Stat Comput (2008) 18: 125–135
DOI 10.1007/s11222-007-9043-x

Bayesian inference for a discretely observed stochastic kinetic model

R.J. Boys · D.J. Wilkinson · T.B.L. Kirkwood

Received: 18 March 2005 / Accepted: 5 November 2007 / Published online: 28 November 2007

Abstract The ability to infer parameters of gene regulatory networks is emerging as a key problem in systems biology. The biochemical data are intrinsically stochastic and tend to be observed by means of discrete-time sampling systems, which are often limited in their completeness. In this paper we explore how to make Bayesian inference for the kinetic rate constants of regulatory networks, using the stochastic kinetic Lotka-Volterra system as a model. This simple model describes behaviour typical of many biochemical networks which exhibit auto-regulatory behaviour. Various MCMC algorithms are described and their performance evaluated in several data-poor scenarios. An algorithm based on an approximating process is shown to be particularly efficient.

Keywords Biochemical networks · Block updating · Lotka-Volterra model · Markov jump process · MCMC methods · Parameter estimation · Reversible jump · Systems biology

1 Introduction

A high current priority in the biological sciences is the development of new techniques for integrative and predictive modelling (Bower and Bolouri 2001; Kitano 2001; Kirkwood et al. 2003). This is based both on the realisation that traditional reductionist approaches need to be complemented by efforts to reconstruct an understanding of how systems function as a whole (i.e. "systems biology") and also on the massive amounts of detailed experimental data being produced by high-throughput technologies, such as gene expression micro-arrays. As knowledge of underlying mechanisms has advanced, it has also become increasingly apparent that there is an important stochastic element inherent in cell and molecular processes. Stochastic variation at this level can have significant impacts even on high-level outcomes such as an organism's development and ageing (Finch and Kirkwood 2000).

To date, relatively little work has addressed the implications of the stochastic nature of the gene regulatory networks in terms of modelling and data analysis (McAdams and Arkin 1997; Arkin et al. 1998). The biochemical reactions involved in gene regulation typically involve very low concentrations of key reactants which interact with each other and with DNA (Guptasarma 1995). Stochastic variation arises both from randomness of molecular diffusion and from effects of chance in the combinatorial assembly of transcription factor complexes at DNA control sequences. Experimental evidence (Zlokarnik et al. 1998) confirms that gene expression occurs in abrupt stochastic bursts. Conventional deterministic chemical kinetics fail to describe the development of systems of coupled biochemical reactions correctly when both concentrations of reactants and reaction rates are low (Zheng and Ross 1991). Recognition of the fact that chemical reaction steps occur discretely and at random times is vital.

One of the most important challenges in developing systems-level models of stochastic gene regulatory processes is how to estimate the values of the key rate parameters. Realistic models involve many parameters of biological interest and importance. Experimental technology is improving rapidly, so that (semi-)quantitative high-resolution single-cell data of the type that is most informative for the build-

R.J. Boys (✉) · D.J. Wilkinson · T.B.L. Kirkwood
School of Mathematics and Statistics, University of Newcastle upon Tyne, Claremont Road, Newcastle upon Tyne, NE1 7RU, UK
e-mail: richard.boys@ncl.ac.uk

ing of stochastic models is now realistically attainable (Pepperkok and Ellenberg 2006). Typically, data is generated via fluorescence microscopy, then processed to extract gene expression time series (Shen et al. 2006).

Traditionally, network models have been given a continuous deterministic interpretation leading to a set of coupled ordinary differential equations. The inference problem then becomes one of estimating the kinetic rate parameters and a variety of techniques are possible ranging from *ad hoc* parameter tuning to sophisticated model-based Bayesian methods; see, for example, Brown and Sethna (2003), Barenco et al. (2006) and Liebermeister and Klipp (2005) for the latter. For intracellular processes, it is well known that stochastic effects are important (Bahcall 2005; McAdams and Arkin 1999) and so methods are required which explicitly account for intrinsic stochastic effects. Another important consideration is that the experimental procedures mentioned in the previous paragraph rarely allow for the simultaneous measurement of more than a small number of the key reactants. Therefore the case of how to make inferences using only partial observation of the system is of particular interest.

Until recently, stochastic gene regulatory models have been too complicated for direct inferential analysis and in current work (see Arkin et al. 1998) parameters are set to biologically plausible starting values and then tuned in an ad hoc manner in an attempt to match experimental data. However, progress in Bayesian stochastic-simulation methodology allows, in principle, direct inference to be made for the parameters of any fully specified model, taking account of prior information about parameter values in the form of probability distributions.

A typical stochastic gene regulatory model describes the evolution of u species $Y_1, Y_2, \ldots, Y_u$ (in thermal equilibrium inside some fixed volume) using a set of v reaction equations $R_1, R_2, \ldots, R_v$. Such systems are represented using chemical reaction notation as follows:

$$\begin{aligned}
R_1: \quad & p_{11}Y_1 + p_{12}Y_2 + \cdots + p_{1u}Y_u \\
& \longrightarrow \quad q_{11}Y_1 + q_{12}Y_2 + \cdots + q_{1u}Y_u \\
R_2: \quad & p_{21}Y_1 + p_{22}Y_2 + \cdots + p_{2u}Y_u \\
& \longrightarrow \quad q_{21}Y_1 + q_{22}Y_2 + \cdots + q_{2u}Y_u \\
\vdots \quad & \qquad \vdots \\
R_v: \quad & p_{v1}Y_1 + p_{v2}Y_2 + \cdots + p_{vu}Y_u \\
& \longrightarrow \quad q_{v1}Y_1 + q_{v2}Y_2 + \cdots + q_{vu}Y_u
\end{aligned}$$

where p_{kj} is the stoichiometry associated with reactant j in reaction k and q_{kj} is the stoichiometry associated with product j in reaction k. Each reaction R_k has a stochastic rate constant θ_k and a rate law $h_k(Y, \theta_k)$, where $Y = (Y_1, Y_2, \ldots, Y_u)$ is the current state of the system measured in numbers of molecules. The rate law describes the instantaneous hazard of reaction R_k occurring under an assumption of mass action kinetics (Gillespie 1977). The effect of reaction R_k is to change the value of each Y_j by $q_{kj} - p_{kj}$. A consequence of the model is that, at time t, the time to the next reaction has an exponential distribution with rate $h_0(Y, \theta) = \sum_{k=1}^{v} h_k(Y, \theta_k)$, and the reaction is of type k with probability $h_k(Y, \theta_k)/h_0(Y, \theta)$. Hence, the process is easily simulated using discrete event simulation methods. Within the chemical kinetics literature, this technique is known as the Gillespie algorithm (Gillespie 1977); see Wilkinson (2006) for further details of stochastic kinetic modelling and its application to systems biology.

A naive approach to parameter inference in this context would be to work with a deterministic approximation to the stochastic model. Parameter estimates can then be obtained by using standard least squares or maximum likelihood approaches. However, Tian et al. (2007) show that this strategy does not work well in general. In this paper, we describe a systematic attempt to conduct rigorous inference for a partially and discretely observed stochastic kinetic model. There have been several attempts in the recent literature to tackle this problem. Reinker et al. (2006) assumed full observation of the system at discrete times but the applicability of their methods are limited due to the extent to which non-Bayesian methods can cope with hidden data. In particular, the parsimony assumptions that they use have the effect of downward-biasing of parameter estimates. Rempala et al. (2006) study a model for gene transcription containing two species. They assume data are obtained with the process in steady state and use the tractability of the steady state distribution for their model to integrate out the unobserved specie. Using these simplifications they develop a Bayesian inference algorithm for the rate constants in their model. The applicability of these techniques are somewhat limited and could not be applied to the non-linear models typically of interest in systems biology (including the model considered in this paper). Golightly and Wilkinson (2006, 2008) develop two very general inference algorithms based on a diffusion approximation to the true discrete stochastic model. Although their approximation captures the intrinsic stochastic variation and they have shown that their method works well for many problems, it nevertheless ignores the discreteness of the underlying process which can be important in low copy number scenarios. This point is highlighted by Tian et al. (2007), who develop an algorithm based on the exact process but which requires observation of all species within the system. Their procedure evaluates the likelihood function and uses a genetic algorithm to search for the maximum likelihood estimate. In contrast, the method developed here provides a fully Bayesian solution to the parameter estimation problem for the exact model in the context in which

not all species are observed. Furthermore, our procedure can be extended in a straightforward way to allow for experimental error in the observation process.

To illustrate the methodology, the system we use is of Lotka-Volterra type (Lotka 1925; Volterra 1926), and describes the time evolution of two species, usually called predator and prey (Renshaw 1991). Although not explicitly a gene-regulatory model, the "species" represented in the model could equally well be molecular species and the model serves to gain insight into how inference might be done in more realistic and complex networks. The Lotka-Volterra system is a basic stochastic process which is sufficiently complex to explore the behaviour typical of many biochemical networks that exhibit auto-regulatory behaviour; see Gillespie (1977) for further background on stochastic and deterministic versions of the Lotka-Volterra model and its chemical kinetic interpretation. Jost and Arditi (2000) describe methods for estimating parameters for predator-prey models from time series but these are not appropriate for sparse observation on models with significant amounts of intrinsic noise. This case is considered by Gilioli et al. (2008) and they describe procedures similar in spirit to those develop by Golightly and Wilkinson (2006) and suffer from the potential disadvantages described previously.

The main contribution of this paper is to show how exact inferences can be made despite the discreteness and partial nature of the data. Section 2 examines how inferences can be made for the kinetic rate constants by using a complete data trace and then, in Sect. 3, we show how this can be achieved when only discretely observed time course data are available. Experimental results often measure only part of the discretised data, for example, by missing some or all of one of the chemical "species". In Sect. 4, we describe how inferences can be obtained in various data-poor scenarios and in Sect. 5, the methods are illustrated using simulated data.

2 Inferential issues for the general model

2.1 Analysis using complete data

Suppose the entire process $\boldsymbol{y}$ is observed over the interval $[0, T]$ and that the ith unit interval $(i, i+1]$ contains $n_i = \sum_{k=1}^{v} r_{ki}$ reactions with reaction times and types (t_{ij}, k_{ij}), $j = 1, 2, \ldots, n_i$, that is, reaction $R_{k_{ij}}$ occurs at time t_{ij}. The likelihood function for the parameters θ is

$$\pi(\boldsymbol{y}|\theta) = \left\{ \prod_{i=0}^{T-1} \prod_{j=1}^{n_i} h_{k_{ij}} \left\{ y(t_{i,j-1}), \theta_{k_{ij}} \right\} \right\} \times \exp\left\{ -\int_0^T h_0\{y(t), \theta\}\, dt \right\}, \tag{1}$$

where $t_{i0} \equiv i$ (Wilkinson 2006). In the case of mass-action kinetic rate laws typically used in this area, the hazard function can be written as

$$h_k(Y, \theta_k) = \theta_k\, g_k(Y), \quad k = 1, 2, \ldots, v. \tag{2}$$

This leads to a convenient factorisation of the likelihood function which in turn permits a conjugate choice of prior distribution for the rate constants *viz.* independent gamma components

$$\theta_k \sim \Gamma(a_k, b_k), \quad k = 1, 2, \ldots, v. \tag{3}$$

Application of Bayes Theorem produces a posterior distribution which retains parameter independence, with for $k = 1, 2, \ldots, v$

$$\theta_k|\boldsymbol{y} \sim \Gamma\left(a_k + r_k,\ b_k + \int_0^T g_k\{y(t)\}\, dt \right), \tag{4}$$

where r_k is the total number of type k reactions occurring in $(0, T]$. Note that the integrals here are simply finite sums as the integrands are piecewise constant functions. Thus, with complete data, parameter inference is straightforward.

2.2 Analysis using discrete data

Experimentally it is not feasible to observe the times and types of all reactions. However, it is often possible to observe the levels of the species at a discrete number of time points. We shall assume that data are observed on a regular grid and scale time so that the data are

$$y = \left\{ y(t) = \left(y_1(t), y_2(t), \ldots, y_u(t)\right)' : t = 0, 1, 2, \ldots, T \right\}.$$

Generalisations to a non-regular grid are straightforward but not considered in this paper.

Assuming the independent gamma prior specification for the rate constants in (3), the posterior distribution for $\theta = (\theta_1, \theta_2, \ldots, \theta_u)'$ given the discrete data y can be determined using an MCMC scheme with two blocks. One block simulates the entire process $\boldsymbol{y}(0, T]$ conditional on the parameters θ and the observed data y, and the other block simulates the parameters given the entire process as in (4).

The probability law for the entire latent process $\boldsymbol{y}(0, T]$ conditional on the observations can be expressed as

$$\pi(\boldsymbol{y}|y, \theta) = \prod_{i=0}^{T-1} \frac{\pi\{\boldsymbol{y}(i, i+1]|y(i), \theta\}}{\pi\{y(i+1)|y(i), \theta\}}, \tag{5}$$

where

$$\boldsymbol{y}(i, i+1] = \{y(t) : t \in (i, i+1]\}$$

denotes the latent process in interval i $(i = 0, 1, \ldots, T-1)$ and $\pi\{y(i+1)|y(i)\}$ represents the conditional distribution

of population levels at the end of an interval conditional on those at the start of the interval. This factorisation shows that, given the population sizes at the interval boundaries y and the reaction rates θ, the latent process can be broken down into a collection of independent intervals. Thus, the problem of simulating an entire latent process can be simplified into one of simulating each interval in turn from $\boldsymbol{y}(i, i+1]|\theta, y(i), y(i+1)$. The following section describes two ways in which this can be achieved in the context of a simple but analytically intractable example.

3 The Lotka-Volterra model

The stochastic kinetic Lotka-Volterra model describes the evolution of two species Y_1 (prey) and Y_2 (predator) using three reaction equations:

$$\begin{aligned} Y_1 &\xrightarrow{\theta_1} 2Y_1 && \text{prey reproduction,}\\ Y_1 + Y_2 &\xrightarrow{\theta_2} 2Y_2 && \text{predator reproduction,}\\ Y_2 &\xrightarrow{\theta_3} \emptyset && \text{predator death.} \end{aligned}$$

More conventionally we can express the probabilistic laws governing the time evolution of the process as: in a (small) interval $(t, t+dt]$, the process evolves according to

$$\begin{aligned} Pr\{Y_1(t+dt) &= y_1(t)+1, Y_2(t+dt) = y_2(t)|y_1(t), y_2(t)\}\\ &= \theta_1 y_1(t)\,dt + o(dt),\\ Pr\{Y_1(t+dt) &= y_1(t)-1, Y_2(t+dt)\\ &= y_2(t)+1|y_1(t), y_2(t)\}\\ &= \theta_2 y_1(t) y_2(t)\,dt + o(dt),\\ Pr\{Y_1(t+dt) &= y_1(t), Y_2(t+dt) = y_2(t)-1|y_1(t), y_2(t)\}\\ &= \theta_3 y_2(t)\,dt + o(dt). \end{aligned}$$

Thus the Lotka-Volterra model is a Markov jump process in which each reaction occurs at a particular rate that depends on the current state of the system. The three possible reactions (*reaction types* 1, 2 and 3) have mass-action reaction rates at time t described by (2), where

$$g_1(t) = y_1(t), \qquad g_2(t) = y_1(t)\,y_2(t), \qquad g_3(t) = y_2(t). \tag{6}$$

Therefore, when the entire process $\boldsymbol{y}$ is observed over the interval $[0, T]$ and we take a prior distribution with independent gamma components (3) for the rate constants, the posterior distribution also has independent components, with

$$\begin{aligned} \theta_1|\boldsymbol{y} &\sim \Gamma\left(a_1 + r_1,\ b_1 + \int_0^T y_1(t)\,dt\right),\\ \theta_2|\boldsymbol{y} &\sim \Gamma\left(a_2 + r_2,\ b_2 + \int_0^T y_1(t)y_2(t)\,dt\right),\\ \theta_3|\boldsymbol{y} &\sim \Gamma\left(a_3 + r_3,\ b_3 + \int_0^T y_2(t)\,dt\right), \end{aligned} \tag{7}$$

where r_k is the total number of type k reactions occurring in $(0, T]$.

The problem of determining the posterior distribution when the data are discretely observed as

$$y = \left\{y(t) = \big(y_1(t), y_2(t)\big)' : t = 0, 1, 2, \ldots, T\right\}$$

rests on how to simulate a latent process within each interval from $\boldsymbol{y}(i, i+1]|\theta, y(i), y(i+1)$.

3.1 Reversible jump method

A complicating feature of simulating the latent process in a particular interval is that not only are the times and types of reaction not known but neither is the total number of reactions that have taken place. A standard way of addressing this problem is to use a reversible jump algorithm (Green 2003) which proposes small changes to the latent process; see, for example, Gibson and Renshaw (1998) and Boys and Giles (2007). Suppose that in interval i there are r_{ki} reactions of type k. These reaction counts are non-negative and must satisfy the population counts of predator and prey at the ends of the interval. This latter requirement imposes two constraints on the three reaction counts and so they may be decomposed as

$$r_{ki} = s_i + s_{ki}, \tag{8}$$

where s_{ki} is the (known) minimal number of type k reactions that must have taken place, and s_i is the (unknown) number of occurrences of all three types of reaction ("triples"). These triples arise because a combination of all three reaction types (in any order) produces no net change in population levels. Determination of the s_{ki} is an integer linear programming problem. However, in this simple case, it is straightforward to write out the solution explicitly *viz.*

$$\begin{aligned} s'_{1i} &= \max\{y_1(i+1) - y_1(i), 0\},\\ s'_{2i} &= \max\{y_1(i) - y_1(i+1), 0\},\\ s'_{3i} &= y_2(i) - y_2(i+1) + s'_{2i};\\ \text{if } s'_{3i} > 0 \text{ then } & s_{1i} = s'_{1i}, s_{2i} = s'_{2i}, s_{3i} = s'_{3i}\\ \text{else } & s_{1i} = s'_{1i} - s'_{3i}, s_{2i} = s'_{2i} - s'_{3i}, s_{3i} = 0. \end{aligned}$$

The reversible jump scheme operates on the number of triples in each interval and consists of three move types:

Birth move: with probability b, a new triple (reaction types 1, 2 and 3) is added with independent reaction times taken uniformly within the interval

Death move: with probability d, a randomly selected triple is deleted within the interval

Shift move: with probability $1-b-d$, a randomly selected reaction is shifted within the interval.

The likelihood function for the latent process $\boldsymbol{y}(i, i+1]$ is

$$\pi\{\boldsymbol{y}|y(i),\theta\} = \left\{\prod_{j=1}^{n_i} h_{k_{ij}}\left\{y(t_{i,j-1}),\theta_{k_{ij}}\right\}\right\} \times \exp\left\{-\int_i^{i+1} h_0\{y(t),\theta\}\,dt\right\}, \tag{9}$$

where $t_{i0} \equiv i$ and the hazards $h_k(Y,\theta)$ are as in (2) and (6). Thus the Metropolis-Hastings acceptance probability for each move that generates proposal $\tilde{\boldsymbol{y}}(i, i+1]$, takes the form $\min(1, A_M)$, $M \in \{B, D, S\}$, where

$$A_B = LR \times \frac{d}{(r_{1i}+1)(r_{2i}+1)(r_{3i}+1)b},$$

$$A_D = LR \times \frac{r_{1i}r_{2i}r_{3i}b}{d}, \qquad A_S = LR,$$

and $LR = \pi\{\tilde{\boldsymbol{y}}|y(i),\theta\}/\pi\{\boldsymbol{y}|y(i),\theta\}$. Note that the simplicity of these acceptance probabilities is due in part to the cancellation of the (complicated) conditional distributions $\pi\{y(i+1)|y(i),\theta\}$ in the acceptance ratio.

3.2 Block updating method

Reversible jump methods can be very inefficient in moving around the state space. An alternative and potentially more efficient strategy is to propose an entire new interval from a closely related process. Such block updating strategies have been shown to be effective in a wide variety of latent process models; see, for example, Shephard and Pitt (1997), Liechty and Roberts (2001), Blackwell (2003) and Wilkinson and Yeung (2004). Here we use a block updating strategy which uses a random walk proposal on the number of type 1 reactions and then proposes reactions times using Poisson process approximations to the reaction processes.

Suppose a new value $\tilde{r}_{1i}$ is proposed for the number of type 1 reactions from $f(\tilde{r}_{1i}|r_{1i})$. This value, together with the population sizes at the ends of the interval, then determines the numbers of type 2 and 3 reactions ($\tilde{r}_{2i}$ and $\tilde{r}_{3i}$) in the proposed new interval. One choice of proposal is a (symmetric) discrete random walk in which the current value is augmented by the difference (y) between two independent Poisson random variables with same mean λ, with probability function $p(y) = e^{-2\lambda} I_y(2\lambda)$, where $I_\nu(\cdot)$ is a regular modified Bessel function of order ν (Johnson and Kotz 1969; Abramowitz and Stegun 1984).

Proposals for the reaction times are made by approximating the reaction process for each reaction type with independent inhomogeneous Poisson processes whose rates $\lambda_{ki}(t)$ vary linearly from the initial hazard to the final hazard for the interval, that is, for $k = 1, 2, 3$

$$\lambda_{ki}(t) = (i+1-t)h_k\{y(i),\theta_k\} + (t-i)h_k\{y(i+1),\theta_k\}, \quad t \in [i, i+1].$$

These processes can easily be simulated (conditional on the number of reactions $\tilde{r}_{ki}$) using a homogeneous Poisson process with mean rate $\{h_k\{y(i),\theta_k\} + h_k\{y(i+1),\theta_k\}\}/2$ and then, for $h_k\{y(i+1),\theta_k\} \neq h_k\{y(i),\theta_k\}$, re-scaling time with

$$t = i + \frac{\sqrt{h_k\{y(i),\theta_k\}^2 + [h_k\{y(i+1),\theta_k\}^2 - h_k\{y(i),\theta_k\}^2](t'-i)} - h_k\{y(i),\theta_k\}}{h_k\{y(i+1),\theta_k\} - h_k\{y(i),\theta_k\}}.$$

Clearly a proposal based on the homogeneous process could be used but the additional complexity associated with the linear inhomogeneous process is minor and leads to improved mixing of the algorithm. Finally, a new proposal for the latent process $\tilde{\boldsymbol{y}}(i, i+1]$ is obtained by combining the events in the three reaction processes.

The use of Poisson process approximations and random walk move in generating the proposal can be corrected for via a Metropolis-Hastings step. Let $\mathbb{Q}\{\boldsymbol{y}|y(i), r_{1i}, r_{2i}, r_{3i}\}$ denote the law of the bivariate stochastic process producing the proposed new interval and $\mathbb{P}\{\boldsymbol{y}|y(i), y(i+1)\}$ denote the true "target" process. Then the new interval is accepted with probability $\min(1, A)$, where

$$A = \frac{\frac{d\mathbb{P}}{d\mathbb{Q}}\{\tilde{\boldsymbol{y}}|y(i)\}}{\frac{d\mathbb{P}}{d\mathbb{Q}}\{\boldsymbol{y}|y(i)\}} \times \frac{f(r_{1i}|\tilde{r}_{1i})p(\tilde{r}_{1i})p(\tilde{r}_{2i})p(\tilde{r}_{3i})}{f(\tilde{r}_{1i}|r_{1i})p(r_{1i})p(r_{2i})p(r_{3i})}, \tag{10}$$

$p(r_{ki})$ is the probability function of a Poisson random variable with mean $\{h_k\{y(i),\theta_k\} + h_k\{y(i+1),\theta_k\}\}/2$ (the distribution of r_{ki} under $\mathbb{Q}$), and the Radon-Nikodym derivative

of the true process with respect to the approximate process is

$$\frac{d\mathbb{P}}{d\mathbb{Q}}\{\boldsymbol{y}|y(i)\} = \left\{\prod_{j=1}^{n_i} \frac{h_{k_{ij}}\{y(t_{i,j-1}), \theta_{k_{ij}}\}}{\lambda_{k_{ij},i}(t_{ij})}\right\} \times \exp\left\{\frac{h_0\{y(i), \theta\} + h_0\{y(i+1), \theta\}}{2} - \int_i^{i+1} h_0\{y(t), \theta\}\, dt\right\}.$$

Note that this derivative is simply a likelihood ratio and measures the accuracy of the linear approximation to the true rate processes.

Finally, note that a quicker (but approximate) version of the block updating method can be obtained by not correcting for the use of approximate processes in the derivation of the proposal. This has the effect of dropping the Radon-Nikodym derivatives in (10) and, as the reaction times are no longer needed, the update is obtained with fewer operations. Of course, the drawback is that the posterior distribution for θ obtained from the resulting MCMC scheme is only approximate. However, as demonstrated in Sect. 5, this approximate posterior distribution may well be accurate enough to enable correct inferences to be made about the rate parameters.

4 Partially observed data

Suppose now that only prey can be observed at each of the observation time points because the number of predators is hard to measure. Thus the data on which to base inferences are

$$y_1 = \{y_1(t) : t = 0, 1, 2, \ldots, T\}.$$

One question of interest is whether it is still possible to make inferences for all three reaction rates (and also the predator numbers), that is, whether the model becomes unidentifiable. This partially observed case requires the additional specification of a marginal model for the initial number of predators; we denote this (prior) distribution by $\pi\{y_2(0)\}$.

A MCMC scheme to simulate the posterior distribution $\pi(\boldsymbol{y}[0, T], \theta|y_1)$ can be constructed in a similar fashion to the previous algorithm but instead of updating single intervals of the latent process one at a time, the scheme updates intervals in pairs. In each pair update, the numbers of prey and predator are fixed at the two ends, but in the middle, the number of prey is fixed and the number of predators allowed to vary. Thus, the entire latent process is updated by simulating from $\boldsymbol{y}(i, i+2]|\theta, y(i), y_1(i+1), y(i+2)$ for $i = 0, 1, \ldots, T-2$. Additional moves are also used to update the number of predators at the beginning and end of the process, that is, $y_2(0)$ and $y_2(T)$. Note that our MCMC scheme has the unusual feature of updating overlapping blocks. This strategy has a sound theoretical basis (Carter and Kohn 1996) and is used here to ensure that the unobserved predator levels are all updated at each iteration of the MCMC scheme.

4.1 Reversible jump method

This strategy requires few alterations to the reversible jump method described earlier. The main change is that, in the birth, death and shift moves, reactions can be added and deleted anywhere within both unit intervals. However, only new paired intervals with the correct number of prey at the middle, $y_1(i+1)$, are accepted. Note that a change in the number of predators at the middle of the paired intervals may necessitate a change in the triple counts s_i and s_{i+1} (in (8)) due to a change in s_{ki} and $s_{k,i+1}$, the (recalculated) minimal number of type k reactions that must have taken place (consistent with numbers of predator and prey at the ends of the intervals). The Metropolis-Hastings acceptance probability for each move type which generates proposal $\tilde{\boldsymbol{y}}(i, i+2]$ takes the form $\min\{1, A_M \times \delta_{\tilde{y}_1(i+1), y_1(i+1)}\}$, $M \in \{B, D, S\}$, where $\delta_{\cdot\cdot}$ is Kronecker's delta function,

$$A_B = LR \times \frac{8d}{(r_{1i} + r_{1,i+1} + 1)(r_{2i} + r_{2,i+1} + 1)(r_{3i} + r_{3,i+1} + 1)b},$$

$$A_D = LR \times \frac{(r_{1i} + r_{1,i+1})(r_{2i} + r_{2,i+1})(r_{3i} + r_{3,i+1})b}{8d}, \qquad A_S = LR,$$

and

$$LR = \frac{\pi\{\tilde{\boldsymbol{y}}(i, i+1]|y(i), \theta\}\pi\{\tilde{\boldsymbol{y}}(i+1, i+2]|y_1(i+1), \tilde{y}_2(i+1), \theta\}}{\pi\{\boldsymbol{y}(i, i+1]|y(i), \theta\}\pi\{\boldsymbol{y}(i+1, i+2]|y(i+1), \theta\}},$$

with $\pi(\cdot|\cdot)$ as in (9). The reversible jump method for dealing with the uncertainty about $y_2(0)$ and $y_2(T)$ again uses the birth, death and shift move types but now these are applied only to type 3 reactions within the unit intervals at each end, as they produce no net change in the number of prey and a single (unit) change in the number of predators. The acceptance probability for a proposed new end interval $\tilde{\boldsymbol{y}}(T-1,T]$ takes the form $\min\{1, A_M\}$, $M \in \{B, D, S\}$, where

$$A_B = LR \times \frac{d}{(r_{3i}+1)b}, \qquad A_D = LR \times \frac{r_{3i}b}{d},$$
$$A_S = LR,$$

and $LR = \pi\{\tilde{\boldsymbol{y}}|y(T-1),\theta\}/\pi\{\boldsymbol{y}|y(T-1),\theta\}$. The equivalent calculations for a proposed new first interval $\tilde{\boldsymbol{y}}[0,1)$ are very similar but with an adjustment for the distribution of the initial number of predators, that is, taking $LR = \pi\{\tilde{\boldsymbol{y}}|y_1(0),\tilde{y}_2(0),\theta\}\pi\{\tilde{y}_2(0)\}/[\pi\{\boldsymbol{y}|y(0),\theta\}\pi\{y_2(0)\}]$. As before, any accepted proposed new interval may require an alteration to the triple counts in these intervals.

4.2 Block updating method

The block updating scheme for pairs of intervals again uses random walk proposals $f(\tilde{r}_{ki}|r_{ki})$. In the first interval of each interval pair, new values $\tilde{r}_{1i}$ and $\tilde{r}_{3i}$ are proposed for the numbers of type 1 and type 3 reactions. These together with the number of prey at the ends of the interval and the number of predators at the start of the interval determines the number of type 2 reactions ($\tilde{r}_{2i}$) in the proposed new interval. In the second interval, the reaction counts are perturbed by proposing a change to the number of type 1 reactions ($\tilde{r}_{1,i+1}$). The numbers of types 2 and 3 reactions ($\tilde{r}_{2,i+1}$ and $\tilde{r}_{3,i+1}$) are then fixed. Full realisations of the intervals are proposed by simulating the first and second interval in each pair from the approximating processes $\mathbb{Q}\{\boldsymbol{y}|y(i),\tilde{r}_{1i},\tilde{r}_{2i},\tilde{r}_{3i}\}$ and $\mathbb{Q}\{\boldsymbol{y}|y_1(i+1),\tilde{y}_2(i+1),\tilde{r}_{1,i+1},\tilde{r}_{2,i+1},\tilde{r}_{3,i+1}\}$ respectively, as described in Sect. 3.2. Finally, the proposed replacements $\tilde{\boldsymbol{y}}_1$ and $\tilde{\boldsymbol{y}}_2$ to the interval pair $\boldsymbol{y}_1 = \boldsymbol{y}(i,i+1]$ and $\boldsymbol{y}_2 = \boldsymbol{y}(i+1,i+2]$ are jointly accepted with probability $\min(1, A)$, where

$$A = \frac{\frac{d\mathbb{P}}{d\mathbb{Q}}\{\tilde{\boldsymbol{y}}_1|y(i)\}}{\frac{d\mathbb{P}}{d\mathbb{Q}}\{\boldsymbol{y}_1|y(i)\}} \times \frac{f(r_{1i}|\tilde{r}_{1i})f(r_{3i}|\tilde{r}_{3i})p(\tilde{r}_{1i})p(\tilde{r}_{2i})p(\tilde{r}_{3i})}{f(\tilde{r}_{1i}|r_{1i})f(\tilde{r}_{3i}|r_{3i})p(r_{1i})p(r_{2i})p(r_{3i})}$$
$$\times \frac{\frac{d\mathbb{P}}{d\mathbb{Q}}\{\tilde{\boldsymbol{y}}_2|y_1(i+1),\tilde{y}_2(i+1)\}}{\frac{d\mathbb{P}}{d\mathbb{Q}}\{\boldsymbol{y}_2|y(i+1)\}}$$
$$\times \frac{f(r_{1,i+1}|\tilde{r}_{1,i+1})p(\tilde{r}_{1,i+1})p(\tilde{r}_{2,i+1})p(\tilde{r}_{3,i+1})}{f(\tilde{r}_{1,i+1}|r_{1,i+1})p(r_{1,i+1})p(r_{2,i+1})p(r_{3,i+1})}, \quad (11)$$

and the mass functions $p(r_{ki})$ and the Radon-Nikodym derivatives are as defined in Sect. 3.2. Note that the mass functions $p(\tilde{r}_{k,i+1})$ are determined conditional on the proposed number of predators $\tilde{y}_2(i+1)$ at the centre of the paired intervals.

Random walk proposals can also be used to make adjustments to the end unit intervals in a similar vein. A new proposal $\tilde{\boldsymbol{y}}_1$ for the final interval $\boldsymbol{y}(T-1,T]$ is determined using the technique described above to construct the first of the interval pairs, and is accepted with a Metropolis-Hastings probability calculated using the first line in (11) with $i = T-1$. A new initial interval $\tilde{\boldsymbol{y}}[0,1)$ is constructed by using a random walk proposal distribution $f\{\tilde{y}_2(0)|y_2(0)\}$ to perturb the initial number of predators and a random walk innovation $\tilde{r}_{1,0}$ for the number of type 1 reactions in the interval. These, together with the numbers of prey at the interval end-points and the number of predator at the end of the interval, determine the other reaction counts $\tilde{r}_{2,0}$ and $\tilde{r}_{3,0}$. A proposed new interval is simulated from $\mathbb{Q}\{\boldsymbol{y}|y_1(0),\tilde{y}_2(0),\tilde{r}_{1,0},\tilde{r}_{2,0},\tilde{r}_{3,0}\}$ and accepted with probability $\min(1, A)$, where

$$A = \frac{\frac{d\mathbb{P}}{d\mathbb{Q}}\{\tilde{\boldsymbol{y}}|y_1(0),\tilde{y}_2(0)\}}{\frac{d\mathbb{P}}{d\mathbb{Q}}\{\boldsymbol{y}|y(0)\}} \times \frac{f\{y_2(0)|\tilde{y}_2(0)\}\pi\{\tilde{y}_2(0)\}}{f\{\tilde{y}_2(0)|y_2(0)\}\pi\{y_2(0)\}}$$
$$\times \frac{f(r_{1,0}|\tilde{r}_{1,0})p(\tilde{r}_{1,0})p(\tilde{r}_{2,0})p(\tilde{r}_{3,0})}{f(\tilde{r}_{1,0}|r_{1,0})p(r_{1,0})p(r_{2,0})p(r_{3,0})}.$$

5 Analysis of simulated data

We illustrate the method using data simulated from a Lotka-Volterra process with rate constants $\theta_1 = 0.5$, $\theta_2 = 0.0025$ and $\theta_3 = 0.3$ and initial population values of $y_1(0) = 71$ prey and $y_2(0) = 79$ predators. These initial values are those obtained after running the process for a short time from some arbitrarily chosen population levels. In order to assess the extent to which the data dominate the prior in this example, we take weakly informative independent exponential prior distributions with mean 100 for the rate constants.

We present here results using data from n unit intervals, each contributing m observations on a regular grid giving a further $T = n \times m$ data points in addition to the initial values. We begin by considering the $(n = 40, m = 1)$ case in which the process is observed at the end of 40 subsequent unit intervals. These data are shown in Fig. 1 and clearly display the oscillatory and interaction patterns between the prey and predator populations typical of this process. Standard diagnostics were used to assess the convergence of the MCMC algorithms. Thinning of the MCMC output was employed to yield a posterior sample of size 20,000 with low autocorrelations. The reversible jump sampler (rj) with move probabilities $b = d = 0.3$ required a burn-in of 50,000 iterations and a thin of 1,000 iterates. In contrast, both the block updating scheme (bu) and its approximation (a) needed only 500 iterations to burn-in and a further thin of 10. These

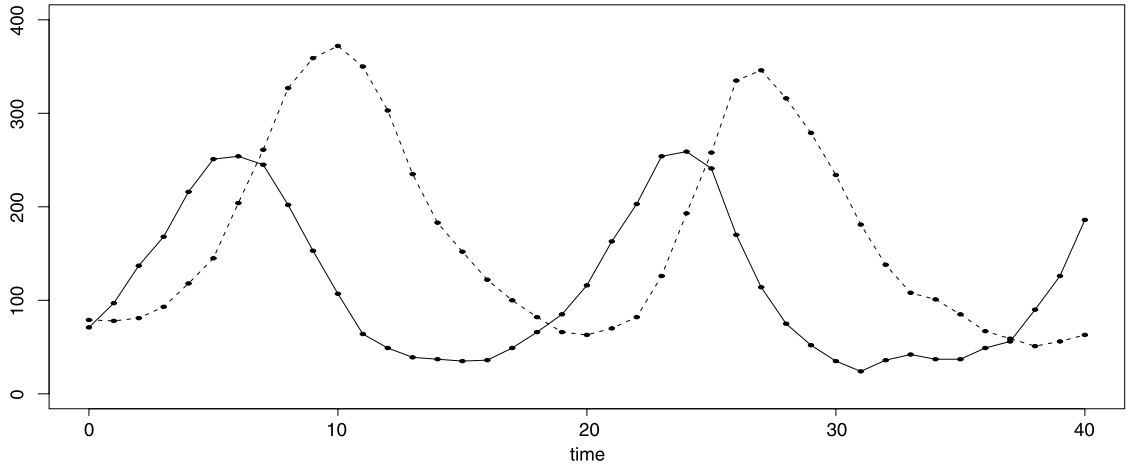

Fig. 1 Simulated observations (·) of prey and predator levels (*solid* and *dashed* lines resp.)

schemes used random walk updates on r_1 with tuning parameter $\lambda = 1 + r_1^2/200$, a choice found to induce good mixing. Here the approximate algorithm refers to simulation from the approximate process $\mathbb{Q}$. Interestingly, the approximate scheme typically demonstrated superior mixing properties to its exact counterpart. The *rj* algorithm was roughly five times faster than the *bu* algorithm (per iteration). Also the approximate scheme was typically over twenty times faster than the *bu* algorithm, with slightly superior mixing behaviour. For comparison purposes, we also include results obtained by using the algorithm based on a diffusion approximation (*d*) described in Golightly and Wilkinson (2008), using the same burn-in and thinning as the *bu* and *a* algorithms (and imputing nine latent observations between each pair of actual observations). Combining iteration speed with the convergence and mixing performances of the algorithms gives job times for the $rj : bu : d : a$ algorithms which scale roughly as $500 : 25 : 5 : 1$. These figures illustrate the efficiency of the block updating schemes and the additional benefit of approximating the underlying process.

Figure 2 contains a graphical analysis of a typical run and includes plots of the traces and densities of converged and thinned MCMC output for the rate constants $\theta = (\theta_1, \theta_2, \theta_3)'$. The marginal means and standard deviations of the posterior distribution are given in Table 1. They clearly show that the two exact algorithms produce results within Monte Carlo error and that there is little loss in inferential accuracy when either approximate algorithm is used. The table also shows results for larger datasets and, in particular, the trade-off between the numbers of intervals observed and the number of observations per interval. For example, the posterior standard deviations determined using single observations from 200 intervals are less than half those obtained from the same total number of observations but measured five times more frequently (except for *d*, which had inflated standard errors in the case of 200 intervals). There is also a suggestion from the tables that algorithm *a* outperforms *d* in the case of high frequency data, and that *d* outperforms *a* in the case of large amounts of low-frequency data. This makes intuitive sense, as the approximation used by *a* is likely to be very accurate in high-frequency scenarios. The table also illustrates the relatively small gain in parameter accuracy achieved by increasing the sampling rate per interval.

Not surprisingly, analyses of partially observed data scenarios present much more of a challenge for the MCMC algorithms due to the additional complication of having to mix over the uncertainty of unobserved predator levels. We illustrate the effect of data reduction on parameter uncertainty for three scenarios based on the data values used for the $n = 40$, $m = 1$ case in Table 1. These scenarios are partly motivated by results from fluorescence microscopy experiments which typically measure only a few biochemical species but sometimes are initialised from known conditions. We begin by assuming that predator levels are only available at the ends of the observation period, then only the initial level is seen and finally, that no predator level is observed. For the first two scenarios, algorithms *bu*, *a* and *d* required a burn-in of 10,000 iterations and a thin of 100 iterations to obtain a posterior sample with tolerable autocorrelations. In contrast, the reversible jump algorithm needed a far longer burn-in than before and mixing was significantly worse for the same level of thinning. The third scenario we have posed is particularly testing for the algorithm because no predator values are observed and we only assume a very diffuse prior on their initial number. Specifically we take $\pi\{y_2(0)\}$ to have an improper uniform distribution on $1, 2, \ldots$. Not surprisingly, this analysis required considerably more computational effort and the algorithms suffered very significant mixing problems. Indeed, CPU time constraints prevented us from obtaining satisfactory results for the *rj* algorithm in this case. Table 2 shows the summaries for a posterior sam-

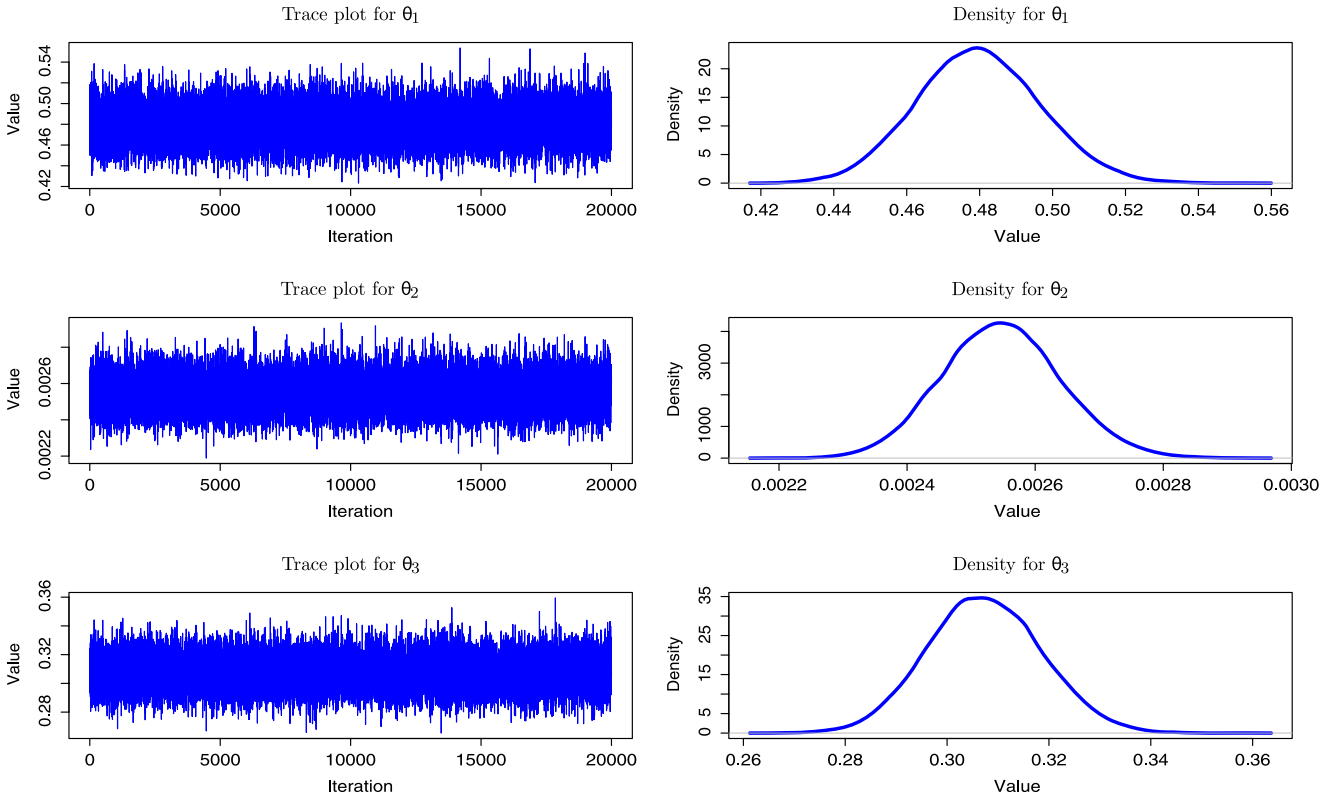

Fig. 2 Plots of traces and densities of converged and thinned MCMC output

Table 1 Posterior means (standard deviations) of rate constants using different algorithms and based on $T+1$ data points, where $T = n \times m$ and $n =$ number of unit intervals and $m =$ number of measurements within each interval. Data simulated using $\theta_1 = 0.5$, $\theta_2 = 0.0025$ and $\theta_3 = 0.3$

	Algorithm	θ_1	θ_2	θ_3
$n = 40, m = 1$	rj	0.4799 (0.0171)	0.00255 (0.000094)	0.3075 (0.0113)
	bu	0.4797 (0.0170)	0.00247 (0.000094)	0.3073 (0.0113)
	a	0.4840 (0.0170)	0.00307 (0.000095)	0.3104 (0.0113)
	d	0.4800 (0.0163)	0.00254 (0.000091)	0.3067 (0.0110)
$n = 200, m = 1$	rj	0.4997 (0.0063)	0.00250 (0.000030)	0.3036 (0.0038)
	bu	0.4998 (0.0062)	0.00251 (0.000031)	0.3036 (0.0038)
	a	0.5071 (0.0063)	0.00254 (0.000039)	0.3082 (0.0039)
	d	0.5028 (0.0104)	0.00252 (0.000051)	0.3058 (0.0065)
$n = 40, m = 5$	rj	0.4929 (0.0163)	0.00262 (0.000091)	0.3143 (0.0107)
	bu	0.4927 (0.0163)	0.00262 (0.000090)	0.3142 (0.0108)
	a	0.4929 (0.0161)	0.00262 (0.000090)	0.3142 (0.0107)
	d	0.4925 (0.0160)	0.00263 (0.000088)	0.3146 (0.0104)

ple of 20,000 values. The results for the reversible jump algorithm are within Monte Carlo error of those of exact block updates. The table shows that, apart from the case where no predator values are observed, the posterior means are not particularly sensitive to the reduction in information and that there is only a modest corresponding increase in parameter uncertainty. However, removing the final remaining predator value had a considerable effect on both the posterior mean and standard deviation. Clearly a wider range of parameter values are consistent with the observed prey levels. The table also includes summaries for the first and middle unobserved predator levels. The large standard deviations (particularly for the unobserved predator case) indicate the very wide range of predator values that are consistent with the observed data and explain the additional mixing problems incurred within the MCMC algorithms. Note that, for all data scenarios, the posterior distributions are consistent with the parameter values from which the data were simu-

Table 2 Posterior means (standard deviations) of rate constants and two unobserved predator levels using partially observed data. Data simulated using $\theta_1 = 0.5$, $\theta_2 = 0.0025$ and $\theta_3 = 0.3$

Data	Algorithm	θ_1	θ_2	θ_3	$y_2(1)$	$y_2(20)$
$y_1 \cup y_2$	*rj*	0.4799 (0.0171)	0.00255 (0.000094)	0.3075 (0.0113)	78 (0)	63 (0)
	bu	0.4797 (0.0170)	0.00247 (0.000094)	0.3073 (0.0113)	78 (0)	63 (0)
	a	0.4840 (0.0170)	0.00307 (0.000095)	0.3104 (0.0113)	78 (0)	63 (0)
	d	0.4800 (0.0163)	0.00254 (0.000091)	0.3067 (0.0110)	78 (0)	63 (0)
$y_1 \cup y_2(0) \cup y_2(40)$	*rj*	0.4715 (0.0236)	0.00242 (0.000133)	0.2831 (0.0172)	72.87 (5.24)	81.35 (12.52)
	bu	0.4715 (0.0237)	0.00242 (0.000134)	0.2827 (0.0172)	72.94 (5.32)	81.40 (12.56)
	a	0.4743 (0.0238)	0.00244 (0.000138)	0.2862 (0.0178)	72.62 (5.36)	80.24 (12.73)
	d	0.4762 (0.0238)	0.00240 (0.000133)	0.2801 (0.0170)	73.40 (5.37)	83.86 (12.36)
$y_1 \cup y_2(0)$	*rj*	0.4688 (0.0308)	0.00244 (0.000192)	0.2870 (0.0272)	72.47 (5.82)	80.25 (18.77)
	bu	0.4693 (0.0311)	0.00244 (0.000193)	0.2868 (0.0273)	72.50 (5.88)	80.27 (18.80)
	a	0.4898 (0.0331)	0.00236 (0.000190)	0.2728 (0.0268)	74.30 (5.87)	91.82 (21.30)
	d	0.4729 (0.0283)	0.00256 (0.000188)	0.3036 (0.0270)	71.91 (5.40)	75.47 (16.47)
y_1	*bu*	0.5718 (0.1088)	0.00201 (0.000417)	0.2357 (0.0504)	159.11 (97.84)	166.05 (102.62)
	a	0.7522 (0.1735)	0.00151 (0.000389)	0.1760 (0.0458)	236.98 (227.10)	369.75 (239.00)
	d	0.4318 (0.0608)	0.00291 (0.000417)	0.3471 (0.0593)	30.76 (39.50)	58.45 (37.49)

lated and the observed predator values in the "full" dataset and are therefore strongly suggestive that the model remains "identifiable" even in the case of no predator observations. Also note that the discrepancies between the algorithms in the final scenario are not within Monte Carlo error and are not due to convergence problems of the MCMC algorithms. Therefore, there is an indication (perhaps unsurprising) that the accuracy of both of the approximate algorithms declines as the proportion of missing data increases. Overall, the table highlights the benefit of using the approximate algorithm (*a*) in that little inferential power is lost and there is a much needed reduction in computational overhead.

6 Conclusions

Although there has been some previous work on inferring rate constants in deterministic networks, we believe this is the first systematic attempt to conduct rigorous "exact" inference for partially and discretely observed stochastic kinetic models. Inferences for the rate constants in the Lotka-Volterra model can be made using MCMC methods in various data-poor scenarios. The model parameters are identifiable even when no measurements are available on one of the species, though parameter uncertainty is considerably reduced when measurements are available on both species. Block updating algorithms are much more efficient than more naive reversible jump methods, and an algorithm based on an approximating process has been shown to perform particularly well. These algorithms readily extend to larger more complex networks but their computational efficiency relative to competing algorithms which exploit other approximations (such as a diffusion approximation) is the subject of on-going work.

Acknowledgements This work was funded by the BBSRC Bioinformatics Initiative (Grant #BIO14452). We would also like to thank Paul Blackwell for suggesting the idea of a Metropolis-Hastings correction to the approximating process, Andrew Golightly for assistance with computations relating to the diffusion approximation, Wan Ng for useful discussions and the referees and associate editor for their constructive comments on an earlier draft.

References

Abramowitz, M., Stegun, I.A. (eds.): Pocketbook of Mathematical Functions. Harri Deutsch, Frankfurt (1984), Chap. 9.6

Arkin, A., Ross, J., McAdams, H.H.: Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. Genetics **149**, 633–648 (1998)

Bahcall, O.G.: Single cell resolution in regulation of gene expression, Mol. Syst. Biol. (2005). doi:10.1038/msb4100020

Barenco, M., Tomescu, D., Brewer, D., Callard, R., Stark, J., Hubank, M.: Ranked prediction of p53 targets using hidden variable dynamic modeling. Genome Biol. **7**, R25 (2006)

Blackwell, P.G.: Bayesian inference for Markov processes with diffusion and discrete components. Biometrika **90**, 613–627 (2003)

Bower, J.M., Bolouri, H. (eds.): Computational Modeling of Genetic and Biochemical Networks. MIT Press, Cambridge (2001)

Boys, R.J., Giles, P.R.: Bayesian inference for SEIR epidemic models with time-inhomogeneous removal rates. J. Math. Biol. **55**, 223–247 (2007)

Brown, K.S., Sethna, J.P.: Statistical mechanical approaches to models with many poorly known parameters. Phys. Rev. E **68**, 021904 (2003)

Carter, C.K., Kohn, R.: Markov chain Monte Carlo in conditionally Gaussian state space models. Biometrika **83**, 589–601 (1996)
Finch, C.E., Kirkwood, T.B.L.: Chance, Development and Aging. Oxford University Press, New York (2000)
Gibson, G.J., Renshaw, E.: Estimating parameters in stochastic compartmental models using Markov chain methods. IMA J. Math. Appl. Med. Biol. **15**, 19–40 (1998)
Gilioli, G., Pasquali, S., Ruggeri, F.: Bayesian inference for functional response in a stochastic predator-prey system. IBull. Math. Biol. (2008). doi: 10.1007/s11538-007-9256-3
Gillespie, D.T.: Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. **81**, 2340–2361 (1977)
Golightly, A., Wilkinson, D.J.: Bayesian sequential inference for stochastic kinetic biochemical network models. J. Comput. Biol. **13**(3), 838–851 (2006)
Golightly, A., Wilkinson, D.J.: Bayesian inference for nonlinear multivariate diffusion models observed with error. Comput. Stat. Data Anal. **52**, 1674–1693 (2008)
Green, P.J.: Trans-dimensional Markov chain Monte Carlo. In: Green, P.J., Hjørt, N.L., Richardson, S. (eds.) Highly Structured Stochastic Systems, pp. 179–198. Oxford University Press, Oxford (2003)
Guptasarma, P.: Does replication-induced transcription regulate synthesis of the myriad low copy number proteins of *Escherichia coli*? BioEssays **17**, 987–997 (1995)
Johnson, N.L., Kotz, S.: Discrete distributions. In: Distributions in Statistics, vol. 1, p. 115. Wiley, New York (1969)
Jost, C., Arditi, R.: Identifying predator-prey processes from time-series. Theor. Pop. Biol. **57**, 325–337 (2000)
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. Mol. Cell Biol. **4**, 243–249 (2003)
Kitano, H. (ed.): Foundations of Systems Biology. MIT Press, Cambridge (2001)
Liebermeister, W., Klipp, E.: Biochemical networks with uncertain parameters. IEE Syst. Biol. **152**(3), 97–107 (2005)
Liechty, J.C., Roberts, G.O.: Markov chain Monte Carlo methods for switching diffusion models. Biometrika **88**, 299–315 (2001)
Lotka, A.J.: Elements of Physical Biology. Williams and Wilkins, Baltimore (1925)
McAdams, H.H., Arkin, A.: Stochastic mechanisms in gene expression. Proc. Natl. Acad. Sci. USA **94**, 814–819 (1997)
McAdams, H.H., Arkin, A.: It's a noisy business! Genetic regulation at the nanomolar scale. Trends Genet. **15**, 65–69 (1999)
Pepperkok, R., Ellenberg, J.: High-throughput fluorescence microscopy for systems biology. Nat. Rev. Mol. Cell Biol. **7**, 690–696 (2006)
Reinker, S., Altman, R.M., Timmer, J.: Parameter estimation in stochastic biochemical reactions. IEE Syst. Biol. **153**(4), 168–178 (2006)
Rempala, G.A., Ramos, K.S., Kalbfleisch, T.: A stochastic model of gene transcription: an application to L1 retrotransposition events. J. Theor. Biol. **242**(1), 101–116 (2006)
Renshaw, E.: Modelling Biological Populations in Space and Time. Cambridge University Press, Cambridge (1991)
Shen, H., Nelson, G., Nelson, D.E., Kennedy, S., Spiller, D.G., Griffiths, T., Paton, N., Oliver, S.G., White, M.R.H., Kell, D.B.: Automated tracking of gene expression profiles in individual cells and cell compartments. J. R. Soc. Interface **3**, 787–794 (2006)
Shephard, N., Pitt, M.K.: Likelihood analysis of non-Gaussian measurement time series. Biometrika **84**(3), 653–667 (1997)
Tian, T., Xu, S., Gao, J., Burrage, K.: Simulated maximum likelihood method for estimating kinetic rates in gene expression. Bioinformatics **23**, 84–91 (2007)
Volterra, V.: Fluctuations in the abundance of a species considered mathematically. Nature **118**, 558–560 (1926)
Wilkinson, D.J.: Stochastic Modelling for Systems Biology. Chapman & Hall/CRC, Boca Raton (2006)
Wilkinson, D.J., Yeung, S.K.H.: A sparse matrix approach to Bayesian computation in large linear models. Comput. Stat. Data Anal. **44**, 493–516 (2004)
Zheng, Q., Ross, J.: Comparison of deterministic and stochastic kinetics for nonlinear systems. J. Chem. Phys. **94**, 3644–3648 (1991)
Zlokarnik, G., Negulescu, P.A., Knapp, T.E., Mere, L., Burres, N., Feng, L., Whitney, M., Roemer, K., Tsien, R.Y.: Quantitation of transcription and clonal selection of single living cells with β-lactamase as reporter. Science **279**, 84–88 (1998)