Published in IET Systems Biology
Received on 4th April 2007
Revised on 15th July 2008
doi: 10.1049/iet-syb:20070031


ISSN 1751-8849

# Moment-closure approximations for mass-action models

*C.S. Gillespie*

*School of Mathematics & Statistics, Newcastle University, Newcastle upon Tyne NE1 7RU, UK*
*E-mail: c.gillespie@ncl.ac.uk*

**Abstract:** Although stochastic population models have proved to be a powerful tool in the study of process generating mechanisms across a wide range of disciplines, all too often the associated mathematical development involves nonlinear mathematics, which immediately raises difficult and challenging analytic problems that need to be solved if useful progress is to be made. One approximation that is often employed to estimate the moments of a stochastic process is moment closure. This approximation essentially truncates the moment equations of the stochastic process. A general expression for the marginal- and joint-moment equations for a large class of stochastic population models is presented. The generalisation of the moment equations allows this approximation to be applied easily to a wide range of models. Software is available from http://pysbml.googlecode.com/ to implement the techniques presented here.

## 1 Introduction

As a result of recent advances in experimental techniques, biology has become much more of a quantitative science. The capacity to answer questions ranging from cell and molecular functions, through to population dynamics requires an increasing ability to acquire, store and manipulate large volumes of raw data in a flexible, efficient manner. Moreover, there is a growing realisation that complex biological processes cannot be understood through the application of ever-more reductionist experimental programmes.

There is a developing perception that mathematical modelling may provide some of the necessary tools required to understand this mass of biological data. Indeed, there are distinct advantages of modelling a biological process with the rigour needed to build a mathematical model. First, when constructing a model, gaps in current knowledge are highlighted quantitatively [1, 2]. Even the very process of model specification will highlight important unknowns. Also, when building a model, verbal hypotheses are made specific and conceptually rigorous [3, 4]. Finally, as [5] shows, models can yield quantitative as well as qualitative predictions.

A model can be generally classed as deterministic or stochastic. A deterministic model is one that takes no account of random variation and therefore gives a fixed and precisely reproducible result. Deterministic models are often mathematically described by sets of differential equations and are most (but not always) appropriate when large numbers of individuals of a species are involved and statistical variations in the average behaviour of the system are relatively unimportant.

However, for many biological systems, this assumption may not be valid. A stochastic model should be used when either the number of a particular species is small or when there is reason to expect random events to have an important influence on the behaviour of the system. The essential difference between a stochastic and a deterministic model is that in a stochastic model, different simulation runs can result from the same initial conditions and model parameters (see [6] for a detailed discussion on stochastic models).

A stochastic model, that is, a continuous time Markov process on discrete state space, takes into account every molecule and every reaction. This poses a large computational problem. Typically, to obtain a realisation of a process of interest, one would implement the Gillespie algorithm [7]. This algorithm chooses at random an event to occur (weighted by the hazard function), and then updates the event time. The Gillespie algorithm has the

useful property of generating exact independent realisations from the underlying stochastic model. The next-reaction method (also known as the Gibson–Bruck algorithm) is a modification of a variant of the Gillespie algorithm known as the first-reaction method [8, 9]. This algorithm can increase the simulation speed by efficiently working out event times. However, the speed-up is likely to depend on the precise structure of the model and the speed of the random number generator used. A number of other approximation simulators have been suggested. For example, Gillespie and Petzold [10] divide the time axis into small discrete chunks and updates multiple species at once. However, when the number of molecules are low, problems can arise since mass updating of species can cause the number of species to become negative. Hybrid simulators combine an exact stochastic framework with traditional deterministic ODE methods, that is, by partioning the species into two groups: a group with a low number of individuals to be treated exactly (stochastically) and a group of species that can be treated with an ODE solver [11, 12]. Alternatively, the model could be formulated as a stochastic differential equation, which is considered valid for large numbers of molecules.

This paper considers the moment-closure approximation. This technique constructs a set of ODEs for the model moments, that is, the mean, variance, skewness and so on. Typically for nonlinear models, the ODE for the $i$th moment will depend on higher-order moments, that is, the $(i+1)$th moment. This results in a set of differential equations that may be impossible to solve analytically or numerically. The usual procedure to circumvent this dependency problem involves setting the moments or cumulants above a certain order equal to zero, and solving the remaining coupled set of differential equations either numerically or algebraically (if possible). For instance Goodman and Whittle [13, 14] set all third- and higher-order cumulants to zero, thereby obtaining a normal approximation. In general, moment-closure techniques have been applied with considerable success in a number of different situations. For example, Matis and Kiffe [15] model the spread of the African honey bee through South America via a logistic process, with moment-closure techniques used to estimate moments, and Marion *et al.* [16] model a nematode infection in ruminants.

All applications of the moment-closure approximation have so far only been applied to models with a relatively small number of interacting populations. In this paper, an efficient method of obtaining moment equations up to any order is developed along with software that performs the moment-closure approximation. The results are illustrated by examining a model that contains 23 reactions and 17 chemical species.

## 2 Approach

Typically moment-closure techniques are applied in a model specific manner, that is, for a given model, higher-order moments are calculated [6, 13, 16–18]. Recently, more general frameworks have been proposed, for example, using symbolic mathematics engines [19] or general second-order moment equations [20, 21]. In this section, general multivariate results are developed, which enable moment equations up to any order to be easily calculated.

Suppose, one has $N$ chemical species $\{X_1, \ldots, X_N\}$ and $L$ reactions $\{R_1, \ldots, R_L\}$. Reaction $R_l$ corresponds to

$$\underline{s}_{l1}X_1 + \cdots + \underline{s}_{lN}X_N \xrightarrow{k_l} \bar{s}_{l1}X_1 + \cdots + \bar{s}_{lN}X_n \qquad (1)$$

where $\underline{s}_l$ and $\bar{s}_l$ are the number of reactants and the products in each species involved in reaction $l$. The reaction occurs with rate $k_l$. In this paper, it is assumed that all reaction rates follow the polynomial rate laws (which would include all mass-action kinetic rate laws).

Denote $\boldsymbol{x}$ to be the column vector of molecule numbers, and $\boldsymbol{s}_l = \bar{\boldsymbol{s}}_l - \underline{\boldsymbol{s}}_l$ and $s_{li} = \bar{s}_{li} - \underline{s}_{li}$ to be the stoichiometric coefficients of species $X_i$ in reaction $R_l$. Then, when reaction $R_l$ occurs, $x_i \rightarrow x_i + s_{li}$.

Let $p(\boldsymbol{x})(t) = p(\boldsymbol{x})$ be the probability of being in state $\boldsymbol{x}$ at time $t$, with initial conditions of $\boldsymbol{x}(0)$. The time evolution of $\boldsymbol{x}$ can be formulated as the chemical master equation (or Kolmogorov's forward equation), viz

$$\frac{\mathrm{d}p(\boldsymbol{x})}{\mathrm{d}t} = \sum_{l=1}^{L} p(\boldsymbol{x} - \boldsymbol{s}_l)a_l(\boldsymbol{x} - \boldsymbol{s}_l) - p(\boldsymbol{x})a_l(\boldsymbol{x}) \qquad (2)$$

where $a_l(\boldsymbol{x})$ is the propensity function of reaction $R_l$. $a_{i,\boldsymbol{j}}$ is denoted to be the coefficient of ${x_1}^{j_1} \times \cdots \times {x_N}^{j_N}$ in reaction $i$. For example, if the first propensity function was

$$a_1(x_1, x_2) = \frac{\lambda x_1 x_2(x_1 - 1)}{2}$$

then, $a_{1,(1,1)} = -\lambda/2$, $a_{1,(2,1)} = \lambda/2$, and 0 otherwise.

The $i$th univariate moment is defined as

$$\mu_i(t) = E[X^i] = \sum_{x=0}^{\infty} p(x)(t)x^i$$

for $i = 0, 1, 2, \ldots$. The associated univariate moment generating function is defined as

$$M(t) \equiv \sum_{x=0}^{\infty} p(x)(t)\mathrm{e}^{\theta x} = \mu_0(t) + \frac{\mu_1(t)\theta^1}{1!} + \frac{\mu_2(t)\theta^2}{2!} + \cdots.$$

Likewise, the multivariate moment generating function is

defined as

$$M(t) \equiv \sum_{x_1,\ldots,x_N=0}^{\infty} p(\boldsymbol{x})(t)\mathrm{e}^{\theta_1 x_1+\cdots+\theta_N x_N} = \sum_{\boldsymbol{x}=0}^{\infty} p(\boldsymbol{x})(t)\mathrm{e}^{\boldsymbol{x}\theta} = \sum_{\boldsymbol{x}=0}^{\infty} \frac{\mu_{\boldsymbol{x}}(t)\theta^{\boldsymbol{x}}}{\boldsymbol{x}!}$$

where

$$\boldsymbol{x}! = x_1!x_2!\ldots x_N!$$

and

$$\mu_{\boldsymbol{x}}(t) = E[X_1^{x_1}\ldots X_N^{x_N}] = \sum_{\boldsymbol{x}=0}^{\infty} p(\boldsymbol{x})(t)\boldsymbol{x}^{\boldsymbol{i}}$$

are the moments. On multiplying (2) by $\mathrm{e}^{\boldsymbol{x}\theta}$ and summing over $\boldsymbol{x}$, one obtains

$$\frac{\partial M(t)}{\partial t} = \sum_{l=1}^{L}\sum_{\boldsymbol{i}} a_{l,\boldsymbol{i}}\frac{\partial^{\boldsymbol{i}} M(t)}{\partial\theta^{\boldsymbol{i}}}[\exp(\boldsymbol{s}_l\theta)-1] \quad (3)$$

where

$$\boldsymbol{i} = \{i_1, i_2, \ldots, i_N\} \quad \text{and} \quad \frac{\partial^{\boldsymbol{i}} M(t)}{\partial\theta^{\boldsymbol{i}}} = \frac{\partial^{i_1+\cdots+i_N} M(t)}{\partial\theta_1^{i_1}\ldots\theta_N^{i_N}}$$

On expanding the partial derivative in (3), one obtains

$$\frac{\partial M(t)}{\partial t} = \sum_{l=1}^{L}\sum_{\boldsymbol{i}} a_{l,\boldsymbol{i}}\sum_{\boldsymbol{j}=\boldsymbol{i}}^{\infty}\frac{\theta^{\boldsymbol{j}-\boldsymbol{i}}}{(\boldsymbol{j}-\boldsymbol{i})!}\mu_{\boldsymbol{j}}(t)\left(\left\{\sum_{\boldsymbol{k}=0}^{\infty}\frac{(\boldsymbol{s}_l\theta)^{\boldsymbol{k}}}{\boldsymbol{k}!}\right\}-1\right) \quad (4)$$

where

$$\frac{(\boldsymbol{s}_l\theta)^{\boldsymbol{k}}}{\boldsymbol{k}!} = \frac{(s_{l1}\theta_1)^{k_1}}{k_1!}\times\cdots\times\frac{(s_{lN}\theta_N)^{k_N}}{k_N!}$$

and

$$\mu_{\boldsymbol{j}}(t) = \mu_{j_1,j_2,\ldots,j_N}(t)$$

Extracting the coefficients of $\theta$ from (4) yields

$$\frac{\partial M(t)}{\partial t} = \sum_{\boldsymbol{n}=0}^{\infty}\theta^{\boldsymbol{n}}\sum_{l=1}^{L}\sum_{\boldsymbol{i}} a_{l,\boldsymbol{i}}\sum_{\boldsymbol{k}=0}^{\boldsymbol{n}}\boldsymbol{s}_l^{\boldsymbol{k}}\binom{\boldsymbol{n}}{\boldsymbol{k}}\mu_{\boldsymbol{n}-\boldsymbol{k}+\boldsymbol{i}}(t) \quad (5)$$

where $k_1+\cdots+k_N \neq 0$

$$\boldsymbol{s}_l^{\boldsymbol{k}}\binom{\boldsymbol{n}}{\boldsymbol{k}} = s_{l1}^{k_1}\binom{n_1}{k_1}\times\cdots\times s_{lN}^{k_N}\binom{n_N}{k_N}$$

and

$$\mu_{\boldsymbol{n}-\boldsymbol{k}+\boldsymbol{j}} = \mu_{n_1-k_1+j_1,\ldots,n_N-k_N+j_N}$$

Thus, to obtain an equation for a particular moment, one extracts the corresponding coefficient of $\theta_1, \ldots, \theta_N$ from (5). Hence, (5) is a general result that can be applied to a wide class of stochastic models. In particular, when there is only a single species in the population, expression (5) simplifies to

$$\frac{\partial M(t)}{\partial t} = \sum_{n=1}^{\infty}\theta_1^n\sum_{l=1}^{L}\sum_{i=0}^{\infty} a_{l,i}\sum_{k=1}^{n} s_{l,1}^k\binom{n}{k}\mu_{n-k+i}(t) \quad (6)$$

Thus, the equation for the first moment is

$$\frac{\mathrm{d}\mu_1(t)}{\mathrm{d}t} = \sum_{l=1}^{L}\sum_{i=0}^{\infty} a_{l,i}\, s_{l,1}\,\mu_i(t) \quad (7)$$

where as for the $n$th moment one obtains

$$\frac{\mathrm{d}\mu_n(t)}{\mathrm{d}t} = \sum_{l=1}^{L}\sum_{i=0}^{\infty} a_{l,i}\sum_{k=1}^{n}\binom{n}{k} s_{l,i}^k\mu_{n-k+i}(t) \quad (8)$$

# 3 Computer software

A Python library, has been constructed which applies expression (5) to a model with polynomial rate laws. It reads in a systems biology mark-up language (SBML) file [22] and outputs a Maple™ file with moment equations up to any order. This library can be used as stand alone software or incorporated into other tools, such as pysbml or PySceS [23, 24]. The library can be downloaded from http://pysbml.googlecode.com/ along with other example SBML models and Maple™ output.

Although the library is written in a scripting language, the module is very fast, portable and extendable. For example, a model with 23 reactions and 17 species takes about five seconds to generate a Maple™ file with all first- and second-order differential moment equations. Maple can then solve these differential equations in a few minutes.

# 4 Examples

## 4.1 Immigration-death model

Consider a system that contains the following two reactions

$$\emptyset \xrightarrow{\alpha} X \quad \text{and} \quad X \xrightarrow{\nu} \emptyset \quad (9)$$

This is simply the classic immigration-death process and has the following chemical master equation

$$\frac{\mathrm{d}p_x(t)}{\mathrm{d}t} = \alpha p_{x-1}(t) + \nu(x+1)p_{x+1}(t) - (\alpha+\nu x)p_x(t) \quad (10)$$




doi: 10.1049/iet-syb:20070031

where $p_x(t)$ is the probability of having $x$ molecules at time $t$ (see [25] for an overview of this process). This system has the following propensity functions

$$a_1(x) = \alpha \quad \text{and} \quad a_2(x) = \nu x$$

thus,

$$a_{i,j} = \begin{cases} \alpha & \text{for } i = 1 \text{ and } j = 0 \\ \nu & \text{for } i = 2 \text{ and } j = 1 \\ 0 & \text{otherwise} \end{cases}$$

The associated stoichiometric coefficients are

$$s_{1,1} = 1 \quad \text{and} \quad s_{2,1} = -1$$

From expression (7), the equation for the first moment, that is, the mean, is

$$\begin{aligned} \frac{\mathrm{d}\mu_1(t)}{\mathrm{d}t} &= \sum_{l=1}^{2} \sum_{i=0}^{1} a_{l,i}\, s_{l,1}\, \mu_i(t) \\ &= a_{1,0}s_{1,1} + a_{2,1}s_{2,1}\mu_1(t) \\ &= \alpha - \nu\mu_1(t) \end{aligned} \tag{11}$$

Likewise, on using expression (8), one obtains equations for higher-order moments, namely

$$\begin{aligned} \frac{\mathrm{d}\mu_2(t)}{\mathrm{d}t} &= \alpha[1 + 2\mu_1(t)] + \nu[\mu_1(t) - 2\mu_2(t)] \\ \frac{\mathrm{d}\mu_3(t)}{\mathrm{d}t} &= \alpha[1 + 3\mu_1(t) + 3\mu_2(t)] - \nu[\mu_1(t) \\ &\quad - 3\mu_2(t) + 3\mu_3(t)] \end{aligned} \tag{12}$$

The equivalent deterministic formulation of the immigration-death process, that is, the deterministic equation corresponding to $a_1(x)$ and $a_2(x)$ is

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t} = \alpha - \nu X(t) \tag{13}$$

where $X(t)$ is the population at time $t$. Since the immigration-death model is linear, (13) and (11) are identical.

## 4.2 Dimerisation process

The simple dimerisation process has the following set of relations

$$2X \xrightarrow{k_1} Y \quad \text{and} \quad Y \xrightarrow{k_2} 2X \tag{14}$$

On noting that

$$X + 2Y = X_0 \tag{15}$$

where $X_0$ is the number of molecules of $X$ that would be present if they were fully disassociated, one obtains the propensity functions

$$a_1(x) = \frac{k_1 x(x-1)}{2} \quad \text{and} \quad a_2(x) = \frac{k_2(x_0 - x)}{2} \tag{16}$$

Hence

$$a_{i,j} = \begin{cases} \dfrac{-k_1}{2} & \text{for } i = 1 \text{ and } j = 1 \\ \dfrac{k_1}{2} & \text{for } i = 1 \text{ and } j = 2 \\ \dfrac{k_2 x_0}{2} & \text{for } i = 2 \text{ and } j = 0 \\ \dfrac{-k_2}{2} & \text{for } i = 2 \text{ and } j = 1 \\ 0 & \text{otherwise} \end{cases}$$

The associated stoichiometric coefficients are

$$s_{1,1} = -2 \quad \text{and} \quad s_{2,1} = 2$$

The chemical master equation for the dimerisation process is

$$\begin{aligned} \frac{\mathrm{d}p_x(t)}{\mathrm{d}t} &= p_{x+2}(t)k_1(x+2)(x+1)/2 + p_{x-2}k_2(x_0 - x + 2)/2 \\ &\quad - p_x[k_1x(x-1)/2 + k_2(x_0 - x)/2] \end{aligned} \tag{17}$$

and the corresponding deterministic model is

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t} = -k_1X(t)[X(t) - 1] + k_2[X_0 - X(t)] \tag{18}$$

On using expression (8), one obtains the following moment equations

$$\frac{\mathrm{d}\mu_1(t)}{\mathrm{d}t} = k_1[\mu_1(t) - \mu_2(t)] + k_2[x_0 - \mu_1(t)] \tag{19}$$

$$\begin{aligned} \frac{\mathrm{d}\mu_2(t)}{\mathrm{d}t} &= -2k_1[\mu_1(t) - 2\mu_2(t) + \mu_3(t)] + 2k_2[x_0(\mu_1(t) + 1) \\ &\quad - \mu_1(t) - \mu_2(t)] \end{aligned} \tag{20}$$

$$\begin{aligned} \frac{\mathrm{d}\mu_3(t)}{\mathrm{d}t} &= k_1[4\mu_1(t) - 10\mu_2(t) + 9\mu_3(t) - 3\mu_4(t)] \\ &\quad + k_2[4(x_0 - \mu_1(t)) + 6(x_0\mu_1(t) - \mu_2(t)) \\ &\quad + 3(x_0\mu_2(t) - \mu_3(t))] \end{aligned} \tag{21}$$

Unlike the moment equation in (11), (19) is not closed, that is, the differential equation for $\mu_1(t)$ depends on the $\mu_2(t)$, likewise the ODE for $\mu_2(t)$ depends on $\mu_3(t)$ and so on. This dependency structure occurs whenever there are nonlinear rate laws in the model.

By rewriting (19) in terms of cumulants, one obtains

$$\begin{aligned} \frac{\mathrm{d}\kappa_1(t)}{\mathrm{d}t} &= -k_1\kappa_1(t)[\kappa_1(t) - 1] + k_2[x_0 - \kappa_1(t)] \\ &\quad - k_1\kappa_2(t) \end{aligned} \tag{22}$$

where the first cumulant $\kappa_1(t)$ is the mean, and second cumulant $\kappa_2(t)$ is the variance. On comparing (22) with (18), it is observed that the stochastic equation for the mean has an additional $\kappa_2(t)$ term. Hence, the deterministic equation in this case can be considered as a first-order approximation to the stochastic where all higher-order cumulants are assumed to be zero.

Fig. 1 shows a stochastic realisation of the dimerisation model with an estimated mean and an approximate 95% confidence interval ($\pm 2$ standard deviations). The estimates for the mean and variance were obtained by setting $\mu_3(t) = 3\mu_2(t)\mu_1(t) - 2\mu_1(t)^3$, that is, the normal approximation that corresponds to setting cumulants of order 3 and above to be 0. It is observed that the realisation oscillates around the mean and remains within the uncertainty bands (except briefly at time $t \simeq 6$).

Third- and fourth-order closure schemes for this model were also investigated. In both cases, the estimates for the mean and variance were almost identical to the standard normal approximation (the results can be obtained from the author's website). By investigating higher-order closure in this manner, the sensitivity of this approximation can be assessed, since similar estimates for the mean and variance across closure levels indicate that the underlying distribution may be approximately normal.

## 4.3 Chaperone model

Molecular chaperones have many cellular functions but are often involved in the folding of nascent proteins, re-folding of denatured proteins and prevention of protein aggregation and in assisting the targeting of proteins for degradation by the proteasome and lysosomes. They also have a role in apoptosis and are involved in modulating signals for immune and inflammatory responses. The induction of heat shock proteins is impaired with age and there is also a decline in the chaperone function. Aberrant/damaged proteins accumulate with age and are implicated in several important age-related conditions (e.g. Alzheimer's disease, Parkinson's disease and cataract). Therefore the balance between damaged proteins and available free chaperones may be greatly disturbed during ageing. Here, this framework is applied to such an example that describes the heat shock system and its implications in ageing (see [26] for further details).

The model is coded in SBML and is available to download from the author's website or from the Biomodels website with model id BIOMD0000000091 (see [27] for a description of the Biomodels website). The system contains 23 reactions and 17 chemical species. To simulate the model up to time $t = 2000$ using an efficient Gillespie simulator would take approximately 50 min. Thus, to calculate the estimated values in Fig. 2 would take approximately 69 days when executed using a single machine with a 3.0 GHz processor. There are eight first-order reactions involving a single species, eight first-order reactions involving two species, two second-order reactions and three zeroth-order reactions.

Fig. 2 show a stochastic realisation of the model for species ADP and NatP (native protein). From the realisation, one can observe that there is an initial dip in both populations; however, the ADP population recovers. The other point to note is the difference in scales. ADP is measured in hundreds of molecules, whereas NatP is measured in millions of molecules. If the number of NatP molecules is increased by a factor of ten, this will cause a large increase in simulation time and possibly make exact simulation impossible.

Also shown in Fig. 2 are the moment-closure approximations to the mean and variance of ADP along with the associated exact values (obtained via 2000 Monte Carlo simulations). There is very little difference between the estimated and the exact moments. It is seen that although the mean is fairly constant throughout, the variance increases in the time period (0, 1000), and then

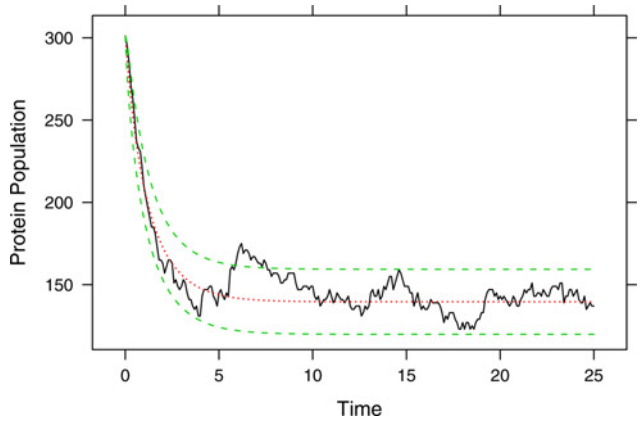


**Figure 1** *Single stochastic realisation of the dimerisation process, where X(0) = 301 and {$k_1$, $k_2$} = {0.00166, 0.2}*

The moment closure approximation for the mean (dotted) and an approximate confidence region of the mean $\pm 2$ standard deviations (dashed) is also shown

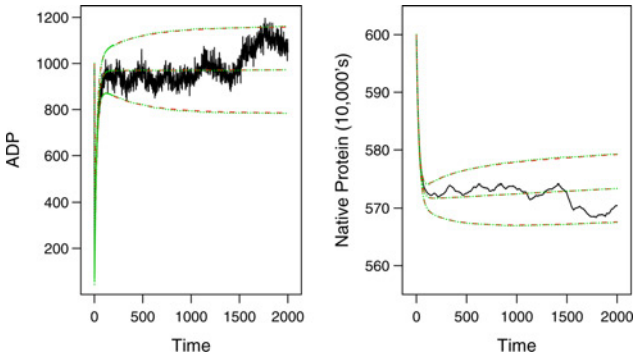


**Figure 2** *Stochastic realisation of the chaperone model (solid)*

The approximate (dashed) and estimated (dotted) mean and $\pm 2$ standard deviations of ADP are also shown
The estimated values were obtained from 2000 realisations of the an exact simulation algorithm




doi: 10.1049/iet-syb:20070031

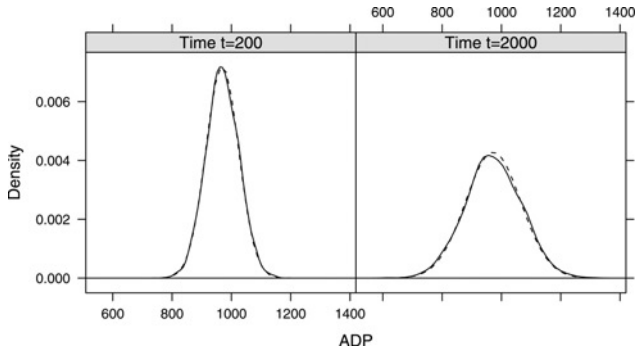


**Figure 3** *Distribution of ADP at times t = 200 and t = 2000*
The filled line is the value obtained by 2000 stochastic realisations
The dashed line is the moment-closure approximation

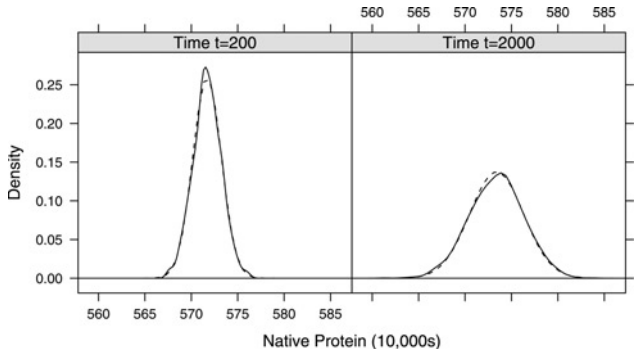


**Figure 4** *Distribution of NatP at times t = 200 and t = 2000*
The filled line is the value obtained by 2000 stochastic realisations
The dashed line is the moment-closure approximation

begins to level off for time $t > 1000$ (although it is still increasing). A similar result is obtained for NatP (shown in Fig. 2).

Whereas the Monte Carlo estimates would take 69 days to run on a single CPU machine, the moment-closure approximation takes a few minutes. Overall, the approximations of the mean and variance are excellent, with the moment-closure approach capturing the initial drop in population at time $t = 10$ (Fig. 2).

It is also possible to approximate the actual distribution at any given time point. Since a normal approximation, has been assumed, it is natural to describe the distribution with a normal distribution. Shown in Fig. 3 is the normal approximation for species ADP compared with the exact distribution (obtained through simulation). Again, the moment-closure approximation seems to capture the key aspects of the distribution. Fig. 4 shows similar results for NatP.

## 5 Discussion and conclusion

Mathematical models can give one a deep insight into real systems; however, with large models, it is difficult to gain an understanding of this behaviour, particularly in a stochastic framework. Nevertheless, stochastic models are essential if one intends to successfully model processes where small fluctuations have a large impact on the system, for example, as in the ageing context [28]. When large stochastic models take a prohibitively long time to simulate a single realisation, then the stochastic approach can lose its appeal. This paper presents a possible solution to explore such large complex models. By developing results for the general multivariate process, a fast approximation to the underlying stochastic process is presented.

Recent developments suggest that one can increase the accuracy of the moment-closure approximation by considering non-zero closure schemes. For example, Krishnarajah *et al.* [29] suggest that a log-normal closure scheme may improve accuracy, whereas Singh and Hespanha [19] suggest a more general moment matching approach for higher-order moments. Either of these techniques can be easily incorporated into the results of this paper.

With the increasing acceptance and usage of standard model formats such as SBML, it is now becoming feasible for modellers to use a variety of tools easily and quickly. This paper presents an approach that will benefit many stochastic modellers and will provide another technique when building large, detailed models.

## 6 Acknowledgment

I wish to thank Prof. Richard Boys, Prof. Darren Wilkinson and Dr. Carole Proctor for comments on the manuscript and helpful discussions. Also, I wish to thank the anonymous referees who have made many useful suggestions.

## 7 References

[1] KOWALD A., KIRKWOOD T.B.L.: 'A network theory of ageing: the interactions of defective mitochondria, aberrant proteins, free radicals and scavengers in the ageing process', *Mutat. Res.*, 1996, **316**, pp. 209–236

[2] SOZOU P.D., KIRKWOOD T.B.L.: 'A stochastic model of cell replicative senescence based on telomere shortening, oxidative stress, and somatic mutations in nuclear and mitochondrial dna', *J. Theor. Biol.*, 2001, **213**, pp. 573–586

[3] BATTOGTOKH D., TYSON J.J.: 'Bifurcation analysis of a model of the budding yeast cell cycle', *Chaos*, 2004, **14**, pp. 653–661

[4] GILLESPIE C.S., PROCTOR C.J., SHANLEY D.P., WILKINSON D.J., BOYS R.J., KIRKWOOD T.B.L.'A mathematical model of ageing in yeast', *J. Theor. Biol.*, 2004, **229**, pp. 189–196

[5] NELSON D.E., IHEKWABA A.E., ELLIOTT M., *ET AL.*: 'Oscillations in nf-κb signaling control the dynamics of gene expression', *Science*, 2004, **306**, pp. 704–708

[6] WILKINSON D.J.: 'Stochastic modelling for systems biology' (Chapman and Hall, CRC Press, 2006)

[7] GILLESPIE D.T.: 'Exact stochastic simulation of coupled chemical reactions', *J. Phys. Chem.*, 1977, **81**, pp. 2340–2361

[8] GILLESPIE D.T.: 'General method for numerically simulating stochastic time evolution of coupled chemical-reactions', *J. Comput. Phy.*, 1976, **22**, pp. 403–434

[9] GIBSON M.A., BRUCK J.: 'Efficient exact stochastic simulation of chemical systems with many species and many channels', *J. Phys. Chem., A.*, 2000, **104**, pp. 1876–1889

[10] GILLESPIE D.T., PETZOLD L.R.: 'Improved leap-size selection for accelerated stochastic simulation', *J. Chem. Phys.*, 2003, **119**, pp. 8229–8234

[11] KIEHL T.R., MATTHEYSES R.M., SIMMONS M.K.: 'Hybrid simulation of cellular behavior', *Bioinformatics*, 2004, **20**, pp. 316–322

[12] PUCHALKA J., KIERZEK A.M.: 'Bridging the gap between stochastic and deterministic regimes in the kinetic simulations of the biochemical reaction networks', *Biophys. J.*, 2004, **86**, pp. 1357–1372

[13] GOODMAN L.A.: 'Population growth of the sexes', *Biometrics*, 1953, **9**, pp. 212–225

[14] WHITTLE P.: 'On the use of the normal approximation in the treatment of stochastic processes', *J. R. Stat. Soc., B.*, 1957, **19**, pp. 481–490

[15] MATIS J.H., KIFFE T.R.: 'On approximating the moments of the equilibrium distributions of a stochastic logistic model', *Biometrics*, 1996, **52**, pp. 980–991

[16] MARION G., RENSHAW E., GIBSON G.: 'Stochastic effects in a model of nematode infection in ruminants', *IMA J. Math. Appl. Med. Biol.*, 1998, **15**, pp. 97–116

[17] GILLESPIE D.T., MANGEL M.: 'Conditioned averages in chemical kinetics', *J. Chem. Phys.*, 1981, **75**, (2), pp. 704–709

[18] KRISHNARAJAH I., MARION G., GIBSON G.: 'Novel bivariate moment-closure approximations', *Math. Biosci.*, 2007, **208**, (2), pp. 621–643

[19] SINGH A., HESPANHA J.P.: 'A derivative matching approach to moment closure for the stochastic logistic model', *Bull. Math. Biol.*, 2007, **69**, pp. 1909–1925

[20] GOMEZ-URIBE C.A., VERGHESE G.C.: 'Mass fluctuation kinetics: capturing stochastic effects in systems of chemical reactions through coupled mean-variance computations', *J. Chem. Phys.*, 2007, **126**, p. 024109–024112

[21] GOUTSIAS J.: 'Classical versus stochastic kinetics modeling of biochemical reaction systems', *Biophys. J.*, 2007, **92**, (7), pp. 2350–2365

[22] HUCKA M., FINNEY A., SAURO H.M., *ET AL.*: 'The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models', *Bioinformatics*, 2003, **19**, pp. 524–531

[23] OLIVIER B.G., ROHWER J.M., HOFMEYR J.-H.S.: 'Modelling cellular systems with PySCeS', *Bioinformatics*, 2005, **21**, pp. 560–561

[24] GILLESPIE C.S., SHANLEY D.P., WILKINSON D.J., BOYS R.J., PROCTOR C.J., KIRKWOOD T.B.L.: 'Tools for the SBML community', *Bioinformatics*, 2006, **22**, pp. 628–629

[25] RENSHAW E.: 'Modelling Biological Populations in Space and Time' (Cambridge University Press, 1991)

[26] PROCTOR C.J., SOTI C., BOYS R.J., *ET AL.*: 'Modelling the actions of chaperones and their role in ageing', *Mech. Ageing Dev.*, 2005, **126**, pp. 119–131

[27] LE NOVERE N., BORNSTEIN B., BROICHER A., *ET AL.*: 'BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems', *Nucleic Acids Res.*, 2006, **34**, (suppl_1), D689–D691

[28] KIRKWOOD T.B.L., BOYS R.J., GILLESPIE C.S., PROCTOR C.J., SHANLEY D.P., WILKINSON D.J.: 'Towards an e-biology of ageing: integrating theory and data', *Nat. Rev. Mole. Cell Bio.*, 2003, **4**, pp. 243–249

[29] KRISHNARAJAH I., COOK A., MARION G., GIBSON G.: 'Novel moment closure approximations in stochastic epidemics', *Bull. Math. Biol.*, 2005, **67**, pp. 855–873



doi: 10.1049/iet-syb:20070031