Model Reduction for Stochastic Reaction Systems

Stephen Smith and Ramon Grima

1 Introduction

Master equations constitute the standard description of stochastic reaction dynamics in well-mixed conditions [1]. For linear systems, the moments can be found exactly in closed-form and sometimes even the probability distribution can be obtained [2]. However, most systems are nonlinear, specifically those involving the interaction of two or more entities. For such systems the master equation can be rarely solved and a different solution strategy becomes necessary. A common approach involves performing stochastic simulations using the stochastic simulation algorithm (SSA) [3]. However, the computational expense involved in gathering accurate statistics can be considerable for many systems of interest, particularly those which involve a high number of reaction events per unit time, a feature of systems with one or more abundant species. One means to circumvent this problem involves deriving a reduced master equation for only the non-abundant species and then obtaining the statistics of the number fluctuations using the SSA. Various methods exist which lead to such a reduction [4–7]. Here we choose to focus on what is perhaps the simplest of such methods, one which is easy to derive for all cases of interest and which leads to accurate and fast stochastic simulations. An additional bonus is that in quite a number of cases, the reduced master equation can be solved exactly.

S. Smith • R. Grima (✉)
School of Biological Sciences, University of Edinburgh, Edinburgh, UK
e-mail: ramon.grima@ed.ac.uk

D. Holcman (ed.), *Stochastic Processes, Multiscale Modeling, and Numerical Methods for Computational Cellular Biology*, DOI 10.1007/978-3-319-62627-7_7

2 The Method

In this section we introduce the rationale behind the model reduction method, by means of a simple chemical reaction system. For a more technical presentation the reader is referred to the original paper which derives the method [8]. This section is self-contained, assuming no knowledge of rate equations or of master equations, and builds the latter mathematical descriptions and the reduction method from the ground up.

Consider the following chemical reaction system:

$$\emptyset \xrightarrow{k_1} A,\ A + A \xrightarrow{k_2} B,\ B \xrightarrow{k_3} \emptyset. \tag{1}$$

This system consists of two proteins, A and B, and three reactions. The first reaction is the creation of a molecule of A out of nothingness ($\emptyset$). This process is, of course, thermodynamically implausible, but this notation is used when A is created at a more-or-less constant rate by a reaction between chemical species which we are not interested in modelling. The second reaction is between two molecules of A to form a molecule B. We call B a *dimer* and the reaction a *dimerisation*. The third reaction is the destruction of B into nothingness. Again, this is shorthand for the conversion of B into products which we do not want to explicitly model. The quantities $k_1,\ k_2,\ k_3$ are the *rates* of the reactions, i.e. a measure of how frequently they occur.

The most common method of modelling systems like (1) is to treat the concentrations of A and B (number of molecules per unit volume) as continuous functions of time, $[A]$ and $[B]$, respectively. These functions are defined as the solution to a set of differential equations called rate equations (RE):

$$\begin{aligned} \frac{d[A]}{dt} &= k_1 - 2k_2[A]^2, \\ \frac{d[B]}{dt} &= k_2[A]^2 - k_3[B]. \end{aligned} \tag{2}$$

The equation for $\frac{d[A]}{dt}$ states that the rate of change of $[A]$ is equal to the rate of creation of A molecules, k_1, minus the rate of destruction of A molecules, $2k_2[A]^2$. The factor of 2 in the latter term refers to the fact that *two* molecules of A are destroyed whenever the reaction occurs; the factor of $[A]^2$ relates to the fact the reaction happens with a frequency proportional to the number of pairs of A molecules, which scales as $[A]^2$. The equation for $\frac{d[B]}{dt}$ states that the rate of change of $[B]$ is equal to the rate of creation of B molecules, $k_2[A]^2$, minus the rate of destruction of B molecules, $k_3[B]$. The term $k_3[B]$ corresponds to the fact that the third reaction happens with a frequency proportional to the number of B molecules, which scales as $[B]$. For more details, a standard reference book can be consulted [1].

The RE system (2) can be solved for $[A]$ and $[B]$ as functions of time, but for simplicity we will consider the equilibrium (steady-state) case, i.e. when $\frac{d[A]}{dt} = \frac{d[B]}{dt} = 0$. Setting the equations in system (2) equal to zero leads to the following simple expressions:

$$[A] = \sqrt{\frac{k_1}{2k_2}},\ [B] = \frac{k_1}{2k_3}. \tag{3}$$

We therefore know how the concentrations of A and B scale with the various reaction rates. However, our entire line of thinking so far has relied on the assumption that the concentrations of A and B are continuous (even differentiable) functions of time, an assumption which is clearly untrue since the number of molecules of A and B must be integer-valued. Following this line of thinking leads us to consider not the concentrations $[A]$ and $[B]$ but the molecule number n_A and n_B. We furthermore observe that chemical kinetics is not deterministic but rather probabilistic. The reason is that the timing of reaction events is random; for example, the precise time at which two molecules of A will meet is unknown because the process which brings the molecules together, Brownian motion, is a stochastic process. We are therefore concerned with the quantity $P(n_A, n_B; t)$, the probability that our system (1) consists of exactly n_A molecules of A and n_B molecules of B at time t.

Just as with the deterministic RE system (2), the probability $P(n_A, n_B; t)$ is described by a differential equation of the form:

$$\begin{aligned}\frac{d}{dt}P(n_A, n_B; t) &= k_1 V \left(P(n_A - 1, n_B; t) - P(n_A, n_B; t)\right) \\ &+ \frac{k_2}{V}\left((n_A + 2)(n_A + 1)P(n_A + 2, n_B - 1; t) - n_A(n_A - 1)P(n_A, n_B; t)\right) \\ &+ k_3\left((n_B + 1)P(n_A, n_B + 1; t) - n_B P(n_A, n_B; t)\right).\end{aligned} \tag{4}$$

This equation is a master equation, specifically a chemical master equation (CME). It describes the rate of change of the probability of the system having n_A, n_B molecules of A and B, respectively, or in the language of statistical physics, the rate of change of the probability that the system is in the *state* (n_A, n_B).

The first term of Eq. (4) concerns how the system could enter or leave the state (n_A, n_B) due to the first reaction. The system could enter the state (n_A, n_B) due to the production of a molecule of A, if the system was previously in the state $(n_A - 1, n_B)$ [hence the probability $P(n_A - 1, n_B; t)$] or else the system could leave the state (n_A, n_B) if it was already in that state [hence the probability $P(n_A, n_B; t)$]. Note that the first reaction happens with a rate $k_1 V$, for reaction volume V, because the production reaction can occur anywhere throughout the reaction volume.

The second term of Eq. (4) concerns how the system could enter or leave the state (n_A, n_B) due to the second (dimerisation) reaction. The system could enter the state (n_A, n_B) if it was previously in the state $(n_A + 2, n_B - 1)$, or else the system could leave the state (n_A, n_B) if it was already in that state. If the system was in state $(n_A + 2, n_B - 1)$, then the second reaction would occur with rate $\frac{k_2}{V}(n_A + 2)(n_A + 1)$, since the number of distinct pairs of A molecules scales as $(n_A + 2)(n_A + 1)$, and the factor V corresponds to the fact that dimerisation reactions are less likely to occur in large volumes (since molecules are less likely to collide). Similarly, if the system was in state (n_A, n_B), then the second reaction would occur with rate $\frac{k_2}{V} n_A(n_A - 1)$, since the number of distinct pairs of A molecules scales as $n_A(n_A - 1)$.

The third term of Eq. (4) concerns how the system could enter or leave the state (n_A, n_B) due to the third reaction. The system could enter the state (n_A, n_B) if it was previously in state $(n_A, n_B + 1)$, or else the system could leave the state (n_A, n_B) if it was already in that state. If the system was in state $(n_A, n_B + 1)$ the reaction would occur with a rate $k_3(n_B + 1)$, since there are $(n_B + 1)$ molecules of B, but there is no volume dependence for this reaction. Similarly, if the system was in state (n_A, n_B), the third reaction would occur with a rate $k_3 n_B$. For more details, see a standard reference book on stochastic methods [1].

While it can be seen that Eq. (4) corresponds to a physically accurate description of the chemical system (1), it is by no means clear how to solve it, even at steady-state. In fact, Eq. (4) cannot be easily solved analytically. Only a small number of CMEs can be solved, and these generally have special properties such as conservation laws [9], detailed balance [10] or no bimolecular reactions [11]. Instead, the most common stochastic approach to systems like (1) is to simulate them using the stochastic simulation algorithm (SSA) [3]. Performing a large number of independent simulations can generate a histogram which is known to agree with the solution of the CME, within sampling error. However, this technique does not tell us how the average number of molecules of A and B (other moments) depend on the various parameters, analogously to Eqs. (3) for the rate equations. If we want this kind of information we must use methods which analytically approximate $P(n_A, n_B; t)$, and we will describe one such method here.

This method [8] makes the assumption that some of the chemical species have a high concentration. For instance, in system (1), suppose that there are a large number of A molecules. By the definition of concentration, we can approximate the number of A molecules as:

$$n_A \approx V[A], \tag{5}$$

where $[A]$ is the concentration given in Eq. (3). In fact, this approximation becomes more accurate for larger concentrations of $[A]$. As shown in Fig. 1, the stochastic behaviour of n_A (as simulated with the SSA) becomes less distinguishable from the constant solution Eq. (3) as n_A increases. It follows that the stochastic behaviour of n_A becomes less relevant as $[A]$ grows, and it also follows that, if $[A]$ is large,

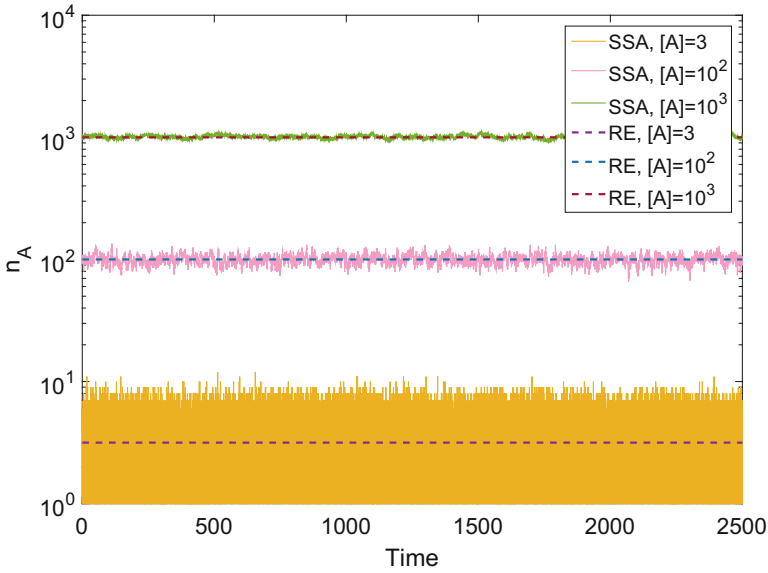

Fig. 1 Stochastic simulations (*yellow*, *pink*, *green*) and deterministic rate-equation solutions (*purple*, *blue*, *red*) for n_A is the system (1). Parameter values are $k_1 = 20$, $k_3 = 1$, $V = 1$ and $k_2 = 1, 10^{-3}, 10^{-5}$ for $[A] = 3, 10^2, 10^3$, respectively. Note that the noise about the mean decreases with increasing mean number of molecules

we can replace every instance of n_A in Eq. (4) with $V[A]$. Also, since n_A is large, $(n_A + i) \approx V[A]$ for small values of i. Hence we have

$$\begin{aligned}\frac{d}{dt}P(V[A], n_B; t) &= k_1 V \left(P(V[A], n_B; t) - P(V[A], n_B; t)\right) \\ &+ \frac{k_2}{V}\left((V[A])^2 P(V[A], n_B - 1; t) - (V[A])^2 P(V[A], n_B; t)\right) \\ &+ k_3 \left((n_B + 1)P(V[A], n_B + 1; t) - n_B P(V[A], n_B; t)\right).\end{aligned} \tag{6}$$

Several simplifications can be made to this equation. The first term, for instance, is now identically zero. Also, notice that since $[A]$ is simply the constant defined in Eq. (3), there is no need to include it as an argument in the probability P, hence we can define a new, simplified probability $\tilde{P}(n_B; t)$, leading to a simplified CME:

$$\begin{aligned}\frac{d}{dt}\tilde{P}(n_B; t) =& \frac{k_1}{2} V \left(\tilde{P}(n_B - 1; t) - \tilde{P}(n_B; t)\right) \\ &+ k_3 \left((n_B + 1)\tilde{P}(n_B + 1; t) - n_B \tilde{P}(n_B; t)\right).\end{aligned} \tag{7}$$

This equation is relatively straightforward, and is known to have the steady-state solution:

$$\tilde{P}(n_B) = \frac{e^{-\frac{k_1 V}{2k_3}} \left(\frac{k_1 V}{2k_3}\right)^{n_B}}{n_B!}. \tag{8}$$

In other words, the number of B molecules approximately satisfies a Poisson distribution with mean $k_1 V/2k_3$.

This approximation can also be made in a simpler (albeit less rigorous) manner immediately from the reaction system (1). As a rule, wherever we see A to the right of an arrow, we simply delete it (replace it with $\emptyset$). Wherever we see A to the left of an arrow, we absorb it as $[A]$ into the reaction rate. This replaces system (1) with the system:

$$\emptyset \xrightarrow{k_1} \emptyset, \ \emptyset \xrightarrow{k_2[A]^2} B, \ B \xrightarrow{k_3} \emptyset. \tag{9}$$

The first reaction here is clearly pointless, and $[A]$ can be replaced by its value given in Eq. (3) to give:

$$\emptyset \xrightarrow{\frac{k_1}{2}} B, \ B \xrightarrow{k_3} \emptyset. \tag{10}$$

It can be shown that the CME for system (10) is identically Eq. (7). We therefore have a very quick way to reduce complex chemical reaction systems to simpler subsystems which we know how to solve.

Yet, although this technique has been derived a priori, a question remains as to the accuracy of the simplified system. How large does $[A]$ have to be before the simplified system is a good approximation to the true system? In fact, for most cases, the simplified system is reasonable even when $[A]$ is not large. For example, in Fig. 2 we show how the Poisson distribution in Eq. (8) agrees very well with the true solution of the CME when $[A]$ is large, but also shows reasonable agreement when $[A]$ is much smaller than $[B]$. In general, however, the error scales as the inverse ratio of the two concentrations, so that the larger $[A]$ is compared to $[B]$, the better the approximation.

The example shown above is illustrative but of limited biological relevance or interest. Next we will therefore demonstrate the power of the analytical approximation method by applying it to a variety of biological systems.

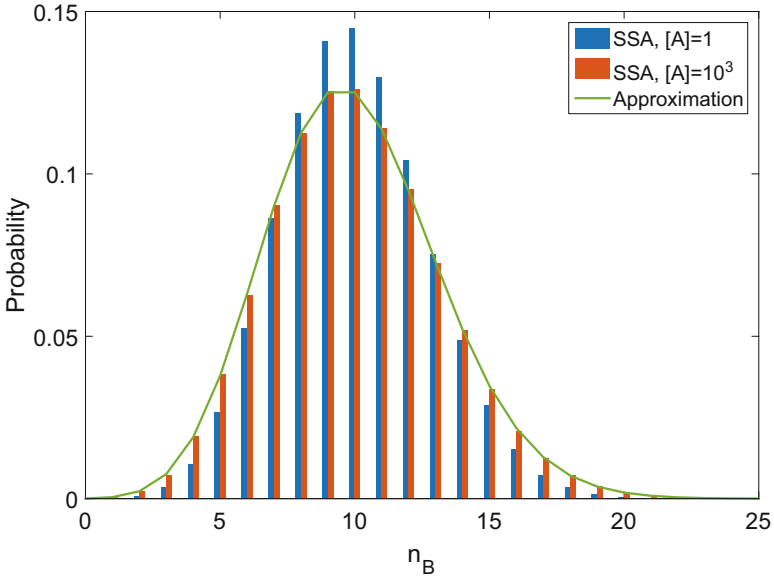

Fig. 2 Probability distributions of n_B obtained from the SSA (histograms) and the analytical approximation assuming abundance of A as given by Eq. (8) (*solid line*). Parameter values are as in Fig. 1 and $k_2 = 10, 10^{-5}$ for $[A] = 1, 10^3$, respectively. Note that the approximation is particularly good when $[A] \gg [B] = 10$

3 Application to Various Biological Systems

3.1 Michaelis–Menten Reaction

Consider the set of chemical reactions:

$$\emptyset \xrightarrow{k_1} S,\ E + S \xrightarrow{k_2} C,\ C \xrightarrow{k_3} E + S,\ C \xrightarrow{k_4} E + P. \tag{11}$$

The system describes how a protein S is created (with rate k_1) and reacts with an enzyme E to form a complex C (with rate k_2). C can subsequently unbind, either back to $E + S$ (with rate k_3) or else to E and new protein P (with rate k_4). This system has been studied extensively with and without protein production, using rate equations and master equations [12, 13]. The first reaction could, for example, model the effective translation of a protein in the cytoplasm while the rest of the reactions model the enzyme-aided catalysis of the protein into another type of protein.

The REs for this system are given by:

$$\begin{aligned}\frac{d[S]}{dt} &= k_1 - k_2[E][S] + k_3[C],\\ \frac{d[E]}{dt} &= -k_2[E][S] + (k_3 + k_4)[C],\\ \frac{d[C]}{dt} &= k_2[E][S] - (k_3 + k_4)[C].\end{aligned} \tag{12}$$

In steady-state, the solutions of this system are

$$[S] = \frac{k_1(k_3 + k_4)}{k_2(E_T k_4 - k_1)},\ [E] = E_T - \frac{k_1}{k_4},\ [C] = \frac{k_1}{k_4}, \tag{13}$$

where $E_T = [E] + [C]$ is the total enzyme concentration which remains constant over time. In order to apply the approximation method, we make the assumption that $[S]$ is large compared to $[E]$ and $[C]$. Then, we can reduce the reaction network to give:

$$E \xrightarrow{k_2[S]} C,\ C \xrightarrow{k_3+k_4} E. \tag{14}$$

The steady-state solution of the chemical master equation for this effective system of reactions is easily found to be a Binomial $\left(E_T V, \frac{k_3+k_4}{k_3+k_4+k_2[S]}\right)$ distribution for n_E. In Fig. 3 we compare this Binomial distribution with SSA histograms for three different values of $[S]$. As $[S]$ increases, it is clear that the approximation improves, though the Binomial distribution seems to be a reasonable estimate of the distribution even when $[S]$ is comparable to E.

3.2 Genetic Network with Feedback

Next we study a simple model of a genetic network with negative feedback:

$$\begin{aligned}&D_{\mathrm{ON}} \xrightarrow{k_1} D_{\mathrm{ON}} + M,\ M \xrightarrow{k_2} M + P_1,\ P_1 \xrightarrow{k_3} P_2,\ P_2 \xrightarrow{k_4} P_3,\ P_3 \xrightarrow{k_5} \emptyset,\\ &D_{\mathrm{ON}} + P_3 \xrightarrow{k_6} D_{\mathrm{OFF}},\ D_{\mathrm{OFF}} \xrightarrow{k_7} D_{\mathrm{ON}} + P_3,\ M \xrightarrow{k_8} \emptyset.\end{aligned} \tag{15}$$

The system describes how a molecule of mRNA M is produced by an active gene D_{ON} with rate k_1 (transcription), which in turn creates a protein P_1 with rate k_2 via

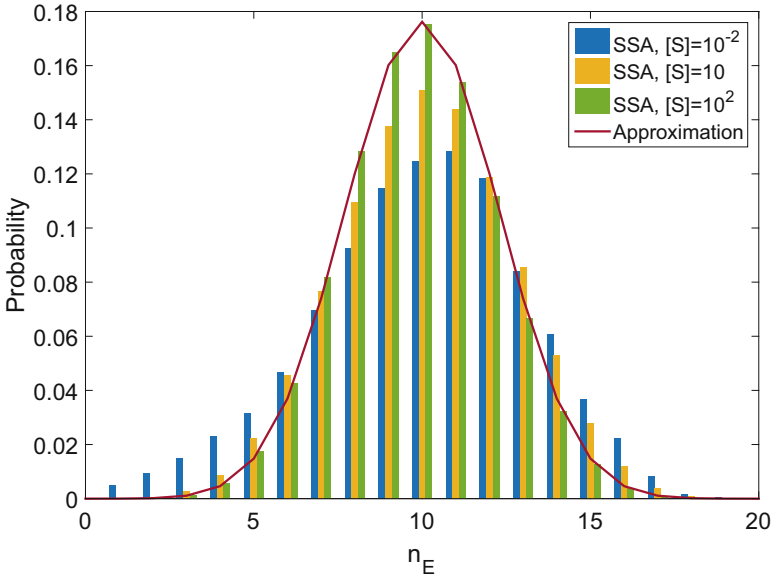

Fig. 3 Probability distribution of n_E is the Michaelis–Menten reaction system (11) obtained from the SSA of the full system (histograms) and the analytical approximation corresponding to the reduced system (14) assuming abundance of S compared to enzyme (*red solid line*). Parameter values are $k_1 = 1,\ k_3 = 9.9,\ k_4 = 0.1,\ E_T = 20,\ V = 1$. k_2 is varied to give different values of $[S]$. Note that the approximation is excellent when the concentration of S is much greater than that of enzyme

the process of translation. P_1 isomerises to P_2 with rate k_3 which isomerises to P_3 with rate k_4. P_3 can decay with rate k_5 or it can bind to the active gene D_{ON} with rate k_6 to convert it to the inactive gene D_{OFF}, and the deactivation can be reversed with rate k_7. Finally, the mRNA can decay with rate k_8. The system possesses negative feedback since the protein produced by the gene in the on state can turn the gene off at high enough concentrations; it has previously been studied as a simple model of a circadian oscillator [14, 15].

The REs for this system are given by:

$$\frac{d[D_{\text{ON}}]}{dt} = -k_6[D_{\text{ON}}][P_3] + k_7[D_{\text{OFF}}],$$

$$\frac{d[D_{\text{OFF}}]}{dt} = k_6[D_{\text{ON}}][P_3] - k_7[D_{\text{OFF}}],$$

$$\frac{d[M]}{dt} = k_1[D_{\text{ON}}] - k_8[M],$$

$$\frac{d[P_1]}{dt} = k_2[M] - k_3[P_1],$$
$$\frac{d[P_2]}{dt} = k_3[P_1] - k_4[P_2],$$
$$\frac{d[P_3]}{dt} = k_4[P_2] - k_5[P_3] - k_6[D_{\text{ON}}][P_3] + k_7[D_{\text{OFF}}]. \tag{16}$$

This system of equations can be solved in steady-state, but the expressions are cumbersome and we will not state them here except for $[P_3]$:

$$[P_3] = \frac{-k_7 + \sqrt{k_7^2 + \dfrac{4k_6k_1k_2k_7N}{k_5k_8}}}{2k_6}, \tag{17}$$

where $N = [D_{\text{ON}}] + [D_{\text{OFF}}]$ is the total gene concentration which remains constant over time. In order to apply the approximation method, we make the reasonable assumption that the protein concentrations $[P_1]$, $[P_2]$, $[P_3]$ are large compared to $[D_{\text{ON}}]$, $[D_{\text{OFF}}]$ and $[M]$. For example, it has been shown that the mean number of proteins per *E. coli* cell is roughly a thousand times that of the mean number of mRNA molecules per cell while the gene copy number is often one [16]. Then, we can reduce the reaction network to give:

$$D_{\text{ON}} \xrightarrow{k_1} D_{\text{ON}} + M, D_{\text{ON}} \xrightarrow{k_6[P_3]} D_{\text{OFF}},\ D_{\text{OFF}} \xrightarrow{k_7} D_{\text{ON}},\ M \xrightarrow{k_8} \emptyset. \tag{18}$$

The chemical master equation for this system can be solved exactly in steady-state using the generating function method [9] . The solution is complex so we will not state it here, but in Fig. 4 we compare the analytical approximation with the SSA distribution, and observe that they agree well. In the inset we show trajectories of the SSA for M (red) and P_1, P_2, P_3 (yellow, blue, purple), which highlight the bimodal nature of this system. The parameter set chosen here highlights two remarkable properties of the approximation method. First, the approximate distribution captures the bimodality of the true distribution. This is surprising because approximation methods are rarely able to deal with bimodality. Second, the distribution is accurate even though for a significant portion of time the system is in the lower state where $[P_1] = [P_2] = [P_3] = 0$, which can hardly be described as a high concentration. It can be shown that the accuracy of the approximation depends only on whether the RE solution is large, irrespective of the stochastic behaviour.

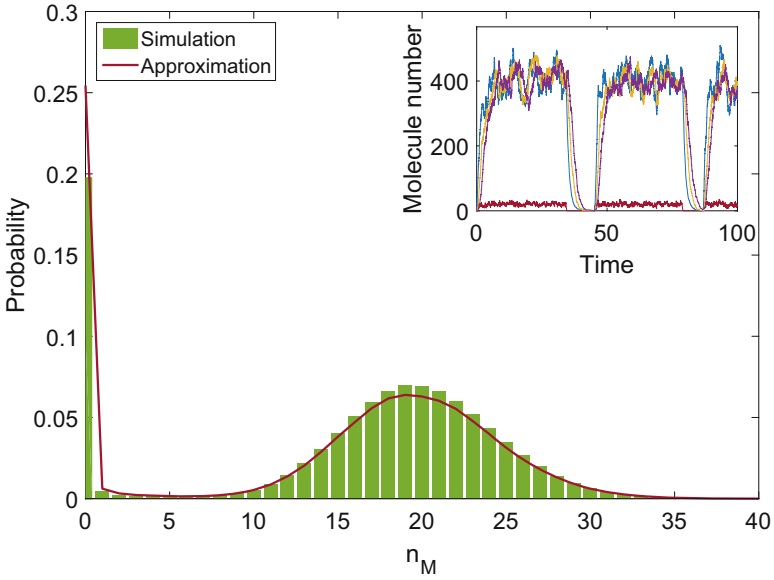

Fig. 4 Probability distribution of the genetic feedback system (15), as obtained from the SSA of the full system (histogram) and the approximation corresponding to the reduced system (18) assuming protein abundance (*red solid line*). Inset: time courses of M (*red*) and P_1, P_2, P_3 (*blue*, *yellow*, *purple*) as obtained from SSA, showing bimodality in both mRNA and protein values and abundance of protein compared to mRNA for most of the time. Parameter values are $k_1 = 100$, $k_2 = 5$, $k_3 = 20$, $k_4 = 1$, $k_5 = 1$, $k_6 = 1$, $k_7 = 10^{-4}$, $k_8 = 0.1$, $N = 1$, $V = 1$

3.3 *Biochemical Switch*

Next we consider a simple model of a biochemical switch [17], which could be used to construct synthetic logic gates:

$$C \xrightarrow{k_1} C + X_1,\ X_1 \xrightarrow{k_2} \emptyset,\ X_1 + B \underset{k_4}{\overset{k_3}{\rightleftharpoons}} A,\ X_2 + A \underset{k_6}{\overset{k_5}{\rightleftharpoons}} B,\ \emptyset \underset{k_8}{\overset{k_7}{\rightleftharpoons}} X_2. \tag{19}$$

The system describes how three enzymes A, B, C catalyse the production and degradation of two proteins X_1 and X_2. The engineering application of this system is that the output, A, is an amplification of the input, C, since a small change in $[C]$

corresponds to a large change in $[A]$, thus functioning as a switch. The REs for this system are given by:

$$\begin{aligned}
\frac{d[A]}{dt} &= k_3[X_1][B] - k_4[A] - k_5[X_2][A] + k_6[B], \\
\frac{d[B]}{dt} &= -k_3[X_1][B] + k_4[A] + k_5[X_2][A] - k_6[B], \\
\frac{d[X_1]}{dt} &= k_1[C] - k_2[X_1] - k_3[X_1][B] + k_4[A], \\
\frac{d[X_2]}{dt} &= -k_5[X_2][A] + k_6[B] + k_7 - k_8[X_2].
\end{aligned} \tag{20}$$

This system can be solved in steady-state, but the expressions are cumbersome and we will not state them here. In order to apply the approximation method, we make the reasonable assumption that the protein concentrations $[X_1]$, $[X_2]$ are large compared to the enzyme concentrations $[A]$ and $[B]$ (similar to the assumption made for the Michaelis–Menten system). Then, we can reduce the reaction network to the effective one:

$$A \underset{k_3[X_1]+k_6}{\overset{k_4+k_5[X_2]}{\rightleftharpoons}} B. \tag{21}$$

Analogously to system (14), the chemical master equation for this system has a steady-state solution given by a Binomial distribution, specifically a Binomial $\left(E_T V, \frac{k_4+k_5[X_2]}{k_4+k_5[X_2]+k_3[X_1]+k_6}\right)$ distribution for n_A, where $E_T = [A] + [B]$ is the total concentrations of enzyme, which remains constant over time.

In Fig. 5 we plot the distribution of n_A for a variety of values of $[C]$, for both the analytical approximation corresponding to the effective system (21) and stochastic simulations using the SSA of the master equation of the full system (19). The switch-like behaviour is clear to see, as the mean of n_A moves swiftly from about 10 to near 100 molecules (almost a ninefold change) as $[C]V$ (the input number of C molecules) is changed from 500 to 1500 molecules (a twofold change). The approximation method here provides a very quick means of modelling a bioengineering component, which would otherwise be time-consuming to simulate.

3.4 Predator–Prey System

Lastly we consider an example from ecology, a Lotka–Volterra-like predator–prey system [18] given by the reactions:

$$R \xrightarrow{k_1} R + R,\ R + F \xrightarrow{k_2} F + F,\ F \xrightarrow{k_3} \emptyset,\ \emptyset \xrightarrow{k_4} R,\ \emptyset \xrightarrow{k_5} F. \tag{22}$$

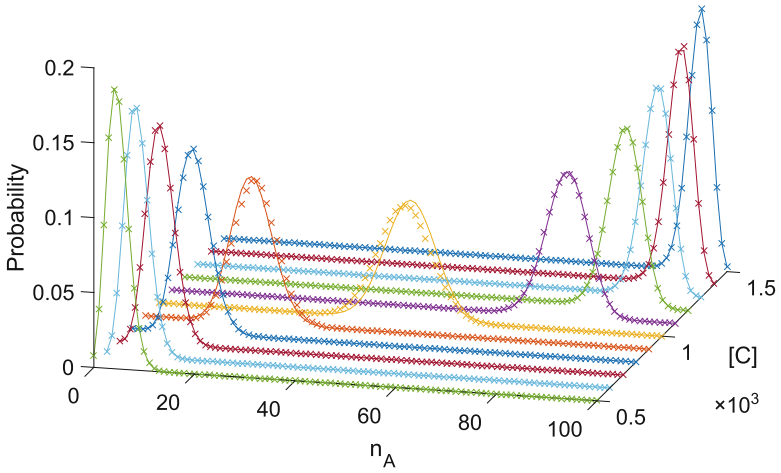

Fig. 5 Probability distributions of n_A in the biochemical switch (19) for different values of $[C]$, obtained from the SSA of the full system (*crosses*) and the analytical approximation corresponding to the reduced system (21) under the assumption of protein abundance (*solid lines*). Parameter values are $k_1 = 0.05$, $k_2 = 0.001$, $k_3 = 0.001$, $k_4 = 1$, $k_5 = 0.001$, $k_6 = 1$, $k_7 = 0.001$, $k_8 = 50$, $V = 1$. Note that a twofold change in the input [C] leads to an almost ninefold change in the mean of the output A

The system describes (in a simplistic way) how a population of foxes F predates on a population of rabbits R. A single rabbit can reproduce with rate k_1 and a fox can eat a rabbit, giving it enough energy to reproduce with rate k_2. Foxes can die with rate k_3, and rabbits and foxes can both immigrate into the environment with rates k_4 and k_5, respectively. The REs for this system are

$$\begin{aligned} \frac{d[R]}{dt} &= k_1[R] - k_2[R][F] + k_4, \\ \frac{d[F]}{dt} &= k_2[R][F] - k_3[F] + k_5. \end{aligned} \tag{23}$$

This system of equations can exhibit oscillatory behaviour for some parameter values and stable behaviour for others.

$$\begin{aligned} [R] &= \frac{2k_3k_4}{k_2k_4 - k_1k_3 + k_2k_5 + \sqrt{(k_2k_4 + k_1k_3 + k_2k_5)^2 - 4k_1k_2k_3k_5}}, \\ [F] &= \frac{k_2k_4 + k_1k_3 + k_2k_5 + \sqrt{(k_2k_4 + k_1k_3 + k_2k_5)^2 - 4k_1k_2k_3k_5}}{2k_2k_3}. \end{aligned} \tag{24}$$

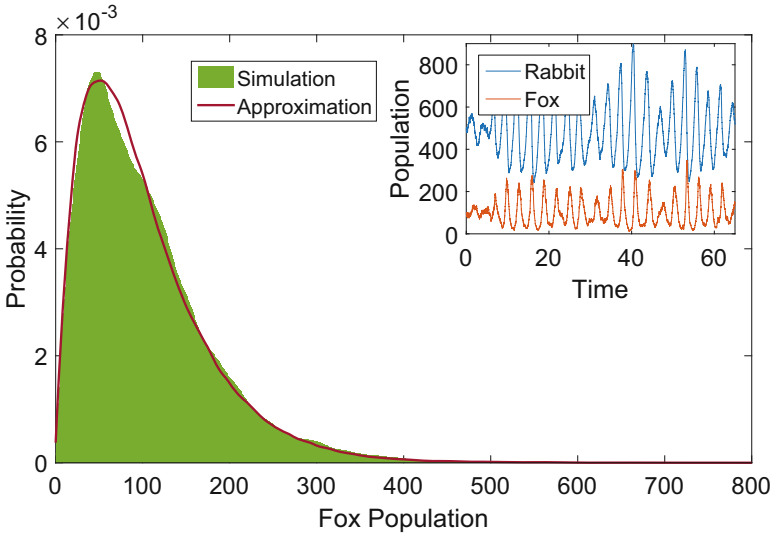

Fig. 6 Probability distribution of the number of foxes, n_F, in the predator–prey model (22) as obtained from the SSA of the full system (histogram) and the analytical approximation corresponding to the reduced system (25) which assumes an abundance of rabbits (*solid line*). Inset: time course data of rabbit (*blue*) and fox (*red*) populations from the SSA of the full system. Parameter values are $k_1 = 1,\ k_2 = 0.01,\ k_3 = 5,\ k_4 = 5,\ k_5 = 10,\ V = 1$

In order to apply the approximation method, we make the reasonable assumption that the concentration of rabbits $[R]$ is large compared to the concentration of foxes $[F]$. Then, we can reduce the reaction network to give:

$$F \xrightarrow{k_2[R]} F + F,\ F \xrightarrow{k_3} \emptyset,\ \emptyset \xrightarrow{k_5} F. \tag{25}$$

The master equation for this effective system of reactions can be solved analytically in steady-state [19] and its solution is a Negative Binomial $\left(\frac{k_5}{k_2[R]}, \frac{k_2[R]}{k_3}\right)$. In Fig. 6 we compare this analytical approximation with the distribution obtained using the SSA of the full system (22), and note that they agree well. In the inset we show a time course of the SSA of the full system for rabbits (blue) and foxes (red). Note that the oscillations are here induced by noise and are not predicted by the deterministic rate equations. It can be shown that the reduced system does not possess noise-induced oscillations; this is because for a one variable system the power spectrum of noise fluctuations cannot exhibit a peak at a non-zero frequency [20]. Hence while the approximation method leads to an excellent approximation for the probability distribution of the full system it cannot always capture other relevant stochastic properties.

4 Conclusion

In this chapter, we have introduced a simple technique of approximating the master equation. This involves reducing the model to one with a smaller number of species interacting via a set of effective reactions. If the number of species in the reduced model is small, then there is a good chance of finding an analytical solution in steady-state, as we have seen for many examples. Even if such an explicit closed-form solution is not possible, stochastic simulation using the SSA of the reduced master equation yields an accurate solution in a time which is typically much shorter than that possible with the SSA of the full master equation.

References

1. N. van Kampen, *Stochastic Processes in Physics and Chemistry*, 3rd edn. (Elsevier, Amsterdam, 2007)
2. C. Gadgil, C. Lee, H. Othmer, A stochastic analysis of first-order reaction networks. Bull. Math. Biol. **67**, 901–946 (2005)
3. D.T. Gillespie, Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. **81**, 2340–2361 (1977)
4. A. Hellander, P. Lötstedt, Hybrid method for the chemical master equation. J. Comput. Phys. **227**, 100–122 (2007)
5. S. Peleš, B. Munsky, M. Khammash, Reduction and solution of the chemical master equation using time scale separation and finite state projection. J. Chem. Phys. **125**, 204104 (2006)
6. P. Thomas, A.V. Straube, R. Grima, The slow-scale linear noise approximation: an accurate, reduced stochastic description of biochemical networks under timescale separation conditions. BMC Syst. Biol. **6**, 39 (2012)
7. T. Jahnke, On reduced models for the chemical master equation. SIAM Multiscale Model. Simul. **9**, 1646–1676 (2011)
8. S. Smith, C. Cianci, R. Grima, Model reduction for stochastic chemical systems with abundant species. J. Chem. Phys. **143**, 214105 (2015)
9. R. Grima, D. Schmidt, T. Newman, Steady-state fluctuations of a genetic feedback loop: an exact solution. J. Chem. Phys. **137**, 035104 (2012)
10. C. Cianci, S. Smith, R. Grima, Molecular finite-size effects in stochastic models of equilibrium chemical systems. J. Chem. Phys. **144**, 084101 (2016)
11. V. Shahrezaei, P.S. Swain, Analytical distributions for stochastic gene expression. Proc. Natl. Acad. Sci. **105**, 17256–17261 (2008)
12. S. Schnell, C. Mendoza, Closed form solution for time-dependent enzyme kinetics. J. Theor. Biol. **187**, 207–212 (1997)
13. R. Grima, Noise-induced breakdown of the Michaelis-Menten equation in steady-state conditions. Phys. Rev. Lett. **102**, 218103 (2009)
14. D. Gonze, J. Halloy, A. Goldbeter, Robustness of circadian rhythms with respect to molecular noise. Proc. Natl. Acad. Sci. **99**, 673–678 (2002)
15. P. Thomas, A.V. Straube, J. Timmer, C. Fleck, R. Grima, Signatures of nonlinearity in single cell noise-induced oscillations. J. Theor. Biol. **335**, 222–234 (2013)
16. Y. Taniguchi, P. Choi, G. Li, H. Chen, M. Babu, J. Hearn, A. Emili, X.S. Xie, Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. Science **329**, 533–538 (2010)

17. J. Ross, I. Schreiber, V. MO, *Determination of Complex Reaction Mechanisms: Analysis of Chemical, Biological, and Genetic Networks* (Oxford University Press, Oxford, 2005)
18. J. Murray, *Mathematical Biology I: An Introduction* (Springer, New York, 2002)
19. S. Smith, V. Shahrezaei, General transient solution of the one-step master equation in one dimension. Phys. Rev. E **91**, 062119 (2015)
20. C. Gardiner, *Handbook of Stochastic Methods* (Springer, Berlin, 1985)