches as a model reduction technique which reduces the huge number of degrees of freedom down to a small fraction. Such approaches have been proposed, e.g., in [29], [8], [16], [9], [15]. The reduction of the problem, however, comes at the cost of a lower accuracy because, in addition to the approximation error caused by solving the differential equations numerically, there is now a modeling error due to the fact that the CME is partially replaced by coarser descriptions. The numerical error can usually be investigated with classical numerical analysis and depends on approximation parameters such as, e.g., the step-size used for time-integration. Hence, the numerical error can—in theory, not in practice—be made arbitrarily small by changing the parameter(s) of the chosen method. In contrast, the modeling error depends only on the problem itself, and there is no way to decrease this error. Moreover, much less is known about this type of error. All hybrid models are based on an intuitively plausible approximation, but a detailed error analysis is often not available.

The goal of this paper is twofold: to provide such an error analysis for two of these reduced models (in section 4) and, based on these results, to propose a new reduced model which yields a more accurate description (in section 5). The results are illustrated by numerical experiments in section 6 and summarized in section 7. First of all, however, two introductory sections about the CME (section 2) and the derivation of reduced models (section 3) are provided.

2. The CME.

In this section the main model for discrete stochastic reaction systems is introduced and discussed. The results presented here are not new and can be found elsewhere in the literature, but since certain properties will frequently be used throughout the article, these facts are compiled for the convenience of the reader.

2.1. Discrete stochastic reaction systems.

We consider a system of $d \in \mathbb{N}$ species $S_1, \ldots, S_d$ which interact through $r \in \mathbb{R}$ reaction channels. Each of the species is a group of discrete units (e.g., humans, cells, molecules, etc.) with the same property (e.g., infected, susceptible, or immune individuals in a model of a viral infection). These discrete units will be referred to as "particles" henceforth.

Let $Z_i(t) \in \mathbb{N}_0$ be the number of particles of the ith species at time $t \geq 0$. The vector $Z(t) = (Z_1(t), \ldots, Z_d(t))$ is assumed to evolve according to a Markov jump process. Jumps between states occur whenever the particle number changes because one of

the reaction channels fires. Each reaction channel is uniquely defined by the associated stoichiometric vector $\xi_k \in \mathbb{Z}^d$ and the propensity function $\gamma_k : \mathbb{N}_0^d \to [0, \infty)$. If $Z(t_1) = z$, then at the time $t_2 > t_1$ of the next reaction event, Z jumps to the new state $Z(t_2) = z + \xi_k$, where k is the number of the corresponding reaction channel. The propensity function indicates, roughly speaking, the reactivity of a channel depending on the current state: Up to terms of $\mathcal{O}(dt^2)$, $\gamma_k(z)dt$ is the probability that the kth reaction channel will fire in the infinitesimal time interval $[t_1, t_1 + dt)$. A rigorous derivation can be found in [11], [12]. Suppose that the reaction scheme of the kth reaction channel is

$$n_1 S_1 + \cdots + n_d S_d \to m_1 S_1 + \cdots + m_d S_d \tag{2.1}$$

with $n_i, m_j \in \mathbb{N}$. Then the propensity function is typically

$$\gamma_k(z) = c_k \binom{z_1}{n_1} \cdot \ldots \cdot \binom{z_d}{n_d} \tag{2.2}$$

with some reaction constant $c_k > 0$. Other propensity functions may be used in special cases, e.g., if inhibition by other species is modeled. The stoichiometric vector of (2.1) is simply $(m_1 - n_1, \ldots, m_d - n_d)$.

Since individual realizations of such a system are random, one is interested in the probability $p(t, z) = \mathbb{P}(Z(t) = z)$ that there are z_i particles of S_i at time t. The probability distribution p is the solution of the CME

$$\partial_t p(t, z) = \sum_{k=1}^{r} (\gamma_k(z - \xi_k) p(t, z - \xi_k) - \gamma_k(z) p(t, z)), \qquad z \in \mathbb{N}_0^d \tag{2.3}$$

(cf. [12]) with $\gamma_k(z - \xi_k) = p(t, z - \xi_k) = 0$ for all $z - \xi_k \notin \mathbb{N}_0^d$. Since $p(t, z)$ has to be computed for each state z in a large and high-dimensional state space, solving the CME is usually highly nontrivial and computationally expensive. On the other hand, the full solution may contain far more information than actually required because one is only interested in a few species, say, $S_1, \ldots, S_{d_1}$, which are considered as the "critical" ones ($0 < d_1 < d$). All other species $S_{d_1+1}, \ldots, S_d$ are only relevant because they interact with the critical species. In this situation, the goal is only to compute the time-dependent *marginal distribution* $\sum_{z_{d_1+1}} \ldots \sum_{z_d} p(t, z)$. Since it is very inefficient (or even impossible) to solve the full CME just in order to obtain a low-dimensional marginal distribution, the question arises if an approximation of the marginal distribution can still be obtained if the behavior of the "secondary" species is modeled in a coarser but cheaper way. Before such model reduction approaches are studied, a closer look on the CME for the full system is necessary.

2.2. Partition of the problem. It is convenient to use the partition

$$\begin{aligned} z = (x, y), \qquad & x = (x_1, \ldots, x_{d_1}) = (z_1, \ldots, z_{d_1}) \in \mathbb{N}_0^{d_1}, \\ & y = (y_1, \ldots, y_{d_2}) = (z_{d_1+1}, \ldots, z_{d_1+d_2}) \in \mathbb{N}_0^{d_2} \end{aligned} \tag{2.4}$$

($0 < d_1 < d, d_2 = d - d_1$) of the state vector z and to decompose the stoichiometric vectors into

$$\begin{aligned} \xi_k = (\nu_k, \mu_k), \qquad & \nu_k = (\nu_{k,1}, \ldots, \nu_{k,d_1}) = (\xi_{k,1}, \ldots, \xi_{k,d_1}) \in \mathbb{Z}^{d_1}, \\ & \mu_k = (\mu_{k,1}, \ldots, \mu_{k,d_2}) = (\xi_{k,d_1+1}, \ldots, \xi_{k,d_1+d_2}) \in \mathbb{Z}^{d_2} \end{aligned} \tag{2.5}$$

for every $k = 1, \dots, r$. Moreover, the propensity of each reaction is assumed to be separable: there are functions $\alpha : \mathbb{N}_0^{d_1} \to [0, \infty)$ and $\beta : \mathbb{N}_0^{d_2} \to [0, \infty)$ such that

$$\gamma_k(z) = \gamma_k(x, y) = \alpha_k(x)\beta_k(y). \tag{2.6}$$

This assumption is satisfied by the standard propensities (2.2) and in a vast majority of applications even by nonstandard propensities. The decomposition (2.6) is not unique because the reaction constant c_k from (2.2) can be either included in α_k or β_k or partly in both. It turns out, however, that this nonuniqueness does not cause any problems because in the CME (2.7) and in the reduced models (3.5)–(3.6), (3.10)–(3.11), and (5.8)–(5.9), α_k or β_k never occur alone, but only appear in products of the type $\alpha_k(\cdots)\beta_k(\cdots)$.

With the partition (2.4)–(2.6), the full CME (2.3) reads

$$\begin{aligned} \partial_t p(t, x, y) &= (\mathcal{A}p)(t, x, y) \\ &= \sum_{k=1}^{r} (\alpha_k(x - \nu_k)\beta_k(y - \mu_k)p(t, x - \nu_k, y - \mu_k) - \alpha_k(x)\beta_k(y)p(t, x, y)). \end{aligned} \tag{2.7}$$

It will always be assumed that the initial data $p_0(x, y) = p(0, x, y)$ are a probability distribution, i.e., $p_0(x, y) \geq 0$ for all $(x, y) \in \mathbb{N}_0^d$ and $\sum_{x \in \mathbb{N}_0^{d_1}} \sum_{y \in \mathbb{N}_0^{d_2}} p_0(x, y) = 1$.

2.3. State space truncation. For numerical computations we consider only the truncated state space

$$\Omega_{(\bar{x},\bar{y})} = \Omega_{\bar{x}} \times \Omega_{\bar{y}}, \quad \Omega_{\bar{x}} = \{x \in \mathbb{N}_0^{d_1} : x < \bar{x}\}, \quad \Omega_{\bar{y}} = \{y \in \mathbb{N}_0^{d_2} : y < \bar{y}\}. \tag{2.8}$$

Here, $(\bar{x}, \bar{y}) \in \mathbb{N}^d$ is some suitably chosen truncation vector, and the relations $x < \bar{x}$ and $y < \bar{y}$ are to be understood entrywise. In the CME (2.7) the propensities and the distribution must be evaluated at $(x - \nu_k, y - \mu_k)$, and for certain (x, y) and (ν_k, μ_k), this state might lie outside of $\Omega_{(\bar{x},\bar{y})}$. Therefore, we formally extend the propensities and the distribution by defining

$$\alpha_k(x) = 0 \quad \text{if } x \notin \Omega_{\bar{x}}, \quad \beta_k(y) = 0 \quad \text{if } y \notin \Omega_{\bar{y}}, \quad p(t, x, y) = 0 \quad \text{if } (x, y) \notin \Omega_{(\bar{x},\bar{y})}. \tag{2.9}$$

The interpretation of (2.9) is that there is no "probability inflow" from the outside of $\Omega_{(\bar{x},\bar{y})}$ into the truncated state space. On the boundaries, we impose the discrete Neumann boundary conditions

$$\alpha_k(x) = 0 \quad \text{if } (x + \nu_k) \notin \Omega_{\bar{x}}, \qquad \beta_k(y) = 0 \quad \text{if } (y + \mu_k) \notin \Omega_{\bar{y}}. \tag{2.10}$$

This means that the Markov jump process cannot leave $\Omega_{(\bar{x},\bar{y})}$ because all reaction channels that would cause a jump from $(x, y) \in \Omega_{(\bar{x},\bar{y})}$ to $(x + \nu_k, y + \mu_k) \notin \Omega_{(\bar{x},\bar{y})}$ are suppressed. On the natural boundary

$$\{(x, y) \in \Omega_{(\bar{x},\bar{y})} : x + \nu_k \not\geq 0 \text{ or } y + \mu_k \not\geq 0 \text{ for some } k = 1, \dots, r\}$$

the Neumann boundary condition (2.10) is always satisfied because propensity functions are always defined in such a way that the particle numbers remain nonnegative.

From now on, we consider the CME (2.7) on the truncated state space unless otherwise stated. The restriction of the operator (2.7) to the finite state space is again denoted

by $\mathcal{A}$. It will be tacitly assumed that $\mathbf{\Omega}_{(\bar{x},\bar{y})}$ is so large that $p(t,\cdot)$ almost vanishes close to the artificial boundary such that the truncation error can be neglected. For functions $q\colon\mathbf{\Omega}_{(\bar{x},\bar{y})} \to \mathbb{R}$ we define the norm

$$\|q\| = \sum_{x\in\mathbf{\Omega}_{\bar{x}}} \sum_{y\in\mathbf{\Omega}_{\bar{y}}} |q(x,y)|. \tag{2.11}$$

2.4. Properties of the CME. An important property of the operator $\mathcal{A}$ on the bounded state space with boundary condition (2.10) is the fact that

$$\sum_{x\in\mathbf{\Omega}_{\bar{x}}} \sum_{y\in\mathbf{\Omega}_{\bar{y}}} (\mathcal{A}q)(x,y) = 0 \tag{2.12}$$

for any function $q\colon\mathbf{\Omega}_{(\bar{x},\bar{y})} \to \mathbb{R}$. This can be checked as follows. With the abbreviation $f_k(x,y) = \alpha_k(x)\beta_k(y)p(x,y)$ we obtain

$$\sum_{x\in\mathbf{\Omega}_{\bar{x}}} \sum_{y\in\mathbf{\Omega}_{\bar{y}}} (\mathcal{A}p)(x,y) = \sum_{k=1}^{r} \sum_{x\in\mathbf{\Omega}_{\bar{x}}} \sum_{y\in\mathbf{\Omega}_{\bar{y}}} (f_k(x-\nu_k, y-\mu_k) - f_k(x,y)).$$

For every fixed $k \in \{1,\dots,r\}$ let

$$\mathring{\mathbf{\Omega}}_k = \{(x,y) \in \mathbf{\Omega}_{(\bar{x},\bar{y})}\colon (x+\nu_k, y+\mu_k) \in \mathbf{\Omega}_{(\bar{x},\bar{y})}\}$$

be the "interior" of $\mathbf{\Omega}_{(\bar{x},\bar{y})}$. Then, according to (2.10),

$$\sum_{x\in\mathbf{\Omega}_{\bar{x}}} \sum_{y\in\mathbf{\Omega}_{\bar{y}}} f_k(x,y) = \sum_{(x,y)\in\mathring{\mathbf{\Omega}}_k} f_k(x,y) \tag{2.13}$$

and, with $\tilde{x} = x - \nu_k$ and $\tilde{y} = y - \mu_k$,

$$\sum_{x\in\mathbf{\Omega}_{\bar{x}}} \sum_{y\in\mathbf{\Omega}_{\bar{y}}} f_k(x-\nu_k, y-\mu_k) = \sum_{\tilde{x}+\nu_k\in\mathbf{\Omega}_{\bar{x}}} \sum_{\tilde{y}+\mu_k\in\mathbf{\Omega}_{\bar{y}}} f_k(\tilde{x},\tilde{y}) = \sum_{(\tilde{x},\tilde{y})\in\mathring{\mathbf{\Omega}}_k} f_k(\tilde{x},\tilde{y}) \tag{2.14}$$

due to (2.9). Since (2.13) and (2.14) coincide, this shows that

$$\sum_{x\in\mathbf{\Omega}_{\bar{x}}} \sum_{y\in\mathbf{\Omega}_{\bar{y}}} (f_k(x-\nu_k, y-\mu_k) - f_k(x,y)) = 0$$

for every $k = 1,\dots,r$, and hence (2.12) follows. As a consequence of (2.12), the total mass $\sum_{x\in\mathbf{\Omega}_{\bar{x}}} \sum_{y\in\mathbf{\Omega}_{\bar{y}}} p(t,x,y)$ of the solution of the CME is conserved because

$$\sum_{x\in\mathbf{\Omega}_{\bar{x}}} \sum_{y\in\mathbf{\Omega}_{\bar{y}}} \partial_t p(t,x,y) = \sum_{x\in\mathbf{\Omega}_{\bar{x}}} \sum_{y\in\mathbf{\Omega}_{\bar{y}}} (\mathcal{A}p)(t,x,y) = 0.$$

Moreover, the solution $p(t,x,y)$ stays nonnegative for all times if this is true for $p(0,x,y)$: If $p(t,x_\star,y_\star) = 0$ for some state $(x_\star,y_\star) \in \mathbf{\Omega}_{(\bar{x},\bar{y})}$ and $p(t,x,y) \geq 0$ for all $(x,y) \in \mathbf{\Omega}_{(\bar{x},\bar{y})}$, then (2.7) implies that $\partial_t p(t,x_\star,y_\star) \geq 0$ such that $p(t,x_\star,y_\star)$ cannot become negative. Together, these properties imply that the solution of the CME is a probability distribution if $p(0,x,y)$ is a probability distribution.

Equation (2.12) also implies that at least one of the eigenvalues of $\mathcal{A}$ is zero, and with the Gershgorin circle theorem (cf., e.g., section 2.3 in the appendix of [4]) it can be

shown that all nonzero eigenvalues have a strictly negative real part. Hence, if $p(0, x, y)$ is a probability distribution, then the solution of the CME will converge to a stationary distribution $\rho = \rho(x, y)$ with $\mathcal{A}\rho = 0$.

2.5. Relation to the reaction-rate equation. Under certain conditions, the traditional reaction-rate equation yields an approximation to the mean

$$\mathbb{E}(t) = \begin{pmatrix} \mathbb{E}_x(t) \\ \mathbb{E}_y(t) \end{pmatrix} = \sum_{x\in\Omega_{\tilde{x}}} \sum_{y\in\Omega_{\tilde{y}}} \begin{pmatrix} x \\ y \end{pmatrix} p(t, x, y).$$

It follows from the CME that

$$\begin{pmatrix} \dot{\mathbb{E}}_x \\ \dot{\mathbb{E}}_y \end{pmatrix} = \sum_{x\in\Omega_{\tilde{x}}} \sum_{y\in\Omega_{\tilde{y}}} \begin{pmatrix} x \\ y \end{pmatrix} (\mathcal{A}p)(t, x, y) = \sum_{k=1}^{r} \begin{pmatrix} \nu_k \\ \mu_k \end{pmatrix} \sum_{x\in\Omega_{\tilde{x}}} \sum_{y\in\Omega_{\tilde{y}}} \alpha_k(x)\beta_k(y)p(t, x, y)$$

(cf. [7]), where we have used that according to (2.9) and (2.10)

$$\begin{aligned} &\sum_{x\in\Omega_{\tilde{x}}} \sum_{y\in\Omega_{\tilde{y}}} \begin{pmatrix} x \\ y \end{pmatrix} \alpha_k(x - \nu_k)\beta_k(y - \mu_k)p(t, x - \nu_k, y - \mu_k) \\ &\quad = \sum_{\tilde{x}+\nu_k\in\Omega_{\tilde{x}}} \sum_{\tilde{y}+\mu_k\in\Omega_{\tilde{y}}} \begin{pmatrix} \tilde{x} + \nu_k \\ \tilde{y} + \mu_k \end{pmatrix} \alpha_k(\tilde{x})\beta_k(\tilde{y})p(t, \tilde{x}, \tilde{y}) \\ &\quad = \sum_{\tilde{x}\in\Omega_{\tilde{x}}} \sum_{\tilde{y}\in\Omega_{\tilde{y}}} \begin{pmatrix} \tilde{x} + \nu_k \\ \tilde{y} + \mu_k \end{pmatrix} \alpha_k(\tilde{x})\beta_k(\tilde{y})p(t, \tilde{x}, \tilde{y}). \end{aligned}$$

In order to obtain a closed differential equation, the mean of the propensity is approximated by the propensity of the mean as follows:

$$\sum_{x\in\Omega_{\tilde{x}}} \sum_{y\in\Omega_{\tilde{y}}} \alpha_k(x)\beta_k(y)p(t, x, y) \approx \alpha_k(\mathbb{E}_x(t)) \cdot \beta_k(\mathbb{E}_y(t)). \tag{2.15}$$

This yields

$$\begin{pmatrix} \dot{\mathbb{E}}_x \\ \dot{\mathbb{E}}_y \end{pmatrix} \approx \sum_{k=1}^{r} \begin{pmatrix} \nu_k \\ \mu_k \end{pmatrix} \alpha_k(\mathbb{E}_x) \cdot \beta_k(\mathbb{E}_y),$$

and hence one may hope that the solution (η_x, η_y) of the reaction-rate equation

$$\begin{pmatrix} \dot{\eta}_x \\ \dot{\eta}_y \end{pmatrix} = \sum_{k=1}^{r} \begin{pmatrix} \nu_k \\ \mu_k \end{pmatrix} \alpha_k(\eta_x) \cdot \beta_k(\eta_y) \tag{2.16}$$

approximates $\mathbb{E}(t)$. If all propensities are at most linear, i.e., if

$$\alpha_k(\eta_x) = c_{k,1}x + c_{k,2}, \quad \beta_k(\eta_y) = c_{k,3} \quad \text{or } \alpha_k(\eta_x) = c_{k,1}, \quad \beta_k(\eta_y) = c_{k,2}y + c_{k,3}$$

for all $k = 1, \ldots, r$ with some constants $c_{k,1}, c_{k,2}, c_{k,3} \in \mathbb{R}$, then the approximation (2.15) is *exact*, and consequently $(\eta_x, \eta_y) = (\mathbb{E}_x, \mathbb{E}_y)$. For more complicated systems, however, the reaction-rate equation does not provide a reasonable approximation to the true dynamics.

3. Model reduction.

3.1. Product approximation. In a first step, the solution $p(t,x,y)$ is supposed to be approximated by the direct product $(u\otimes v)(t,x,y)=u(t,x)v(t,y)$ of two probability distributions $u(t,x)$ and $v(t,y)$, which depend only on time and x or y, respectively. Clearly, it cannot be expected that $u\otimes v$ is an exact solution of the CME (2.7), i.e.,

$$\partial_t(u\otimes v)(t,x,y)\neq\mathcal{A}(u\otimes v)(t,x,y). \tag{3.1}$$

In order to derive differential equations for u and v we impose the condition that the "marginals" of both sides of (3.1) coincide as follows:

$$\sum_y\partial_t(u\otimes v)(t,x,y)=\sum_y\mathcal{A}(u\otimes v)(t,x,y)\quad\text{for all } x\in\mathbf{\Omega}_{\tilde{x}}, \tag{3.2}$$

$$\sum_x\partial_t(u\otimes v)(t,x,y)=\sum_x\mathcal{A}(u\otimes v)(t,x,y)\quad\text{for all } y\in\mathbf{\Omega}_{\tilde{y}}. \tag{3.3}$$

Here and below, the shorthand notation $\sum_x$ and $\sum_y$ is used for $\sum_{x\in\mathbf{\Omega}_{\tilde{x}}}$ and $\sum_{y\in\mathbf{\Omega}_{\tilde{y}}}$, respectively. We assume (and verify below) that

$$\sum_x\partial_t u(t,x)=0\quad\text{and}\quad\sum_y\partial_t v(t,y)=0 \tag{3.4}$$

for all $t\geq 0$, which implies $\sum_x u(t,x)=1$ and $\sum_y v(t,y)=1$ for all $t>0$, provided that this is true for $t=0$. As a consequence, the left-hand side of (3.2) reduces to

$$\sum_y\partial_t(u\otimes v)(t,x,y)=\sum_y\partial_t u(t,x)\cdot v(t,y)+\sum_y u(t,x)\cdot\partial_t v(t,y)=\partial_t u(t,x),$$

whereas the right-hand side of (3.2) is

$$\begin{aligned}\sum_y\mathcal{A}(u\otimes v)(t,x,y)=&\sum_{k=1}^r\Big(\sum_y\beta_k(y-\mu_k)v(t,y-\mu_k)\Big)\alpha_k(x-\nu_k)u(t,x-\nu_k)\\&-\sum_{k=1}^r\Big(\sum_y\beta_k(y)v(t,y)\Big)\alpha_k(x)u(t,x).\end{aligned}$$

As before, one obtains with (2.9), (2.10), and $\tilde{y}=y-\mu_k$ that

$$\sum_y\beta_k(y-\mu_k)v(t,y-\mu_k)=\sum_{\tilde{y}+\mu_k}\beta_k(\tilde{y})v(t,\tilde{y})=\sum_{\tilde{y}}\beta_k(\tilde{y})v(t,\tilde{y}).$$

The same steps can be carried out mutatis mutandis for (3.3). This yields a coupled system of evolution equations for $u(t,x)$ and $v(t,y)$:

$$\partial_t u(t,x)=\sum_{k=1}^r\Big(\sum_y\beta_k(y)v(t,y)\Big)(\alpha_k(x-\nu_k)u(t,x-\nu_k)-\alpha_k(x)u(t,x)), \tag{3.5}$$

$$\partial_t v(t,y)=\sum_{k=1}^r\Big(\sum_x\alpha_k(x)u(t,x)\Big)(\beta_k(y-\mu_k)v(t,y-\mu_k)-\beta_k(y)v(t,y)). \tag{3.6}$$

Similar to section 2.4 it can easily be shown that the solution of (3.5)–(3.6) does indeed satisfy the condition (3.4). Suitable initial data for the new differential equations are $u(0,x) = \sum_y p(0,x,y)$ and $v(0,y) = \sum_x p(0,x,y)$.

LEMMA 1. *If $u(0,x)$ and $v(0,y)$ are probability distributions, then the solutions $u(t,x)$ and $v(t,y)$ of the reduced model* (3.5)–(3.6) *remain a probability distribution for all times.*

Proof. Equation (3.4) implies the mass conservation $\sum_x u(t,x) = 1$ and $\sum_y v(t,y) = 1$ for all $t \geq 0$. It remains to prove that $u(t,x) \geq 0$ and $v(t,y) \geq 0$ for all $t \geq 0$ and $(x,y) \in \boldsymbol{\Omega}_{(\bar{x},\bar{y})}$. Suppose that there is a $t^\star$ and state $x^\star$ such that

$$u(t^\star, x^\star) = 0, \quad u(t^\star, x) \geq 0 \quad \text{for all } x \neq x^\star, \quad v(t^\star, y) \geq 0 \quad \text{for all } y.$$

Then we have

$$\begin{aligned}\partial_t u(t^\star, x^\star) &= \sum_{k=1}^{r} \left(\sum_y \beta_k(y) v(t^\star, y) \right) (\alpha_k(x^\star - \nu_k) u(t^\star, x^\star - \nu_k) - \alpha_k(x^\star) u(t^\star, x^\star)) \\ &\geq 0\end{aligned}$$

because, by assumption, $\alpha_k(x^\star)u(t,x^\star) = 0$ and all other terms are positive. Hence, $u(t,x^\star)$ cannot decrease below 0. The same argument applies to all other x and to $v(t,y)$. ∎

Discussion. While the full CME (2.7) involves $\prod_{i=1}^{d_1} \bar{x}_i \cdot \prod_{j=1}^{d_2} \bar{y}_j$ unknowns, the reduced model (3.5)–(3.6) contains only $\prod_{i=1}^{d_1} \bar{x}_i + \prod_{j=1}^{d_2} \bar{y}_j$ degrees of freedom, which is a significant reduction. The *linear* CME is now replaced by two *nonlinear* equations, but this is not really an inconvenience because each of these equations inherits the structure of a CME. In (3.5) the propensities α_k are multiplied by the factor $\sum_y \beta_k(y) v(t,y)$ and vice versa for (3.6). The two equations are only coupled via these factors such that for *fixed* $v(t,y)$ the differential equation (3.5) is linear, and for *fixed* $u(t,x)$ (3.6) is linear. Hence, the reduced model can be conveniently solved with the second-order Strang splitting method:

1. Choose a step-size $h > 0$, and set $n = 0$.
2. Half-step in $u(t,x)$: Compute an approximation $u_{n+1/2}(x)$ by solving (3.5) from t to $t + h/2$ while keeping $v_n(y)$ fixed.
3. Full step in $v(t,y)$: Compute an approximation $v_{n+1}(x)$ by solving (3.6) from t to $t + h$ while keeping $u_{n+1/2}(x)$ fixed.
4. Half-step in $u(t,x)$: Compute an approximation $u_{n+1}(x)$ by solving (3.5) from $t + h/2$ to $t + h$ while keeping $v_{n+1}(y)$ fixed.
5. Set $n = n + 1$, and go to 2.

In each substep only a lower-dimensional linear CME has to be solved.

The product approximation is akin to the *dynamical low-rank approximation*; cf. [27], [28]. In that approach, the solution of an evolution equation is represented by a sum of rank-1 tensor products which are propagated along with the solution. This strategy has been applied to the CME in [21].

3.2. Hellander–Lötstedt model. The number of degrees of freedom can be further decreased if instead of $v(t,y)$ only its approximative expectation

$$\eta(t) \approx \sum_y y v(t,y) \tag{3.7}$$

is considered. Taking the derivative on both sides of (3.7) yields via (3.6)

$$\dot\eta(t) \approx \sum_{k=1}^{r}\Big(\sum_x \alpha_k(x)u(t,x)\Big)\sum_y y(\beta_k(y-\mu_k)v(t,y-\mu_k)-\beta_k(y)v(t,y))$$

$$= \sum_{k=1}^{r}\Big(\sum_x \alpha_k(x)u(t,x)\Big)\mu_k\sum_y \beta_k(y)v(t,y), \tag{3.8}$$

where we have used again that (2.9) and (2.10) imply

$$\sum_y y\beta_k(y-\mu_k)v(t,y-\mu_k) = \sum_{\tilde y}(\tilde y+\mu_k)\beta_k(\tilde y)v(t,\tilde y)$$

(cf. section 2.5). Now we have to approximate $\sum_y \beta_k(y)v(t,y)$ in terms of $\eta(t)$:

$$\sum_y \beta_k(y)v(t,y) \approx \beta_k\Big(\sum_y yv(t,y)\Big) \approx \beta_k(\eta(t)). \tag{3.9}$$

The first approximation is essentially the same as in (2.15), and the second follows from (3.7). Substituting (3.9) into (3.8) and (3.5) yields coupled differential equations for $w(t,x)\approx u(t,x)$ and $\eta(t)$:

$$\partial_t w(t,x) = \sum_{k=1}^{r}\beta_k(\eta(t))(\alpha_k(x-\nu_k)w(t,x-\nu_k)-\alpha_k(x)w(t,x)), \tag{3.10}$$

$$\dot\eta(t) = \sum_{k=1}^{r}\Big(\sum_x \alpha_k(x)w(t,x)\Big)\mu_k\beta_k(\eta(t)). \tag{3.11}$$

This approach for reducing the CME has been proposed by Hellander and Lötstedt in [16] using a different derivation. Although similar models are known in other application areas (cf. [3], [13], [29]), (3.10)–(3.11) will be called the *Hellander–Lötstedt model* henceforth. In this model, the marginal distribution w solves a CME with propensities depending on η, whereas η solves a classical reaction-rate equation similar to (2.16), but including factors depending on w. The total number of degrees of freedom is now reduced to $\prod_{i=1}^{d_1}\bar x_i + d_2$. As before, the evolution equations can be solved with the Strang splitting method. In [16] a method based on stochastic simulations has been analyzed.

LEMMA 2. *Let $w(t,x)$ and $\eta(t)$ be the solution of the Hellander–Lötstedt model* (3.10)–(3.11). *If $w(0,x)$ is a probability distribution and $\eta(0)\ge 0$, then $w(t,x)$ remains a probability distribution and $\eta(t)\ge 0$ for all times $t>0$.*

Proof. The fact that $w(t,x)$ remains a probability distribution can be proved by adapting the arguments from the proof of Lemma 1. Suppose that for some $t^\star\ge 0$ we have $\eta_i(t^\star)=0$ and $\eta_j(t^\star)\ge 0$ for all $j\ne i$. Then (3.11) yields

$$\dot\eta_i(t^\star) = \sum_{k=1}^{r}\Big(\sum_x \alpha_k(x)w(t^\star,x)\Big)\mu_{k,i}\beta_k(\eta(t^\star)),$$

where $\mu_{k,i}$ denotes the ith entry of the kth stoichiometric vector. If $\mu_{k,i}<0$, then $\beta_k(\eta(t^\star))=0$ due to (2.10). Hence $\dot\eta_i(t^\star)\ge 0$, and $\eta_i(t)$ cannot decrease below 0. ∎

4. Error estimates for the reduced models. The derivation in the previous section raises the question of how accurately the reduced models (3.5)–(3.6) or

(3.10)–(3.11) approximate the solution of the full CME (2.7). This question is now investigated. In applications, the evolution equations of each model have to be solved by applying a suitable numerical method (such as, e.g., the Strang splitting) which causes an additional *numerical error*. In this article, however, only the *modeling error* of the reduced models is analyzed.

4.1. Accuracy of the product approximation.

4.1.1. Exact solutions in special cases.

PROPOSITION 1. *Suppose that the exact solution of the CME* (2.7) *admits an exact product representation: there are probability distributions* $\hat{u}(t,x)$ *and* $\hat{v}(t,y)$ *such that* $p(t,x,y) = (\hat{u} \otimes \hat{v})(t,x,y) = \hat{u}(t,x)\hat{v}(t,y)$. *Let* $u(t,x)$ *and* $v(t,y)$ *be the solution of the product approximation* (3.5)–(3.6). *If* $u(0,x) = \hat{u}(0,x)$ *and* $v(0,y) = \hat{v}(0,y)$, *then the solution of the product approximation is exact:*

$$u(t,x) = \hat{u}(t,x), \quad v(t,y) = \hat{v}(t,y), \quad u(t,x)v(t,y) = p(t,x,y)$$

for all $t \geq 0$, $(x,y) \in \Omega_{(\bar{x},\bar{y})}$.

Proof. Since $p(t,x,y) = (\hat{u} \otimes \hat{v})(t,x,y)$ is the exact solution of the CME, we have (in contrast to (3.1))

$$\partial_t(\hat{u} \otimes \hat{v})(t,x,y) = \mathcal{A}(\hat{u} \otimes \hat{v})(t,x,y).$$

Taking sums with respect to x or y, respectively, shows that $\hat{u}$ and $\hat{v}$ solve (3.5)–(3.6). Since $u(0,x) = \hat{u}(0,x)$ and $v(0,y) = \hat{v}(0,y)$, the uniqueness of the solution implies $u(t,x) = \hat{u}(t,x)$ and $v(t,y) = \hat{v}(t,y)$. ∎

Proposition 1 applies not only to the trivial case where the subsystems represented by the variables x and y are decoupled, but also to monomolecular reaction systems. A reaction system is called monomolecular if and only if all of its reaction channels are monomolecular, and a reaction channel is called monomolecular if it belongs to one of the following three classes ($j \neq i$):

Reaction channel		$\alpha(x)$	$\beta(y)$	
Conversion	$S_j \to S_i$	$c \cdot x_j$	1	if $1 \leq j \leq d_1$
		1	$c \cdot y_{j-d_1}$	if $d_1 < j \leq d$
Degradation	$S_j \to \varnothing$	$c \cdot x_j$	1	if $1 \leq j \leq d_1$
		1	$c \cdot y_{j-d_1}$	if $d_1 < j \leq d$
Inflow	$\star \to S_j$	c	1	if $1 \leq j \leq d_1$
		1	c	if $d_1 < j \leq d$

Of course, the reaction constant $c > 0$ may have different values in different reaction channels. Here and below, $\star$ denotes a constant "source" and $\varnothing$ is the group of "dead" or "disappeared" particles. Monomolecular reaction systems are special because the solution of the associated CME is explicitly known (cf. [20]) which can be used, e.g., to construct efficient stochastic simulation methods (cf. [23]). For the tensor product model, we have the following corollary.

COROLLARY 1. *Consider the CME* (2.7) *and the product approximation* (3.5)–(3.6) *without any truncation of the state space (i.e., let* $(\bar{x},\bar{y}) = (\infty,\ldots,\infty)$ *in* (2.8)*). Suppose that all reaction channels are monomolecular and that* $p_0(x,y)$ *is a product Poisson*

distribution. If $u(0,x)=\sum_y p_0(x,y)$ *and* $v(0,y)=\sum_x p_0(x,y)$*, then the product approximation reproduces the exact solution of the full CME, i.e.,*

$$u(t,x)v(t,y)=p(t,x,y)\quad \textit{for all } t\geq 0, x\geq 0, y\geq 0.$$

Proof. Under the above assumptions the solution of the full CME (2.7) remains a product Poisson distribution for all times with a time-dependent parameter; cf. Proposition 2 of [20]. Hence, Proposition 1 applies and yields the assertion. ∎

4.1.2. A posteriori error bound. In most applications, an exact tensor product representation $p(t,x,y)=(\hat{u}\otimes\hat{v})(t,x,y)$ of the CME solution is not possible, and the reduced model (3.5)–(3.6) will only yield an approximation. Before we derive an error bound for this approximation, we first give a geometric interpretation of the differential equations (3.5)–(3.6).

Let $\mathcal{M}$ be the manifold of all tensor products of probability distributions:

$$\mathcal{M}=\left\{(q_1\otimes q_2)(x,y)=q_1(x)q_2(y)\middle|\sum_x q_1(x)=\sum_y q_2(y)=1,\, q_1(x),\, q_2(y)\geq 0\right\}.$$

The tangent space in $q_1\otimes q_2\in\mathcal{M}$ is

$$\begin{aligned}\mathcal{T}_{q_1\otimes q_2}\mathcal{M}&=\Bigg\{\delta\varphi:\Omega_{(\bar{x},\bar{y})}\to\mathbb{R}\big|\text{there is a differentiable curve }\varphi:\mathbb{R}\to\mathcal{M}\\&\qquad\text{such that }\varphi(0)=q_1\otimes q_2\text{ and }\delta\varphi=\frac{d\varphi}{dt}(0)\Bigg\}\\&=\left\{\delta q_1\otimes q_2+q_1\otimes\delta q_2\middle|\sum_x\delta q_1(x)=\sum_y\delta q_2(y)=0\right\}.\end{aligned}$$

Let $f(x,y)$ have the property that $\sum_x\sum_y f(x,y)=0$. Then a projection of f into the tangent space of $q_1\otimes q_2\in\mathcal{M}$ is defined by

$$\mathcal{P}_{q_1\otimes q_2}:f\mapsto\left(\sum_y f\right)\otimes q_2+q_1\otimes\left(\sum_x f\right).$$

In particular, the condition $\sum_x\sum_y f(x,y)=0$ holds for $f(x,y)=(\mathcal{A}q)(x,y)$ with any $q(x,y)$; cf. (2.12).

If u and v are the solutions of the product approximation (3.5)–(3.6), then $u\otimes v$ solves the differential equation

$$\partial_t(u\otimes v)=\mathcal{P}_{u\otimes v}\mathcal{A}(u\otimes v)\tag{4.1}$$

because by definition of $\mathcal{P}_{u\otimes v}$

$$\begin{aligned}\mathcal{P}_{u\otimes v}\mathcal{A}(u\otimes v)&=\left(\sum_y\mathcal{A}(u\otimes v)\right)\otimes v+u\otimes\left(\sum_x\mathcal{A}(u\otimes v)\right)\\&=\left(\sum_y\partial_t(u\otimes v)\right)\otimes v+u\otimes\left(\sum_x\partial_t(u\otimes v)\right)\\&=\partial_t u\otimes v+u\otimes\partial_t v\;=\;\partial_t(u\otimes v).\end{aligned}$$

THEOREM 1. *Let $p(t,x,y)$ be the solution of the full CME $\partial_t p = \mathcal{A}p$, and let $u(t,x)$ and $v(t,y)$ be the solution of* (3.5)–(3.6). *Then the error $\varepsilon(t) = \varepsilon(t,x,y) = p(t,x,y) - u(t,x)v(t,y)$ is bounded by*

$$\|\varepsilon(t)\| \le \|\varepsilon(0)\| + \int_0^t \|(I - \mathcal{P}_{(u\otimes v)(s)})\mathcal{A}(u\otimes v)(s)\|\,ds, \tag{4.2}$$

where $\|\cdot\|$ is the norm defined in (2.11).

Proof. Comparing the full CME $\partial_t p = \mathcal{A}p$ with (4.1) shows that the error solves

$$\partial_t \varepsilon = \mathcal{A}\varepsilon + (I - \mathcal{P}_{u\otimes v})\mathcal{A}(u\otimes v)$$

with initial data $\varepsilon(0,x,y) = p(0,x,y) - u(0,x)v(0,y)$. The variation-of-constant formula yields (upon omitting the arguments x and y)

$$\varepsilon(t) = \exp(t\mathcal{A})\varepsilon_0 + \int_0^t \exp((t-s)\mathcal{A})(I - \mathcal{P}_{(u\otimes v)(s)})\mathcal{A}(u\otimes v)(s)\,ds, \tag{4.3}$$

and with $\|\exp(t\mathcal{A})\| = 1$ the estimate (4.2) follows. ∎

Discussion. In the product approximation the vector field generated by the CME operator is projected into the tangent space. This ensures that the solution $u\otimes v$ stays on the manifold $\mathcal{M}$. Theorem 1 states that integrating the projection error $\|(I - \mathcal{P}_{u\otimes v})\mathcal{A}(u\otimes v)\|$ gives an upper bound for the error of the product approximation. The above result is an a posteriori bound; i.e., the solution $u\otimes v$ appears on the right-hand side of the error bound. In section II.1.6 of [30] a very similar result has been shown in the context of the time-dependent Schrödinger equation.

A refined estimate is possible by taking into account that every solution of the CME on $\mathbf{\Omega}_{(\bar{x},\bar{y})}$ converges to a stationary distribution $\rho = \rho(x,y)$; cf. section 2.4. For simplicity, now we assume that ρ is unique, and thus $\lim_{t\to\infty}\exp(t\mathcal{A})p_0 = \rho$ for all probability distributions p_0. Then the limit operator $\mathcal{S} = \lim_{t\to\infty}\exp(t\mathcal{A})$ is given by

$$\mathcal{S}\colon q(x,y) \mapsto \rho(x,y)\sum_{\xi}\sum_{\zeta} q(\xi,\zeta).$$

This can be seen as follows. For any state $(\xi,\zeta)\in\mathbf{\Omega}_{(\bar{x},\bar{y})}$ we define

$$\delta_{(\xi,\zeta)}(x,y) = \begin{cases} 1 & \text{if } x=\xi \quad \text{and} \quad y=\zeta, \\ 0 & \text{else.} \end{cases}$$

Then every function $q\colon\mathbf{\Omega}_{(\bar{x},\bar{y})}\to\mathbb{R}$ can be represented as a (finite) linear combination

$$q(x,y) = \sum_{\xi}\sum_{\zeta} q(\xi,\zeta)\cdot\delta_{(\xi,\zeta)}(x,y). \tag{4.4}$$

Since every $\delta_{(\xi,\zeta)}\colon\mathbf{\Omega}_{(\bar{x},\bar{y})}\to\mathbb{R}$ is a probability distribution, our assumption implies that $\lim_{t\to\infty}\exp(t\mathcal{A})\delta_{(\xi,\zeta)} = \rho$, and thus

$$\lim_{t\to\infty} \exp(t\mathcal{A})q = \sum_\xi \sum_\zeta q(\xi,\zeta) \cdot \lim_{t\to\infty} \exp(t\mathcal{A})\delta_{(\xi,\zeta)}$$
$$= \sum_\xi \sum_\zeta q(\xi,\zeta) \cdot \rho = \mathcal{S}q.$$

Since the state space $\mathbf{\Omega}_{(\bar{x},\bar{y})}$ is finite, we obtain $\mathcal{S} = \lim_{t\to\infty} \exp(t\mathcal{A})$ in the operator norm induced by $\|\cdot\|$.

The operator $\mathcal{S}$ is a projection because $\sum_{(\tilde{\xi},\tilde{\zeta})\in\mathbf{\Omega}_{(\bar{x},\bar{y})}} \rho(\tilde{\xi},\tilde{\zeta}) = 1$, and hence

$$\mathcal{S}\mathcal{S}q = \mathcal{S}\rho\left(\sum_\xi \sum_\zeta q(\xi,\zeta)\right) = \rho\left(\sum_{(\tilde{\xi},\tilde{\zeta})\in\mathbf{\Omega}_{(\bar{x},\bar{y})}} \rho(\tilde{\xi},\tilde{\zeta})\right)\left(\sum_\xi \sum_\zeta q(\xi,\zeta)\right)$$
$$= \rho\left(\sum_\xi \sum_\zeta q(\xi,\zeta)\right) = \mathcal{S}q.$$

This projection maps every probability distribution to the stationary distribution ρ, whereas all functions $q\colon\mathbf{\Omega}_{(\bar{x},\bar{y})} \to \mathbb{R}$ with the property that $\sum_\xi \sum_\zeta q(\xi,\zeta) = 0$ are mapped to zero. These properties can now be used to prove the following error bound.

THEOREM 2. *Consider the situation from Theorem* 1*, and, in addition, let* $\gamma_1, \gamma_2 > 0$ *be constants such that*

$$\|\exp(t\mathcal{A}) - \mathcal{S}\| \le \gamma_1 e^{-\gamma_2 t} \quad \textit{for all } t \ge 0. \tag{4.5}$$

Then the error $\varepsilon(t) = \varepsilon(t,x,y) = p(t,x,y) - u(t,x)v(t,y)$ *is bounded by*

$$\|\varepsilon(t)\| \le \|\varepsilon_0\| + \frac{\gamma_1}{\gamma_2}(1 - e^{-\gamma_2 t}) \cdot \max_{\tau\in[0,t]} \|(I - \mathcal{P}_{(u\otimes v)(\tau)})\mathcal{A}(u\otimes v)(\tau)\|.$$

Remark. This error bound accounts for the fact that the error stops growing when the solution of the CME converges to a stationary distribution sufficiently fast.

Proof. It follows from the definition of $\mathcal{S}$ that

$$\mathcal{S}(I - \mathcal{P}_{(u\otimes v)(s)})\mathcal{A}(u\otimes v) = 0$$

because $\sum_\xi \sum_\zeta f(\xi,\zeta) = 0$ for $f = \mathcal{A}(u\otimes v)$ or $f = \mathcal{P}\mathcal{A}(u\otimes v)$. Hence, starting from (4.3) we obtain

$$\|\varepsilon(t)\| \le \|\exp(t\mathcal{A})\varepsilon_0\| + \int_0^t \|\exp((t-s)\mathcal{A})(I - \mathcal{P}_{(u\otimes v)(s)})\mathcal{A}(u\otimes v)(s)\|\mathrm{d}s$$
$$\le \|\varepsilon_0\| + \int_0^t \|(\exp((t-s)\mathcal{A}) - \mathcal{S})(I - \mathcal{P}_{(u\otimes v)(s)})\mathcal{A}(u\otimes v)(s)\|\mathrm{d}s$$
$$= \|\varepsilon_0\| + \int_0^t \gamma_1 e^{-\gamma_2 s}\mathrm{d}s \cdot \max_{\tau\in[0,t]} \|(I - \mathcal{P}_{(u\otimes v)(\tau)})\mathcal{A}(u\otimes v)(\tau)\|$$
$$= \|\varepsilon_0\| + \frac{\gamma_1}{\gamma_2}(1 - e^{-\gamma_2 t}) \cdot \max_{\tau\in[0,t]} \|(I - \mathcal{P}_{(u\otimes v)(\tau)})\mathcal{A}(u\otimes v)(\tau)\|. \qquad \square$$

4.2. Accuracy of the Hellander–Lötstedt model. When discussing the accuracy of the Hellander–Lötstedt model, one has to keep in mind that $w(t,x)$ and $\eta(t)$ do

not approximate the joint probability distribution $p(t,x,y)$ itself, but only provide an approximation to the marginal distribution of the x-variables and the expectation with respect to the y-variables, i.e.,

$$w(t,x) \approx \sum_y p(t,x,y), \qquad \eta(t) \approx \sum_x \sum_y yp(t,x,y).$$

An approximation of $p(t,x,y)$ can be obtained by assuming that

$$p(t,x,y) \approx w(t,x)\boldsymbol{\phi}(y,\eta(t)), \tag{4.6}$$

where $\boldsymbol{\phi}(y,\eta(t))$ is a suitably chosen distribution with mean $\eta(t)$. In [16] a Gaussian was chosen (in spite of the discreteness of the state space), but one could also use, e.g., a Poisson distribution (in case of an unbounded state space) or a multinomial distribution (in case of a bounded state space). Since the question of which choice of $\boldsymbol{\phi}$ is the best in a particular application can often not be answered a priori, the ansatz (4.6) will not be used here. Instead, we assume that the species represented by the y-variables are of minor interest such that an approximation of the mean $\sum_x \sum_y yp(t,x,y)$ provides a sufficient description of their dynamics.

4.2.1. The monomolecular case. As a first result, we prove that the Hellander-Lötstedt model is exact in the monomolecular setting.

PROPOSITION 2. *Consider the CME* (2.7) *and the Hellander–Lötstedt model* (3.10)–(3.11) *without any truncation of the state space (i.e., let* $(\bar{x},\bar{y}) = (\infty,\ldots,\infty)$ *in* (2.8)*). Suppose that all reaction channels are monomolecular and that* $p_0(x,y)$ *is a product Poisson distribution. If* $w(0,x) = \sum_y p_0(x,y)$ *and* $\eta(0) = \sum_x \sum_y yp_0(x,y)$*, then the product approximation is exact in the sense that*

$$w(t,x) = \sum_y p(t,x,y) \quad \textit{and} \quad \eta(t) = \sum_x \sum_y yp(t,x,y) \quad \textit{for all } t \geq 0, x \geq 0.$$

Proof. It follows from Corollary 1 that the exact marginal distributions $p_1(t,x) = \sum_y p(t,x,y)$ and $p_2(t,y) = \sum_x p(t,x,y)$ are the solutions of (3.5)–(3.6), i.e.,

$$\partial_t p_1(t,x) = \sum_{k=1}^{r} \left(\sum_y \beta_k(y) p_2(t,y) \right) (\alpha_k(x-\nu_k)p_1(t,x-\nu_k) - \alpha_k(x)p_1(t,x)),$$

$$\partial_t p_2(t,y) = \sum_{k=1}^{r} \left(\sum_x \alpha_k(x) p_1(t,x) \right) (\beta_k(y-\mu_k)p_2(t,y-\mu_k) - \beta_k(y)p_2(t,y)).$$

Let $\mathbb{E}_y(t)$ be the exact expectation with respect to y. Since all reaction channels are monomolecular, we have either $\beta_k(y) = c$ or $\beta_k(y) = cy_j$ for some $j \in \{1,\ldots,d_2\}$. This implies

$$\sum_y \beta_k(y) p_2(t,y) = \beta_k(\mathbb{E}_y(t)), \tag{4.7}$$

and in particular

$$\partial_t p_1(t,x) = \sum_{k=1}^{r} \beta_k(\mathbb{E}_y(t))(\alpha_k(x-\nu_k)p_1(t,x-\nu_k) - \alpha_k(x)p_1(t,x)). \tag{4.8}$$

Substituting the differential equation for $p_2(t,y)$ into $\dot{\mathbb{E}}_y(t) = \sum_y y\partial_t p_2(t,y)$ yields with (4.7) and similar arguments as in section 2.5 that

$$\begin{aligned}
\dot{\mathbb{E}}_y(t) &= \sum_{k=1}^{r} \left(\sum_x \alpha_k(x)p_1(t,x)\right) \sum_y (\beta_k(y-\mu_k)p_2(t,y-\mu_k) - \beta_k(y)p_2(t,y)) \\
&= \sum_{k=1}^{r} \left(\sum_x \alpha_k(x)p_1(t,x)\right) \mu_k \sum_y \beta_k(y)p_2(t,y) \\
&= \sum_{k=1}^{r} \left(\sum_x \alpha_k(x)p_1(t,x)\right) \mu_k \beta_k(\mathbb{E}_y(t)). \qquad (4.9)
\end{aligned}$$

Equation (4.8) and (4.9) are exactly the equations of the Hellander–Lötstedt model, and since by assumption $w(0,x) = p_1(0,x)$ and $\eta(0) = \mathbb{E}_y(0)$, this proves the assertion. ∎

Remark. An alternative proof could be given by using the explicit solution formula from Proposition 2 in [20].

4.2.2. The general case: Assumptions and consequences. According to the derivation, one can expect that the Hellander–Lötstedt model yields a very good approximation to the solution of the full CME (2.7) if

- an approximate product representation $p(t,x,y) \approx u(t,x)v(t,y)$ exists, and
- $v(t,x)$ is essentially represented by its expectation.

Both conditions hold if the marginal distribution $p_2(t,y) = \sum_x p(t,x,y)$ is localized around the mean value $\mathbb{E}_y(t) = \sum_y y p_2(t,y)$, i.e., if the covariance matrix $\sum_x \sum_y (y - \mathbb{E}_y(t))(y-\mathbb{E}_y(t))^T p(t,x,y)$ is sufficiently small and higher moments vanish. The conjecture that a small covariance with respect to y improves the accuracy of the Hellander–Lötstedt model has already been made in [16]. Proposition 2 shows, however, that this condition is not always necessary since here the exact result is obtained, although the covariance matrix can be arbitrarily large. In this subsection we prove an error bound which explains this observation. Only the error

$$\|u(t,\cdot) - w(t,\cdot)\| + \left\| \sum_y y v(t,y) - \eta(t) \right\|$$

will be considered. An error bound for the Hellander–Lötstedt approximation to the full CME solution can be obtained via the triangle inequality and Theorems 1 or 2.

From now on let

$$|\omega| = \sqrt{\sum_{i=1}^{d_2} \omega_i^2}$$

denote the 2-norm of a vector or the associated matrix norm, respectively. Note that this does not contradict the earlier use of $|\cdot|$ for the absolute value of a real number.

Let $u(t,x)$ and $v(t,y)$ be the solutions of the product approximation (3.5)–(3.6), and let $w(t,x)$ and $\eta(t)$ be the solutions of the Hellander–Lötstedt model (3.10)–(3.11). The following assumptions are made:

(A1) For all $k = 1, \dots, r$ we assume that either

$$\beta_k(y) = c_k \quad \text{or } \beta_k(y) = c_k y_i \quad \text{or } \beta_k(y) = c_k y_i y_j \quad \text{or } \beta_k(y) = c_k \frac{y_i(y_i - 1)}{2}$$

for some $i, j \in \{1, \dots, d_2\}$, where $c_k > 0$ is the reaction constant of the kth reaction. This implies that the Hessian $\nabla^2 \beta_k(y) = \nabla^2 \beta_k$ is constant and $\nabla^3 \beta_k(y) = 0$. Moreover, it implies that β_k is Lipschitz continuous: There is a constant $L_1 > 0$ such that

$$|\beta_k(y) - \beta_k(\tilde{y})| \le L_1 |y - \tilde{y}|$$

for all $0 \le y < \bar{y}$ and $0 \le \tilde{y} < \bar{y}$ and, in particular, that $\beta_k(y) \le L_1 |\bar{y}|$ is bounded.

(A2) There is a constant $L_2 > 0$ such that

$$\left\| \sum_{k=1}^{r} \beta_k(\eta(t)) \cdot (\tilde{\mathcal{A}}_k u(t, \cdot) - \tilde{\mathcal{A}}_k w(t, \cdot)) \right\| \le L_2 \| u(t, \cdot) - w(t, \cdot) \|,$$

$$\| \tilde{\mathcal{A}}_k u(t, \cdot) \| \le L_2,$$

where

$$\tilde{\mathcal{A}}_k u(t, x) = \alpha_k(x - \nu_k) u(t, x - \nu_k) - \alpha_k(x) u(t, x).$$

(A3) There is a constant $L_3 > 0$ such that for all $k = 1, \dots, r$

$$\left| \sum_x \alpha_k(x)(u(t, x) - w(t, x)) \right| \le L_3 \| u(t, \cdot) - w(t, \cdot) \|,$$

$$\left| \sum_x \alpha_k(x) w(t, x) \right| \le L_3,$$

$$\left| \sum_x \alpha_k(x) u(t, x) \right| \le L_3.$$

Discussion of the assumptions. (A1) is satisfied if all reaction channels involve at most two particles of the "deterministic" species $S_{d_1+1}, \dots, S_{d_1+d_2}$ and standard propensities (2.2) are used. This is reasonable because reactions with more than two reactants are rare in biological systems. With some additional efforts and assumptions, our error analysis could be extended to the case of nonstandard propensities. (A2) and (A3) can be seen as regularity assumptions. On a bounded state space these conditions are *always* satisfied because all propensity functions are bounded. Hence, formulas for L_2 and L_3 can easily be derived, but the corresponding expressions depend on $\max_{k=1,\dots,r} \max_{x < \bar{x}} \alpha_k(x)$ and are usually far too pessimistic. Assumptions (A2) and (A3) are to be understood in the sense that constants L_2 and L_3 exist, which are much smaller than the worst-case estimate.

4.2.3. The general case: Error bound.

THEOREM 3. *Let the expectation and the covariance matrix of $v(t, \cdot)$ be denoted by*

$$\mathbb{E}_v(t) = \sum_y y v(t, y), \quad \Sigma_v(t) = \sum_y (y - \mathbb{E}_v(t))(y - \mathbb{E}_v(t))^T v(t, y).$$

Under the assumptions (A1), (A2), *and* (A3), *there is a constant* $\hat{C}$ *depending on* $r, L_1, L_2, L_3, |\bar{y}|$ *and* t_{end} *such that for all* $t \in [0, t_{\text{end}}]$

$$|\mathbb{E}_v(t) - \eta(t)| + \|u(t,\cdot) - w(t,\cdot)\| \leq \hat{C} \sum_{k=1}^{r} \int_0^t |R_k(\Sigma_v(s))| ds$$

with

$$R_k(\Sigma_v(t)) = \frac{1}{2} \sum_y (y - \mathbb{E}_v)^T (\nabla^2 \beta_k)(y - \mathbb{E}_v) v(t,y) = \frac{1}{2} \operatorname{tr}(\Sigma_v(t) \nabla^2 \beta_k), \tag{4.10}$$

where $\operatorname{tr}(M) = m_{11} + \cdots + m_{nn}$ *denotes the trace of a matrix* $M = (m_{ij}) \in \mathbb{R}^{n \times n}$.

Proof of Theorem 3.

Step 1*: Error in the marginal distribution.* Subtracting (3.10) from (3.5) yields

$$\partial_t u(t,x) - \partial_t w(t,x) = \sum_{k=1}^{r} \left(\sum_y \beta_k(y) v(t,y) \right) \tilde{\mathcal{A}}_k u(t,x) - \sum_{k=1}^{r} \beta_k(\eta) \tilde{\mathcal{A}}_k w(t,x). \tag{4.11}$$

Following the approach in [7], we use a Taylor expansion of β_k about $\mathbb{E}_v$ and obtain

$$\beta_k(y) = \beta_k(\mathbb{E}_v) + \nabla \beta_k(\mathbb{E}_v)(y - \mathbb{E}_v) + \frac{1}{2}(y - \mathbb{E}_v)^T (\nabla^2 \beta_k)(y - \mathbb{E}_v)$$

since all higher-order terms vanish according to assumption (A1). This yields

$$\begin{aligned} \sum_y \beta_k(y) v(t,y) &= \beta_k(\mathbb{E}_v) + \frac{1}{2} \sum_y (y - \mathbb{E}_v)^T (\nabla^2 \beta_k)(y - \mathbb{E}_v) v(t,y) \\ &= \beta_k(\mathbb{E}_v) + R_k(\Sigma_v(t)) \end{aligned} \tag{4.12}$$

with the remainder term from (4.10). Substituting (4.12) into (4.11) shows that

$$\partial_t u(t,x) - \partial_t w(t,x) = \sum_{k=1}^{r} \beta_k(\mathbb{E}_v)\tilde{\mathcal{A}}_k u(t,x) - \sum_{k=1}^{r} \beta_k(\eta)\tilde{\mathcal{A}}_k w(t,x) + \sum_{k=1}^{r} R_k(\Sigma_v(t))\tilde{\mathcal{A}}_k u(t,x).$$

Since $u(0,x) = w(0,x)$ by assumption, we obtain with (A1) and (A2)

$$\begin{aligned}
\|u(t,\cdot) - w(t,\cdot)\| &= \left\| \int_0^t (\partial_s u(s,\cdot) - \partial_s w(s,\cdot))\, ds \right\| \\
&\le \sum_{k=1}^{r} \int_0^t |\beta_k(\mathbb{E}_v(s)) - \beta_k(\eta(s))| \cdot \|\tilde{\mathcal{A}}_k u(s,\cdot)\|\, ds \\
&\quad + \left\| \sum_{k=1}^{r} \int_0^t \beta_k(\eta(s)) \cdot (\tilde{\mathcal{A}}_k u(s,\cdot) - \tilde{\mathcal{A}}_k w(s,\cdot))\, ds \right\| \\
&\quad + \sum_{k=1}^{r} \int_0^t \|R_k(\Sigma_v(s))| \cdot \|\tilde{\mathcal{A}}_k u(s,\cdot)\|\, ds \\
&\le r L_1 L_2 \int_0^t |\mathbb{E}_v(s) - \eta(s)| + \|u(s,\cdot) - w(s,\cdot)\|\, ds \\
&\quad + L_2 \sum_{k=1}^{r} \int_0^t |R_k(\Sigma_v(s))|\, ds. \qquad (4.13)
\end{aligned}$$

Step 2*: Error in the expectations.* The exact expectation evolves according to the equation

$$\begin{aligned}
\dot{\mathbb{E}}_v(t) &= \sum_y y \partial_t v(t,y) \\
&= \sum_{k=1}^{r} \left(\sum_x \alpha_k(x) u(t,x) \right) \sum_y y(\beta_k(y-\mu_k) v(t,y-\mu_k) - \beta_k(y) v(t,y)) \\
&= \sum_{k=1}^{r} \mu_k \left(\sum_x \alpha_k(x) u(t,x) \right) \left(\sum_y \beta_k(y) v(t,y) \right).
\end{aligned}$$

Comparing with (3.11) and substituting the Taylor expansion (4.12) shows that

$$\begin{aligned}
\dot{\mathbb{E}}_v(t) - \dot{\eta}(t) &= \sum_{k=1}^{r} \mu_k \Big(\sum_x \alpha_k(x) u(t,x) \Big) \Big(\sum_y \beta_k(y) v(t,y) \Big) \\
&\quad - \sum_{k=1}^{r} \mu_k \Big(\sum_x \alpha_k(x) w(t,x) \Big) \beta_k(\eta(t)) \\
&= \sum_{k=1}^{r} \mu_k \Big(\sum_x \alpha_k(x) u(t,x) \Big) (\beta_k(\mathbb{E}_v(t)) + R_k(\Sigma_v(t))) \\
&\quad - \sum_{k=1}^{r} \mu_k \Big(\sum_x \alpha_k(x) w(t,x) \Big) \beta_k(\eta(t)) \\
&= \sum_{k=1}^{r} \mu_k \Big(\sum_x \alpha_k(x) (u(t,x) - w(t,x)) \Big) \beta_k(\mathbb{E}_v(t)) \\
&\quad + \sum_{k=1}^{r} \mu_k \Big(\sum_x \alpha_k(x) w(t,x) \Big) (\beta_k(\mathbb{E}_v(t)) - \beta_k(\eta(t))) \\
&\quad + \sum_{k=1}^{r} \mu_k \Big(\sum_x \alpha_k(x) u(t,x) \Big) R_k(\Sigma_v(t)).
\end{aligned}$$

Integrating the time-derivatives and using the assumptions (A1) and (A3) yields the error bound

$$\begin{aligned}
|\mathbb{E}_v(t) - \eta(t)| &\le \sum_{k=1}^{r} \int_0^t |\mu_k| \cdot \Big| \sum_x \alpha_k(x) (u(s,x) - w(s,x)) \Big| \cdot |\beta_k(\mathbb{E}_v(s))| ds \\
&\quad + \sum_{k=1}^{r} \int_0^t |\mu_k| \cdot \Big| \sum_x \alpha_k(x) w(s,x) \Big| \cdot |\beta_k(\mathbb{E}_v(s)) - \beta_k(\eta(s))| ds \\
&\quad + \sum_{k=1}^{r} \int_0^t |\mu_k| \cdot \Big| \sum_x \alpha_k(x) u(s,x) \Big| \cdot |R_k(\Sigma_v(s))| ds \\
&\le C r L_1 L_3 |\bar{y}| \int_0^t \|u(s,\cdot) - w(s,\cdot)\| + |\mathbb{E}_v(s) - \eta(s)| ds \\
&\quad + L_3 \sum_{k=1}^{r} \int_0^t |R_k(\Sigma_v(s))| ds. \qquad (4.14)
\end{aligned}$$

Step 3*: Gronwall lemma.* Combining (4.13) and (4.14) provides the estimate

$$\begin{aligned}
&|\mathbb{E}_v(t) - \eta(t)| + \|u(t,x) - w(t,x)\| \\
&\qquad \le C r L_1 L_2 L_3 |\bar{y}| \int_0^t |\mathbb{E}_v(s) - \eta(s)| + \|u(s,x) - w(s,x)\| ds \\
&\qquad + (L_2 + L_3) \sum_{k=1}^{r} \int_0^t |R_k(\Sigma_v(s))| ds.
\end{aligned}$$

Applying the continuous Gronwall lemma finally gives the error bound

$$|\mathbb{E}_v(t) - \eta(t)| + \|u(t,x) - w(t,x)\| \le \hat{C} \sum_{k=1}^{r} \int_0^t |R_k(\Sigma_v(s))| ds$$

for all $t \in [0, t_{\text{end}}]$ with $\hat{C} = (L_2 + L_3) e^{C r L_1 L_2 L_3 |\bar{y}| t_{\text{end}}}$. ∎

As we have expected, Theorem 3 states that replacing the product approximation (3.5)–(3.6) by the Hellander–Lötstedt model (3.10)–(3.11) causes an additional error which depends on the covariance matrix of the y-variables. However, Theorem 3 also explains why under the conditions of Proposition 2 the *exact* result is reproduced, although here the covariance matrix of the y-variables is not necessarily small.

COROLLARY 2. *Suppose that for all $k = 1, \dots, r$ the propensity $\beta_k(y)$ is at most linear, i.e., that either $\beta_k(y) = c_k$ or $\beta_k(y) = c_k y_i$, where $c_k > 0$ is the reaction constant of the kth reaction. Then, under the assumptions (A2) and (A3), the Hellander–Lötstedt model coincides with the product approximation in the sense that*

$$\mathbb{E}_v(t) = \eta(t), \qquad u(t,x) = w(t,x).$$

Proof of Corollary 2. Since $\beta_k(y)$ is at most linear, it follows that $\nabla^2 \beta_k = 0$, and hence $R_k(\Sigma_v(t)) = 0$ for all $k = 1, \dots, r$ and $t \in [0, t_{\text{end}}]$. According to Theorem 3, this means that $|\mathbb{E}_v(t) - \eta(t)| + \|u(t,x) - w(t,x)\| = 0$. ∎

The above condition requires only $\beta_k(y)$ to be at most linear, but *not* the entire propensity. Suppose, for example, that the kth reaction channel is

$$S_i + S_j \to [\dots],$$

where $[\dots]$ denotes an arbitrary reaction product and $i \neq j$. If $i \in \{1, \dots, d_1\}$ and $j \in \{d_1 + 1, \dots, d_1 + d_2\}$, then $\alpha_k(x) = c_k \cdot x_i$ and $\beta_k(y) = y_j$. If $i, j \in \{1, \dots, d_1\}$, then $\alpha_k(x) = c_k \cdot x_i x_j$ and $\beta_k(y) = 1$. In both cases, $\beta_k(y)$ is at most linear, although the total propensity $\alpha_k(x)\beta_k(y)$ is *not*.

We can now easily reprove Proposition 2 as follows: Under the corresponding conditions

$$w(t,x) \overset{(\star)}{=} u(t,x) = \sum_y u(t,x)v(t,y) \overset{(\diamond)}{=} \sum_y p(t,x,y),$$
$$\eta(t) \overset{(\star)}{=} \mathbb{E}_v(t) = \sum_x \sum_y y u(t,x) v(t,y) \overset{(\diamond)}{=} \sum_x \sum_y y p(t,x,y),$$

where the equations marked by $(\star)$ follow from Corollary 2 and the equations marked by $(\diamond)$ follow from Corollary 1. Combining Proposition 1 with Theorem 3 implies even a slightly more general result: The Hellander–Lötstedt model reproduces the exact result if the solution of the CME admits a tensor product representation and propensities $\beta_k(y)$ are at most linear. However, these conditions are very restrictive.

5. Model reduction based on conditional expectations. The derivation of the Hellander–Lötstedt model proceeds in two steps:

- product approximation of the CME (cf. section 3.1), and
- reducing the marginal distribution $v(t,y)$ to its expectation (cf. section 3.2).

The analysis in the previous section indicates that often the first step is the critical one because Corollary 2 shows that the Hellander–Lötstedt model attains the same accuracy as the product approximation if $\beta_k(y)$ is at most linear. This condition is indeed satisfied in many applications. The product approximation, in contrast, often turns out to be far too crude because the solution of the full CME simply does not admit a reasonable approximation of the form $p(t,x,y) \approx u(t,x)v(t,y)$. This observation motivates a model reduction based on conditional expectations which is presented now.

5.1. Derivation. The derivation is based on the exact decomposition

$$p(t,x,y) = p_1(t,x)p_2(t,y|x) \quad \text{for all } (x,y) \in \Omega_{(\bar{x},\bar{y})}, \tag{5.1}$$

where $p_1(t,x) = \sum_y p(t,x,y)$ is the marginal distribution and $p_2(t,y|x)$ is the conditional distribution[2] that at time t there are y_j particles of S_j given that there are x_i particles of S_i $(i \in \{1,\dots,d_1\}, j \in \{d_1+1,\dots,d_1+d_2\})$. If $p_1(t,x) \neq 0$, then we simply have $p_2(t,y|x) = p(t,x,y)/p_1(t,x)$. Summing both sides of the full CME (2.7) with respect to y yields

$$\begin{aligned}\partial_t p_1(t,x) = &\sum_{k=1}^{r} \left(\sum_y \beta_k(y) p_2(t,y|x-\nu_k)\right)\alpha_k(x-\nu_k)p_1(t,x-\nu_k)\\ &- \sum_{k=1}^{r}\left(\sum_y \beta_k(y)p_2(t,y|x)\right)\alpha_k(x)p_1(t,x).\end{aligned}$$

This equation is still exact but depends on $p_2(t,y|x)$, which is not supposed to be computed. Similar as before, we replace $p_2(t,y|x)$ by substituting the approximation

$$\sum_y \beta_k(y)p_2(t,y|x) \approx \beta_k(\eta(t|x)), \tag{5.2}$$

where $\eta(t|x) \approx \sum_y y p_2(t,y|x)$ approximates the mean of the conditional expectation. This gives an equation similar to (3.10):

$$\begin{aligned}\partial_t p_1(t,x) \approx &\sum_{k=1}^{r}(\beta_k(\eta(t|x-\nu_k))\alpha_k(x-\nu_k)w(t,x-\nu_k)\\ &-\beta_k(\eta(t|x))\alpha_k(x)w(t,x)).\end{aligned} \tag{5.3}$$

Next we derive an evolution equation for the new variable $\eta(t|x) \approx \sum_y y p_2(t,y|x)$. Applying the product rule to $\sum_y y p(t,x,y) \approx \eta(t|x)p_1(t,x)$ shows that

$$p_1(t,x)\partial_t \eta(t|x) \approx \sum_y y\partial_t p(t,x,y) - \eta(t|x)\partial_t p_1(t,x). \tag{5.4}$$

Now $\sum_y y\partial_t p(t,x,y)$ has to be approximated in terms of $\eta(t|x)$ and $p_1(t,x)$. Substituting the CME (2.7) yields

$$\begin{aligned}&\sum_y y\partial_t p(t,x,y)\\ &\quad= \sum_{k=1}^{r}\sum_y y(\alpha_k(x-\nu_k)\beta_k(y-\mu_k)p(t,x-\nu_k,y-\mu_k) - \alpha_k(x)\beta_k(y)p(t,x,y))\\ &\quad= \sum_{k=1}^{r}\left(\sum_y (y+\mu_k)\beta_k(y)p_2(t,y|x-\nu_k)\right)\alpha_k(x-\nu_k)p_1(t,x-\nu_k)\\ &\quad\quad - \sum_{k=1}^{r}\left(\sum_y y\beta_k(y)p_2(t,y|x)\right)\alpha_k(x)p_1(t,x).\end{aligned} \tag{5.5}$$

[2]Note that $p_2(t,y|x)$ differs from $p_2(t,y)$ defined in section 4.

In order to replace $p_2(t, y|x)$ we substitute the Taylor expansion

$$\beta_k(y) \approx \beta_k(\eta(t|x)) + \langle \nabla \beta_k(\eta(t|x)), y - \eta(t|x) \rangle$$

which gives

$$\begin{aligned}
&\sum_y y\beta_k(y)p_2(t, y|x) \\
&\quad \approx \sum_y y\beta_k(\eta(t|x))p_2(t, y|x) + \sum_y y\langle \nabla \beta_k(\eta(t|x)), y - \eta(t|x)\rangle p_2(t, y|x) \\
&\quad \approx \beta_k(\eta(t|x))\eta(t|x) + \sum_y (y - \eta(t|x))\langle \nabla \beta_k(\eta(t|x)), y - \eta(t|x)\rangle p_2(t, y|x) \qquad (5.6) \\
&\quad \approx \beta_k(\eta(t|x))\eta(t|x). \qquad (5.7)
\end{aligned}$$

In (5.6) we have used that $\eta(t|x) \approx \sum_y y p_2(t, y|x)$ implies

$$\sum_y \langle \nabla \beta_k(\eta(t|x)), y - \eta(t|x)\rangle p_2(t, y|x) \approx 0,$$

and in (5.7) it was assumed that the term

$$\begin{aligned}
&\sum_y (y - \eta(t|x))\langle \nabla \beta_k(\eta(t|x)), y - \eta(t|x)\rangle p_2(t, y|x) \\
&\quad = \sum_y (y - \eta(t|x))(y - \eta(t|x))^T p_2(t, y|x)
\end{aligned}$$

is so small that it can be neglected. This term approximates the covariance matrix of $p_2(t, y|x)$ if $\eta(t|x)$ approximates the conditional expectation. With (5.4), (5.3), and (5.7), we obtain the following set of equations for $w(t, x) \approx p_1(t, x)$ and $\eta(t|x) \approx \sum_y y p_2(t, y|x)$:

$$\begin{aligned}
\partial_t w(t, x) = &\sum_{k=1}^{r} (\beta_k(\eta(t|x - \nu_k))\alpha_k(x - \nu_k)w(t, x - \nu_k) \\
&- \beta_k(\eta(t|x))\alpha_k(x)w(t, x)), \qquad (5.8)
\end{aligned}$$

$$\begin{aligned}
\partial_t \eta(t|x) \cdot w(t, x) = &-\eta(t|x) \cdot \partial_t w(t, x) \\
&+ \sum_{k=1}^{r} (\eta(t|x - \nu_k)\beta_k(\eta(t|x - \nu_k))\alpha_k(x - \nu_k)w(t, x - \nu_k) \\
&- \eta(t|x)\beta_k(\eta(t|x))\alpha_k(x)w(t, x)) \\
&+ \sum_{k=1}^{r} \mu_k \beta_k(\eta(t|x - \nu_k))\alpha_k(x - \nu_k)w(t, x - \nu_k). \qquad (5.9)
\end{aligned}$$

This reduced model will be denoted by *MRCE* as an abbreviation of model reduction by conditional expectations. We remark, however, that model reduction based on conditional moments is not a new idea. In fact, the MRCE equations (5.8)–(5.9) are closely related to a model reduction approach proposed in [29], [8] for the Fokker–Planck equation to certain techniques in the context of polymerization kinetics

(cf. [19]) and to the approach from [34]. After submission of the manuscript, the author came across the references [17] and [31], where similar concepts are applied to the CME, but with a different derivation and slightly different evolution equations.

As in the Hellander–Lötstedt model, the differential equation for the x-variables is a lower-dimensional CME with propensities depending on η. In contrast to (3.11), however, (5.9) does not really look like a variant of the reaction-rate equation, and since now $\eta = \eta(t|x)$ depends on x, this equation is more complicated. Moreover, the differential equation (5.9) has to be handled with care. Formally, dividing by $w(t,x)$ and replacing $\partial_t w(t,x)$ via (5.8) would yield a differential equation for $\eta(t|x)$, but this can only be done for those x where $w(t,x) \neq 0$. If $w(t,x) = 0$, then $\partial_t \eta(t|x)$ is not defined, and for $w(t,x) \approx 0$, the equation is ill-conditioned. Hence, solving the system (5.8)–(5.9) numerically requires a method which can cope with this problem. Such a method will be derived in section 5.2. For the time being, we assume for simplicity that $w(t,x) \neq 0$ for all $x \in \mathbf{\Omega}_{\bar{x}}$. Under this condition, the MRCE has the following properties:

1. If $\eta(0|x) \geq 0$ for all $x \in \mathbf{\Omega}_{\bar{x}}$ and if $w(0,\cdot)$ is a probability distribution, then $\eta(t|x) \geq 0$ for all $x \in \mathbf{\Omega}_{\bar{x}}$ and $w(t,\cdot)$ is a probability distribution for all $t \geq 0$. This can be shown by adapting the proof of Lemma 2.
2. The Hellander–Lötstedt model is recovered as a special case of MRCE: If $\eta(t|x) = \tilde{\eta}(t)$ is *constant* with respect to x, then (5.8)–(5.9) is equivalent to (3.10)–(3.11). In this case, we obtain from (5.9) and using $\sum_x \partial_t w(t,x) = 0$ that

$$\begin{aligned}\frac{d}{dt}\tilde{\eta}(t) &= \frac{d}{dt}\tilde{\eta}(t)\sum_x w(t,x)\\ &= \sum_x \partial_t \eta(t|x)\cdot w(t,x)\\ &= 0 + 0 + \sum_{k=1}^{r} \mu_k \sum_x \beta_k(\eta(t|x-\nu_k))\alpha_k(x-\nu_k)w(t,x-\nu_k)\\ &= \sum_{k=1}^{r} \mu_k \beta_k(\tilde{\eta}(t))\sum_x \alpha_k(x)w(t,x).\end{aligned}$$

   Hence, $\tilde{\eta}$ and $w(t,x)$ is the solution of the Hellander–Lötstedt model (3.10)–(3.11).
3. Under the conditions of Proposition 2 (monomolecular reaction channels and product Poisson initial data) MRCE is exact, i.e.,

$$w(t,x) = \sum_y p(t,x,y) \quad \text{and} \quad \eta(t|x)w(t,x) = \sum_x \sum_y y p(t,x,y).$$

   The reason is that under these assumptions $\eta(t|x)$ is indeed independent of x. Hence, the solution of MRCE coincides with the solution of the Hellander–Lötstedt model, which, according to Proposition 2, reproduces the exact result.
4. In Theorem 3 the error of the Hellander–Lötstedt model was bounded in terms of the covariance matrix of the marginal distribution of the y-variables. We conjecture that a similar result could also be shown for MRCE, but a detailed error analysis is beyond the scope of this article.

5.2. An adaptive integrator for MRCE. It was pointed out in the previous section that (5.9) provides only an evolution equation for those $\partial_t \eta(t|x)$ with $w(t,x) > 0$ and that the case $w(t,x) \approx 0$ may cause numerical instabilities. Now we sketch a strategy how to cope with this problem.

Preliminaries. The following method computes approximations $w_n(x) \approx w(t_n, x)$ and $\eta_n(x) \approx \eta(t_n|x)$ of the solution of (5.8)–(5.9) at discrete times $t_n = t_{n-1} + h_n \in [0, t_{\text{end}}]$ with $t_0 = 0$. The step-size h_n can be chosen adaptively or kept constant.

Initial data. Let $p(0, x, y)$ be the (known) initial distribution of the full CME (2.7), and let $w_0(x) = \sum_y p(0, x, y)$. For those states with $w_0(x) \geq \varepsilon$ the initial value for the conditional expectations is given by

$$\eta_0(x) = \frac{\sum_y y p(0, x, y)}{w_0(x)} \quad \text{for all } x \in \mathbf{\Omega}_{\bar{x}} \quad \text{with } w_0(x) \geq \varepsilon.$$

Extrapolation of the expectations. Let w_n be an approximation of the marginal distribution $w(t_n, \cdot)$. A state $x \in \mathbf{\Omega}_{\bar{x}}$ is called *active* if $w_n(x) \geq \varepsilon$ and *passive* otherwise. In our method $\eta_n(x)$ will be only computed for the active states. For all passive states the values of $\eta_n(x)$ are not of any practical interest because it does not really matter what happens in a state x that is only visited with a very low probability. Nevertheless, $\eta_n(x)$ must be defined for *all* $x \in \mathbf{\Omega}_{\bar{x}}$ in order to evaluate the right-hand side of (5.8) and (5.9). Thus, we define an estimate $E[\eta_n](x) \approx \eta_n(x)$ for the passive states by extrapolation. This approach is based on the assumption that $x \mapsto \eta_n(x)$ is smooth (in a discrete sense). The extrapolation procedure operates as follows:

Extrapolation of the expectations.

0. Input: $\eta_n(x)$ for all active states x.
1. Define $j = 0$, $\chi^{(0)}(x) = 1$ if x is active and $\chi^{(0)}(x) = 0$ otherwise. The function $\chi^{(j)}(x)$ indicates whether or not $\eta_n(x)$ has been defined for state x after j iterations.
2. Define $\hat{\eta}_n^{(0)}(x) = \eta_n(x)$ for all active states x.
3. For any state $x \in \mathbf{\Omega}_{\bar{x}}$ let

$$\mathcal{B}(x) = \{\tilde{x} \in \mathbf{\Omega}_{\bar{x}} \colon \max_{i=1,\ldots,d_1} |x_i - \tilde{x}_i| = 1\}$$

be the set of all neighboring states, and let

$$\Gamma^{(j)} = \{x \in \mathbf{\Omega}_{\bar{x}} \colon x \text{ is passive, but there is an active state } \tilde{x} \in \mathcal{B}(x)\}$$

be the "border" of the set of those states for which $\hat{\eta}_n^{(j)}(x)$ is not yet defined.

4. If $\Gamma^{(j)} = \emptyset$, then stop. The result $E[\eta_n] = \hat{\eta}_n^{(j)}$ is the desired extrapolation. If $\Gamma^{(j)} \neq \emptyset$, then do the following for every $x \in \Gamma^{(j)}$: Let

$$\{x^1, \ldots, x^\kappa\} = \mathcal{B}(x) \cap \{\tilde{x} \in \mathbf{\Omega}_{\bar{x}} \colon \chi^{(j)}(\tilde{x}) = 1\}$$

be those adjacent states of x, where $\hat{\eta}_n^{(j)}(x)$ is already available. By the definition of $\Gamma^{(j)}$ we have $1 \leq \kappa \leq 3^{d_1} - 1$. Define $\hat{\eta}_n^{(j+1)}(x)$ by averaging over these states:

$$\hat{\eta}_n^{(j+1)}(x) = \frac{1}{\kappa} \sum_{l=1}^{\kappa} \hat{\eta}_n^{(j)}(x^l).$$

Set $\chi^{(j+1)}(x) = 1$ for all $x \in \Gamma^{(j)}$, $j = j + 1$, and go to step 3.

It is worth mentioning that the extrapolation strategy does only give a crude estimate for $\eta_n(x)$. However, this does not cause any problems because wherever $\eta_n(x)$ appears in

the right-hand side of (5.8) and (5.9), it is multiplied with $w(t,x)$, and since $w(t,x) < \varepsilon$ for the passive states, the rather coarse approximation hardly affects the accuracy.

Time-integration by the Strang splitting. In order to keep the notation simple, we reformulate MRCE as

$$\partial_t w(t,x) = \mathcal{H}(\eta) w(t,x), \tag{5.10}$$

$$\partial_t \eta(t|x) \cdot w(t,x) = \mathcal{G}(\eta, w)(x), \tag{5.11}$$

where $\mathcal{H}(\eta)w$ and $\mathcal{G}(\eta, w)$ are abbreviations for the right-hand side of (5.8) and (5.9), respectively.

A Strang splitting method is used for the time-integration; i.e., in each substep only one of the two variables is updated while the other one is kept fixed. One time-step consists in the following substeps:

0. Input: $w_n(x)$ for all $x \in \mathbf{\Omega}_{\bar{x}}$ and $\eta_n(x)$ for all active states.
1. Compute the extrapolation $E[\eta_n]$ of $\eta_n(x)$ to the entire state space $\mathbf{\Omega}_{\bar{x}}$.
2. Half-step in the expectations: For all active states propagate (5.11) with a half-step of the implicit midpoint rule, i.e.,
$$\eta_{n+1/2}(x) = \eta_n(x) + \frac{h_n}{2w_n(x)} \mathcal{G}\left(\frac{E[\eta_n] + E[\eta_{n+1/2}]}{2}, w_n\right)(x),$$
where $E[\eta_{n+1/2}]$ and $E[\eta_n]$ are the extrapolations of $\eta_{n+1/2}$ and η_n to the entire state space $\mathbf{\Omega}_{\bar{x}}$. The factor $1/w_n(x)$ does not cause any problems here because $w(t,x) \geq \varepsilon$ for all active states. The nonlinear equation can be solved by fixed-point iteration. For all passive states, $\eta_{n+1/2}(x)$ is not defined.
3. Full step in the marginal distribution: Solve the reduced master equation (5.8) on the interval $[t_n, t_{n+1}]$, i.e., compute
$$w_{n+1} = \exp(h_n \mathcal{H}(E[\eta_{n+1/2}]))w_n.$$
4. Update the active-passive partition: A state $x \in \mathbf{\Omega}_{\bar{x}}$ is active if $w_{n+1}(x) \geq \varepsilon$ and passive otherwise.
5. Half-step in the expectations: For all active states compute $\eta_{n+1}(x)$ via
$$\eta_{n+1}(x) = \eta_{n+1/2}(x) + \frac{h_n}{2w_{n+1}(x)} \mathcal{G}\left(\frac{E[\eta_{n+1/2}] + E[\eta_{n+1}]}{2}, w_{n+1}\right)(x).$$

In the above formulation of the algorithm it was tacitly assumed that the reduced state space $\mathbf{\Omega}_{\bar{x}}$ is so small that the steps 2, 3, and 5 can be performed with no additional difficulties. In large applications, however, this is not possible anymore, and we come back to this issue in section 7.

6. Numerical examples. The three models are illustrated by a simple but significant test problem. Only $d = 2$ species are involved, and we choose $d_1 = d_2 = 1$ in the partition from section 2.2. This means, e.g., that the Hellander–Lötstedt model approximates the marginal distribution of S_1 and the expectation of S_2. The propensities and stoichiometric vectors are listed in Table 6.1.

A possible interpretation of this system is that S_1 carries an infectious disease and may infect the particles of S_2 via the fifth reaction channel. The other reaction channels correspond to the death of particles ($k = 1, 2$) and the birth or inflow of new particles

TABLE 6.1
Reaction channels of the test problem. The stoichiometric vectors ν_k and μ_k are scalars because $d_1 = d_2 = 1$.

k	Reaction channel	$\alpha_k(x)$	$\beta_k(y)$	ν_k	μ_k
1	$S_1 \to \emptyset$	$c_1 x$	1	-1	0
2	$S_2 \to \emptyset$	1	$c_2 y$	0	-1
3	$\star \to S_1$	c_3	1	1	0
4	$\star \to S_2$	1	c_4	0	1
5	$S_1 + S_2 \to S_1 + S_1$	$c_5 x$	y	1	-1

$(k = 3, 4)$. The first four reaction channels are monomolecular and leave S_1 and S_2 decoupled, but this fact does not give an advantage to any of the models if $c_5 \neq 0$.

The reason for considering this simple system is twofold. For a suitable choice of parameters, the state space of the problem is so small that the full CME (2.7) can be solved with high precision and provides a reference solution which allows us to investigate the accuracy of the hybrid models. On the other hand, the reaction system is complicated enough to generate bimodal solution profiles for certain configurations (see below).

All numerical experiments were implemented in MATLAB. The CME and the differential equations (3.5)–(3.6) and (3.10)–(3.11) were solved with the ODE solver `ode15s` with a very small tolerance such that the numerical error is negligible in comparison to the modeling error. For the time-integration of MRCE the adaptive Strang splitting method introduced in the previous section has been applied with step-size $h = 0.001$ and $\varepsilon = 10^{-8}$. In step 3 of the algorithm the matrix exponential was evaluated exactly with MATLAB's `expm`-command.

In the first test, the parameters

$$c_1 = 2, \quad c_2 = 2.5, \quad c_3 = 25, \quad c_4 = 75, \quad c_5 = 0.05, \quad t_{\text{end}} = 2 \tag{6.1}$$

are used, and the initial distribution is the projection of a product Poisson distribution with mean at $(4, 20)$ on the finite state space $\mathbf{\Omega}_{(\bar{x},\bar{y})}$. Snapshots of the solution[3] of the full CME (2.7) at different times are shown in Figure 6.1 and the left panel of Figure 6.2. Figure 6.2 shows that the product approximation (3.5)–(3.6) yields quite a reasonable approximation in this first example. The marginal distributions computed with the hybrid models are compared with the exact marginal distribution $p_1(t, x)$ in Figure 6.3. Although the Hellander–Lötstedt model uses only approximately half as many degrees of freedom as the product approximation, it achieves the same accuracy. This is not a surprise: since all $\beta_k(y)$ are at most linear, Corollary 2 states that $w(t, x)$ from (3.10) coincides with $u(t, x)$ from (3.5). MRCE yields nearly the exact result but uses almost twice as many degrees of freedom as the Hellander–Lötstedt model.

For the second test the parameters c_3 and c_5 were set to $c_3 = 0$ and $c_5 = 0.2$, respectively. This changes the behavior of the system completely. Now the only way for S_1 to reproduce is the reaction channel $S_1 + S_2 \to S_1 + S_1$, but this requires that at least one particle of S_1 is still left (i.e., $x \neq 0$). If all particles of S_1 have died, then no new

[3]In Figures 6.1, 6.2, 6.5, and 6.4 the two-dimensional solutions are visualized by contour plots, although the corresponding distributions are actually only defined in discrete states $(x, y) \in \mathbb{N}_0^2$.

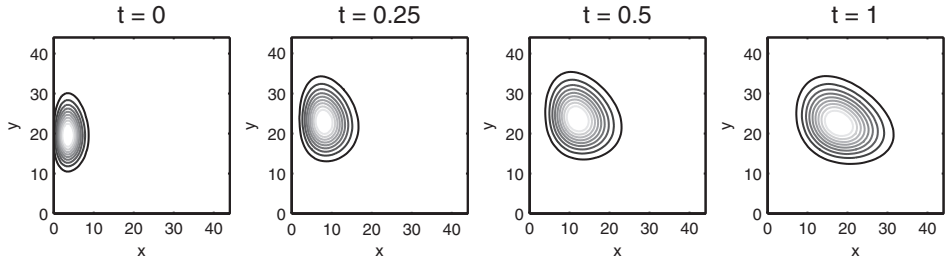

FIG. 6.1. *Contour plot of the solution of the full CME* (2.7) *with parametrization* (6.1) *at different times.*

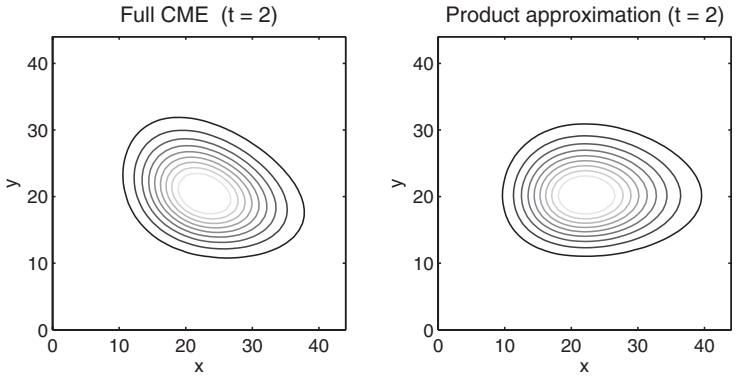

FIG. 6.2. *Left: Solution of the full CME* (2.7) *with parametrization* (6.1) *at time* $t = 2$. *Right: Product approximation at time* $t = 2$.

particles can appear because for $c_3 = 0$ there is no inflow via the reaction channel $\star \to S_1$. The snapshots in Figure 6.4 and the left panel of Figure 6.5 show that now the solution of the CME develops two distinct peaks as time evolves. The peak centered around (23,10) at $t = 2$ corresponds to the situation where S_1 and S_2 attain a coexistent equilibrium, whereas the peak on the axis $x = 0$ centered around $(0, 30)$ indicates the probability for extinction of S_1. Such a bimodal solution profile appears in many

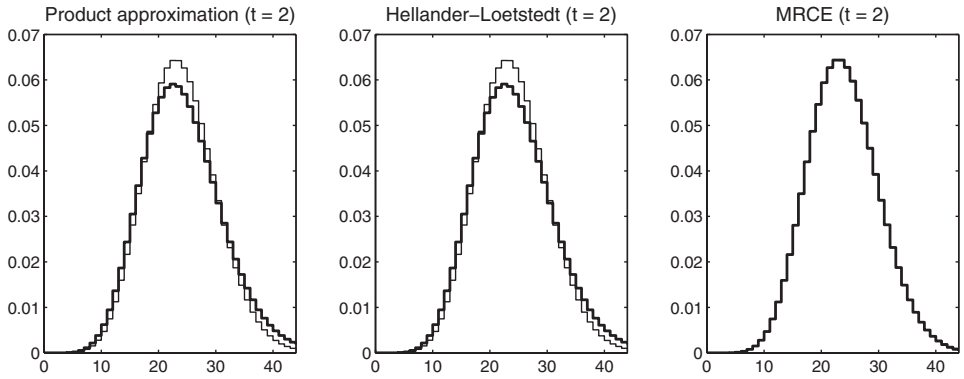

FIG. 6.3. *Comparison of the exact marginal distribution* $p_1(t, x)$ *(thin line) with the corresponding approximations obtained with the reduced models (bold line) for the parameters from* (6.1).

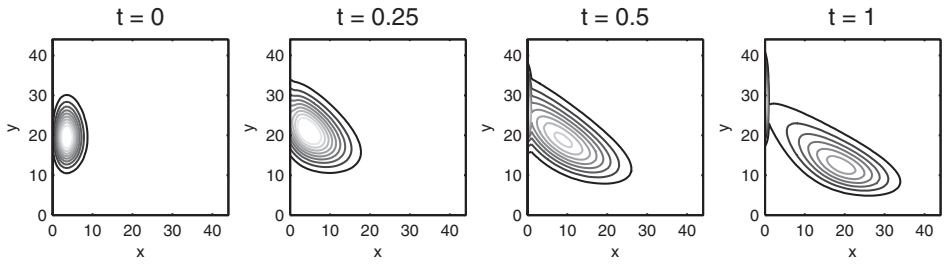

FIG. 6.4. *Contour plot of the solution of the full CME* (2.7) *with modified parametrization* ($c_3 = 0$, $c_5 = 0.2$) *at different times.*

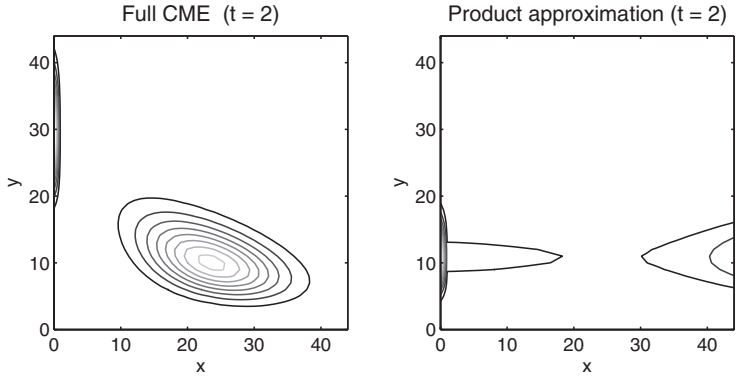

FIG. 6.5. *Left: Solution of the full CME* (2.7) *with modified parametrization* ($c_3 = 0$, $c_5 = 0.2$) *at time* $t = 2$. *Right: Product approximation at time* $t = 2$.

biological reaction systems and indicates that due to the stochastic evolution different scenarios are possible.

In the second example the product approximation gives a completely wrong result, which can be seen from Figure 6.5. The reason is that the essential rank of the distribu-

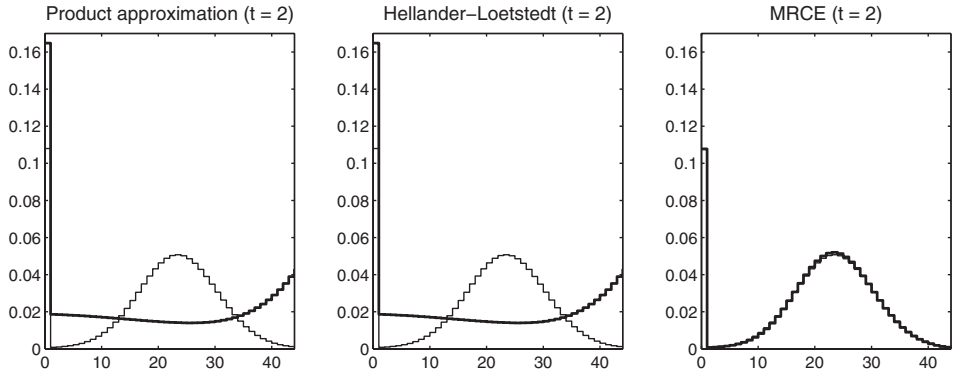

FIG. 6.6. *Comparison of the exact marginal distribution* $p_1(t, x)$ *(thin line) with the corresponding approximations obtained with the reduced models (bold line) with modified parametrization* ($c_3 = 0$, $c_5 = 0.2$).

tion $p(t, x, y)$ is now so high such that the solution cannot be approximated with one single tensor product. In Figure 6.6 the exact marginal distribution $p_1(t, x)$ is compared to the results of the hybrid models. Since Corollary 2 applies again, the Hellander–Lötstedt model cannot give a better result than the product approximation. The right panel shows, however, that MRCE reproduces the marginal distribution very accurately. Hence, the additional efforts required for MRCE (solving more complicated differential equations with more unknowns) pays off.

It should be pointed out, however, that the above example is rather simplistic and does not allow a fair assessment of the different models. One of our future research goals is to apply the MRCE models to considerably larger and more complicated reaction systems, but such applications are beyond the scope of the present work because they will require further development of the basic algorithm proposed in section 5.2 (see also the discussion in the next section).

7. Conclusions. The CME is crucial for the understanding of discrete stochastic reaction systems, but solving this equation numerically is usually computationally expensive or even impossible due to the huge size of the state space. This motivates the idea to replace the full CME by a reduced model which represents *some* of the species by a reduced CME coupled to other differential equations for the remaining species. Three different model reductions are discussed in this paper. Error estimates for the product approximation (3.5)–(3.6) and for the Hellander–Lötstedt model (3.10)–(3.11) are proven in Theorems 2 and 3, respectively. The analysis revealed that the main source of error in both models is the assumption that the solution of the CME can be approximated by a direct product. In the MRCE approach introduced in section 5, this condition is removed by an ansatz based on conditional expectations. This leads to a set of differential equations which involve approximately twice as many unknowns as the Hellander–Lötstedt model and which are more difficult to handle numerically and analytically. The numerical examples in section 6 indicate, however, that the additional efforts pay off and that MRCE may produce a very accurate result even in cases where the product approximation and the Hellander–Lötstedt model fail, in particular when the CME distribution splits into several modes.

The results presented in this paper give rise to a number of new questions. An error analysis of MRCE is missing so far. The derivation in section 5.1 suggests, however, that the approximation error could be bounded in terms of the covariances of the conditional expectations of the y-variables, and it is one of our future research goals to prove a corresponding error bound. Moreover, the numerical method outlined in section 5.2 is based on the assumption that the operator exponential in step 3 can be computed and that the nonlinear equations in steps 2 and 5 can be solved in a straightforward way. This is not possible when larger reaction systems are considered where even the reduced state space $\mathbf{\Omega}_{\bar{x}}$ is still large. In this case, these steps have to be carried out in an approximative sense by introducing a suitable spatial approximation of w_n and η_n, which allows us to reduce the number of degrees of freedom significantly. An extension of the adaptive wavelet method proposed in [24], [25] to the MRCE system (5.8)–(5.9) is one of our future research goals. We hope that such a method will allow us to approximate the dynamics of reaction systems which are way out of the scope of the existing methods.

Acknowledgments. The author thanks Per Lötstedt, Andreas Hellander, Tudor Udrescu, and Michael Kreim for their valuable comments on a preliminary draft of this paper, and the two anonymous referees for their helpful remarks and suggestions.

REFERENCES

[1] A. Alfonsi, E. Cancès, G. Turinici, B. Di Ventura, and W. Huisinga, *Adaptive Simulation of Hybrid Stochastic and Deterministic Models for Biochemical Systems*, ESAIM Proc. 14, EDP Sciences, Les Ulis, France, 2005, pp. 1–13.

[2] K. Ball, T. G. Kurtz, L. Popovic, and G. Rempala, *Asymptotic analysis of multiscale approximations to reaction networks*, Ann. Appl. Probab., 16 (2006), pp. 1925–1961.

[3] F. A. Bornemann and C. Schütte, *On the singular limit of the quantum-classical molecular dynamics model*, SIAM J. Appl. Math., 59 (1999), pp. 1208–1224.

[4] P. Brémaud, *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*, Texts Appl. Math., Springer, New York, 1999.

[5] K. Burrage, T. Tian, and P. Burrage, *A multi-scaled approach for simulating chemical reaction systems*, Prog. Biophys. Molec. Biol., 85 (2004), pp. 217–234.

[6] Y. Cao and L. Petzold, *Slow scale tau-leaping method*, Comput. Methods Appl. Mech. Engrg., 197 (2008), pp. 3472–3479.

[7] S. Engblom, *Computing the moments of high dimensional solutions of the master equation*, Appl. Math. Comput., 180 (2006), pp. 498–515.

[8] L. Ferm and P. Lötstedt, *Numerical method for coupling the macro and meso scales in stochastic chemical kinetics*, BIT, 47 (2007), pp. 735–762.

[9] L. Ferm, P. Lötstedt, and A. Hellander, *A hierarchy of approximations of the master equation scaled by a size parameter*, J. Sci. Comput., 34 (2008), pp. 127–151.

[10] C. W. Gardiner, *Handbook of Stochastic Methods*, 4th ed., Springer Ser. Synergetics, Springer, Berlin, 2009.

[11] D. T. Gillespie, *A general method for numerically simulating the stochastic time evolution of coupled chemical reactions*, J. Comput. Phys., 22 (1976), pp. 403–434.

[12] D. T. Gillespie, *A rigorous derivation of the chemical master equation*, Phys. A., 188 (1992), pp. 404–425.

[13] M. Griebel and L. Jager, *The BGY3dM model for the approximation of solvent densities*, J. Chem. Phys., 129 (2008), pp. 174511–174525.

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

[15] M. Hegland, A. Hellander, and P. Lötstedt, *Sparse grids and hybrid methods for the chemical master equation*, BIT, 48 (2008), pp. 265–284.

[16] A. Hellander and P. Lötstedt, *Hybrid method for the chemical master equation*, J. Comput. Phys., 227 (2007), pp. 100–122.

[17] T. A. Henzinger, M. Mateescu, L. Mikeev, and V. Wolf, *Hybrid numerical solution of the chemical master equation*, in Proceedings of the 8th International Conference on Computational Methods in Systems Biology (CMSB10), ACM, New York, 2010.

[18] D. J. Higham, *Modeling and simulating chemical reactions*, SIAM Rev., 50 (2008), pp. 347–368.

[19] P. D. Iedema, M. Wulkow, and H. C. J. Hoefsloot, *Modeling molecular weight and degree of branching distribution of low-density polyethylene*, Macromolecules, 33 (2000), pp. 7173–7184.

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

[21] T. Jahnke and W. Huisinga, *A dynamical low-rank approach to the chemical master equation*, Bull. Math. Biol., 70 (2008), pp. 2283–2302.

[22] T. Jahnke, *On hybrid models in stochastic reaction kinetics*, in Numerical Analysis and Applied Mathematics: International Conference on Numerical Analysis and Applied Mathematics 2009: Volume 1 and Volume 2, AIP Conf. Proc. 1168, T. E. Simos, G. Psihoyios, and Ch. Tsitouras, eds., Amer. Inst. Phys., Melville, NY, 2009, pp. 1540–1543.

[23] T. Jahnke and D. Altintan, *Efficient simulation of discrete stochastic reaction systems with a splitting method*, BIT, 50 (2010), pp. 797–822.

[24] T. Jahnke, *An adaptive wavelet method for the chemical master equation*, SIAM J. Sci. Comput., 31 (2010), pp. 4373–4394.

[25] T. Jahnke and T. Udrescu, *Solving chemical master equations by adaptive wavelet compression*, J. Comput. Phys., 229 (2010), pp. 5724–5741.

[26] R. Khanin and D. J. Higham, *Chemical master equation and Langevin regimes for a gene transcription model*, Theoret. Comput. Sci., 408 (2008), pp. 31–40.

[27] O. Koch and C. Lubich, *Dynamical low rank approximation*, SIAM J. Matrix Anal. Appl., 29 (2007), pp. 434–454.

[28] O. Koch and C. Lubich, *Dynamical tensor approximation*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 2360–2375.
[29] P. Lötstedt and L. Ferm, *Dimensional reduction of the Fokker–Planck equation for stochastic chemical reactions*, Multiscale Model. Simul., 5 (2006), pp. 593–614.
[30] C. Lubich, *From Quantum to Classical Molecular Dynamics: Reduced Models and Numerical Analysis*, Zurich Lect. Adv. Math., Zürich: European Mathematical Society (EMS), Zurich, 2008.
[31] S. Menz, J. C. Latorre, C. Schütte, and W. Huisinga, *Hybrid Stochastic-Deterministic Solution of the Chemical Master Equation*, preprint, Freie Universität, Berlin, 2011.
[32] C. V. Rao and A. P. Arkin, *Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm*, J. Chem. Phys., 118 (2003), pp. 4999–5010.
[33] H. Salis and Y. Kaznessis, *Accurate hybrid simulation of a system of coupled chemical or biochemical reactions*, J. Chem. Phys., 122, (2005), 54103.
[34] O. Schreiber, *Dimensionsreduktionsmethoden für die chemische Mastergleichung*, Diploma thesis, Freie Universität, Berlin, 2010.
[35] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry*, North–Holland Personal Library, 3rd ed., North-Holland, Amsterdam, 2001.
[36] E. Weinan, D. Liu, and E. Vanden-Eijnden, *Nested stochastic simulation algorithms for chemical kinetic systems with multiple time scales*, J. Comput. Phys., 221 (2007), pp. 158–180.